Boston University

OpenBU **http://open.bu.edu**

Boston University Theses & Dissertations Boston University Theses & Dissertations

2015

Evaluating a peak height based method to determine the number of contributors in a DNA mixture and a study of DNA recovery using laser microdissection

https://hdl.handle.net/2144/16071

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

EVALUATING A PEAK HEIGHT BASED METHOD TO DETERMINE THE NUMBER OF CONTRIBUTORS IN A DNA MIXTURE AND A STUDY OF DNA RECOVERY USING LASER MICRODISSECTION.

by

GENESIS TEJADA

B.A., University of Vermont, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2015

© 2015 by
Genesis Tejada
All rights reserved

Approved by

First Reader

Catherine Grgicak, Ph.D.
Assistant Professor, Biomedical Forensic Sciences

Second Reader

Robin Cotton, Ph.D.
Associate Professor, Biomedical Forensic Sciences
Director of Biomedical Forensic Sciences

ACKNOWLEDGMENTS

I would like to express my eternal gratitude to my thesis advisor, Dr. Grgicak, for her continuous confidence in my abilities and her boundless patience throughout the research and writing process. I would also like to thank Robin Cotton and Elisse Coronado for serving on my thesis committee.

Lastly, I would like to thank my family and friends for their relentless support over the past two years.

EVALUATING A PEAK HEIGHT BASED METHOD TO DETERMINE THE NUMBER OF CONTRIBUTORS IN A DNA MIXTURE AND A STUDY OF DNA RECOVERY USING LASER MICRODISSECTION.

GENESIS TEJADA

ABSTRACT

Forensic laboratories process evidentiary samples found at crime scenes. These evidentiary samples may contain limited amounts of DNA and DNA from more than one individual. When data from these sample types are acquired, the data may be so complicated that efficient interpretation is prohibitive. Thus, there is a need for optimal DNA processing during testing. The forensic analysis of DNA involves; extraction, to purify the DNA; quantification, to determine the amount of DNA; amplification, to replicate DNA fragments of interest; separation and detection of the DNA fragments; and interpretation of genetic profiles. With each process there is error introduced, such as pipetting error and stochastic effects, which introduce the need to consider stutter artifacts and preferential amplification during interpretation. The result is difficult and complex profile analysis and interpretation. Since the work performed in a forensic laboratory involves human identification for purposes of resolving a legal argument, it is of importance to minimize error and obtain the most accurate and precise result for comparison. Thus, the first part of this study explores recovery rates for an alternative sample preparation method: laser microdissection. Laser

microdissection is proclaimed to be less time consuming compared to other methodologies, but more expensive; and many crime laboratories may deem the investment in such a system prohibitive. In this study, the majority of the DNA recovery rates of LMD samples were less than 10%, which is much lower than previously published recovery rates of 16-32%. Therefore, further studies optimizing the LMD method should be performed prior to the implementation in forensic casework.

The second part of this study assessed the possibility of using average peak height in lieu of template mass when evaluating the number of contributors. The evaluation was done using the probabilistic software NOC*It*, which is a computational tool created to aid analysts in determining the likely number of contributors (i.e. NOC) to a sample. The software currently uses target mass as a basis for its models and calculations. However, it has been suggested that target mass may be a suboptimal explanatory variable. In an effort to ascertain whether the use of target mass negatively impacts results, average peak height was analyzed to examine whether it could be used as the independent variable within NOC*It* and to discern whether the results were comparable to those obtained when target mass was the independent variable. Prior to use of average peak height in NOC*It*, an examination of the models that characterize alleles, noise, allele drop-out, reverse and forward stutter, and reverse and forward stutter drop-out was conducted. This study demonstrated concordance between the class of models which use target mass and the models that use average peak height,

indicating that average peak height could be used as the independent variable in NOC*It*.

Further, since NOC*It* models both noise and forward stutter, an evaluation of the effects of characterizing forward stutter as noise ensued. The results show that at low target masses the effects of characterizing forward stutter as noise are negligible; however an effect is seen at higher template mass. Thus, the model which does not characterize forward stutter as noise was preferred since it can be utilized for both low-template and high-template masses. Prior to the assessment of NOC*It* utilizing APH (average peak height), an examination of the software reproducibility using TM (target mass) was performed to ensure that multiple runs would estimate the same NOC with a similar *a posteriori* probability (APP). For this study, 128 samples were analyzed in quintuplicates; and, only 12 samples showed a different estimate for NOC. A repeatability graph plotting the range of APP against the median APP for quintuplicate samples illustrated that when the median APP was high (approximately 0.999), the range of APP was small (approximately 0.1). The data suggest that as the median APP decreases and the range of APP increases, there is less of a chance that the actual NOC can be determined with certainty. After repeatability results were determined, the use of the APH as the independent variable was analyzed. The accuracies were compared to those obtained when TM was used. If the max probability was taken as an indicator of NOC, the accuracy was 48 % and 47 % for APH and TM, respectively, illustrating that APH could be used by NOC*It* as a proxy for TM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

°C – degrees Celsius

µL – microliter

APH – average sum of peak heights

APP – *a posteriori* probability

AT – analytical threshold

bps- base pairs

DNA – deoxyribonucleic acid

dNTPs- deoxynucleotide triphosphates

EDTA- ethylenediaminetetraacetic acid

IPC – internal PCR control

K- kilo

LMD – laser microdissection

MAC- maximum allele count

mL – milliliter

MLE – maximum likelihood estimator

ng – nanograms

NOC – number of contributors

PCR – polymerase chain reaction

qPCR – quantitative polymerase chain reaction

rcf – relative centrifugal force

RFU – relative fluorescence units

RMSE – root mean square error

RNA – ribonucleic acid

STR- short tandem repeat

TE – Tris-EDTA

TM – target mass

1.0 Introduction

The discovery of short tandem repeats (STRs) by Jeffreys et al. (1) and the development of the polymerase chain reaction (PCR) by Mullis and Faloona (2) have revolutionized forensic DNA analysis. Due to the work of these pioneers, the ability to generate genetic information with high discriminating power from minute samples is now possible. Although there have been great improvements in analysis methods used to interpret forensic samples, there are still known limitations.

1.1 The Forensic DNA Laboratory Process

The DNA analysis process starts with the extraction of DNA from a substrate. The substrate could be any item suspected of containing DNA, such as clothing or a swab of a stain. Various methods of extraction can be performed: including the classic organic extraction (phenol/chloroform), Chelex extraction, or a solid-phase extraction (3). The extraction method utilized is typically chosen based on the amount of sample available; the condition of the sample; or whether there is the possibility of contamination or degradation. If there is a suspected mixture of sperm and non-sperm cells, a differential extraction can be utilized to separate the two fractions (3). After extraction, the amount of DNA in the extract is estimated through real-time PCR (also known as quantitative PCR or qPCR). The mechanism of the TaqMan based qPCR assay –a commonly utilized method in forensics–involves a probe, labeled with a quencher and reporter dye, which hybridizes to the DNA between the two primer binding sites on the target strand.

While the quencher is attached to the probe, it prevents detection of the fluorescent signal from the reporter dye. During the polymerization step of qPCR, the polymerase cleaves the probe, which releases the reporter and quencher dyes; and fluorescence from the reporter is detected (4). Therefore, as the number of PCR-cycles increases, and the amount of DNA product increases, the fluorescence increases. The cycle number at which the fluorescence reaches a specified intensity, usually set by the laboratory, is the cycle threshold. By comparing the cycle thresholds obtained from an unknown to those obtained from standards, the concentration of the unknown can be approximated (4). Obtaining the concentration of a sample is important since the result may be used to dictate which downstream forensic process to use. The qPCR result may also be used to decide whether the sample is amplified at all. Additionally, qPCR can provide other valuable information, such as the concentration of male DNA and/or the presence of potential inhibitors (4). The presence of inhibitors is assessed via the internal PCR control (IPC), which is a fragment of synthetic DNA that is co-amplified with the sample. An increase in the IPC cycle threshold indicates the presence of inhibitors. If there is no indication of inhibition, the concentration values obtained can be used to unambiguously target a mass of DNA for STR analysis. A forensic STR locus, usually 4-5bps in repeat unit length, is a region of DNA that repeats numerous times within the human genome (3). Different numbers of repeats at a location are designated as alleles. Forensic and human identification utilize these tetra- and penta- nucleotide repeats for

purposes of identification. Typically, forensic STR analysis is accomplished through multiplex PCR. In multiplex PCR, more than one region of DNA is amplified (5), through the addition of more than one primer set. Since STRs are hypervariable within a population (1), multiplex PCR can be used to differentiate between individuals by co-amplifying multiple loci and determining which alleles are present. After amplification, the STR fragments are separated and detected to identify allele types by capillary electrophoresis, which separates the alleles based on size. The result is a genetic profile, composed of peaks, which is used for comparison purposes. Analysis and interpretation of genetic profiles is accomplished by comparing the profile obtained from an item of evidence to the profile obtained from a known individual.

1.2 Variability in the DNA Process

During every step of the DNA process error is introduced, which can affect analysis. For example, when a differential extraction is performed, the complete separation of female and male DNA is not always possible and can interfere with later interpretation (3). Chelex extraction is faster and involves less transfer steps, but since there is no washing, purification, or concentration, it results in large volume extracts, which may be detrimental for samples containing low numbers of cells. Other shortcomings of Chelex extraction include ineffectively reducing the amount of contaminants in a sample and degradation of DNA after long-term storage and multiple freeze-thaw cycles. This was evidenced by the work of Greenspoon et al. who observed signal loss in 30% of samples that were

stored for approximately one year at -20°C, and subjected to several freeze thaws (6).

Solid-phase extraction involves the binding of DNA to substrates, such as a silica membrane (7). The benefit of a solid-phase extraction is the implementation of wash steps, which can help in the reduction of contaminants within a sample (3). Although the wash steps are beneficial for purifying the DNA sample, the DNA may not be bound efficiently to the silica layer resulting in DNA loss.

The error associated with extraction results in extracts with less than the desired 100% DNA yield. In fact, the average yield has been shown to be as low as 56% for low-template samples (8). Further, as modeled by Gill et al., there may be stochastic effects associated with the extraction process (9). Since forensic samples are usually of limited quantity, an extraction method that recovers a high amount of DNA is preferred.

Another source of error originates from the measurements associated with qPCR (10-12). Results obtained by Grgicak et al. illustrate the effect pipetting error has on qPCR (10). During absolute qPCR quantification, a dilution series is created by sequentially diluting a sample to obtain desired concentrations. The inability to pipette accurately or precisely can have a substantial effect on the standard curve used to estimate the amount of DNA in the extract. For example, Grgicak et al. showed significant impacts on standard curve generation with the use of an uncalibrated pipette, as well as the use of two different pipettes (10).

Koukoulas et al. observed differences in the standard curve generated by two standards provided by the same supplier; and, determined through spectrophotometry that the DNA standards were not the concentration specified by the manufacturer (11). Any error involved in the generation of the standard curve for qPCR greatly affects the downstream processing decisions and resultant STR signal. Since the curve is utilized to estimate the amount of DNA in the extract, it must result in measurements that are as accurate and precise as possible. In addition to measurement errors affiliated with absolute quantification using qPCR, Stenman and Orpana demonstrated that the accuracy of PCR is notably decreased when less DNA is available (12). Specifically, they estimated target copy number ranges using a Poisson distribution. They showed that when there were an estimated 8 target molecules, the 95% confidence interval ranged from 3-13 resulting in a deviation of 62%, but when there were an estimated 1,000 target molecules, the 95% confidence interval ranged from 938-1,062, resulting in a lower deviation of 6% (12). Similarly Timken et al. simulated starting copy number ratios of heterozygote alleles using a Poisson distribution and illustrated that the pre-PCR stochastic allelic sampling resulted in a higher peak height ratio dispersion compared to the dispersion obtained when only PCR sampling was simulated (13). For example, at an average template mass of 52.8 pg the range of ratios obtained when only PCR error was considered was approximately 0.85-0.95. In contrast, samples which were subjected to only pre-PCR sampling effects resulted in a range of 0.55-0.85 (13). Additionally, Timken

et al., like Stenman and Orpana, showed that dispersion increased with a decrease in template mass.

Though Timken et al. suggested most of the peak height ratio outcomes were due to sampling effects, a small but significant amount of error originated from the PCR. Timken et al. modeled this phenomenon with a Poisson distribution with constant PCR efficiency of 0.85 (13). However, PCR is usually modeled by the following equation:

$$N_t = N_0(E+1)^t \text{ (Eq. 1)}$$

where N_t is the number of amplicons at cycle t, N_0 is the initial number of target molecules, and E is the amplification efficiency (14). However, the amplification efficiency (E) not only varies at every cycle, but is also known to decrease with an increase in amplicon number (15). Further, it is unclear as to whether the variation in the PCR efficiency increases as E decreases. This may result in an accumulation of inconstant error throughout PCR cycling. Another explanation for the variable number of amplicons within an amplified product is preferential amplification. Preferential amplification is where one allele amplifies with a higher efficiency than the other in a heterozygous pair and can be explained by several mechanisms (16). Other sources of peak height imbalance within a locus are due to the formation of stutter product. Stutter is an artifact produced from the PCR reaction and is typically one repeat unit smaller than the allele (3). A proposed mechanism for the formation of stutter is the slipped strand mispairing model described by Walsh et al. (17). Generally, during replication, the template strand

loops out causing the new strand to differ by one repeat unit (17). Though stutter typically occurs in the reverse direction (n-1), it also occurs in the forward direction (n+1). The combination of inherent PCR error, preferential amplification, and stutter can result in complicated genetic signal.

Given the complexities associated with PCR processing, and the errors introduced at every step of the PCR process it is of interest to; 1) examine alternative methodologies for sample preparation and extraction designed to minimize peak height variability in forensic samples; and 2) study the forensic laboratory process in a systematic manner using well defined samples so a detailed understanding of the causes and effects of laboratory protocols can be better understood, modeled, and simulated. If the entire forensic process is well understood, then optimization of laboratory protocols and forensic interpretation may result. Therefore, the first part of this study assesses the DNA recovery using laser microdissection.

Laser microdissection (LMD) is an alternative method to sample preparation that utilizes a short laser pulse to isolate single cells by selecting cells and cutting through the slide membrane and tissue. The separated tissue together with the membrane is released from the remaining slide membrane by either being dropped directly in a microcentrifuge tube or collected with the adhesive polymer cap of the reaction tube (Figure 1). Recently, a significant amount of work utilizing laser microdissection has been published. Sanders et al. collected groups of 75, 150, and 300 sperm and epithelial cells by gravity into the

cap of a PCR tube which was placed below the stage; cells were stained and extracted with Qiagen QIAamp®, MicroLYSIS®, or Lyse-N-Go™ and amplified using AmpFlSTR® Profiler Plus Kit®. The study evaluated three different DNA isolation methods and showed that with 150 buccal cells, the QIAamp® detected 90% of the loci, Lyse-N-Go™ detected 74% and microLYSIS performed poorly (18). Also the QIAamp® extraction method produced RFU values 75% higher than Lyse-N-Go™ (18). In contrast, Vandewoestyne et al. employed an LMD method where 25, 50, 100, and 200 cells were catapulted into a collection device in triplicate. They extracted with DNA IQ™, PicoPure, or a short alkaline extraction method, quantified with Qubit® dsDNA HS assay, and amplified with AmpFlSTR® ProfilerPlus® (19). This study illustrated that the short alkaline method was ineffective for LMD samples because no profiles were recovered and the DNA yield of the extract was below the detection limit; DNA IQ™ resulted in DNA yields of 15-22% and was shown to work well as long as there was greater than 100 cells; and PicoPure successfully generated profiles with samples containing at least 25 cells, and resulted in DNA yields of 70-97% (19).

Not only can LMD be used to research the efficacy of a laboratory process, the ability to isolate single cells during casework processing would be advantageous to forensic DNA practitioners. For example, previous work has shown that the minor male component can be successfully separated from the female component using LMD (17, 20-23). Elliot et al. examined the difference between sperm cells separated using LMD and differential extraction by

comparing the resultant likelihood ratios and showed that LMD outperformed differential extraction for 15 of 16 samples (23). In contrast, Vandeweoestyne et al. examined female and male mixtures utilizing buccal samples and suspension fluorescence in situ hybridization (21). This study focused on using LMD to isolate groups of 5 and 10 male buccal cells and analyzed the profiles to determine that the recovery of a full DNA profile was possible with as little as 10 cells (21). Current work by Ballantyne et al. shows that LMD samples can also be used in conjunction with probabilistic interpretation tools (24). The process could be semi-automated and would be less time consuming. Hence, the efficient use of LMD would benefit crime laboratories.

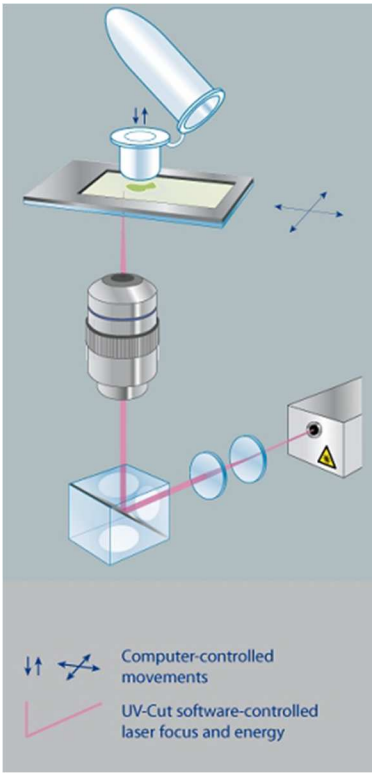

Figure 1: Diagram of the LMD process with an adhesive polymer cap [25]

1.3 Interpreting Genetic Profiles

Evidentiary samples collected from a crime scene could contain trace amounts of DNA, degraded DNA, or a mixture of contributors. These factors make the comparison of evidentiary genetic profiles against known genetic profiles (from a person of interest or a database) difficult. The development of a universally accepted methodology for the interpretation of low template DNA or mixture analysis has recently garnered much attention (24,26-36). Currently, an assumption regarding number of contributors (NOC) to a sample must be made (38) before comparison to a known profile. The predominant method to do this is the maximum allele count (MAC) method. Since an individual will have a maximum of two alleles at each locus (excluding very rare exceptions where there may be three alleles at a locus (39)) the minimum number of contributors can be determined by counting the number of alleles at every locus, taking the maximum number of alleles observed, dividing by two, and rounding up. This approach only provides information on the minimum NOC, which may not be equivalent to the actual NOC. Further, it may not take into account the quantitative data that are present with utilizing peak heights, and effects from stutter or allele drop-out (27). Another factor not taken into account by MAC is allele sharing. Allele sharing causes the underestimation of the number of contributors within a mixture as shown by Paoletti et al. who determined a 3.18%-3.39% mischaracterization rate of three-person mixtures to two-person mixtures. Similarly, Buckleton et al. demonstrated a 3.3%-6.2%

mischaracterization rate (27-28). Additionally, as the true NOC increased, both Paoletti et al. and Buckleton et al. showed the mischaracterization rates sharply increase. Thus, when two or more contributors are indicated by the allele counts in the mixture, it is suggested the report stipulates that it is the minimum NOC that has been determined, rather than the actual NOC (27).

Recently, there has been a surge in the development of tools to aid in complex mixture interpretation and comparisons (24,26,29-37). Examples of these tools include Lab Retriever (29,32), LRmix (32, 35-37), STRmix™ (31-33), and TrueAllele® (24, 30, 34). Because these methods rely on the user to provide information about the NOC, methods which assess the likelihood that a certain NOC describe the evidence have been developed. For example, a maximum likelihood estimator (MLE) method to determine the NOC has been proposed. Haned et al. compared the MLE method to the MAC method for 1,000 simulated mixtures ranging in 2-5 NOC and observed a decrease in accuracy with increasing NOC for both MAC and MLE; but the decrease was more prominent for the MAC method (35). For example, when examining simulated 2-person mixtures, MAC correctly identified the NOC 100% of the time and only correctly identified the actual NOC in 5-person mixtures 2% of the time. Contrastingly, MLE correctly identified 99% of the 2-person mixtures and 64% of the 5-person mixtures (35). Though MLE outperformed MAC, the study was conducted on 1000 simulated mixtures. Therefore, other interferences such as stutter and allelic drop-out were not considered. Further the MLE method does not account

for peak height information. Therefore, it was of interest to evaluate a method that utilizes the peak height information. As a result, the second part of this study involves the probabilistic DNA software NOC*It*−a computational tool that estimates the *a posteriori* probability (APP) that a certain NOC gave rise to the signal. NOC*It* incorporates both quantitative data (the heights of the peaks) and qualitative data (i.e. allele frequencies) to make its estimation (26). The tool has been validated with target mass (TM) as the independent variable. Work conducted by Swaminathan et al. has shown that NOC*It* results in accuracy rates greater than 80% when the target mass of the DNA sample is greater than 0.05ng (26). However, since TM is estimated via qPCR and thus incorporates a substantial level of uncertainty in the metric, it was of interest to examine whether an alternative explanatory variable could be used as a basis for the models used in NOC*It*. In response to this, the average sum of the peak heights (APH) is used to assess whether a) APH can be used as the independent variable within NOC*It* and b) APH results in accuracy rates that are at least consistent with those obtained when target mass is utilized.

2. Study of DNA Recovery using Laser Microdissection

2.1 Methods

Sample Preparation

Liquid saliva samples were used to prepare the slide for microdissection. The procedure was as follows: 6 µL of saliva was washed in 500 µL of TE Buffer (10mM Tris pH 8.0, 0.1mM EDTA). To pellet the cells, the sample was centrifuged at maximum angular velocity for 5 minutes and all but 10 µL of the supernatant was removed. The pellet was then reconstituted in the remaining supernatant and the full volume was pipetted onto an mmi MembraneSlide (Molecular Machine & Industries AG, Glattbrugg, Switzerland). The slide was incubated at 75°C for 20 minutes. After incubation, the slide was inverted and placed on a clean, glass slide to prevent contamination.

LMD Process

Cells were microdissected using an mmi CellCut Plus® laser microdissection system (Molecular Machine & Industries AG, Glattbrugg, Switzerland). The cells were collected on either transparent or diffuser mmi IsolationCaps® (Molecular Machine & Industries AG, Glattbrugg, Switzerland), which are 0.2 mL reaction tubes with a special adhesion cap. A total of 6 replicates of 20 cells and 4 replicates of 100 and 200 cells were collected. To demonstrate that the cells adhered to the cap, images of the caps were acquired before and after the LMD process (Figure 2).

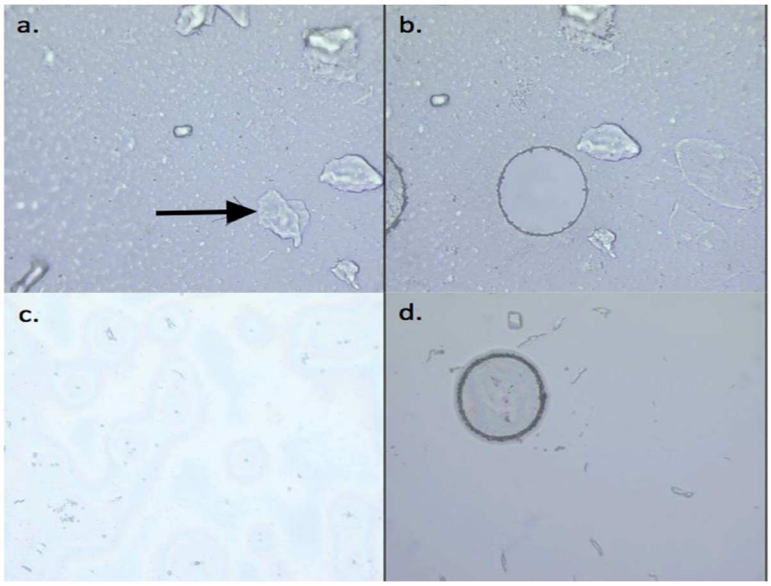

Figure 2: **a.** Slide prior to cutting, with cell of interest indicated with black arrow **b.** Slide after cutting of the cell **c.** Adhesive cap prior to pick up **d.** Adhesive cap after pick up (All images were photographed at 40x magnification)

QIAGEN Extraction

The QIAamp DNA Investigator Kit® extraction was performed using the manufacturer's recommended protocol for *Isolation of Total DNA from Surface and Buccal Swabs* (40). In accordance with the manufacturer's recommendations for small amounts of sample, carrier RNA was added to Buffer AL (40). The lysis buffer (Buffer AL, Qiagen, Germantown, MD) contained 1 µL of carrier RNA for every 400 µL of buffer. Modifications to the protocol were as follows: 1) The cap-insert was removed from the reaction tube with tweezers and placed into a 2 mL microcentrifuge tube 2) After lysis with Buffer AL and Proteinase K the samples were spun in order to remove liquid from the cap and to mimic the processes typically encountered in forensic laboratories. Specifically, the insert was removed from the microcentrifuge cap and placed into a perforated basket. The

basket was inserted into a 2 mL microcentrifuge tube, and the sample was centrifuged for one minute at maximum angular velocity. The basket was then removed, and the insert was discarded. 3) Samples were eluted in 20 µL of Buffer ATE, then the sample was allowed to evaporate until a final volume of 13 µL was reached. If the volume was less than 13 µL, deionized water was added to bring up the volume to the desired 13 µL. An extraction positive of whole blood or saliva was prepared for each experiment and showed expected results (data not shown).

Quantitative PCR

Samples were quantified with the Quantifiler® Duo DNA Quantification Kit (Applied Biosystems, Foster City, CA) using the ABI 7500 Sequence Detector (Applied Biosystems, Foster City, CA). Amplifications were performed in a 25 µL final reaction volume using the manufacturer's recommended laboratory protocol (4) and a single validated external calibration curve (10). The data obtained from qPCR were used to calculate DNA recovery for each sample. DNA percent recovery was calculated by multiplying the DNA concentration results from qPCR and the sample volume measured after extraction as described with the following equation:

$$\boldsymbol{DNA\ \%\ Recovery} = \frac{\left[\boldsymbol{DNA\ Concentration}\left(\frac{\boldsymbol{ng}}{\boldsymbol{\mu L}}\right)\right]\ \boldsymbol{x}\ [\boldsymbol{Eluted\ Volume\ (\mu L)}]}{[\boldsymbol{Number\ of\ Cells\ in\ Sample}]\ \boldsymbol{x}\ [\frac{\boldsymbol{0.0063ng}}{\boldsymbol{cell}}]}\ \boldsymbol{x}\ \mathbf{100}\% \quad \text{(Eq. 2)}$$

where 0.0063 ng is the estimated mass of DNA in one cell (41-42).

2.1.1 *Method Validation*

Slide Preparation with KPIC Staining

To prepare the slide for microdissection, 3 µL of saliva was washed in 500 µL of TE Buffer. The sample was centrifuged and all but 10 µL of the supernatant was removed. The pellet was then reconstituted in the remaining supernatant and pipetted onto an mmi MembraneSlide. The slide was incubated at 75°C for 60 minutes. After incubation, the slide was stained with KPIC (Kernechtrot-Picroindigocarmine) stain (Serological Research Institution, Richmond, CA). Briefly, 2 drops of nuclear fast red were placed on the slide and then the slide was incubated in a moisture chamber for 15 minutes. After gently rinsing with deionized water, 2 drops of picroindigocarmine were placed on the slide. After 30 seconds, the slide was rinsed with 100% ethanol and left to air dry.

QIAGEN Extraction with Amicon® Filtration

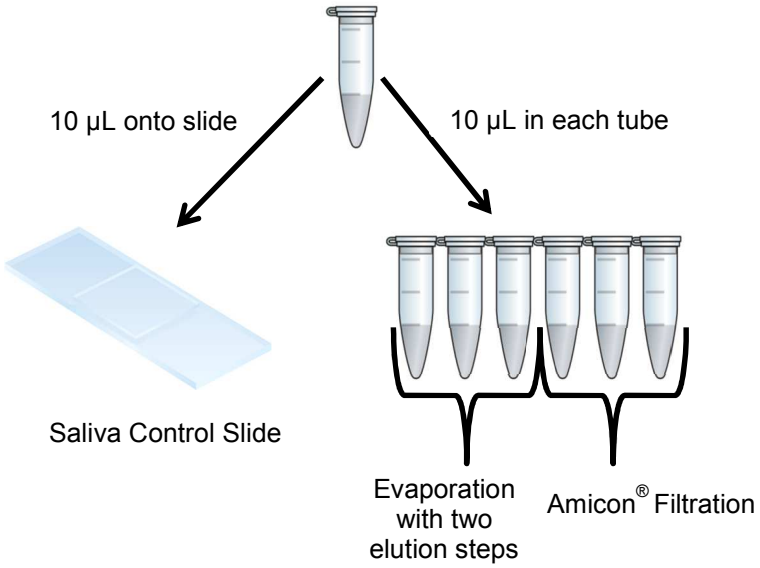

Figure 3: Diagram of samples concentrated by evaporation or Amicon® filtration

Samples were prepared as described previously and processed as depicted in Figure 3. The sample for the saliva control slide was prepared using the staining method. The number of cells on the slide was counted and used as a reference for the number of cells in the liquid saliva samples. The remaining samples were extracted using the QIAGEN extraction described above. One set of samples was subjected to a double elution with 20 µL of Buffer ATE in two separate tubes. These samples were then evaporated as previously described in order to reach a final extract volume of 13 µL. The other set of samples was subjected to an elution with 55 µL of Buffer ATE and concentrated using Amicon® Ultra-0.5 50K filtration units (EMD Millipore, Billerica, MA) following the manual's protocol (42).

ZyGEM Extraction

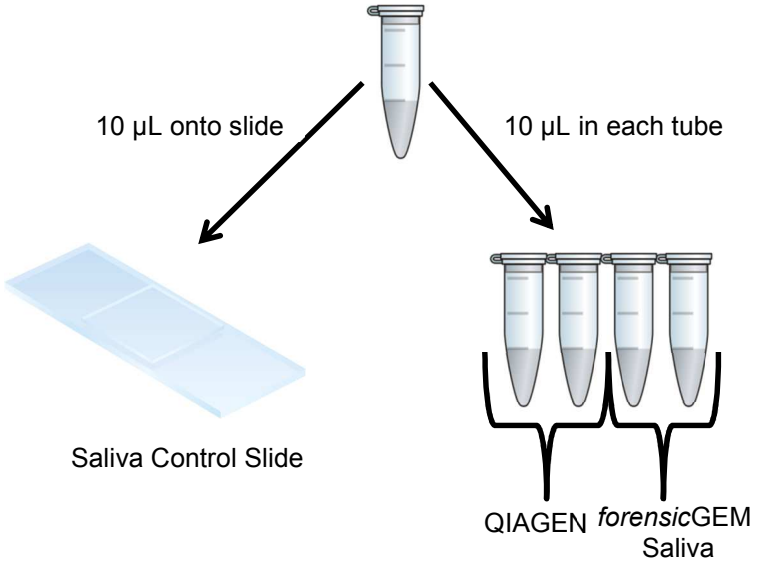

Figure 4: Diagram of samples extracted by QIAGEN or *forensic*GEM Saliva

Samples were prepared as described previously and processed as portrayed in Figure 4. The sample for the saliva control slide was prepared using the staining method. The number of cells on the slide was counted and used as a reference for the number of cells in the liquid saliva samples. One set of samples was subjected to the QIAGEN extraction previously described. The other set of samples was extracted using *forensic*GEM® Saliva (ZyGEM, Lane Hamilton, New Zealand) following the manufacturer's recommended protocol (44).

LMD Process

In the previously described method, individual cells were microdissected until the desired number was reached. For the following samples, groups of cells were dissected in order to obtain of the desired number of cells on the cap (Figure 5). Once dissected, DNA extraction and quantification ensued as described in Section 2.1.

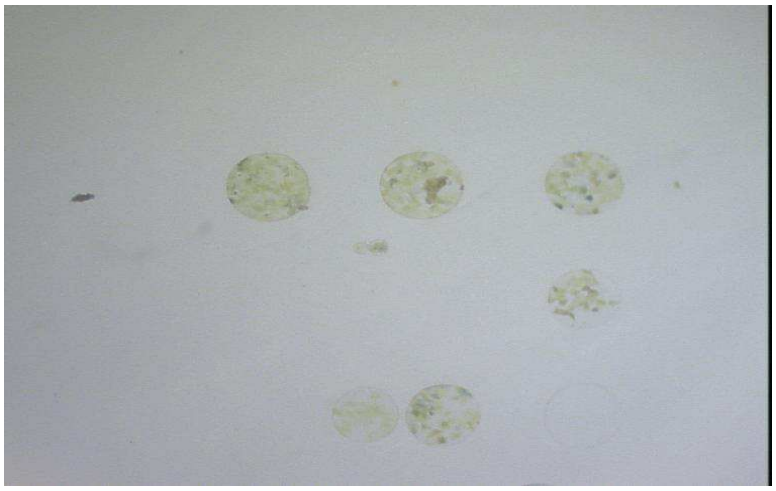

Figure 5: Adhesive cap after pick up of groups of cells (Image photographed at 4x magnification)

2.2 Results and Discussion

2.2.1 LMD Method Optimization and Evaluation of Sources of DNA Loss

The first set of results obtained from non-stained LMD samples resulted in three samples containing no detectable DNA and one sample exhibiting DNA recovery of 60% (Table 1). Since the majority of the samples did not yield detectable quantities of DNA, a validation of the LMD process commenced.

2.2.1.1 Effect of KPIC Stain on DNA Recovery

The first step in the process to be evaluated was the sample preparation step. Though the number of cells on the cap was counted and the presence of cells on the cap was confirmed through visual inspection, the identification of useable cells was susceptible to human error. Nucleated epithelial cells were targeted; however, without staining it is possible to mistakenly collect anucleated cells or other debris of the same size and shape as the epithelial cells targeted. Presumably, it is advantageous to utilize cells which have not undergone the staining process as it is desirable to minimize extraneous downstream PCR inefficiencies caused by staining reagents. To evaluate the effects of staining on DNA results, the KPIC stain, which stains the nucleus of a cell red and the cytoplasm green was used during this study because it is commonly employed in forensic laboratories, and it provides a strong contrast between the cytoplasm and the nucleus of the cell. The stained samples resulted in recovery rates ranging from 3-37% (Table 1). Though the quantity of DNA recovered was highly variable−regardless of whether a stain was utilized−the stained samples resulted

in detectable DNA quantities for all replicates, indicating some minor improvement in performance. This is likely due to the enhanced ability to visualize the cells, rather than an increase in PCR efficiency, since previous studies have indicated that KPIC may induce a decrease in PCR efficiency. Specifically, Sanders et al. examined LMD samples (groups of 75, 150, and 300 sperm and epithelial cells) stained with hematoxylin/eosin (H&E), nuclear fast red/picroindigocarmine (CTS), methyl green (MG), Wright's stain (WRT), or acridine orange (AO) (18). Data obtained from profiles was evaluated and exhibited a decrease in RFU peak heights for stained cells compared to the unstained control. H&E samples exhibited RFU values that were 62% of those determine with the unstained control, while CTS/KPIC stained samples exhibited RFU values that were 43% of those determined with the unstained control. The AO stain produced no amplified product in any samples tested. The MG and WRT were not further analyzed due to poor morphological distinction between sperm and epithelial cells (18). Consequently, for purposes of this research, all ensuing LMD samples were stained with KPIC.

Though there was DNA recovered from the stained LMD cells, the majority of the samples recovered less than 10% of the expected quantity of DNA. This is slightly less than previously reported DNA recovery rates, which range from approximately 16%-32% using various extraction techniques (45-46). Specifically Colussi et al. examined the average DNA yield of a swab quarter containing 1/5 dilution of semen, extracted with DNA IQ using a differential lysis protocol and

quantified with RT-PCR (45). The reference sample used for comparison was the liquid semen dilution and the study showed that the average DNA yield for the 3 dilution samples ranged from 16% to 20% (45). In contrast, Kishore et al. obtained a slightly higher DNA yield of 32% by examining the extraction efficiency of the BioRobot® EZ1 compared to organic extraction utilizing whole blood and semen extracts that were quantified using the nuTh01 assay for RT-PCR (46).

Table 1: Comparison of DNA recovery for unstained and stained LMD samples

Number of LMD Cells	Cells Stained	DNA Recovery (%)
20	No	ND
20	No	61
20	No	ND
20	No	ND
20	Yes	7
20	Yes	37
100	Yes	3
100	Yes	5

ND = no detectable DNA

Before continuing with additional LMD experiments, the concentration process was also tested to evaluate effects of the concentration methodology.

2.2.1.2 Effects of concentration and elution volume methodology on DNA recovery

After extraction, forensic laboratories may concentrate and purify DNA samples prior to further analysis (47). A commonly employed method is one that

utilizes centrifugal filters, such as Amicon® filters (47). During this method, the sample is centrifuged and a filter allows salts and detergents to pass through while retaining the DNA. In this work, an evaporation step was used rather than Amicon® filtration in order to alleviate concerns related to DNA loss on the filter. For example, Iacona showed that the average DNA loss after filtration at 14,000 rcf for 50 minutes was 46% (48). Another factor to consider during method development was that of elution volume. During extraction, a larger elution volume has been shown maximize DNA yield, but dilute the final sample extract (8). Considering that the primary goal of this work was the extraction of individual microdissected cells, studies that characterize which−if any−laboratory processes unnecessarily decrease yield were needed. Thus, three replicates of whole liquid saliva (3 µL) were extracted and concentrated using both the evaporation and filtration methods to determine whether one method was preferable. The samples that were evaporated were eluted using a 20 µL volume, while the samples concentrated by filtration used a 55 µL elution volume. The samples that were to undergo evaporation were also subjected to a second elution to determine whether the DNA was effectively desorbing from the silica membrane. Figure 6 shows percent recovery (Equation 2) for the three replicate samples subjected to filtration and evaporation. The samples that were evaporated and eluted twice resulted in DNA recovery of approximately 10% on the first elution. The second elution showed no detectable DNA for the majority of the samples; with one sample containing 9% DNA recovery (data not shown),

suggesting that one elution removes the majority of detectable DNA from the silica column. Further, two of the samples concentrated by filtration gave a DNA recovery of approximately 10% and one gave a DNA recovery of 20% (Figure 6). These data indicate that samples concentrated by the two methods resulted in similar recoveries; therefore, the evaporation process was deemed a suitable method for concentration of DNA samples and was henceforth utilized as the method to concentrate the extracts. However, since the DNA recovery was approximately 10% lower than the range reported in the literature, the effect of the QIAGEN extraction method on the DNA recovery was studied.

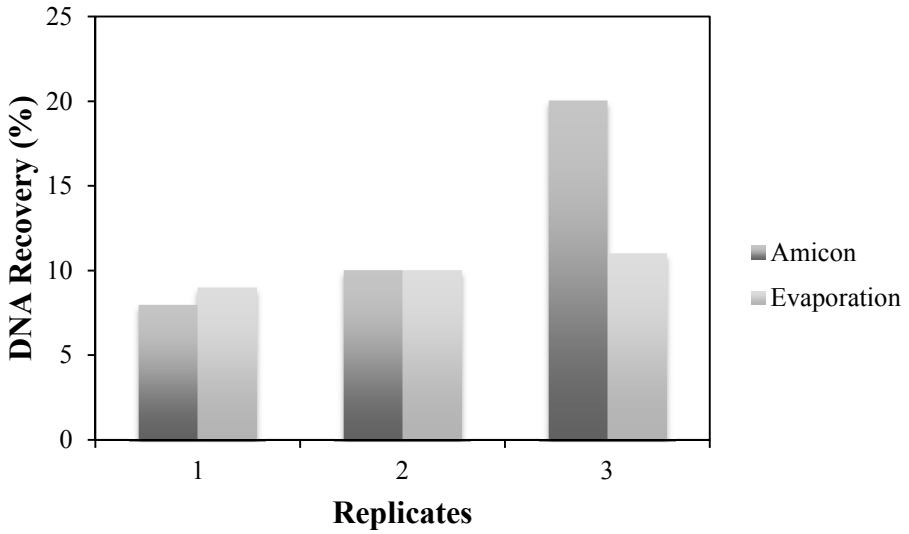

Figure 6: Comparison of DNA recovery for samples concentrated with evaporation or Amicon® filtration

2.2.1.3 Effects of extraction method on DNA recovery

The QIAGEN extraction method is a solid phase extraction technique that uses a silica column to adsorb DNA (39). The DNA is physisorbed onto the

membrane under low pH, high alcohol conditions. The presence of a chaotrope also promotes physisorption of DNA by modifying the entropy of the system. Desorption occurs under neutral pH, low alcohol conditions, and when no chaotrope is present (7). Prior to desorption of DNA from the membrane, the column/sample is washed multiple times to remove any contaminants. It is hypothesized that these wash steps, as well as the imperfect adsorption and desorption efficiencies, may contribute to DNA loss. Previous studies which focused on characterizing QIAGEN recovery efficiencies showed that an average 7-44% of DNA can be lost during silica-based DNA extraction (8, 49). Specifically, Phillips examined whole and swabbed blood and saliva using QIAamp and showed and average DNA loss of 12-24% (8). Similarly, Mygind et al. examined atherosclerotic tissue using five extractions methods, including the silica based method DNeasy Tissue kit, and showed that the average DNA loss was 26% (49). One way to ammeliorate DNA loss is by using single-tube extraction methodologies. One such commercially available method is the *forensic*GEM® Saliva protocol, which allows for a single tube extraction by utilizing a proprietary thermostable proteinase to lyse the cell and destroy nucleases without inhibiting downstream amplification (44). Befittingly, there is only one transfer step during the *forensic*GEM® Saliva extraction process, which would presumably decrease DNA loss. Figure 7 shows percent recovery for two replicate whole saliva samples extracted with QIAGEN and *forensic*GEM®. The DNA recovery using QIAGEN ranged from approximately 26-47%, while DNA

recovery using *forensic*GEM® ranged from approximately 78-97%, indicating a substantial and consistent improvement in DNA recovery when a one step extraction methodology is employed. The results are consistent with those obtained by Kaeser who showed that *forensic*GEM® samples had higher mean peak heights compared to QIAGEN samples (50). For example, at 0.0625 ng *forensic*GEM® resulted in a mean peak height of 146 RFU, while QIAGEN resulted in a mean peak height of 101 RFU (50). Evaluation of the data presented in Figures 7 and 8 suggests that the QIAGEN extraction causes a significant loss in DNA levels. However, this does not explain the variation and near 0 % recovery rates shown in Table 1.

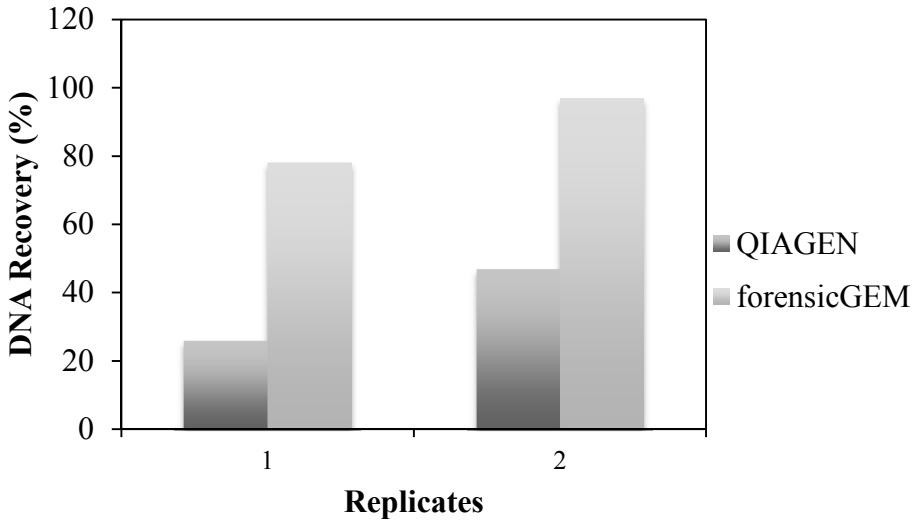

Figure 7: Comparison of DNA recovery for samples extracted with QIAGEN or *forensic*GEM Saliva

Therefore, to test whether the LMD process itself impacts recovery, LMD samples were processed along with a liquid saliva control. As previously

observed, the LMD samples had a DNA recovery ranging from 0% (no detectable DNA) to 9%, while the liquid saliva control had a DNA recovery ranging from 9% to 38% (Figure 8), which is consistent with the data shown in Figure 7 and is within the range of previously reported DNA recovery levels. However, according to Figure 8 LMD samples produce a lower than expected DNA recovery, suggesting that a factor within the LMD process itself may be hindering the recovery.

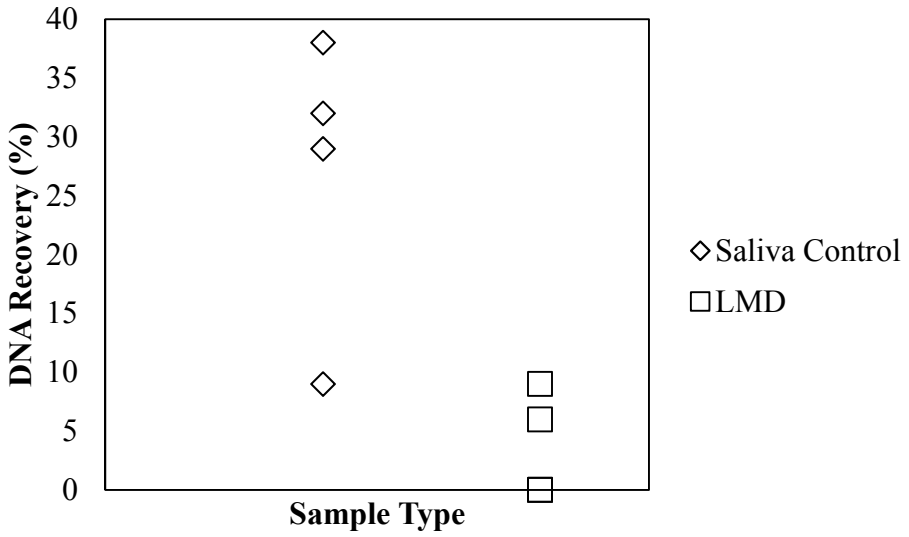

Figure 8: Comparison of DNA recovery between LMD samples and the liquid saliva controls using QIAGEN

Although the results obtained in this study do not suggest that LMD should be used for forensic purposes, previous studies, using LMD for forensic samples, have produced favorable results (20, 22-24). Sanders et al. showed that LMD samples (75, 150, and 300 cells) extracted with QIAGEN QIAamp DNA Investigator Kit® gave a DNA recovery of approximately 25-45% (18).

Vandeweostyne et al. showed that LMD samples (25, 50, 100, and 200 cells) extracted with DNA IQ™ (Promega Corporation, Madison, WI) and PicoPure (Arcturus, Mountain View, CA) resulted in DNA recoveries ranging from 15% to 95% (19). Ballantyne et al. demonstrated that LMD samples (1-5, 10, and 20 cells) extracted with forensicGEM® gave average partial or full profiles with recovery rates ranging from 69-99% (24). One plausible reason for the inconsistencies between studies is the type of LMD method used. The LMD technique used in this study requires an adhesive cap attached to a collection tube, located above the slide, to which the cells adhere. The LMD technique used by Sanders et al. and Ballantyne et al. does not require a cap and the collection tube is located below the slide (18,24). The cells are microdissected and fall into the collection tube. Vandeweostyne et al. used laser pressure catapulting where the cells were launched into a collection device (19). In both instances the cells are placed in the tube and do not need to be retrieved from a surface. Thus, the LMD process utilized in this study may need further optimization or evaluation before it can be recommended for purposes of human identification.

3. Evaluating a peak height based method to determine the number of contributors

3.1 Methods

The samples used for analysis in this study were processed as described by Swaminathan et al. (26). Single source samples were extracted using standard organic procedures, quantified using the Quantifiler® Duo DNA Quantification Kit, amplified using the AmpFlSTR® Identifiler® Plus Amplification Kit (Applied Biosystems, Foster City, CA) targeting 0.0078 ng-0.5 ng of DNA, separated using a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA), and analyzed using GeneMapper IDX v1.1.1 (Life Technologies, Inc., Foster City, CA) at an RFU of 1.

3.1.1 Modeling of Variables using Average Peak Height

The variables studied were allele signal, baseline noise, allele dropout, reverse and forward stutter ratio, and reverse and forward stutter dropout using the calibration samples described in Swaminathan et al. (26). For each variable, the mean and standard deviation were computed using both the target mass (TM) and average peak heights (APH). Plots were created with Igor Pro v.6.36 (WaveMetrics, Inc., Lake Oswego, OR). The models were based on the ones NOC*lt* uses when target mass is the independent variable (Table 2) (26). The TM ranged from 0.0078 ng-0.5 ng, while APH ranged from 0 RFU-5500 RFU. To calculate the APH of a sample, off-ladder (OL) peaks were filtered. The OL peaks are peaks that do not fall into a pre-defined allelic bin (5). The average baseline

noise height for a given color channel (Table 3) −determined by evaluating electropherogram signal obtained by negatives−was subtracted from the heights of all peaks. All heights within a locus were then summed, and averaged across loci. The averages were subsequently rounded using the following rules: if the APH was below 500 RFU, the APH was rounded to the nearest 50. For example, an APH of 135.3 RFU would be rounded to 150 RFU. If the APH was greater than 500 RFU, the number was rounded to the nearest 500. As an example, an APH of 3273.6 RFU would be would be rounded to 3500 RFU. Any APH greater than 3500 RFU (for example 7290 RFU) was rounded to 5500 RFU.

Table 2: Model assumptions for variables [26]

Variable	Model Description	Distribution of the Variable
Mean of true peak heights (μ_t)	Line with a positive slope	$\mu_t = ax+b$, where x is the DNA mass in the true peak
Standard deviation of true peak heights (σ_t)	Line with a positive slope	$\sigma_t = ax+b$, where x is the DNA mass in the true peak
Mean of noise peak heights (μ_n)	Line with a positive slope	$\mu_n = ax+b$, where x is the DNA mass that the sample was amplified with
Standard deviation of noise peak heights (σ_n)	Line with a positive slope	$\sigma_n = ax+b$, where x is the DNA mass that the sample was amplified with
Drop-out rate of alleles (α)	Exponentially decreasing curve	$\alpha = ae^{bx}$, where x is the DNA mass from the contributor with the allele
Mean of Stutter Ratios (μ_s)	Exponentially decreasing curve	$\mu_s = ae^{bx}+c$, where x is the DNA mass in the parent allele that gives rise to stutter
Standard Deviation of Stutter Ratios (σ_s)	Exponentially decreasing curve	$\sigma_s = ae^{bx}+c$, where x is the DNA mass in the parent allele that gives rise to stutter
Rate of non-occurrence of stutter (β)	Exponentially decreasing curve	$\beta = ae^{bx}$, where x is the DNA mass in the parent allele that gives rise to stutter

Table 3: Noise correction based on the dye channel

Dye Channel	Average Noise Height (RFU)
Blue	3
Green	4
Yellow	7
Red	7

3.1.2 Effect of forward stutter on noise distribution

The data obtained with samples generated using methods described in Section 3.1.1 were filtered to create two data sets: In both data sets the true peaks, reverse stutter peaks, and OL peaks were removed. However, in one data set, the forward stutter peaks were also filtered. Histograms were prepared in Igor Pro v.6.36 using a manual binning method with bin-centered x-values, such that each bin was 1 RFU in width and the bins extended from 1 RFU up to the maximum observed noise peak (51). Since Bregu et al. illustrated that the lognormal fit was a better representation of baseline noise distribution, the histograms were fitted with a lognormal curve (52). A KS-test was performed to determine whether the two data sets originate from the same distribution using a freely available web service, which can be accessed at: *http://www.physics.csbsju.edu/stats/KS-test.n.plot_form.html*. For data sets exceeding 1,024 points, the following web service was utilized: *http://scistatcalc.blogspot.com/2013/11/kolmogorov-smirnov-test-calculator.html*. The KS-test was utilized since it is non-parametric and makes no assumptions on the distribution of the data.

3.1.3 Reproducibility Study

A subset of 1-, 2-, 3-, 4-, 5-person experimental samples (n=128) previously tested by Swaminathan et al. (26) were chosen and run in quintuplicate on NOC*lt* v.1.5. The calibration file consisted of 654 single source samples with known genotypes, ranging from 0.0078 ng-0.5 ng of DNA. The

allele frequencies used were those of the US Caucasian population provided with the AmpFlSTR® Identifiler® Plus Amplification Kit (5). The number of times that NOC*lt* estimated the same NOC or a different NOC within sample runs was counted and graphed in Microsoft® Excel 2011 (Microsoft Corporation, Redmond, WA). A repeatability chart was created by plotting the range of the *a posteriori* probability (APP) values against the median APP value, for the quintuplicate runs. The range is defined as the difference between the maximum APP value and minimum APP value.

3.1.4 Testing of NOCIt using Average Peak Height

A subset of the aforementioned experimental samples (n=64) was tested again, but this time APH was used instead of TM. The average sum of peak heights for the test samples were calculated and binned using the method described in Section 3.1.1. It should be noted the exact samples utilized to create the calibration file using TM were employed for the calibration file using APH. Further, the same allele frequencies used in the reproducibility study were used in this section. The accuracy of NOC*lt* was calculated using two methods: maximum probability and 1% probability (26). For maximum probability, if the true NOC in a sample gave the highest probability, then the output was classified as correct. For the 1% probability metric, if the true NOC had an APP output of at least 1%, then it was taken to indicate a reasonable probability that the sample originated from the actual NOC; therefore it was deemed a correct result. For example, Table 4 illustrates the APP outputs for a sample where the actual NOC

was 5. When the max probability metric was utilized, the sample would be considered incorrect (max NOC = 4), but when using the 1% probability the sample would be deemed correct (Table 4). Swaminathan et al. explains the rationale for choosing 1% as the threshold for an accurate result (26). In short, since NOC*It* gives a range of where the NOC is most likely to lie, a NOC with a probability of at least 1% indicates a non-negligible or significant probability that sample originated from *n* contributors and should not be disregarded.

Table 4: The APP outputs for a representative sample where the actual NOC was 5

	n=0	*n*=1	*n*=2	*n*=3	*n*=4	*n*=5
APP	0	0	3.48 E-113	4.48 E-04	0.934	0.065

3.2 Results and Discussion

3.2.1 Modeling of Variables using Average Peak Height

The relationship between target mass and peak height is expected to be linear if the PCR efficiency remains constant and close to 1 throughout all cycles, as per (Eq. 3):

$$(\mathrm{RFU}) = \kappa N_0 (E+1)^C \quad \text{(Eq.3)}$$

where κ is the proportionality constant, N_0 is the initial number of target molecules, E is the amplification efficiency, C is the last cycle of PCR, and RFU is representative of signal intensity. Since Equation 3 suggests a direct proportionality between target mass and peak height, and assuming the

proportion of molecules electrokinetically injected does not change with amplicon concentration, APH may be taken to be a good proxy for DNA mass. If the APH is taken as a good representation, then the models utilized to describe allele and stutter drop-out, allele and noise height, and stutter ratios should remain the same; albeit with differing parameters. To confirm this assertion, the average and standard deviation of the allele peak heights were plotted using TM or APH. Since, target mass and peak height are proportional, a positive linear trend is expected according to Equation 3. Figure 9 displays the equation and root mean square error (RMSE) for the linear regression using the D18S51 and D5S818 loci. The RMSE indicates the quality of the fit to the data: the lower the RMSE the better the fit. For both loci, APH and TM show that a linear trend is supported (Figure 9). Table 5 exhibits the RMSE values for all loci. When examining the average of the allele peak heights, APH has smaller RMSE values than TM for 12 of 15 loci. However, when examining the standard deviation, TM has smaller RMSE values than APH for 14 of 15 loci (Table 5). Further, the standard deviation seems to be approximately 1/2 of the average peak height when TM is used (Figure 9). This is higher than the standard deviation when APH is utilized; here the standard deviation is approximately 1/3 of the average peak height.

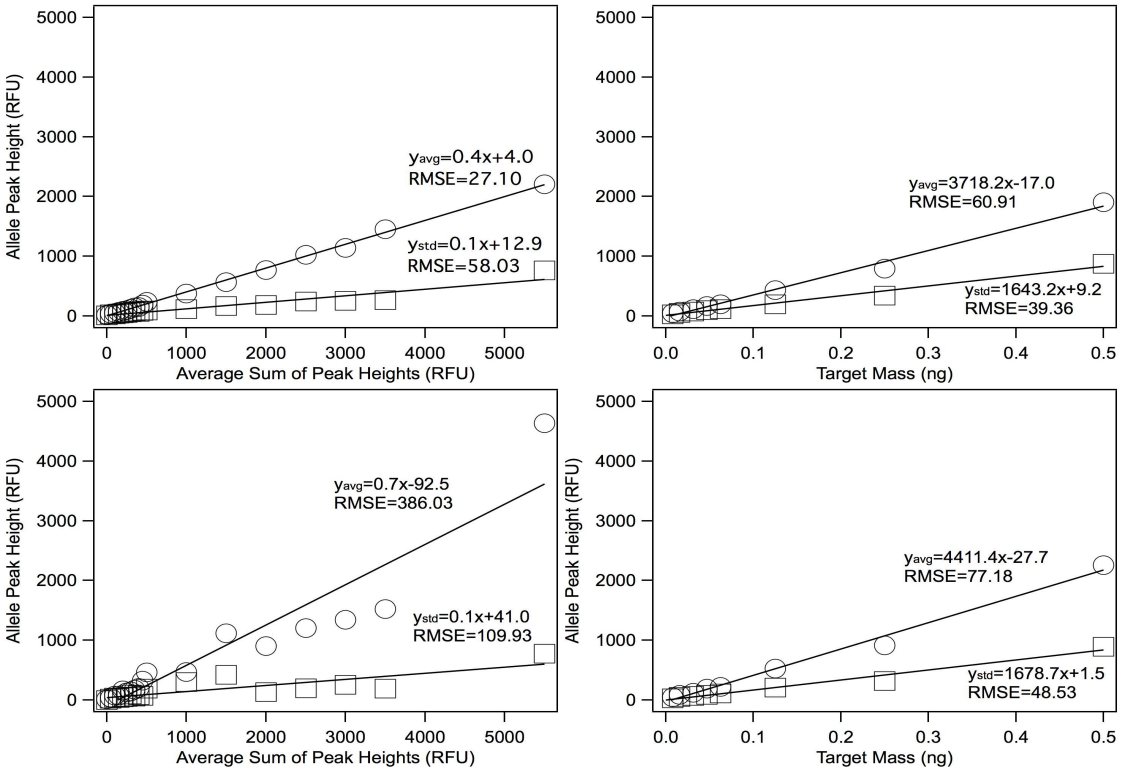

Figure 9: The average (○) and standard deviation (□) for allele peak height plotted against target mass and average sum of peak heights for the D18S51 (top) and D5S818 (bottom) loci

Table 5: Root mean square error for the average and standard deviation for allele peak height plotted against target mass and average sum of peak heights for all loci

	Average		Standard Deviation	
Locus	APH	TM	APH	TM
CSFIPO	192.42	49.11	102.26	52.81
D13S317	72.48	150.60	92.69	72.95
D16539	52.23	95.25	71.06	54.58
D18S51	27.10	60.91	58.03	39.36
D19S433	236.49	90.65	97.80	30.56
D21S11	28.97	72.87	57.55	31.79
D2S1338	36.16	64.58	84.30	47.59
D3S1358	33.83	96.51	47.60	26.09
D5S818	386.03	77.18	109.93	48.53
D7S820	27.69	61.67	54.94	29.40
D8S1179	81.53	100.12	47.98	53.55
FGA	40.19	65.42	41.88	33.22
TH01	73.22	145.83	79.01	46.23
TPOX	130.42	136.50	105.46	113.88
vWA	33.63	94.58	59.87	32.13

The next variable examined was baseline noise. Gilder et al. examined positive, negative, or reagent blank controls and suggested that baseline noise from the capillary electrophoresis instrument was more variable between runs than within runs (53). Bregu et al. also examined noise and observed that as the amount of input DNA mass increases, so does the baseline noise (52). Wellner took this assessment further and suggested that a sigmoidal model may be used to describe noise (54). The sigmoidal model would be indicative of an increase of noise with target mass until a large amount is reached at which point the noise levels off and becomes constant (54). Since the data used for this study did not include high-template samples, the sigmoidal model could not be evaluated nor

confirmed. Figure 10 shows noise peak height versus TM and APH, displayed with the equation and RMSE values for the D7S820 and vWA loci. For both loci, APH and TM show the same trend, where the average and the standard deviation of the noise seem to remain constant until large quantities (i.e., greater than 0.25 ng) of DNA are reached (Figure 10). Table 6 shows the RMSE values when the data is approximated as linear. The RMSE values for average and standard deviation for TM are smaller than APH for 11 of the 15 loci (Table 6). Furthermore, the standard deviation is generally 1/2 of noise for both APH and TM (Figure 10).

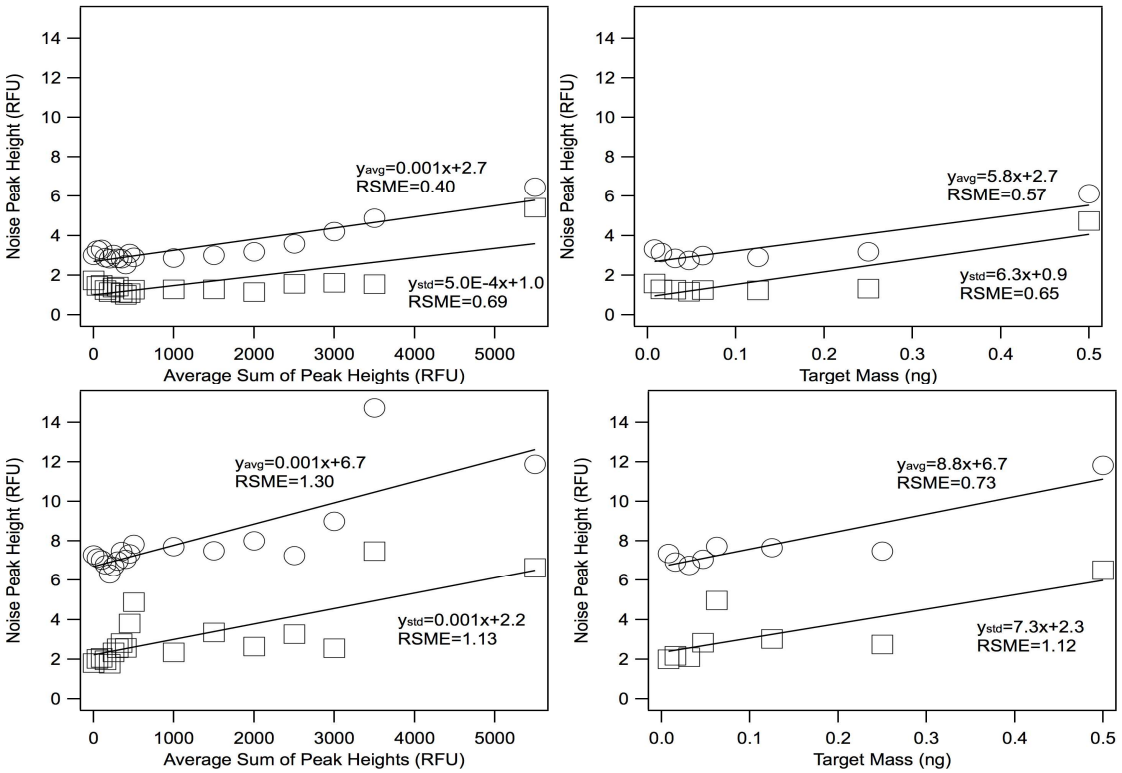

Figure 10: The average (○) and standard deviation (□) for noise peak height plotted against target mass and average sum of peak heights for the D7S820 (top) and vWA (bottom) loci

Table 6: Root mean square error for the average and standard deviation for noise peak height plotted against target mass and average sum of peak heights for all loci

	Average		Standard Deviation	
Locus	APH	TM	APH	TM
CSFIPO	1.09	0.82	0.62	0.49
D13S317	1.05	0.95	2.00	2.21
D16539	0.62	0.45	0.67	0.33
D18S51	0.82	0.38	0.74	0.38
D19S433	1.04	1.20	1.34	1.50
D21S11	0.98	0.15	4.86	1.52
D2S1338	0.96	0.70	0.81	0.60
D3S1358	0.74	0.92	0.78	0.43
D5S818	0.63	0.54	0.62	0.58
D7S820	0.40	0.57	0.69	0.65
D8S1179	0.55	0.34	1.17	0.90
FGA	0.67	0.33	0.64	0.26
TH01	0.97	0.98	2.38	2.37
TPOX	0.86	0.45	0.60	0.25
vWA	1.30	0.73	1.13	1.12

The next variable examined was allele drop-out. Allele drop-out is an extreme form of heterozygous imbalance (55). Heterozygote imbalance has been shown to increase as the amount of DNA decreases (55). An allele is said to have “dropped out” when it is below the analytical threshold (AT). The AT, usually set by a laboratory through validation, is the signal value used to reliably differentiate a signal from baseline noise. It has been demonstrated that as the amount of DNA increases, the frequency of allele drop-out decreases because a larger number of alleles will be above the AT (55-56). Swaminathan et al. chose the exponential model to describe the probability of allele drop-out, instead of the

logistic fit previously suggested because according to their dataset there was no discernible difference between the two models (57). Similarly, Norsworthy et al. examined five different methods to characterize allele drop-out and determined that both logistic and exponential regression are appropriate methods (58). Thus, the exponential fit was used to model the frequency of allele drop-out when APH is used as a proxy for TM. Figure 11 shows the frequency of allele drop-out using APH and TM with an exponential fit for the FGA and D7S820 loci. For APH, the exponential model seems to fit the data well with RMSE values ranging from 7.56E-03 to 5.27E-02, which are similar to the RMSE values obtained from TM that range from 1.79E-03 to 1.86E-02 (Table 7).

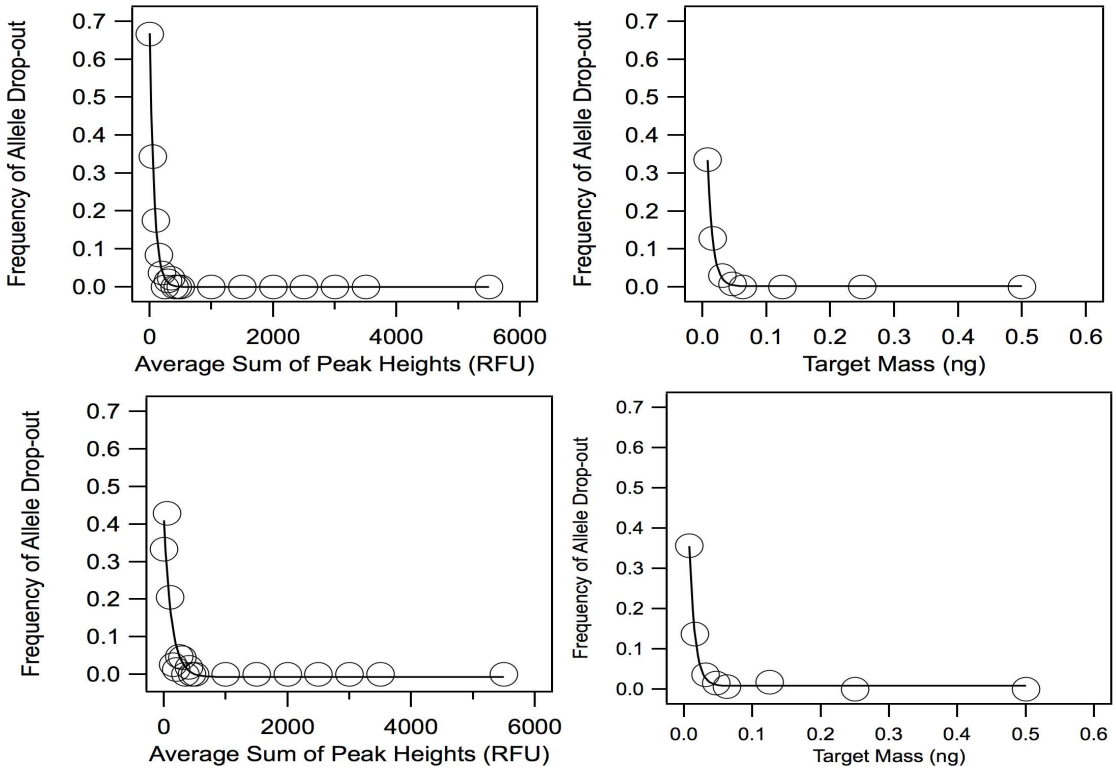

Figure 11: The frequency of allele drop-out (O) plotted against target mass and average sum of peak heights for the FGA (top) and D7S820 (bottom) loci

Table 7: Root mean square error for the frequency of allele drop-out plotted against target mass and average sum of peak heights for all loci

Locus	APH	TM
CSFIPO	3.04E-02	1.22E-02
D13S317	2.50E-02	1.20E-02
D16539	4.25E-02	1.79E-03
D18S51	3.09E-02	1.01E-02
D19S433	2.61E-02	1.86E-02
D21S11	2.32E-02	4.63E-03
D2S1338	3.19E-02	1.01E-02
D3S1358	1.15E-02	2.32E-03
D5S818	1.05E-02	1.21E-02
D7S820	5.27E-02	7.39E-03
D8S1179	1.79E-02	5.89E-03
FGA	7.56E-03	4.50E-03
TH01	1.01E-02	3.49E-03
TPOX	8.44E-03	7.22E-03
vWA	1.18E-02	1.93E-03

As previously discussed, stutter is an artifact produced during PCR. There have been numerous studies which characterize both forward and reverse stutter; though reverse stutter has been widely studied (59-63). One way to evaluate stutter peak intensity is to use the stutter ratio, described by the following equations:

$$RS_R = \frac{X_{a-1}}{X_a} \text{ (Eq. 4)}$$

$$FS_R = \frac{X_{a+1}}{X_a} \text{ (Eq.5)}$$

where RS_R is the reverse stutter ratio and FS_R is the forward stutter ratio, X_a is the height of the parent allele, and X_{a-1} and X_{a+1} are the heights of the reverse and forward stutter, respectively. Figures 12 and 13 are plots of the average and

standard deviation for reverse and forward stutter ratios against TM and APH fitted to exponential curves. These figures demonstrate that both the average and standard deviation of the stutter ratios decrease as the TM and APH increase. Additionally, it seems that the standard deviation is approximately equal to the average of both the reverse and forward stutter ratios for APH and TM. It is also observed that when the average and standard deviation for the reverse stutter ratios become constant, the value is comparable for both APH and TM. For example, when examining the FGA locus (top two graphs of Figure 12) the average stutter ratio remains constant at approximately 0.06 for APH, while for TM the average remains constant at 0.07; the standard deviations for both APH and TM remain constant at 0.03. This trend is also generally observed for forward stutter ratio (Figure 13). Tables 8 and 9 illustrate the small RMSE values for the average and standard deviations, indicating that the exponential model is an acceptable model for characterizing stutter ratios. In a similar fashion, Seo et al. also evaluated stutter artifacts using 25 pg, 50 pg, and 200 pg samples amplified using the Identifiler® kit with 32 cycles and observed a decrease in average stutter ratios as template mass increased (64). Gibb et al. studied the characterization of forward stutter by plotting the percentage height of parent peak against the parent peak height (63). Though the study only examined 1 ng samples amplified with SGM Plus ™, a comparable trend is observed where the percentage height of parent peak decreases as the parent peak height increases (63). In contrast, Bright et al. examined four amplification

kits with the following DNA targets: 1 ng for NGM Select™ and Identifiler™, 0.5 ng for PowerPlex®, and 0.4 ng for GlobalFiler ™, and observed little or no effect of parent allele height on forward stutter ratio for all four multiplexes (60). Further, Bright et al. examined the use of two models; gamma and exponential. It was determined that the gamma distribution placed 7.6% of forward stutters above the detection threshold (30 RFU), while the exponential distribution would place 0.2% of forward stutters above the detection threshold. Comparing these two values with data obtained from samples amplified with the NGM Select™ kit, which had 10.3% of its observations showing detectable forward stutter, the gamma distribution was chosen as the representative model. Since reverse stutter is approximately 15% the height of the main allele, it is expected that stutter will not be detected 100% of the time at low-template levels. Further, if stutter is a consequence of strand slippage as proposed by Walsh et al. (16) then PCR sampling effects may result in a preponderance of the non-detection of stutter (or stutter drop-out). Hence, characterization and subsequent use of the rates of stutter detection would be prudent during low-template mixture interpretation. Thus, another variable, the frequency of reverse and forward stutter drop-out, was examined for various target masses. Figures 14 and 15 are plots of the frequency of reverse and forward stutter drop-out against TM and APH fitted to exponential curves. The results show that as the TM and APH increase, the frequency of stutter drop-out, decreases for both reverse and forward stutter (Figures 14, 15) which indicates that as TM and APH increases

the observed stutter increases. These data show that reverse stutter is consistently observed at a TM greater than 0.1 ng or an APH of greater than 1000 RFU. However reverse stutter drop-out rapidly increases, when APH is less than 450 RFU or TM is less than 0.1 ng (Figure 14). Tables 10 and 12 show the frequency of reverse stutter drop-out for APH and TM; the general trend is observed for all loci. In contrast, forward stutter is not always observed, even at higher TM's or APH's (Figure 15). In fact, there is always a probability of forward stutter drop-out with an APH as high as 3500 RFU and TM of 0.5 ng (Tables 11 and 13). This coincides with Gill et al., who examines serial dilutions of semen ranging from 0.8 pg to 1 ng, amplified with SGM Plus ™ to determine the presence of reverse stutter in a profile (59). By calculating stutter as a proportion of success−measured as the percent of alleles observed across all loci−Gill et al. demonstrates that the presence of stutter increases with template mass (59).

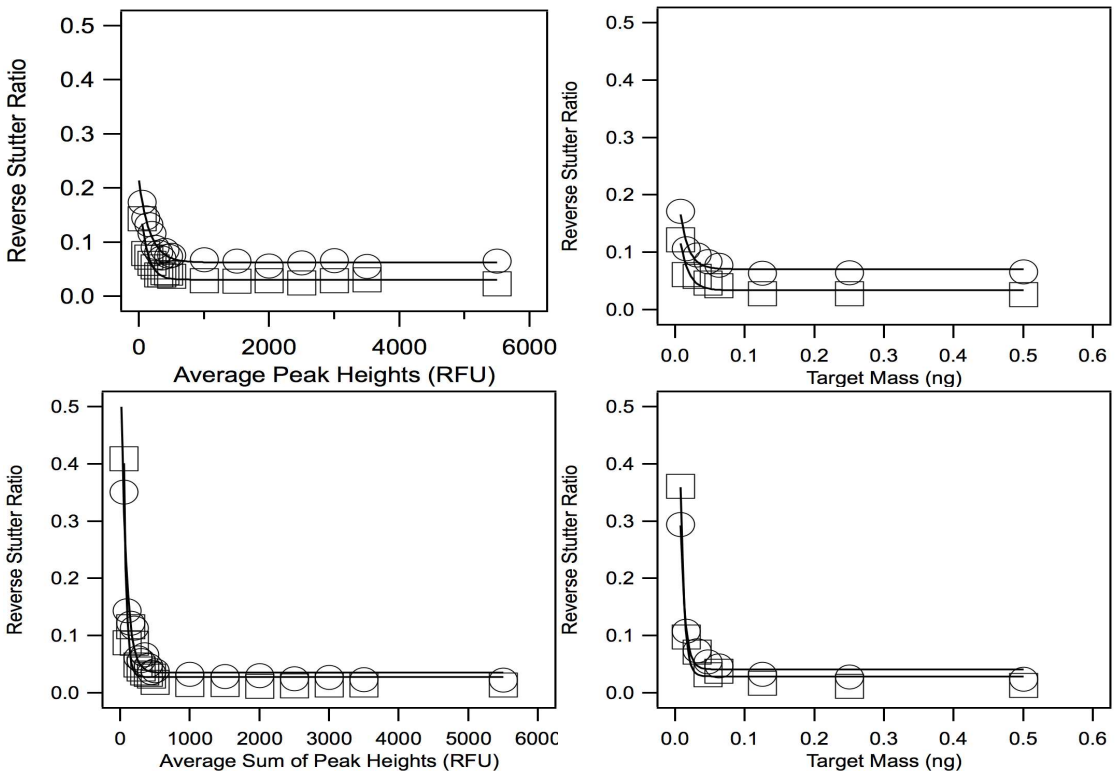

Figure 12: The average (○) and standard deviation (□) of reverse stutter ratio plotted against target mass and average sum of peak heights for the FGA (top) and TPOX (bottom) loci

Table 8: Root mean square error for the average and standard deviation of reverse stutter ratio plotted against target mass and average sum of peak heights for all loci

	Average		Standard Deviation	
Locus	APH	TM	APH	TM
CSFIPO	1.31E-02	9.56E-03	3.35E-02	3.86E-02
D13S317	1.02E-02	3.33E-03	5.04E-02	1.18E-03
D16539	7.87E-03	3.57E-03	1.44E-02	8.50E-03
D18S51	1.14E-02	9.18E-03	1.12E-02	2.09E-02
D19S433	6.83E-03	7.33E-03	6.30E-03	3.46E-03
D21S11	9.03E-03	7.01E-03	2.76E-02	2.84E-02
D2S1338	1.77E-02	8.51E-03	7.04E-02	4.49E-02
D3S1358	1.40E-02	4.60E-03	3.94E-03	3.48E-03
D5S818	1.04E-02	8.39E-03	1.22E-02	1.11E-02
D7S820	1.48E-02	3.53E-03	3.26E-02	1.16E-02
D8S1179	9.91E-03	5.23E-03	9.01E-03	1.22E-02
FGA	5.96E-03	1.01E-02	6.76E-03	1.07E-02
TH01	9.98E-03	6.35E-03	4.02E-02	1.07E-02
TPOX	1.81E-02	1.57E-02	2.52E-02	1.96E-02
vWA	1.05E-02	3.00E-03	1.92E-02	7.19E-03

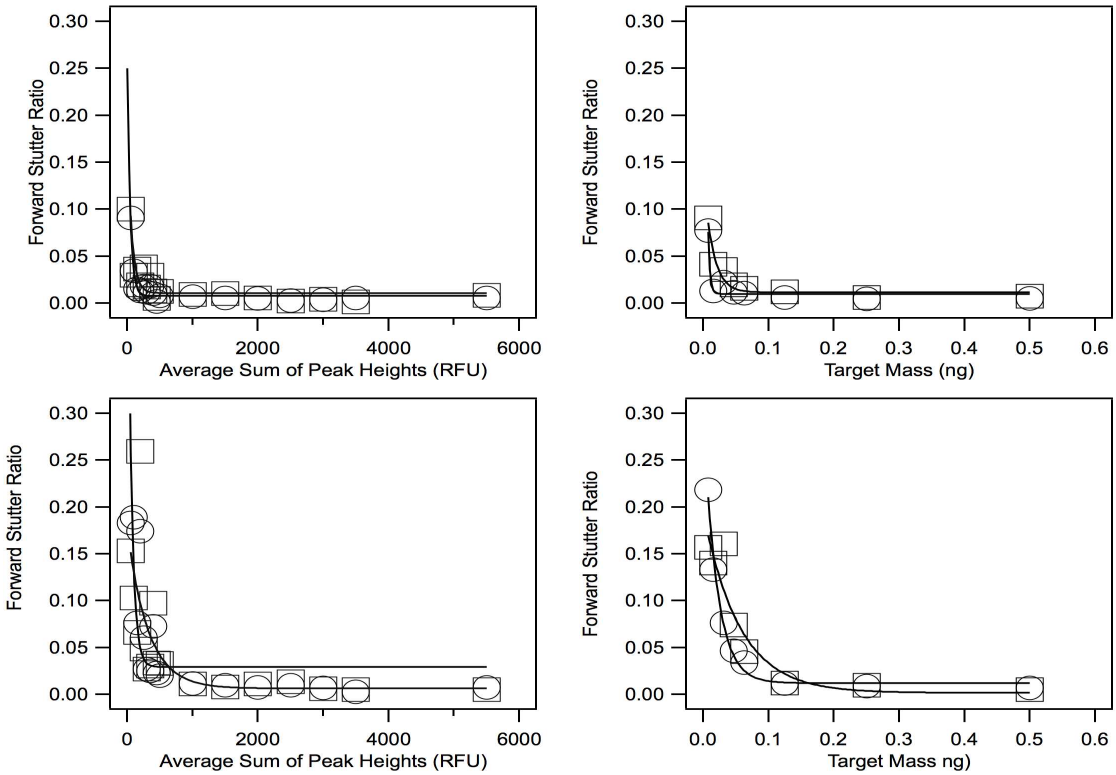

Figure 13: The average (○) and standard deviation (□) of forward stutter ratio plotted against target mass and average sum of peak heights for the representative D7S820 (top) and FGA (bottom) loci

Table 9: Root mean square error for the average and standard deviation of forward stutter ratio plotted against target mass and average sum of peak heights for all loci

	Average		Standard Deviation	
Locus	APH	TM	APH	TM
CSFIPO	8.14E-03	4.75E-03	7.06E-03	4.55E-03
D13S317	5.82E-03	2.55E-03	1.01E-02	3.08E-03
D16539	2.39E-02	2.02E-02	6.41E-02	6.82E-02
D18S51	1.64E-02	1.46E-02	1.79E-02	2.18E-02
D19S433	2.25E-02	8.29E-03	7.36E-02	1.85E-02
D21S11	1.86E-02	2.66E-02	7.58E-02	1.09E-01
D2S1338	1.09E-02	9.29E-03	1.48E-02	1.77E-02
D3S1358	4.63E-03	4.48E-03	7.38E-03	6.49E-03
D5S818	1.67E-02	1.52E-02	3.10E-02	2.91E-02
D7S820	4.44E-03	6.05E-03	1.03E-02	7.77E-03
D8S1179	6.69E-03	3.32E-03	1.11E-02	5.00E-03
FGA	4.57E-02	8.22E-03	4.73E-02	2.31E-02
TH01	5.10E-03	8.80E-03	8.57E-03	3.21E-02
TPOX	4.38E-02	1.41E-02	3.81E-02	1.88E-02
vWA	3.47E-02	1.71E-02	5.97E-02	2.13E-02

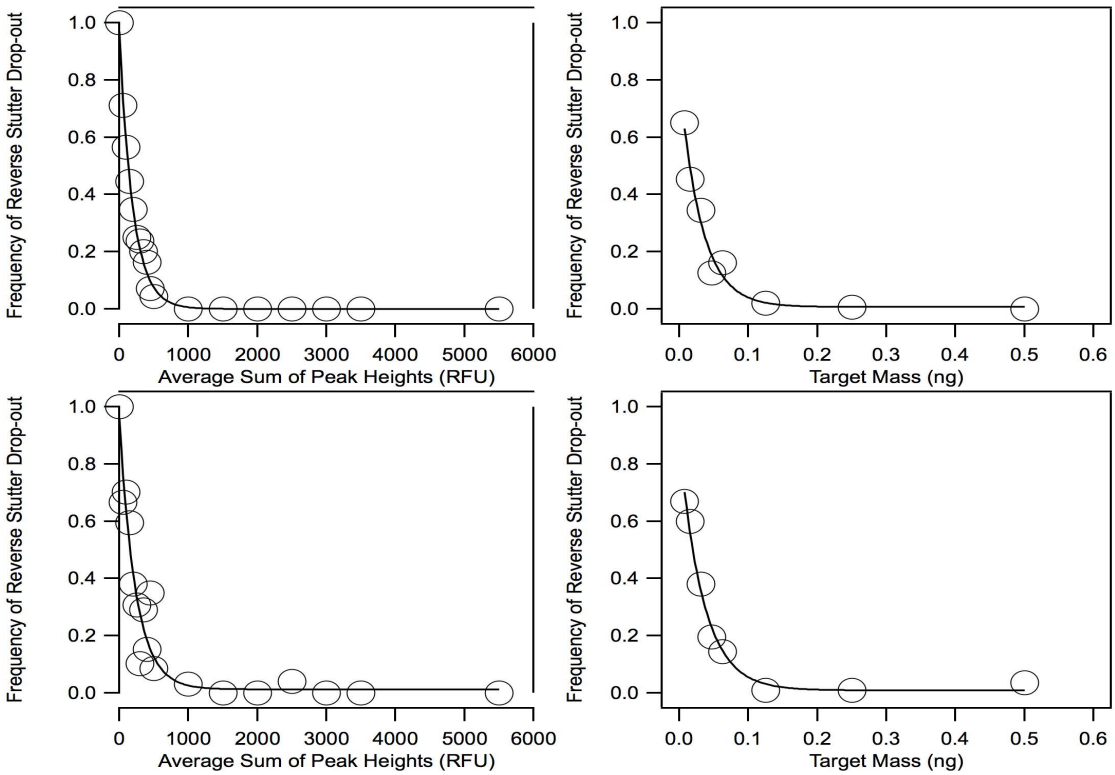

Figure 14: The frequency of reverse stutter drop-out (○) plotted against target mass and average sum of peak heights for the representative D2S1338 (top) and CSF1PO (bottom) loci

Table 10: The frequency of reverse stutter drop-out for average sum of peak heights, for all loci

APH	CSFIPO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	TH01	TPOX	vWA
0	1.00	1.00	1.00	0.67	1.00	0.75	1.00	0.33	1.00	0.67	0.67	1.00	1.00	1.00	1.00
50	0.67	0.66	0.77	0.84	0.83	0.81	0.71	0.53	0.83	0.86	0.66	0.85	0.85	0.83	0.85
100	0.70	0.65	0.59	0.82	0.67	0.58	0.56	0.47	0.79	0.70	0.57	0.68	0.79	0.81	0.84
150	0.60	0.56	0.49	0.63	0.48	0.50	0.45	0.22	0.62	0.71	0.29	0.67	0.78	0.87	0.60
200	0.38	0.49	0.37	0.59	0.40	0.36	0.35	0.16	0.61	0.56	0.25	0.59	0.74	0.80	0.56
250	0.31	0.33	0.40	0.60	0.39	0.30	0.25	0.11	0.44	0.39	0.13	0.52	0.62	0.82	0.48
300	0.10	0.35	0.26	0.46	0.28	0.35	0.24	0.10	0.42	0.50	0.06	0.34	0.57	0.80	0.45
350	0.29	0.10	0.21	0.34	0.48	0.13	0.20	0.06	0.36	0.31	0.03	0.33	0.41	0.66	0.42
400	0.15	0.20	0.17	0.25	0.23	0.16	0.16	0.03	0.36	0.24	0.03	0.34	0.52	0.72	0.37
450	0.35	0.17	0.00	0.24	0.18	0.09	0.07	0.00	0.28	0.25	0.00	0.29	0.43	0.73	0.21
500	0.09	0.09	0.05	0.20	0.20	0.14	0.04	0.02	0.22	0.18	0.00	0.14	0.37	0.50	0.19
1000	0.03	0.03	0.00	0.06	0.04	0.13	0.00	0.00	0.06	0.05	0.00	0.01	0.11	0.38	0.11
1500	0.00	0.01	0.00	0.04	0.00	0.18	0.00	0.00	0.00	0.02	0.00	0.03	0.03	0.20	0.01
2000	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.02	0.00	0.00	0.08	0.07	0.00
2500	0.04	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.00
3000	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.21	0.00
3500	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.13	0.00
5500	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.03	0.19	0.00	0.00

Table 11: The frequency of forward stutter drop-out for average sum of peak heights, for all loci

APH	CSF1PO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	TH01	TPOX	vWA
0	1	0.80	0.67	1.00	1.00	1.00	1.00	0.67	1.00	0.67	0.67	0.83	1.00	0.75	1.00
50	0.86	0.83	0.88	0.94	0.93	0.86	0.89	0.84	0.87	0.84	0.85	0.86	0.93	0.84	0.86
100	0.84	0.90	0.86	0.84	0.88	0.79	0.90	0.82	0.85	0.87	0.95	0.83	0.88	0.85	0.88
150	0.88	0.81	0.94	0.92	0.83	0.71	0.85	0.92	0.88	0.94	0.75	0.91	0.92	0.94	0.93
200	0.82	0.79	0.85	0.91	0.88	0.79	0.89	0.82	0.88	0.88	0.82	0.87	0.83	0.92	0.95
250	0.64	0.89	0.79	0.87	0.91	0.89	0.94	0.74	0.94	0.80	0.69	0.76	0.93	0.89	0.73
300	0.77	0.83	0.91	0.93	0.83	0.73	0.90	0.80	0.74	0.81	0.72	0.86	0.91	0.80	0.87
350	0.77	0.71	0.71	0.95	0.85	0.88	0.80	0.94	0.82	0.84	0.70	0.85	0.88	0.97	0.89
400	0.70	0.84	0.83	0.82	0.77	0.75	0.84	0.80	0.94	0.82	0.71	0.85	0.93	0.85	0.88
450	0.80	0.75	0.90	0.80	0.95	0.91	0.82	0.86	0.78	0.95	0.68	0.88	0.86	0.91	0.79
500	0.66	0.78	0.82	0.89	0.85	0.66	0.87	0.72	0.79	0.78	0.64	0.88	0.93	0.86	0.87
1000	0.69	0.67	0.63	0.79	0.92	0.54	0.92	0.71	0.80	0.80	0.44	0.79	0.89	0.88	0.88
1500	0.48	0.60	0.67	0.76	0.87	0.41	0.90	0.64	0.64	0.70	0.30	0.77	0.87	0.89	0.72
2000	0.36	0.53	0.45	0.66	0.87	0.16	0.80	0.48	0.55	0.55	0.29	0.83	0.94	0.87	0.83
2500	0.32	0.42	0.12	0.64	0.77	0.14	0.79	0.54	0.41	0.54	0.13	0.62	0.87	0.79	0.77
3000	0.55	0.38	0.31	0.57	0.93	0.17	0.88	0.50	0.50	0.38	0.06	0.75	0.81	1.00	0.75
3500	0.33	0.13	0.22	0.50	1.00	0.00	1.00	0.38	0.33	0.50	0.29	0.89	1.00	0.75	0.86
5500	0.10	0.08	0.13	0.37	0.80	0.03	0.81	0.25	0.14	0.23	0.00	0.31	0.88	0.75	0.67

Table 12: Frequency of reverse stutter drop-out for target mass for all loci

TM	CSFIPO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	TH01	TPOX	vWA
0.0078	0.67	0.63	0.72	0.82	0.74	0.68	0.65	0.51	0.83	0.83	0.68	0.79	0.85	0.83	0.86
0.0156	0.60	0.52	0.54	0.68	0.66	0.52	0.45	0.30	0.71	0.66	0.37	0.72	0.76	0.84	0.66
0.0313	0.38	0.40	0.31	0.55	0.45	0.31	0.35	0.14	0.45	0.46	0.19	0.45	0.65	0.77	0.49
0.047	0.20	0.22	0.18	0.34	0.31	0.14	0.13	0.03	0.37	0.31	0.04	0.36	0.52	0.68	0.38
0.0625	0.14	0.23	0.17	0.31	0.22	0.21	0.16	0.05	0.30	0.35	0.05	0.25	0.40	0.66	0.32
0.125	0.01	0.05	0.02	0.06	0.03	0.15	0.02	0.00	0.06	0.07	0.01	0.04	0.14	0.34	0.10
0.25	0.01	0.01	0.00	0.04	0.00	0.15	0.01	0.00	0.00	0.01	0.00	0.00	0.03	0.18	0.02
0.5	0.04	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.03	0.22	0.00	0.00

Table 13: Frequency of forward stutter drop-out for target mass for all loci

TM	CSF1PO	D13	D16	D18	D19	D21	D2	D3	D5	D7	D8	FGA	TH01	TPOX	vWA
0.0078	0.85	0.89	0.84	0.93	0.89	0.86	0.90	0.83	0.90	0.83	0.86	0.85	0.92	0.83	0.87
0.0156	0.83	0.83	0.88	0.83	0.88	0.77	0.87	0.90	0.83	0.93	0.86	0.88	0.85	0.87	0.92
0.0313	0.73	0.82	0.89	0.89	0.89	0.77	0.93	0.87	0.87	0.86	0.75	0.86	0.90	0.93	0.85
0.047	0.74	0.83	0.83	0.83	0.86	0.83	0.84	0.76	0.85	0.80	0.68	0.85	0.90	0.90	0.87
0.0625	0.79	0.75	0.82	0.92	0.84	0.67	0.84	0.72	0.82	0.78	0.69	0.83	0.95	0.85	0.88
0.125	0.53	0.70	0.68	0.79	0.89	0.51	0.88	0.67	0.74	0.79	0.48	0.80	0.87	0.88	0.81
0.25	0.46	0.52	0.43	0.68	0.87	0.27	0.86	0.54	0.55	0.59	0.22	0.77	0.95	0.85	0.80
0.5	0.18	0.14	0.18	0.45	0.83	0.07	0.81	0.33	0.17	0.35	0.06	0.39	0.78	0.80	0.72

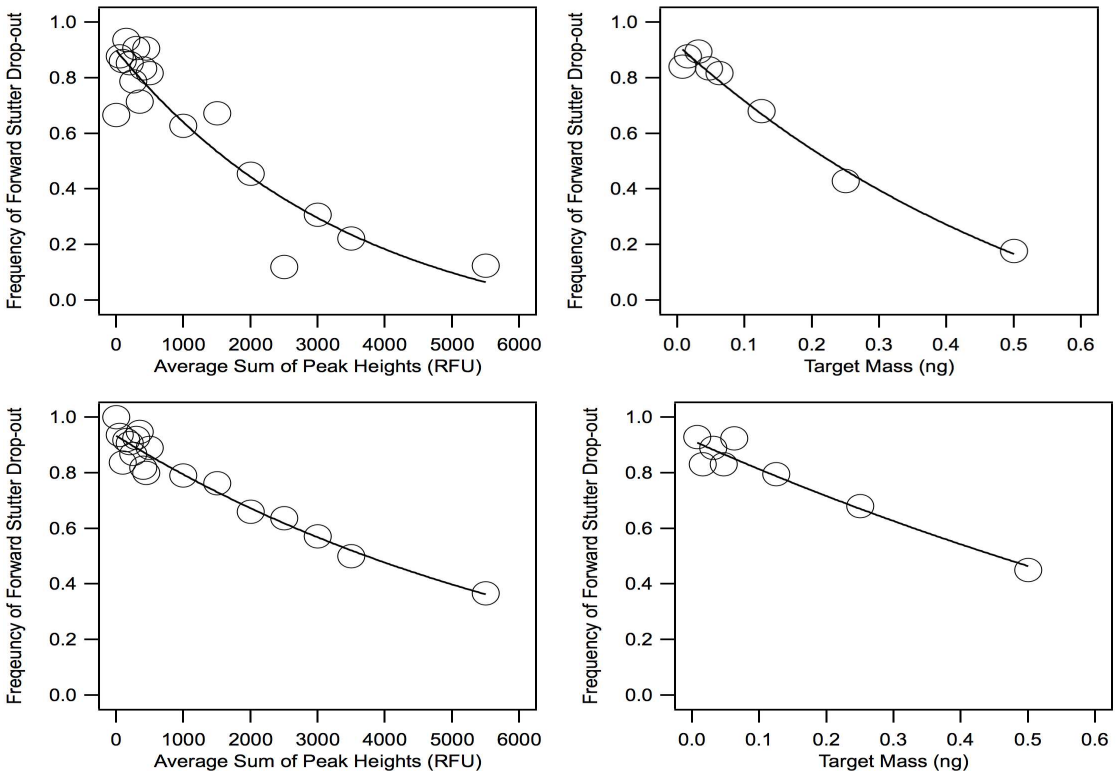

Figure 15: The frequency of forward stutter drop-out (○) plotted against target mass and average sum of peak heights for the D16S539 (top) and D18S51 (bottom) loci

Overall, the models for APH show concordance with the models for TM which indicates that APH is comparable to TM as an independent variable. Since APH is shown to fit the models used by the NOC*It* algorithm, samples can be tested on NOC*It* using APH.

3.2.2 Effect of forward stutter on noise distribution

Many interpretation schemes do not incorporate forward stutter ratios into the modeling and as such the forward stutter is typically classified as noise. This in turn would affect the noise distribution and may be one of the reasons why noise has been shown to increase with target mass (52,54). Thus, it was of interest to examine the effects of classifying forward stutter as noise on the noise distribution in an attempt to evaluate whether or not there is a significant implication to modeling noise and determining an analytical threshold. Figures 16 and 17 show the noise distributions with forward stutter categorized as noise and without forward stutter categorized as noise, per dye channel. The 0.0078 ng (Figure 16) and 0.125ng (Figure 17) masses were chosen to represent a low template sample and an average sample that may be encountered in crime laboratories. Table 14 shows the *p*-values of the KS-test for all target masses. For the lower target mass, the *p*-value is 1, for all colors, indicating that the two datasets are similar (Table 14), which is corroborated visually by examining the tails of the lognormal distribution; the tails for both datasets overlap (Figure 16). However, as the target mass increases, the *p*-values begin decreasing indicating that the datasets begin to differ (Table 14). For example, at 0.125 ng for the blue channel, the p-value is 0 (Table 14). Graph B of Figure 17 illustrates this trend visually; the tails of the two datasets diverge with the distribution for the noise with forward stutter removed resulting in a lower tail than the noise with forward stutter. The trend is more pronounced in the blue and green dye channels

compared to the red and yellow dye channels. This may be due to the fact that the red and yellow channels have been shown to have a higher baseline noise compared to the blue and green dye channel (52). If there are significantly more noise peaks in the red and yellow channels, the effects of whether forward stutter is present would have a minimal impact on the overall distribution.

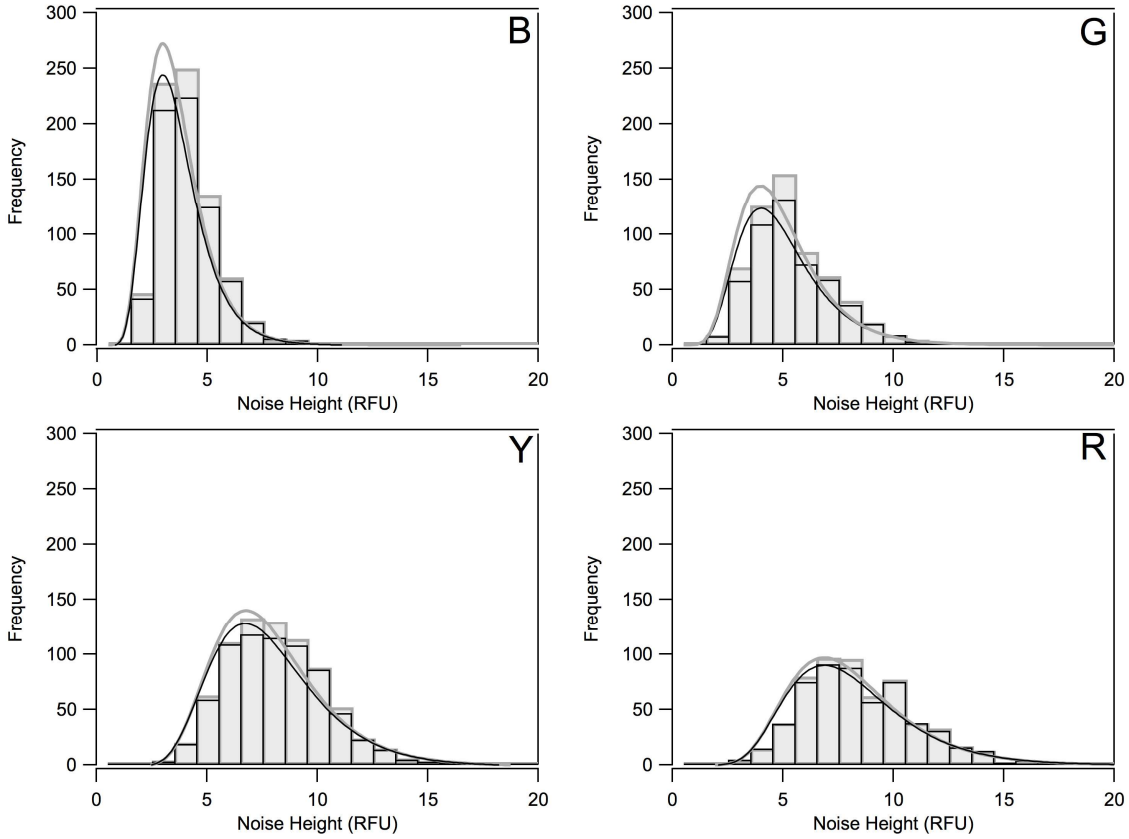

Figure 16: Noise distribution with forward stutter (grey) and without forward stutter (black) fitted with lognormal distribution (solid line) for 0.0078ng, separated by dye channels (B = Blue, G = Green, R = Red, Y = Yellow).

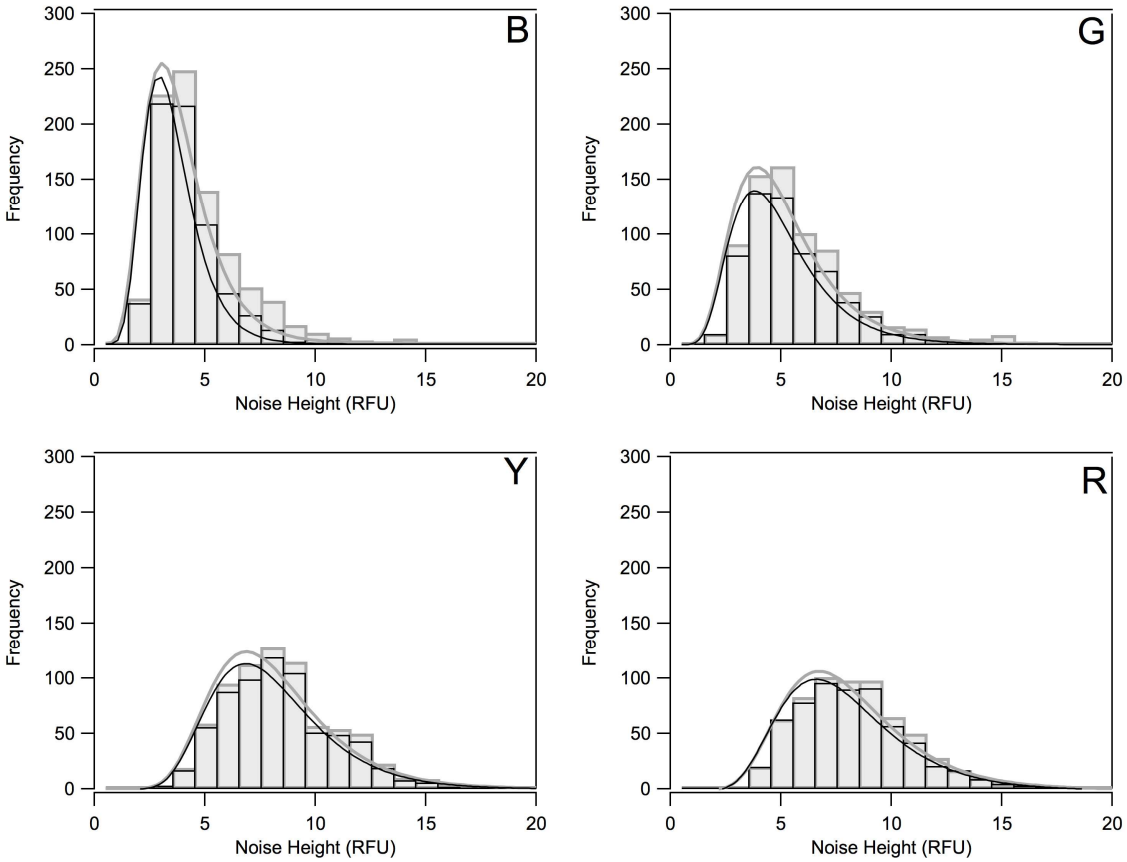

Figure 17**:** Noise distribution with forward stutter (grey) and without forward stutter (black) fitted with lognormal distribution (solid line) for 0.125ng, separated by dye channels (B = Blue, G = Green, R = Red, Y = Yellow).

Table 14 shows that misclassification of forward stutter as noise does not seem to have an impact on noise when data from lower target masses are evaluated. This is presumably the result of high probability of forward stutter drop-out as seen in Figure 15. However, since high template masses seems to be affected by presence of forward stutter, noise data which excludes forward stutter from the set would be preferred as it could be used for samples that contain both low and high masses of DNA.

Table 14: *p* values from KS-test for all target masses

Dye Channel	Mass (ng)	KS Test (*p)*
Blue	0.0078	1
	0.0156	1
	0.0313	1
	0.047	1
	0.0625	0.661
	0.125	0
	0.25	0
	0.5	0.001
Green	0.0078	1
	0.0156	1
	0.0313	1
	0.047	1
	0.0625	0.890
	0.125	0.541
	0.25	0.002
	0.5	0.057
Yellow	0.0078	1
	0.0156	1
	0.0313	1
	0.047	1
	0.0625	0.962
	0.125	0.998
	0.25	0.640
	0.5	0.879
Red	0.0078	1
	0.0156	1
	0.0313	1
	0.047	1
	0.0625	0.999
	0.125	0.999
	0.25	0.326
	0.5	0.050

3.2.3 Reproducibility Study

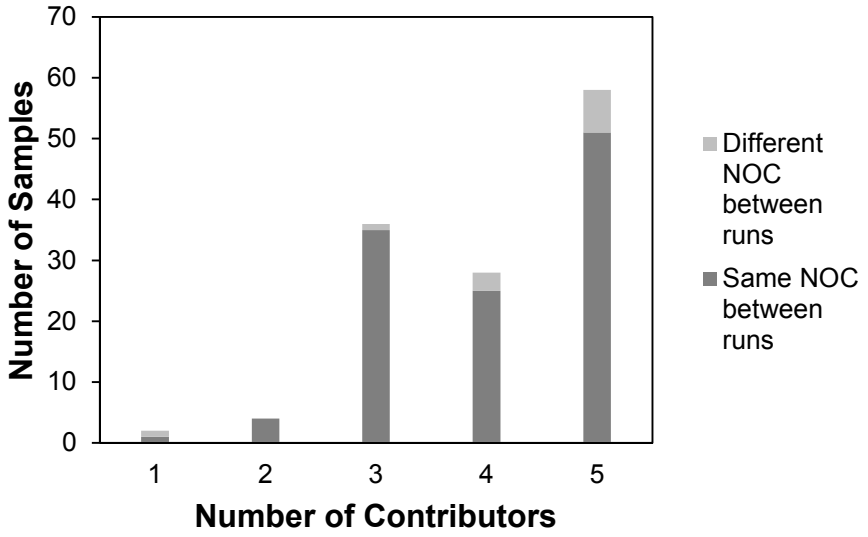

Figure 18: The precision of NOC*lt* displayed by mixture type

Before NOC*lt* could be tested with APH, a reproducibility study was performed. Figure 18 shows the number of NOC*lt* estimates, separated by the actual NOC in a sample. Ideally, if a sample were tested multiple times, NOC*lt* would estimate the same NOC each time and result in similar probabilities for each *n*. Figure 18 shows that for the 5-person mixtures there were more instances of NOC*lt* imprecisely estimating the NOC, compared to the low order mixtures. Figure 19 plots the range of APP against the median APP of the five runs per sample. The range was calculated by subtracting the minimum APP from the maximum APP. The highest APP NOC*lt* can assign to an *n* is 1; the data show that the majority of the median APP values range from 0.9-0.99 with an APP range of less than 0.2 (Figure 19). The outlier with a range of

approximately 0.3 is a five-person sample that exhibited more variability in APP values than the majority of the samples. The other two outliers were samples where NOC*It* estimated the same NOC for 4 of the 5 replicates and imprecisely estimated the NOC for 1 of the 5 replicates resulting in a large variation for the APP range (Figure 19). However, when the median APP is less than approximately 0.85, the range is more variable. This suggests that the APP output obtained for a sample may be used to indicate whether the sample will consistently produce the same NOC. Perlin et al. examined the precision of TrueAllele utilizing duplicate runs of 101 profiles and observed that the variation for log (LR) between computer runs was greater at the medium LR values, but the larger LR values (and therefore more informative) were very reproducible (30). Comparably, Taylor et al. examined STRmix™ reproducibility by analyzing a 2-person mixtures run 10 times and observed that the LRs obtained from separate runs were highly reproducible (33). Similarly, Bright et al. determined the reproducibility of STRmix™ by examining 10 replicates of a 2-person mixture sample, artificially created based on two single source Identifiler™ profiles, and comparing the LR values (32). The coefficient of variation for the major contributor was 0.24%, while for the minor contributor it was 12.85% (32). Since the majority of the samples tested with NOC*It* resulted in the same NOC with similar APP values, the software was deemed reproducible.

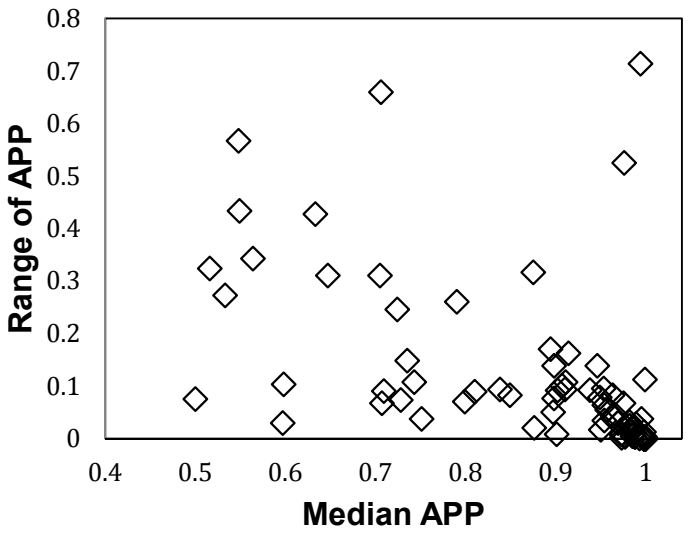

Figure 19: The range of APP plotted against the median APP of the quintuplicate runs

3.2.4 Testing of NOCIt using Average Peak Height

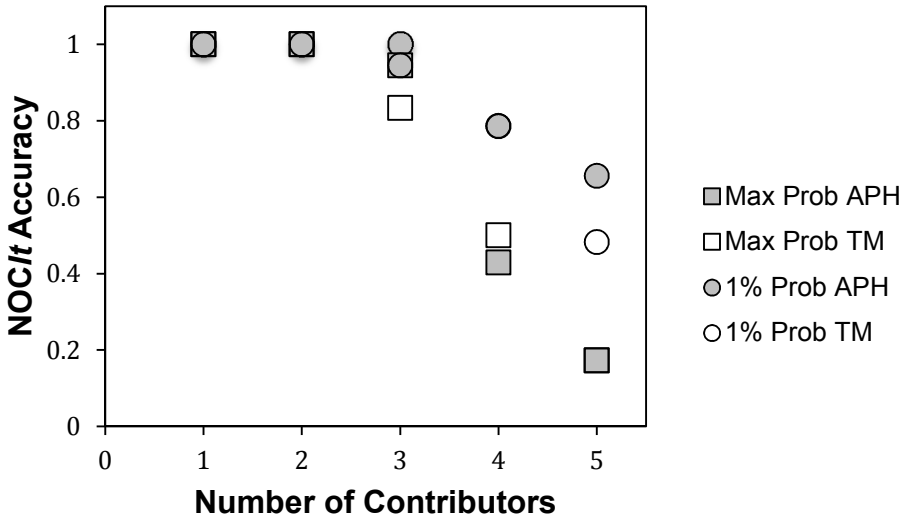

Figure 20: The accuracy of NOC*It* for average sum of peak heights (APH) and target mass (TM) displayed by mixture type

The overall accuracy of NOC*It* using the max probability is 48% for APH and 47% for TM and with the 1% probability it is 80% for APH and 70% for TM.

The trend observed for all data sets is a decrease in accuracy as the number of contributors within a sample increases (Figure 20). The 1- and 2- person samples for APH and TM show an accuracy of 100% for both the max and 1% accuracy. The small sample size may be a factor for the high accuracy, although a small deviation from 100% would be expected for these samples, as observed in Swaminathan et al. (26). For the 3- person samples, the accuracy of the max probability for APH was 94% and for TM was 83%, while the accuracy of the 1% probability for APH was 100% and for TM was 94%. For the 4-person samples, the accuracy of the max probability for APH was 43% and for TM was 50%, while the accuracy of the 1% probability for both was 79%. For the 5-person samples, the accuracy of the max probability for both was 17%, while the accuracy of the 1% probability for APH was 66%and for TM was 48%. The accuracy results obtained for TM are lower than the recorded accuracy for TM in Swaminathan at al. (26). This is expected, as the subset chosen for this analysis contained samples with extreme mixture ratios and targets as low as 0.002 ng for the minor contributor. These results concur with Taylor et al. who examined the effect of mass of minor contributor on LR utilizing STRmix™ by analyzing 127 artificially constructed 2-person mixtures of known genotypes (33). The study shows that as the minor template decreases, so does the LR. Further, as the ability to resolve peaks belonging to the major or minor contributor is lost, the weightings for the genotypes becomes equal and the probability is spread, further reducing the LR (33).

The results of this study illustrates that for purposes of estimating the NOC that gave rise to the stain, APH is as accurate as TM, and either can be utilized successfully.

4. Conclusion

In the first part of this study, LMD was evaluated to determine DNA recovery. The DNA recovery for LMD was suboptimal. The previously reported recovery rates range from 16-32% (45-46). In contrast, the recovery rates obtained for LMD samples during this study ranged from 0%-61%, 0-6%, and 0-9%, with the majority of the samples recovering less than 10%. These results indicate that the LMD methodology utilized in this study would need further optimization prior to use in forensic laboratories.

In the second part of this study, the use of peak heights as a means to estimate the NOC was evaluated. It was determined that both TM and APH can be successfully utilized to produce the models used by NOC*lt*. Since one of the variables modeled by NOC*lt* is forward stutter, the effects of forward stutter on the noise distribution were also evaluated. Results obtained utilizing a KS-test show that at lower target masses there is not a discernible difference between distributions, but at higher target masses the forward stutter seems to impact the noise distribution. Prior to analyzing samples with APH on NOC*lt*, the software reproducibility was assessed. The results illustrate that the software is reproducible with only 12 samples estimating different NOC within runs out of 128 samples. The APP values were also shown to be repeatable when the median APP was greater than 0.85. Testing of the software using APH illustrated accuracy levels comparable to TM; the overall accuracy of NOC*lt* using the max probability is 48% for APH and 47% for TM and with the 1% probability it is 80%

for APH and 70% for TM. It is concluded that APH could be used by probabilistic software, such as NOC*It*, and produce reasonable results.

References

1. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature 1985:314;67-73.
2. Mullis KB and Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods in Enzymology 1987:155;335-350.
3. Butler JM. Fundamentals of forensic DNA typing. Boston, MA: Academic Press; 2010.
4. Applied Biosystems. Quantifiler® Duo DNA Quantification Kit User's Manual. Foster City, CA:2014.
5. Applied Biosystems. AmpFlSTR® Identifiler® Plus Amplification Kit User's Manual. Foster City, CA:2014.
6. Greenspoon SA, Scarpetta MA, Drayton ML, Turek SA. QIAamp spin columns as a method of DNA isolation for forensic casework. Journal of Forensic Sciences 1998;43(5):1024-1030.
7. Melzak KA, Sherwood CS, Turner RFB, Haynes CA. Driving forces for DNA adsorption to silica in perchlorate solutions. Journal of Colloid and Interface Sciences 1996;181:635-644.
8. Phillips SM. A comparative study of DNA extraction methodologies: variation in DNA yield and effects on downstream PCR analysis [Masters Thesis]. Boston, MA: Boston University School of Medicine; 2009.
9. Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucleic Acids Research 2005;33(2):632-643.
10. Grgicak CM, Urban ZM, Cotton RW. Investigation of reproducibility and error associated with qPCR methods using Quantifiler® Duo DNA quantification kit. Journal of Forensic Sciences 2010;55(5):1331-1339.

11. Koukoulas I, O'Toole FE, Stringer P. Quantifiler™ observations of relevance to forensic casework. Journal of Forensic Sciences 2008;53(1):135-141.
12. Stenman J and Orpana A. Accuracy in amplification. Nature Biotechnology 2001;19:1011-1012.
13. Timken MD, Klein SB, Buoncristiani MR. Stochastic sampling effects in STR genotyping: Implications for analysis and interpretation. Forensic Science International: Genetics 2014;11:195-204
14. Rutledge RG and Côté C. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Research 2003;31:1-6.
15. Booth CS, Pienaar E, Termaat JR, Whitney SE, Louw TM, Viljoen HJ. Efficiency of the polymerase chain reaction. Chemical Engineering Science 2010;65(17):4996–5006.
16. Walsh PS, Erlich HA, Higuchi R. Preferential PCR amplification of alleles: mechanisms and solutions. Genome Research 1992;1(4):241–50.
17. Walsh SP, Fildes NJ, Reynolds R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Research 1996;24(14):2807-2812
18. Sanders CT, Sanchez N, Ballantyne J, Peterson DA. (2006). Laser microdissection separation of pure spermatozoa from epithelial cells for short tandem repeat analysis. Journal of Forensic Sciences 2006;51(4):748-757.
19. Vandewoestyne M, Nieuwerburgh FV, Hoofstat DV, Deforce D. Evaluation of three DNA extraction protocols for forensic STR typing after laser capture microdissection. Forensic Science International: Genetics, 2012;6:258-262.
20. Axler-DiPerte G, Orans S, Singh A, Caragine T, Prinz M, Budimlja ZM. Comparison and optimization of DNA recovery form sperm vs. epithelial cells using laser capture microdissection technology and an

immunofluorescent staining system. Forensic Science International: Genetics Supplement Series 2011;3:e224-e225

21. Vandewoestyne M, Hoofstat DV, Nieuwerburgh FV, Deforce D. Suspension fluorescence in situ hybridization (S-FISH) combined with automatic detection and laser microdissection for STR profiling of male cells in male/female mixtures. International Journal of Legal Medicine 2009;123:441-447.
22. Budimlija ZM, Lechpammer M, Popiolek D, Fogt F, Prinz M, Bieber FR. Forensic applications of laser capture microdissection: Use in DNA-based parentage testing and platform validation. Croatian Medical Journal 2005;46(4):549-555.
23. Elliott K, Hill DS, Lambert C, Burroughes TR, Gill P. Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides. Forensic Science International 2003;137:28-36.
24. Ballantyne J, Hanson EK, Perlin MW. DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. Science and Justice 2013;53:103-114.
25. Molecular Machines & Industries AG. mmi CellCut Plus® User Manual. Glattbrugg, Switzerland:2007.
26. Swaminathan H, Grgicak CM, Medard M, Lun DS. NOCIt: A computational method to infer the number of contributors to DNA samples analyzed by STR genotyping. Forensic Science International: Genetics 2015;16: 172-180.
27. Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE. Empirical Analysis of the STR profiles resulting from conceptual mixtures. Journal of Forensic Sciences 2005;50(6):1-6.

28. Buckleton JS, Curran JM, Gill P.Towards understanding the effect of uncertainty in the number of contributors to DNA stains. Forensic Science International: Genetics 2007;1:20-28.
29. Lohmueller KE and Rudin N. Calculating the weight of evidence in low-template forensic DNA casework. Journal of Forensic Sciences 2013;58(S1):S243-249.
30. Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S. TrueAllele casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. PloSone 2014;9(3):1-15.
31. Cooper S, McGovern C, Bright J, Taylor D. Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics 2015;16:121-131.
32. Bright J, Evett IW, Taylor D, Curran JM, Buckleton. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics 2015;14:125-131.
33. Taylor D, Bright J, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics 2013;7:516-528.
34. Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW. Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences 2011;56(6):1430-1447.
35. Haned H, Pène L, Lobry JR, Dufour AB, Pontier D. Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? Journal of Forensic Sciences 2011;56(1):23-28.

36. Haned H, Pène L, Sauvage F, Pontier D. The predictive value of maximum likelihood estimator of the number of contributors to a DNA mixture. Forensic Science International: Genetics 2011;5:281-285.
37. Haned H, Benschop CCG, Gill PD, Sijen T. Complex DNA mixture analysis in a forensic context: evaluating the probative value using a likelihood ratio model. Forensic Science International: Genetics 2015;16:17-25.
38. Scientific Working Group on DNA Analysis Methods. SGWDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Approved 01/14/10.
39. Picanco JB, Raimann PE, Paskulin GA, Alvarez L, Amorim A, Batista Dos Santos SE, Alho CS.Tri-allelic pattern at the TPOX locus: a familial study. Gene. 2014;535(2):353-358.
40. QIAGEN. QIAamp® DNA Investigator Handbook. Germantown, MD:2012.
41. Tiersch TR, Chandler RW, Wachtel SS, Elias S. Reference standards for flow cytometry and application in comparative studies of nuclear DNA content. Cytometry 1989;10:706-710.
42. Shapiro HS. Deoxyribonucleic acid content per cell of various organisms. In: Handbook of Biochemistry and Molecular Biology, Fasman GD (ed). CRC Press, Cleveland 1976;2:284-306.
43. Millipore. Amicon® Ultra-0.5 Centrifugal Filter Devices User Guide. Billerica, MA:2011.
44. ZyGEM. DNA Extraction Using forensicGEM® Saliva Quick-Start Guide. Lane Hamilton, New Zealand:2013.
45. Colussi A, Viegas M, Beltramo J, Lojo M. Efficiency of DNA IQ system® in recovering semen DNA from cotton swabs. Forensic Science International: Genetics Supplement Series 2009;2:87-88.
46. Kishore R, Hardy WR, Anderson VJ, Sanchez NA, Buoncristiani MR. Optimization of DNA extraction from low-yield and degraded samples

using the BioRobot® EZ1 and BioRobot ® 48. Journal of Forensic Sciences 2006;51(5):1055-1061.

47. Garvin AM, Fritsch A. Purifying and concentrating genomic DNA from mock forensic samples using Millipore Amicon filters. Journal of Forensic Sciences 2013;58(S1):173-175.
48. Iacona JR. Genomic DNA isolation from amplified product for recursive genotyping of low-template DNA samples [Masters Thesis]. Boston, MA: Boston University School of Medicine; 2013.
49. Mygind T, Østergaard L, Birkelund S, Lindholt JS, Christiansen G. Evaluation of five DNA extraction methods for purification of DNA from atherosclerotic tissue and estimation of prevalence of *Chlamydia pneumonia* in tissue from a Danish population undergoing vascular repair. BMC Microbiology 2003; 3(19): 1-12.
50. Kaeser JC. Investigating the origin of stochastic effects in low-template DNA samples by developing a single-tube extraction protocol [Masters Thesis]. Boston, MA: Boston University School of Medicine; 2013.
51. WaveMetrics, Inc. IgorPro: Version 6.36 User's Guide. Lake Oswego, OR:2014.
52. Bregu J, Conklin D, Coronado E, Terrill M, Cotton RW, Grgicak CM. Analytical thresholds and sensitivity: establishing RFU thresholds of forensic DNA analysis. Journal of Forensic Sciences 2013;58(1):120-129
53. Gilder JR, Doom TE, Inman K, Krane DE. Run-specific limits of detection and quantiation for STR-based DNA testing. Journal of Forensic Sciences 2007;52(1):97-101.
54. Wellner G. The examination of baseline noise and the impact on the interpretation of low-template DNA samples [Masters Thesis]. Boston, MA: Boston University School of Medicine; 2014.

55. Gill P, Puch-Solis R, Curran J. The low-template-DNA (stochastic) threshold- its determination relative to risk analysis for national DNA databases. Forensic Science International: Genetics 2009;3:104-111.
56. Rakay CA, Bregu J, Grgicak CM. Maximizing allele detection: Effects of analytical threshold and DNA levels on rates of allele and locus drop-out. Forensic Science International: Genetics, 2012;6:723-728.
57. Swaminathan H, Lun D, Médard M, Norsworthy S, Rowan K, Wellner, G, Grgicak C. NOCIt: A Method for Determining the Number of Contributors Technical Manual: 2014.
58. Norsworthy S. Characterizing rates of allelic dropout and the impact on estimating the number of contributors. Presented at the American Academy of Forensic Sciences 66th Annual Meeting; Seattle, WA February 2014.
59. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. (2000). An investigation of the rigor of interpretation rules for STRs derived from less than 100pg of DNA. Forensic Science International 2000;112:17-40.
60. Bright J, Buckleton JS, Taylor D, Fernando MACSS, Curran JM. Modeling forward stutter: toward increased objectivity in forensic DNA interpretation. Electrophoresis 2014;35:3152-3157.
61. Bright J, Taylor D, Curran JM, Buckleton, JS. Developing allelic and stuter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics 2013;7:296-304.
62. Brookes C, Bright J, Harbison SA, Buckleton J. Characterising stutter in forensic STR multiplexes. Forensic Science International: Genetics 2012;6:58-63.
63. Gibb AJ, Huell AL, Simmons MC, Brown RM. Characterisation of forward stutter in the AmpFlSTR® SGM Plus® PCR. Science and Justice 2009;49:24-31.

64. Seo SB, Ge J, King JL, Budowle B. Reduction of stutter ratios in short tandem repeat loci typing of low copy number DNA samples. Forensic Science Interntational: Genetics 2014;8:213-218

CURRICULUM VITAE