DEVELOPING A FORENSICALLY RELEVANT SINGLE-CELL INTERPRETATION STRATEGY FOR HUMAN IDENTIFICATION

By

AMANDA J. GONZALEZ

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Master of Science

Graduate Program in Biology

Written under the direction of

Dr. Catherine Grgicak

And Approved by

_____

Catherine M. Grgicak, M.S.F.S., Ph.D.

_____

Daniel Shain, Ph.D.

_____

Nir Yakoby, Ph.D.

Camden, NJ

May 2019

THESIS ABSTRACT

Developing a Forensically Relevant Single-cell Interpretation Strategy for Human Identification

by AMANDA J. GONZALEZ

Thesis Director:
Dr. Catherine Grgicak

Biological evidence submitted to the forensic DNA laboratory contains cells from an unknown number of contributors in unknown proportions, resulting in profiles that are difficult to interpret.

Thus, recent efforts have focused on developing single-cell forensic DNA pipelines to deconvolve mixture signal by separating cells at the front end of processing. Single-cell signal, however, are often obfuscated by the presence of confounding signal such as false negative detection of alleles (i.e., drop-out); stutter, a polymerase chain reaction (PCR) artifact; and false positive detection of alleles (i.e., drop-in). Given the need to provide the weight-of-evidence against the accused, probabilistic characterization of the confounding single-cell artifacts is a necessity.

As such, 556 single-source, single-cell samples of known genotype were analyzed. The data were evaluated to determine if distributions associated with allele detection, stutter, and allelic drop-in were significantly different from those of bulk-

processed samples. The results demonstrate that, in contrast to bulk-processed samples, allele detection is cell dependent. Like bulk-processed samples, stutter in the single-cell regime was found to be locus dependent; however, single-cell samples resulted in higher stutter ratios. As predicted, the frequency of allelic drop-in appeared consistent with that of bulk-processed samples. These findings suggest current state-of-the-art probabilistic systems are ill-equipped to evaluate single-cell evidence and new probabilistic constructs are required. The results of this study form the foundation from which these new inference systems may be developed.

Not only is probabilistic characterization of single-cell signal a necessity; practical implementation of a single-cell pipeline (i.e., that the cells can be effectively desorbed from common collection material such as cotton swabs) must be verified. Therefore, the second phase of this work focused on developing and accessing a protocol to desorb buccal cells from cotton-tipped applicators. To measure its efficacy, hemocytometry was used to determine the percent of cells recovered. The percent recovery of buccal cells appeared consistent with that of bulk-mixture extraction, demonstrating that a single-cell strategy is a viable alternative to the traditional forensic DNA pipeline.

ACKOWLEGEMENTS

I would like to acknowledge and thank the many people who have encouraged and supported me throughout my academic career.

Dr. Catherine Grgicak has been an invaluable mentor and advisor. Thanks to her knowledge and expertise, I have developed and expanded on new skills such as: fragment analysis, Visual Basic Application (VBA) coding, sample preparation, hemocytometry, and statistical analysis. Her instruction has not only provided me with the unique privilege of learning forensic DNA analysis; it has promoted exponential learning and growth as a new forensic scientist, and I will be forever in her debt.

I would also like to thank Kimberlee Moran, Director of the inaugural Rutgers-Camden M.S. Forensic Science Program for paving the way for new forensic scientists. She laid the foundation for my career in forensic science and encouraged me to return to Rutgers-Camden to pursue my master's degree. She has provided me with many academic and professional opportunities throughout the years, and words cannot express how grateful I am for all the mentorship she has bestowed upon me.

I am grateful for my thesis committee, including Dr. Daniel Shain and Dr. Nir Yakoby, who have been indispensable in the development and finalization of my written thesis.

I would like to thank my husband, Juan and four young children: Gianni, Genevieve, Giorgio, and Giuliette for their patience and support through the long days and nights of research. They were my absolute leading light.

To my friends and peers, I couldn't have done it without their unrelenting support in the final writing process, especially my research companion, Ms. Laura Malek.

Finally, a special thank you to the Forensic Sciences Foundation (FSF) for funding my research endeavors via the Jan S. Bashinski Criminalistics Graduate Thesis Assistance Grant. Without funding, my project would not have been possible, and I look forward to presenting my research at the 2020 72nd American Academy of Forensic Sciences (AAFS) Annual Meeting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1. Background Information

Forensic science is the application of scientific principles and techniques to criminal or civil proceedings as they relate to the collection, examination, and analysis of physical evidence. Forensic DNA analysis is a process by which scientists analyze and interpret DNA profiles generated from biological evidence that has been collected from a crime scene. In many cases, the data generated from evidence sample is compared to that of a suspect standard, and the weight-of-the evidence for a suspect is reported to the court in terms of a match statistic.

The interpretation of these DNA profiles is complicated for mixtures as the number of cells and the number of contributors for a sample are unknown and DNA from the cells are bulk-processed from the collection material (i.e., cotton swab) for downstream processing. This results in a DNA profile that contains comingled allelic information of more than one contributor, ultimately affecting interpretation and the match statistics.

Given the complications with bulk-mixture interpretation, this study aims to explore the possible implementation of a single-cell system. In a single-cell system, cells are first eluted from the collection material and pico-pipetted into individual microcentrifuge tubes for downstream processing. This results in the deconvolution of the mixture at the front-end of processing, where the DNA profiles can be analyzed and interpreted, cell by cell.

Existing inference systems are used by forensic scientists to interpret and compute match statistics; however, these systems make assumptions based on bulk-processed samples.

The purpose of this study is the to characterize single-cell signal to determine if existing inference systems can be applied to single-cell data and whether the common cotton-swab collection technique is a reasonable method of collection for a single-cell system.

1.1.1. Defining Short Tandem Repeats (STRs)

The most commonly used genetic markers in forensic DNA analysis are short tandem repeats (STRs). STRs, also known as microsatellites, are repetitive tandem nucleotide sequence arrays, usually repeating between 15 and 30 times [1];[2]. These polymorphic DNA sequences are dispersed throughout the human genome and given their hypervariability, are the most commonly used markers for human identification [1]. An STR contains several tandemly repeating nucleotide sequences that are typically 1–6 base pairs in length, i.e. mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats [2]. An individual inherits two of these STR sequences in the form of two alleles per locus (the location of the allele on chromosome): one from their mother and one from their father. Therefore, an individual can exhibit STR homozygosity if the child inherited the same allele from their mother and father or heterozygosity when two different repeat unit lengths are inherited.

Figure 1 is a schematic of an STR region for two alleles. Note that there are five and seven tandemly repeating tetranucleotides contained in the target STR regions with conserved sequences in flanking regions on each side.

STR typing is the most commonly used method for human identification, and is, therefore, the method employed throughout this study.

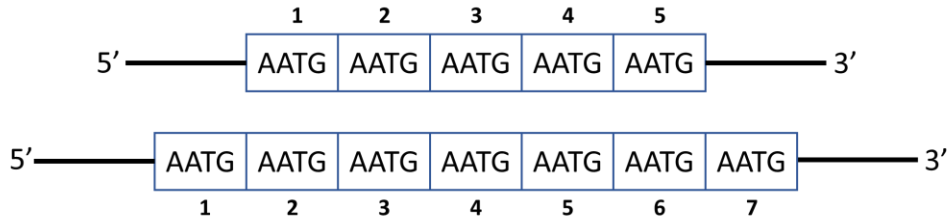

Figure 1: A schematic of STR regions of two alleles. Note that there are five and seven tandemly repeating tetranucleotides contained in the target regions with conserved flanking regions on each side. The STR genotype of this individual is G= 5, 7.

1.1.2. Mixture Interpretation for Bulk Processed Samples

Biological evidence submitted to the forensic DNA laboratory contains cells from an unknown number of contributors in unknown proportions, making mixture interpretation from interleaved, bulk-processed samples an arduous task.

The current state of interpretation is to compute the likelihood ratio (*LR*), which is the ratio of probabilities given two mutually exclusive propositions:

$$\mathrm{LR} = \frac{\mathrm{P(E|H}_p\mathrm{)}}{\mathrm{P(E|H}_d\mathrm{)}} \quad \text{(Equation 1)}$$

In the forensic DNA context, the two propositions are: H_d—that n random individuals contributed to the evidence, *E*; and H_p—that the suspect and $(n-1)$ random individuals contributed. Though much effort over the last decade has focused on the development of software programs to compute the *LR* in an automated fashion [3];[4], significantly improving interpretation [5], issues with mixture interpretation remain. For example, Bille et al. demonstrated that applying significantly different probabilistic models to the same data during inference can significantly affect the result [6].

Swaminathan et al. takes this a step further [7], demonstrating that even small changes to the model assumptions may have significant ramifications to the *LR* outcomes, suggesting that despite the sophisticated mathematical methods used to compute the weight-of-evidence (i.e., *LR*), forensic mixture interpretation has yet to be completely solved without material modifications to the forensic pipeline. Not only do the propositions and models found to the right of "|" in Equation 1 impact the *LR*, the authors of [8] demonstrated that the quality and information content contained in *E* also has profound effects. Briefly, Peters et al. tested thirty-six mixture profiles while varying the analytical threshold and showed that as the limit of detection approached one molecule of DNA, the *LR* for true contributors increased while the probability for a non-contributor decreased resulting in a $LR>1$, indicating that the information content and quality of data contained in *E* plays an integral role in determining *LR* outcomes [8].

In the forensic context, the data, *E*, takes the form of an electropherogram, containing information about the length of the DNA amplicon and the number of amplicons produced during PCR (polymerase chain reaction). It is the end result of a laboratory pipeline that includes: 1) the extraction of a mixture of cells from an unknown number of contributors; 2) quantification; 3) amplification of a set of forensically relevant STR (short tandem repeat) loci; and 4) the fragment analysis of these STR amplicons through capillary electrophoresis and laser-induced fluorescence. If the sample contains many contributors other than the POI (person-of-interest), the POI's signal is obfuscated by these interference contributors and the quality of *E* may be too low or too complicated for interpretation.

1.1.3. Single-cell Methods: Deconvolving Mixture Signal

One potential path forward is the introduction of single-cell systems into the forensic laboratory. If successful, single-cell methods have the potential to deconvolve the mixture signal by separating cells at the front end of processing as demonstrated by [9];[10].

Figure 2 illustrates the potential of single-cell techniques for mixture deconvolution. Figure 2(a) is the resultant STR electropherogram when a 3-person mixture is processed using a traditional bulk-processing pipeline, where the cells are lysed while in the same tube and the DNA from multiple contributors is co-amplified. Figure 2(b) through 2(f), however, are STR electropherograms from the same cellular admixture, but here the DNA has been extracted and amplified one cell at a time. The 3-person, bulk-processed mixture profile (Figure 2(a)) could be explained by hundreds of genotype combinations, decreasing the weight against any true contributor and increasing the probability that a non-contributor is not excluded. In contrast, single-cell profiles unambiguously indicate that the allelic information contained in each of the electropherograms represents only one of the contributors to the 3-person admixture.

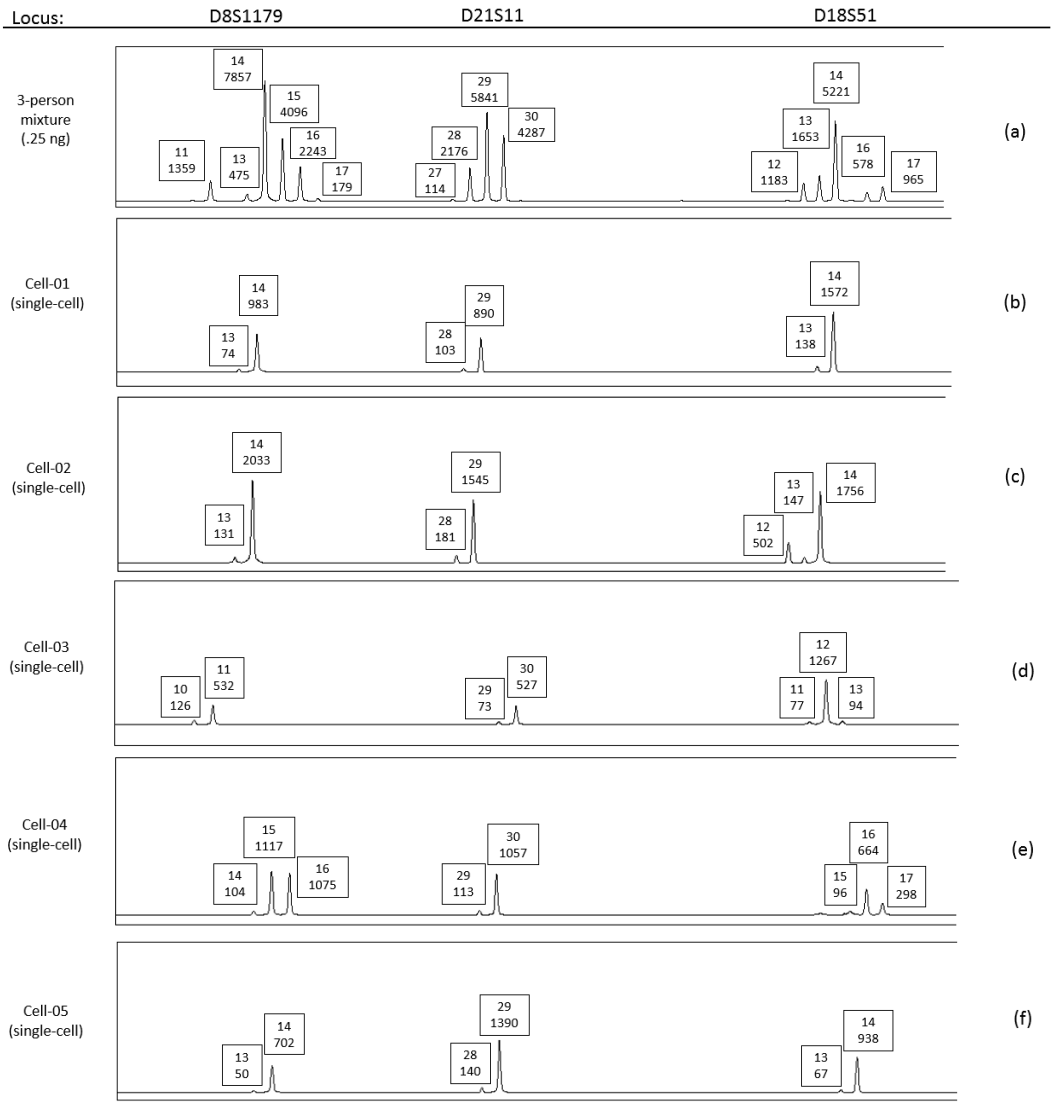

Figure 2: Six STR electropherograms from three representative loci: D8S1179, D21S11, and D18S51 (laboratory conditions: Globalfiler™ amplification, for 30 PCR cycles, 25-second injection on 3500 CE). Figure 2(a) is an electropherogram from a 0.25ng, 3-person (1:1:1) bulk-processed mixture. Figures 2(b) through 2(f) are electropherograms resulting from five individually processed cells from the same 3-person admixture. Note that some of the signal represented in the individually processed cells is in stutter position and is attributed to PCR stutter artifact.

1.2. Defining Allelic Drop-Out, Stutter, and Allelic Drop-In

Despite its popularity, DNA comparisons between profiles acquired from the crime scene to that of the suspect is an arduous task made ever more difficult by the presence of the confounding signal due to allelic drop-out, stutter or allelic-drop-in.

1.2.1. Allelic Drop-out

Allelic drop-out is the false negative detection of an allele and is observed when only a few cells worth of DNA are available for testing. Allelic drop-out is caused by: 1) stochastic sampling error, 2) poor quality of the template DNA due to degradation, 3) primer annealing inefficiency, or 4) detection effects (see Figure 3) [11]. Figure 3(a) depicts stochastic sampling error, where the two red alleles are not contained in the volume fraction aliquoted for PCR. Figure 3(b) depicts the case when PCR extension does not proceed due to the degraded state of the template DNA. Figure 3(c) depicts primer annealing inefficiency due to primer-binding site mutations that may occur in the typically nonpolymorphic flanking regions of the template DNA, while Figure 3(d) depicts the false negative detection of alleles resulting from the application of an analytical threshold.

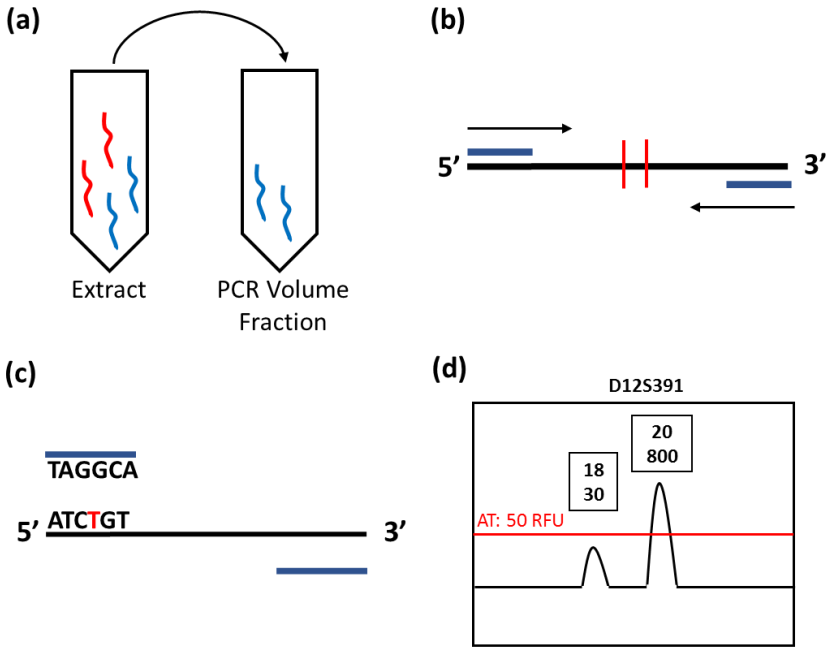

Figure 3: A schematic illustration of (a) stochastic sampling error, (b) poor quality of the template DNA due to degradation, (c) primer annealing inefficiency, and (d) detection effects.

1.2.2. Stutter

Stutter (reverse stutter) signal is signal in a position that is one STR unit shorter than the suspected allele. Stutter product is a known STR artifact generated during amplification, also known as polymerase chain reaction (PCR) [12] and is hypothesized to occur because there is random strand slippage of the copied strand during replication (see Figure 4). In some instances, the stutter products themselves undergo strand slippage, resulting in what is termed double-back stutter—a PCR product which is two STR units shorter than the targeted template molecule. In terms of signal, if present, reverse stutter signal is observed to the left of the suspected parent allele and, typically, results in peaks heights that are significantly lower [12].

Forward stutter, like reverse stutter is also attributed to strand slippage, but in this instance it is the synthesized strand which is thought to loop or slip during the extension phase of PCR (see Figure 4) [12]. Though the synthesis of forward stutter is mechanistically similar to that of reverse stutter, the frequency of its occurrence is not, resulting in fewer forward stutter strands and less impact on the total signal. For example, Bright et al. demonstrated that the amplified DNA of several single-source samples across four different multiplexes, produced very few forward stutter products per locus, ranging from 0.0% to 3.2% , with the highest rate of forward stutter being attributed to the trinucleotide, D22S1045 locus [13]. Bright et al. also demonstrated that 61% of all forward stutter was below an 50 RFU—an analytical threshold commonly used by the forensic laboratory [13].

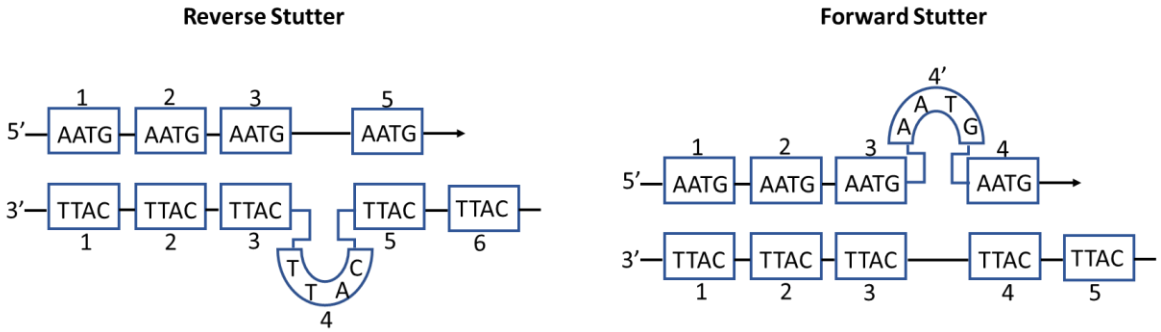

Figure 4: A schematic illustration of strand slippage that occurs either in the forward or reverse direction during PCR, adapted from [14];[15].

1.2.3. Allelic Drop-in

Drop-in (extraneous signal) is a stochastically occurring phenomenon that occurs when extraneous DNA, not original to the sample, is co-amplified with it.

1.3. Signal Impacts of Allelic Drop-out, Stutter, and Allelic Drop-in

The DNA commission of the International Society of Forensic Genetics (ISFG) convened at the 21st congress of the International Society for Forensic Genetics to recommend guidelines for the best practice of mixture interpretation and low copy number (LCN) reporting [16], in which Gill et al. highlights the need for continuing education and research in the aforementioned areas [16].

1.3.1. Allelic Drop-out

Figure 5 is a visual representation of the way in which these products impact STR electropherograms (EPGs). Figure 5(a) demonstrates the effects of allelic drop-out on the resultant EPG, wherein the known genotype for the TPOX locus is 8,11. Since evidentiary samples are of unknown origin, the genotype must be inferred. If the phenomenon of drop-out is not considered and we assume the number of contributors (NOC) is one, the STR genotype G=11,11 would be inferred and the true contributor with a genotype of 8,11 would erroneously seem to have not contributed. Alternatively, if

drop-out is considered, the inferred genotype from the evidence would be reported as a genotype of 11,11 or 11,*O*, where *O* is any allele not detected. Within this interpretation paradigm, the accused with a genotype of 8,11 would be included as a potential contributor, but with significantly less weight (i.e., the LR would be closer to 1). Peters et al. tested thirty-six mixture profiles while varying the analytical threshold (AT) and showed that as the AT increased, the *LR* for true contributors decreased while the probability that a non-contributor resulted in a *LR*>1 increased, indicating that allelic drop-out due to detection effects can significantly impact *LR* outcomes [17]. According to the ISFG, if drop-out of an allele is required to explain the evidence under Hp, e.g., Suspect= a,b; Evidence=a, then the signal intensity, i.e., peak height of allele *a* must be small enough to justify the potential drop-out of allele *b* [16]. Conversely, if a full evidence profile is obtained where the signal intensity of allele *a* is significantly higher than that of background noise, the probability of drop-out for the undetected partner allele approaches 0, and Hp cannot be supported [16]. More specifically, if allele, *a* is observed at a locus, in a mixture, with signal intensity close to background noise level, this indicates a minor contributor with the possible undetected partner allele, which has implications for the suspect with an *ab* genotype at that locus [16]. As the intensity of the signal for *a* increases, the probability of drop-out approaches zero, the suspect can be excluded ($LR \approx 0$), and probability that the source is *aa* is more likely [16].

1.3.2. Stutter

Figure 5(b) shows the impact of stutter product on the known genotype for the D5S818 locus is 10,11. Additional signal observed at position 9 is one STR less than the 10 allele, and is indicative of a reverse stutter artifact. Signal is also observed at position

8 which is two STRs less than the 10 allele, and is indicative of double-back stutter artifact. Additional signal is also observed at position 12 which is one STR unit larger than the 11 allele—this is indicative of forward stutter artifact. According to the ISFG, the origin of signal in stutter position can be from an allele, stutter, or a mixture of both [16]. For example, if the number of contributors assumption for the EPG in Figure 5(b) is 2, then it is possible that the signal in positions 8 and 9 belong to a minor contributor and signal in the 10 and 11 positions belong to a major contributor. However, signal in position 9 falls into stutter position of allele 10. In this case, stutter signal cannot be distinguished from allelic signal. Therefore, if signal in stutter position cannot be distinguished from allelic signal, then the probability of stutter must be considered in the numerator of Equation 1 when determining *LR*.

1.3.3. Allelic Drop-In

In Figure 5(c), the known genotype for the D12S391 locus is 18,23; however, additional signal is observed in the 25 position and is not consistent with stutter signal position. According Gill et al., drop-in is restricted to 1 or 2 alleles per profile, and if multiple alleles are observed at more than two loci, then an additional contributor or contamination must be considered [18]. According to the ISFG, a uniform assumption is made regarding number of contributors (NOC) across loci [16], and when drop-in is not considered, additional contributors are assumed. The maximum allele count (MAC) approach is commonly used in forensic science laboratories to determine NOC, where the maximum number of alleles at a locus is divided by two and rounded up. This provides an accurate estimate on the NOC when the sample is a simple mixture at high copy numbers, but only provides the lower bound on the NOC for more complex DNA

evidence [19]. For example, given the extraneous signal (position 26) present on the EPG in Figure 5(c), a maximum of three peaks must be considered for the D12S391 locus. When divided by 2 (maximum number of expected alleles per locus), the minimum number of contributors assumption for the whole profile is rounded to 2. This significantly impacts the LR ratio, as a NOC of 2 affects genotype inferences and *LR* outcomes. For this reason, the ISFG recommends that drop-in be considered for *LR* calculations when less than two loci exhibit additional signal [18]. In this case, if drop-in is considered, the number of contributors assumption remains 1, and the probability of drop-in is factored into *LR* calculations. Since the NOC has a significant impact on *LR* outcomes, much effort has also been applied to building probabilistic systems that infer the NOC [20];[21];[22].

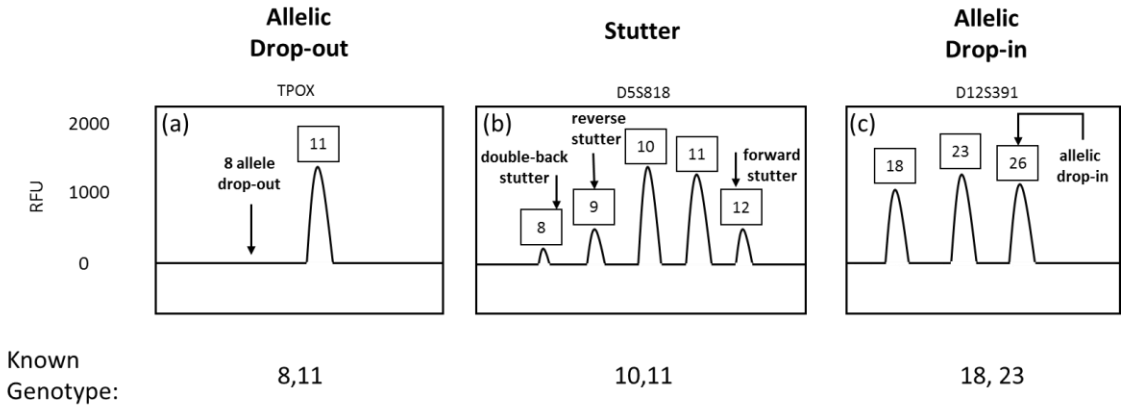

Figure 5: A visual representation of (a) allelic drop-in, (b) stutter (reverse, forward, and double-back) and (c) allelic drop-in as it appears on an electropherogram given the NOC is 1.

1.4. Bulk Mixture versus Single-copy Confounding Signal

1.4.1. Allelic Drop-out

As indicated in the previous section, considerations associated with allelic drop-out can have significant impacts on genotype inference, and these negative impacts

exponentially grow in the presence of mixtures. For example, peaks A,B, and C at locus *l*, where drop-out is considered and where it is not considered is depicted in Table 1. Here we see that the number of genotypes that explain the EPG greatly increases within the NOC assumption. As the NOC assumption increases, so does the number of genotypes. As the number of genotypes that explain the evidence increases, the *LR* decreases.

Table 1: The inferred genotype combinations for three observed peaks (A,B, and C) when allelic drop-in is considered versus when allelic drop out is not considered, given the number of contributors assumption is 2. Note that *O* represents any allele not detected.

	Person 1	Person 2
Inferred Genotype Combinations If Drop-Out *Not* Considered (A,B,C)	A,A	B,C
	A,B	A,C or B,C or C,C
	A,C	A,B or B,B or B,C
	B,B	A,C
	B,C	A,A or A,B or A,C
	C,C	A,B
Inferred Genotype Combinations If Drop-Out *Is* Considered (A,B,C,O)	A,A	B,C
	A,B	A,C or B,C or C,C
	A,C	A,B or B,B or B,C
	B,B	A,C
	B,C	A,A or A,B or A,C
	C,C	A,B
	A,O	B,C
	A,B	C,O
	A,C	B,O
	B,C	A,O
	B,O	A,C
	C,O	A,B

Thus, though there are numerous probabilistic interpretation systems that can evaluate the likelihood of the electropherogram across these genotype combinations, the ability to definitively report that a person was a contributor is wholly dependent upon the data quality (i.e. drop-out) itself. As such, single-cell methods have recently garnered attention in the forensics realm [10] since it may to lead to improved inference outcomes.

Despite its potential, single-cell pipelines are not without their challenges and have been shown to render data with significant levels of allelic drop-out [10].

Even if single-cell methods are found to be more informative, state-of-the-art probabilistic systems assume allele drop-out is a cell-independent phenomenon. Thus, if single-cell pipelines are to be integrated into current forensic pipelines, it is necessary to explore and, perhaps, confirm allelic drop-out is cell independent. If it is not, then new interpretation paradigms must be engineered for these sample-types.

The amplification of DNA acquired from one cell has been explored, in the forensic context, since 1997 [9] and may, incorrectly, be considered the equivalent process to amplifying technical replicates. The amplification of DNA from a multitude of independently and identically distributed (i.i.d.) DNA molecules, that are evenly dispersed in solution can be described using a binomial distribution:

$$T_{o,a} \sim \text{Binomial}(T_{ext}, \frac{V_{PCR}}{V_{ext}}) \quad \text{(Equation 2)}$$

where $T_{o,a}$ is the total number of target molecules of allele *a* at PCR cycle 0, $\frac{V_{PCR}}{V_{ext}}$ is the volume fraction dispensed into the PCR tube from the extract tube, and T_{ext} is the total number of DNA copies available to be pipetted from the extract tube.

Therefore, $T_{o,a}$ represents the initial copy number of allele *a* in the PCR tube which survived pre-PCR processing and is available for amplification. After *c* rounds of PCR, the total number of PCR fragments synthesized during amplification is dependent upon PCR efficiency and whether stutter fragments are produced in lieu of full amplicons [17]. Notably, Equation 2 assumes that the DNA molecules are identically and independently distributed, throughout the extract, which is reasonable given that DNA is often eluted into and stored in aqueous solutions. Indeed, comparisons between

experimental and simulated data demonstrates that the signal is well described in this manner [17];[8].

For the single-cell case, however, direct PCR is used as the extraction technique. Here, the cell is aliquoted into the PCR tube, lysed, and the PCR reagents are aliquoted directly into the extraction tube, thereby theoretically negating sampling effects associated with pipetting volume fractions from the extract tube, effectively rendering $\frac{V_{PCR}}{V_{ext}} = 1$. Thus in the single-cell case, allelic drop-out will not be a random phenomenon dependent upon $\frac{V_{PCR}}{V_{ext}}$, but on a value that is correlated with the cell (i.e. drop-out will be cell-dependent).

Preliminary examination suggests this may, indeed, be the case. Recall, that direct PCR is part of the single-cell pipeline; as such, the expectation is that allelic drop-out is a rare event. The data in Figure 6, indicates that this may not the case. Figure 6 shows four representative loci of three single-cell samples from the same contributor. These three samples were amplified on the same 96-well plate under the same laboratory conditions; however, the resultant EPGs indicate varying allelic drop-out rates.

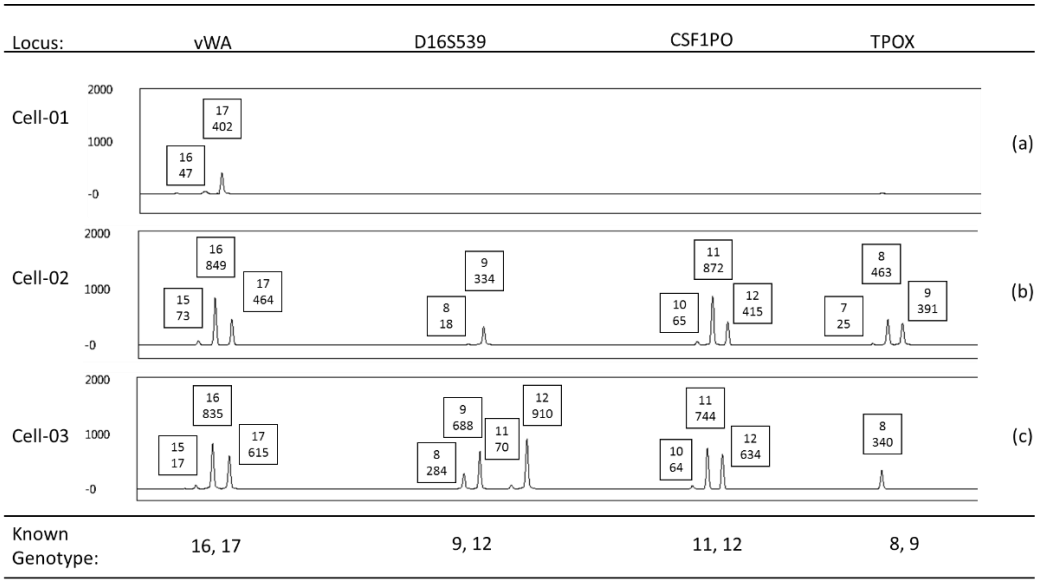

Figure 6: Three electropherograms from four representative loci: vWA, D16S539, CSF1PO, and TPOX (laboratory conditions: Globalfiler™ amplification, for 30 PCR cycles, 25-second injection on 3500 CE). Cell-01 through Cell-03 are electropherograms resulting from three individually processed cells of the same known contributor (single-source samples). Note the that allelic drop-out seems to vary between cells and that stutter artifact appears in the EPGs of all three cells (see table 2).

The cause of these observations is unknown, but have previously been attributed to ineffective extraction strategies [10];[23], though previous work in this laboratory demonstrated that modification to the extraction technique did not significantly impact allele drop-out rates (C.M. Grgicak, personal communication, August 8, 2018) [24]. Since full resolution of mixtures can only occur at the front end of processing, a new and relevant interpretation strategy for single-cell data is justified. To accomplish this, characterization of allelic drop-out patterns are a necessity.

If allele drop-out is independent and not impacted by detection effects [17], then the probability of drop-out is:

$$\Pr\left(\text{Binomial}\left(T_{ext}, \frac{V_{PCR}}{V_{ext}}\right)\right) = 0 \qquad \text{(Equation 2)}$$

In single-cell analysis, $T_{ext} = 1$ and if direct PCR is used, then $\frac{V_{PCR}}{V_{ext}}$ is 1 and the probability of drop-out would be 0. Even if one assumes small volume loss, for example

20%, such that $\frac{V_{PCR}}{V_{ext}}$ = 0.80, the probability of drop-out should be randomly distributed amongst STR loci and should not exceed 20%. Thus, for the common 24 locus STR profile, only nine to ten are expected to randomly drop out. However, as previously shown in Figure 6(a), exploratory examination of single-cell profiles demonstrated that drop-out greatly exceeds this value and may be cell-dependent. This work will, therefore, endeavor to statically evaluate whether allelic drop-out is cell independent.

1.4.2. Stutter

Table 2 summarizes the allele drop-out rate and stutter ratios for each of the three cells described in Figure 6. The stutter ratio (*SR*) is determined by dividing the peak height of the peak in stutter position, PH_{ST}, by the peak height of the allele, PH_{AL}, given the known genotype, which will be discussed in more detail in Section 2.1.2.

Thus, the second key feature from preliminary data analysis that shall be explored is that the stutter ratios for single-cell data may be significantly larger and more varied than stutter ratios obtained from high-template samples [17]. If the stutter ratio distribution for single-cell data are significantly impacted by PCR branching effects, then this is additional evidence demonstrating that current interpretation strategies, based on bulk-mixture models [3];[25], would need to be replaced by newly engineered probabilistic systems. Thus, this work will endeavor to perform an investigative, empirically-driven analysis using experimentally derived samples, to explore stutter trends.

Table 2: Summary of the allele drop-out rates ($\widehat{Pr(DO)}$) and stutter ratios (SR) for each of the three cells represented in Figure 6.

Replicate	Allele Drop-out Rate ($\widehat{Pr\ (DO)} = \frac{N_{PH<5RFU}}{N_{exp}}$)	Stutter Ratio ($SR = \frac{PH_{ST}}{PH_{AL}}$)
Cell-01	0.71	0.17
Cell-02	0.09	0.52
Cell-03	0.09	0.41

Note: the stutter filter threshold is 0.15 based on validation studies from the manufacturer [26].

To investigate potential branching effects on stutter production, an *in-silico* PCR laboratory model is often utilized [17];[27]. Figure 7 is one such model, and demonstrates the ways in which new amplicons, stutter, and double-back stutter artifacts are produced during PCR. In this model, the *Stock Tube* is a memory bank wherein any *New Amplicons* produced during cycle, *c* of PCR are stored. The probability of successfully copying a DNA molecule is the efficiency of the PCR, and was set to 95% such that $T_{NewAmplicon} \sim$ Binomial($T_{StockTube}$, 95) at every PCR cycle. Of the new amplicons produced, approximately 0.55% become stutter, which are themselves amplified in subsequent cycles. Of the *Stutter 1* copies produced, approximately 0.55% of those undergo strand slippage, generating double-back stutter.

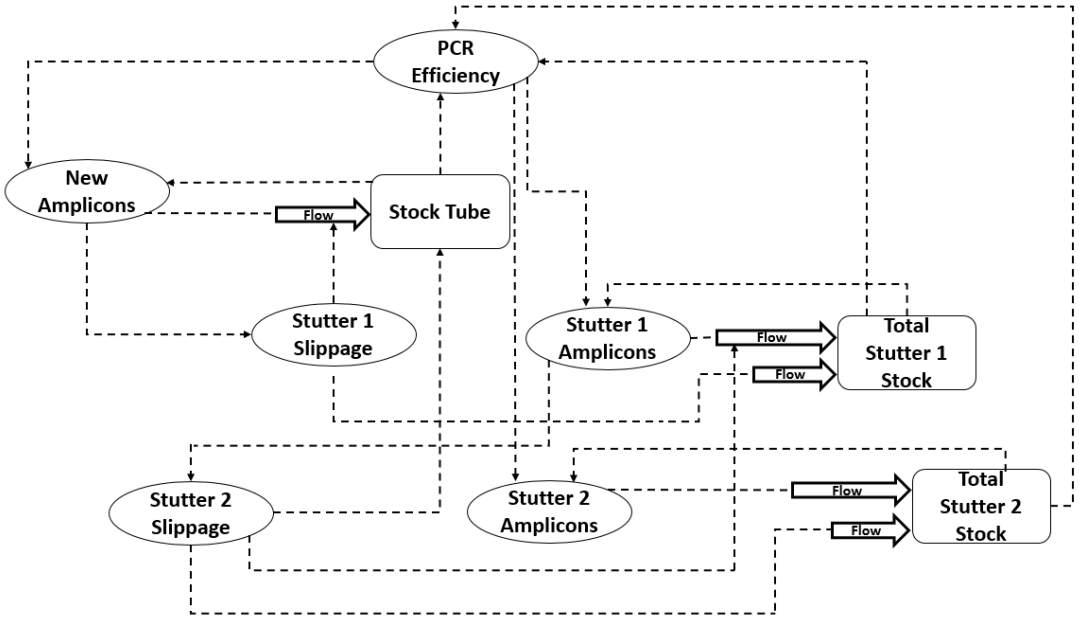

Figure 7: The causal-loop diagram explicating the relationship between new amplicon, stutter, and double-back stutter production during STR amplification.

Thus, total *Stutter 1* or *2* products and the total number of amplicons produced over *c* cycles is representative of peak heights one would expect to observe in the electropherogram. This, therefore, is a reasonable model which can be used to explore expected results. As such, twenty PCR simulations were conducted, for 29 cycles, with template copy numbers of $T_{o,a} = 1$ and $T_{o,a} = 40$ copies, respectively, to examine potential impacts on initial copy numbers on stutter and allele peak heights.

Table 3 contains the total number of end cycle (29th cycle) amplicons and stutter products for 20 PCR simulations, where $T_{o,a} = 1$. Relative Fluorescent Units (RFUs) were then calculated for the total new amplicons and total stutter products by multiplying the respective copy number by the sensitivity of the instrument: 1.4045×10^{-6} RFU/copy, giving the RFU of the total amplicons (peak height) and the RFU of each total stutter product (stutter height) for each PCR simulation [8]. The stutter ratio was calculated by dividing the total stutter height by the total peak height for each PCR simulation.

Table 3: A table containing the total number of end cycle (29th cycle) amplicons and stutter products for 20 PCR simulations, where $T_{o,a}$ = 1. Relative Fluorescent Units (RFUs) were then calculated for the total new amplicons and total stutter products by multiplying the respective copy number by the sensitivity of the instrument: 1.4045x10^{-6} RFU/copy, giving you the RFU of the total amplicons (peak height) and the RFU of each total stutter product (stutter height) for each PCR simulation. The stutter ratio was calculated by dividing the total stutter height by the total peak height for each PCR simulation.

PCR Run Number	Total Amplicons (Allele 1: $T_{o,a}$ = 1, 29th end cycle)	RFU of Total Amplicons (Peak Heights)	Total Stutter (Allele 1: $T_{o,a}$ = 1, 29th end cycle)	RFU of Total Stutter (Stutter Height)	Stutter Ratio ($SR = \frac{PH_{ST}}{PH_{AL}}$)
1	189,250,720	266	14,489,477	20	0.08
2	164,909,966	232	14,145,153	20	0.09
3	205,453,777	289	12,502,898	18	0.06
4	226,512,108	318	15,854,655	22	0.07
5	156,937,252	220	29,632,942	42	0.19
6	282,266,495	396	19,775,519	28	0.07
7	286,349,828	402	14,619,220	21	0.05
8	234,642,597	330	53,257,120	75	0.23
9	213,815,509	300	18,438,685	26	0.09
10	196,893,624	277	14,542,591	20	0.07
11	254,288,034	357	21,827,797	31	0.09
12	273,199,355	384	24,427,148	34	0.09
13	277,347,484	390	18,061,567	25	0.07
14	188,491,084	265	9,482,713	13	0.05
15	192,931,897	271	15,577,728	22	0.08
16	205,496,263	289	21,015,022	30	0.10
17	249,789,804	351	14,872,196	21	0.06
18	202,900,157	285	17,059,098	24	0.08
19	112,678,425	158	13,083,033	18	0.12
20	192,592,712	270	9,666,048	14	0.05

Table 4 contains the total number of end cycle (29th cycle) amplicons and stutter products for 20 PCR simulations, where $T_{o,a}$ = 40. Relative Fluorescent Units (RFUs) were then calculated for the total new amplicons and total stutter products by multiplying the respective copy number by the sensitivity of the instrument: 1.4045x10^{-6} RFU/copy, giving the RFU of the total amplicons (peak height) and the RFU of each total stutter

product (stutter height) for each PCR simulation [8]. The stutter ratio was calculated by dividing the total stutter height by the total peak height for each PCR simulation.

Table 4: A table containing the total number of end cycle (29th cycle) amplicons and stutter products for 20 PCR simulations, where $T_{o,a}$ = 40. Relative Fluorescent Units (RFUs) were then calculated for the total new amplicons and total stutter products by multiplying the respective copy number by the sensitivity of the instrument: 1.4045×10^{-6} RFU/copy, giving you the RFU of the total amplicons (peak height) and the RFU of each total stutter product (stutter height) for each PCR simulation. The stutter ratio was calculated by dividing the total stutter height by the total peak height for each PCR simulation.

PCR Run Number	Total Amplicons (Allele 2: $T_{o,a}$ = 40, 29th end cycle)	RFU of Total Amplicons (Peak Heights)	Total Stutter (Allele 2: $T_{o,a}$ = 40, 29th end cycle)	RFU of Total Stutter (Stutter Height)	Stutter Ratio ($SR = \frac{PH_{ST}}{PH_{AL}}$)
1	10,300,000,000	14,466	825,452,706	1,159	0.08
2	9,940,000,000	13,961	631,640,787	887	0.06
3	8,620,000,000	12,107	783,638,943	1,101	0.09
4	7,720,000,000	10,843	875,124,433	1,229	0.11
5	11,200,000,000	15,730	725,564,274	1,019	0.06
6	8,220,000,000	11,545	772,888,306	1,086	0.09
7	9,650,000,000	13,553	669,643,897	941	0.07
8	9,110,000,000	12,795	812,197,567	1,141	0.09
9	10,200,000,000	14,326	704,975,298	990	0.07
10	10,000,000,000	14,045	755,866,771	1,062	0.08
11	9,490,000,000	13,329	761,578,131	1,070	0.08
12	9,000,000,000	12,641	863,455,918	1,213	0.10
13	10,700,000,000	15,028	669,519,233	940	0.06
14	8,750,000,000	12,289	812,012,169	1,140	0.09
15	8,500,000,000	11,938	682,735,216	959	0.08
16	11,200,000,000	15,730	842,687,840	1,184	0.08
17	10,000,000,000	14,045	1,150,000,000	1,615	0.12
18	8,590,000,000	12,065	914,389,846	1,284	0.11
19	10,700,000,000	15,028	919,050,003	1,291	0.09
20	9,870,000,000	13,862	851,635,247	1,196	0.09

Figure 8 depicts a box plot of the stutter ratios shown in Tables 3 and 4.

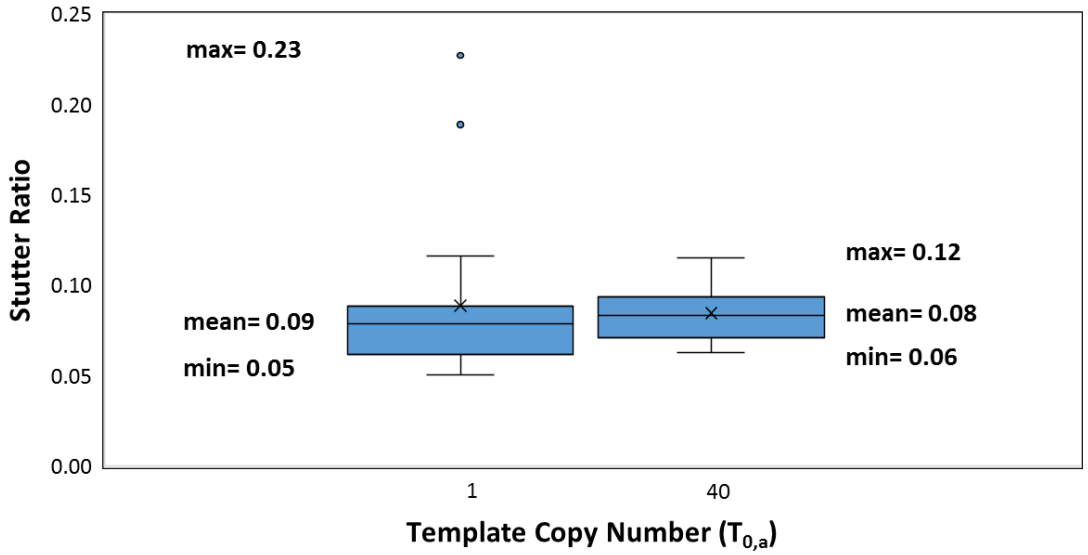

Figure 8: A box plot of the stutter ratios of 20 PCR simulations for Allele 1 and Allele 2 against their respective initial copy number/s at cycle zero: $T_{o,a}$ = 1 and $T_{o,a}$ = 40. The mean, maximum, and minimum values are reported.

In general, the simulation demonstrates that although the means of the stutter ratios were relatively close in value, the maximum stutter ratio when $T_{o,a}$ = 1 is expected to be significantly higher than at $T_{o,a}$ = 40. Upon further inspection, the difference in the maximum stutter ratio is attributed to the occurrence of strand slippage early in the amplification process. Though early strand slippage occurs in both the $T_{o,a}$ = 1 and 40 cases, in the $T_{o,a}$ = 40 scenario one or two template molecules that produce stutter early in cycling are masked by the remaining 38 or 39 template copies that did not produce stutter product early in the PCR process, negating the effects of the former on signal interpretation. Therefore, as per the *in-silico* model, the expectation is to observe higher stutter ratios, more often, in data obtained from single-cell samples.

1.4.3. Allelic Drop-in

As for allelic drop-in, preliminary analysis demonstrates that only a small number of samples exhibited signal, which may be due to allele drop-in. By characterizing the allelic drop-in for all of the samples and comparing it to drop-in rates procured from non-single-cell approaches, we will elucidate whether drop-in rates increase or decrease at the single-cell regime. If drop-in increases, then an assessment of drop-in independence will be performed.

Mitchell et al. accessed the allelic drop-in of 700 saliva samples (combinations of 85 contributors) of varying template masses, in duplicate and triplicate, with an applied analytical threshold of 75 RFU [28]. Any peaks above 75 RFU, not attributed to the known genotype, were considered drop-in, regardless of stutter position [28]. According to the study, drop-in rates were calculated for per locus, sample type, and 28 and 31 PCR cycle protocols. Drop-in was calculated per locus, per replicate; however, the number of single drop-in alleles and the number of drop-in events involving two or more peaks were counted separately [28]. The results of the study indicated that drop-in rates did not vary significantly by locus, number of contributors, or DNA template mass [28] demonstrating that it is not related to the sample, but dependent upon the laboratory within which the DNA sample is processed.

1.5. Aims of Study

The purpose of this study is to determine whether confounding signal observed in forensically relevant electropherograms garnered from bulk-mixtures is significantly different from the signal acquired from single cells. The specific aims are to:

1.) Analyze the electropherograms (EPGs) of 556 single cells.

2.) Characterize the prevalence of known confounding factors such as extraneous signal (allelic drop-in), stutter, and allelic drop-out in single-cell EPGs.

3.) Statistically assess if the distribution of drop-in, stutter, and drop-out effects are significantly different from bulk-processed, high-template samples.

4.) Evaluate methods of collection to establish if single-cell methods are viable for casework-type samples.

2. Methods

2.1. Single-cell Signal Characterization

The first aim of the project was to analyze 556 previously run, single-cell samples using the fragment analysis software OSIRIS 2.10.3 with an analytical threshold of 5 relative fluorescent units (RFU)— the lowest peak height amplitude allowed by the OSIRIS detection software. During this phase, the signal of each cell was compared to the known genotypes to authenticate the sample and to confirm there was no sample switch. If observed, signal artifacts such as spikes and dye blobs were disabled. Spikes appear on electropherograms as several peaks in the same approximate position across all fluorescent dye channels [15] and are believed to be attributed to sharp increases in current during capillary electrophoresis, while dye blobs are the result of unincorporated primer dye that co-migrates during capillary electrophoresis and appear on electropherograms as blunt, wide peaks in the same approximate position across samples [29].

Upon completion of fragment analysis, the data were exported as a data table containing the following: the short tandem repeat (STR) number (i.e., allele designation), approximate fragment size, and fluorescent peak heights for each peak detected.

The second aim of the project was to characterize the prevalence of the following confounding signal: allelic drop-out, allelic drop-in, and stutter from the aforementioned data tables. Each data table was filtered using an automated system used to filter and remove artifacts such as bleed through (pull-up) or minus A from the data tables and is fully described in [30].

Homozygous loci and loci containing Y-specific markers (Y-loci) were removed

from the data set since frequencies of drop-out cannot be effectively measured at these loci [31].

2.1.1. Allelic Drop-out

Recall, allelic drop-out is a false negative detection where one or more alleles are not detected throughout the profile and within a locus [32]. The probability of drop-out was approximated by calculating the false negative detection rate for each contributor at each locus as per,

$$\widehat{\Pr(DO)} = 1 - \left(\frac{N_{PH>5RFU}}{N_{exp}}\right) \quad \text{(Equation 3)}$$

where $N_{PH<5RFU}$ is the number of times the peak heights, PH were less than five RFU (relative fluorescent units) in a known allele position. N_{exp} is the total number of expected alleles, given the known genotype. Only heterozygous loci were used.

Since we examine $N_{PH<5RFU}$, and some of the lower level signal is attributable to noise, this $\widehat{\Pr(DO)}$ is taken to represent the minimum possible value of $\Pr(DO)$. Increasing the signal threshold from 5 RFU will decrease interference from labeling noise but is known to significantly artificially increase $\Pr(DO)$ [33]. As such, we computed the $\Pr(DO)$ at 30 RFU as well, since previous work suggested that 30 RFU is an optimal analytical threshold for reducing instrument noise detection in EPGs [34]. This is corroborated by Peters et al. who expanded this by demonstrating optimal ATs range from 25-45 RFU [8].

Given the hypotheses that drop-out is 1) cell-dependent and 2) sample-dependent, histograms of the number of heterozygous alleles detected for each contributor were plotted. The distributions of the data for each contributor were then compare to each other, one at a time, using the Kolmogorov–Smirnov test (KS test)— a statistical test that

can be used to compare one sample with a reference distribution or compare the distributions of two samples [35]. In this case, the KS test was used to access whether or not two data sets were of the same distribution. Briefly, the KS test determines this by computing *d*, the maximum vertical distance between the cumulative frequency distributions of the two samples and the corresponding p-value (*p*), the probability of the *d* statistic given the null hypothesis is true. Generally, the greater the *d* statistic (and the lower the p-value), the evidence there is to reject that the two data are of the same distribution.

2.1.2. Stutter

Recall, stutter is signal that is often observed in stutter position (one STR unit short of the parent allele) and is attributed to strand slippage that occurs during polymerization of a new DNA strand. Stutter prevalence was characterized by calculating the stutter ratio (*SR*) across all samples, for each locus, as per,

$$SR = \frac{PH_{ST}}{PH_{AL}} \quad \text{(Equation 4)}$$

where PH_{ST} is the peak height of the peak in stutter position and PH_{AL} is the peak height of the allele, given the known genotype.

Therefore, peaks that fell into allele positions that corresponded to the known genotype with peak heights ≥ 30 RFU [8] were stored, as were the corresponding stutter peak heights. Notably, peaks where the stutter positions overlapped with known allele positions were excluded from analysis. For example, if the known genotype at locus, *l*, was 11,12, only stutter from the 11 (i.e., located in position 10) would be used to determine the stutter ratio.

Given that the STR sequence's influence on stuttering during PCR [12], the

stutter ratio distribution was examined on a per-locus basis. Histograms of the stutter ratios were plotted for each locus, and a KS test was employed to determine whether each data set was normally distributed [35].

2.1.3. Allelic Drop-in

Recall, allelic drop-in (extraneous signal) is a false positive detection where signal is observed in a position other than that of a known genotype or stutter position [32]. It is a stochastically occurring phenomenon that is hypothesized to occur when extraneous DNA from the environment is inadvertently co-amplified with the sample [36].

Thus, allelic drop-in was characterized by counting the number of times extraneous signal was detected (see Figure 5(c)) across all 556 samples and by calculating the false positive detection rate as per [37],

$$\Pr(DI) = \frac{\sum_{i=1}^{5} n_i}{\sum_{i=1}^{5} E_i} \quad \text{(Equation 5)}$$

where n is the number of drop-in peaks observed (signal not in stutter position or consistent with the known genotype) per contributor and *E* is the expected number of empty signal positions for each contributor [36].

A histogram of the drop-in's peak heights across all 556 samples were plotted, and a KS test was employed to determine if the data set were normally distributed [35].

2.2. Methods of Collection: Single-cell Desorption

The final aim of the project was to examine the percent recovery for the desorption of cells from cotton-tipped applicators, the substrate most commonly used to collect biological samples.

2.2.1. Accessing the Cell Density in Whole Saliva Samples

Eight distinct single-source saliva samples were used throughout the course of this study.

Each sample was vortexed and 6 µL of whole saliva was pipetted into one of four chambers located on the Bulldog Bio® 4-chip disposable hemocytometer (Figure 9) [Bulldog Bio, Portsmouth, NH].

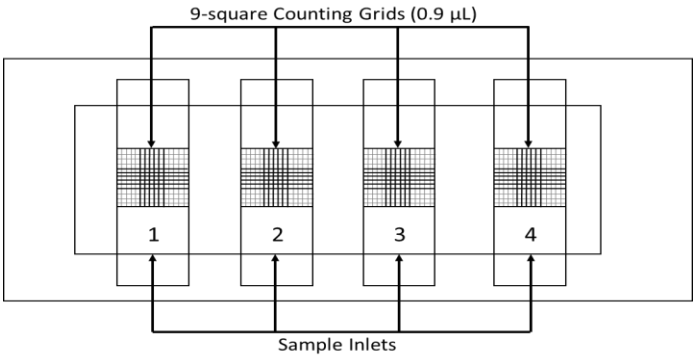

Figure 9: A visual representation of the 4-chip disposable hemocytometer adapted from the Bulldog Bio® 4-chip Disposable Hemocytometer User Manual [38]. There are four chambers located on the hemocytometer, each including one sample inlet and one 9-square grid pattern (0.9 µL).

The buccal cells suspended in whole saliva were manually counted using a Nepagene® Compass Video Microscope at x52 magnification. The cells that fell within the entire, 0.9 µL, 9-square patterned area on the Bulldog Bio® 4-chip disposable hemocytometer (see Figure 9) were counted [38].

Upon microscopic examination, if the sample exhibited cell-clumping (i.e. cell counting was not viable), a dilution of the sample was prepared and used for cell counting.

2.2.2. Desorption of Cells from Cotton and Cell Counting

Each saliva sample was vortexed and 150 µL of whole saliva, which is the maximum volume absorbed by the cotton, was pipetted onto one sterile, Fisherbrand® cotton-tipped applicator, and allowed to dry for at least 24-hours.

The entire dried, cotton tip was cut from the wooden applicator using a sterile, disposable scalpel and transferred to a 2 mL microcentrifuge tube with 450 µL of 10 mM Tris, 1 mM EDTA TE buffer (pH 8.0). The sample was briefly vortexed, placed on an orbital shaker, and allowed to incubate with continuous shaking for one hour at 25ºC.

The cotton substrate was removed from the microcentrifuge tube with sterilized tweezers, and placed into a sterile Costar® Spin-X® insert [Corning Incorporated, Tewksbury, MA]. The insert containing the cotton substrate was placed back into the original microcentrifuge tube containing the eluate and centrifuged for two minutes at maximum angular velocity.

The mesh insert now containing the dried cotton substrate was discarded. Each sample was vortexed and 6 µL of eluate was aliquoted into one of the four chambers located on the Bulldog Bio® 4-chip disposable hemocytometer as described previously.

The buccal cells desorbed from the cotton were then counted, microscopically, using a Nepagene® Compass Video Microscope at x52 magnification. The cells that fell within the entire, 0.9 µL, 9-square area on the Bulldog Bio® 4-chip disposable hemocytometer (see Figure 9) were counted.

Baseline concentrations of buccal cells/µL were determined by dividing the cell count, *h*, by 0.9 µL (the total volume of the of the 9-square grid pattern). Note that any samples exhibiting cell clumping were diluted by adding one volume fraction of saliva to

two volume fractions of buffer; therefore, if necessary, concentrations were back-calculated by multiplying $\frac{h}{0.9\ \mu L}$ by a factor of 3. Similarly, one volume fraction of saliva was aliquoted onto the cotton-tipped applicator and incubated in two volume fractions of buffer. Therefore, the concentration of buccal cells desorbed from cotton-tipped applicators were also back-calculated by multiplying $\frac{h}{0.9\ \mu L}$ by a factor of 3.

Percent recovery was determined by dividing the concentration of buccal cells desorbed from the cotton substrate by the respective baseline concentration of the sample.

3. Results and Discussion

3.1. Statistical Assessment of the Distribution of Confounding Signal

3.1.1. Allelic Drop-out

The probability of drop-out was approximated by calculating the false negative detection rate for each contributor as per, Equation 4 (see Table 7).

Table 6 contains the probability of drop out for each contributor (Persons A through E (referenced in Figure 10) that were calculated by dividing the total number of heterozygous repeat alleles detected (over 5 RFU) by the total number of expected heterozygous alleles based on the known genotype of the contributor.

Table 5: A table containing the probability of drop out for each contributor (Persons A through E referenced in Figure 10) that were calculated by dividing the total number of heterozygous repeat alleles detected (over 5 RFU) by the total number of expected heterozygous alleles based on the known genotype of the contributor.

Person (A- E)	Total No. of Heterozygous Repeat Alleles Detected (over 5 RFU)	Total No. of Expected Heterozygous Alleles	Probability of Drop-Out $(\widehat{\Pr(DO)} = 1 - (\frac{N_{PH>5RFU}}{N_{exp}})$
A	2,729	3,876	.30
B	2,748	3,648	.25
C	1,471	3,638	.60
D	1,773	3,876	.54
E	2,460	4,280	.43

* The total number of expected heterozygous alleles varied per contributor due to varying numbers of heterozygous loci and number of samples per contributor.

Figure 10 are histograms for the number of heterozygous alleles detected for a given cell, separated by contributor, where Contributor A, C and D (shown in Figures 10(a), (c), and (d)) had 34 potential heterozygous alleles that could be detected. Based on the known genotype, Contributors B and E (shown in Figures 10(b) and (e)) had a maximum of 32 and 40 heterozygous alleles that could be detected, respectively. The

number of cells tested per contributor ranged from 107 to 114. The expected binomial distribution using the empirically determined drop-out rate (see Table 6, column 4) is also plotted.

Qualitatively, the histograms exhibit some interesting features. First, it is notable that obtaining a full profile is a rare event; that is of the 556 single-cell buccal cells, only 28 (5%) resulted in full profile representations, suggesting that allelic drop-out is not a rare event. This is consistent with the findings of Williamson et al. [10], but inconsistent with the findings of Geng et al. who produced full profiles [39]. The difference between the two studies were the cell types and conditions applied. Specifically, Geng et al., extracted and amplified the DNA of human lymphoid cells that were grown in a medium under controlled conditions, while Williamson et al. extracted and amplified the DNA of buccal, sperm, and blood cells contained on cotton-tipped applicators. The samples tested by Williamson et al. are more consistent with sample qualities received by the forensic laboratory. Similar to the Williamson et al. study, Findlay et al. [9] analyzed the STR profiles of 226 single buccal cells from 4 different contributors amplified for 34 cycles, and showed an overall drop-out rate of 39% where only 114 of 226 (50%) cells produced full profiles [9]. Findlay et al. also observed complete drop-out in 20 of 226 (9%) of single-cell profiles, which is consistent with the second feature observed in Figure 10, where 81 of 556 (15%) cells resulted in complete profile drop-out and 28% resulted in fewer than eight alleles detected.

The third feature is that, interestingly, the non-zero mode for Contributors 1 through 5 is 31, 29, 30, 32, and 34, suggesting that the information content contained in a single-cell electropherogram is likely to be very high or very low. Thus, even the most

temperamental of cell types (i.e., buccal cells) is likely to provide enough signal to produce adequate weights against true contributors as long as, at least, a few cells are made available for testing. These finding were also supported by Findlay et al. [9], where ≥4 alleles were detected in 64% of the profiles, and <4 alleles were detected in 27% of the profiles.

The last interesting feature is that there seem to exist differences between the distributions across contributors. The left tailing and the high frequencies at 0 in Figure 10(a) through (e) suggest that these data cannot be easily explicated by random sampling as predicted by Equation 2. If the drop-out of alleles was an independent phenomenon, then the expectation is that most of the cells would exhibit allele drop-out on 32, 34, or 40 trials and the empirical non-detect rates shown in Table 6. These expectations are also plotted on Figure 10(a) through (e).

A KS test was performed to determine whether the allele detection rate (blue bars, Figure 10) for Persons A through E were of the same distribution as the expected binomial distribution (orange bars, Figure 10), using the $\Pr(DO)$ reported in Table 5. All of the p-values were determined to be zero, indicating that they were not of the same distribution, respectively (see Figure 10).

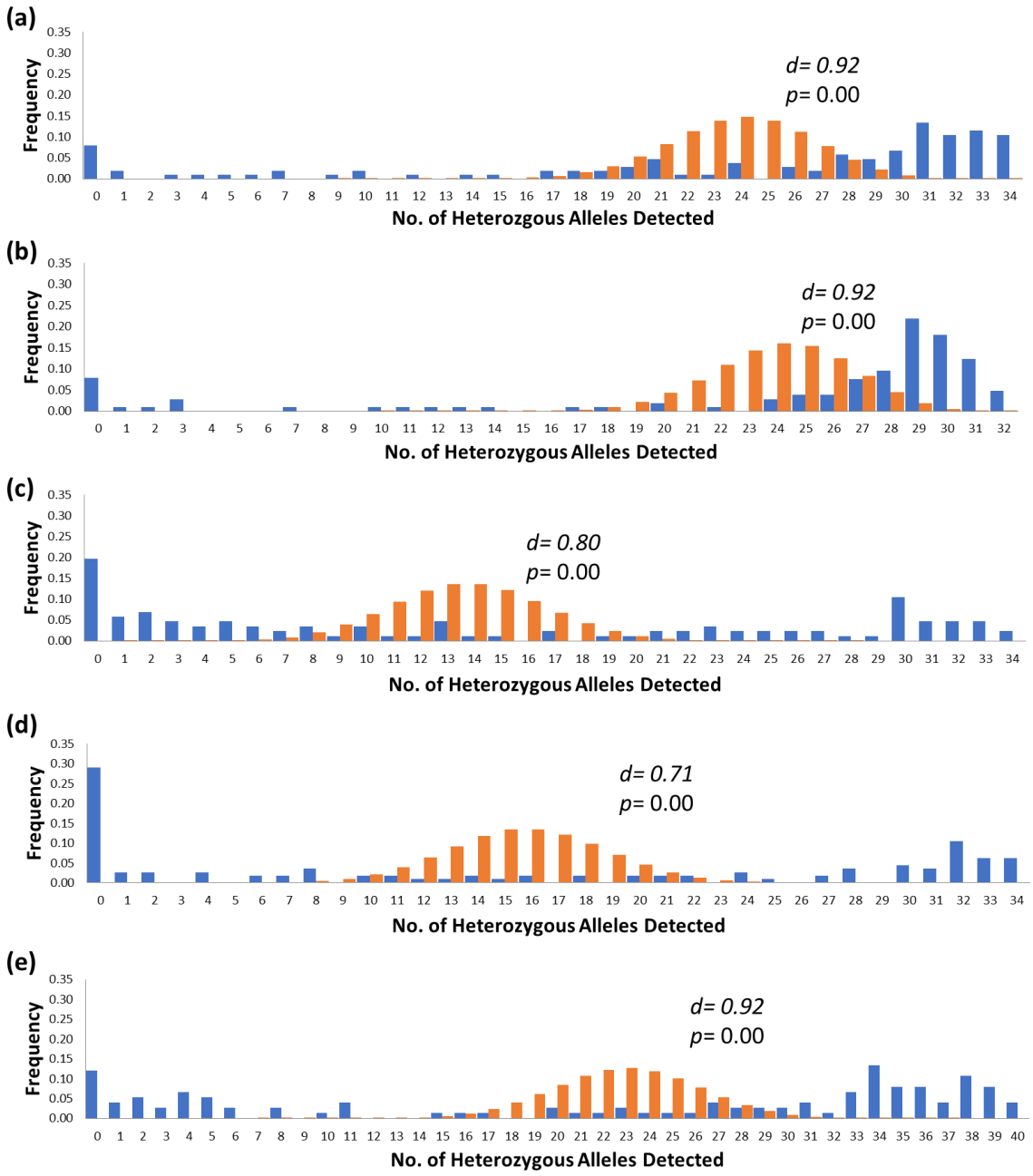

Figure 10: A plot of frequency versus number of heterozygous alleles detected for five contributors: A thru E, respectively (laboratory conditions: Globalfiler™ amplification, for 30 PCR cycles, 25-second injection on 3500 CE) from 556 single-cell, single- source samples. Figures 10(a), 10(c), and 10(d) depict allele detection for contributors with 34 heterozygous alleles per profile (■). Figures 10(b), 10(e) depict allele detection for contributors with 32 and 34 heterozygous alleles per profile, respectively(■). The theoretical, binomial distribution (■) is plotted and statistically compared to actual allele detection distributions for each contributor, respectively. The results of the KS test (i.e. d-statistic and p-value) are shown in Figures 10(a)-(e).

A KS test was also performed to determine whether the allele detection rates *between* contributors was of the same distribution. Figure 11 illustrates four representative plots (cumulative (Σ) frequency versus number of heterozygous alleles), generated from the KS test, comparing the data of two individuals at a time (Persons A/B or A/E). In Figure 11(a), the cumulative frequency of allele detection is compared between Person A and Person B, where d=0.26 and p=0.00. In Figure 11(b), the cumulative frequency of allele detection is compared between Person A and Person E, where d=0.30 and p=0.00. When the data sets of each contributor were compared to each other, two at a time, the *d*-statistic ranged from 0.13 to 0.47 and the p-values ranged between 0.00 and 0.25 (see Table 7), suggesting there is some evidence to support the proposition that allelic drop-out from buccal cells may be person-dependent. Further studies would be required before probabilistic model recommendations can be reported.

Table 6: A table containing the of KS test results: d-statistic and p-values, for all five contributor samples that where compared to each other, one at a time.

Samples	d-statistic	p-value
A/B	0.26	0.00
A/C	0.39	0.00
A/D	0.33	0.00
A/E	0.30	0.00
B/C	0.47	0.00
B/D	0.39	0.00
B/E	0.44	0.00
C/D	0.13	0.25
C/E	0.38	0.00
D/E	0.33	0.00

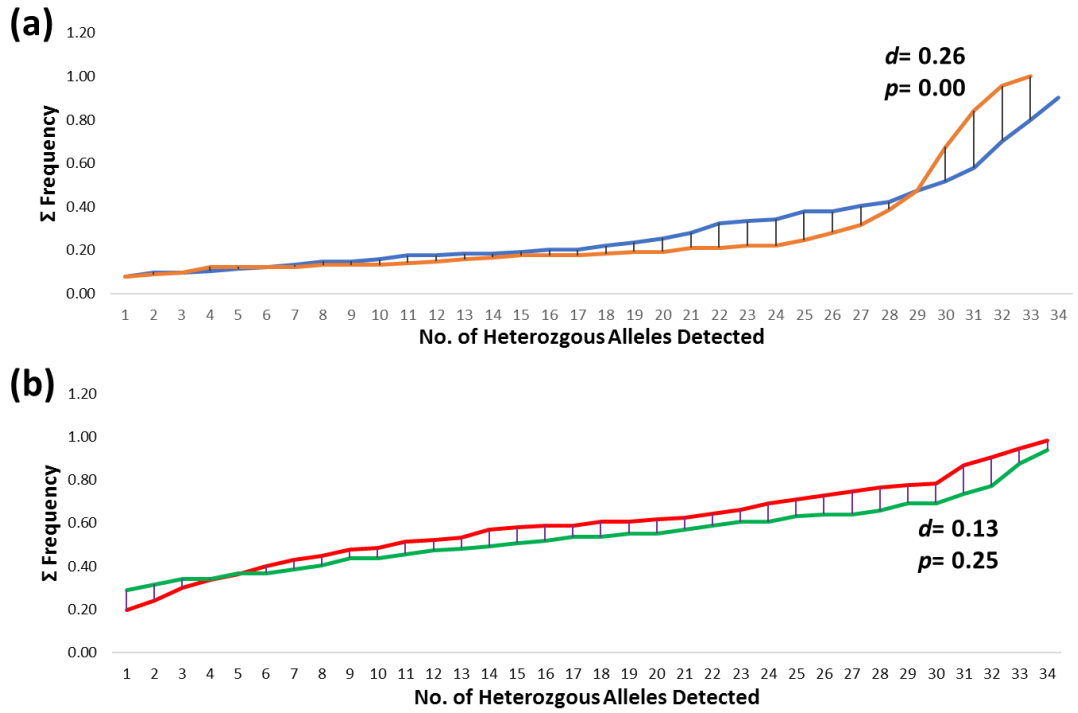

Figure 11: Two plots (cumulative (Σ) frequency versus number of heterozygous alleles), generated from the KS test, comparing the data of two individuals at once ((■)Person A; (■)Person B; (■)Person B ; and (■)Person D). In Figure 11(a), the cumulative frequency of allele detection is compared between Person A and Person B, where *d*=0.26 and *p*=0.00. In Figure 11(b), the cumulative frequency of allele detection is compared between Person C and Person D, where *d*=0.13and *p*=0.25.

Given the variability of allele detection for each replicate within a sample, these data support the hypothesis that allelic drop-out is not cell independent, indicating the need of a new interpretation strategy for single-cell pipelines.

3.1.2. Stutter

Previous research suggested that the Adenine-Thymine (A-T) base pair content of a sequence influences the amount of stutter product due to the weaker A-T double hydrogen bond compared to the triple hydrogen bonding of Guanine-Cytosine (G-C). It was also suggested that stutter is correlated to the length of the allele (repeat number) [26];[13]. Brooks et al. expanded on this research and demonstrated that the longer the uninterrupted sequence (LUS) within an STR-type, the higher the stutter ratios [12]. For

example, the TPOX locus contains a simple repeat structure of consecutive, uninterrupted tetranucleotides of the same sequence, and depending on the length of the target fragment (number of STRs for the individual), it is predicted that the TPOX locus will exhibit more stutter signal than that of shorter STR-types. Another factor known to impact stuttering is the length of the repeat unit itself. According to Mulero et al., stutter products decrease as the number of base pairs in the core repeat unit become longer; therefore, trinucleotides tend to stutter more than their tetra-, penta-, and hexanucleotide counterparts; however, stutter also increases as the number of STR repeats increase [40]. A study conducted by Mulero et al., demonstrated that reverse stutter of the trinucleotide, DYS392 locus increased as the number of T-A-T STR repeats increased [40].

Table 7 contains the repeat structure of exemplar STR sequences of four forensically relevant STR loci: TPOX, SE33, D22S1045, and vWA, for Contributors A through E. The TPOX locus contains a simple repeat structure of consecutive, uninterrupted tetranucleotides of the same sequence ranging from 8 to 11 repeats, depending on the contributor. The SE33 locus has a very complex repeat structure: four tetranucleotide repeats with several dinucleotide inserts, including a rather long LUS. D22S1045 has a repeat structure containing alternating trinucleotide repeats with a variable LUS length, while the vWA locus contains a moderately complex sequence with two alternating tetranucleotide repeats and a relatively long LUS.

Table 7: A table containing the repeat structure of the four representative loci depicted in Figure 12: TPOX, SE33, D22S1045, and vWA; observed alleles for each contributor (A-E); and the associated repeat structures.

Locus	Observed Alleles (Contributors A-E)	Repeat Structure
TPOX	8, 9, 10 ,11	**[AATG]$_{8\text{-}11}$**
SE33	**16**, 19, 20, 22, 25.2, 26.2, 27.2, 31.2	$[AAAG]_2$ AG $[AAAG]_3$ AG **[AAAG]$_{16}$** G $[AAAG]_3$ AG**
D22S104	11, 15, 16, 17	**[ATT]$_{8,12\text{-}15}$** ACT $[ATT]_2$
vWA	15, 16, 17, 18, 19	TCTA $[TCTG]_4$ **[TCTA]$_{10\text{-}14}$**

Bold = representative LUS for Contributors A-E. **The repeat structure for the SE33 locus is highly complex and only the 16 allele is represented for SE33. For the structure for each allele see the National Institute of Standards and Technology (NIST) Short Tandem Repeat DNA Internet Database [41].

Given that the STR sequences influence stuttering during PCR, we examined stutter ratio distribution on a per-locus basis. In Figure 12, are stutter ratios for four representative loci: vWA, TPOX, D22S1045, and SE33—each including a y-axis expanded figure insert to better exhibit frequencies associated with the maximum stutter ratios. Figure 12(a) is a histogram of stutter ratios for vWA, a locus with a moderately complex sequence structure. Figure 12(b) illustrates the stutter ratio frequencies of TPOX, a locus with a simple sequence structure. Figure 12(c) illustrates the stutter ratio frequencies of D22S1045, a locus with a trinucleotide sequence structure. Figure 12(d) illustrates the stutter ratio frequencies of SE33, a locus with a complex sequence structure. For all plots, there is a very high frequency of stutter ratios between 0.00 and 0.01, which represents stutter signal not observed—suggesting most of the single-cell allele signal produced stutter that did not exceed the maximum detectable signal associated with the process. The result is contrasted by the stutter signal acquired at the high template regime, wherein nearly all samples, exhibited non-zero stutter events [42].

Thus, unlike bulk-mixture probabilistic systems that assume stutter is regularly detected, single-cell interpretation strategies will need to employ probabilities of not detecting stutter or employ strategies which are impervious to such effects.

Although TPOX has a relatively simple repeat structure, the known genotypes of the contributors were on average 9 repeat units in length. This shorter LUS length may account for why a lower instance of stutter was observed in TPOX (median stutter ratio: 5%) compared to SE33 (median stutter ratio: 13%) that has a more complex repeat structure, but includes a rather large LUS averaging 23 repeats for these data (see Figures 12(b) and 12(d)). Similarly, vWA (median stutter ratio: 11%) has a moderately complex sequence with a larger LUS of approximately 15 STRs which may attribute to the increase in stutter ((see Figure 12(a)). D22S1045 (median stutter ratio: 10%) has a trinucleotide STR with an average length of 15 per contributor, which may account for the increased stutter ratios compared to that of TPOX; thus, like currently available forensic interpretation systems for bulk-mixture samples, single-cell samples will also require a per-locus or even per-allele model for inference.

Another important feature shown in Figures 12(a) through 12(d), is that all the distributions appear to have a right tail for the non-zero stutter values. In particular, Figure 12(d) shows one sample had an abnormally high stutter ratio of >100%. In other loci, not depicted in Figure 12, the maximum stutter ratio was as high as 2.41, meaning stutter signal amplitude was 241% higher than that of the parent allele. This is inconsistent with high-template stutter ratios reported by the Globalfiler™ PCR Amplification User Guide [26], where the stutter ratios that ranged from 3.9% to 16.26%, depending upon the STR length of the locus. According to a study conducted by Brochu

et al., high template buccal samples, directly amplified with GlobalFiler™ Express PCR Amplification Kit at 26 cycles and 5-25 second 1.2 kV CE 3500 injection, demonstrated average stutter ratios of 7% to 9% across samples [43].

As previously discussed and predicted using an *in-silico* model (see Table 2), this may be attributed to an early occurrence of stutter during the amplification process where stutter products were amplified in the numerous subsequent PCR cycles, and comparatively few true-sized amplicons are simultaneously synthesized. Recall, in the case of high-template samples, the stutter peaks are not expected to impact the signal in the same way, since stuttering early in cycling is expected to be a rare event.

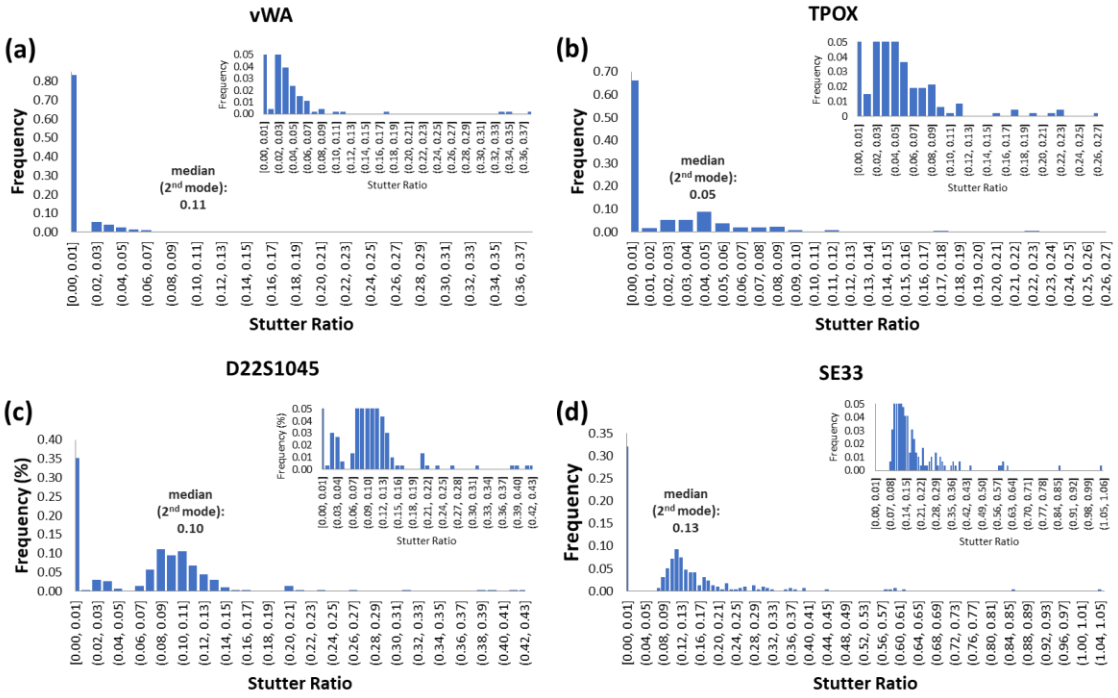

Figure 12: A plot of frequency against stutter ratio for four representative loci: vWA, TPOX, D22S1045, and SE33 (laboratory conditions: Globalfiler™ amplification, for 30 PCR cycles, 25-second injection on 3500 CE) from 556 single-cell, single- source samples across five contributors—each including a zoomed figure insert to show the maximum stutter ratios. Figure 12(a) illustrates the stutter ratio frequencies of vWA. Figure 9(b) illustrates the stutter ratio frequencies of TPOX. Figure 12(c) illustrates the stutter ratio frequencies of D22S1045. Figure 12(d) illustrates the stutter ratio frequencies of SE33. Note: stutter ratios between 0.00 and 0.01 represent signal that was not observed above 5 RFU— the lowest peak height amplitude allowed by the OSIRIS detection software.

To statistically confirm the distributions were not normal, a KS test [35] was employed for the distribution of stutter ratios for twenty loci and all of the p-values were determined to be zero, indicating that distribution of the stutter ratios for all twenty loci were, as expected, not normal.

3.1.3. Allelic Drop-in

In this study, the probability of allelic drop-in was determined to be 1.4×10^{-3} across the 556 single-source, single-cell samples, as calculated by Equation 6. Since the probability of drop-in has been characterized as a rare event in bulk-processed systems [18];[28];[44], it appears to be consistent with that of the single-cell regime. According to Gill et al., there is no absolute method to determine whether drop-in or contamination has occurred, however, drop-in can be distinguished from contamination by examining the electropherograms themselves. Contamination often appears as two or more allele signals per locus from a single contributor; whereas, drop-in alleles appear as random signal from different contributors [18].

As previously described by [28], drop-in does not vary significantly by locus, number of contributors, or DNA template mass in high template samples [28], and is a rare event [36];[44].

Of all 556 single-cell signal examined in this study, 120 peaks extraneous to the profile exhibited peak heights greater than 5 RFU. Figure 13 plots the histogram of the peak heights associated with these 120 drop-in peaks. The plot appears to have a right tailed distribution with a mode of 11 RFU and median of 35 RFU. Recall, that it is difficult to distinguish allele signal from noise signal under 30 RFU [34] and since 5 RFU is the lowest detectable peak height amplitude [26], any peak height between 6 and 30

RFU may be attributed to noise. Similarly, extraneous signal with peak heights over 30 RFU are more likely to contain allele signal. We note that the peak heights associated with these random signals can be quite substantive, reaching intensities of thousands of RFUs.

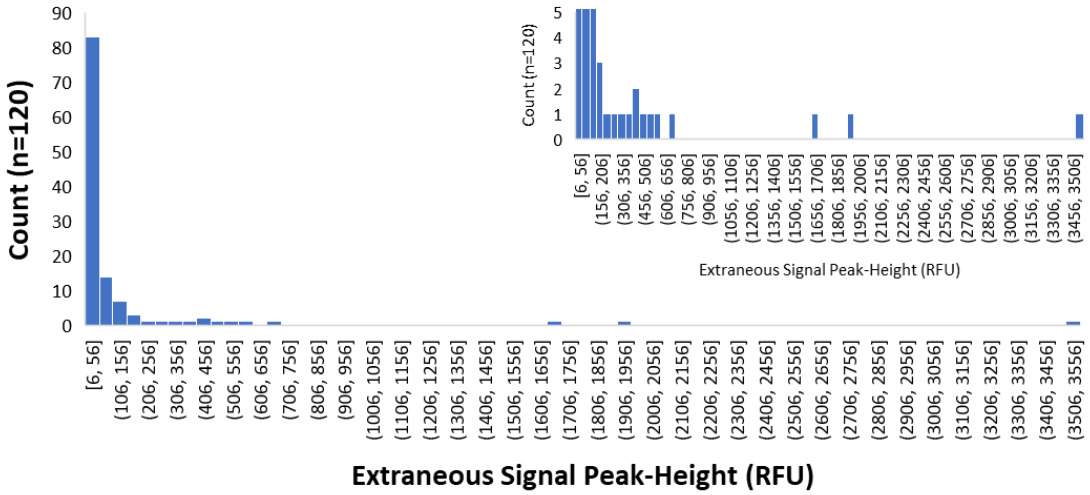

Figure 13: A plot of count (out of 120 observed extraneous signal) against extraneous signal peak heights (laboratory conditions: Globalfiler™ amplification, for 30 PCR cycles, 25-second injection on 3500 CE) from 556 single-cell, single- source samples across five contributors, including a zoomed figure insert to show the maximum extraneous signal peak heights.

Recall, the KS test can be used to compare one sample with a reference distribution to determine normality [35]. This test was conducted for the distribution of the extraneous signal peak heights and the p-value of 0.00—indicating that the distribution was not normal. Thus, any single-cell interpretation system will likely be required to use methods which are non-parametric or use a log-normal distribution assumption as shown by [45]. In brief, the aforementioned study explored potential nuances associated with single-cell signal and demonstrated that there is a very good chance a single-cell will render interpretable, forensically relevant STR results; however, current probabilistic genotyping systems are ill-equipped to properly handle the full

interpretation of these signals as they assume drop-out is independent; reverse stutter detection is necessarily detectable; and the probability of drop-in is constant [46].

3.2. Percent Recovery of Single Cells from Cotton-Tipped Applicators

Though single-cell forensic DNA analysis has recently resurged as an area of interest, these results suggest appropriate interpretation paradigms will need to be constructed if single-cell pipelines are to translate to operation. Integration of strategies into forensic operations will also require studies to confirm success of or the re-design typical cell collection.

As such, the second phase of this work examines the cell recovery rates associated with the most common of collection strategies— collection using sterile, cotton-tipped applicators.

Figure 14 is a box plot of the percent recovery of buccal cells desorbed from cotton-tipped applicators for eight different contributors (09 through 16), labeled with the baseline concentration of buccal cells/µL for each contributor. The percent recovery varied between 34 and 79 percent for most of the contributors' samples, except for that of Contributors 12 and 15 which demonstrated instances of zero percent recovery and a far broader range of cells recovered.

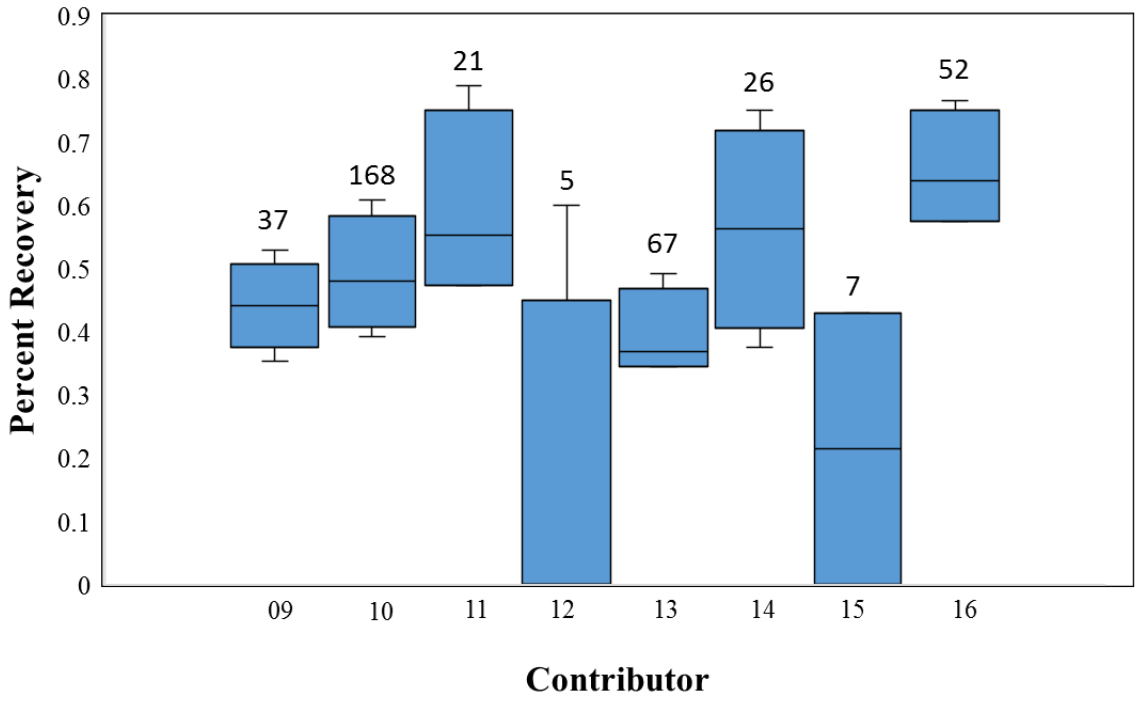

Figure 14: A box plot of the percent recovery of buccal cells desorbed from cotton-tipped applicators for eight different contributors (09 through 16). The number above the box plot is the concentration of buccal cells/µL of whole saliva and represents the baseline concentration of cells for the contributor.

Upon initial observation (see Figure 14), there appeared to be correlation between the baseline concentration of buccal cells/µL whole saliva and percent recovery. To further investigate the potential correlation between the two variables, a scatter plot (see Figure 15) of percent recovery against the baseline concentration of buccal cells (cells/µL saliva) for whole saliva was plotted. The correlation coefficient, r, of the two variables is 0.27—indicating little correlation between the percent recovery and initial starting concentration of cells. This figure also demonstrates the variability of all contributor's baseline concentration of buccal cells, with two of the contributors exhibiting a very low concentration of cells.

Varying buccal cell concentrations may be attributed to individuals based on gender, age, and/or overall health according to [47]. This 2007 study accessed the DNA yield and purity of buccal cells collected from both male and female subjects (via cotton swab), of varying ethnicities from ages 28.5–92.5, some of which were reported to be taking diuretic drugs [47]. The buccal cell samples were then organically extracted and the DNA yield was quantified via spectrophotometry. The total DNA yield ranged from 0.08 to 1078.0 µg (median 54.3 µg; mean 82.2 µg ± SD 92.6 with samples originating from male subjects yielding significantly more DNA (median 58.7 µg) than that of female subjects (median 44.2 µg) [47]. Diuretic drug users had significantly lower DNA purity (median 1.92), and older age was also reported to be associated with lower DNA purity [47]. Given that each buccal cell is estimated to contain .0063 ng of DNA, this study indicates cell quantity and quality vary across individuals.

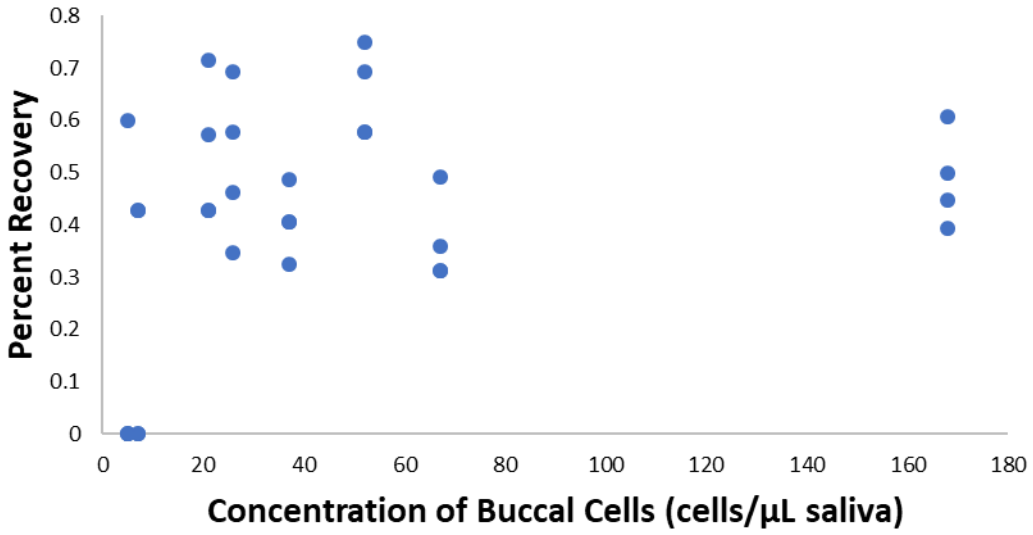

Figure 15: A scatter plot of percent recovery (desorbed from cotton-tipped applicators for eight different contributors, four repeats each) against the baseline concentration of buccal cells (cells/µL saliva) for whole saliva. The correlation coefficient, *r*, of the two variables is 0.27.

If the data of Contributors 12 and 15 are excluded as outliers given low baseline concentrations of buccal cells; the mean, minimum, maximum, and standard deviation of the percent recoveries are 0.52, 0.34, 0.79, and 0.13, respectively.

The mean 52% recovery rate for single cells is consistent with an analogous study exploring the percent recovery of bulk-processed extracted samples [48]. In the study conducted by Adamowicz et al., a known concentration of DNA from a set volume of buccal cell suspension was extracted (QIAamp™ extraction) from cotton swabs, with similar incubation conditions outlined in Section 2.2.2. The concentration of the extract was quantitatively determined to be 0.54 ± 0.06 [48].

Although the cotton-swab collection method is one of the mostly commonly used techniques, other methods of evidence collection methods have been explored, e.g., adhesive tape collection [49]. Li et al. examined the recovery of DNA from cuttings of hydrophilic adhesive tape (HAT) used to collect surface cells from various parts of the human body, i.e., the ankle, the arm, behind the ear, fingers, and back of the neck of several individuals. The DNA was extracted from the cuttings with Chelex® and amplified using 4-loci, COfiler™ and Profiler Plus™ amplification kits, where overall success rates where based on the rate of allele detection of four loci: vWA, TH01, F13A1, and FES, [49]. Allele detection rates varied between sample-collection locations: 30.6% (arm), 58.3% (neck), 65.5% (finger) and 100% (ear) [49], and the authors hypothesize that the low success rates were attributed to a low yields of template DNA and stochastic effects [49]. Although this proposed methodology may have practical applications for trace and/or touch DNA evidence from victims, it may not be applicable for the collection of evidence from other more common crime scene surfaces, and given

the vast fluctuation of success rates, it may not be a reliable in its application to forensic casework.

Stouder et al. explored another method of trace evidence collection, e.g., scraping, where various items of clothing, previously laundered, were worn by participants for one day and trace evidence debris was scraped from these items, transferred to the item's respective pillbox, where it was later swabbed for downstream processing [50]. These same items were also swabbed for trace evidence using sterile cotton swabs and sterile water, and were referred to by the authors as friction swabs [50]. The DNA from both evidence sets were organically extracted, quantified using ACES™ Human DNA Quantification Probe Plus Kit, amplified using the AmpFlSTR® Profiler Plus™ Amplification Kit, and typed by capillary electrophoresis on an ABI Prism™ 310 Genetic Analyzer [50]. The study demonstrated that both the DNA recovered from friction swabs and trace evidence debris from the pillboxes contained suitable quantities of DNA for downstream DNA analysis, and that the quantity of DNA recovered from the pillbox swab was equal to or greater than that of the friction swab in 9 out of the 11 items analyzed, and the average amount of DNA recovered from the friction swabs was 4 ng compared to the 21 ng of DNA recovered from the pillboxes [50]. Although the scraping method appears to be an ideal method of trace evidence collection in the laboratory, it may not be practically applied in a crime scene setting due to risks of airborne contamination.

Given the results of other methods-of-collection, the cotton-swab method seems to be applicable to a wider range of crime scene samples and surfaces, though it remains that a 48% loss of cells was observed in this study and a 46% loss of DNA was observed

in an analogous study [48], suggesting continued efforts to improve cell collection and recovery by modifying the material properties of the collection material are justified.

Voorhees et al. explored the elution of cells from cotton-tipped applicators [51]. The authors used scanning electron microscopy to demonstrate that the dimensions of sperm and epithelial cells are quite small compared that of cotton-fibers, and that it was unlikely that the cells were physically blocked from eluting [51]. Rather, Voorhees et al. attributed the poor cell recovery rates to the fragility of the cell's nuclear membrane and lysis during incubation and elution [51] .

Given its versatility and unrivaled ability to adsorb forensically relevant biological material from any number of substrates, collection with cotton swabs seems a reasonable means by which to collect evidence.

4. Conclusions

Given the complexity of bulk-mixture interpretation, single-cell analysis has presented the forensic science community with a means to fully deconvolve DNA mixtures. In order to determine if single-cell systems are a viable alternative, confounding signal such as allelic drop-out, stutter, and allelic drop-in were explored and characterized.

State-of-the-art probabilistic genotyping software such STRMix™ and TrueAllele™ are currently available to the forensic science community and probabilistically compute the weight-of-evidence against a person-of-interest; however, they make assumptions which are based on bulk-processing [52];[4]. Our exploratory research indicates these softwares are unlikely to be effectively applied to single-cell systems in their current form. Thus, information regarding the cell dependency of allelic-drop out, increased stutter, and the characterization of extraneous signal in the single-cell regime are all crucial in the development of single-cell inference systems. If successful, the implementation of a single-cell strategy would lead to strengthened likelihood ratios/match statistics, potentially improving inclusion and exclusion decisions for comparisons with reference samples.

Though improvements to collection techniques are always justified, preliminary results described in Section 3.2. confirms that typical collection techniques are a reasonable method by which to collect cells from the environment—as approximately 52% of the cells are made available to downstream processing by these methods. Given only singles of cells are needed to produce a near-full forensic profile, the cotton-swab collection methods are reasonable for operations' purposes.

In summary, this work demonstrates that successful single-cell implementation into forensic operations requires probabilistic model development and the engineering of new inference constructs.

BIBLIOGRAPHY

1. Koreth, J., J.J. O'Leary, and O.D.M. J, *Microsatellites and PCR genomic analysis.* J Pathol, 1996. **178**(3): p. 239-48.

2. Fan, H. and J.Y. Chu, *A brief review of short tandem repeat mutation.* Genomics Proteomics Bioinformatics, 2007. **5**(1): p. 7-14.

3. Swaminathan, H., et al., *CEESIt: A computational tool for the interpretation of STR mixtures.* Forensic science international. Genetics, 2016. **22**: p. 149-160.

4. Perlin, M.W., et al., *Validating TrueAllele(R) DNA mixture interpretation.* J Forensic Sci, 2011. **56**(6): p. 1430-47.

5. *Report to the President Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. 2016; September 20, 2006:[Available from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

6. Bille, T.W., et al., *Comparison of the performance of different models for the interpretation of low level mixed DNA profiles.* Electrophoresis, 2014. **35**(21-22): p. 3125-33.

7. Swaminathan, H., et al., *Four model variants within a continuous forensic DNA mixture interpretation framework: Effects on evidential inference and reporting.* PLOS ONE, 2018. **13**(11): p. e0207599.

8. Peters, K.C., et al., *Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: Standardizing the forensic validation process.* Forensic Sci Int Genet, 2017. **31**: p. 160-170.

9. Findlay, I., et al., *DNA fingerprinting from single cells.* Nature, 1997. **389**(6651): p. 555.

10. Williamson, V.R., et al., *Enhanced DNA mixture deconvolution of sexual offense samples using the DEPArray™ system.* Forensic Science International: Genetics, 2018. **34**: p. 265-276.

11. Chaolong, W., N.A. Rosenberg, and K.B. Schroeder, *A maximum-likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes .(Report)(Author abstract).* Genetics, 2012. **192**(2): p. 651.

12. Brookes, C., et al., *Characterising stutter in forensic STR multiplexes.* Forensic Sci Int Genet, 2012. **6**(1): p. 58-63.

13. Bright, J.A., et al., *Modeling forward stutter: Toward increased objectivity in forensic DNA interpretation.* ELECTROPHORESIS, 2014. **35**(21-22): p. 3152-3157.

14. Walsh, P.S., Fildes, N. J., Reynolds, R., *Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA.* Nucleic acids research, 1996. **24**(14): p. 2807-2812.

15. Butler, J.M., *Advanced topics in forensic dna typing : interpretation*. 2015, Oxford, England ;: Academic Press.

16. Gill, P., et al., *DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures.* Forensic Sci Int, 2006. **160**(2-3): p. 90-101.

17. Duffy, K.R., et al., *Exploring STR signal in the single- and multicopy number regimes: Deductions from an in silico model of the entire DNA laboratory process.* Electrophoresis, 2017. **38**(6): p. 855-868.

18. Gill, P., et al., *DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods.* Forensic science international. Genetics, 2012. **6**(6): p. 679-688.

19. Haned, H., Pene, L., Lobry, J. R., Dufour, A. B., Pontier, D., *Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count?* J Forensic Sci, 2011. **56**(1): p. 23-8.

20. Marciano, M.A., Adelman, Jonathan D., *PACE: Probabilistic Assessment for Contributor Estimation— A machine learning-based assessment of the number of contributors in DNA mixtures.* Forensic Science International: Genetics, 2017. **27**: p. 82-91.

21. Taylor, D., Bright, J. A., Buckleton, J., *Interpreting forensic DNA profiling evidence without specifying the number of contributors.* Forensic Sci Int Genet, 2014. **13**: p. 269-80.

22. Swaminathan, H., Grgicak, C. M., Medard, M., Lun, D. S., *NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping.* Forensic Sci Int Genet, 2015. **16**: p. 172-180.

23. Taranow, L., Grgicak, Catherine M., Forry, Erin, Swaminathan, Harish, *Exploring the sources of peak height reduction during low-template, compromised DNA data analysis*. 2016, ProQuest Dissertations Publishing.

24. Grgicak, C.M., *Personal Communication*. 2018, August 8.

25. Bright, J.-A., P. Gill, and J. Buckleton, *Composite profiles in DNA analysis.* Forensic Science International: Genetics, 2012. **6**(3): p. 317-321.

26. Thermo Fisher Scientific, *GlobalFiler™ PCR Amplification Kit: User Guide*. 2015, Life Technologies Corporation: Carlsbad, CA.

27. Gill, P., Curran, James, Elliot, Keith, *A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci.* Nucleic Acids Research, 2005. **33**(2): p. 632-643.

28. Mitchell, A.A., Tamariz, Jeannie, O'Connell, Kathleen, Ducasse, Nubia, Budimlija, Zoran, Prinz, Mechthild and T. Caragine, *Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in.* Forensic Science International: Genetics, 2012. **6**(6): p. 749-761.

29. Butler, J.M., *Fundamentals of Forensic DNA Typing*. 2010, Amsterdam: Academic Press/Elsevier.

30. Alfonse, L.E., et al., *The Development and Release of a Collection of Computational Tools and a Large-Scale Empirical Data Set for Validation: The PROVEDIt Initiative*. 2016.

31. Johnson, P.C.D. and D.T. Haydon, *Maximum-likelihood estimation of allelic dropout and false allele error rates from microsatellite genotypes in the absence of reference data.* Genetics, 2007. **175**(2): p. 827-842.

32. Haned, H., K. Slooten, and P. Gill, *Exploratory data analysis for the interpretation of low template DNA mixtures.* Forensic Science International: Genetics, 2012. **6**(6): p. 762-774.

33. Rakay, C.A., Bregu, Joli, Grgicak, Catherine M., *Maximizing allele detection: Effects of analytical threshold and DNA levels on rates of allele and locus drop-out.* Forensic Science International: Genetics, 2012. **6**(6): p. 723-728.

34. Taylor, D., et al., *Validating multiplexes for use in conjunction with modern interpretation strategies.* Forensic Science International: Genetics, 2016. **20**: p. 6-19.

35. Chakravarti, I.M., R.G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*. Vol. 1. 1967, New York: John Wiley and Sons.

36. Hansson, O. and P. Gill, *Characterisation of artefacts and drop-in events using STR-validator and single-cell analysis.* Forensic Science International: Genetics, 2017. **30**: p. 57-65.

37. Broquet, T. and E. Petit, *Quantifying genotyping errors in noninvasive population genetics.* Molecular Ecology, 2004. **13**(11): p. 3601-3608.

38. *Bulldog Bio® 4-chip Disposable Hemocytometer User Manual*. 2017.

39. Geng, T., Novak, Richard., Mathies, Richard A., *Single-cell forensic short tandem repeat typing within microfluidic droplets.* Analytical chemistry, 2014. **86**(1): p. 703-712.

40. Mulero, J.J., C.W. Chang, and L.K. Hennessy, *Characterization of the N+3 stutter product in the trinucleotide repeat locus DYS392*. J Forensic Sci, 2006. **51**(5): p. 1069-73.

41. National Institute of Standards and Technology. *Short Tandem Repeat DNA Internet DataBase [Database]*. 2018; Available from: https://strbase.nist.gov/index.htm.

42. Aponte, R.A., Gettings, Katherine B., Duewer, David L., Coble, Michael D., Vallone, Peter M., *Sequence-based analysis of stutter at STR loci: Characterization and utility*. Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e456-e458.

43. Brochu, E., Swaminathan, Harish, Grgicak, Catherine M., *Forensic DNA Collection: Extraction of Molecular Information from Buccal Cells Using Direct Amplification*. 2017, ProQuest Dissertations Publishing.

44. Balding, D.J., Buckleton, John, *Interpreting low template DNA profiles.* Forensic Science International: Genetics, 2009. **4**(1): p. 1-10.

45. Mönich, U.J., Duffy, Ken, Médard, Muriel, Cadambe, Viveck, Alfonse, Lauren E., Grgicak, Catherine, *Probabilistic characterisation of baseline noise in STR profiles.* Forensic Science International: Genetics, 2015. **19**: p. 107-122.

46. Taylor, D., Buckleton, John, *Do low template DNA profiles have useful quantitative data?* Forensic Science International: Genetics, 2015. **16**: p. 13-16.

47. van Wieren-de Wijer, D.B.M.A., et al., *Determinants of DNA yield and purity collected with buccal cell samples*. Vol. 24. 2009. 677-82.

48. Adamowicz, M.S., et al., *Evaluation of Methods to Improve the Extraction and Recovery of DNA from Cotton Swabs for Forensic Analysis.* PLOS ONE, 2015. **9**(12): p. e116351.

49. Li, R.C., Harris, Ha, *Using Hydrophilic Adhesive Tape for Collection of Evidence for Forensic DNA Analysis.* Journal of Forensic Sciences, 2003. **48**(6): p. 1-4.

50. Stouder, S.L., et al., *Trace evidence scrapings: a valuable source of DNA?(Research and Technology).* Forensic Science Communications, 2001. **3**(4).

51. Voorhees, J.C., J.P. Ferrance, and J.P. Landers, *Enhanced Elution of Sperm from Cotton Swabs Via Enzymatic Digestion for Rape Kit Analysis* *. Journal of Forensic Sciences, 2006. **51**(3): p. 574-579.

52. Buckleton, J.S., et al., *The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity*. J Forensic Sci, 2018.