

TECHNICAL NOTE

Criminalistics

High-quality data from a forensically relevant single-cell pipeline enabled by low PBS and proteinase K concentrations

Nidhi Sheth MS¹ | Ken R. Duffy PhD² | Catherine M. Grgicak PhD^{1,3} 

¹Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey, USA

²Hamilton Institute, Maynooth University, Maynooth, Ireland

³Department of Chemistry, Rutgers University, Camden, New Jersey, USA

Correspondence

Catherine M. Grgicak, Department of Chemistry, Rutgers University, 315 Penn Street R306C, Camden, NJ 08102, USA.
Email: c.grgicak@rutgers.edu

Funding information

This work was partially supported by NIJ2020-R2-CX-0032 and NIJ2018-DU-BX-K0185 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not reflect those of the Departments of Justice.

Abstract

Interpreting forensic DNA signal is arduous since the total intensity is a cacophony of signal from noise, artifact, and allele from an unknown number of contributors (NOC). An alternate to traditional bulk-processing pipelines is a single-cell one, where the sample is collected, and each cell is sequestered resulting in n single-source, single-cell EPGs (scEPG) that must be interpreted using applicable strategies. As with all forensic DNA interpretation strategies, high quality electropherograms are required; thus, to enhance the credibility of single-cell forensics, it is necessary to produce an efficient direct-to-PCR treatment that is compatible with prevailing downstream laboratory processes.

We incorporated the semi-automated micro-fluidic DEPArray™ technology into the single-cell laboratory and optimized its implementation by testing the effects of four laboratory treatments on single-cell profiles. We focused on testing effects of phosphate buffer saline (PBS) since it is an important reagent that mitigates cell rupture but is also a PCR inhibitor. Specifically, we explored the effect of decreasing PBS concentrations on five electropherogram-quality metrics from 241 leukocytes: profile drop-out, allele drop-out, allele peak heights, peak height ratios, and scEPG sloping. In an effort to improve reagent use, we also assessed two concentrations of proteinase K. The results indicate that decreasing PBS concentrations to 0.5X or 0.25X improves scEPG quality, while modest modifications to proteinase K concentrations did not significantly impact it. We, therefore, conclude that a lower than recommended proteinase K concentration coupled with a lower than recommended PBS concentration results in enhanced scEPGs within the semi-automated single-cell pipeline.

KEYWORDS

direct-to-PCR extraction, forensic DNA, PBS, PCR efficiency, phosphate buffer saline, scEPG, single-cell electropherogram, single-cell forensic analysis

Highlights

- Direct-to-PCR workflows producing high quality single-cell electropherograms (scEPGs) were constructed.
- Impacts of PBS and proteinase K concentrations on scEPGs were studied.

Abbreviations: b.p., base pairs; CE, capillary electrophoresis; EPG, electropherogram; ILS, internal lane standard; LR, likelihood ratio; PBS, phosphate buffer saline; PCR, polymerase chain reaction; Pro K, proteinase K; RFU, relative fluorescent unit; STR, short tandem repeat.

Presented at the PITTCON 2020 Conference, March 1–5, 2020, in Chicago, IL; the 73rd Annual Scientific Conference of the American Academy of Forensic Sciences, February 15–19, 2021, held virtually; and the 32nd International Symposium on Human Identification Conference, September 12–16, 2021, in Orlando, FL.

- Lower PBS concentrations improved scEPG qualities.
- Changes to proteinase K concentrations did not impact scEPG qualities.
- Median allele detection rate per cell was 90% for the best performing treatment.

1 | INTRODUCTION

Forensic DNA interpretation has shifted from relying on binary decisions of peak presence/absence [1] to working within a likelihood ratio paradigm and applying continuous models [2–4]. Though statistically sound [5], and reliable [6], LR-based interpretation is contextual rather than absolute. That is, for involved case scenarios with complex EPGs, numerous LRs using a variety of assumptions may be required, increasing computational and analyst burdens. If a partial set of assumptions is applied, for example, the weight of evidence can be incomplete, possibly significantly so [7]. Thus, despite computational advances interpreting forensic DNA mixture samples remains arduous, requiring substantive resources to validate [8], train personnel [9], and support.

The drive to adopt intricately designed interpretation systems is a result of complex mixture EPGs, which are representations of potentially partial profiles from an unknown NOC who may or may not be related, and whose DNA may or may not be expressed in the data (e.g., victim DNA). Mixture EPGs are garnered from a multi-step workflow consisting of sample collection, DNA extraction, quantification, amplification of forensically relevant loci, and fragment analysis. For well-preserved, uninhibited samples, it is possible to optimize each step of the process [10–12] and obtain high-fidelity data—that is, data where each allele in the PCR tube is represented by good signal intensities. Even for these sample types, however, drop-out can occur since fractionation of the extract volume into two parts—one that is amplified and another that is stored—can cause partial profiles when the starting DNA copy number is small. As illustrated in the outer loop of Figure 1, in the bulk pipeline, only a small volume fraction (usually less than 15 μL) [13,14], V_{PCR} , is transferred from the extract volume, V_{ext} , to the PCR tube for amplification. The act of extracting DNA decouples alleles that are paired while the cell is intact, resulting in variability in the number of DNA molecules of each allele type that are found within V_{ext} . Thus, when V_{PCR} is transferred only a subset of DNA molecules will be sampled. If the starting copy number of DNA is low it is possible that zero copies of a given allele will be transferred, leading to complete allele drop-out. Thus, complete allele drop-out is inherent in the laboratory process and cannot be relieved by enhancing post-PCR conditions alone. Unfortunately, since forensic samples contain DNA from any number of individuals in any mixture proportion it is nearly impossible to assemble a bulk mixture pipeline devoid of volume fractionation effects [15].

A single-cell strategy overcomes these challenges by treating one cell at a time using a direct-to-PCR extraction/amplification approach. As illustrated in the inner loop of Figure 1, in a forensically relevant single-cell pipeline cells are sequestered before extraction and, thus, both alleles are necessarily transferred to the extraction/amplification tube. It is at this point that DNA

extraction, amplification, and fragment analysis ensue, resulting in a set of scEPGs where each renders a profile from an individual cell. Notably, in the single-cell strategy the extract is not fractionated, which circumnavigates its negative effects. Applications of single-cell methodology to forensics are not new [16], nor is it devoid of its own challenges. Single-cell laboratory systems, for example, require direct-to-PCR reagents that are compatible with downstream PCR ones. In bulk mixture samples, adverse effects of reagent compatibility are less obvious, but in the single-cell regime small PCR inefficiencies can have a significant impact on profile qualities [17].

Single-cell adaptations to meet forensic requirements were described as early as 1997 [16], where the authors summarized that of the 226 micro-dissected cells, 114 resulted in a full profile, and the allele dropout rate was a manageable 39%. More recently, semi-automated procedures have been used to garner single-cell data in a mechanized fashion [18–21]. The myriad ways single-cell analysis can be implemented in the forensic laboratory can be divided into three broad categories: (a) manual micromanipulation techniques that rely on significant analyst engagement and include pico-pipetting, laser micro-dissection, and other micromanipulators; (b) semi-automated methods that require the analyst to review and select cells for processing, though collection is managed by the instrumentation through microfluidics [19,20]; and (c) fully automated methods that do not rely on significant labor, and use micro- or nano-liters of reagents [22].

With the aim of optimizing semi-automated approaches, this work describes efforts to characterize and enhance a pipeline that incorporates the DEPArray™ instrument into the forensic single-cell system. Here, cell identification and selection are performed using immunostaining and bright field imaging. Fluorescently labeled cells are loaded onto a silicon chip consisting of ca. 300,000 micro-electrodes that apply alternative current in-phase or counter-phase voltages within the cartridge, trapping cells into field cages while acquiring high-resolution images for each cell. After the cells are selected, they are collected in 0.2 mL tubes using DEPArray™ Buffer. Next, the cells are washed in Phosphate Buffer Saline (PBS), an isotonic solution.

PBS is a salt solution containing sodium phosphate and potassium phosphate, which balances salt concentrations around the cell, reducing the chance it will lyse. Though it is useful in this regard, PBS is a known PCR inhibitor [23–25] having detrimental effects on results [26–31]. This study, therefore, aims to optimize a semi-automated single-cell pipeline by focusing on the PBS and proteinase K concentrations within it. Lower concentrations of proteinase K were investigated for the purposes of reducing reagent usage. An optimized procedure is one defined as producing high quality single-cell EPG signal, wherein quality was assessed using five performance metrics: (a) profile drop-out; (b) rates of allele drop-out; (c) allele peak heights; (d) peak height ratios; and (e) peak heights as a

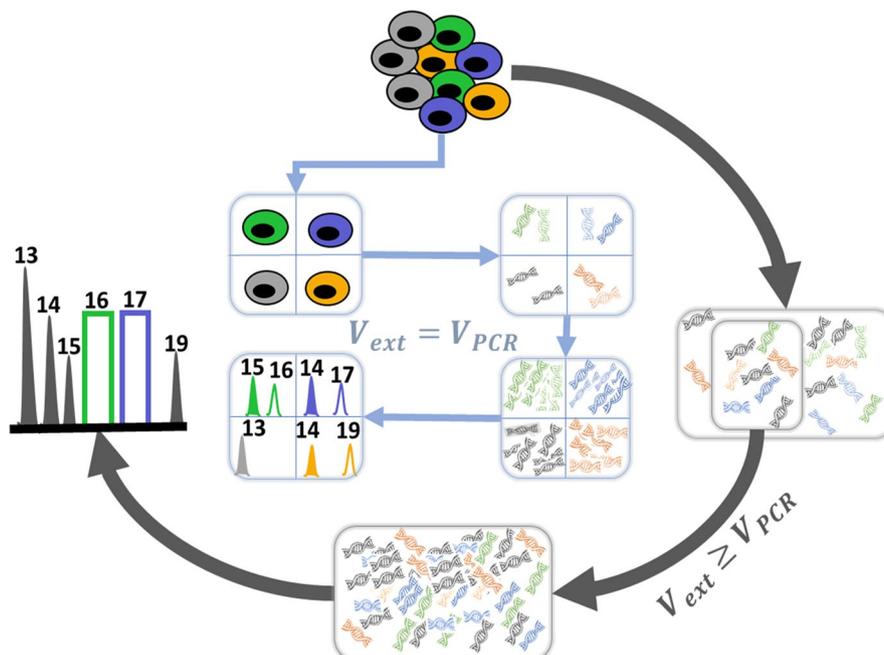


FIGURE 1 Schematic of the single-cell (inner) and traditional bulk (outer) laboratory pipelines. Each contributor's cell is distinguished by color. A heterozygous genotype contains two alleles (i.e., one dark shade, and one light shade), while homozygous loci contain one allele (i.e., dark shade). In traditional bulk mixture systems the cells are lysed, and the DNA is extracted and purified into a final volume, V_{ext} . A portion of V_{ext} is transferred to a PCR tube, V_{PCR} , to which PCR reagents are added. It is during this step that some alleles may not be sampled, driving allele drop-out (e.g., alleles 16 and 17 from the green and blue contributor, respectively). Amplification, electrophoresis and fragment analysis ensue resulting in one EPG containing, potentially, incomplete signal. The minimum number of contributors that could explain the evidence is two, though ground truth is four. Interpretation, therefore, requires computational systems that estimate the likelihoods of the data given certain genotypes are present in the signal, which requires a series of assumptions (e.g., the NOC or if a known contributor's genotype is represented in the data). An alternative is the single-cell system, which is depicted in the inner loop. Here, the cells are separated into distinct vessels prior to lysis and DNA extraction. Lysis and extraction are conducted using direct-to-PCR reagents followed by the addition of PCR reagents to the entire V_{ext} , avoiding fractionation. Fragment analysis follows, resulting in an scEPG for each cell, where the light shaded alleles are represented as unfilled peaks and the dark alleles are depicted in the electropherogram as solid peaks. Each scEPG, therefore, is composed of a single contributor's DNA, circumnavigating the need to deconvolve genotypes

function of molecular weight (i.e., EPG sloping) across 241 single leukocytes. Treatments producing many EPGs with high quality signal were taken to be those supporting good single-cell extraction and amplification efficiencies and were, thus, implemented.

2 | MATERIAL AND METHODS

White blood cells (WBC) were obtained from whole blood samples of two unknown individuals, in accordance with the ethical standards of the Institutional Review Board, Protocol number Pro2018002536. A total of 241 single-cells were sequestered, extracted, and analyzed as described in the following sections.

2.1 | Cell staining and isolation using DEPArray™ technology and STR profiling

All WBCs were collected using the DEPArray™ benchtop system using the manufacturer's recommendations [32]. Briefly, single-cell

identification and selection were performed using immunostaining and bright field imaging. The WBCs were stained using the Forensic Sample Prep kit [32] according to recommendations, which stains the leukocytes with anti CD45 PE, and nuclei with DAPI. The fluorescently tagged cells were then loaded onto a silicon chip consisting of over 300,000 micro-electrodes that applied alternative current in-phase or counter-phase voltages, trapping cells into field cages while acquiring images for each cell as depicted in Figure 2A. Once a maximum of 47 WBCs were selected, they were routed to a parking chamber on the DEPArray™ cartridge and each were then collected in 0.2 mL MicroAmp® Reaction Tubes for washing and volume reduction, direct-to-PCR extraction, STR amplification, electrophoresis, and data analysis.

2.2 | Volume reduction and extraction

According to the manufacturer's volume reduction recommendations [33], once a single-cell is dispensed into the 0.2 mL tube, the next step is to administer 100 μ L of 1X (10 mM) Ca^{2+} and Mg^{2+}

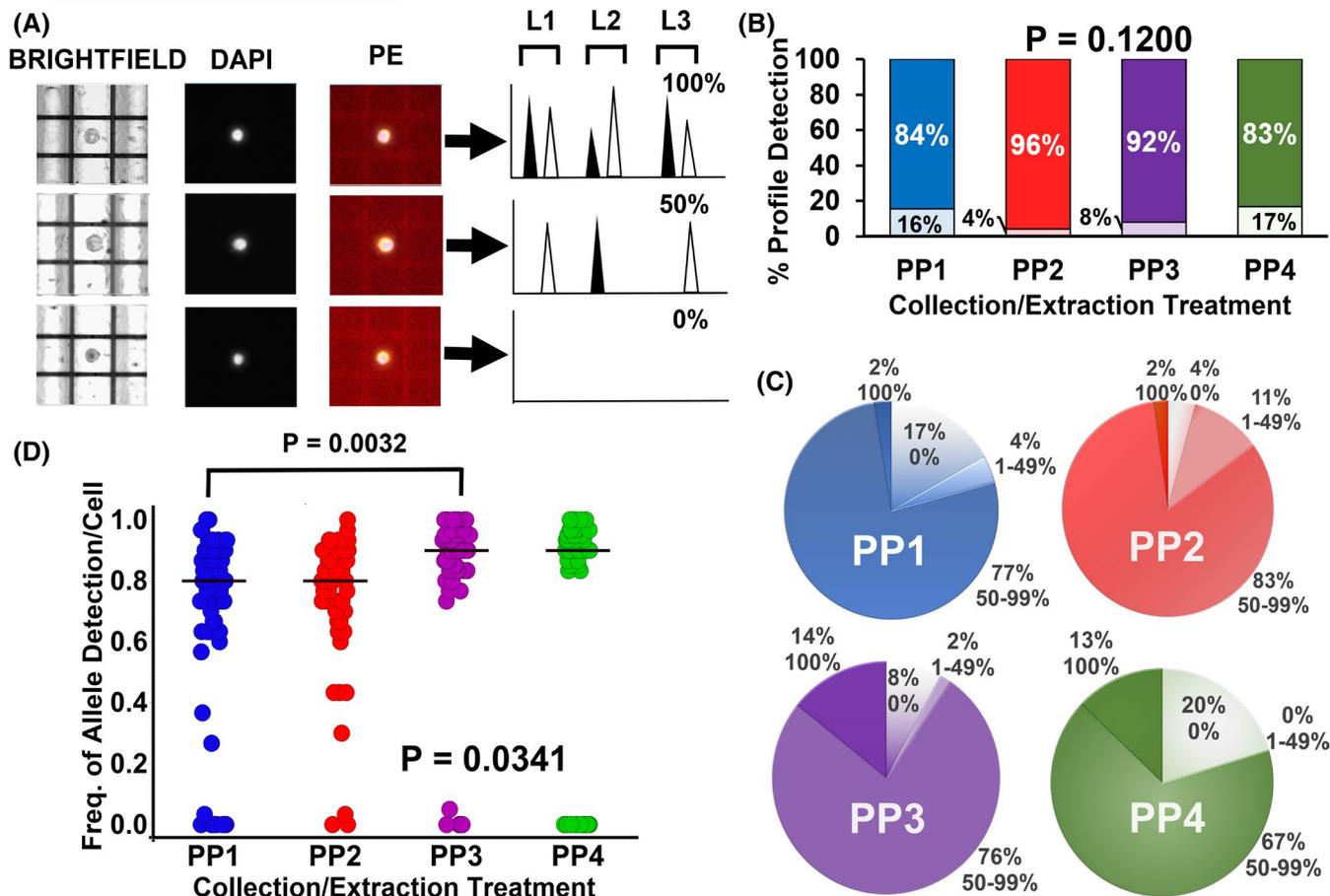


FIGURE 2 A, Brightfield, DAPI and PE images of white blood cells from the DEPAarray™ instrument. Based on visual inspection a maximum of 47 cells were collected, washed, extracted, and amplified using direct-to-PCR methods, where the amplification reagents were added directly to the extract, circumnavigating fractionation and its effects. A portion of the amplified product was injected into a 3500 Genetic Analyzer using post-PCR conditions with a limit of detection of one copy of DNA. Since there was no extract fractionation and the limit of detection is one copy, any allele drop-out across any STR locus, L, was the result of the cell quality or reagent incompatibilities. In some cases, allele drop-out may be so prevalent as to induce full profile drop-out—that is, 0% of the alleles (heterozygote or homozygote) are detected. B, Stacked plots depicting the percentage of profiles exhibiting profile drop-out versus those with at least one allele with a peak height exceeding the AT of 30 RFU, across the four treatments. If profile drop-out is more prevalent in one or more treatments it could indicate early cell lysis, inefficient cell collection, or cell/DNA loss during volume reduction. Though the cause of profile drop-out cannot be determined by this statistical analysis, exploring full profile drop-out is an important factor when optimizing. C, Pie charts showing the percentage of profiles exhibiting full (100%), most (50%–99%), few (1%–49%), and no alleles across the heterozygous alleles, where the known alleles are at least two STR units apart. D, A scatterplot depicting the frequency of allele detection per cell for all heterozygous alleles where the known alleles are at least two repeat units apart with median values 0.8, 0.8, 0.9, and 0.9 for PP1, PP2, PP3, and PP4, respectively. Significant *p*-values are shown

free PBS. The aim of volume reduction is to reduce this relatively large volume (i.e., ca. 100–150 μL) to ca. 2 μL for DNA extraction. The volume was reduced using the VRNxt™ instrument, which uses centrifugal force to collect and trap unwanted liquid in a dedicated cap for disposal, mitigating user-specific variable results and decreasing the risk of cell loss. To the 2 μL of PBS containing the cell is added 3 μL of PicoPure™ extraction mix for a total of 5 μL direct-To-PCR extraction volume. The extraction mix was prepared by adding reconstitution buffer into one vial of lyophilized proteinase K enzyme [34]. Since preliminary experiments using the manufacturer's recommendations showed lower than expected signal quality when compared with those acquired from micromanipulation methods [17], four collection/extraction reagent modifications were prepared and tested on

TABLE 1 Description of the four collection/extraction treatments across 241 single leukocytes collected using the DEPAarray™ system and coupled with PicoPure™ extraction

Treatment	PBS/proteinase K concentration	Number of single-cells
PP1	1X/high	77
PP2	1X/low	47
PP3	0.5X/low	63
PP4	0.25X/low	54

single leukocytes collected from the DEPAarray™ instrument. As depicted in Table 1, three concentrations of PBS—that is, 1X (10 mM), 0.5X (0.5 mM), 0.25X (0.25 mM)—were trialed, as were

two proteinase K concentrations. The DNA extracted using a “high” (i.e., recommended) concentration of proteinase K were those using a PicoPure™ extraction mix where 93 μL of extraction buffer was added to the tube containing a pre-determined, unknown amount of lyophilized proteinase K. Notably, 93 μL is not the manufacturer's volume recommendation. However, the recommended protocol as outlined within the manual [34] does not presume any pre-existing volume (i.e., 2 μL) in the extraction tube and stipulates a full 5 μL be aliquoted directly to the vessel containing the cells. By reconstituting the proteinase K with 93 μL rather than the stated 155 μL , the proteinase K concentration of the extract remains equal to the manufacturer's. More specifically, the manufacturer's recommendations stipulate that 5 μL of $x[\mu\text{g Pro K}]/155 [\mu\text{L}]$ extraction buffer be pipetted into a dry tube. To account for volume differences, we pipetted 3 μL of $x[\mu\text{g Pro K}]/93 [\mu\text{L}]$ extraction buffer into a tube already containing 2 μL of PBS, a necessary outcome of integrating the DEPAArray™ system into the single-cell pipeline. In contrast, the “low” proteinase K treatment was one that included the addition of 155 μL of reconstitution buffer to the lyophilized proteinase K. The use of 1X PBS and the addition of 93 μL of reconstitution buffer (i.e., high proteinase K) represents each of the manufacturer's recommendations [32,34] used in tandem, and is named PP1. Though treatment combinations using low PBS and high proteinase K concentrations were possible, we focused on exploring effects of low proteinase concentrations since it represents a process using only one tube of proteinase K per 47 cells and, therefore, represented a more efficient laboratory pipeline.

After addition of extraction buffer, the sample was centrifuged at 300g for 3 min and then incubated at 65°C for 3 h. To inactivate the proteinase K, the sample was incubated at 95°C for 10 min, and multiplex STR amplification followed.

2.3 | STR analysis

Direct single-cell STR amplification was performed in a total reaction volume of 12.5 μL using the GlobalFiler™ PCR Amplification kit [13] (Life Technologies Corporation), where 7.5 μL of amplification reaction mix was directly added to each well containing 5 μL of the extract (i.e., 2 μL of solution containing one cell of DNA in addition to the 3 μL of extraction mix). All thermal cycling was completed in a GeneAmp® 9700 PCR thermal cycler (Applied Biosystems) with ramp speeds, temperatures, and incubation times following the manufacturer's recommendations [13]. A total of 30 PCR cycles was used. In preparation for fragment separation, 1 μL of PCR product was added to 9.7 μL of HiDi formamide and 0.3 μL of GeneScan™ 600 Liz® (60–460) Size Standard v2.0 (Applied Biosystems). Products were separated and detected using an injection potential and time of 1200 V and 25 s, respectively, on an Applied Biosystems 3500 Genetic Analyzer. All profiles were analyzed using OSIRIS Version 2.10.3. An analytical threshold (AT) of 30 relative fluorescence units (RFU) was applied.

2.4 | Data preparation and data clean-up

In this study, the genotype of single-cell DNA samples is known as determined by extracting the DNA from 100 μL of whole blood using the QIAamp Investigator DNA Extraction kit [35]. A mass of 0.25 ng of DNA was amplified using the recommended cycling parameters (i.e., 29 cycles) for the GlobalFiler™ STR assay [13]. STR fragment analysis occurred as previously described for the single-cell samples.

The first step in the analysis process was to quality control check each single-cell sample and confirm that the apparent alleles of the single-cell samples were consistent with the known genotype, and no inconsistencies were noted. Samples with more than two loci exhibiting drop-in [36] were removed from further analysis, where drop-in was defined as an extraneous peak that did not fall within a stutter or allele position and that exceeded the AT of 30 RFU.

2.5 | Statistical analysis

2.5.1 | Profile drop-out

Since PBS is used to maintain the integrity of the cell, we tested the effects of the collection/extraction treatments on the propensity to observe full profile drop-out. High rates of profile drop-out were taken as an indication that the combination of collection and extraction reagents was sub-optimal. An EPG was classified as exhibiting full profile drop-out if no known allele's peak height crossed the AT of 30 RFU. Allelic drop-out, generally, originates from two sources: (a) sampling effects, which are the result of zero copies of a given allele surviving the pre-PCR process [10]; and (b) detection effects, which occur when DNA targets survive the pre-PCR stage and are amplified but the number of amplified copies are too low to illicit adequate fluorescence [12]. Simulations by the authors of References [12,37] show that for well-preserved DNA, carefully chosen post-PCR laboratory settings can significantly reduce negative effects from the latter while simultaneously minimizing complete drop-out. That work was experimentally corroborated in [10]. Given the PCR cycle number was set to 30 and the injection time set to 25 s on a 3500 Genetic Analyzer, a limit of detection of one amplifiable copy of DNA [10,38] was obtained. Thus, within the context of this work signal loss and intensity differences can be attributed to pre-PCR treatments and their compatibility with downstream amplifications. We recorded the proportion of sCEPGs with 0 detected alleles and the proportion of samples for which there was at least one allele whose height exceeded the AT. Because we were interested in reviewing the entire profile, both heterozygous and homozygous loci were interrogated. We applied the chi-square for independence (JMP® Pro 15) to statistically test the null hypothesis that full profile dropout is independent of the pre-PCR treatment applied. A *p*-value threshold of 0.05 was used to reject the null hypothesis. When performing post-hoc tests, we use the Bonferroni corrected *p*-value threshold of 0.008 [39].

2.5.2 | Frequency of allele detection

To alleviate confounding effects from stutter, when determining allele detection rates only heterozygous loci with STRs at least two repeat units apart were used. This represented 10 and 15 loci for the two individuals. In total, 3380 heterozygous alleles meeting this criterion were evaluated.

The frequency of allele detection was performed on a per cell basis since previous work demonstrated that there is inconsistency in the quality of the signal of EPGs generated from distinct cells and that a single likelihood of allele drop-out does not effectively characterize those data [17]. Specifically, rates of allele drop-out were determined by counting the number of heterozygous alleles crossing the AT of 30 RFU and dividing that value by the total number of expected heterozygous alleles based on the known genotype. We statistically assessed if the rates of allele drop-out are interchangeable between treatments by applying the permutation test function in JMP® Pro 15. In the permutation test, the frequency of alleles detected per cell is randomly shuffled among treatments, under the null hypothesis that the treatment has no effect on the frequency of allelic detection. A *p*-value was calculated using the F-ratio obtained from the data as compared to the F-ratios obtained after performing 10⁵ simulations under the null hypothesis. A *p*-value threshold of 0.05 was used to reject the null hypothesis. The post-hoc *p*-value threshold was 0.008. Only significant *p*-values are expressed in the figures.

2.5.3 | Average peak height and their ratios

To compute the average peak height ratios and average peak heights we, once again, only use the heterozygous alleles that are at least two repeat units apart. The average peak height per cell was determined by summing the heights of the heterozygous allele peaks and dividing by the number of heterozygous alleles. The average peak height ratio per cell was determined by taking the average peak height ratio across candidate loci, which was calculated by dividing the height of the less intense allele by the height of the more intense one. If an allele was not detected, a peak height of 1 RFU was assigned. Hypothesis testing was conducted with a permutation test, using the F-ratio statistic. A *p*-value threshold of 0.05 and 0.008 was used to reject the null hypothesis for the multi-factor and pair-wise tests, respectively.

2.5.4 | EPG sloping

Another measure of EPG quality is EPG sloping, where the peaks exhibit a downward trend as the molecular weight of the target fragment increases [40]. The statistic representing EPG sloping is the exponential parameter, β , in a model describing the tendency of the peak height, H_l , associated with the known genotypes at locus l , to exponentially decrease with the average base pair size of the STR alleles at that locus, s_l ,

$$H_l = Ae^{\beta s_l} \quad (1)$$

If there is significant sloping due to inhibition or DNA degradation, β will take a negative value. In contrast, if there is good signal balance across all loci, indicating good reagent compatibility and high DNA quality, β will be close to zero. When calculating the parameters both heterozygous and homozygous alleles are used. To test if the collection and extraction treatments have a significant effect on β , a permutation test using the F-ratio statistic and the same *p*-values as previously described was applied.

3 | RESULTS AND DISCUSSION

With the aim of producing a semi-automated single-cell process capable of yielding high quality signal, 241 single WBC profiles from two individuals were studied across four collection/extraction treatments (Table 1).

3.1 | Profile drop-out

In Figure 2A, we show the size and shape of a typical leukocyte chosen for further processing. To be selected, a spherically shaped object must be centered in one of the DEPAArray™ cages when viewed in brightfield and PE mode. Choosing objects that fall in the center of a cage improve the chance of effectively routing it. Under DAPI filters, the size and shape of the object must be spherical, but with a dense bright spot in the same location as the object observed using PE filters. These characteristics indicate that the object-of-interest is a whole WBC, with an intact nucleus. Any images suggesting two or more cells are in a cage or that the cells/nucleus are compromised are invalidated, and not selected for collection.

In contrast to bulk pipelines that produce a single EPG per crime-scene sample, the single-cell pipeline produces n scEPGs per sample—one for every cell collected. Despite the sensitive post-PCR conditions and the absence of fractionation from the pipeline, scEPGs are known to vary in their quality [16,17,19], wherein some display peaks at every allele position across all loci, while others exhibit partial or full profile drop-out, as illustrated in Figure 2A. As described in section 2.2, the cell is collected into a 0.2 mL amplification tube in proprietary DEPAArray™ Buffer. At the completion of the collection step, 100 μ L of PBS is added to the tube. The cell, collection buffer and PBS buffer undergo volume reduction, decreasing the volume from hundreds of microliters to ca. 2 μ L. Given the series of steps, cell loss or rupture could have occurred during cell collection or volume reduction. Though it is impossible to determine the cause of profile drop-out when it occurs, comparing its rate across treatments can challenge the hypothesis that no significant difference exists, and expose which treatments are associated with an increase in profile drop-out. Specifically, Figure 2B are stacked plots presenting the

proportion of samples for which profile drop-out was observed and the proportion of samples for which it was not. The figure shows four interesting features. First, the p -value of 0.12 indicates that there is no evidence that profile drop-out is dependent on the collection/extraction treatments employed. Next, when PP2, PP3, and PP4—low proteinase K concentrations with decreasing PBS concentrations—are examined, the trend is suggestive that lower PBS concentrations impact profile drop-out. However, when viewed in conjunction with the results from PP1 (i.e., 1X PBS) and PP4 (0.25X PBS), which show similar profile drop-out rates, no definitive trend emerges. When coupled with the change in profile drop-out rates of 16% and 4% between PP1 and PP2, we conclude that the four reagent combinations had no significant or consistent effect on profile drop-out, suggesting lower than recommended PBS and proteinase K concentrations may be viable laboratory options when developing a forensically relevant single-cell process.

3.2 | Frequency of allele detection

To explore the effects of pre-amplification reagents on profile qualities, we examined each treatment by assessing the percentage of scEPGs containing all (100%), most (50%–99%), some (1%–49%), or no (0%) heterozygous alleles. The pie charts of Figure 2C demonstrate that although the percentage of samples rendering allele detection rates greater than 50% are similar across treatments, there is a noticeable difference in the percentage of samples exhibiting 100% of their heterozygous alleles. More specifically, PP3 and PP4 resulted in 14% and 13% of the profiles with full allele detection as compared to 2% for PP1 and PP2. In addition, the results of Figure 2D show that the frequency of allele detection per cell have medians of 0.8, 0.8, 0.9, and 0.9 for PP1, PP2, PP3, and PP4, respectively, and a p -value of 0.034. Post-hoc multiple comparisons suggest, in particular, that the allele drop-out rates are not statistically the same for PP1 and PP3. Taken together, the PP3 and PP4 treatments again emerge as viable options for the collection and direct-to-PCR analysis of single-cells.

3.3 | Average peak height and their ratios

Developing a forensically relevant single-cell process is reliant on high extraction efficiency and good direct-to-PCR reagent compatibility. Good reagent compatibility will be borne out in the quality of the scEPG, which can be explored using several metrics. Though not as important as peak detection, peak heights indicate a well-assembled pipeline since the height of an allele is a proxy for the number of synthesized copies of DNA procured during PCR, which is a function of the efficiency of the PCR reaction [10,11]. If the PCR efficiency is low, the peak heights will also be low. Though forensic DNA samples are known to carry PCR inhibitors causing a variety of unwanted effects [28,31,41–43], the single-cells utilized in this

study were from whole-blood, and so any overall differences in peak heights between treatments will necessarily be a reflection of the amplification efficiency for a set of STR targets.

As shown by Zhu et. al [23], higher concentrations of PBS can have detrimental effects on PCR efficiency, which would be reflected in lower peak heights. Figure 3A shows density plots of the high and lower peak intensity [RFU] of heterozygous alleles. In a well-performing system, most of the heterozygous alleles will be associated with peaks that are of the same and higher intensity, resulting in a plot with more points falling near the $x = y$ line, and away from the origin. Unlike treatments PP1 and PP2, those of PP3 and PP4 show a higher density of heterozygous peaks near $x=y$ and away from the origin. These data are further explored in Figure 3B,C. Figure 3B is the scatterplot of the average peak height per cell grouped by treatment, and the overall p -value demonstrates significant differences. Post-hoc pairwise testing shows all statistically significant results are seen between groups of cells collected with high concentrations of PBS and those collected with lower than recommended concentrations. Notably, the median average peak height rendered by PP3—that is, low proteinase K and 0.5X PBS—is 667 RFU, which is consistent with the median peak heights obtained with single epithelial cells collected in the absence of PBS [17]. The peak heights obtained from PP1 and PP2 are similar to those obtained when using an unrelated, though analogous, treatment (i.e., LysePrep extraction and a 31 cycle PowerPlex Fusion 6c amplification) [19] showing that, in general, forensically relevant single-cell pipelines are readily implemented using a variety of treatments and improved results can be obtained by systematically evaluating the effect of each component on EPG results.

Though permutation tests suggested peak height ratios were also influenced by the pre-amplification components (Figure 3B), there was no post-hoc pair rendering p -values less than the Bonferroni-corrected significance threshold. When evaluating any trends between PP1 and PP2 (i.e., same 1X PBS, with changing proteinase K), and between PP2, PP3, and PP4, no clear trends emerge. Despite this result, improved peak heights and, thus, amplification efficiencies result from reducing PBS concentrations. In addition, the peak height and ratio outcomes between PP1 and PP2 suggest that modestly decreasing proteinase K concentrations do not negatively impact results.

3.4 | EPG sloping

The last of the EPG features studied was that of sloping, which is represented by the exponential parameter, β (Equation 1). As shown in Figure 4A, large negative β s means there is a substantive decrease in signal intensity across base pair lengths, while β values near zero suggest there is peak balance across the size of the fragments. The scatter plot of Figure 4B shows β s for each cell across treatments. Permutation testing confirms what can be surmised by visual inspection: that the collection and extraction conditions influence EPG results, with PP3 and PP4 showing significant improvement over

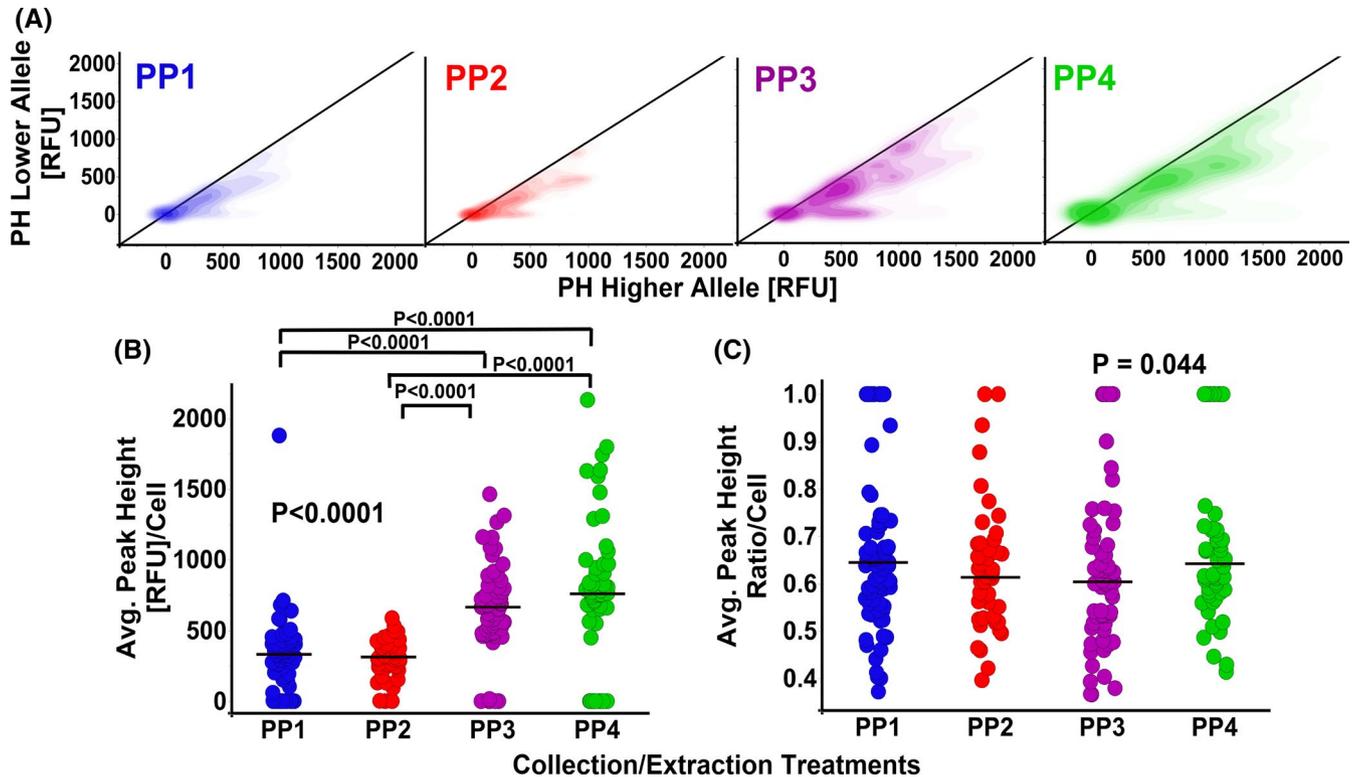


FIGURE 3 A, Density plots of the less intense peak heights plotted against heights of more intense allele peaks for all heterozygous alleles across four treatments. Also shown is the $x = y$ line. B, Scatter plot depicting the average peak height per cell, in RFU, for all the heterozygous alleles across the four PBS and proteinase K treatments with median values 333, 314, 667, and 761 for PP1, PP2, PP3, and PP4, respectively. The permutation test indicates that the peak height is significantly different. Pairwise tests giving p -values below the Bonferroni corrected threshold of 0.008 are also shown, where we see the peak heights between PP1 and PP2 are of the same distribution, as are those of PP3 and PP4. C, Scatter plot depicting the average peak height ratio per cell for the heterozygous alleles separated by treatment, with associated median values 0.64, 0.61, 0.60, and 0.64 for PP1, PP2, PP3, and PP4, respectively. Significant p -values are shown

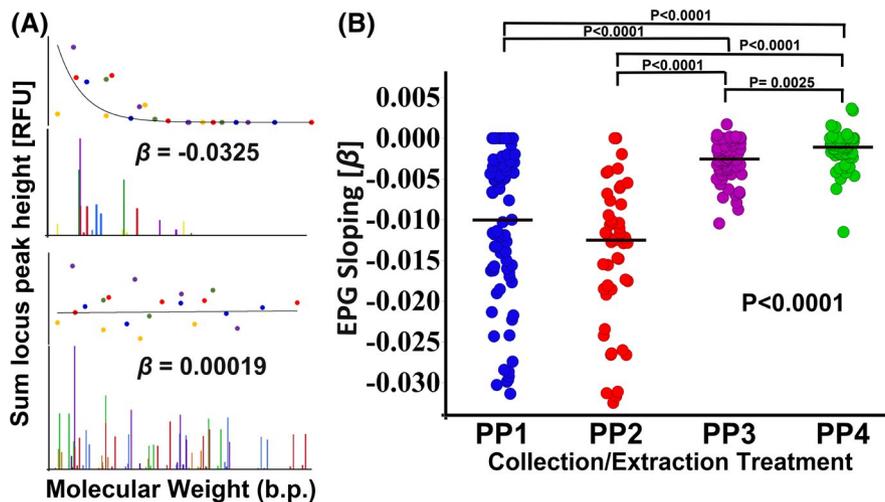


FIGURE 4 A, Representative EPGs obtained when amplifying DNA procured from a single white blood cell for 30 cycles using the GlobalFiler™ assay, which contains 21 STR tri- or tetra-nucleotide repeats ranging from ca. 80 to 400 b.p. The first exemplar EPG obtained under PP2 conditions (i.e., 1X PBS and lower Pro K) shows severe EPG sloping with a β -value of -0.0325 . The second EPG shows one with a near-0 β -value, indicative of the good peak height balance across base-pair size. This EPG was obtained under PP3 conditions (i.e., 0.5X PBS and lower Pro K). The treatment rendering more samples with β values near zero, the more conducive that treatment is to single-cell forensics. B, Scatter plot depicting the β -values across treatments with medians -0.01 , -0.0125 , -0.0025 , and -0.0011 for PP1, PP2, PP3, and PP4, respectively. Pairwise p -values less than the Bonferroni corrected threshold of 0.008 are included

collection conditions using 1X PBS. Notably, the p -value between PP3 and PP4 also indicates improvements to sloping when the PBS concentration is 0.25X, but the medians suggest the improvement is marginal. The largest enhancement is seen between PP2 and PP3, where the median β value increased by a factor of 5.

Though PBS is a valuable component to single-cell pipelines, its effect on the availability of Mg^{2+} [26], which is an important co-factor of Taq Polymerase, can clearly impact EPG quality. An alternative to decreasing PBS concentrations is to maintain the PBS's concentration at 1X but to supplement the PCR amplification reaction mix with additional Mg^{2+} by the addition of $MgCl_2$. This treatment was trialed, but the resulting EPGs contained high noise levels (likely the result of nonspecific amplification), which did not induce further analysis (data not shown).

Taken together, the reasonable rates of profile drop-out, large peak heights, and very good sloping values suggest that modest modifications to PBS concentrations from 1X to 0.5X, or even 0.25X, with modest reductions in proteinase K concentrations render many good quality scEPGs suitable for forensic purposes.

4 | CONCLUSION

Whether a semi-automated or manual single-cell sequestration technique is adopted, signal quality from the direct-to-PCR extraction and amplification of single-cells is reliant on reagent compatibility across the variety of steps required of this type of system. This is especially true when coupling components, reagents, or steps from independent vendors. Since PBS was deemed a necessary reagent, explorations into the effects of PBS concentrations ensued and showed that PBS concentrations at 0.5X or 0.25X improve EPG results without negatively impacting full-profile drop-out. In addition, decreasing the proteinase K concentration had no detrimental effects. Since barriers to implementation of single-cell technologies into crime laboratory environments will be driven by successful single-cell EPG procurement, reasonable reagent costs and the development of sound interpretation techniques, this work supplements existing literature by demonstrating most scEPGs render good quality profiles when the individual system components are well-joined and compatible [16,19–21,44,45]

ORCID

Catherine M. Grgicak  <https://orcid.org/0000-0003-3745-0577>

REFERENCES

- Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genet.* 2016;17:15. <https://doi.org/10.1186/s12863-016-0429-7>
- Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele(R) DNA mixture interpretation. *J Forensic Sci.* 2011;56(6):1430–47. <https://doi.org/10.1111/j.1556-4029.2011.01859.x>
- Bright JA, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix (TM) – a multi laboratory response to PCAST. *Forensic Sci Int Genet.* 2018;34:11–24. <https://doi.org/10.1016/j.fsigen.2018.01.003>
- Bleka O, Storvik G, Gill P. EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Sci Int Genet.* 2016;21:35–44. <https://doi.org/10.1016/j.fsigen.2015.11.008>
- Curran JM, Triggs CM, Buckleton J, Weir BS. Interpreting DNA mixtures in structured populations. *J Forensic Sci.* 1999;44(5):987–95. <https://doi.org/10.1520/JFS12028J>
- Buckleton JS, Pugh SN, Bright J-A, Taylor DA, Curran JM, Kruijver M, et al. Are low LRs reliable? *Forensic Sci Int Genet.* 2020;49:7. <https://doi.org/10.1016/j.fsigen.2020.102350>
- Grgicak CM, Duffy KR, Lun DS. The a posteriori probability of the number of contributors when conditioned on an assumed contributor. *Forensic Sci Int Genet.* 2021;54:9. <https://doi.org/10.1016/j.fsigen.2021.102563>
- Riman S, Iyer H, Vallone PM. Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset. *PLoS One.* 2021;16(9):e0256714. <https://doi.org/10.1371/journal.pone.0256714>
- Kalafut T, Pugh S, Gill P, Abbas S, Semaan M, Mansour I, et al. A mixed DNA profile controversy revisited. *J Forensic Sci.* 2021;00:1–8. <https://doi.org/10.1111/1556-4029.14912>
- Duffy KR, Gurram N, Peters KC, Wellner G, Grgicak CM. 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):855–68. <https://doi.org/10.1002/elps.201600385>
- Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Res.* 2005;33(2):632–43. <https://doi.org/10.1093/nar/gki205>
- Hedell R, Hedman J, Mostad P. Determining the optimal forensic DNA analysis procedure following investigation of sample quality. *Int J Legal Med.* 2018;132(4):955–66. <https://doi.org/10.1007/s00414-017-1635-1>
- ThermoFisherScientific. GlobalFiler™ PCR amplification kit: user guide. Carlsbad, CA: Life Technologies Corporation; 2015. <http://tools.thermofisher.com/content/sfs/manuals/4477604.pdf> Accessed Dec 6, 2021
- PowerPlex® Fusion System for use on the Applied Biosystems® genetic analyzers. Madison, WI: Promega Corporation; 2020. <https://www.promega.com/~media/Files/Resources/Protocols/Technical%20Manuals/101/PowerPlex%20Fusion%20System%20Protocol.pdf> Accessed 6 Dec 2021
- An Z. Advanced modeling of circadian systems and forensic DNA interpretation. New Brunswick, NJ: Rutgers, The State University of New Jersey; 2021.
- Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A. DNA fingerprinting from single cells. *Nature.* 1997;389(6651):555–6. <https://doi.org/10.1038/39225>
- Sheth N, Swaminathan H, Gonzalez AJ, Duffy KR, Grgicak CM. Towards developing forensically relevant single-cell pipelines by incorporating direct-to-PCR extraction: compatibility, signal quality, and allele detection. *Int J Legal Med.* 2021;135(3):727–38. <https://doi.org/10.1007/s00414-021-02503-4>
- Williamson VR, Laris TM, Romano R, Marciano MA. Enhanced DNA mixture deconvolution of sexual offense samples using the DEPArray™ system. *Forensic Sci Int Genet.* 2018;34:265–76. <https://doi.org/10.1016/j.fsigen.2018.03.001>
- Watkins DRL, Myers D, Xavier HE, Marciano MA. Revisiting single cell analysis in forensic science. *Sci Rep.* 2021;11(1):7054. <https://doi.org/10.1038/s41598-021-86271-6>

20. Hansson O, Gill P. Characterisation of artefacts and drop-in events using STR-validator and single-cell analysis. *Forensic Sci Int Genet.* 2017;30:57–65. <https://doi.org/10.1016/j.fsigen.2017.04.015>
21. Anslinger K, Graw M, Bayer B. Deconvolution of blood-blood mixtures using DEPAArray(TM) separated single cell STR profiling. *Rechtsmedizin.* 2019;29(1):30–40. <https://doi.org/10.1007/s00194-018-0291-1>
22. Geng T, Novak R, Mathies RA. Single-cell forensic short tandem repeat typing within microfluidic droplets. *Anal Chem.* 2014;86(1):703–12. <https://doi.org/10.1021/ac403137h>
23. Zhu Y, Zhang Y-X, Liu W-W, Ma Y, Fang Q, Yao B. Printing 2-dimensional droplet array for single-cell reverse transcription quantitative PCR assay with a microfluidic robot. *Sci Rep.* 2015;5(1):9551. <https://doi.org/10.1038/srep09551>
24. Martin NC, Pirie AA, Ford LV, Callaghan CL, McTurk K, Lucy D, et al. The use of phosphate buffered saline for the recovery of cells and spermatozoa from swabs. *Sci Justice.* 2006;46(3):179–84. [https://doi.org/10.1016/s1355-0306\(06\)71591-x](https://doi.org/10.1016/s1355-0306(06)71591-x)
25. Chan LLY, Rice WL, Qiu J. Observation and quantification of the morphological effect of trypan blue rupturing dead or dying cells. *PLoS One.* 2020;15(1):17. <https://doi.org/10.1371/journal.pone.0227950>
26. Wilson IG. Inhibition and facilitation of nucleic acid amplification. *Appl Environ Microbiol.* 1997;63(10):3741–51. <https://doi.org/10.1128/aem.63.10.3741-3751.1997>
27. Matsumura S, Matsusue A, Waters B, Kashiwagi M, Hara K, Kubo S. Effects of PCR inhibitors on mRNA expression for human blood identification. *Leg Med.* 2018;32:113–9. <https://doi.org/10.1016/j.legalmed.2018.04.002>
28. Demeke T, Jenkins GR. Influence of DNA extraction methods, PCR inhibitors and quantification methods on real-time PCR assay of biotechnology-derived traits. *Anal Bioanal Chem.* 2010;396(6):1977–90. <https://doi.org/10.1007/s00216-009-3150-9>
29. Kemp BM, Bingham B, Frome R, Labonte M, Palmer E, Parsons ES, et al. Subduing the influence of PCR inhibitors on amplifying aged, degraded, and low copy number DNA: PCR enhancer cocktail-p and rescue PCR. *PLoS One.* 2020;15(6):13. <https://doi.org/10.1371/journal.pone.0234745>
30. Radstrom P, Knutsson R, Wolffs P, Lovenklev M, Lofstrom C. Pre-PCR processing – Strategies to generate PCR-compatible samples. *Mol Biotechnol.* 2004;26(2):133–46. <https://doi.org/10.1385/mb:26:2:133>
31. Alaeddini R. Forensic implications of PCR inhibition – a review. *Forensic Sci Int Genet.* 2012;6(3):297–305. <https://doi.org/10.1016/j.fsigen.2011.08.006>
32. DEPAArray™ Forensic sample prep kit user manual. Bologna, Italy: Menarini Silicon Biosystems S.p.A.; 2019. https://cdn2.hubspot.net/hubfs/304284/User%20Manuals/IFU_1002-R4.0%20-%20DEPAArray%20Forensic%20SamplePrep%20-%202019-02-04%20-%20For%20Print.pdf Accessed Dec 6, 2021
33. VR NxT volume reduction instrument instruction manual. Bologna, Italy: Menarini Silicon Biosystems; 2019. <https://fcc.report/FCC-ID/2APP5-VRNXT01/4229353.pdf> Accessed Dec 6, 2021
34. ARCTURUS® PicoPure® DNA Extraction Kit. Carlsbad, CA: Life Technologies Corporation; 2010. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_086062.pdf Accessed Dec 6, 2021
35. QIAamp® DNA investigator handbook. Hilden, Germany: QIAGEN®; 2020 Jan. <https://www.qiagen.com/us/resources/resource-detail?id=26ef8f2c-7c2a-49e6-b2d2-39d4e130b3cc&lang=en> Accessed Dec 6, 2021
36. Taylor D, Bright J-A, McGoven C, Hefford C, Kalafut T, Buckleton J. Validating multiplexes for use in conjunction with modern interpretation strategies. *Forensic Sci Int Genet.* 2016;20:6–19. <https://doi.org/10.1016/j.fsigen.2015.09.011>
37. Hansson O, Egeland T, Gill P. Characterization of degradation and heterozygote balance by simulation of the forensic DNA analysis process. *Int J Legal Med.* 2017;131(2):303–17. <https://doi.org/10.1007/s00414-016-1453-x>
38. Peters KC, Swaminathan H, Sheehan J, Duffy KR, Lun DS, Grgicak CM. Production of high-fidelity electropherograms results in improved and consistent DNA interpretation: standardizing the forensic validation process. *Forensic Sci Int Genet.* 2017;31:160–70. <https://doi.org/10.1016/j.fsigen.2017.09.005>
39. Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt.* 2014;34(5):502–8. <https://doi.org/10.1111/opo.12131>
40. Bright JA, Taylor D, Curran JM, Buckleton JS. Degradation of forensic DNA profiles. *Aust J Forensic Sci.* 2013;45(4):445–9. <https://doi.org/10.1080/00450618.2013.772235>
41. Opel KL, Chung D, McCord BR. A study of PCR inhibition mechanisms using real time PCR. *J Forensic Sci.* 2010;55(1):25–33. <https://doi.org/10.1111/j.1556-4029.2009.01245.x>
42. Funes-Huacca ME, Opel K, Thompson R, McCord BR. A comparison of the effects of PCR inhibition in quantitative PCR and forensic STR analysis. *Electrophoresis.* 2011;32(9):1084–9. <https://doi.org/10.1002/elps.201000584>
43. Kuffel A, Gray A, Daeid NN. Impact of metal ions on PCR inhibition and RT-PCR efficiency. *Int J Legal Med.* 2021;135(1):63–72. <https://doi.org/10.1007/s00414-020-02363-4>
44. Perié L, Hodgkin Philip D, Naik Shalin H, Schumacher Ton N, de Boer RJ, Duffy KR. Determining lineage pathways from cellular barcoding experiments. *Cell Rep (Cambridge).* 2014;6(4):617–24. <https://doi.org/10.1016/j.celrep.2014.01.016>
45. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science.* 2015;347(6226):1138–42. <https://doi.org/10.1126/science.aaa1934>

How to cite this article: Sheth N, Duffy KR, Grgicak CM. High-quality data from a forensically relevant single-cell pipeline enabled by low PBS and proteinase K concentrations. *J Forensic Sci.* 2021;00:1–10. <https://doi.org/10.1111/1556-4029.14956>