2022

Classification and Differentiation of Forensically Relevant Cell Populations using Imaging Flow Cytometry

Arianna D. DeCorte
Virginia Commonwealth University

Classification and Differentiation of Forensically Relevant Cell Populations using Imaging Flow Cytometry

Arianna DeCorte
Fall 2020 – Spring 2022
March 16th, 2022

Mentor: Christopher J. Ehrhardt, Ph.D.
Ehrhardt Research Laboratory
Department of Forensic Science
Virginia Commonwealth University Science

A thesis/dissertation submitted in partial fulfillment of the requirements for the degree of Master of Forensic Science at Virginia Commonwealth University.

Acknowledgements

I would like to thank Dr. Christopher Ehrhardt for his constant support, patience and help throughout the duration of this research project. His guidance and ability to inspire critical thinking was priceless and I will continue to use the tools he equipped me with throughout my continued education and career. I would also like to thank my committee members, Katherine Philpott J.D., and Dr. Sarah Seashols Williams for their assistance throughout this research. A big thank you to Julie Farnsworth and Dr. XinYan Pei from the MCV Flow Core for their help throughout this project as I processed my samples. Lastly, thank you to my wonderful family and friends, especially my parents, Anne and Diego DeCorte, for their constant love and support. Your confidence in me is everything.

Abstract

Forensic casework relies heavily on DNA profiles which may be time-consuming to generate and difficult to interpret when biological mixtures are present. Additionally, there are significant challenges in using sub-source data alone to answer the activity-level questions often most pertinent in criminal cases. Source level information can add critical probative value, however, methods that can provide information as to the source tissue of evidentiary cell populations are limited. There remains a need for new methods that can differentiate between and classify various cell populations, particularly for vaginal and epidermal tissue, since there are no conventional serological techniques specific to these sources. One promising but unexplored approach is to use flow cytometry on the front-end of the DNA workflow. Flow cytometry is a high throughput and non-destructive method for characterizing physical and biological attributes of individual cells through autofluorescence profiles. This study aimed to develop a new forensic signature system to increase the probative value of DNA profiles generated from specific types of sexual assault evidence samples: cellular mixtures resulting from digital penetration, consisting of trace hand epidermal cells and vaginal cells, and from mouth-to-skin contact, consisting of trace hand epidermal cells and cellular components of saliva (i.e. buccal cells). Cell characterizations were performed using imaging flow cytometry (IFC) and subsequent analysis with data imaging and statistical analysis software. The data collected was used to create cell filters within the data imaging software, resulting in a method to effectively filter all of the hand cells out, potentially isolating a vaginal cell signature. Cells filtered with the image analysis filter were then run through a series of discriminant function analyses to enhance classification and to predict group membership of unknown cells from mock forensic samples. The results of this study showed that the vast majority of these unknown cells classified into the expected group, with few misclassifications. Further, correct classifications were supported by high posterior probabilities, in stark contrast to the posterior probabilities accompanying misclassifying cells. These results indicate that this method can successfully differentiate between the cell populations in question, as well as flag possible false positive misclassifications, and may provide a promising new method for front-end analysis prior to DNA profiling. The ability to identify the components of a biological mixture prior to STR analysis will improve efficacy and decrease bias that can often confuse the activity-level conclusions in trace mixture sample analysis.

Keywords: Trace biological samples, Cell populations, Imaging flow cytometry

Introduction

Sexual assault casework relies heavily on DNA profiles generated from evidentiary samples which can include blood stains, semen stains, vaginal swabs, saliva stains, epidermal cells, and, often, a mixture of more than one of these tissue/body fluid types. One of the most studied cell mixture systems in forensics involves vaginal epithelial cells and sperm cells. Validated serological methods exist for this mixture system to indicate the presence of semen and to microscopically confirm the presence of sperm cells[1]. Most importantly, a validated DNA isolation method exists to separate these tissue types found in a sexual assault mixture for DNA profile interpretation, i.e. differential extraction[2,3]. Despite this, other sexual assault mixtures involving vaginal epithelial cells, epidermal cells from hands and/or saliva, such as those resulting from digital penetration or mouth-to-skin contact, have limited or no validated presumptive *or* confirmatory methods for body fluid differentiation/separation. When these sample types are analyzed, DNA mixture interpretation may be the only method to infer the presence of multiple cell types and attribute contributor profiles[4].

Moreover, lack of serological information for DNA samples can hinder associated activity level-analyses. Where DNA testing can confirm *whose* biological material is present in a sample (i.e. sub-source data), and serological testing can confirm *what types* of biological material are present in a sample (i.e. source data), activity-level analysis aims to answer *how* the sample was deposited[5,6]. Though activity-level analysis can be a helpful contextualizing tool, there are significant challenges in using sub-source data alone to answer the activity-level questions often most pertinent in criminal cases[5]. Additionally, increasing emphasis has been put on the analysis of low-template DNA samples due to increases in sensitivity in forensic methodology[7,8]. “Touch” or trace DNA samples, consisting of epidermal cells that have been

transferred from a person to another person or object during physical contact, often contain low levels of DNA[9,10], which may be difficult to interpret when biological mixtures are present, and can then be controversial in court due to the questionable reliability of the profiles obtained[7,8]. In cases where high levels of DNA is obtained from both contributors, such as a swab from an alleged digital penetration, one might infer the presence of a DNA-rich bodily fluid. However, with lower levels of DNA, that kind of inference is unreliable or even impossible, compounding the challenges surrounding probative value. Methods that are capable of providing source level information for a range of cell types can add critical probative value and strengthen activity-level analyses, however, such methods are currently limited[11]. The reliability and probative value of conclusions drawn from sexual assault samples, especially with cases involving trace or low-template DNA, would improve greatly with the development of such a method.

Serological testing using body fluid presumptive and confirmatory tests is a crucial task in the processing of sexual assault samples. Many of these techniques rely on the subjective interpretation of microchemical reactions and colorimetric changes caused by enzymatic activity, are considered presumptive rather than confirmatory, and are not accompanied with any statistical metrics that could further support the confidence with which an indication is made[1,12,13]. For example, body fluid presumptive tests exist for both saliva (Phadebas® Amylase test, Starch-iodine test, RSID™ Saliva) and vaginal fluid (Acid-Schiff reagent, immunohistochemical tests for estrogen receptors, rRNA amplification from bacteria found in the vagina)[1,12], but these tests screen for enzymes, other proteins and/or nucleic acids that are not specific to saliva or vaginal fluid. This means that these body fluids can be indicated, but the tests will experience false positives when used on body fluids with similar or shared biological compositions[1,12].

There remains a significant need for new approaches that differentiate between and/or identify each of these target sexual assault cell populations, i.e. vaginal epithelial cells, hand/epidermal cells, and saliva/buccal cells. One promising but unexplored approach is to use imaging flow cytometry (IFC) on the front-end of the DNA workflow for rapid characterization of cell types. IFC is a high-throughput and non-destructive method for characterizing physical as well as optical attributes of individual cells through autofluorescence profiles. Thousands of microscopic images of individual cells can be captured in bright field, dark field and multiple fluorescent channels in a matter of minutes, allowing for analysis of morphology, spatial arrangement, and autofluorescent signatures[11,14]. In sexual assault cases, using imaging flow cytometry workflow may be useful in differentiating hand epidermal cells from vaginal cells and buccal cells due to the intrinsic biological differences between these cell populations. Epidermal cells, which are directly exposed to the environment, are protected (and protect our bodies) from exogenous materials via keratinization[15,16]. Cell types that are not exposed to the environment, however, tend to be either partially or not keratinized at all, and include epithelium that lines the buccal and vaginal cavities. Differences in keratinization lead to very different morphologies of each cell type[15], which can be captured using IFC. Additional differences between these cell types that can be captured by IFC, such as size and thickness, can also be used to differentiate and characterize cells. The development of this new forensic signature system could increase the probative value of DNA profiles derived from these cells and increase the reliability of associated activity-level analyses.

Materials and Methods

Obtained Samples for Mock Casework Analysis

For this project samples were collected at both Virginia Commonwealth University (VCU) and at the Ministry of the Solicitor General (SOLGEN) Ontario Centre for Forensic Science in Ontario, Canada. The samples obtained from SOLGEN's forensic laboratory include four reference vaginal samples, nine reference hand epidermal samples, two reference saliva samples, and ten digital penetration samples. All samples were collected, dried and stored at room temperature for approximately two weeks. They were then packaged and shipped to our laboratory at VCU, where they stayed packaged until analysis.

VCU samples were collected following VCU-IRB approved protocol HM20000454_CR8, and include eight reference hand samples, seven hand swabs taken after saliva had been deposited and dried ~10 minutes (saliva-on-hand samples) and five reference saliva samples (buccal swabs). All samples were collected and dried at room temperature, swab up, in a fume hood from one day up to one week prior to analysis.

Mock Casework Sample Extraction

Elution of all dried mock-casework and reference swabs was performed by an initial incubation in 1.5mL of sterile, 1X phosphate-buffered saline (PN: 119-069-131; Quality Biological™; Gaithersburg, MD, USA) for 10-15 minutes. Following incubation, all swabs were vortexed for one minute, and filtered into 1.5mL microcentrifuge tubes with 100μm filter paper. To complete elution, samples were centrifuged at 21,130xg for 10 minutes and concentrated down to 25-50μL by removing the supernatant, taking care not to disturb the cell pellet. After the supernatant was removed, the cell pellet was resuspended by vortexing prior to IFC analysis.

Imaging Flow Cytometry using AMNIS Image Stream

Samples were analyzed using an Amnis® Imagestream X Mark II (Luminex Corp.; Seattle, WA, USA) equipped with 405nm, 488nm, 561nm, and 642nm lasers. Voltages were set to 120mW, 100mW, 100mW, and 150mW. Images of individual events were captured at 40x magnification in five fluorescent detector channels and a brightfield channel. The channels were labeled: channel 1 (430-505nm; violet), channel 2 (505-560nm; green), channel 3 (560-595nm; yellow), channel 5 (640-745nm; red) and channel 6 (745-780nm; side scatter). Channel 4 was used to capture brightfield images. Prior to insertion onto the flow cytometer, concentrated samples were vortexed to ensure resuspension of cells. In Amnis® Imagestream software, samples were loaded onto the IFC by clicking "load samples". Once loaded, "5,000 observations" was selected and "acquire" was clicked. The IFC ran until 5,000-10,000 images were acquired for each sample. Once all samples were analyzed, the resultant raw image (.rif) files were exported into IDEAS® Image Data Exploration and Analysis Software (Luminex Corp.; Seattle, WA, USA) for analysis of imaged cells.

IFC Data Analysis

Cell populations used for digital penetration mock casework analysis (reference vaginal and reference hand samples, digital penetration hand swabs, and saliva-skin hand swabs) were processed with a series of 'gates' to identify and isolate an informative subpopulation of cells with features specific to vaginal cells (to the exclusion of hand epidermal cells). These gates targeted cells 600-5,000μm^2 or ~28-80μm in diameter, in-focus cells with > 50 Gradient RMS values, a 0-10 pixel contrast range, and a second area filter of within a 1170-3810μm^2 area range or ~38-70μm in diameter. Cells passing all four filters were tabulated in Table 1, and all of the feature values of these cells were exported into Microsoft Excel (Microsoft Corp.; Redmond,

WA, USA) which were used in downstream analysis of digital penetration mock casework samples through discriminant function analysis.

A complete description of all feature values captured by IFC and exported for analysis can be found in the IDEAS® Image Data Exploration and Analysis Software User's Manual (Luminex Corp.; Seattle, WA, USA)[17].

Statistical Analysis of Cell Feature Values

Reference vaginal, reference hand epidermal, mock casework digital penetration, and mock casework saliva-skin sample feature values were analyzed using a discriminant function analysis (DFA) in IBM SPSS v28 (IBM Inc.; Armonk, NY, USA). The cell feature values assessed for vaginal reference, hand epidermal reference and digital penetration samples were run as three separate groups. Digital penetration and hand epidermal reference samples were coded as the two cell groups in the DFA, sequentially blinding each digital penetration sample. This was done changing the frequency values for the selected sample from one to zero in the SPSS software, which allows the software to treat the samples with a frequency of zero as an unknown sample. Blinding of the samples was done in order to create a mock forensic unknown sample, to see if the blinded sample could be classified correctly with all of the reference samples for that cell population. Blinding of each digital penetration sample was followed by sequential blinding of each reference hand sample, of which the resulting cell counts are tabulated in Table 2 (19 DFAs total, one for each blinded sample). Digital penetration and saliva-on-hand samples were analyzed using SPSS, and were run in a two-group DFA, sequentially blinding digital penetration, then saliva-on-hand samples, the results of which can be viewed in Table 3 (12 DFAs total, one for each sample).

In addition to tabulation of cells classifying as digital penetration, calculated posterior probabilities[18] and multivariate distances were also tabulated (for each cell, as well as averages across each sample).

Correlation between IFC Data and DNA Yield

Seven replicate samples (14 total) were obtained per the swabbing methods listed in Table 4 with diH_2O-moistened sterile cotton swabs. One replicate set was set aside on a tube rack, swab-up, and dried at room temperature in a laminar flow cabinet, in preparation for later IFC analysis. The other set of replicates were cut and added to 1.5mL microcentrifuge tubes and subjected to DNA extraction via DNA IQ™ System (PN: DC6701; Promega Inc.; Madison, WI, USA). In addition to the replicates, a reagent blank was made by moistening a sterile, cotton swab with diH_2O. Once the replicates and reagent blank were cut and placed in their respective 1.5mL microcentrifuge tubes, 300µL of pre-mixed DNA IQ Lysis Buffer was added to all samples. The samples were then incubated at 70°C for 30 minutes. After incubation the samples were vortexed (approximately three seconds), and the cutting was transferred from the tube to a spin basket (PN: V122A; Inc.; Madison, WI, USA). The spin basket was then placed back in the tube, cap left open, and the samples were centrifuged at 21,130xg for two minutes. After centrifugation, the baskets containing the swab cuttings were removed and discarded, and 100µL of DNA IQ Lysis Buffer was added to each sample. The stock DNA IQ Resin was vortexed for 10 seconds, 7µL of which was added to all samples. The samples were vortexed and let stand at room temperature for five minutes, vortexing every minute, then quickly vortexed again before the tubes were placed on the DNA IQ magnetic stand. Without disturbing the resin:DNA pellet, the supernatant was removed from all tubes and discarded. This step was repeated once more.

Once the lysis buffer steps were complete, 100µL of DNA IQ Wash Buffer was added to each tube, followed by vortexing for approximately two seconds. The tubes were then placed back on the DNA IQ magnetic stand and the supernatant was carefully discarded. This wash step was repeated two more times for a total of three washes with the DNA IQ Wash Buffer. After the washes, the tubes were opened and let stand on the DNA IQ magnetic stand to air-dry for five minutes at room temperature. After drying, 100µL of DNA IQ Elution Buffer was added to the pellet and the lids were closed, followed by an approximately two second vortex and incubation at 65°C for five minutes. After incubation, the samples were briefly vortexed and placed on the DNA IQ magnetic stand. The isolated DNA in solution was removed and placed into new, labeled 1.5mL microcentrifuge tubes, and set aside at room temperature briefly (~10 minutes) prior to beginning the qPCR protocol.

To begin the qPCR protocol, 13 0.2mL Chai Bio PCR tubes and caps (PN: S02132; Chai Bio; Santa Clara, CA, USA) were obtained. Five of these PCR tubes were labeled according to the Human DNA standards in the Femto™ Human DNA Quantification kit (PN: E2005; Zymo Research; Irvine, CA, USA): Standard 1 (20ng), Standard 2 (2ng), Standard 3 (0.2ng), Standard 4 (0.02ng) and a negative control (sterile diH_2O). The remaining eight tubes were labeled to correspond to the replicate samples and reagent blank. Once thawed, 18µL of Femto premix was added to all 13 PCR tubes, followed by addition of 2µL of each sample / standard to their respective labeled tubes. The qPCR program was as follows: 95°C for 10 minutes (initial denaturation) followed by 40 cycles of 95°C for 30 seconds, 59°C for 30 seconds and 72°C for 1 minute, and a final extension at 72°C for 7 minutes.

Once the reaction was complete, a standard curve was created from the standard quantitation cycle (Cq) values, which were used to determine concentrations of DNA in each

sample (ng/100µL). The second set of duplicates were run on the IFC according to the extraction and IFC methods listed above, and counts for large cells, large and in-focus cells, large, in-focus and low contrast cells, and large, in-focus, low contrast and largest area cells were obtained according to the analysis of .rif files workflow listed above. We noted that in the reagent blank of one experiment (Sample 8), 200pg was quantified, showing some level of exogenous DNA in the PCR reaction.

Statistical Analysis Comparing DNA Yields and IFC Data

DNA yields obtained from the qPCR reactions and cell feature values obtained from the IFC were analyzed in RStudio v1.4.1717 (RStudio Team 2021; Boston, MA, USA), using Pearson correlation coefficients. Correlation coefficients for dozens of feature values predicted to have the strongest correlation to DNA yields were analyzed (intensity, intensity ratios, brightness detail intensity R3 pixel increment and R7 pixel increment, area, compactness, circularity, aspect ratio, etc.). The two averaged feature values (area and compactness) with the strongest correlations were plotted in simple linear regressions with DNA yield as the response variable.

Results and Discussion

Digital Penetration Mock Casework Mixtures and the Image Analysis Filter

The first step in this new workflow was to identify the subpopulation of cells hypothesized to have the most probative and forensically relevant feature values which could be used to distinguish between the mixture of cells present in a sample. For digital penetration mixture analysis, we created a series of morphological filters to selectively differentiate hand epidermal cells from vaginal cells. The first two gates within the filter select for cells that are

large (e.g., 600-5,000μm^2 or ~28-80μm in diameter) and cells that are in focus (> 50 Gradient RMS, which measures the focus quality of an image). These initial two gates serve to remove out-of-focus images, small biological debris, swab fibers, and bacteria, to ensure that the only intact cells are being compared across all samples.

The next two gates were designed to preferentially select for cells displaying sizes and morphological features represented more frequently in vaginal cells compared to epidermal cells. Analysis of vaginal reference cell populations in IDEAS® software indicated that individual cells had lower optical contrast pixel values in brightfield images, whereas epidermal cells had higher contrast pixel values (8.4 and 15.5 values in Contrast Channel 4 for vaginal and epidermal cell types respectively). This is consistent with previous observations[11] as well as histological surveys of each source tissue[19,20,21] which indicate that thinner cells, such as vaginal cells, will correspond with lower optical contrast, and thicker cells, such as hand epidermal cells, will have greater optical contrast. To target this unique characteristic, the next gate in the image analysis filter selects for cells within a 0-10 pixel contrast range to include low contrast cells, which were inferred to be predominantly vaginal cells through the survey of reference vaginal average contrast values cell in our data set.

Finally, vaginal superficial and intermediate squamous cells are larger (1,250-1,620μm²) than hand epidermal corneocytes (900-1,000μm²)[11], which was supported by a survey of reference vaginal and reference hand average area feature values in the brightfield channel (1546μm² and 949μm² values in Area Channel 4 for vaginal and epidermal cell types respectively). To target this difference in size, the last gate in the image analysis filter targeted cells falling within a 1170-3810μm² area range (approximately 38-70μm in diameter), to further select for cells falling within an area range specific to vaginal cells. The results of the image

analysis filter on reference vaginal and reference hand samples, are listed in the first two sections of Table 1.

There is a clear difference between the number of cells passing each gate for vaginal and hand reference samples, e.g., 120 - 1,105 cells passing all four gates for the four vaginal reference samples and 0-10 cells passing all four gates for epidermal reference samples. This demonstrates that a very small subset of hand cells are large and low contrast, with six out of eight hand samples having one or fewer cells possessing features specific to vaginal cells within the image analysis filter. These cells were visually consistent with aggregations of cells, consistent with the larger cell sizes detected.

Overall, these results indicate that the image analysis filter is successfully selecting for a subpopulation of cells that differentiate vaginal cells from epidermal cells present in a dried sample.

Next, the image analysis filter was applied to mock digital penetration samples, and the number of cells selected in these samples ranged between 0-745, with five out of eight samples showing over 60 cells. These counts are higher from the reference hand sample baseline, and indicate not only the presence of vaginal cells, but also that the filter is still successfully selecting for vaginal cells present on the hand surface and filtering out the hand cell baseline within the mixtures.

Four of the digital penetration samples retained comparatively fewer cells (20A, 21B). This may be a function of swab size and/or pre-sampling activities. These digital penetration samples were all processed with only a half swab for IFC analysis rather than full swabs, which could account for fewer cells. Additionally, sample documentation for 21B, 12 states that the hand used in the digital penetration act was washed multiple times prior to sampling, therefore

any traces of vaginal cells could have been removed prior to analysis. Sample 21B, 7 has documented “unclear” activity, which could also implicate hand washing or a sampling time much later than minutes after a digital penetration event. Pre-sampling activity may have impacted digital penetration sample 23B, 28 which was sampled three hours after a digital penetration act and showed only 26 cells. Conversely, samples 14A – 30B are all full swabs and were all collected 5-15 minutes after a digital penetration act, and do in fact show a large increase in cell counts. However, these samples were purposely collected minutes after a digital penetration act to maximize cell yields, whereas true forensic evidentiary samples may not be collected minutes after an assault, and therefore may not have as many cells present as these generated mock casework samples. Because of this, variation in swab sample sizes and pre-sampling activities for digital penetration mixture samples may skew initial conclusions from the image analysis filter, making comparisons between digital penetration cell counts and reference hand cell counts difficult and unreliable if cell counts are low in both groups.

Overall, these results demonstrate that the image analysis filter was able to select for vaginal cells, both in reference vaginal samples and mock casework digital penetration samples. When smaller swab samples were used, and with longer time spans between deposition and sampling, diminished cell yields were observed. Even though ten or fewer cells from each of the reference hand samples made it through the filter, comparing between reference hand samples and digital penetration samples with similarly low cell yields does not give enough information to indicate the presence of vaginal cells. To enhance differentiation of these cell subpopulations, additional methods for characterization were tested.

Characterization of Digital Penetration Mock Casework Samples using Discriminant Function Analysis

To differentiate vaginal cells from hand cells, a DFA was performed using all cells that passed through the large, in-focus and low contrast vaginal filter (column two of Table 1). A DFA is a statistical ordination technique used for two purposes: to model variation between user-defined sample groups based on their response variables, and to classify unknown or new observations into the user-defined groups[22]. Unlike other ordination techniques, a DFA models variation between groups by *maximizing* inter-group differences between selected groups. It also predicts membership of unknown samples and where potential areas of overlap may be, making it a powerful tool when used in a for unknown forensic sample classification[22,23,24].

Cell populations from the digital penetration samples were organized so that the user-defined sample groups corresponded to vaginal tissue and hand palmar surface. The first DFA is a three-group DFA between reference hand samples, reference vaginal samples, and mock digital penetration samples (results in Figure 1). The resultant scatter plot (where each point represents a single cell) shows three distinct clusters, with areas of overlap, especially between reference hand and mock digital penetration mixtures. Multivariate differences between each group were significant (Wilks lambda = 0.065, p-value < 0.0001). The results show the vaginal cell population is the most distinct group, displaying the least amount of overlap between all three clusters, and separation from the other two clusters achieved on both the x and y-axis. The results also show that the reference hand and mock digital penetration cell clusters are more similar to one another, with much more overlap between the clusters, and differentiation only occurring on the y-axis. Similarity between reference hand and digital penetration samples is expected due to hand cells being present in both cell populations (both samples are generated from hand swabs).

These results show that the modeled variation between the groups is distinct enough that prediction of unknowns into these groups is possible, even if there is some overlap.

The other important difference between a DFA and other ordination techniques is that posterior probabilities/likelihoods can be calculated using the Mahalanobis distances for each data point/cell. In the context of this research, the posterior probabilities for individual cells reflect the likelihood with which a cell is correctly classifying to a group[18]. For example, a cell from a blinded digital penetration sample that is correctly classified by the DFA into the digital penetration cell population with a posterior probability close (0.90 - 0.99) or equal to 1 indicates high likelihood in the group classification of the cell being correct. The Mahalanobis distance, otherwise known as the distance to a group's centroid, is another metric to assess the multivariate relationship between an unknown cell and a reference cell cluster. This value measures the similarities between unknown and known observations through a distance from the group "centroid" or mean value of the discriminant score for a given cell population[25,26,27]. Cells with small distances from a group centroid reflect a greater similarity to that sample group than cells with large distances. Use of posterior probabilities and Mahalanobis distances can therefore infer the likelihood with which the DFA classified an observation correctly, as well as the similarity of a classified unknown observation to the rest of the known observations in a group.

To test the accuracy of predictive classification into these groups, DFAs between reference hand and mock digital penetration samples were run with blinds to simulate an unknown forensic sample. Each mock digital penetration sample was sequentially blinded to determine if the DFA could correctly classify the blinded sample with the rest of the digital penetration cell population, rather than misclassify it with reference hand samples. The same procedure was then carried out with the reference hand samples as the blinds.

Table 2 displays the eight metrics for the classification: number of cells classifying as digital penetration, the number of cells classifying as digital penetration with posterior probabilities greater than 0.99, 0.95 and 0.90, number of cells classifying as digital penetration with posterior probabilities greater than 0.95 and Mahalanobis distances less than 1.0, and finally mean posterior probability, and mean and median Mahalanobis distances for all cells classifying as digital penetration. The results displayed in Table 2 show a large number of cells classifying as digital penetration (2,395 cells across seven samples for an average of 342 cells classifying correctly per sample). The posterior probabilities accompanying correctly classifying digital penetration cells have a small range (0.86-1.0). The distances accompanying digital penetration cells also have a small range (0.256-5.012; Std deviation 1.62). There are two samples not aligned with this trend (23B, 28 and 21B, 12). These two samples are the only samples with longer sampling gaps and washing of the hands, and have comparatively fewer cells correctly classifying, with fewer cells accompanied with large posterior probability values.

Table 2 also shows the results of sequentially blinded reference hand samples, which demonstrate a drastically smaller number of cells from reference hand samples misclassifying as digital penetration (163 cells across 12 samples for an average of 13 cells misclassifying per sample). The posterior probabilities accompanying the misclassifying hand cells range from high likelihood (0.99) to no likelihood (0) and have a standard deviation of 0.27; the same pattern is reflected in the distance values for the hand cell population: the distances accompanying the misclassifying hand cells range from 1.087 - 64.565 to the group centroid and have a standard deviation of 18.13.

The differences in cell counts, posterior probabilities, and distances between these two cell population populations indicate that even though misclassifications (i.e. false positives) are

possible, vaginal cells can be differentiated from reference hand cells in digital penetration mixtures. Furthermore, results from the cell classification and posterior probabilities obtained in Table 2 indicate the possibility to define an interpretation threshold for hand cells misclassifying as vaginal cells, which could eliminate false positives altogether.

Differentiation of Digital Penetration Mock Casework Samples from Saliva-on-Hand Mock Casework Samples

The next step in this workflow was designed to differentiate vaginal cells from the cellular components of saliva (e.g. buccal cells). Differentiation of vaginal cells from saliva/buccal cells is forensically relevant in cases where one party claims that deposited cells or DNA are present because of a different, unrelated activity that was not an assault[9,10]. For example, the defense may claim that the victim's DNA or cells are present on the defendant, not because of a sexual assault event, but because the victim needed assistance that lead to touching of the face or mouth (e.g. in cases of vomiting). The cellular components of saliva are histologically similar to vaginal cells[11,20,21], and share many of the same cell signatures that can be targeted by IFC, as demonstrated by the large number of cells from mock saliva-on-hand samples that evaded the image analysis filter (Table 1). Further, a survey of average area and contrast feature values from the reference buccal samples in our data set showed that buccal cells averaged 1720μm^2 in area and 6.2 in contrast, which are similar to the vaginal averages reported earlier (1546μm^2 and 8.4 respectively) and within the range that the last two gates in the image analysis filter is targeting. Seven saliva-on-hand samples were generated by inserting index and middle fingers into the mouth, and then letting saliva dry on the hand for ten minutes prior to sampling. The image analysis filter used on these generated mock saliva-on-hand samples shows

that a range of 12 - 693 cells are being retained through all four gates within the filter, with five out of seven samples retaining more than 50 cells, resulting in a very high overlap in cell counts between reference vaginal, mock digital penetration, and saliva-on-hand samples. This overlap is expected due to the similarities between vaginal cells and buccal cells mentioned previously.

Results from the image analysis filter on saliva-on-skin samples indicate that vaginal cells and buccal cells require a differentiation method with much stronger classification abilities than just comparison of cell counts resulting from the image analysis filter. These results reinforce the conclusion from earlier: the filter is not, by itself, enough to provide a robust differentiation and characterization framework for epidermal, buccal, and vaginal cells. To enhance differentiation of buccal and vaginal cells, additional methods for characterization are warranted.

The same sequentially blinded 2-group DFA classification scheme was used to determine if vaginal cells can be differentiated from saliva by using digital penetration and saliva-on-hand samples as the known groups. The results of these analyses are listed in Table 3. This table includes all criteria from Table 2 as well as a new criterion: cells classifying as saliva-on-hand – this criterion was included to compare the rate of misclassification within both cell populations.

The results in Table 3 show hundreds of cells correctly classifying across the majority of digital penetration samples, ranging from 9 to 1353 cells per sample; five out of seven of these samples had 140 cells or more classifying as digital penetration. In addition, the number of cells misclassifying as saliva-on-hand within these samples is comparatively very small, ranging from 0 to 155 cells, with five out of seven samples having fewer than 30 cells misclassifying with the saliva-on-hand samples.

Digital penetration cells classifying correctly had a range of mean posterior probabilities from 0.821 to 1 (SD = 0.09) and a range of mean distances from 0.5 to 23 (SD = 8.15). In five out of seven digital penetration mixtures, a greater number of cells correctly classified as digital-penetration than misclassified as saliva-on-hand, with posterior probabilities all greater than 0.90, indicating a high likelihood that these classifications are correct. The distances for these cells are all small distances from the group centroid (<1.81), except for an outlier, sample 30B, with a distance of 22.986. Though the distance is large, the classification scheme identified the sample correctly, demonstrating that even when variance between cells in the same population is high, correct classifications can still be made.

Samples 14A and 16A showed markedly lower similarities to vaginal cell populations (mean posterior probability = 0.821 and 0.776, dist. = 3.547 and 2.951, respectively) and had more saliva misclassifications than digital penetration correct classifications. Lower degrees of confidence and similarity for cells correctly classifying as digital penetration, and a high number of cells misclassifying as saliva could be an indicator that saliva is actually present on the samples collected, which is plausible when considering that saliva is often used as a sexual lubricant.

In comparison, saliva-on-hand samples had much fewer cells misclassifying as digital penetration, and a greater number of correct classifications. Buccal cells misclassifying as vaginal had a range of posterior probabilities from 0.742 to 0.880 (SD = 0.06), and a range of distances from 2.611 to 4.625 (SD = 0.84). Misclassifying saliva-on-hand cells display posterior probabilities that are all less than 0.90 and distances that are all greater than 2.00 from the group centroid, indicating that the decision on where to set an interpretation threshold for these cell types may be relatively simple.

The differences in cell counts, posterior probabilities, and distances between these two cell population populations indicate that even though misclassifications and false positive results are possible, vaginal cells present in a digital penetration mixture can be differentiated from saliva deposited on skin. Furthermore, results from the cell classification and posterior probabilities obtained in Table 3 indicate the possibility to define an interpretation threshold for buccal cells misclassifying as vaginal cells. Misclassifying buccal cells have accompanying low posterior probabilities and large distances that signal a false positive even more clearly than with hand cells, which further supports the conclusion that saliva can be differentiated from vaginal cells.

Overall, the results from the image analysis filter used in conjunction with the DFA classification scheme may be a promising new method for indicating hand cells, components of saliva (e.g. buccal cells) and most importantly in this context, vaginal cells. Clear differences in cell counts, posterior probabilities and multivariate distances allows for effective discrimination of true positives and false positives. Further testing on mock casework samples will assist in building larger reference databases, defining interpretation thresholds, and concomitantly narrowing the false positive rate for this method to a point that it might one day be used in forensic casework laboratories.

Using IFC Data to Predict DNA Concentration in Forensic Casework Samples

The goal of this experiment was to determine whether morphological and/or autofluorescence data obtained from the IFC could be used to predict DNA content in sexual assault forensic evidentiary samples. More specifically, it aimed to determine if the number of cells obtained from samples involving saliva deposited on skin, specifically fingers, could be

used to indicate potential DNA yield for a given sample. To test this, seven different types of saliva samples were generated, each of which had varying depositional activities (summarized in Table 4).

The results of the qPCR and cell counts from the IFC are listed in Table 5. The standard curve for the qPCR reaction was calculated from four standards: 20ng, 2ng, 0.2ng and 0.02ng of input DNA, with a resulting R^2 of 0.9947, regression equation y = -4.922x + 18.386 (standard error of measurement +/- 0.01ng). The results show that the calculated concentrations may correlate to the activity levels in each sample: Sample 1 involved full insertion of two fingers into the mouth and was meant to be the highest concentration sample, which it was by far, with a DNA yield of 31.6ng. The next three samples (2-4) all had gradual decreases in activity level (less of the finger being licked from sample to sample) and correlating decreases in DNA concentration (4.7ng, 3.4ng, 1.95ng, respectively). The last three samples had the lowest concentrations overall and were meant to simulate cases where activity level was low enough that evidence of saliva deposition may not be obvious, due to a lack of abundance of cells detected with saliva/buccal morphology in the sample. The sample with the lowest DNA concentration (Sample 5) was the sample with the least amount of finger contact with the tongue/mouth, and also the smallest swabbed surface area (pinky finger) with a DNA yield of 0.45ng. The lower DNA yield is consistent with the smallest surface area being swabbed, and is also consistent with the activity level being so low that the potential for deposited buccal/saliva cells was also very low. The concentrations for Samples 6 and 7 were slightly higher than Sample 5, with yields of ~1ng of DNA. Sample 6 was a variation of Sample 5, with the same activity level but swabbed with more surface area, and Sample 7 was a variation of Sample 6, with slightly more contact between tongue and finger (aiming for more saliva deposition) and a

swab of the whole hand (larger surface area swabbed). Because the concentrations of these two samples were so similar, this could possibly indicate that the slightly different saliva depositions may have had little to no effect on overall DNA yield, meaning that the activity levels were potentially not different enough to see markedly different DNA concentrations in these samples. These results may also indicate that a potential source of the DNA obtained from Samples 5, 6 and 7 are from extracellular DNA rather than deposited saliva. We predicted that different activity levels should result in a clear difference in DNA yields, but if the results of our experiment followed this prediction, Sample 7 would have a larger amount of DNA than Sample 6 and 5. A possible explanation for these unexpected results from Sample 7 may be because the whole hand was swabbed, rather than just a finger: in swabbing the whole hand, it is possible that saliva was spread onto other parts of the hand and left behind, resulting in lower DNA yields. Further testing with many more samples with similar activity levels and swabbing techniques is required to truly illuminate the role that activity level plays in DNA concentration in samples like these.

The results from the cell counts in Table 5 do not show strong evidence of a linear relationship with DNA yield. Sample 1 retains the most cells through all 4 gates and has the highest concentration of DNA, however, the rest of the samples do not follow this pattern. Samples with lower concentrations are still retaining cells through all four filters, some with even more cells present than samples with higher concentrations. The one factor that does show correlation with the DNA yield were the number of visually obvious squamous cells present in the cell population. For example, Sample 1 is the highest concentrated sample (31.6ng) and visually, is clearly a mixture of both epidermal and buccal cells. Overall, visual analysis of these cells shows that there seem to be more epidermal cells present than buccal cells prior to any

filtering, but as epidermal cells are being filtered out in the later gates of the image analysis filter, the cells present become almost completely buccal. Moving onto Sample 2, with the second-highest DNA concentration (4.7ng), a mixture of buccal and epidermal cells is still obvious through visual analysis, however the ratio of buccal to epidermal is not as high as in Sample 1, i.e. there are many more epidermal cells present in comparison to buccal cells prior to filtering. Many of the epidermal cells filter out by the final gates, leaving a predominant buccal cell signature once all four gates have been applied. This trend continues with Samples 3 and 4, with buccal cells visually present but being sequentially diluted or drowned out by the number of epidermal cells present in the samples as well. Once the DNA concentrations match those obtained in Samples 5, 6, and 7 (0.45ng, 1.08ng and 1.12ng, respectively), visual analysis becomes much more difficult, as there are far fewer cells that morphologically stand out as true buccal cells. Cells present through all four gates in these lower concentrated samples are still predominantly epidermal cells, with possibly one or two buccal cells present, but with only visual analysis to assess this, it is impossible to be sure.

This trend demonstrates that there may not be a direct relationship between the types and abundance of cells detected with IFC and amplifiable DNA. This may be partly explained by incomplete differentiation of epithelial cell types by the three criteria used for the image analysis filter (i.e., focus, size, contrast). Aggregations of epidermal cells can cause overlap in cell populations, particularly in calculated area values. A more in-depth survey of morphological signatures specific to components of saliva that can be captured by IFC is required to truly understand the trend between cell counts and DNA concentration.

A final experiment on the correlation between DNA concentration and specific morphological and/or fluorescence variables was performed via simple linear regression (Figure

2). The DNA concentrations for the seven samples were used as the response variable. Cell feature value averages from all samples were calculated and used as the predictor variables. After calculating the Pearson correlation coefficient for dozens of averaged feature values, two were selected as the strongest explanatory variables: cell area in brightfield channel ($r = 0.89$), and cell compactness in brightfield ($r = -0.82$). Interpretation of the correlation coefficients indicates that 89% of change seen in DNA concentration can be explained by cell area, and that 82% of the change seen in DNA concentration can be explained by cell compactness. These results are consistent with squamous cells deposited from mouth-to-skin contact driving DNA concentration: as cell area increases and compactness decreases (squamous cells are larger and less compact than epidermal cells), the concentration of DNA increases. The results of this linear regression taken together with the conclusions from visual analysis of the IFC cell images indicate that squamous cells present on epidermal substrates will drive up DNA concentration in samples and can be explained/correlated with information obtained from the IFC datastream.

Conclusion

The results of the projects in this research indicate that vaginal cells, components of saliva/buccal cells, and hand epidermal cell populations can be successfully differentiated and classified into correct cell groupings.

This methodology may provide a promising new presumptive technique for front-end analysis of sexual assault mixtures prior to DNA profiling. Independently, the image analysis filtering method and discriminant function analyses could not accurately group vaginal, buccal/saliva and epidermal cells into their correct cell populations without large numbers of misclassifications from the overlapping target cell types in question. When the image analysis

filter and the DFA are used together, however, the image analysis filter was selective enough to isolate targeted, forensically probative cell subpopulations that could be further characterized using a DFA.

The cell counts from the DFA between reference hand samples and digital penetration samples suggest that up to approximately 70 cells may misclassify in reference hand samples. Using this method in an operational setting, the decision framework for differentiating hand cells from vaginal cells could set an interpretation threshold at 20 cells classifying as digital penetration with posterior probabilities greater than 0.93. This interpretation threshold would effectively eliminate every single reference hand sample from Table 2 as a digital penetration cell, and retain five out of seven digital penetration samples. The ability to set an interpretation threshold that effectively retains the majority of correct classifications and eliminates all false positives indicates that this method could in fact work as a highly accurate presumptive test for digital penetration events. This is further evidenced by the cell counts from the DFA between digital penetration and saliva-on-skin samples, which suggest that the up to 100 cells originating from saliva deposition may misclassify. Again, if this method were to be implemented in an operational setting, the decision framework for differentiating buccal cells from vaginal cells could have an interpretation threshold set at 20 cells classifying as digital penetration with posterior probabilities greater than 0.90. This threshold would eliminate all of the saliva-on-skin samples, and retain four out of the seven digital penetration samples.

We have demonstrated that our data can be used to set interpretation thresholds that would effectively retain the majority of our true positives, and eliminate all of our false positives in an operational setting. False positives are possible in this classification scheme and in many other presumptive serological tests as well, but the defining difference between this method and other

serological methods is that characterizations from the DFA are accompanied by posterior probabilities and Mahalanobis distances, which allow for a classification of an unknown with an accompanying measure of similarity to a known cell population and a likelihood that the classification is correct. This also means that when misclassifications do occur, they are more likely to be addressed and excluded by a set interpretation threshold.

Further efforts into testing the efficacy of this classification scheme will require a validation effort using hundreds of mock samples with varying activity-levels to set an interpretation threshold that could work in real forensic casework.

The DNA prediction results demonstrate that DNA concentration may have some correlation with activity level through IFC data in samples including squamous epithelial cells through visual analysis of IFC images and through linear regression. This conclusion may be used in casework to determine if a statistically significant number of squamous cells are found on a suspect, and to infer if the DNA concentration correlated to activity level upholds a suspect or victim's explanation for how DNA or squamous cells were deposited. Further research with more samples, improved signatures for IFC analysis of buccal cells, and varying activity levels is required to determine a more in-depth understanding of the relationship between DNA concentration and IFC data.

References

1. Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic science international*, *188*(1-3), 1-17.
2. Yoshida, K., Sekiguchi, K., Mizuno, N., Kasai, K., Sakai, I., Sato, H., & Seta, S. (1995). The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. *Forensic science international*, *72*(1), 25-33.
3. Norris, J. V., Evander, M., Horsman-Hall, K. M., Nilsson, J., Laurell, T., & Landers, J. P. (2009). Acoustic differential extraction for forensic analysis of sexual assault evidence. *Analytical chemistry*, *81*(15), 6089-6095.
4. An, Ja Hyun, Kyoung-Jin Shin, Woo Ick Yang, and Hwan Young Lee. "Body fluid identification in forensics." (2012): 545-553.
5. Biedermann, A., Champod, C., Jackson, G., Gill, P., Taylor, D., Butler, J., ... & Taroni, F. (2016). Evaluation of forensic DNA traces when propositions of interest relate to activities: analysis and discussion of recurrent concerns. *Frontiers in Genetics*, *7*, 215.
6. Taylor, D., Kokshoorn, B., & Biedermann, A. (2018). Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic Science International: Genetics*, *36*, 34-49.
7. Gill, P., Haned, H., Bleka, O., Hansson, O., Dørum, G., & Egeland, T. (2015). Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—twenty years of research and development. *Forensic Science International: Genetics*, *18*, 100-117.
8. Balding, D. J. (2013). Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proceedings of the National Academy of Sciences*, *110*(30), 12241-12246.
9. Williamson, A. L. (2012). Touch DNA: forensic collection and application to investigations. *J Assoc Crime Scene Reconstr*, *18*(1), 1-5.
10. Burrill, J., Daniel, B., & Frascione, N. (2019). A review of trace "Touch DNA" deposits: Variability factors and an exploration of cellular composition. *Forensic Science International: Genetics*, *39*, 8-18.
11. Brocato, E. R., Philpott, M. K., Connon, C. C., & Ehrhardt, C. J. (2018). Rapid differentiation of epithelial cell types in aged biological samples using autofluorescence and morphological signatures. *PloS one*, *13*(5), e0197701.
12. Sakurada, K., Watanabe, K., & Akutsu, T. (2020). Current methods for body fluid identification related to sexual crime: focusing on saliva, semen, and vaginal fluid. *Diagnostics*, *10*(9), 693.
13. Mayes, C., Seashols-Williams, S., & Hughes-Stamm, S. (2018). A capillary electrophoresis method for identifying forensically relevant body fluids using miRNAs. *Legal Medicine*, *30*, 1-4.
14. Doan, M., Vorobjev, I., Rees, P., Filby, A., Wolkenhauer, O., Goldfeld, A. E., ... & Hennig, H. (2018). Diagnostic potential of imaging flow cytometry. *Trends in biotechnology*, *36*(7), 649-652.

15. Wang, C., Stanciu, C. E., Ehrhardt, C. J., & Yadavalli, V. K. (2017). Nanoscale characterization of forensically relevant epithelial cells and surface associated extracellular DNA. *Forensic science international*, *277*, 252-258.
16. Schröder, J. M. (2010). The role of keratinocytes in defense against infection. *Current opinion in infectious diseases*, *23*(2), 106-110.
17. Luminex. (2019). IDEAS® Image Data Exploration and Analysis Software User's Manual. Luminex Corporation. Seattle, WA, USA.
18. Zenkov, V. V. (2019). Evaluation of the posterior probability of a class with a series of Anderson discriminant functions. *Automation and Remote Control*, *80*(3), 447-458.
19. Marks, R., & Barton, S. P. (1983). The significance of the size and shape of corneocytes. In *Stratum corneum* (pp. 161-170). Springer, Berlin, Heidelberg.
20. Schroeder, H. E. (1981). *Differentiation of human oral stratified epithelia*. Karger Publishers.
21. van Eyk, A. D. (2004). Human vaginal mucosa as a model of buccal mucosa for in vitro permeability studies: an overview. *Current Drug Delivery*, *1*(2), 129-135.
22. Huberty, C. J. (1994). Why multivariable analyses?. *Educational and Psychological Measurement*, *54*(3), 620-627.
23. Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic science international*, *261*, 165-e1.
24. Chiu, A. S., & Donlon, D. (2000). Anthropological and forensic aspects of odontological variation in two contemporary Australian populations. *Dental Anthropology Journal*, *14*(2), 20-37.
25. Queen, J. P., Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge university press.
26. Galeano, P., Joseph, E., & Lillo, R. E. (2015). The Mahalanobis distance for functional data with applications to classification. *Technometrics*, *57*(2), 281-291.
27. McLachlan, G. J. (1999). Mahalanobis distance. *Resonance*, *4*(6), 20-26.
28. DiGangi, E. A., & Moore, M. K. (Eds.). (2012). *Research methods in human skeletal biology*. Academic Press.

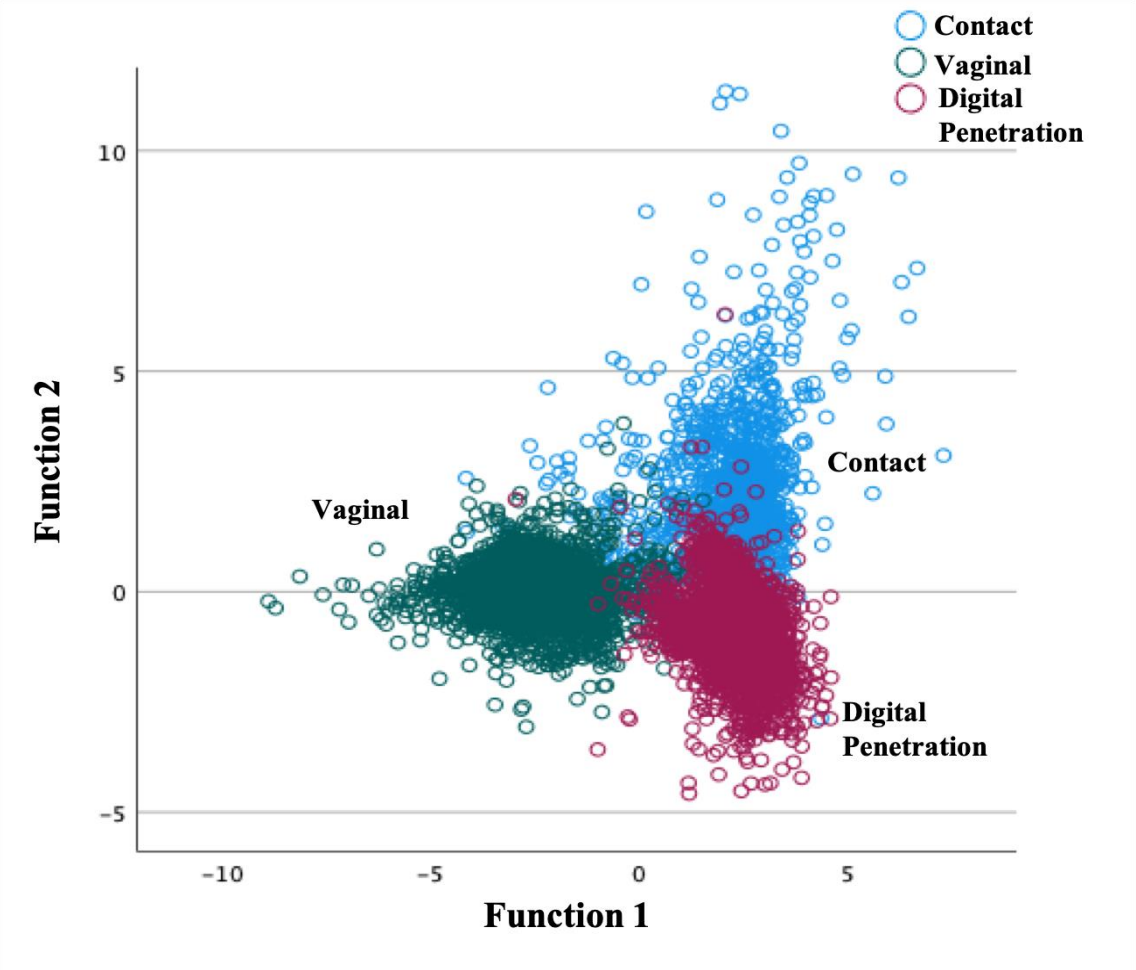
Table 1. Image Analysis Filter Results

Sample	Number of Cells Low Contrast	Number of Cells Low Contrast and Largest Area
Reference Vaginal		
Vaginal 23A	1,168	933
Vaginal 20B, 22	188	120
Vaginal 11	1,298	1,105
Vaginal 10	427	220
Reference Hand Epidermal		
Hand 1	28	8
Hand 5	39	10
Hand 6	43	10
Hand 8	27	4
Hand 9	32	6
Hand 11	28	6
Hand 20A, 1	5	0
Hand 23B, 27	75	9
Hand 19B, 7	6	0
Mock Digital Penetration		
23B, 28; 3hrs post vaginal penetration, no wash	193	26
20A, 31; 5 mins post vaginal penetration	2	0
21B, 7; vaginal penetration, details unclear	2	1
20A, 2; 5 mins post vaginal penetration, no wash	10	2
21B, 12; 20hrs post vaginal penetration, washed multiple times	14	0
14A, 1; 5-15mins post vaginal penetration, no wash	449	62

16A, 2; 5-15mins post vaginal penetration, no wash	503	169
19B, 4; 5-15mins post vaginal penetration, no wash	269	168
23B, 3; 5-15mins post vaginal penetration, no wash	1,115	745
30B, 3; 5-15mins post vaginal penetration, no wash	181	110
Mock Saliva-On-Hand		
A24; 10-mins dried, no wash	716	275
D61; 10-mins dried, no wash	544	289
P94; 10-mins dried, no wash	123	52
K47; 10-mins dried, no wash	905	693
I66; 10-mins dried, no wash	258	60
N27; 10-mins dried, no wash	220	35
H71; 10-mins dried, no wash	95	12

Resultant cell counts obtained from use of the image analysis filter on reference vaginal, reference hand epidermal, mock digital penetration and mock saliva-on-hand samples.

Figure 1. Discriminant Function Analysis Results: All Target Cell Populations



DFA between three target cell populations: Reference hand epidermal cells (blue), reference vaginal cells (teal), and mock digital penetration cells (burgundy).

Table 2. Discriminant Function Analysis Results: Mock Digital Penetration and Reference Hand Samples

Sample Blind	# Cells classif. DP	# Cells > .99PP	# Cells > .95PP	# Cells > .90PP	# Cells > .95PP, < 1 Dist.	Mean PP	Mean Distance	Median Distance
Digital Penetration								
14A	407	167	290	326	109	0.94	1.974	1.902
16A	373	101	226	277	118	0.94	1.537	1.433
19B	258	243	252	254	222	0.96	0.466	0.198
21B, 12 (20hrs)	9	2	7	9	2	0.97	1.456	1.577
23B (5min)	1381	1196	1351	1370	1318	0.99	0.256	0.088
23B (3hrs no wash)	151	15	53	78	4	0.86	2.891	2.912
30B	222	219	220	221	33	1.00	5.012	4.416
Ref. Hand Epidermal								
Hand 1	4	1	2	2	1	0.912	1.904	2.152
Hand 5	20	11	13	16	11	0.921	1.657	0.989
Hand 6	3	1	1	1	1	0.824	2.445	2.557
Hand 8	1	0	0	0	0	0.749	3.351	3.351
Hand 9	9	4	4	6	2	0.907	1.920	2.161
Hand 11	2	1	1	1	0	0.757	64.565	64.565
Hand 19B	4	3	4	4	1	0.994	1.897	0.931
Hand 20A	4	3	3	3	3	0.962	1.087	0.681
Hand 23B	43	19	24	29	12	0.890	2.580	2.035
Hand BFF, A24	0	0	0	0	0	0	0	0
Hand BFF, D61	71	11	16	20	3	0.773	15.269	4.603
Hand BFF, P94	2	0	0	0	0	0.633	3.747	3.747

Results of first series of DFAs between reference hand epidermal samples and mock digital penetration samples, with number of cells classifying as digital penetration, number of digital penetration-classifying cells with posterior probabilities greater than 0.99, 0.95, 0.95, number of digital penetration-classifying cells with posterior probabilities greater than 0.95 and distances less than 1.0, mean posterior probabilities for all cells classifying as digital penetration, and mean and median distances for all cells classifying as digital penetration.

Table 3. Discriminant Function Analysis Results: Mock Digital Penetration and Mock Saliva-On-Hand Samples

Sample Blind	# Cells Classif. DP	# Cells > .99PP	# Cells > .95PP	# Cells > .90PP	# Cells > .95PP, < 1 Dist.	Mean PP	Mean Distance	Median Distance	# Cells Classif. S-o-H
Digital Penetration									
14A	252	18	47	67	3	0.821	3.547	3.723	155
16A	21	2	4	5	2	0.776	2.951	3.121	352
19B	258	247	256	256	214	0.997	0.523	0.288	0
21B, 12	9	6	8	9	7	0.984	0.585	0.379	0
23B, 5min	1353	737	1104	1206	1095	0.962	0.534	0.264	28
23B, 3hrs no wash	140	31	69	96	36	0.904	1.811	1.774	11
30B	221	221	221	221	30	1.000	22.986	17.720	0
Saliva-On- Hand									
A24	62	6	14	18	8	0.774	2.611	3.002	640
D61	17	5	6	7	3	0.822	4.625	3.006	527
I66	20	10	13	13	4	0.880	3.609	2.242	188
K47	99	6	13	21	4	0.742	3.964	3.802	789
P94	0	0	0	0	0	N/A	N/A	N/A	123

Results of second series of DFAs between mock digital penetration samples and mock saliva-on-hand samples, with number of cells classifying as digital penetration, number of digital penetration-classifying cells with posterior probabilities greater than 0.99, 0.95, 0.95, number of digital penetration-classifying cells with posterior probabilities greater than 0.95 and distances less than 1.0, mean posterior probabilities for all cells classifying as digital penetration, mean and median distances for all cells classifying as digital penetration, and number of cells classifying as saliva-on-hand cells.

Table 4. Descriptions of Sampling Activity Levels for Correlating IFC Data and DNA Yield Experiment

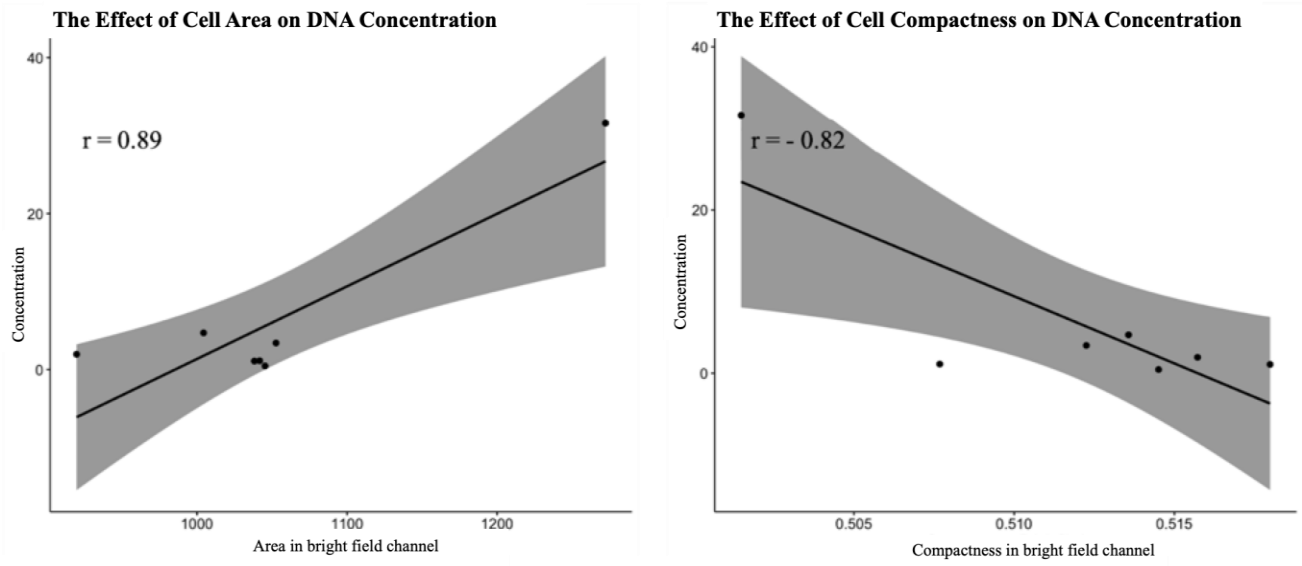
Sample	Notes
1	Left index & middle fingers fully inserted into mouth 1-2 seconds
2	Left ring finger: full finger lick on one side of all three finger sections
3	Right ring finger: lick on one side of top two finger sections
4	Right middle finger: lick on one side of only top finger section/finger pad
5	Right pinky: barely licked top of finger
6	Right index finger: barely licked top of finger
7	Left pinky: lick on one side of top of finger section only, whole hand swab
8	Reagent blank with water-moistened sterile swab

Table 5. Comparative Results between qPCR and IFC Data on seven sample duplicates

Sample	Conc. (ng/100μL)	Large Cells	Large & In-focus Cells	Large & In-focus & Low Contrast	Large & In-focus & Low Contrast & Largest Area
1	31.6	2591	560	309	161
2	4.7	1346	424	218	45
3	3.4	1730	633	373	90
4	1.95	1071	386	164	19
5	0.45	1833	649	318	68
6	1.08	2164	778	218	44
7	1.12	661	209	109	32
8 (RB)	0.2				

Resultant DNA concentrations obtained from one set of seven sample duplicates and reagent blank; resultant cell counts obtained from use of the image analysis filter on the other set seven sample duplicates.

Figure 2. Simple Linear Regressions of the Effect of Cell Area and Cell Compactness on DNA Concentration



Results of two simple linear regressions using DNA concentration as the response variable and cell area and cell compactness in Ch04 (brightfield channel) as the explanatory variables. The gray bands represent a 95% confidence interval.

Vita

Arianna Daisy DeCorte was born on December 19th, 1997 in Manhattan, New York. She attended Albemarle High School in Charlottesville, Virginia, as well as the University of Virginia. In 2020 she graduated from UVA with a Bachelor of Arts in Biology and Italian. After graduating from UVA, Arianna pursued a Master's degree in Forensic Science at Virginia Commonwealth University, concentrating in Molecular Biology. While at VCU, Arianna worked as a graduate research assistant to Dr. Christopher Ehrhardt. Upon graduating from VCU, Arianna will be pursuing a Doctorate degree in Integrated Life Sciences at VCU, staying on with Dr. Ehrhardt's research laboratory.