National Center for Forensic Science
University of Central Florida
P.O. Box 162367 · Orlando, FL 32826
Phone: 407.823.4041
Fax: 407.823.4042
Web site: http://www.ncfs.org/

Biological Evidence

De-convolution of Body Fluid Mixtures: Cell Type Identification and Genetic Profiling of Micro-Dissected Cells

FINAL REPORT

July 31, 2013

Department of Justice, National Institute of justice
Award Number: 2008-DN-BX-K007
(1 October 2008 – 28 February 2014)

Principal Investigator:
Jack Ballantyne, PhD
Professor
Department of Chemistry
Associate Director for Research
National Center for Forensic Science
P.O. Box 162367
Orlando, FL 32816-2366
Phone: (407) 823 4440
Fax: (407) 823 4042
e-mail: Jack.Ballantyne@ucf.edu

ABSTRACT

Mixtures are frequently encountered in forensic DNA analysis due to the involvement of multiple persons during the commission of certain crimes. Mixtures are also often found associated with 'touch DNA' or with other crime scene samples due to unavoidable contamination with previously deposited DNA. Analysis and interpretation of these admixed DNA samples is often quite complex and therefore extremely challenging even more so if the number of donors exceeds two. In certain situations, the presence of a mixture can be anticipated and the cells separated prior to analysis, such as is the case with a 'differential extraction' of sexual assault evidence where sperm cells are present. In many cases, however, the cellular components of a mixture cannot be separated due to the presence of non-easily distinguishable epithelial cells or leukocytes from more than one donor. Interpretation of these mixtures is sometimes further complicated because of degraded DNA or artifacts that arise during the PCR-based analyses. A limited number of tools are available to operational forensic laboratories in order to aid in the interpretation or handling of mixtures. The current work sought to provide a novel strategy for the de-convolution of body fluid mixtures containing non-distinguishable cell types (e.g. epithelial cells and leukocytes) through an isolation of individual cells using laser capture micro-dissection and recovery of genetic material (RNA and DNA) for cell type identification (RNA profiling methods) and donor identification (autosomal STR analysis). Thus the cells comprising mixtures would be separated out according to body fluid source and single source profiles obtained from each of the individual mixture contributors. Uniquely, such an approach would permit the attribution of each DNA profile in a mixture to a particular contributing body fluid.

In this study, we demonstrate the ability to isolate single epithelial cells using laser capture micro-dissection (LCM) and also using a novel micro-manipulation approach. We have developed highly robust and sensitive DNA lysis and amplification strategies, including the use of micro-volume PCR reactions as well as single-tube combined lysis and amplification reactions, for the enhanced analysis of single (or few) epithelial cells isolated from admixed epithelial cell mixtures which can aid in the de-convolution of epithelial cell admixtures. While the routine use of single cell profiling still requires additional work, we also report the development of a novel strategy for a more efficient and informative genotyping of two person body fluid mixtures containing non-distinguishable cell types (e.g. epithelial cells and leukocytes) at approximately equal concentrations through a physical isolation of multiple samplings of groups of cells using LCM and subsequent DNA profiling of the cell population samples. Genotype inference and the attachment of statistical weight are accomplished via the TrueAllele® quantitative computer interpretation system. The combination of multiple binomial sampling of the total mixed cell population to create separate sub-populations with differing weight ratios, together with quantitative computer-based statistical modeling, can provide a significant information gain compared to the standard analytical approach. Additionally, we also report the development of preliminary RNA profiling strategies to provide a determination of the cell type of origin of epithelial cells (buccal, vaginal and skin), which permits further individualization of epithelial cells from admixed samples.

In summary, the methodologies developed in the current work may help to resolve forensic mixtures and permit facile genotyping of individual contributors.

TABLE OF CONTENTS

ABSTRACT 2
EXECUTIVE SUMMARY 4
I. INTRODUCTION 10
A. Statement of the Problem 10
B. Literature Review 10
C. Statement of Hypothesis or Rationale for the Research 12
D. Project Aims 12
II. METHODS 14
DNA Profiling 14
RNA Profiling 18
III. RESULTS 23
A. Statement of Results 23
1. DNA Profiling of Micro-Dissected Cells 23
1.1 STR Profiling of Single and Few Epithelial Cells 23
1.2 DNA Mixture Genotyping by Probabilistic Computer Interpretation[55] 26
1.3 Limiting Dilutions 31
1.4 Enhanced Single Cell Analysis 32
2. Cell Type Identification (RNA Profiling) of Micro-Dissected Cells 35
2.1 Initial RNA Isolation Method Evaluation 36
2.2 Identification of Novel mRNA Biomarkers for Cell Type Identification 38
2.3 Enhanced RNA Isolation and Profiling Methods for Cell Type Identification 45
3. Initial Development of DNA/RNA Co-Isolation Methods 47
3.1 RNA Hybridization Probes for Cell Type Identification 47
3.2 DNA/RNA Co-Isolation: AllPrep Micro kit (QIAGEN) 49
B. TABLES 51
C. FIGURES 63
IV. CONCLUSIONS 95
A. Discussion of findings 95
B. Implications for policy and practice 97
C. Implications for further research 97
V. REFERENCES 99
VI. DISSEMINATION OF RESEARCH FINDINGS 104
A. Publications 104
B. Presentations 104
VII. RESPONSE TO REVIEWERS 106

EXECUTIVE SUMMARY

Introduction

1. It is now a matter of routine for the forensic scientist to obtain the genetic profile of an individual from DNA recovered from a biological stain deposited at the crime scene. However, during the commission of a violent crime two or more people are typically present and may deposit biological material at the crime scene or onto an individual or individuals, thus resulting in a mixture of genetic material. Moreover the shed biological material may be deposited on a substrate that already has previously deposited DNA present, thus contaminating the crime scene deposits with other 'legitimate' sources. This is especially so with so-called 'touch DNA' evidence.

2. With the high sensitivity of PCR based analysis methods, such as STR analysis, the appearance of low level minor contributor profiles are frequently encountered.

3. In some cases it is not possible to deduce the component profiles in admixed samples thus significantly reducing the probative nature of the evidence. Often, mixtures can be easily identified due to morphological differences between cell types, such as in sexual assault evidence that is typically comprised of sperm and vaginal epithelial cells. Several strategies have been developed to aid in the interpretation of this type of mixture. The most commonly used strategy involves the use of a differential extraction which allows for a physical separation of sperm and non-sperm cells prior to analysis. More recently, the use of laser capture micro-dissection (LCM) has been evaluated to isolate sperm cells from non-sperm cells.

4. The above-mentioned strategies are limited to mixtures involving sperm cells. No such strategy or methodology has been developed thus far to physically separate the contributors in a non-easily-distinguishable cell type mixture.

5. Genotype attributions in such mixtures are often complex and time consuming, particularly if the contributors are present in approximately equal portions or the number of donors exceeds two. The identification of a mixture typically relies on a visual examination of the number and relative fluorescence intensity of alleles present at each locus, with the presence of three or more alleles and/or significant allelic imbalance indicating the presence of a mixture. Individual contributor profiles can sometimes be discerned through an assessment of peak height ratios if differences in signal intensities identify major and minor contributors. However this type of basic interpretation can be confounded in forensic casework without prior knowledge of the number of contributors, the ratio in which the contributors are present within the mixture and the extent of degradation. Low copy number samples comprising part of the admixture present additional challenges, with significant peak imbalances, increased stutter peaks and allelic-drop in and drop-out frequently observed.

6. Numerous statistical approaches to the interpretation of mixtures can be applied such as likelihood ratios (LR) or the combined probability of exclusion (Pr(*Ex*), or random man not excluded, RMNE). Despite attempts to provide consolidated guidelines for mixture interpretation, there is currently no single method that accommodates all the challenges of forensic casework samples and that always permits simple statistical calculations to be made.

7. The inability to clearly resolve individual contributors in admixed epithelial cell mixtures and the need for the use of complex statistical interpretations necessitates further research aimed at developing new methodologies to aid operational crime laboratories in the analysis of mixture samples.

Current Study

8. The current study therefore sought to develop novel methods for the de-convolution of epithelial cell mixtures through cell type identification and genetic profiling of micro-dissected cells. This study firstly sought to determine whether DNA and RNA can be co-isolated from individual cells micro-dissected from epithelial and/or leukocyte cell mixtures. Secondly we sought to utilize the RNA isolated from micro-dissected epithelial and/or leukocyte cells to determine the source of the collected cell (blood, skin, saliva, or vaginal secretions) through RNA profiling techniques. Lastly, attempts were made to further individualize the micro-dissected cells, including single cells, by obtaining DNA profiles which can then be attributed to a particular individual. The use of a single-cell typing approach to determine single source profiles would by definition eliminate the need to use statistical methods to deconvolute the admixed profile.

DNA Profiling

9. We developed a direct lysis protocol for DNA isolation from micro-dissected cells using a new direct lysis reagent, forensicGEM® (ZyGEM). This direct lysis kit utilizes a thermostable proteinase for DNA extraction in as little as 20 minutes. It is specifically formulated to be compatible with STR multiplex reactions. Using this lysis method, it was determined that 1-20 cells would provide a DNA profile (partial or full) in the majority of instances ($\geq$ 70%). As expected increasing amounts of full DNA profiles were obtained as the number of micro-dissected cells is increased. Only a small percentage of full profiles were recovered from single cells. However, the success rate for partial and/or full profiles from single cells was over 70%. This indicated that while full profiles may not always be recovered, potentially probative genetic information can still be obtained from single cells.

10. Rather than relying solely on the use of single cell profiles to aid in mixture de-convolution, we also explored the use of the quantitative computer-based TrueAllele® interpretation system to aid in mixture de-convolution (collaboration with Mark Perlin, CyberGenetics). We hypothesised that LCM-mediated isolation of multiple groups of cells ('binomial sampling') from the admixture would create separate cell sub-populations with differing constituent weight ratios. Furthermore we predicted that interpreting the resulting DNA profiling data by the TrueAllele® interpretation system would result in an efficient recovery of the constituent genotypes due to newfound abilities to compute a maximum LR from sub-samples with skewed weight ratios, and to jointly interpret all possible pairings of sub-samples using a joint likelihood function.

11. As a proof of concept, 10 separate cell samplings of size 20 recovered by LCM from each of two 1:1 buccal cell mixtures were DNA-STR profiled using a specifically developed LCN methodology, with the data analyzed by the TrueAllele® Casework system. In accordance with the binomial sampling hypothesis, the sub-samples exhibited weight ratios that were well dispersed from the 50% center value (50 $\pm$ 35 % at the 95% level). The maximum log(LR) information for a genotype inferred from a single 20 cell sample was 18.5 ban, with an average log(LR) information of 11.7 ban. Co-inferring genotypes using a joint likelihood function with two sub-samples essentially recovered the full genotype information. We demonstrated that a similar gain in genotype information can be obtained with standard (28-cycle) PCR conditions using the same joint interpretation methods.

12. We also evaluated an alternative approach to generating variation in relative cell ratios from two donors involving a standard DNA extraction of a mixed stain (1:1 ratio), followed by preparation of sub-samples by limiting dilution. These 'Poisson' random aliquot samples would be expected to contain varying proportions of the starting constituents, just as in binomial sampling. While we did not observe extremes in weight ratios among sub-samples, there was sufficient variation to provide an initial indication that such an approach might warrant further investigation (although significant optimization will be required in order to determine optimal inputs and number of replicates needed for analysis).

13. Additional optimization of the developed single cell methods was needed in order to improve profile recovery from single cells. We evaluated the use of next generation STR kits (e.g. Identifiler® Plus) as well as novel micro-volume PCR reactions. While the use of the micro-volume PCR reactions are currently still being fully evaluated and optimized, the initial studies demonstrate the initial success in profile recovery using the developed micro-volume PCR reactions. In addition to an increase in profile recovery, these reactions provide the additional benefit of reduced cost of analysis since significantly less amplification reagents are required per reaction.

14. The development of the micro-volume PCR reactions necessitated the development of alternative epithelial cell collection strategies as they were not compataible with the LCM instrument since a larger reaction volume is needed to cover the surface of the PCR tube cap for efficient cell recovery. We therefore utilized a micro-manipulation approach rather than LCM collections for the recovery of single or few epithelial cells. This approach provides an alternative for laboratories that may not possess, or have the ability to obtain, LCM instrumentation.

15. We also developed a one-step single tube micro-volume combined lysis and STR amplification reaction that reduces the need for additional sample manipulation or tube opening, thereby reducing the potential for contamination.

16. Despite the improvements made to single cell analysis using the developed enhanced DNA typing strategies, we acknowledge that such an approach could not be implemented into routine casework at this time. We are hopeful that this work represents a succesful proof of principle study that augurs well for its possible operational use after further optimization and validation.

RNA Profiling

17. The goal of this study was to not only individualize isolated epithelial cells through the recovery of the genetic profile (i.e. autosomal STR) of the cell donor but to also determine the body fluid or tissue source of origin of the cell. We proposed to accomplish cell type identification using RNA profiling. Cell type identification experiments were mainly focused on buccal, vaginal and skin epithelial cells. We performed extensive experiments in order to discover novel vaginal secretion and skin mRNA biomarkers for use in cell type identification experiments and identified numerous highly specific biomarkers for both skin and vaginal secretions.

18. While RNA profiling of individual cells proved to be rather challenging, we were able to obtain preliminary results demonstrating the ability to determine the cell type of origin of directly sampled epithelial cells. We have developed a preliminary single-step micro-volume cell lysis/reverse transcription reaction for RNA isolation and cDNA synthesis from isolated epithelial cells. We additionally developed simple gene-specific pre-amplification reactions in order to improve the sensitivity of subsequent detection assays. We also evaluated the use of an alternative RNA detection method that we developed involving high-resolution melt (HRM) analysis. This approach utilizes differing melt temperatures of the mRNA biomarker amplification products, each with their own unique melt temperature. This method is more rapid and inexpensive than traditional CE detection (~2 hours to complete analysis after reverse transcription and involves the use of unlabeled primers for detection). Using a combination of the developed methods, we were able to successfully detect body-fluid specific biomarkers from as few as 10 isolated epithelial cells.

19. We began an evaluation of several approaches to combined DNA/RNA profiling strategies including the use of RNA FISH (fluorescence in situ hybridization) probes and commercially available DNA/RNA co-isolation kits. Using the RNA FISH approach, multiple contiguous fluorescently labeled probes that bind along a target mRNA sequence are used in order to produce sufficient signal (without PCR amplification) to be detected using fluorescence microscopy. The use of this approach would ideally permit an identification of cell type prior to collection, and thereby allow us to determine which body fluid cells were present and collect cells from each body fluid donor separately. The collected cells could then be used solely for DNA analysis since the cell type would have been determined visually prior to collection thus permitting a direct association between the DNA profile and its the cellular origin.

20. While neither of these single cell RNA profiling approaches have been fully developed, these methods represent a suitable avenue for future development of sensitive and robust DNA/RNA co-profiling strategies for single or few isolated epithelial cells (the ultimate goal of the current work).

Summary of Findings

21. In summary, in this study we demonstrate the ability to isolate single and few epithelial cells isolated from admixed samples using either laser capture micro-dissection or micro-manipulation and recover genetic profiles from these samples using enhanced typing strategies such as micro-volume PCR reactions and single-tube one-step lysis and amplification reactions. The efficient ('smart') recovery and STR genotyping of individual cells by LCM described in detail in this report is achieved by (i) lysing the cells using a thermostable protease, (ii) performing the lysis and amplification in a single tube and (iii) using an increased PCR cycle number (34 cycles). The latter is strictly not essential since 28 cycles also yields comprehensive genotype information with TrueAllele®. This LCM approach is not a high throughput method, but it could be used 'as is' for lab-determined high priority mixture cases with a heightened requirement to maximize the genotype information recovered.

22. Using a combination of the developed methods, we were able to successfully detect body-fluid specific biomarkers from as few as 10 isolated epithelial cells permitting a determination of the tissue source of origin of isolated epithelial cells. We also began an evaluation of several approaches to combined DNA/RNA profiling strategies.

23. While some of the developed approaches require additional optimization and validation, they successfully demonstrate our ability to 'get more (genetic information) from less' (a fundamental aim of the current work).

Implications for Policy and Practice

24. Interpretation of admixed STR profiles from multiple donors present a complex challenge for forensic laboratories. Few studies have been carried out to aid the scientific community in the physical de-convolution of non-distinguishable cell type mixtures. The proposed work sought to help the scientific community in this regard by developing a novel strategy to isolate individual epithelial or leukocyte cells from admixed samples using laser capture micro-dissection. This will permit the both the positive identification of the cell type by mRNA expression profiling and the individual by STR profiling. Statistical analysis will be simplified and strengthened by requiring only single source profile estimates instead of likelihood ratio or combined probability of exclusion calculations. Such an approach, should the testing laboratory choose to employ it in a particular case, is expected to yield more probative information from evidentiary items from cases that contain biological material from more than one person.

Implications for Future Research

25. Additional studies need to be carried out to determine the efficacy of alternative binomial sampling strategies to optimize information recovery while providing flexibility when processing cases that may have different numbers of available cells for sub-sampling. Although this proof of concept work used $n = 20$, other sample sizes (e.g. 5-15, >20, including a range of differing sub-sample sizes from the same stain) as well as different numbers of sub-samples (e.g. 5-9) should be tested using the same LCM recovery and typing strategy. Optimal sampling parameters (m, n) should be determined for DNA mixtures incorporating weight ratios different from the (albeit worse-case scenario) 1:1 studied here, as well as those with more than 2 donors. Lastly, the

number of jointly examined sub-samples necessary to compute an optimal joint likelihood function should be empirically determined for different values of the *m* and *n* parameters.

26. We believe that the preliminary work described here has implications for forensic practice that go beyond the specialized use of binomial sampling/LCM for mixture genotype inference. The binomial strategy relies on the deliberate generation of variation in the proportion of the two different cell constituents comprising the mixture. The same effect can be obtained and exploited in routine casework much more easily than with LCM. There are two different ways this could be done in the routine setting: 1) multiple micro-geographical samplings of such stains could result in sub-samples with different component proportions; 2) standard DNA extraction of a mixed stain, followed by preparation of sub-samples by limiting dilution (an very preliminary evaluation provided in this report). These 'Poisson' random aliquot samples would be expected to contain varying proportions of the starting constituents, just as in binomial sampling.

27. Further optimization and validation of the RNA profiling techniques will be needed in order to improve sensitivity. However, we are hopeful that with continued development of our novel HRM RNA body fluid identification assays will result in a sensitive, robust and cost-effective for cell type identification of single or few epithelial cells to aid in the individualization of individual cells in admixed samples.

Dissemination of Research Findings

28. This work has resulted in 5 publications (4 published and 1 additional manuscript in preparation) as well as 3 related publications. Ten presentations at scientific conferences (including international conferences) have been given in order to present the findings of this study to the scientific community.

I. INTRODUCTION

A. Statement of the Problem

It is now a matter of routine for the forensic scientist to obtain the genetic profile of an individual from DNA recovered from a biological stain deposited at the crime scene. However, during the commission of a violent crime two or more people are typically present and may deposit biological material at the crime scene or onto an individual or individuals, thus resulting in a mixture of genetic material. Moreover the shed biological material may be deposited on a substrate that already has previously deposited DNA present, thus contaminating the crime scene deposits with other 'legitimate' sources. This is especially so with so-called 'touch DNA' evidence. Interpretation of mixture profiles without first separating out the component cell types prior to analysis can be complex and time consuming [1-6]. In some cases it is not possible to deduce the component profiles thus significantly reducing the probative nature of the evidence. Often, mixtures can be easily identified due to morphological differences between cell types, such as in sexual assault evidence that is typically comprised of sperm and vaginal epithelial cells. Several strategies have been developed to aid in the interpretation of this type of mixture. The most commonly used strategy involves the use of a differential extraction which allows for a physical separation of sperm and non-sperm cells prior to analysis [7]. However, this procedure is limited to mixtures involving sperm cells. No such strategy or methodology has been developed thus far to physically separate the contributors in a non-easily-distinguishable cell type mixture. Therefore, this study investigated a novel method for the de-convolution of epithelial cell mixtures through cell type identification and genetic profiling of micro-dissected cells. Through the current work, we sought to determine: 1) whether DNA and RNA can be co-isolated from individual cells micro-dissected from epithelial and/or leukocyte cell mixtures, 2) whether the RNA isolated from micro-dissected epithelial and/or leukocyte cells could be used to determine the source of the collected cell (blood, skin, saliva, or vaginal secretions) through RNA profiling techniques, and 3) if autosomal STR profiles could be obtained from the micro-dissected cells which could then be attributed to a particular individual.

B. Literature Review

DNA mixtures derived from two or more contributors are frequently identified in biological evidence recovered from the crime scene, participating individuals or weapons. Mixed stains may comprise blood, saliva, skin cells, semen and vaginal secretions or combinations thereof. Moreover, with the high sensitivity of PCR based analysis methods, such as STR analysis, the appearance of low level minor contributor profiles are frequently encountered. In some instances, such as in sexual assault evidence, morphological differences between cell types allows for an identification of the presence of a mixture and also allows for physical separation of sperm and non-sperm cells prior to analysis [8-10]. More recently, the use of laser capture micro-dissection (LCM) has been evaluated to isolate sperm cells from non-sperm cells [11-13]. Laser capture micro-dissection can be accomplished using a variety of platforms, which can be classified as direct versus non-direct cell contact. In the direct contact method a thermoplastic membrane polymer is melted by the laser over the cell(s) of interest and the captured cells

removed by adhesion. This approach is less desirable as the direct contact of the laser could damage the genetic material in the targeted cells. The non-direct contact LCM method requires the dispersal of the cell suspension onto a proprietary membrane and the targeted cells are then circumscribed by, and cut out, by the laser. Single sperm cells can be visually identified and then isolated from a sample that contains both sperm and epithelial cells.

While the differential extraction and LCM strategies above provide the ability to separate components in admixed samples, they are limited to mixtures involving sperm cells. No such strategy or methodology has been developed thus far to physically separate the contributors in a non-easily-distinguishable cell type mixture, such as epithelial cell mixtures. Genotype attributions in such mixtures are often complex and time consuming, particularly if the contributors are present in approximately equal portions or the number of donors exceeds two [5,14,15]. The identification of a mixture typically relies on a visual examination of the number and relative fluorescence intensity of alleles present at each locus, with the presence of three or more alleles and/or significant allelic imbalance indicating the presence of a mixture. Individual contributor profiles can sometimes be discerned through an assessment of peak height ratios if differences in signal intensities identify major and minor contributors [14]. However this type of basic interpretation can be confounded in forensic casework without prior knowledge of the number of contributors, the ratio in which the contributors are present within the mixture and the extent of degradation. Low copy number samples comprising part of the admixture present additional challenges, with significant peak imbalances, increased stutter peaks and allelic-drop in and drop-out frequently observed. An assessment of all possible genotypes at each locus can be made, and with the aid of reference samples inclusion/exclusion conclusions can often be drawn. However, inclusion conclusions need to be bolstered by a statistical evaluation of the significance of the inclusion. Numerous statistical approaches to the interpretation of mixtures can be applied such as likelihood ratios (LR) or the combined probability of exclusion (Pr(*Ex*), or random man not excluded, RMNE) [1,5,6,16-19]. Despite attempts to provide consolidated guidelines for mixture interpretation, there is currently no single method that accommodates all the challenges of forensic casework samples and that always permits simple statistical calculations to be made [1,20,21]. In some complex mixture cases involving low copy number samples it may not be prudent to proffer a statistical evaluation of the evidence, especially in the absence of known samples from potential contributors. Indeed some scientists go as far as to express the belief that a statistical evaluation of admixture evidence is only appropriate when one has a potential contributor's genotype to compare with. Even when statistical evaluation of a mixed sample is possible the probative value of the evidence is often significantly reduced compared to a single source sample due to the statistical uncertainties of attributing specific genotypes to individual contributors.

The inability to clearly resolve individual contributors in admixed epithelial cell mixtures and the need for the use of complex statistical interpretations necessitates further research aimed at developing new methodologies to aid operational crime laboratories in the analysis of mixture samples. Laser capture micro-dissection allows for the isolation and collection of individual epithelial cells. If individual epithelial cells could be collected and analyzed, all contributors in an admixed sample could be identified with single source genetic profiles, thus eliminating the need for the rigorous interpretation guidelines and strategies described above. However, no studies have demonstrated the ability to consistently recover full autosomal STR profiles from single cell samples, which is the routine method of DNA analysis in forensic casework.

This study seeks to provide a novel method for the de-convolution of epithelial cell mixtures through cell type identification and genetic profiling of micro-dissected cells. This study firstly seeks to determine whether DNA and RNA can be co-isolated from individual cells micro-dissected from epithelial and/or leukocyte cell mixtures. Secondly the current work seeks to utilize the RNA isolated from micro-dissected epithelial and/or leukocyte cells to determine the source of the collected cell (blood, skin, saliva, or vaginal secretions) through RNA profiling techniques. Lastly, attempts will be made to further individualize the micro-dissected cells by obtaining DNA profiles which can then be attributed to a particular individual. The latter will be accomplished using various methods including an enhanced DNA typing approach and probabilistic computer interpretation of the resulting profiles.

C. Statement of Hypothesis or Rationale for the Research

During the commission of a violent crime two or more people are typically present and may deposit biological material at the crime scene or onto an individual or individuals, thus resulting in a mixture of genetic material. In some cases it is not possible to deduce the component profiles thus significantly reducing the probative nature of the evidence. Often, mixtures can be easily identified due to morphological differences between cell types, such as in sexual assault evidence. However, this procedure is limited to mixtures involving sperm cells. No such strategy or methodology has been developed thus far to physically separate the contributors in a non-easily-distinguishable cell type mixture. The inability to clearly resolve individual contributors in admixed epithelial cell mixtures and the need for the use of complex statistical interpretations necessitates further research aimed at developing new methodologies to aid operational crime laboratories in the analysis of mixture samples. Therefore, this study sought to determine whether RNA and DNA profiling techniques could be applied to single micro-dissected cells isolated from body fluid mixtures in order to identify cell type and recover a single source genetic profile of the donor thus providing a novel strategy for the de-convolution of epithelial cell admixtures.

D. Project Aims

Aim 1. Co-isolation of DNA and RNA from Micro-Dissected Cells of Sufficient Quality and Quantity for Genetic Profiling.

Aim 2. Cell Type Identification of Captured Cells

Aim 2A. Identification of RNAs specific for vaginal, oral, and skin epithelial cells and leukocytes (blood).

Aim 2B. Detection of cell-type specific RNAs in individual micro-dissected cells.

*Aim 2C. Development of RNA pre-amplification methods to improve sensitivity of cell-type identification methods***.**

Aim 3. DNA Profiling of Captured Cells.

*Aim 3A. STR profiling using standard and modified amplification conditions***.**

Aim 3B. Improving the sensitivity of STR profile recovery.
Aim 3B (i). Increased cycle number and alternative DNA polymerases**.**
Aim 3B (ii). Whole genome amplification (WGA)**.**
Aim 3B (iii). Sample Pooling.

II. METHODS

DNA Profiling

Sample Collection

Buccal samples were collected from volunteers using procedures approved by the University of Central Florida's institutional review board. Informed written consent was obtained from each donor. Buccal samples were collected from donors using sterile swabs by swabbing the inside of each donor's mouth. Swabs were dried overnight at room temperature and stored at 4°C (short term storage) or -20°C (long term storage) until needed. Each swab was agitated in sterile water. Samples were centrifuged at 14,000 rpm (16,000g) to obtain an epithelial cell pellet. The cell pellet was re-suspended in ~300-500 µl sterile water.

Sample Collection – Micro-Manipulation

Buccal samples were collected from volunteers using procedures approved by the University of Central Florida's institutional review board. Informed written consent was obtained from each donor. Buccal samples were collected from donors using sterile swabs by swabbing the inside of each donor's mouth. Swabs were used immediately or dried overnight and then stored at 4°C (short term storage) or -20°C (long term storage) until needed. For swabs used immediately, the moist swab was smeared onto Gel-Film™ (x0, x4, x8 retention level, Gel-Pak®, Hayward, CA) to transfer cells onto Gel-Film™ surface. Alternatively, a cell suspension was prepared as described above and allowed to dry on the Gel-Film™ surface (~6 hours – overnight). If staining was required, the cells were stained with tryphan blue (0.4% solution, Sigma-Aldrich, St. Louis, MO) for 1 min and gently rinsed with nuclease free water. The samples were then air dried before used.

Samples/LMD Slide Preparation

Aliquots of cell suspensions prepared as described above were placed onto a Leica 0.2 µM PEN slide (Micro Optics, Davie, FL) and heat fixed at ~65°C until dry. Mixtures were prepared by combining appropriate numbers of cells from each single donor cell suspensions in order to obtain the desired ratio. Cell counts were determined with the use of a hemacytometer. Admixed cell suspensions (~300-500 µl) were placed onto a Leica 0.2 µM PEN slide and heat fixed at ~65°C until dry. Slides were stored at 4°C in covered slide boxes until needed.

Laser Capture Micro-dissection (LCM)

Non-contact laser micro-dissection was performed using the Leica AS LMD instrument (Leica Microsystems). Cells were identified and collected using 400X magnification (40x objective) with the following laser settings: power – 34, speed – 8, specimen balance – 45 (slight modifications to laser settings were made between samples if needed). A predetermined number of cells were collected by laser ablation around the cell and subsequent gravitational capture of the released cell into the caps of 0.2 ml PCR flat-cap tubes (Phenix Research, Candler, NC) containing 10 µl of lysis buffer (see below). After collection, the tubes were capped and centrifuged briefly. Direct lysis was performed immediately after collection (see below). A "0 cell" sample (collection of an empty section of the slide) was used as a negative control for all LCM collections.

Micro-Manipulation

A tungsten needle is used to collect a "ball" of 3M water soluble adhesive at the end of the needle. The water soluble adhesive is then touched to the surface of the Gel-Film™ in order to collect the desired number of cells (see Gel-Film™ sample preparation above). All collections were performed using a Leica M205C stereomicroscope. The needle (with adhesive and cells attached) is then placed into solution (lysis or amplification reaction) inside a 0.2ml PCR tube. The dissolution of the adhesive and resulting cell transfer is viewed under the microcope. A "0 cell" sample consisting of adhesive touched to an empty portion of the Gel-Film™ surface was used as a negative control for all micro-manipulation collections.

DNA Isolation (Direct Lysis) - Standard

DNA isolation was performed using the forensicGEM® Saliva kit (ZyGEM™, VWR International, Suwanee, GA) according to the manufacturer's recommended conditions, with minor modifications. The reaction volume was decreased to 10 µl (2.1X buffer – blue (final concentration), 10% forensicGEM® reagent, sterile water). All lysis reactions consisted of incubation of samples at 75ºC for 15 minutes and 95ºC for 5 minutes and were performed using an Applied Biosystems 9700 GeneAmp(R) PCR system thermal cyler (Applied Biosystems, Foster City, CA). Modifications to the standard direct lysis protocol will be described in the relevant Statement of Results section.

PCR Amplification

Amplification of the cell lysates was performed using various amplification kits: AmpFlSTR® Profiler Plus®, AmpFlSTR® Identifiler, AmpFlSTR® Identifiler Plus, AmpFlSTR® Yfiler (Applied Biosystems/Life Technologies, Foster City, CA) and PowerPlex® 16 HS (Promega, Madison, WI) according to manufacturer's recommended conditions. Amplifications were performed using low copy number (LCN) analysis (i.e. 34-36 cycles) or standard (i.e 28 – 30 cycles, depending on the kit) conditions. All amplifications were performed using an Applied Biosystems 9700 GeneAmp® PCR system thermal cycler (Applied Biosystems/Life Technologies). Amplification controls were included with each amplification (positive control – 9947a (provided in kit), negative control – amplification blank (water)). Reduced volume amplifications (1, 2, 2.5, 3, 3.5, 4, 4.5, 5, 10 and 12.5µl) were also performed using appropriately scaled reaction volumes. Any modifications to these reaction volumes or compositions will be described in relevant results sections.

PCR Production Detection

An aliquot (1 µl) of the amplified product was added to 9.7 µl of deionized formamide (Applied Biosystems/Life Technologies) and 0.3 µl GeneScan™ 500 LIZ® size standard (Applied Biosystems/Life technologies). Samples were injected onto an ABI Prism 3130 Genetic Analyzer using Module G5 (16 sec injection, 15kV, 60ºC) and analyzed with GeneMapper analysis software v 4.0 (Applied Biosystems/LifeTechnologies). A peak detection threshold of 10 RFUs was used for allele detection.

Quantitative Methods

Quantitative Likelihood

STR analysis produces quantitative electrophoretogram (EPG) data that reflects the underlying amount of DNA template [22]. The post-PCR peak height at a given base pair length

is roughly proportional to the number of pre-PCR allele copies [23]. Amplification artifacts, such as stutter or imbalance, also affect the peak height data pattern [24].

The likelihood of a hypothesis given data and a specific model is proportional to the conditional probability Pr{data | hypothesis}, where the proportionality constant is arbitrary [25]. A likelihood function (LF) explains the observed data through a set of model parameters [25]. Parameter values that better explain the data are assigned a higher likelihood. Parameters that express data uncertainty are represented as random variables and reported as probability distributions [26].

We can write the likelihood of an unknown genotype X for a locus EPG peak height mixture pattern y as

$$\Pr\{y|X_1 = x_1, X_2 = x_2,\ldots\}$$

where $x_1 = [a_1, b_1]$ is the allele pair value of the first contributor's genotype X_1, $x_2 = [a_2, b_2]$ is the allele pair of the second contributor's genotype X_2, and "…" refers to other relevant random variables, such as mixture weight, DNA quantity, peak uncertainty, PCR stutter, and so on. The precise form of the LF depends on the mathematical probability model being used [27,28]; here we follow Perlin et al [28].

Joint Likelihood

Most problems in science generate more than one data point, each a sample from a set of possible observations [29]. A complete solution that uses all the data tries to simultaneously explain all of the observed data. The joint likelihood function (JLF) for two mixture experiments y_1 and y_2 is then

$$\Pr\{y_1, y_2|X_1 = x_1, X_2 = x_2,\ldots\}$$

When the data are sampled independently, the JLF factors into a product of likelihoods as

$$\Pr\{y_1|X_1 = x_1, X_2 = x_2,\ldots\} \cdot \Pr\{y_2|X_1 = x_1, X_2 = x_2,\ldots\}$$

In exploring genotype and other parameters, those values that better jointly explain both data patterns will give a higher probability product, and thus have a greater JLF.

Genotype Inference

Before we see any sampled STR data, the prior probability of a genotype allele pair value $x = [a, b]$ is based on allele frequencies through the product rule as

$$\Pr\{X = x\} = \begin{cases} p_a^2, & a = b \\ 2p_a p_b, & a \neq b \end{cases}$$

where p_a is the probability of seeing allele a in a population. The LF updates this prior to a posterior genotype probability that explains the observed (joint, say, with two experiments) data y_1 and y_2 as

$$\Pr\{X = x | y_1, y_2, \ldots\} \propto \Pr\{X = x\} \times \Pr\{y_1 | X = x, \ldots\} \times \Pr\{y_2 | X = x, \ldots\}$$

using Bayes theorem [30].

When there are many variables, as found in an expressive hierarchical probability model [31], statistical Markov chain Monte Carlo (MCMC) computer search [32,33] is used to explore

the parameter values. The output of this computation is a posterior probability distribution for each model variable, including the genotype of each contributor at every locus, and the mixture weight of each contributor. Note that the posterior genotype here is a probability distribution over possible allele pair outcomes.

Identification Information

The purpose of forensic DNA science is to preserve the identification information present in biological evidence, and express this information understandably in court. The standard measure of information and science in law is the likelihood ratio (LR) [34,35]. The LR gives the evidentiary support for a probative hypothesis concerning the presence of the suspect's DNA. The LR also removes prejudicial preconceptions about guilt or innocence.

There are many mathematically equivalent ways to formulate the LR [36]. One favored form in forensic science is the balance of (conditional probability) likelihoods [34]

$$LR = \frac{\Pr\{evidence | suspect\ is\ a\ contributor\}}{\Pr\{evidence | suspect\ is\ not\ a\ contributor\}}$$

Equivalently [36], in terms of an inferred evidence genotype compared with a known suspect genotype, relative to a reference population, the LR is [37]

$$LR = \frac{\Pr\{match\ between\ suspect\ and\ evidence\ genotypes\}}{\Pr\{coincidental\ genotype\ match\}}$$

The LR can thus be understandably stated as "a match between the suspect and the evidence is (some number) times more probable than coincidence." We shall use $\log_{10}$(LR) to measure how well a genotype inference has preserved identification information. The base ten logarithm of the LR is the standard "ban" information unit [38] developed by Alan Turing during the World War II Enigma code cracking project [39]. To account for co-ancestry [40], a theta value of 1% is applied.

TrueAllele® Casework

Cybergenetics TrueAllele® Casework is a computer system for objectively inferring probabilistic genotypes from DNA evidence. TrueAllele® examines peak height data patterns to infer the constituent genotypes in DNA mixtures that contain multiple contributors (e.g., 2, 3, 4, 5 or more), and models many of the STR process variables (e.g., PCR stutter, relative amplification, mixture weight ω, and DNA degradation). Stochastic effects in parameter values are modeled through hierarchical variance parameters [41], including peak height variation, baseline dectector variation, and mixure variation between loci [42]. Following genotype inference, TrueAllele® can also be used to compare genotypes and form a LR match statistic.

STR data files are first uploaded to the TrueAllele® database. Interpretation questions about the data are then visually asked by a workstation operator. A central TrueAllele® server computer performs a MCMC statistical search to solve these problems, and records the inferred (probabilistic) genotypes onto the TrueAllele® database. Visual review of these genotype answers and LR match statistics is performed on an operator workstation.

In this work we primarily examined single experiment LFs and two experiment JLFs. For one data set, joint interpretation examined 10 experiments simultaneously. The interpretation requests were prepared in the operator VUIer™ interface following standard procedures, and processed in TrueAllele® Casework version 3.25.3915.1. To solve each TrueAllele® request, the computer search ran 25,000 MCMC burn in cycles to first move the random variables into their

posterior distribution, conditioned on the data. The computer then ran for an additional 25,000 readout cycles to sample from the posterior distribution.

RNA Profiling

Preparation of Body Fluid Stains

Body fluids were collected from volunteers using procedures approved by the University of Central Florida's Institutional Review Board as well as the Institute of Legal Medicine, Forensic Genetics, Zurich, Switzerland. Informed written consent was obtained from each donor. Blood samples were collected by venipuncture into additive-free vacutainers and 50 µl aliquots were placed onto cotton cloth and dried at room temperature. Freshly ejaculated semen was provided in sealed plastic tubes and stored frozen until being dried onto sterile cotton swabs. Buccal samples ('saliva') were collected from donors using sterile swabs by swabbing the inside of the donor's mouth. Semen-free vaginal secretions and menstrual blood were collected using sterile cotton swabs. Penile swabs, taken before and after sexual intercourse, were collected using a sterile cotton swab (moistened with sterile water). Surface swabs of a male finger and an inserted foreign object were collected using sterile cotton swabs (moistened with sterile water) after insertion into the vagina of a female donor. All samples were stored at -20°C until needed. A 50 µl stain or a single cotton swab was used for RNA isolation.

Human skin total RNA was obtained from commercial sources: Stratagene/Agilent Technologies (Basel, Switzerland or Santa Clara, CA), Biochain® (Hayward, CA), Zenbio (Research Triangle Park, NC), Zyagen (San Diego, CA), and AMS Biotechnology (Bioggio-Lugano, Switzerland). Human tissue total RNA (adipose, bladder, brain, cervix, colon, esophagus, heart, kidney, liver, lung, ovary, placenta, prostate, skeletal muscle, small intestine, spleen, testes, thymus, thyroid, and trachea) was also obtained from commercial sources (FirstChoice® Human Total RNA Survey Panel, Ambion by Life Technologies or Rotkreuz, Switzerland). Skin samples tested included scraped or swabbed skin from friction ridge skin and hairy skin, and mock casework samples (swabs from various surfaces and objects (e.g. door handle, keyboard, telephone receiver)). Swabs of these samples were collected by swabbing the skin or object surface with a pre-moistened (sterile Millipore water) sterile swab. All body fluid samples were dried prior to storage. All body fluid samples were stored at -20°C (UCF) or at room temperature protected from light (Zurich) until needed. Total RNA samples (skin and tissues) were stored at -80°C. A 10 - 50 µl stain (cotton cloth) or a single cotton swab was used for RNA isolation.

RNA Isolation

Total RNA was extracted from blood, semen, saliva, vaginal secretions and menstrual blood with guanidine isothiocyanate-phenol:chloroform (Ambion by Life Technologies) and precipitated with isopropanol [43]. Briefly, 500 µl of pre-heated (56°C for 10 minutes) denaturing solution (4M guanidine isothiocyanate, 0.02M sodium citrate, 0.5% sarkosyl, 0.1M β-mercaptoethanol) was added to a 1.5mL Safe Lock tube extraction tube (Eppendorf, Westbury, NY) containing the stain or swab. The samples were incubated at 56°C for 30 minutes. The swab or stain pieces were then placed into a DNA IQ™ spin basket (Promega), re-inserted back into the original extraction tube, and centrifuged at 14,000 rpm (16,000 x g) for 5 minutes. After centrifugation, the basket with swab/stain pieces was discarded. To each extract the following was added: 50 µl 2 M sodium acetate and 600 µl acid phenol:chloroform (5:1), pH 4.5 (Ambion by Life Technologies). The samples were placed at 4°C for 30 minutes to separate the layers and

then centrifuged for 20 minutes at 14,000 rpm (16,000 x g). The RNA-containing top aqueous layer was transferred to a new 1.5 ml microcentrifuge tube, to which 2 µl of GlycoBlue™ glycogen carrier (Ambion by Life Technologies) and 500 µl of isopropanol were added. RNA was precipitated for 1 hour at -20ºC. The extracts were then centrifuged at 14,000 rpm (16,000 x g). The supernatant was removed and the pellet was washed with 900 µl of 75% ethanol/25% DEPC-treated water. Following a centrifugation for 10 minutes at 14,000 rpm (16,000 x g), the supernatant was removed and the pellet dried using vacuum centrifugation (56ºC) for 3 minutes. Twenty microliters of pre-heated (60ºC for 5 minutes) nuclease free water (Ambion by Life Technologies) was added to each sample followed by an incubation at 60ºC for 10 minutes. Samples were used immediately or stored at -20ºC until needed. An extraction blank (reagents only, no sample) was included with every extraction.

Alternatively, total RNA was extracted from blood, semen, saliva, vaginal secretions and menstrual blood using other extraction types including a commercially available extraction kits (RNeasy® Mini kit and AllPrep DNA/RNA Mini kit (QIAGEN®, Hombrechtikon, Switzerland)) as previously described [44]. Additionally, the use of direct lysis methods will be described in the results section.

RNA Isolation - Pinpoint™ Slide RNA Isolation System II

Skin cells were collected from fingerprints on glass slides and from touched surfaces and objects (e.g. door handle, pen, cup). Some objects were not touched for several months before sampling. A thin layer of Pinpoint™ Solution was applied to the area of choice (1-2 cm^2), using a sterile pipette tip. The Pinpoint™ Slide RNA Isolation System II extraction (Zymo Research/ Lucerne Chem) was performed according to the manufacturer's protocol with the following adaptations: the paraffin removal steps were omitted. Two to three times the amount of RNA digestion buffer and proteinase K was used, to cover the whole sample with reagents and to reduce the viscosity of the Pinpoint Solution (prevents spin column plugging). RNA was eluted in 12 µl RNA Elution Buffer.

DNase I Digestion

DNase digestion was performed using the Turbo DNA-*free*™ kit (Applied Biosystems by Life Technologies) according to the manufacturer's protocol. Briefly, 1X Turbo DNase Buffer and 1 µl TURBO DNase was added to the 20 µl RNA extracts and incubated at 37ºC for 30 minutes. After incubation, 2.3 µl DNase inactivation reagent was added and samples were incubated for 5 minutes at room temperature. Following a centrifugation for 1.5 minutes at 10,000 x g, extracts were transferred to new 1.5mL tubes.

RNA Quantitation

RNA extracts were quantitated with Quant-iT™ RiboGreen® RNA Kit (Invitrogen by Life Technologies, Carlsbad, CA) as previously described [44-47]. Fluorescence was determined using a Synergy™ 2 Multi-Mode microplate reader (BioTek Instruments, Inc., Winooski, VT).

Whole Transcriptome Sequencing (RNA-Seq)

Total RNA was isolated from two vaginal swabs (26 yr old female; 30 yr old female) as described above. Two human skin total RNA samples were obtained from commercial sources (Biochain, Hayward, CA; Zyagen, San Diego, CA). Transcriptome sequencing (RNA-seq) was

performed by Ambry Genetics (Aliso Viejo, CA) using 54 bp paired-end sequencing on the Illumina Genome Analyzer IIx. All samples were sequenced in duplicate.

cDNA Synthesis

RNA samples were reversed transcribed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems by Life Technologies) according to the manufacturer's protocols. A reverse transcription (RT) negative reaction (containing total RNA and reaction buffer but no reverse transcriptase enzyme mix) was performed for each sample. A reverse transcription positive control (previously tested body fluid sample) was also included with each RT reaction.

Polymerase Chain Reaction and PCR Product Detection – Vaginal Secretions

Standard reaction–2 µl of the RT-reaction was amplified in a total reaction volume of 25 µl. The reaction mixture contained buffer, 1.6 - 2.4 µM primers (see below), 250 µM dNTPs, 1X PCR Buffer II (10mM Tris-HCl, pH 8.3, 50 mM KCl), 3.25mM $MgCl_2$, and 2 units AmpliTaq Gold® (Applied Biosystems by Life Technologies). Amplification blanks (reaction mix with water added in place of sample) and amplification positive controls (previously tested RT products) were included with each amplification.

Primers - PCR primer sequences were designed using Primer 3 v. 0.4 [48]. Primers were custom synthesized by Invitrogen and Applied Biosystems by Life Technologies. For capillary electrophoresis detection, the forward primer of each primer set was labeled with a fluorescent phosphoramidite dye (Applied Biosystems or Invitrogen by Life Technologies). The primer concentrations for initial candidate screening experiments were 1.6µM. The primer concentrations for the six candidates used in further testing were as follows: SFTA2 – 1.6µM; FUT6 – 1.6µM; DKK4 – 1.6µM; IL19 – 2.4µM; MYOZ1 – 1.6 µM; CYP2B7P1 – 1.6µM.

Cycling conditions – (1) 95ºC 11 min, (2) 35 cycles: 94ºC 20 sec, 60ºC 30 sec, 72ºC 40 sec; and (3) final extension at 72ºC for 45 min. The thermal cycler used was a GeneAmp® 9700 PCR System (Applied Biosystems).

Housekeeping Gene (B2M) – 2 µl of the RT-reaction was amplified in a total reaction volume of 25 µl. The reaction mixture contained buffer, 0.8 µM primers, 250 µM dNTPs, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 3.25 mM $MgCl_2$, and 1.25 units AmpliTaq Gold® (Applied Biosystems by Life Technologies). The cycling conditions for the B2M singleplex amplification were as follows: (1) 95ºC 11 min, (2) 35 cycles: 94ºC 20 sec, 58ºC 30 sec, 72ºC 40 sec; and (3) final extension at 72ºC for 45 min.

PCR Product Detection-Amplified fragments were detected with the ABI Prism 3130 Genetic Analyzer capillary electrophoresis system (Applied Biosystems by Life Technologies). A 1.0 µL aliquot of the amplified product was added to 9.7 µL of deionized formamide (Applied Biosystems by Life Technologies) and 0.3 µL of GeneScan™ 500 LIZ™ internal lane standard (Applied Biosystems by Life Technologies). Samples were injected using Module G (5-s injection, 1.2 kV injection voltage, 15 kV run voltage, 60°C) and analyzed with GeneMapper® Analysis Software v4.0 using Filter Set G (Applied Biosystems by Life Technologies). A peak detection thresholds of 50 RFUs was used.

Polymerase Chain Reaction and PCR Product Detection – Skin

Primers - PCR primer sequences were designed using Primer 3 v.0.4.0. Primers were custom synthesized by Invitrogen, Applied Biosystems, or Microsynth (St. Gallen, Switzerland).

The forward primer of each primer set was labeled with a fluorescent phosphoramidite dye (AB or Invitrogen by Life Technologies). The reverse primer sequence for LCE1C was modified during the course of this work to try to eliminate the presence of a second unidentified amplification product. For any studies that utilized the original reverse primer (5'-ACACTTTGGGGGACACTTTG-3'), the results are designated as LCE1C*. The current LCE1C primers are provided in Table 1. A housekeeping gene (UBC, ubiquitin C) was also used in the evaluation of human skin and touched objects (Section 3.4 and 3.5, Zurich laboratory) to ensure the presence of total RNA (UBC-F: 5'-GGGTCGCAGTTCTTGTTTGT-3', R: 5'-TCCAGCAAAGATCAGCCTCT-3').

Singleplex Reactions (UCF) – 2 µl of the RT-reaction was amplified in a total reaction volume of 25 µl. The reaction mixture contained 0.16 – 2.4 µM primers (LCE1C* - 1.6 µM, LCE1C – 0.32 µM, LCE1D – 2.4 µM, LCE2D – 1.2 µM, IL1F7 – 0.16 µM, CCL27 – 1.2 µM), 250 µM dNTPs, 1x PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 3.25 mM $MgCl_2$, and 2 units AmpliTaq Gold® (AB). All amplifications were performed on a GeneAmp® 9700 PCR System (AB): (1) 95ºC 11 min, (2) 35 cycles: 94ºC 20 sec, 58ºC 30 sec, 72ºC 40 sec; and (3) final extension at 72ºC for 45 min. Amplification blanks (reaction mix with water added in place of sample) and amplification positive controls (previously tested RT products) were included with each amplification.

Singleplex Reactions (Zurich) - Minor modifications were made to the standard singleplex reactions for Zurich-amplified samples. For amplification of LCE1C*, LCE1D, LCE2D, and CCL27, the following conditions were used: 2 µl of the RT-reaction was amplified in a total reaction volume of 25 µl. The reaction mixture contained 0.8 µM primers, 1 mM dNTPs (AB), 1xPCR Buffer I (AB), and 1.25 units AmpliTaq® Gold (AB). All amplifications were performed on a GeneAmp® 9700 PCR System (AB): (1) 95ºC 11 min, (2) 35 cycles: 94ºC 20 sec, 55ºC 30 sec, 72ºC 40 sec; and (3) final extension at 72ºC for 5 min. For amplification of IL1F7, the following conditions were used: 2 µl of the RT-reaction was amplified in a total reaction volume of 25 µl. The reaction mixture contained 0.8 µM primers, 1 mM dNTPs (AB), 1xPCR Buffer II (AB), 3.5 mM $MgCl_2$ (AB) and 1.75 units AmpliTaq Gold® (AB). All amplifications were performed on a GeneAmp® 9700 PCR System (AB): (1) 95ºC 11 min, (2) 35 cycles: 94ºC 20 sec, 58ºC 30 sec, 72ºC 40 sec; and (3) final extension at 72ºC for 60 min. Amplification blanks (reaction mix with water added in place of sample) and amplification positive controls (previously tested RT products) were included with each amplification.

PCR Product Detection - Capillary Electrophoresis - Amplified fragments were detected with the ABI Prism 3130 or 3130xl Genetic Analyzer capillary electrophoresis system (AB). Using the 3130 Genetic Analyzer, a 1.0 µL aliquot of the amplified product was added to 9.7 µL of Hi-Di™ formamide (AB) and 0.3 µL of GeneScan™ 500 LIZ® size standard (AB) using the following electrophoresis conditions: 16 sec injection time, 1.2 kV injection voltage, 15 kV run voltage, 60°C, 20 min run time, dye set G5. Using the 3130xl Genetic Analyzer, a 1.0 µl aliquot of the amplified product was added to 12.75 µl Hi-Di™ Formamide (AB) and 0.25 µl of GeneScan™ 500 LIZ® size standard (AB) using the following electrophoresis conditions: 18 sec injection time, 1.2 kV injection voltage, 15 kV run voltage, 60ºC, 20 min run time, dye set G5. All samples were analyzed with GeneMapper® Software v4.0 or ID-X v1.1.1 (peak detection thresholds of 25 RFUs for the 3130 and 100 RFUs for the 3130xl).

Combined Lysis – Reverse Transcription

A 5µl combined cell lysis/reverse transcription (RT) reaction was performed using the *RNA*GEM™ Tissue kit (ZyGEM™, VWR International, Suwanee, GA). The 5µl reaction

volume included: 3.55µl of lysis reaction mix (8.5 µl sterile water, 1.0 µl 10x silver buffer, 0.5µl *RNA*GEM™ enzyme); 1.45 µl of RT reaction mix (High Capacity RT kit (Life Technologies); 0.5µl 10x primer mix, 0.2µl 25x dNTP mix, 0.5µl 10x buffer, 0.25µl MultiScribe RT enzyme). Cells are collected into the lysis reaction mixture and then incubated at 75°C for 5 min in an Applied Biosystems 9700 GeneAmp(R) PCR system thermal cyler. After incubation, the RT reaction mix is then added and the following conditions used for reverse transcription: 25°C 10 min, 37°C 120 min, 85°C 5 min, 4°C hold. All RT reactions were performed using an Applied Biosystems 9700 GeneAmp(R) PCR system thermal cyler.

RNA Pre-Amplification

The 25µl reaction consisted of the following: 5µl RT (see above), 12.5µl 2x QIAGEN Multiplex PCR master mix (QIAGEN, Multiplex PCR kit), 2.5µl 1µM primer mix (singleplex reaction), and 5µl nuclease free water. The pre-amplification was performed using Applied Biosystems 9700 GeneAmp® PCR system thermal cyler: 95°C 5 min; 14 cycles of 95°C 15 sec, 60°C 2 min; 4oC hold. Amplification blanks (reaction mix with water added in place of sample) were included with each pre-amplification.

Polymerase Chain Reaction – mRNA Body Fluid Multiplex

The 25µl reaction consisted of the following: 12.5µl of 2x QIAGEN Multiplex PCR Master mix, 2.5µl Q-solution, 2.5µl of a proprietary primer mix, 5.5µl nuclease free water and 2µl of cDNA (standard input of 25ng). All amplifications were performed on ABI 9700 thermal cyclers using the following conditions: 95°C 15 min; 33 cycles of 94°C 30 sec, 55°C 90 sec (+0.2°C/cycle), 72°C 45 sec; 72°C 30 min; 4°C hold. The multiplex system used included 12 mRNA biomarkers (2 biomarkers per body fluid or tissue): blood – ALAS2, ANK1 [47,49]; semen - PRM2, TGM4 [46,50]; saliva – HTN3, STATH [45,46,50]; vaginal secretions – CYP2B7P1 [51], VAG2 (novel unpublished vaginal secretions biomarker so gene name not provided at this time); menstrual blood – MMP10, LEFTY2 [46,52]; skin – LCE1C, CCL27 [53,54]. Singleplex amplifications were also performed using this same protocol with modified primer mixes. Amplification blanks (reaction mix with water added in place of sample) and amplification positive controls (previously tested RT products) were included with each amplification.

High Resolution Melt Analysis (HRM)

The 25µl reaction consisted of the following: 5µl RNA pre-amplification product (see above), 12.5µl 2x QIAGEN HRM PCR Master Mix (QIAGEN, Type-It HRM PCR kit), 1.75µl 40µM primer mix (singleplex reaction), and 5.75µl nuclease free water. HRM detection was performed using the QIAGEN RotorGene Q real time PCR instrument: 95°C 5 min; 45 cycles - 95°C 10 sec, 57°C 40 sec, 72°C 20 sec; HRM 73-90°C. Amplification blanks (reaction mix with water added in place of sample) and amplification positive controls (previously tested RT products) were included with each amplification.

III. RESULTS

A. Statement of Results

1. DNA Profiling of Micro-Dissected Cells

1.1 STR Profiling of Single and Few Epithelial Cells

Analysis and interpretation of admixed DNA samples is often quite complex and therefore extremely challenging even more so if the number of donors exceeds two. In certain situations, the presence of a mixture can be anticipated and the cells separated prior to analysis, such as is the case with a 'differential extraction' of sexual assault evidence where sperm cells are present. However, cellular components of an epithelial cell mixture cannot be separated based on cell type or morphology. Interpretation of these mixtures is sometimes further complicated because of degraded DNA or artifacts that arise during the PCR-based analyses. A limited number of laboratory strategies are available to operational forensic laboratories in order to aid in the interpretation or resolution of these types of mixtures. The proposed work sought to provide a novel strategy for the de-convolution of body fluid mixtures containing non-distinguishable cell types (e.g. epithelial cells) through an isolation of individual cells using laser capture micro-dissection and recovery of genetic material (RNA and DNA). Using this type of approach, isolation and profile recovery from a single cell would by definition eliminate the presence of an admixed profile. Therefore, a significant focus of the current study was to develop robust and sensitive DNA profiling methods for the analysis of single epithelial cells. For ease of method development, the current work was limited to only two-donor mixtures. We acknowledge that single cell analysis may raise serious concerns for operational laboratories due to the extreme low template nature of these samples. Additionally, we do not anticipate such an approach being necessary for a majority of cases. Therefore, the goal in the current work was to begin to evaluate such a strategy to demonstrate the possible future use of such an approach in selected cases in which admixed profiles cannot be resolved by conventional means.

In previous and separate work, we had evaluated the use of laser capture micro-dissection for the analysis of sperm/vaginal cell mixtures (i.e. sexual assault evidence). We therefore had previously developed laser capture micro-dissection and direct lysis methods that could initially be evaluated for use with epithelial cell mixtures. Our previous method involved the use of a no-contact laser micro-dissection instrument, the Leica AS LMD system. Cells of interest are outlined on the screen thus providing a path for the laser to cut. The laser then cuts around the cell and the cell of interest falls by gravity into a collection tube below the stage. An optimized lysis buffer is added to the collection tube prior to collection. This lysis buffer (10 μl) consists of a commercially available buffer (Quick Extract, Epicenter) and DTT. The cells are lysed in the collection tube using a thermocycler (65ºC 8 minutes, 98ºC 2 min). Subsequent STR or whole genome amplification reactions are then performed in the same tube. The use of this lysis method resulted in the need for an enzyme substitution (Expand High Fidelity polymerase (Roche Applied Sciences, Indianapolis, IN) instead of Taq Gold) in subsequent STR amplifications.

To initially evaluate the above described method, the number of cells to be collected for each mixture sample and the number of amplification cycles used in STR amplifications needed

to be determined. For these initial experiments, 15-cell samples were prepared through the collection of a known number of cells from single source samples into the same collection tube (donor 1-donor 2 in the following ratios: 15-0, 14-1, 13-2, 12-3, 11-4, 10-5, 9-6, 8-7, 7-8, 6-9, 5-10, 4-11, 3-12, 2-13, 1-14 and 0-15). This type of artificial mixture was initially used in order to evaluate the various genotypes that would be recovered from the different combination of cell numbers from both donors. This allowed us to evaluate the minimum number of minor component cells that were necessary for detection. The analysis of these mixture samples was performed using Profiler Plus (32 amplification cycles) and the Expand High Fidelity polymerase. The number of cells collected from each donor was varied in order to obtain the full range of cell number combinations (15-0, 14-1, 13-2, 12-3, 11-4, 10-5, 9-6, 8-7, 7-8, 6-9, 5-10, 4-11, 3-12, 2-13, 1-14, 15-0). The profiles obtained from the evaluation of these 15-cell mixtures can be seen in Table 1. It can be seen from this table, that even with 2 or 3 cells collected from the minor contributor that only a single source profile of the major contributor can sometimes be observed. However, as few as one cell from the minor contributor can sometimes provide enough genetic material in order to allow for a minor contributor profile to be observed (14-1 cell mixture, Table 1). Typically, when four or more cells of one donor is present, a profile from that donor will be detected. The presence or absence of a contributor where 1-3 cells are present from the minor donor is unpredictable in this 15 cell system. With unknown mixture samples, the ratio of minor and major contributor cells will not be known *a priori*. However, these initial experiments were performed in order to evaluate the ability to detect mixture profiles using small numbers of cells. The resulting genotypes for the 15-cell samples (FAM channel only) are shown in Figure 1. Since the genotypes of both donors were known, the alleles belonging to each donor have been designated. At each individual locus, there are several combinations of alleles that can be observed depending on the genotypes of the donors. The presence of one, two, three or four alleles may be observed. If a single allele is present in all samples, this would indicate that both donors are homozygous and share the same allele. If two alleles are present, there could be three possible genotype scenarios. The first is that both donors are homozygous and do not share a common allele. The second is that both are heterozygous and both have the same alleles identical by state. The third is that one donor is homozygous and the other heterozygous with one allele shared with the first donor. An example of this third scenario can be seen in Figure 2. In this case, the female donor was a 9,12 and the male donor a 12,12. A mixture sample could therefore be indicated by an un-balanced 9,12 profile such as seen with the 11-4 and 6-9 mixtures. However, the assessment of the presence of a mixture cannot be solely be made based on the results of a single locus. As a result of the use of low template samples (15 cells) it is possible that an imbalanced 9,12 profile could be obtained from a single source profile due to stochastic effects of the amplification process. An evaluation of the rest of the STR profile should, however, provide sufficient evidence of the presence of a mixture.

Figure 3 shows the genotypes that could be observed at a three-allele locus in a two-donor mixture where one of the donors is homozygous and the other is heterozygous with no shared alleles between the two donors. A mixture for this type of locus would be indicated by the presence of the three alleles. Figure 4 also shows an example of a 3-allele locus. However, in this 3-allele profile, both of the donors are heterozygous with one shared allele between the two donors. A mixture would be indicated by the presence again of three alleles, however the peak height of the shared allele should be higher than that of the two non-shared alleles. This can be seen, for example, in the 11-4, 7-8 and 5-10 samples. Figure 5 shows an example of a four allele genotype obtained from a two donor mixture. The presence of a mixture would clearly be

indicated by the presence of the four alleles. However, as a result of the use of low template samples, occasionally an allele drop-out can be observed resulting in the presence of a three allele genotype. This makes profile interpretation more difficult. However, if multiple 15-cell samples were used in the analysis, it would be unlikely that such an allele drop-out would occur in each collection. Therefore if numerous samples provided a four-allele genotype and one provided a three-allele genotype, it could be concluded that a possible allele drop-put occurred in the one sample.

The results of the above described experiments suggest that mixture profiles of sufficient quality can be obtained from the 15-cell cell samples. The next step was to evaluate actual two-donor epithelial cell mixtures. As a result of binomial sampling variation, multiple 15-cell samples will contain different numbers of cells from the two donors. Based on the resulting profiles, it may be possible to de-convolute the mixtures. Additionally, it is also possible that a single source profile for one or both of the donors may also be obtained as the only or major component in one or more of the multiple 15-cell samples. This would allow for a de-convolution of the mixture by subtracting out the deduced single source profile. An example of an attempt to de-convolute a 1:1 two-donor mixture using multiple 15-cell samples (x 10) is provided in Table 2. As can be seen from the obtained genotypes, a mixture was obtained for most of the 15-cell replicates. However, one of the 15-cell samples provided a single source profile. This single source profile could then be used to de-convolute the mixture and to additionally obtain the genotype of the second donor.

The initial experiments described above involved the analysis of a 1:1 epithelial cell mixture. This same lysis and amplification strategy was used on ten replicate 15-cell samples collected from 1:1, 2:1, 3:1 and 4:1 female-male mixture samples. Samples were collected from two different mixture sets with different female donors used in each set. Each of the 10 replicate data sets was initially evaluated and genotyped by hand. Typically, for each of the data sets a single source profile was obtained for one of the ten replicate samples and this allowed for an easier determination of the second donor's profile. The number of observed single source profiles increased as the mixture ratio increased. For the 4:1 mixture set five of the ten replicates were single source profiles. Upon evaluation of the mixture data sets, a consensus profile was generated for both of the donors in the mixture. For each of the mixture ratios, a consensus genotype could be discerned for a majority of the loci (upon comparison to the known genotypes of both donors obtained through reference samples). A consensus genotype for a few of the loci could not be determined. For example, for the first female-male mixture the genotypes at the D5 locus were 11,11 and 10, 11, respectively. In the ten replicate samples for the 1:1 mixture from this donor couple, a single source profile of the female was obtained for one of the samples indicating her genotype as 11,11. Alleles 10 and 11 at D5 were observed for a majority of the remaining replicate samples. However, it could not be determined from the mixtures if the minor component genotype was a 10,10 or a 10,11 and therefore a consensus genotype for this locus could not be determined. A determination of the consensus genotypes at other loci were complicated by peak imbalance issues. However, this first attempt at mixture de-convolution using this type of approach was reasonably successful.

While we were performing these initial studies using our existing protocol, we were able to evaluate a new direct lysis reagent, forensicGEM® (ZyGEM). This kit utilizes a thermostable proteinase for DNA extraction in as little as 20 minutes. It is specifically formulated to be compatible with STR multiplex reactions. To initially evaluate this lysis kit, we collected varying numbers of cells from a single source sample and were able to obtain high quality STR profiles.

We adjusted the volumes of the kit components as it is designed for use with whole buccal swabs. We kept the amount of the forensicGEM reagent constant, but reduced the amount of buffer and water used. We were able to keep the lysis reaction volume to 10 µl so that we could use the entire lysate in a subsequent STR amplification. We decided to use the AmpFlSTR® Identifiler® STR amplification kit since it contained a larger number of loci that Profiler Plus®. We also attempted to use the standard AmpliTaq Gold® polymerase which had previously been incompatible with our lysis buffer. The obtained profiles were generated with the Identifiler® kit and Taq Gold polymerase. The quality and reproducibility of the obtained profiles appeared to be improved compared to those obtained with the Quick Extract protocol.

In order to further evaluate the use of the forensicGEM® lysis strategy, we collected 1, 2, 3, 4, 5, 10 and 20 cell samples (10 replicates of each cell number) from a total of 10 individuals. Figures 6 and 7 show the success rates for various cell numbers for any type of profile (both partial and/or full profiles) (Figure 6) and for full profiles only (Figure 7). A partial profile here includes any samples where one allele or more was recovered. Overall, it can be seen that 1-20 cells will provide a DNA profile (partial or full) in the majority of instances ($\geq$ 70%). As expected increasing amounts of full profiles are obtained as the number of micro-dissected cells is increased (Figure 7). Only a small percentage of full profiles were recovered from single cells. However, the success rate for partial and/or full profiles from single cells was over 70%. This indicates that while full profiles may not always be recovered, potentially probative genetic information can still be obtained from single cells. Additional factors such as allelic signal intensity, the number of allele and locus drop outs, the number of allele drop ins and stutter percentages were evaluated for this data set. The allelic signal intensity increased as the cell number increased which is expected. The rate of allele and locus drop out was higher for smaller cell numbers and was reduced as the cell number increased. The amount of allele drop in was variable amongst the cell numbers. The average percent stutter was generally less than 20% for all loci. A few loci exhibited an aberrant increased stutter greater than the height of the true allele. This occurred only for samples containing 1-4 cells and did not occur frequently.

The ultimate goal of mixture analysis is to precisely discern the constituent genotypes. The work described above was focused mainly on the development of amplification strategies for the analysis of single cells since that by definition would permit a determination of individiual constituent genotypes. However, at this stage in current work, further work was required in order to improve the success rate of profile recovery from single cells (see 'enhanced single cell analysis' section below). Therefore in addition to our single cell analysis method development we also evaluated possible alternative strategies for mixture de-convolution of epithelial cell mixtures.

1.2 DNA Mixture Genotyping by Probabilistic Computer Interpretation[55]

From our previous initial development work using the forensicGEM® lysis (described above), we had 20-cell replicates from 1:1 and 3:1 mixtures (male-female and female-female) (Identifiler® amplification kit). Rather than relying solely on the use of single cell profiles to aid in mixture de-convolution, we also explored the use of the quantitative computer-based TrueAllele® interpretation system to aid in mixture de-convolution (collaboration with Mark Perlin, CyberGenetics). We hypothesised that LCM-mediated isolation of multiple groups of cells ('binomial sampling') from the admixture would create separate cell sub-populations with

differing constituent weight ratios. Furthermore we predicted that interpreting the resulting DNA profiling data by the TrueAllele® interpretation system would result in an efficient recovery of the constituent genotypes due to newfound abilities to compute a maximum LR from sub-samples with skewed weight ratios, and to jointly interpret all possible pairings of sub-samples using a joint likelihood function.

As a proof of concept, 10 separate cell samplings of size 20 recovered by LCM from each of two 1:1 buccal cell mixtures were DNA-STR profiled using a specifically developed LCN methodology, with the data analyzed by the TrueAllele® Casework system. In accordance with the binomial sampling hypothesis, the sub-samples exhibited weight ratios that were well dispersed from the 50% center value (50 $\pm$ 35 % at the 95% level). The maximum log(LR) information for a genotype inferred from a single 20 cell sample was 18.5 ban, with an average log(LR) information of 11.7 ban. Co-inferring genotypes using a joint likelihood function with two sub-samples essentially recovered the full genotype information. We demonstrate that a similar gain in genotype information can be obtained with standard (28-cycle) PCR conditions using the same joint interpretation methods.

Concept of Binomial Sampling of Mixtures

The concept of binomial sampling is illustrated in Figure 8, in which microscopically indistinguishable epithelial cells from two genetically distinct individuals (victim-blue and perpetrator-white) are shown admixed at a ratio of 1:1. In real world crime situations, such two person mixtures are expected to be somewhat heterogeneous (in the micro-geospatial sense) due to cell clumping and inefficient mixing of the two cell types during the formation and deposition of the mixed stains. As shown in the left hand side of the diagram, standard mixture analysis involves sampling the whole stain (or a substantial portion thereof) for subsequent DNA extraction. This has the intended result of homogenizing the mixture such that a single weight ratio (ω, the relative proportion of either constituent) is obtained, which in the captioned instance is expected to be 0.5 due to equal quantities of cells from each individual and a relatively large sample size ($N \approx 100$).

In contrast, 'binomial sampling' refers to the situation whereby random single cells or cell clumps viewed under the microscope are recovered via laser capture (laser capture microdissection, LCM) and placed into a single reaction vessel to deliberately concoct a set of sub-samples with differing weight ratios. Thus, depending upon the precise *a priori* admixture ratios created during the original deposition of the stain, multiple ten-cell samplings (e.g. 10 samples of ten cells) will contain differing numbers of cells from each contributor. For example in a two donor mixture comprising equal numbers of cells from each donor (50:50, the 'worse case' scenario for mixture de-convolution analysis) a ten cell sample can be thought of roughly as representing multiple Binomial trials resulting in a binomial distribution (where $N = 10$ and $p = 0.50$).

The probability distribution of the sub-sample weight ratios is expected to approximately mimic the relevant binomial distribution *Binom*(10, 0.5). Since the cells are removed without replacement, the hypergeometric distribution is probably a more apt description of the physical process but, heuristically, we still refer to the sampling as binomial and, in the case where the sampling takes place in the background of hundreds or thousands of cells, is well approximated by the binomial.

The concept is to obtain sub-samples that possess measurably different weight ratios that should permit an increase in the genotype information content of a mixture compared to the

standard single sampling approach. Manual genotyping of a mixture may be efficient at ω = 2:1 or higher (but not at 1:1). Thus if a ten cell sub-sample from the 1:1 admixture actually comprises a 3:7 ratio or less (2:8, 1:9, 0:10) or its converse (7:3, 8:2, 9:1, 10:0), expected to arise in ~50% of the sub-samples based upon the binomial distribution, manual mixture genotyping should be facilitated. Here, instead of manual methods, we employ quantitative computer interpretation, which can more thoroughly explore all the explanatory variables when inferring constituent genotypes with sub-samples that possess minor differences in ω.

In summary, the conceptual hypothesis of this work is that multiple LCM-mediated samplings of groups of individual cells from a mixed stain will result in an information gain as evidenced by higher LRs compared to a single homogenized sampling of the whole stain. This would be due to (i) it being more likely that a beneficial (i.e. more informative) weight ratio will be obtained in at least one or more of the sub-samples (e.g. ω = 0.2 is better than 0.5) and (ii) that joint likelihood functions that combine data from the different samplings will further increase the information.

Genotyping Efficacy of Laser Captured Cells

The binomial sampling strategy subsumes the efficacy and accuracy of the genotyping procedure used subsequent to the recovery of cells by laser capture (LCM). In brief, the LCM–profiling method developed by us for single (or few) cell analysis includes, *inter alia*, laser capture of unstained buccal epithelial cells deposited on a proprietary polymer coated slide, gravity capture of individual laser-enscribed cells, thermostable protease-mediated cellular lysis, amplification of 15 autosomal STR loci with increased cycle number (AmpFlSTR® Identifiler® at 34 cycles) and separation and detection of the STR alleles using capillary electrophoresis. Cell recovery is efficient by LCM: cells visualized under a stereomicroscope (400 X) (Figure 9A) are marked for capture (Figure 9B) and, after application of the laser, are checked to see whether the appropriate cells have been removed (Figure 9C).

To test the efficacy of the combined cell recovery and genotyping system (LCM-LCN), buccal epithelial cells from 10 individuals (5 female and 5 male) were deposited on polymer-coated slides. From each of these individuals varying numbers of cells were captured for analysis (1-, 2-, 3-, 4-, 5-, 10- and 20-cell samples). Apart from these 'biological replicates', 'technical replicates' comprising ten replicates of each cell sample size/donor combination were also collected. Low copy number (LCN) DNA analysis, which employed 34 PCR cycles, was performed using multiplex autosomal STR analysis with 15 loci.

The LCM-LCN method is sensitive and yields accurate DNA profiles from single donors with varying degrees of success, depending upon the number of cells. The results are summarized in Table 3. Single cells yielded some genetic information from the donor 74% of the time (1-29 alleles) with 4% providing a full profile (30 alleles). Twenty cell samples, on the other hand, virtually always (99%) provided some genetic information from the donor with 80% providing a full profile. Intermediate cell numbers (2-10) yielded intermediate levels of genetic information (e.g. full profiles observed 19-66% of the time).

Binomial Sampling and Analysis of Two Donor (1:1) Mixtures of Epithelial Cells

As a proof of concept for the proposed mixture genotyping approach involving binomial sampling/LCM-LCN analysis and quantitative computer interpretation, two different 2 person (50:50) mixtures were prepared (one female:female (F2/F3) and one male:female (M1/F1) mixture with a different female donor, the genotypes of which are listed in Table 4). From each

mixture 10 x 20 cell sub-samples were collected and each of the sub-samples subjected to separate DNA analysis. The resulting electropherograms demonstrated the presence of admixed DNA as indicated by the presence of more than 2 alleles at a locus and significant allelic signal imbalance. An example of a mixed profile with the expected characteristics obtained from a single 20 cell sub-sample is shown in Figure 10. In accordance with the binomial sampling hypothesis, the sub-samples exhibited different weight ratios of the two donors comprising the mixture (Figure 11).

Computer-Inferred LCN Genotypes

We ran the TrueAllele® Casework system on each LCM LCN binomial sample, each of which contained exactly 20 cells sampled from a two-person mixture. There were 2 different 50:50 mixture proportion preparations (M1:F1 and F2:F3), each one sampled 10 times, for a total of 20 data items. The TrueAllele® computer inferred 2 probabilistic genotypes from each item, for a total of 40 inferred genotypes. These genotypes were subsequently matched against the known contributor genotypes, yielding 37 informative genotypes that had positive log(LR) information. The computer also inferred the posterior probability distribution of mixture weight for each genotype contributor [42].

Binomial Sampling Mixture Weight

Standard processing of N = 200 cells sampled from a 50:50 DNA mixture extraction produces a mixture weight close to 50% for each contributor. More precisely, the binomial sampling theorem tells us that with 0.95 probability, the weight would lie within the range 50 ± 7%, since $\mu = N/2$ and $\sigma^2 = N \times 0.5 \times 0.5$ for a $(\mu \pm 2\sigma)/N$ fraction. Mixtures near 50% lose considerable identification information because their STR peak heights are less able to facilitate differentiation between contributing genotypes.

Dividing the entire DNA template into 10 groups of 20 cells should proportionately increase the mixture weight variance 10-fold. This factor of $\sqrt{10}$ standard deviation gain would expand the mixture weight range to 50 ± 22%, since 3.162 × 7% = 22%. Thus, if binomial sampling operates properly, we should see appreciable deviation from mixtures having a 50% contributor weight.

The mixture weights of the 37 informative genotypes are shown in a histogram (Figure 12), indicating dispersion away from 50%. The mean of these TrueAllele®–measured weights is 49.6%, with a standard deviation of 17.7%. The observed 0.95 interval of 50 ± 35% is well dispersed from the 50% center value, as theory predicts.

Identification Information

The TrueAllele®–measured identification information is shown for each of the 37 informative genotypes in log(LR) ban units, together with its mixture weight (Figure 13). We observe that the genotypes can be very informative, particularly with imbalanced mixtures of 25% and 75% where peak height differences help in separating out the two contributors. The maximum information for a genotype inferred from a single 20 cell sample was 18.49 ban, with an average log(LR) information of 11.73 ban (Table 5).

Binomial Pairs

Statistical theory predicts that combining data from multiple experiments should increase information. Intuitively, this gain occurs because each additional data item imposes further

constraints on possible genotype solutions, sharpening the probability distribution of those allele pairs that are most consistent with all the data patterns. When the current information is not near the maximum, doubling the data quantity should increase information by about a $\sqrt{2}$ factor. We tested this information gain hypothesis by having TrueAllele® jointly interpret all possible pairings of binomial DNA samples from the same mixture using a joint likelihood function.

For each LCN sample, TrueAllele® requests were set up for all pairings with the other 9 binomial samples drawn from the same DNA mixture. TrueAllele® then solved these 45 interpretation requests using a two sample JLF, inferring 2 probabilistic genotypes for each pairing. Following objective genotype inference, comparison was then made between the inferred genotypes and the two known contributor genotypes to compute LR values.

Pairing one DNA sample with another generally increased identification information, but could also elicit no change or even decrease information. One factor was how well the pair's data complemented each other. For example, dissimilar mixture weights can reverse the roles of major and minor contributor, providing complementary information for separating genotypes [56]. We show the information resulting from pairing binomial sample 6 with its nine companions, for the matching inferred genotype (Table 6). Note that relative to the singly inferred sample information of 12.13 ban, the pairwise jointly inferred genotypes have match information ranging from 9.02 to 17.65 ban.

The greatest information gain (over 5 ban) for sample 6 was seen in pairwise combinations with samples 1, 4 and 10. These samples have respective minor contributor mixture weights of 21.0%, 20.5% and 19.5% that are complementary to sample 6's major 62.6% mixture weight. Combining sample 6 with sample 1 reproducibly increased information by 5.52 ban to a maximal total of 17.65 ban, producing a single match information number for the 9 jointly inferred genotypes.

We recorded the log(LR) information for each binomial sample (relative to a corresponding reference) as the maximum joint LR match value, over its 9 sample pairings. The bar chart (Figure 14) shows the original match information from a single sample (blue bar), the information gained by pairing with the most complementary second sample in a joint genotype interpretation (red bar), and the total recovered information for the sample from joint inference (combined stack height).

Inferring genotypes using a joint likelihood on just two samples, the computer recovered essentially the full genotype information for most of the binomial samples. The average information of pairwise joint interpretation was 17.22 ban (Table 5). The pair–to–single information ratio averaged 1.468, which is close to the 1.414 (i.e., $\sqrt{2}$) predicted theoretical information gain.

Standard PCR Data Combination

The TrueAllele® joint interpretation methods are also applicable to standard PCR conditions. At 28 cycles, the LCM binomial samples of 20 cells showed highly variable peak height patterns with many low peaks under 50 RFU (Figure 15). However, when interpreted jointly, these separate weakly informative patterns acted in concert to constrain the possible genotypes.

TrueAllele® computer jointly processed all ten 28 cycle binomial samples together in a single interpretation run on the mixture pair F2:F3. The 10 inferred mixture weight probability histograms are shown (Figure 16). These distributions have centers that show a diverse range of binomial sample mixture weights, with values falling far from 50%.

The two inferred contributor genotypes were both uniquely determined by TrueAllele® at the 99% probability level. These definite genotypes reflected the certainty of the combined data, and yielded full log(LR) match information equivalent to an RMP single source match statistic. The full match information was also reproducibly recovered by joint interpretation with just three binomial samples, as found using two disjoint sample subsets. Therefore, while simultaneous analysis of all the samples can be informative and feasible, in practice far fewer samples are needed to preserve identification information.

Summary of Binomial Sampling and Probabilistic Computer Interpretation of Two Person Mixtures

The data confirms that, as predicted, binomial sampling of two person (1:1) epithelial cell admixtures by LCM produces sub-samples with differing weight ratios. The recovery of different sub-samples containing different proportions of each constituent necessarily and demonstrably facilitates DNA mixture genotype inference. We showed that quantitative computer interpretation using the TrueAllele® system efficiently recovers the full genotype information from the original sample using the sub-samples' genotype data. Conceptually, then, it is possible using the binomial sampling method to construct more informative data from the same evidence, compared to standard biological sampling strategies.

1.3 Limiting Dilutions

An alternative approach to generating variation in relative cell ratios from two donors could involve a standard DNA extraction of a mixed stain, followed by preparation of sub-samples by limiting dilution. These 'Poisson' random aliquot samples would be expected to contain varying proportions of the starting constituents, just as in binomial sampling. While the use of such an approach could not be fully evaluated within the scope of the current work, we were able to perform an initial evaluation of this approach.

To begin our analysis of the "limiting dilution" approach described above, we first prepare 1:1 mixtures and evaluated them using standard methods to ensure that they exhibited DNA profiles consistent with bona fide 1:1 mixtures. We used the AmpFlSTR® COfiler® kit for this initial testing to simplify analysis and reduce cost (numerous COfiler kits were available in our laboratory for use in this experiment). The two donor mixtures consisted of a male and female donor so that we could use amelogenin for reference. We prepared sufficient volumes of these mixtures so that future experiments would be performed from these samples and not have to be re-made. We amplified 1ng of the 1:1 mixtures and also amplified each individual donor to obtain reference profiles. Four mixtures were prepared using different donors in each mixture. After this initial evaluation of the mixture samples, we selected the best 1:1 mixtures (X:Y ratios of 3:1).

Using the selected mixture, we amplified 20 replicates of a 50pg (from a 1:10 dilution of the original amplification) sample. We used Identifiler® Plus and standard cycle number (28 cycles) for this experiment. Each of the obtained profiles was evaluated to determine average donor weight ratios. For the purposes of this initial study, we will describe the analysis using only loci for which four distinct alleles from the two heterozygous donors were present. The obtained weight ratio averages for this experiment are provided in Table 5. The standard deviation for each of the averages is also provided. We were looking for variation among the weight averages which would indicate that in some samples donor 1 (male) was the major contributor, but in others donor 2 (female) was the major contributor. If this occurred, it may be

possible to easily discern the genotypes of both donors using these samples. The results from this experiment are provided in Table 7 and Figure 17. While we did not observe extremes in weight averages as we might have expected (more similar to that observed with the 20-cell LCM samples), there was variation in weight ratio averages that provide an initial indication that such an approach might warrant further investigation (although significant optimization will be required in order to determine optimal inputs and number of replicates needed for analysis). The male donor average weight ratios (ω) amongst four allele loci ranged from 0.41 to 0.67, thereby including sub-samples in which the male donor was the minor and major profile contributor. While it was difficult in this initial experiment to evaluate these differences across entire profiles, an evaluation of individual loci demonstrates the potential future success of such an approach. Figure 17 shows an example of differing weight ratios at the same locus (D8S1179; male genotype 12,16; female genotype 13,15) that could aid in mixture de-convolution. This would especially be the case if a joint likelihood function approach was used in combination with probabilistic computer interpretation as described above for the binomial sampling strategy.

1.4 Enhanced Single Cell Analysis

Despite the successful development of the probabilistic computer interpretation (TrueAllele®) approach to mixture de-convolution of admixed epithelial cell mixtures [55], we nevertheless continued to develop our single cell analysis methods since single cells would provide the ultimate mixture de-convolution that would permit direct typing of the constituent single source DNA genotypes of the individuals present. All of our previous experiments had been performed using the Identifiler® amplification kit with increased cycle number. We therefore also evaluated the Identifiler® Plus (not previously available for use in initial studies) and PowerPlex® 16 HS amplification kits in order to determine whether these next generation kits with improved chemistry would result in superior profile recovery of single micro-dissected cells. This analysis was performed using 5 of the 10 donors previously tested (ten 1-cell replicates; 34 amplification cycles). A comparison of the success rates for profile recovery using the three amplification kits (Identifiler®, Identifiler® Plus and PowerPlex® 16 HS) is provided in Figure 18. As can be seen from this figure, the Identifiler® Plus kit resulted in an increased success rate compared to Identifiler® and PowerPlex® 16 HS. A similar success rate was obtained for single cells between Identifiler® and Identifiler® Plus. No full profiles were obtained from single cells using the PowerPlex® 16 HS system in this small scale study.

In addition to profile recovery success, additional factors were compared between the three amplification kits: 1) average RFU values (Figure 19); 2) average heterozygous peak balance (Figure 20); 3) average locus drop-out rate (Figure 21); and 4) average allele drop-out rate (Figure 22). An increase in allelic signal intensity for almost all loci was obtained with the Identifiler® Plus amplification kit (Figure 19), although sufficient peak heights (well above detection thresholds) were obtained for all kits. Despite the use of single cells (extreme low template input amount), the heterozygous peak height balance was >60% for most loci across the three amplification kits (Figure 20) (although some heterozygous loci were affected by allele drop out and therefore are not included in the analysis). The average locus drop-out rates for each locus are provided in Figure 21. Higher locus drop-out rates are observed for the larger sized amplimers (ex: CSF1PO, D16, D2, D18 and FGA). This same increase in drop-out rate for larger sized amplimers was not observed for single allele, as opposed to locus, drop-out (Figure 22). We additionally evaluated single cell samples with the Yfiler® amplification kit. Y-STR loci

(with the exception of the bi-local locus DYS385) are not affected by heterozygous peak imbalance and the formation of pseudo-homozygotes due to drop out. Figure 23 shows the success rate of profile recovery for single cells using the Yfiler® amplification kit. As can be seen from this figure, no improvement in the recovery of full profiles was obtained using the Yfiler® kit. However, again the success rate of partial and full profile recovery was greater than 70%. Since there was no improvement in the success rate for single cells and autosomal STR analysis is the gold standard in DNA profiling, we did not include Yfiler® in subsequent studies. It should be noted that since this work was completed next generation Y-STR kits have become available and these may be better suited than the first generation ones for such single cell analyses. We did not have the opportunity to test these in this study. Based on the overall results of the comparison studies of the three amplification kits for single micro-dissected epithelial cells, the Identifiler® Plus kit was selected as the most suitable for use with single epithelial cells.

1.5 Micro-Volume PCR reactions

Despite the improvement in profile recovery with the Identifiler® Plus kit, additional optimization of the developed methods were needed in order to improve profile recovery from single cells. All of our previous work with single cells has generally been performed using increased cycle amplifications but still using standard reaction volumes. However, the use of standard reaction volumes may not be ideal for use with single cells. It is possible that low-volume PCR reactions may be more efficient for this purpose.

The Advalytix AmpliSpeed Slide Cycler appeared to be an ideal option for performing reduced volume amplifications since it was specifically designed for 1µl reaction volumes. We began our initial attempts at the use of these micro-volume 1µl reactions using this system. However, we were obtaining inconsistent results with assays performed on this instrument. Upon contacting the company for a possible repair, we were told that the instrument and consumables were no longer manufactured. While we possessed the instrument and some consumables, since this was not a long-term viable option for use, we discontinued all studies using the Advalytix system.

We continued to study micro-volume PCR reactions using standard equipment (ABI 9700 thermal cyclers). Development of these amplifications on standard equipment would provide a possible protocol for operational crime laboratories to use without the requirement to purchase additional equipment. To begin our development of low volume PCR reactions, we amplified a small number of DNA samples (1ng input) using decreasing PCR reaction volumes. We evaluated the use of 10 µl, 5µl, 2.5µl and 1µl reaction volumes for the Identifiler®, Identifiler® Plus and Yfiler® amplification kits. We utilized the amplification programs, including cycle number, recommended by the manufacturer (28 cycles for Identifiler®, 29 cycles for Identifiler® Plus and 30 cycles for Yfiler®) since we were using 1ng of input. We were able to obtain successful results using the 2.5 – 10µl reaction volumes. However, we had variable success with the 1µl Identifiler® amplifications. Initial experiments were performed using just the 1µl reactions without any kind of sealing solution over the reaction. We observed complete sample evaporation for several samples. The use of 1µl reaction volumes in standard thermal cyclers was therefore not ideal as it required the use of mineral oil to prevent sample evaporation which then also resulted in the need for post-PCR purification which ended up diluting the amplified product since it could not be eluted into the same 1µl volume. Therefore, after much optimization of these reactions, we determined that it was not suitable for use.

We next evaluated the use of additional micro-volume PCR reaction volumes (2.5 – 5µl) using 5 pg of input DNA (single cell equivalents) using both standard (28) and increased (36) cycle number. We evaluated profile recovery using each reaction volume, but also evaluated evaporation rates (volume remaining after amplification compared to starting volume) and average RFU values. The results from this experiment are provided in Figure 24. Surprisingly, similar profile recovery values were obtained using standard and increased cycle numbers, indicating the potential improvement in sensitivity simply from using a micro-volume PCR reaction rather than having to increase cycle numbers. As can be seen from Figure 24, even with the increase to 2.5 µl reaction volumes the sample evaporation rate was high. We ultimately decided on the use of 3.5 µl reaction volumes based on profile recovery success rates, low evaporation rates and high RFU values.

These initial experiments were performed using extracted DNA and not isolated epithelial cells. It was not possible to transfer these protocols to the LCM instrument since a large reaction volume was needed to cover the surface of the PCR tube cap for efficient cell recovery. We therefore needed to utilize a micro-manipulation approach rather than LCM collections for the recovery of single or few epithelial cells. In this approach, a tungsten needle is used to collect a "ball" of 3M water soluble adhesive at the end of the needle. The water soluble adhesive is then touched the surface of the Gel-Film™ in order to collect the desired number of cells. All work is performed using a Leica M205C stereomicroscope. The needle (with adhesive and cells attached) is then placed into solution (lysis or amplification reaction) inside a 0.2ml PCR tube. The adhesive dissolution and cell transfer into solution is viewed under the microcope. Since the previous experiments involved extracted DNA, no lysis step was required. However, since we were now evlauating isolated epithelial cells, we needed to determine if the use of a lysis (forensicGEM®) step would be required. Based on the results of preliminary data of the analysis of 1-10 buccal peithelial cells isolated using micro-manipulation (standard reaction volumes) (Figure 25), the use of a lysis step rather than direct PCR amplification resulted in higher profile recovery success. Rather than performing the lysis first and subseuqently adding the PCR reaction mix, we were able to develop a combined lysis/amplification reaction. The lysis reaction involves incubation at 75ºC for 15 min followed by enzyme inactivation at 95ºC for 5 min. Therefore, a 75ºC 15 min step was included in the beginning of the amplification program and the initial hot start for the STR amplification was also used to inactivate the lysis enzymes. The lysis/amplification is performed in a single tube reaction without the need for additional sample manipulation or tube opening, thereby reducing the potential for contamination. We next evaluated the use of the combined lysis/amplification using a 3.5µl reaction volume using 1, 2, 5 and 10 isolated buccal epithelial cells (Figure 26). The use of the 3.5µl combined lysis/amplification method was shown to be suitable for use with isolated buccal epithelial cells and resulted in improved profile recovery compared to standard reaction volumes for most cell numbers.

Since only a small number of donors were used in the initial experiments, we next evaluated profile recovery from single buccal epithelial cells from additional donors (8 donors, four female and four male; 10 replicates from each donor). The results from this experiment are shown in Figure 27. For six of the eight donors, an increase in profile recovery success was observed. In addition to the analysis of buccal epithelial cells, we have also succesfully utilized the 3.5µl combined lysis/amplification reaction for the analysis of skin bio-particles isolated from worn clothing items using micro-manipulation (Figure 28, shirt collar).

While the use of the micro-volume PCR reactions are currently still being fully evaluated and optimized, the initial studies demonstrate the initial success in profile recovery using the developed micro-volume PCR reactions. In addition to an increase in profile recovery, these reactions provide the additional benefit of reduced cost of analysis since significantly less amplification reagents are required per reaction.

2. Cell Type Identification (RNA Profiling) of Micro-Dissected Cells

As described above, we had good success with DNA profiling from individual or few epithelial cells isolated using laser-capture micro-dissection or micro-manipulation. However, the goal of this study was to not only individualize isolated epithelial cells through the recovery of the genetic (autosomal STR profile) of the cell donor but to also determine the body fluid or tissue source of origin of the cell as well. We proposed to accomplish cell type identification using RNA profiling. Cell type identification experiments were mainly focused on buccal, vaginal and skin epithelial cells. We performed extensive experiments in order to identify novel vaginal secretion and skin mRNA biomarkers for use in cell type identification experiments. However, as with DNA profiling of individual or few epithelial cells, we needed to develop robust and sensitive "RNA-friendly" lysis, amplification and typing strategies to permit an identification of the cell type of origin using mRNA profiling. The ability to perform DNA and RNA profiling from the same cell was ideally what we hoped to accomplish through the current work. However, RNA profiling of single or few epithelial cells from dried forensic body fluid samples proved to be quite challenging. Therefore, as we will describe in the subsequent sections, our initial method development included RNA-only isolation and detection methods. The initial combined DNA/RNA isolation methods that we have begun to perform will also be described.

For any developed RNA isolation and detection strategies, additional factors need to be considered beyond that required for any of the DNA isolation strategies. RNA is highly susceptible to RNase degradation and it is possible that direct lysis reactions may not provide sufficient inactivation of RNases. Therefore, the use of RNase inhibitors may need to be present in the initial preparation of the slide, and/or in the lysis buffer. Our sample preparation strategies involved the use of heat fixation (LCM sample preparation) or the use of sample smears or air drying for extended periods of time (micro-manipulation sample preparation). It is unknown whether such heating/drying has a detrimental effect on RNA, in that during these processes it is possible that cell membranes are perforated or "burst" causing RNA to "leak" from the cell. Since DNA is contained in the cell nucleus, outer cell membrane disruptions would not be as detrimental to its recovery. Additionally, the sensitivity of current RNA profiling methods has not been evaluated in individual or isolated cells. Typically, RNA profiling assays involve the use of body fluid stains or swabs where a sufficient amount of material is present. Therefore, a failure to detect RNA biomarkers using any individual RNA isolation strategy may simply be due to the sensitivity limitations of detection strategies and not the lysis/isolation method itself. Additionally, it is possible that the selected mRNA biomarkers may be found in significant abundance in the "liquid" portion of body fluids and less in the cellular environment. This presents a challenge when cells are directly isolated for analysis and the fluid component of the body fluid is essentially removed during sample preparation. In developing RNA direct lysis and

detection strategies for individual cells we took the above challenges into consideration when trying to optimize the system.

2.1 Initial RNA Isolation Method Evaluation

***Note – Extensive evaluation and optimization experiments for RNA isolation and typing were performed throughout the current work. However, for the purposes of this report, only brief summaries of initial testing is provided since initial methods were not successful. More detailed descriptions of the recently developed (and more successful) methods are provided in subsequent sections.*

Initially, we evaluated five lysis buffers for RNA isolation: 1) Quick Extract (Epicentre); 2) Cells Direct (Invitrogen); 3) CytosALL (Thermo-Scientific); 4) Cells to Ct (Applied Biosystems/Ambion/Life Technologies); 5) RNA Quick Extract (Epicentre). We evaluated each lysis method according to the manufacturer's instructions (with minor modification of reaction volumes if needed). Attempts were made to reverse transcribe the cell lysates and amplify them using gene specific mRNA's (CE-based assays). Despite attempts at optimization of these lysis methods, no results were obtained from any of the cell lysates.

Since we did not have any initial success with RNA isolation using direct lysis approaches, we performed some basic experiments in order to determine the sensitivity of our existing RNA extraction and profiling methods. We routinely extract RNA from dried body fluid stains. However, we had not performed experiments to determine if RNA of sufficient quantity and quality for analysis can be recovered from cells rather than whole body fluid samples. Possibly RNA is present in the non-cellular or fluid portions of body fluids and therefore the sensitivity of our RNA profiling strategies might be reduced if we examined only the cellular component. In order to determine the sensitivity limitations of our existing methods, we prepared cell suspensions from buccal and vaginal swabs. This involved the pelleting of cells using centrifugation and the removal of the supernatant containing and fluid components of the sample. Cell counts were performed using a hemacytometer in order to prepare a serial dilution of an aliquot of the cell suspension. The dilution series included a range of cell numbers, from ~1 – 8,000 cells for the saliva/buccal samples and from ~1 – 15,000 cells for the vaginal samples. These samples were applied to the tip of a cotton swab and dried overnight. We dried the samples in order to simulate the type of samples that would be encountered in forensic casework and to ensure that the drying process did not affect our ability to successful recover RNA from these samples. The samples were extracted using standard organic RNA extraction (see Methods). The amount of RNA obtained from these samples was determined using a RiboGreen quantitation. The quantitation results from the cell dilution series can be found in Table 8. As can be seen from these results, sufficient amounts of RNA were obtained to perform a reverse transcription reaction. At the time of this study, our mRNA profiling assays required the use of 50 ng for standard RT reactions. Numerous samples did not permit 50ng of total RNA to be added to the RT reaction and therefore the maximum amount of RNA (based on the volume of extract allowed to be added based on the reaction volume limitations) was used. The expression of saliva and vaginal secretion specific mRNAs was then evaluated in each of the samples. Figure 29 (top panel) shows the results from the detection of the salivary biomarker HTN3 in the dilution series samples. HTN3 was successfully detected in samples with >100 cells. Additionally, a vaginal specific marker (MUC4) was also detected in the vaginal cell dilution series for samples with >100 cells (Figure 29, bottom panel). The results of these experiments indicated that it would be possible to recover sufficient quantities of RNA for analysis from cells. However, the sensitivity of the RNA extraction and profiling method utilized was not as good as

we had hoped. These experiments were performed using a standard extraction method. It is possible that the multiple physical manipulations required during this extraction adversely affected the sensitivity due to sample loss. We reasoned that the use of a direct lysis approach would likely be better suited for use with samples containing small numbers of cells (<100).

In addition to the sensitivity experiments, we decided to perform some basic experiments to try to determine if during the cell drying process (whether air-drying or heat fixing) outer cell membranes were rupturing with an associated loss of RNA from the cells themselves. This would not be a problem if liquid saliva or buccal swabs were used because the RNA lost outside of the cells would still be in the sample or on the swab and therefore included in the extraction. However, isolation of individual cells would be greatly affected by a loss of integrity of the cell membrane since cytosolic RNA might no longer be present in the cells that were recovered. The first experiment we performed involved a comparison of results from liquid and dry saliva samples. We placed liquid saliva in varying volumes (1-100 μl) onto cotton swabs and allowed it to dry overnight. We placed aliquots of the same volume and from the same extract into 1.5ml tubes for direct comparison. We initially performed this experiment using our standard organic RNA extraction (rather than a direct lysis approach) since this extraction protocol had been extensively used and we knew it was an effective method for RNA recovery. We quantitated the RNA extractions and for a majority of the samples, RNA quant values were not detectable. With increased volumes (25-100μl), some quantitation values were obtained but they were relatively low. This was not unexpected as the quantitation from full buccal swabs can often be quite low. The extracts were reverse transcribed and analyzed using a B2M (housekeeping gene) and HTN3 singleplex reaction. The results from this study can be seen in Table 9. B2M and HTN3 were successfully detected from a majority of samples. The results of this experiment indicate that we could successfully recover RNA from both liquid and dry samples.

The above experiment indicated on a larger scale that drying did not have an immediate detrimental effect on RNA recovery. However, we again extracted a swab portion for the dried samples, so if there was cell-free RNA it would still be included in the samples. So for the next experiment, we tried to determine if cell-free RNA was present. For this, we used a 15 μl liquid saliva sample and a 15 μl aliquot from the same liquid sample that had been dried. For the liquid sample, we immediately centrifuged the sample and isolated the supernatant. We then performed a total of 10 washes and retained the supernatant for washes 1, 2, 5 and 10. The purpose of this was to determine if RNA would be detected in the supernatant, which could indicate the presence of cell-free RNA. While it is expected that some cells will be ruptured during the centrifuge steps, by progressively monitoring the contents of various wash steps we could evaluate whether RNA was lost with each wash or if a significant amount of RNA was present in the initial wash steps. For the dried sample, the swab piece was added to an initial volume of water in order to remove the sample from the cotton. The same 10 washings were performed for this sample as well. All fractions (cells, initial supernatant, and then washes 1, 2, 5 and 10) were extracted using our standard organic RNA extraction. Samples were treated with DNase, reverse transcribed and then amplified using B2M and HTN3 singleplexes. The results from this experiment can be seen in Table 10. For the liquid sample, B2M and HTN3/HTN1 were detected in the "cell" fraction and only B2M was detected in the initial supernatant. This result was promising since it demonstrated that HTN3 can be detected in cells and that lower amounts were present in the supernatant. B2M and HTN3 were detected in each of the wash supernatant fractions which may indicate that cells were being lysed with each wash step and therefore RNA from cell leakage was recovered in these supernatant fractions. However, this was the liquid saliva sample and

most forensic cases will not involve liquid samples. For the dried saliva sample, B2M and HTN3/HTN1 were detected in the original "cell" fraction. This demonstrates that saliva-specific RNA is found in the cellular fraction even after sample drying. We are hopeful, therefore, that some cells remain essentially intact (or at least impermeable to macromolecules) after drying and therefore we should be able to successful recover RNA from cells and potentially identify the body fluid or tissue origin of these cells. Limited detection was observed in the wash fractions for the dried saliva sample. For the dried sample, there was greater RNA detection in the original supernatant compared to the liquid sample. So it is possible that any lysis that took place had already occurred during drying and less took place during subsequent steps. Overall the results of the experiment indicated that while some cell lysis is probably occurring during drying, RNA is still present in significant quantities in dried cells and we should therefore be able to isolate RNA from dried saliva stains.

After performing these somewhat fundamental RNA experiments and determining that RNA should be recoverable from cell fractions of body fluid samples, an RNA isolation kit became available from ZyGEM® (same manufacturer of our DNA lysis buffer). Since we had great success with the forensicGEM® lysis method for DNA, we wanted to evaluate the use of the *RNA*GEM Tissue kit for the isolation of RNA from isolated epithelial cells. We performed a similar experiment to that described above using the cell serial dilution samples for both buccal and vaginal swabs. We used the recommended protocol which is a simple 5 minute lysis step at 75^{o}C. The entire cell lysate was reverse transcribed without any additional purification (standard RT method, see Methods). The cDNA was then amplified using primers specific for both a body-fluid specific mRNA (HTN3 – saliva, MUC4 – vaginal) and a housekeeping gene (B2M). The housekeeping gene results for the buccal cell dilution series can be seen in Figure 30. The housekeeping gene was detected with as little as 49 cells (although some cell suspensions, probably for some unknown technical reason gave no results). This was an improved sensitivity over that obtained with the standard RNA extraction. However, the body fluid specific mRNAs (HTN3, MUC7) were detected in fewer of the samples then with the standard RNA extraction. HTN3 was still detected in the 49 cell sample. So while the results were not as consistent compared to the housekeeping gene, the same sensitivity was observed. Similar results were obtained for the housekeeping gene for the vaginal cell dilution series, however, no results for the body fluid specific mRNA (MUC4) were obtained (data not shown). Based on these preliminary results, it appeared that the *RNA*GEM lysis may provide the ability to isolate RNA from micro-dissected cells. However as a result of other mRNA profiling studies in our laboratory, it became apparent that new biomarkers for the identification of vaginal secretions were needed. The specificity of MUC4 was not sufficient (i.e. cross-reactivity with saliva). Additionally, at this time no biomarkers for the identification of skin were available. Therefore, we performed extensive biomarker identification and characterization experiments in order to identify novel biomarkers for the identification of vaginal secretions and skin. These new biomarkers could then be used in subsequent further development of RNA isolation and typing strategies.

2.2 Identification of Novel mRNA Biomarkers for Cell Type Identification

Previous studies conducted by our laboratory had led to the development of messenger RNA profiling assays for body fluid identification that exhibit a high level of tissue type

specificity [45-47]. We identified numerous candidate mRNAs that appear to be specific to particular body fluids including blood (16 candidates), saliva (6 candidates), vaginal secretions (10 candidates), menstrual blood (24 candidates), and skin (19 candidates). Importantly we demonstrated that mRNA from these candidate genes can be recovered in sufficient quality and quantity for analysis from dried body fluid stains or tissues [45-47,57,58]. However, we sought to identify novel biomarkers, particular for vaginal secretions and skin, that would provide improved specificity and potentially sensitivity for use with RNA profiling from single or few isolated epithelial cells that are often found in admixed stains.

2.2.1 Vaginal mRNA Biomarkers[51]

Identification of vaginal secretions (VS)-specific gene candidates

The human transcriptome of vaginal secretions was interrogated using a 'Next Generation Sequencing' platform (RNA-Seq) [59]. Whole transcriptome sequencing of human skin was also performed for comparison in an attempt to identify those mRNA biomarkers expressed at a higher level (or even exclusively) in vaginal epithelia compared to other epithelial cell types (exemplified by skin epidermal tissue). Paired-end sequencing (54 bp average length) was performed on total RNA isolated from vaginal secretions (two donors) and human skin (two donors). Sequencing data was obtained for all samples (average matched read counts of 17 million for skin samples and 12 million for vaginal samples). Merged sequencing reads were aligned to the human reference sequence assembly (GRCch37/hg19), which permitted the maximum count numbers for each of over 42,000 gene segments to be determined in each sample. The maximum count values were then compared between the vaginal and skin samples in order to identify possible highly abundant and specific vaginal secretions candidates.

Initial Specificity Screening of VS Gene Candidates

Expression of the RNA-Seq candidates was evaluated in total RNA samples from a screening panel of biological relevant fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin). RT-PCR was conducted using gene specific primers and the products detected using gel electrophoresis (data not shown). Six genes (from 102 screened candidates) demonstrated a high degree of VS specificity. High specificity was indicated by a gene transcript's presence in VS alone or co-expressed in menstrual blood (that will commonly contain VS) or by a significantly higher expression in VS compared to the other tested body fluids. The six putative VS specific genes included surfactant associated 2 (SFTA2), fucosyltransferase 6 (FUT6), dickkopf homolog 4 (DKK4), interleukin 19 (IL19), myozenin 1 (MYOZ1) and cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1 (CYP2B7P1). The RNA-Seq maximum count values for these genes are provided in Table 11. As can been seen from the data, these genes were found in low abundance or virtually undetected in skin samples. The PCR primer sequences and expected RNA (and DNA) product sizes for the six selected candidates are provided in Table 12.

Detailed Evaluation of Six Putative Vaginal Secretions-Specific Gene Candidates

Capillary Electrophoresis (CE)-based Detection

Singleplex assays suitable for CE-based separation and detection were developed for each of the six candidates (Figure 31A). Primers for all candidates were designed to span

different exons such that amplification of DNA would result in amplimers larger in size than the expected mRNA or would not be amplifiable under the specific conditions used (Table 12). Only IL19 produced a detectable product (260 bp) when genomic DNA was tested, which was easily distinguishable from the *bona fide* mRNA product (109 bp) (Table 12, Figure 31B). Moreover, no potentially confounding processed pseudogenes from any of the candidates were observed in genomic DNA. Single amplification products of the expected size were obtained for all VS samples tested (n = 20) (Figure 31A, left panels). No products were present in controls when reverse transcriptase was omitted, indicating that mRNA was being detected and not contaminating genomic DNA (Figure 31A, right panels).

Sensitivity

The sensitivity of the CE-based singleplex reactions for each of the six candidates was evaluated using varying quantities of input total RNA (25pg– 25ng) from five different donors and the results are summarized in Figure 32. Product detection at each input level is indicated by the presence of a colored square, with each color representing a different RFU level to indicate signal strength. A majority of positive results were obtained using 5ng or more of input total RNA. Significant detection was still observed for most candidates using 250pg – 1ng of input total RNA, although the results were more variable in this range. Three of the candidates (SFTA2, IL19, and MYOZ1) could be detected in at least one donor with as little as 50pg of total RNA. In order to minimize potential false negative results in subsequent testing, an input amount of 5 ng of total RNA was selected.

Specificity

More extensive testing of the specificity of the six vaginal secretions candidates was carried out using additional body fluid samples from different donors. Using 5ng of total RNA input, additional sample sets from different individuals comprising blood, semen, saliva, VS and menstrual blood were evaluated (Table 13). None of the six candidates were detected in any of the blood or semen samples. Five of the six vaginal candidates (with the exception of FUT6) were detected in one or two of the menstrual blood samples. However, this result is not surprising due to the expected presence of vaginal secretions in menstrual blood samples. Since it is critical that a putative VS biomarker be able to differentiate between VS and saliva, particular emphasis was placed on determining the candidate genes' expression in saliva stains. Thus, fifteen saliva samples were tested and, as can be seen from the Table, no cross reactivity was observed for IL19, MYOZ1 and CYP2B7P1. Some cross-reactivity with saliva was observed for SFTA2 (30% of samples), FUT6 (7% of samples) and DKK4 (20% of samples). However their expression (as indicated by RFU values) was significantly less in saliva, with average 9, 14 and 29 fold expression differences between VS and saliva for FUT6, DKK4 and SFTA2, respectively. Due to this low but measurable cross reactivity, these VS candidates may be better suited to a quantitative real time RT-PCR assay format. IL19, while not exhibiting any saliva cross-reactivity, was relatively insensitive since it was detected in only 2 of the 5 vaginal samples. Based upon these results, the best two candidates for VS identification employing a qualitative CE-based assay platform appeared to be MYOZ1 and CYP2B7P1, and all subsequent work using this qualitative assay format was performed with these two markers.

MYOZ1 and CYP2B7P1 were tested against increased input amounts of saliva total RNA (10 – 100ng), the larger amounts of which would not be expected to be recovered from, nor subjected to RNA analysis in, dried saliva stains. Expression of CYP2B7P1 was not detected in

saliva even with 100ng RNA input (Table 13). However, a small amount of cross-reactivity with saliva was observed for MYOZ1 with increased input amounts (7% and 13% of samples tested using 10 and 25ng of input saliva total RNA) (Table 13). A small number of saliva samples were tested with 50 and 100ng of total input RNA and one of the two samples was positive for MYOZ1 at each input level (Table 13). However the level of expression with this increased, but somewhat unrealistic, input level of total RNA from saliva was always 4-5 fold less than in VS (data not shown).

In addition to the specificity testing of forensically relevant biological fluids (blood, semen, saliva, and menstrual blood), the increased occurrence of 'touch DNA' evidence warrants the inclusion of skin in specificity testing experiments. The markers described here were selected based upon their non-expression in skin (Table 11) using highly purified skin total RNA. The latter skin RNA samples were of high quality since we were able to successfully detect and identify skin-specific transcripts therein (see below). Additionally, potential skin expression of CYP2B7P1 and MYOZ1 was tested in RNA recovered from swabs of the surface of a male forearm as well as from a recently touched computer mouse. No expression of CYP2B7P1 or MYOZ1 was detected in either skin sample (data not shown) despite both of these samples being positive for the skin-specific marker LCE1C [54] and the B2M housekeeping gene (data not shown).

Mock Sexual Assault Casework Samples

The performance of the two best VS candidates with simulated casework samples was evaluated. The simulated casework samples were designed to represent possible casework scenarios including digital penetration, sexual assault (vaginal intercourse) and assault with a foreign object. The simulated casework samples therefore included the following: 1) swab of the surface of male fingers after digital penetration of a female participant; 2) penile swab before and after intercourse with the 'before' swab serving as a control to ensure that no prior vaginal secretions were present; 3) swab of the inside of male underwear worn 3 hours after intercourse; 4) swab of the surface of a foreign object after insertion into the vaginal canal.

The results from the simulated casework samples are provided in Figure 33. Both CYP2B7P1 and MYOZ1 were detected strongly in the swab of the male fingers following digital penetration (Figure 33A) indicating evidence of a transfer of VS onto the fingers. CYP2B7P1 and MYOZ1 were also detected on the penile swab after sexual intercourse (Figure 33C) but not detected on the control swab taken immediately prior to intercourse (Figure 33B). The presence of VS biomarkers (albeit with a weaker MYOZ1 signal) on the male participant's underwear, sampled 3 hours after vaginal intercourse, was also demonstrated (Figure 33D). Lastly, both CYP2B7P1 and MYOZ1 were detected on the surface of a vaginally inserted foreign object (Figure 33E). Collectively, these results indicate the potential ability of the CYP2B7P1 and MYOZ1 biomarkers to detect and identify vaginal secretions that may be transferred to the body of perpetrators and to foreign objects used during sexual assaults. More detailed follow up studies should address factors affecting the transfer and persistence of the markers over time.

2.2.2 Skin mRNA Biomarkers

Identification of skin-specific gene candidates

'Touch DNA' evidence is generally perceived to be the result of DNA obtained from shed skin cells transferred from donor to an object or person during physical contact [60-65].

Although published studies and common practice have demonstrated the ability to obtain genetic profiles from this type of trace biological evidence, the tissue source of the profile is rarely determined. Therefore, a main objective of the present study was to identify novel and highly specific gene candidates to enable the positive identification of a skin source of origin of trace biological evidence in 'touch DNA' samples. Previously, we have identified and extensively validated numerous body fluid-specific genes for use in messenger RNA (mRNA) profiling assays for the identification of forensically relevant biological fluids (blood, semen, saliva, vaginal secretions, and menstrual blood) [44-47,49,66,67]. Importantly we have demonstrated that mRNA from these candidate genes can be recovered in sufficient quantity and quality for analysis from dried body fluid stains or tissues. Due to the success of this previous work, we reasoned that the identification of skin would be amenable to this mRNA profiling approach too.

In order to identify potential skin-specific gene candidates, we performed literature searches and targeted 83 specific genes that are, or are likely to be, involved in skin development and/or regulation based upon an *a priori* understanding of skin physiological processes. The second approach to candidate identification was by means of deep sequencing of the skin cell transcriptome using a 'Next Generation Sequencing' platform (RNA-Seq) [59]. Whole transcriptome sequencing from another related external epithelial tissue type of forensic interest, namely vaginal secretions (VS), was also carried out for comparison purposes in an attempt to facilitate the finding of expressed genes in skin that are not present in VS. Fifty-four base pair (bp) paired-end sequencing was performed on total RNA isolated from vaginal secretions (two donors) and human skin (two individual donors). Sequencing data was obtained for all samples (average matched read counts of 17,000,000 for skin samples and 12,000,000 for vaginal samples). Sample data were mapped to the RefSeq human reference sequence assembly (GRCch37/hg19), which permitted the maximum count numbers for each of over 42,000 gene segments to be determined in each sample, thus indicating expression levels of each gene ('DGE' or digital gene expression). The maximum count values were then compared between the vaginal and skin samples in order to identify putative highly abundant and skin-specific candidates. In this manner 20 potential skin candidates were targeted for further investigation.

The 103 gene candidates identified using both approaches were screened by gel electrophoresis of singleplex end-point PCR products to provide an initial indication of transcript abundance and relative expression in a small panel of forensically relevant biological fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin) (data not shown). Candidates were rejected if: (i) no amplification products were obtained; or (ii) cross-reactivity with other non-skin body fluids was observed; or (iii) non-specific amplification products were present. Based on the results of this preliminary screening, five putative skin-specific candidates were identified including late cornified envelope genes 1C, 1D and 2D (LCE1C, LCE1D, LCE2D), interleukin 1 family member 7 (IL1F7, also known as IL37), and chemokine (c-c motif) ligand 27 (CCL27). Table 14 lists the primers, expected amplimer sizes, and the average RNA-Seq count values for each candidate in skin and vaginal samples. As can be seen from the average count values for each of the four samples (two vaginal, two skin), all candidates were found to be in higher abundance in skin, with significant lower count values observed in the vaginal samples. This digital gene expression data supported the initial screening results indicating a high degree of specificity for skin. The attributed biological functions of the five candidates are consistent with a high degree of skin specificity. Thus the late cornified envelope (LCE) gene cluster is part of the epidermal differentiation complex and encodes for stratum-corneum proteins [68]. Studies have shown that the LCE cluster contains "groups" (1, 2 and 3)

displaying differential expression with groups 1 and 2 dominant in skin (external epithelial) and down-regulated or undetectable in internal epithelial [68,69]. CCL27 is involved in the recruitment of CLA+ memory T cells to normal or inflamed skin. The most abundant expression of CCL27 is observed in keratinocytes of epidermal basal layers [70,71]. IL1F7 plays a predominant role in suppression of inflammatory responses and is synthesized by, inter alia, keratinocytes [72,73]. Multiple splice variants have been discovered for IL1F7, a-e [74]. Our primers were designed for isoform IL1F7b (also referred to as variant 1).

Subsequent to the initial screening, the five candidates were assayed using a laser-induced fluorescence-capillary electrophoresis detection platform, the latter of which was employed in all subsequent studies. Representative electropherograms from singleplex amplifications of each candidate are shown in Figure 34. During early testing of LCE1C, a product of the expected size (160 bp) was observed, but often a second co-amplification product (145 bp) was also obtained. The original primer set (designated LCE1C* in this report) was utilized for some studies before the primer set was re-designed in order to eliminate the appearance of this second product. Due to the large number of members in the LCE family, it is possible that an additional isoform was being amplified. A re-design of the reverse primer permitted an amplification of a significantly smaller sized product (58 bp), that was often accompanied by another product of unknown origin ~2 bases smaller (56 bp) than the expected product. The double peaks do not interfere with the interpretation of the results presented in this report but will be more fully characterized in subsequent work.

An initial preliminary evaluation of the skin specificity of the five biomarkers was carried out with a range of body fluids and skin. The results indicated that the biomarkers exhibited strong expression in skin samples and a high degree of specificity for skin. Varying amounts of input total RNA from each of the body fluids was used (~1 ng – 25 ng) along with some samples that had not been quantitated. No cross-reactivity with blood or semen was observed for any of the five candidates (4 – 10 donors tested per candidate). No cross-reactivity was observed for saliva or buccal swabs for LCE1C, LCE1D, LCE2D or CCL27 (10 - 13 donors tested per candidate). A minor IL1F7 peak (161 RFUs) was detected in one of eleven saliva/buccal samples tested. Cross-reactivity for a small number of samples was observed for vaginal secretions (LCE1C – 3/10 donors; LCE2D – 2/10 donors) and menstrual blood (LCE1C – 2/9 donors; LCE1D – 1/10 donors; LCE2D – 1/10 donors; IL1F7 – 1/6 donors). Despite the detection of some of the biomarkers in some non-skin samples, the intensity of the observed products was 10-45 times less than that of skin and many of these products were present in samples that had not had the extracted RNA quantitated.

Sensitivity of skin biomarkers

A forensically relevant assay should be evaluated for 'analytical sensitivity' using varying levels of input analyte to determine the assay's limit of detection (LOD) and dynamic range. In this paper we, like some others, refer to biomarker 'sensitivity' when really describing 'assay-specific sensitivity.' We are not measuring the absolute abundance of the biomarker in these sensitivity studies, since differently configured assays for the same biomarker may exhibit different LODs. The optimal input of total RNA for each of the five skin candidates was determined in singleplex assays using a range of input RNA (5 pg to 25 ng) and the results are shown in Table 15. The highest sensitivity was observed for LCE1C and CCL27 with detection down to 0.005 ng (5 pg). While not all donors were detected at this input level, the gene product signal intensity for those that were detected (>500 RFUs for approximately half of the donors

tested) was significantly above the signal to noise ratio. IL1F7 was also detected down to 5 pg of input total RNA but with slightly reduced signal intensities than that observed for LCE1C and CCL27 at this input level. LCE2D was detected in more than half of the 100 pg samples (8/13), with only sporadic detection of a few donors with less input. LCE1D proved to be the least sensitive marker, requiring 5 ng of input for detection in the majority of samples. Overall, the high sensitivity of most of the candidates indicated their potential suitability for use in the analysis of 'touch DNA' samples.

Confirmation of the specificity of the skin biomarkers using optimal input amounts

The sensitivity testing indicated that significantly higher input RNA amounts than what was optimal for each candidate was used in the preliminary specificity testing described above. Therefore, additional specificity testing was performed for each of the five candidates using a more optimal and realistic input amount, given their potential utility for touch samples (LCE1C - 25 pg; LCE2D, CCL27, IL1F7 - 250 pg; LCE2D - 5 ng). As can be seen from Table 16, a majority of all skin samples were detected at this input whereas no detection was observed in any of the other body fluid samples. This confirmed the high degree of skin specificity of each of the candidates and also further demonstrated the need to use appropriate input amounts of RNA in order to maintain this level of specificity. In addition to the body fluid samples, the expression of each candidate was evaluated in a panel of twenty human tissues (Table 16). Significantly, no cross-reactivity was observed for a majority of tissues. A minor weak product (52 RFUs) was observed in kidney tissue for IL1F7 and a stronger CCL27 product (769 RFUs) was observed in thyroid tissue. However, these tissues are unlikely to be encountered at crime scenes and therefore are not likely to interfere with the identification of skin in forensic analyses.

Detection of skin biomarkers in swabs of human skin and touched objects

The expression of the five skin markers was evaluated in a variety of extraneous human skin swabs or scrapings of human skin (e.g. forearm, knee, palm of hand, back of hand, face, sole of foot, leg, forehead, arm, wrist, finger, and neck) as well as touched objects from the indoor environment (e.g. mouse, telephone, coffee pot, keyboard, gloves, pencils, car keys and scissors). Each marker was detected in human skin samples and touched object samples (Figure 35), although not in all samples tested (~50% success rate) and expression of only one or two markers in each individual sample was often observed (Figure 35). Although co-expression of all five markers was not observed for any of the samples tested, some did express three or four of the markers. Significantly, the most sensitive marker (LCE1C) was detected in 100% of an additional set of samples tested (N = 12), which included swabs of human skin (e.g. cheek, arm, leg and forehead) as well as touched objects (e.g pencil, mouse, phone, door handle). The significantly higher sensitivity and rate of detection in touch samples with LCE1C was promising. However, it needed to be demonstrated that this ultra-high sensitivity did not inadvertently result in "false positive" results (i.e. cross reactivity with non-skin touched samples). Thus locations within our laboratory presumed to be negative for skin contact were swabbed and tested for LCE1C expression. Thirteen areas were tested with negative results except for two areas that, upon further investigation, could indeed have been touched (data not shown).

2.3 Enhanced RNA Isolation and Profiling Methods for Cell Type Identification

With suitable highly sensitive and specific biomarkers identified for vaginal secretions and skin, we were able to continue our efforts to develop suitable RNA isolation and typing strategies for use with isolated epithelial cells.

Previously, we began an evaluation of the *RNA*GEM Tissue kit (direct lysis). While we were able to obtain sufficient results using cell serial dilutions (using larger numbers of cells), we were not initially successful in the ability to detect mRNA markers in the micro-dissected cell samples (LCM) (data not shown). Since we were concerned about the possibility of the heat fixation step affecting cell integrity, we wanted to try alternative methods to sample preparation that eliminated the need for the heat fixation step. As described previously for our DNA experiments, we can utilize the Gel-Film™ adhesive material for epithelial cell recovery. Therefore, we decided to evaluate its use in RNA recovery protocols as well. A fresh buccal swab was smeared onto the Gel-Film™ surface. We used a trypan blue stain for better visualization. We collected 10, 25 and 50 buccal cells. For this initial RNA experiment, 10 µl of *RNA*GEM lysis buffer was used and the cells were collected directly into the lysis buffer. The samples were used by incubation at 75ºC for 5 min. Following lysis, the samples were reverse transcribed and then evaluated for expression of both a housekeeping gene (B2M) and a saliva mRNA (HTN3). We decided to utilize a singleplex reaction rather than our mRNA body fluid multiplex in order to ensure that we had a highly efficient amplification with no interference from other markers. Weak detection of B2M was detected in the 50 cell samples (only 57 RFUs), but HTN3 was not detected in any of the samples. In the previous experiment, we did not include a DNase step following cell lysis. Therefore, we repeated this experiment and included a DNase step in case that was affecting the efficiency of reverse transcription. We also included a purification step following DNase treatment in order to try to remove any cell debris or other impurities prior to reverse transcription. B2M was observed in the 25 cell samples, but HTN3 was not detected in any of the samples. We then tried modifying the lysis protocol, such as using increased temperatures (from 75 to 95ºC) for the lysis. None of the modifications tried seemed to improve detection of the housekeeping gene or saliva mRNA marker. It was therefore possible that the failure to detect the body fluid markers was due to the small amounts of RNA present in these samples and would be necessary to try to "pre-amplify" the small amounts of starting material in order to successfully detect the body fluid markers.

As described previously in the development of enhanced DNA profiling strategies, we demonstrated improved typing success with the use of micro-volume PCR reactions. RNA profiling from single or few epithelial cells up to this point had proven to be very challenging, so we wanted to evaluate the use of micro-volume RNA profiling reactions as well in order to determine if the same increase in typing success would be obtained for the RNA profiling assays as was obtained for the DNA profiling. Our initial experiments with the use of reduced volume reverse transcription and body fluid multiplex amplifications using total RNA (standard extracts) demonstrated an improvement in profiling success with the reduced volume amplifications when compared to standard reaction volumes (performed using total RNA extracts not cells; data not shown). This was evidenced by improved signal intensity and also sensitivity (biomarkers detected with lower total RNA input amounts) with the reduced volume reactions. After optimization of both reduced volume reactions, it was determined that the most suitable reactions were 5µl for both the reverse transcription reaction and the body fluid multiplex amplification.

With the success of the reduced volume approach, we next evaluated the use of these reduced volume reactions with isolated buccal and epithelial cells. At this time in the work, on a

separate but related project we were evaluating the use of RNA profiling with touch DNA samples (although this work is relevant here since skin is one of our target epithelial cell containing tissues of interest) including the use of reduced volume reactions. Initially, we extracted the touch samples using our standard RNA isolation method to obtain highly purified total RNA. Using this approach we were able to successfully detect the skin biomarker LCE1C in a variety of samples using the reduced volume amplification reactions. However, we also obtained results using the standard reaction volumes. These results were highly promising since we were successfully detecting the skin biomarker from touch DNA samples and not purified skin RNA. These initial experiments involved the extraction of a full swab used to collect trace material from touched objects. Prior to collection of bio-particles using our micro-manipulation approach, we attempted to prepare cell suspensions using touch samples and count the number of bio-particles in the cell suspension using a hemacytometer. However, there were few bio-particles detected if any. We proceeded with analysis using smaller aliquots (serial dilutions prepared) of the cell suspension in order to determine if we could obtain results using decreasing amounts of bio-particles. However, there was no detection of the LCE1C biomarker in these samples. We therefore evaluated possible strategies for improving the sensitivity of the detection reactions and therefore profiling success. We decided to try a pre-amplification reaction using the QIAGEN Multiplex PCR kit (from the manufacturer). While not specifically designed for use with isolated cells, we believed this would be an attractive approach to try with our isolated bio-particles. We therefore developed a short pre-amplification reaction using 10-14 amplification cycles and the unlabeled LCE1C primers (same primers used in the CE body fluid identification multiplex). With this pre-amplification step, we were able to detect LCE1C in these serial dilution samples (previously no LCE1C was detected without prior pre-amplification).

With the success of the pre-amplification method, we employed this approach to bio-particles collected from touch samples using our developed micro-manipulation protocol. We collected 10, 25, 50, 100 and 0 bio-particle samples from various clothing items including a sweater sleeve, a sock, a shirt collar, a shirt (back side) and a lab coat sleeve. A 5µl reduced volume RT reaction was performed. The entire 5µl RT product was then added to the pre-amplification reaction (25µl). A 5µl aliquot of the pre-amplification reaction was then used in the standard body fluid identification multiplex (the standard reaction volume was used to allow for an increased amount of the pre-amplification product to be used if needed). Using this approach we were able to successfully detect LCE1C in as few 10 and 25 bio-particle samples (Figure 36).

As these methods were being developed for skin (and successful results were being obtained), we also were performing related experiments with buccal and vaginal cells. We had initially tried standard lysis and reverse transcription experiments with micro-dissected buccal and vaginal cells (1-100 cells collected) with no successful results. Since we had lysate available from these experiments, we tried pre-amplification reactions as that approach was developed. There was still limited success even with pre-amplification. We then incorporated the reduced volume RT reactions and again had limited success. However, all experiments through this point included a simple heating step for lysis. While this seemed to be sufficient for touch particles, buccal and vaginal cells may be more structurally intact and therefore might require additional lysis strategies beyond simple heating. Since we had previous success with the *RNA*GEM lysis, we were hopeful that it could be incorporated into the developed reduced volume and pre-amplification reactions in order to provide the additional sensitivity that we needed. Our DNA profiling work involved the use of a combined, single-tube lysis/amplification and therefore we

wanted to try a similar approach for RNA profiling as well. We incorporated the *RNA*GEM lysis into a reduced volume 5µl RT reaction. Since we were using this reduced volume reaction, cells were collected using micromanipulation. The cells (10, 25, 50 and 100 cells) were collected into the 5µl reaction consisting of lysis and reverse transcription reagents. The standard RT program was used and then the entire 5µl cDNA product was used for subsequent HTN3 pre-amplification. We analyzed the pre-amplification samples using our standard RNA body fluid identification multiplex (CE) and obtained only a weak HTN3 result for the 25 cell sample. We next evaluated the use of an alternative detection method that we have also developed in our laboratory involving high-resolution melt (HRM) analysis. This approach utilizes differing melt temperatures of the mRNA biomarker amplification products, each with their own unique melt temperature. We have recently developed a 6plex HRM multiplex assay that permits identification of all forensically relevant body fluids (manuscript submitted). This method is more rapid and inexpensive than traditional CE detection (~2 hours to complete analysis after reverse transcription and involves the use of unlabeled primers for detection). These assays are currently performed on a RotorGene Q real time PCR system that has high resolution melt capabilities. We therefore used an HTN3 singleplex HRM reaction and we were able to successfully detect HTN3 with as few as 25 cells (Figure 37). We next repeated this analysis with vaginal cells (10, 25, 50 and 100 cells) using several vaginal secretions biomarkers – IL19 and MYOZ1, plus an additional two unpublished vaginal biomarkers (gene names not reported here; will be the subject of an upcoming publication). Using all four biomarkers, we successfully detected the biomarker in all four samples (10, 25, 50 and 100 cells) (Figure 38, IL19 results shown). Nothing was detected in the 0 cell samples for either the saliva or vaginal assays. As can be seen from this figure, the Tm values for the cell samples are slightly higher than the expected temperature. This occurred for all of the biomarkers tested. The source of this shift is being evaluated, but is not of immediate concern since the cell samples could be influenced by the differing reagents used for their isolation compared to the standard RNA extracts.

The results of these experiments demonstrate (for the first time!) the ability to successfully use RNA profiling to identify the body fluid or tissue source of origin of dried epithelial cell containing fluids/tissues – a major objective of this project and one that has proven to be quite challenging. We are currently further evaluating this approach and also working on multiplex pre-amplification and HRM assays for epithelial cells so that single assays will permit an identification of vaginal secretions, buccal cells and skin. Additional optimization of these developed assays will be performed in order to determine if biomarker detection is possible from < 10 cells. Currently, the use of > 1 epithelial cell may result in the detection of multiple body fluids for an individual sample (if admixed body fluid mixtures are present, e.g. saliva/vaginal). However, this is not a significant concern since it will still permit the determination of the body fluids present in the admixed samples. We are hopeful that with the improved sensitivity afforded by the use of the pre-amplification and HRM assays that we will be able to improve assay sensitivity in order to determine the cell type of origin from individual cells.

3. Initial Development of DNA/RNA Co-Isolation Methods

3.1 RNA Hybridization Probes for Cell Type Identification

**Note – This approach initially began as a collaboration with Rafal Wierzchoslawski (Agencja Bezpieczenstwa Wewnetrznego, Poland). However, that collaboration could not be continued and any work, including probe and method development, described below was performed solely by our laboratory.*

A goal of this project is to be able to determine the tissue source of origin of epithelial cells in admixed body fluid samples and obtain a genetic profile of the donor of the biological material. While we successfully developed suitable RNA profiling strategies for use with isolated epithelial cells (skin, buccal and vaginal), these were RNA-only profiling strategies and would require separate samplings to be collected from individual samples in order to perform both DNA and RNA profiling assays. However, co-isolation of DNA and RNA from single or few cells will be challenging due to sensitivity limitations of the profiling and detection methodologies. Ideally if we could identify the cell type prior to collection, this would allow us to determine which body fluids where present (not necessarily the number of different donors within that body fluid however) and allow us to collect cells from each body fluid separately. The collected cells could then be used solely for DNA analysis since the cell type would have been determined visually prior to collection. We began the development of such an approach using custom Stellaris® RNA FISH (fluorescence in situ hybridization) probes (Biosearch Technologies, Petaluma. CA). Using this approach, multiple contiguous fluorescently labeled probes that bind along a target mRNA sequence are used in order to produce sufficient signal to be detected using fluorescence microscopy. Software is available that designs suitable probe sets for the desired target. We began our analysis with a vaginal mRNA marker, CYP2B7P1. This vaginal biomarker has demonstrated the highest specificity for vaginal secretions (as well as a high abundance in vaginal secretions) in our CE based assays and therefore we decided to select this marker for initial testing with the RNA hybridization probes. The gene sequence was sufficiently long to permit a suitable number of probes to be designed along its sequence.

The manufacturer provides recommended protocols, although none of them specifically designed for dried forensic samples. There were two approaches that we could follow – one for tissue samples and one for cells in solution. Since our standard LCM analysis is performed using cells that have been heat fixed to the Leica proprietary membrane slides, we decided to continue to use this sample preparation protocol and then follow the tissue sample protocol. The protocol recommends testing three different final probe concentrations (50nM, 250nM, and 1.25μM) for initial experiments in order to determine the most suitable probe concentration. The protocol also recommends hybridization overnight at 37^{o}C (although some protocols say that this can be completed in 4 hours rather than overnight). For all initial experiments we utilized an overnight incubation in an incubator so it would be protected from light. The samples needed to be placed in a humidified chamber. This was accomplished by using a plastic petri dish with a wet Kim-Wipe in the bottom of the dish. The probe set (varying volumes to obtain the desired concentrations) was added to the hybridization buffer (instructions for preparation of this buffer were provided by the manufacturer) and 100μl of this solution was placed over the sample and a cover slip added. The slides were then placed in the petri dishes and incubated overnight at 37^{o}C. The slides were washed (30 minutes at 37^{o}C) with the prepared wash buffer (instructions on preparation provided by the manufacturer). The slides were then viewed on the Leica LMD system using the fluorescence option. All of the images were very grainy and high resolution images were difficult to obtain. We tried various options for improving the resolution. We tried cleaning and alignment of our LMD system and an improvement in image quality was obtained. However, as we proceeded with our testing, we were not seeing the expected results from the probes (small fluorescent "dots" within individual cells to indicate the presence of the mRNA. We tried other protocols including the cell suspension protocol and also tested the probes against buccal cells and blank Leica slides. Overall on all samples there was just a red "haze" but no specific results that we could identify.

We are now in the process of trying to optimize this protocol. There are several issues that may be causing an image failure, including the following: 1) inefficient or absent probe hybridization, 2) incorrect buffers, 3) CYP2B7P1 abundance or 4) technical challenges with the Leica LMD system. The manufacturer's protocol suggested a cell permeabilization step. We originally did not think this was necessary since it is likely that dried epithelial cells from forensic samples will already be to some degree permeable or "leaky" (disruptions in the cell membrane that would permit the probe to enter the cell). However, we are evaluating the use of a permeabilization step prior to probe hybridization in case this improves probe hybridization. The buffers (hybridization and wash) were prepared according to the manufacturer's protocols. They recommended specific catalog numbers for the buffer components. While we purchased a majority of the specific items listed in the protocol, we did substitute a few items since we had equivalents for these items already in the lab. While the protocol stated that substitutions can be used, it is possible that our substitutions were not compatible or suitable. Therefore, we are purchasing all of the recommended items and will re-make all hybridization and wash buffers. Additionally, all of this initial testing involved the use of only the CYP2B7P1 probe set. It is possible that this probe set is not ideal or there is not sufficient sensitivity in order to observe the expected fluorescent signal. In order to ensure that the problem is not specific to CYP2B7P1, we have also designed probe sets for additional biomarkers. We have designed probe sets for both HTN3 (saliva) and LCE1C (skin). We are hopeful that these two biomarkers will be found in high enough abundance for detection. Therefore, while at the time of this report we have not fully developed this approach, we are hopeful that continued efforts to develop a successful protocol using these probes will permit cell type identification prior to DNA analysis.

3.2 DNA/RNA Co-Isolation: AllPrep Micro kit (QIAGEN)

In other work in our laboratory, we have been utilizing the AllPrep Micro co-extraction kit, which is a semi-automated protocol (QIAGEN, QIAcube protocol) for the co-isolation of DNA and RNA. The manufacturer reports that this kit is suitable for use with micro-dissected samples including cells. While this was not specifically developed for single or few epithelial cells from dried forensic samples, we decided to evaluate the use of this kit with micro-dissected cells. Again, we began our evaluation using buccal epithelial cells. Micro-dissected buccal cells (25, 50, 100 and 0 cells) were collected into 20µl of RLT Plus buffer (provided in the kit) (β-ME added after collection). The samples were then placed on the QIACube robot for lysis and DNA/RNA isolation. Our lysis step involved incubation with the RLT Plus buffer for 1 hour at 56ºC. Following this incubation, the standard AllPrep Micro DNA and RNA isolation protocol was followed. For this initial experiment we collected two sets of buccal epithelial cells so that carrier RNA could be added to one of the sets. The use of carrier RNA is recommended if less than 500 cells are used. Therefore, we evaluated its use with one of the sample sets to determine if would improve DNA/RNA recovering and typing success. For the RNA fractions, a 14µl elution volume was used (the minimum volume allowed). The reverse transcription reaction also permits the addition of the full 14µl since the reaction volume for the master mix is only 5.8µl of a 20µl reaction. The samples were reverse transcribed using a standard RT reaction. The resulting cDNA was then used in the previously developed RNA pre-amplification method. A portion of the pre-amplification product was then used for amplification with our current body fluid multiplex. We also amplified a portion of cDNA that had not been pre-amplified. However, saliva mRNA biomarkers (HTN3, STATH) were not detected in any of the samples. We next

evaluated the use of the alternative detection method involving high-resolution melt (HRM) analysis (previously described). We performed singleplex HRM assays using both of our saliva mRNA biomarkers: HTN3 and STATH. While neither biomarker was detected in 100% of the samples, HTN3 was detected in a 100 cell sample (no carrier RNA and no pre-amplification) and STATH was detected in a 50 cell sample with and without pre-amplification. These results were quite surprising since they were obtained using a standard extraction kit and represented an improvement over our standard mRNA profiling method involving our CE body fluid multiplex system. Since the extraction used was a DNA/RNA co-isolation method we also had DNA fractions from each of these samples. We were able to obtain DNA profiles (partial) from the 100 cell sample in which HTN3 was detected and from the 50 cell sample in which STATH was detected. The recovery of the STR profiles, although partial, was significant since this was the first time we were able to obtain DNA and RNA profiling results from the same cell sample.

In summary, while neither DNA/RNA co-isolation approach has been fully developed, these initial experiments demonstrate the potential future development of suitable DNA/RNA co-isolation strategies for single or few isolated epithelial cells (the ultimate goal of the current work). We therefore can continue to optimize these approaches in order to determine if we can improve sensitivity and therefore increase the success of DNA and RNA profiling success for isolated epithelial cells.

B. TABLES

Table 1. Genotypes Recovered From Artificially Created 15-Cell Mixtures

controls	D3		vWA		FGA		AMEL		D8		D21		D18		D5		D13		D7	
male donor	16	16	14	18	19	21	X	Y	10	14	29	31.2	18	19	10	11	12	12	10	11
female donor	16	18	17	18	20	24	X	X	13	14	28	33.2	15	15	11	12	9	12	10	10

No. cells (Female-Male)	D3	vWA	FGA	AMEL	D8	D21	D18	D5	D13	D7
15-0 (donor 1)	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
14-1	16,18	14,17,18	20,21,24	X	10,13,14	28,29,31.2,33.2	15	10,11,12	9,12	10
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2	16	14	21	X	10,14	29,31.2		10,11	12	10
13-2	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor1	16,18	17,18	20,24	X	13,14	28,33.2	14	11,12	9,12	10
12-3	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
11-4	16,18	14,17,18	19,20,21,24	X,Y	10,13,14	28,29,31.2,33.2	15,18,19	10,11,12	9,12	10,11
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2	16	14,18	19.21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
10-5	16,18	14,17,18	19,20,21,24	X,Y	10,13,14	28,29,31.2,33.2	15	10,11,12	9,12	10,11
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2	16	14,18	19,21	X,Y	10,14	29,31.2		10,11		10,11
9-6	16,18	14,17,18	19,20,21,24	X,Y	10,13,14	28,29,31.2,33.2	15,18,19	10,11,12	9,12	10,11
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2	16	14	19,21	X,Y	10,14	29,31.2	18,19	10	12	10,11
8-7	16	14,18	19	X,Y	10,14	29,31.2		10,11	12	10,11
donor2	16	14,18	19	X,Y	10,14	29,31.2		10,11	12	10,11
7-8	16,18	14,17,18	19,20,21,24	X,Y	10,13,14	28,29,31.2,33.2	15,18,19	10,11,12	9,12	10,11
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2		14,18	19.21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
6-9	16,18	14,17,18	19,20,21,24	X,Y	10,13,14	28,29,31.2,33.2	15,18,19	10,11,12	9,12	10,11
donor1	16,18	17,18	20,24	X	13,14	28,33.2	15	11,12	9,12	10
donor2	16	14,18	19.21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
5-10	16,18	14,17,18	19,20,21	X,Y	10,13,14	29,31.2,33.2	18,19	10,11,12	12	10,11
donor1	16,18	17,18	20,24	X	13,14	33.2		12		10
donor2	16	14,18	19	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
4-11	16	14,17,18	19,21,24	X,Y	10,14	29,31.2	18,19	10,11	9,12	10,11
donor1		17	24	X					9,12	10
donor2	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
3-12	16,18	14,17,18	19,21,24	X,Y	10,14	28,29,31.2	18,19	10,11,12	12	10,11
donor1	16,18	17,18	24	X	14	28		12		10
donor2	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
2-13	16,18	14,17,18	19,21,24	X,Y	10,14	29,31.2,33.2	18,19	10,11	9,12	10,11
donor1	16,18	17	24	X		33.2			9,12	10
donor2	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
1-14	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
0-15 (donor 2)	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11

single source male
single source female
mixture

Table 2. De-Convolution of Two-Donor Mixture Using Ten 15-Cell Samples

controls	D3		vWA		FGA		AMEL		D8		D21		D18		D5		D13		D7	
male donor	16	16	14	18	19	21	X	Y	10	14	29	31.2	18	19	10	11	12	12	10	11
female donor	14	16	16	18	22	23	X	X	12	12	29	30	15	15	11	11	9	12	12	12

No. cells (Female-Male)	D3	vWA	FGA	AMEL	D8	D21	D18	D5	D13	D7
15-cell (1)	14,16	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18	22,23	X,X	12	29,30	15	11	9 or 9,12	12
15-cell (2)	14,16,17,19	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	?	16,18	22,23	X,X	12	29,30	15	11	9 or 9,12	12
15-cell (3)	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor1										
donor2										
15-cell (4)	14,16	14,16,18	21,22,23	X,Y	10,12,14	29,30,31.2	15	10,11	9,12	11,12
donor1	16	14,18	*,21	X,Y	10,14	29,31.2	*,*	10,11	12	*,11
donor2	14 or 14,16	16,18	22,23	X,X	12	29,30	15	11	9	12
15-cell (5)	14,16	14,16,18	19,21	X,Y	10,12,14	29,31.2	18,19	10,11	12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18		X,X	12 or 10,12			11		12 or 10,12
15-cell (6)	16	14,16,18	23	X,Y	10,12,14	29,30	15	10,11	9,12	10,11,12
donor1	16	14,18	*,*	X,Y	10,14	29,*	*,*	10,11	12	10,11
donor2		16,18	23	X,X	12	30	15	11	9 or 9,12	12
15-cell (7)	14,16	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18	22,23	X,X	12	29,30	15	11	9 or 9,12	12
15-cell (8)	14,16	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29.31,2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18	22,23	X,X	10,12	29,30	15	11	9 or 9,12	12
15-cell (9)	14,16	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18	22,23	X,X	12	29,30	15	11	9 or 9,12	12
15-cell (10)	14,16	14,16,18	19,21,22,23	X,Y	10,12,14	29,30,31.2	15,18,19	10,11	9,12	10,11,12
donor1	16	14,18	19,21	X,Y	10,14	29,31.2	18,19	10,11	12	10,11
donor2	14 or 14,16	16,18	22,23	X,X	12	29,30	15	11	9 or 9,12	12

Male = donor 1
Feale = donor 2

Table 3. DNA Profile Recovery from Single Source Micro-Dissected Epithelial Cells (Identifiler, 34 cycles)

Number of Micro-dissected Cells	Avg. % Partial/Full Profile	Avg. % Full Profiles	Avg. % No Profile
1	74 (± 5%)	4 (± 2%)	26 (± 5%)
2	69 (± 5%)	19 (± 8%)	30 (± 5%)
3	81 (± 3%)	20 (± 5%)	20 (± 5%)
4	81 (± 3%)	39 (± 6%)	19 (± 3%)
5	81 (± 3%)	46 (± 7%)	18 (± 4%)
10	90 (± 3%)	66 (± 6%)	11 (± 3%)
20	99 (± 1%)	80 (± 6%)	1 (± 1%)

*Std error values listed in parentheses next to each percentage
**Average % profile recovery determined by averaging the success rates (number of profiles recovered from 10 replicates) of the 10 individual donors.

Table 4. Identifiler® Reference Genotypes of the Donors in the Epithelial Cell Mixtures

Locus	M1	F1	F2	F3
D8S1179	10,14	12,12	13,13	14,15
D21S11	29,31.2	29,30	29,31.2	27,29
D7S820	10,11	12,12	8,11	8,9
CSF1PO	10,10	11,12	10,11	10,12
D3S1358	16,16	14,16	16,18	16,16
TH01	6,9	6,9.3	6,9.3	7,9.3
D13S317	12,12	9,12	10,12	12,12
D16S539	9,13	11,12	12,13	11,11
D2S1338	20,25	17,23	21,25	17,20
D19S433	12,13	14,15	13,13	13,15
vWA	14,18	16,18	16,18	14,17
TPOX	8,11	8,11	8,11	8,11
D18S51	18,19	15,15	12,14	12,15
AMEL	X,Y	X	X	X
D5S818	10,11	11,11	9,12	12,13
FGA	19,21	22,23	20,23	20,22

Table 5. Information statistics for single and pairwise TrueAllele® computer interpretation of 37 LCM LCN 20 cell samples, measured in log_{10}(LR) ban units.

Information (ban)	Single	Pairwise
mean	11.732	17.218
standard deviation	3.996	2.386
minimum	1.764	10.699
maximum	18.488	20.722

*Information for a sample's pairwise combinations is the maximum over its group of 9 pairs.

Table 6. Information statistics for TrueAllele® computer interpretation of LCM LCN 20 cell sample 6, both singly and in pairwise combination with the other 9 samples, measured in log_{10}(LR) ban units.

Sample 1	Sample 2	Mixture	Run 1	Run 2	Difference	Average	Change
6		67.5%	11.815	12.441	0.626	12.128	0.000
6	1	21.0%	17.637	17.666	0.029	17.652	5.524
6	2	57.0%	15.187	15.270	0.083	15.229	3.101
6	3	54.0%	16.793	16.385	-0.408	16.589	4.461
6	4	20.5%	17.646	17.615	-0.031	17.631	5.503
6	5	39.0%	13.931	13.574	-0.357	13.753	1.625
6	7	35.0%	14.112	15.293	1.181	14.703	2.575
6	8	49.5%	8.868	9.177	0.309	9.023	-3.106
6	9	46.0%	15.683	15.374	-0.309	15.529	3.401
6	10	19.5%	16.881	17.489	0.608	17.185	5.057
					minimum	9.023	-3.106
					mean	15.255	3.127
					median	15.529	3.401
					maximum	17.652	5.524

Table 7. Average weight ratios for 50pg Replicates of a 1:1 male-female mixture (limiting dilutions)

Mixture amount	Four Allele Loci	
	Weight Avg	St Dev
1ng	0.55	0.06
50pg-1	0.67	0.07
50pg-2	0.57	0.23
50pg-3	NA	NA
50pg-4	0.58	0.04
50pg-5	0.63	0.16
50pg-6	0.52	0.07
50pg-7	0.57	0.17
50pg-8	0.52	0.17
50pg-9	0.41	0.07
50pg-10	0.56	0.08
50pg-11	0.64	0.06
50pg-12	0.56	0.10
50pg-13	0.57	0.08
50pg-14	0.63	0.10
50pg-15	0.52	0.08
50pg-16	0.61	0.02
50pg-17	0.51	0.07
50pg-18	0.54	0.10
50pg-19	0.50	0.09
50pg-20	0.62	0.09

Average	0.56
Std err	0.01
Male/Female max	.67/.33
Female/Male max	.59/.41

Table 8. RNA Recovery from Cell Suspension Dilution Series Using a Standard Manual Organic RNA Extraction

SALIVA			
No. Cells	Quantity (ng/ul)	Total ng (20ul)	ng into RT
8750	6.4	128.0	51.2
4375	1.3	26.0	10.4
2188	4.5	90.0	36.0
1094	2.3	46.0	18.4
547	3.7	74.0	29.6
273	2.5	50.0	20.0
137	6.2	124.0	49.6
68	1.0	20.0	8.0
34	0.5	10.0	4.0
17	5.2	104.0	41.6
9	5.7	114.0	45.6
4	3.0	60.0	24.0
2	1.9	38.0	15.2
1	2.3	46.0	18.4

VAGINAL			
No. Cells	Quantity (ng/ul)	Total ng (20ul)	ng into RT
15050	6	120	48
7525	4.6	92	36.8
3763	6.4	128	51.2
1881	1.6	32	12.8
941	0.9	18	7.2
470	0.7	14	5.6
235	3	60	24
118	1.3	26	10.4
59	3.9	78	31.2
29	1	20	8
15	0.6	12	4.8
7	1.4	28	11.2
4	2.8	56	22.4
2	3.9	78	31.2
1	1.6	32	12.8

Table 9. Comparison of RNA Recovery and Typeability from Liquid and Dried Saliva Samples (Standard Organic RNA Extraction)

Donor	µl saliva	Form	B2M	HTN3	HTN1
1	1	Liquid	9120	9083	6423
		Dry	8906	9074	9199
	5	Liquid	8920	8888	8706
		Dry	8886	9602	9426
	10	Liquid	8576	8416	9382
		Dry	782		454
	25	Liquid	8384	9152	5749
		Dry	8557	9411	1204
	50	Liquid	9011	8721	8518
		Dry	8610	8770	8699
	100	Liquid	8638	8187	2420
		Dry	8136	8324	3487
Donor	µl saliva	Form	B2M	HTN3	HTN1
2	1	Liquid	1088	9310	4921
		Dry	910		
	5	Liquid	1196	9620	5133
		Dry	9005	8800	6881
	10	Liquid	8941	9122	6833
		Dry	9007	8818	9344
	25	Liquid	9032	8459	9603
		Dry	8887	8725	8940
	50	Liquid	9022	8970	8891
		Dry	57	7289	9257
	100	Liquid	8993	8652	9504
		Dry	9096	8678	9463

*Numbers listed in table refer to signal intensity (relative fluorescence units (RFUs)

Table 10. Detection of RNA in cells vs. supernatant fractions

Source	Sample	B2M	HTN3	HTN1
liquid (15µl)	cells	8730	6526	1105
	supernatant	742		
	wash 1	8914	9124	3570
	wash 2	8774	9643	883
	wash 5	5673	7156	539
	wash 10	1260	1640	398
Source	**Sample**	**B2M**	**HTN3**	**HTN1**
dried (15µl)	cells	4216	3651	787
	supernatant	5965	5090	349
	wash 1		535	568
	wash 2	388	203	
	wash 5	323		
	wash 10			

*Numbers listed in table refer to signal intensity (relative fluorescence units (RFUs)

Table 11. RNA-Seq Results (MaxCount Values) for Six VS mRNA Candidates

Gene	Abbrev	Vag 1-1	Vag1-2	Vag2-1	Vag2-2	Skin 1-1	Skin 1-2	Skin 2-1	Skin 2-2
Surfactant Associated 2	SFTA2	321	512	699	517	0	0	1	0
Fucosyltransferase 6	FUT6	371	753	744	491	5	8	72	34
Dickkopf homolog 4	DKK4	13	24	32	19	0	2	1	0
Interleukin 19	IL19	5	6	6	4	0	0	0	0
Myozenin 1	MYOZ1	6	15	30	28	2	3	3	4
Cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1	CYP2B7P1	934	1039	915	1065	27	19	3	4

Table 12. List of VS Candidate Genes and Primer Sequences

Gene	Accession Number	Product Size (bp)		Primer Sequence (5'-3')
		RNA	DNA	
SFTA2	NM_205854	183	422	F*: FAMCGGGACTTCAGTGTCTCCTC R: TTCCAGGAAGCTGGACTCAT
FUT6	NM_000150	162	3711	F*: FAMCCTCTTCCACGTCCTCAGAA R: CTAGAGGGACTCGGGGATCT
DKK4	NM_014420	168	934	F*: FAMACGGACTGCAATACCAGAAA R: AATATTGGGGTTGCATCTTC
IL19	NM_153578	109	260	F*: PETAACCACGGTCTCAGGAGATG R: TGACATTTGGGAAGGTGTCC
MYOZ1	NM_021245	81	503	F*: FAM GGGTTGGTGAGACAGGATCA R: TCCCATGGGGAAATATAGGT
CYP2B7P1	NR_001278	198	318	F*: FAMTCCTTTCTGAGGTTCCGAGA R:TTTCCATTGGCAAAGAGCAT
B2M**	NM_004048	120	3930	F*: FAM GGCATTCCTGAAGCTGACA R:AAACCTGAATCTTTGGAGTACG

* indicates fluorescently labeled primer
** housekeeping gene

Table 13. Specificity of VS Genes

The number of samples detected out of the number of samples tested is shown. Input represents the amount of total RNA used for cDNA synthesis in the RT reaction. Dark grey squares represent detection in >75% of samples tested, light grey squares present detection in ≤50% of samples tested and white squares indicate no detection. NT = not tested.

Input	Gene	Vaginal	Menstrual	Saliva	Semen	Blood
5ng	SFTA2	4/5	1/5	5/15	0/5	0/5
	FUT6	5/5	0/5	1/15	0/5	0/5
	DKK4	4/5	1/5	3/15	0/5	0/5
	IL19	2/5	2/5	0/15	0/5	0/5
	MYOZ1	4/5	1/5	0/15	0/5	0/5
	CYP2B7P1	5/5	2/5	0/15	0/5	0/5
10ng	MYOZ1	23/25	0/5	1/15	0/5	0/5
	CYP2B7P1	25/25	4/5	0/15	0/5	0/5
25ng	MYOZ1	21/25	4/5	3/24	0/5	0/5
	CYP2B7P1	25/25	5/5	0/24	0/5	0/5
50ng	MYOZ1	NT	NT	1/2	NT	NT
	CYP2B7P1	NT	NT	0/2	NT	NT
100ng	MYOZ1	NT	NT	1/2	NT	NT
	CYP2B7P1	NT	NT	0/2	NT	NT

Table 14. Skin-Specific Gene Characteristics

Gene	Accession #	Primers (5'-3')	Size (bp)	RNA-seq – Max Count Values*			
				Skin-1	Skin-2	Vag-1	Vag-2
LCE1C	NM_178351	F: GCTGAAGGACCCTGTGCT R: CAGGACATCTTGGTGGCG	56/58	842	1703	8	51
LCE1D	NM_178352	F: CCTGTGCTGCCTGTGACT R: GGCACTTAGGGGGACATTTA	142	166	102	1	2
LCE2D	NM_178430	F: TCTGTGCTTTTGCATGTGAC R: GGACCACAGCAGGAAGAGAC	193	194	428	90	127
CCL27	NM_006664	F: AGCACTGCCTGCTGTACTCA R: TTCAGCCCATTTTCCTTAGC	254	964	1897	0	7
IL1F7	NM_173203	F: CCAGTGCTGCTTAGAAGACC R: TCACCTTTGGACTTGTGTGAA	92	93	421	0	2

*Max count values for each donor (1 and 2) are an average of two replicates

Table 15. Skin biomarker sensitivity

Shading indicates the average RFUs obtained for each biomarker (dark grey >1000 RFUs; medium grey 500-999 RFUs; light grey 51-499 RFUs; white <50 RFUs). The number of samples in which the biomarker was detected out of the total number of samples tested is indicated. Average RFU values were determined using only the samples in which the marker was detected.

RNA input (ng)	Average Signal Intensity (RFU)				
	LCE1C	LCE1D	LCE2D	CCL27	IL1F7
25	NT	7/8	6/6	6/6	6/6
10	NT	13/15	6/6	6/6	6/6
5	NT	12/15	6/6	6/6	6/6
1	NT	4/8	6/6	6/6	6/6
0.5	NT	2/8	6/6	6/6	6/6
0.25	10/11	1/8	10/13	6/6	13/13
0.1	9/11	1/8	8/13	12/13	12/13
0.05	9/11	0/8	1/13	6/13	12/13
0.025	10/11	0/8	2/13	5/13	11/13
0.015	5/7	0/8	0/13	5/13	10/13
0.005	4/7	0/8	1/13	7/13	7/13

NT = not tested

Table 16. Skin biomarker specificity

Shading indicates the average RFUs obtained for each biomarker (dark grey >1000 RFUs; medium grey 500-999 RFUs; light grey 51-499 RFUs; white <50 RFUs). The number of samples in which the biomarker was detected out of the total number of samples tested is indicated.

Body Fluid	LCE1C	LCE1D	LCE2D	CCL27	IL1F7
	25 pg	5 ng	250 pg	250 pg	250 pg
Skin	10/11	12/15	10/13	6/6	13/13
Blood	0/4	0/5	0/4	0/4	0/4
Semen	0/4	0/5	0/4	0/4	0/4
Saliva/Buccal	0/4	0/5	0/4	0/4	0/4
Vaginal Secretions	0/4	0/5	0/4	0/4	0/4
Menstrual Blood	0/4	0/5	0/4	0/4	0/4
Tissue	LCE1C	LCE1D	LCE2D	CCL27	IL1F7
	~1 ng	~1 ng	~1 ng	~1 ng	~1 ng
Adipose	0/1	0/1	0/1	0/1	0/1
Bladder	0/1	0/1	0/1	0/1	0/1
Brain	0/1	0/1	0/1	0/1	0/1
Cervix	0/1	0/1	0/1	0/1	0/1
Colon	0/1	0/1	0/1	0/1	0/1
Esophagus	0/1	0/1	0/1	0/1	0/1
Heart	0/1	0/1	0/1	0/1	0/1
Liver	0/1	0/1	0/1	0/1	0/1
Lung	0/1	0/1	0/1	0/1	0/1
Kidney	0/1	0/1	0/1	0/1	1/1
Placenta	0/1	0/1	0/1	0/1	0/1
Prostate	0/1	0/1	0/1	0/1	0/1
Ovary	0/1	0/1	0/1	0/1	0/1
Skeletal Muscle	0/1	0/1	0/1	0/1	0/1
Small Intestine	0/1	0/1	0/1	0/1	0/1
Spleen	0/1	0/1	0/1	0/1	0/1
Testes	0/1	0/1	0/1	0/1	0/1
Thymus	0/1	0/1	0/1	0/1	0/1
Thyroid	0/1	0/1	0/1	1/1	0/1
Trachea	0/1	0/1	0/1	0/1	0/1

C. FIGURES

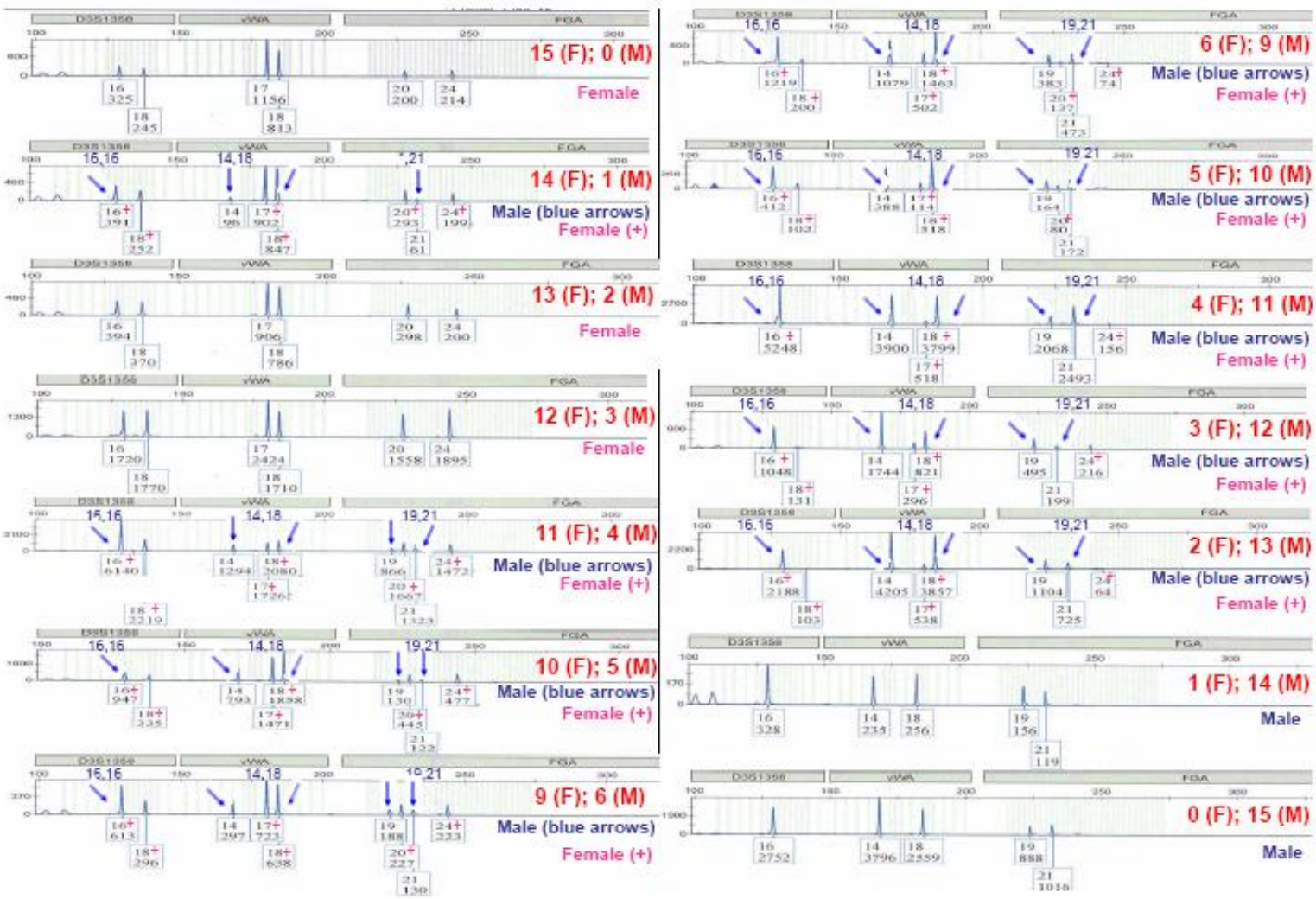


Figure 1. Profiles (FAM, Profiler Plus) Recovered From Artificially Created 15-Cell Mixtures

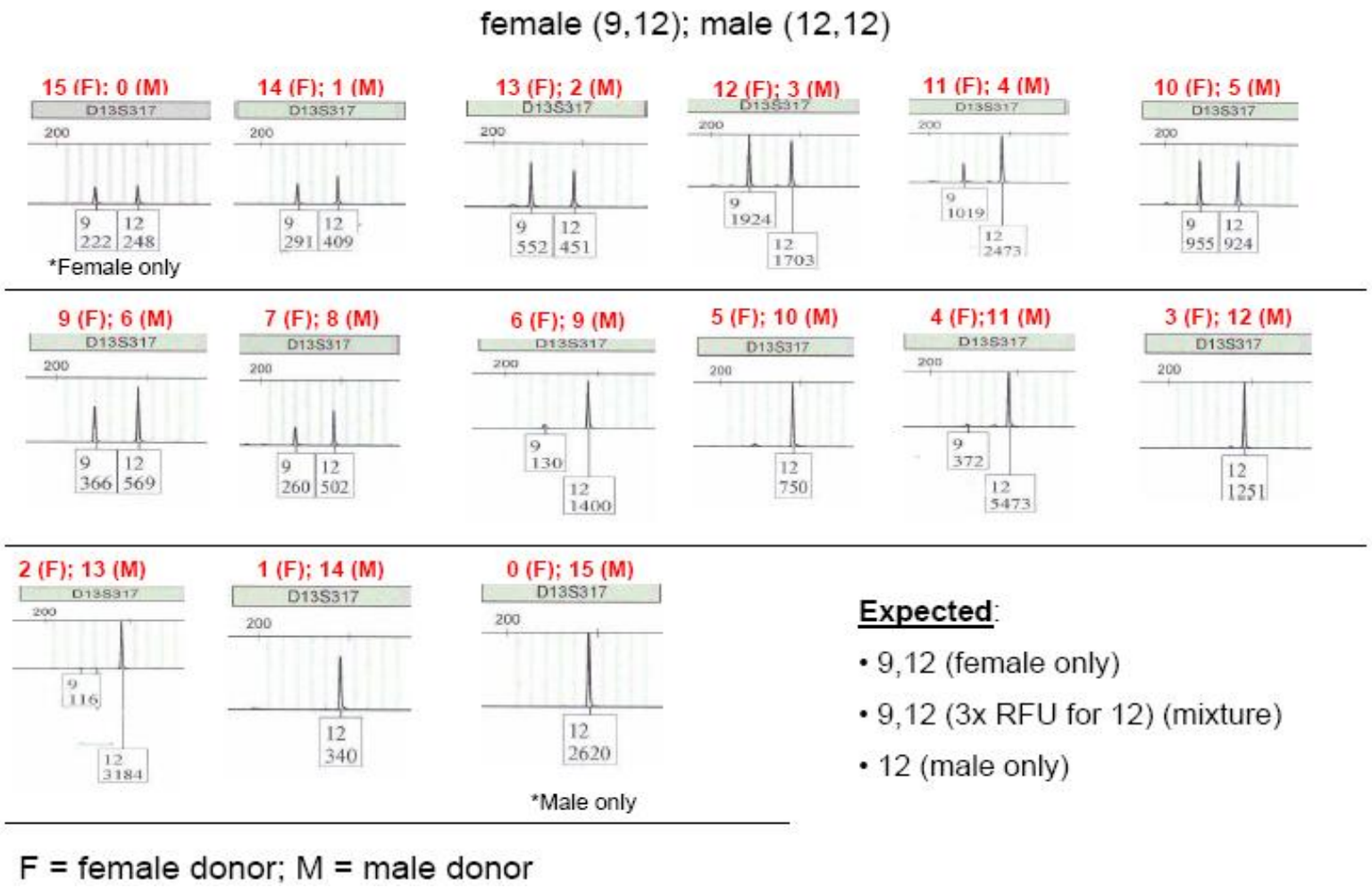


Figure 2. Genotypes Recovered from a Two-Allele Locus in a Two-Donor 15-Cell Mixture

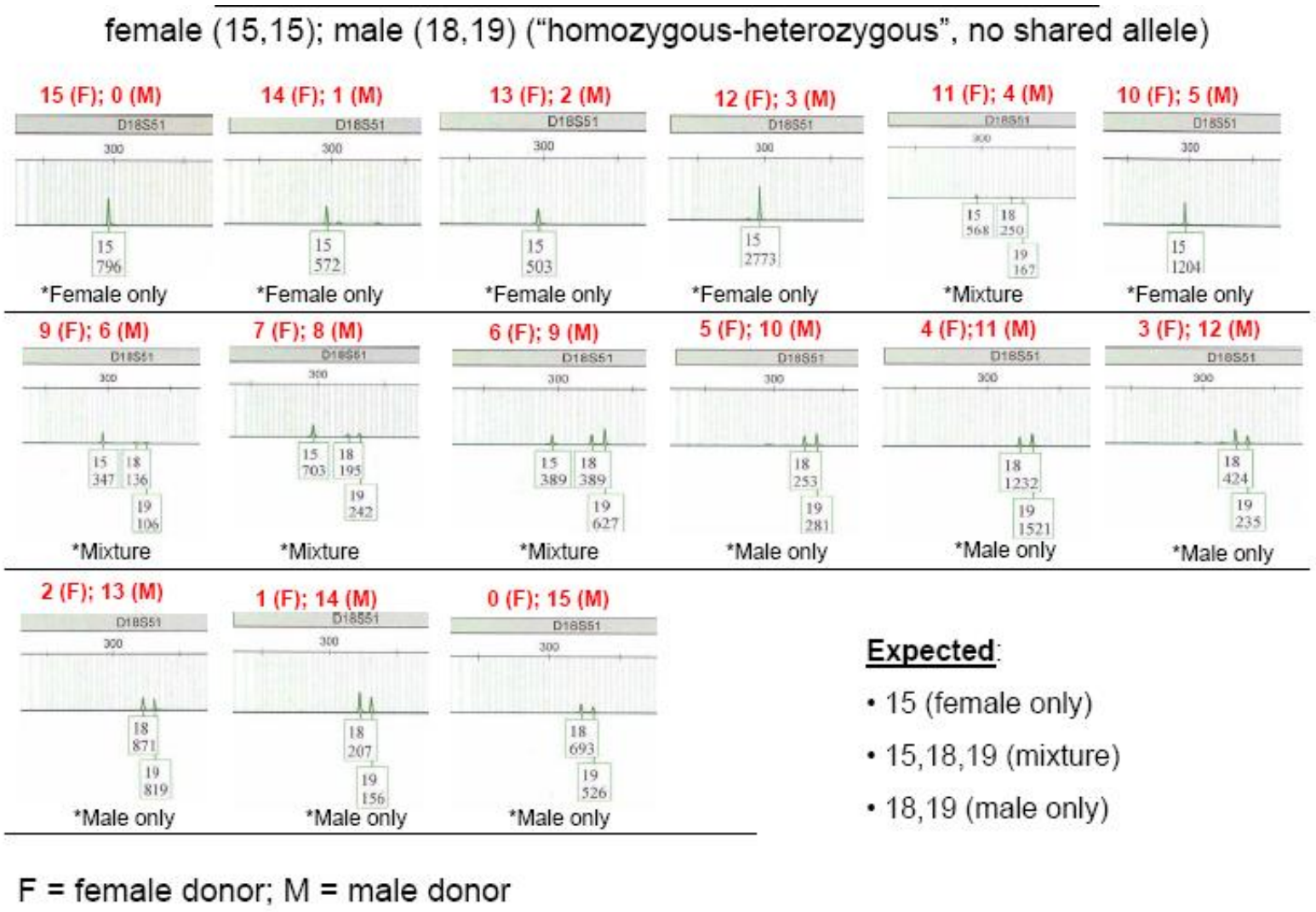


Figure 3. Genotypes Recovered from a Three-Allele (Non-Shared) Locus in a Two-Donor 15-Cell Mixture

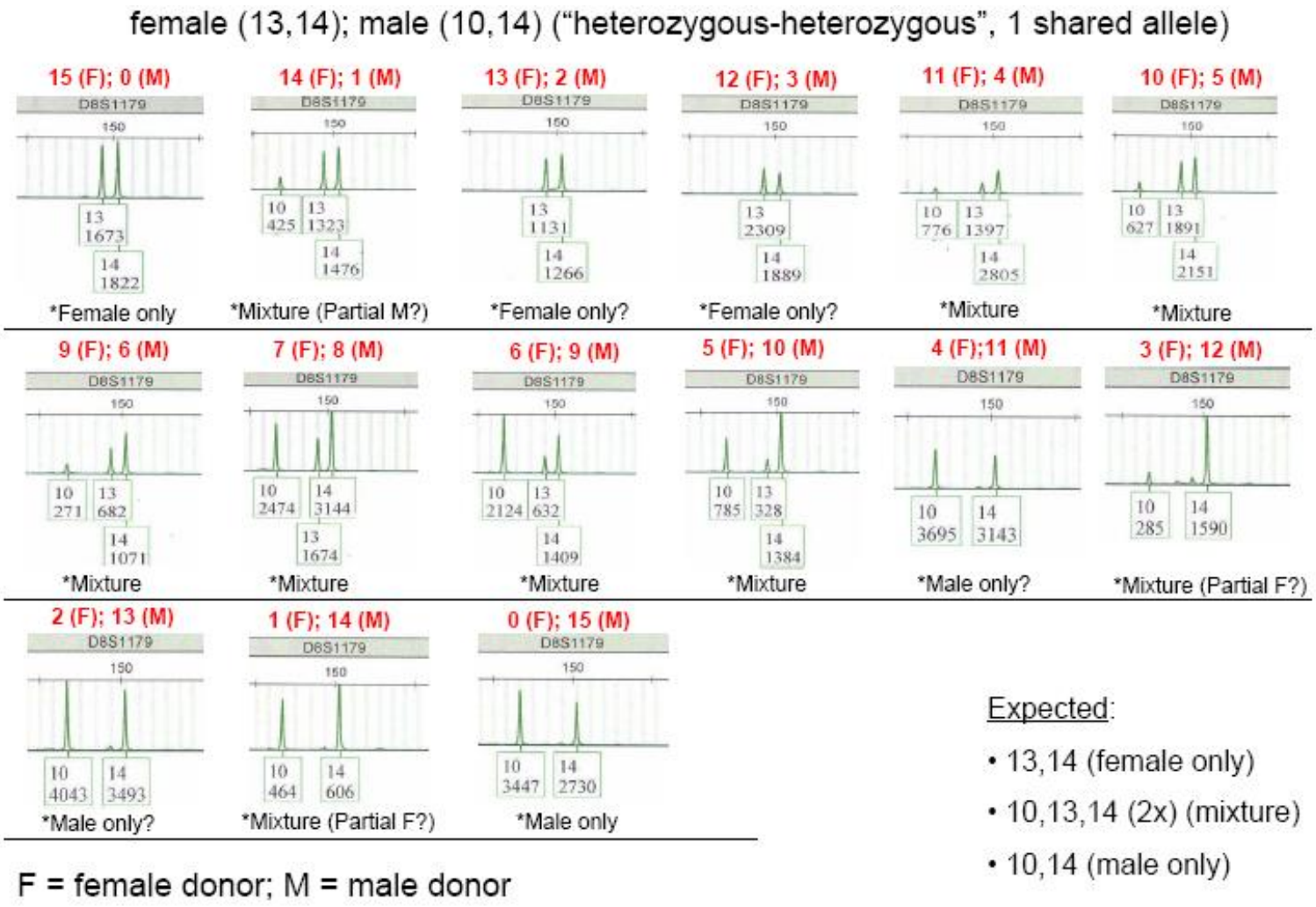


Figure 4. Genotypes Recovered from a Three-Allele Locus (One Shared Allele) in a Two-Donor 15-Cell Mixture

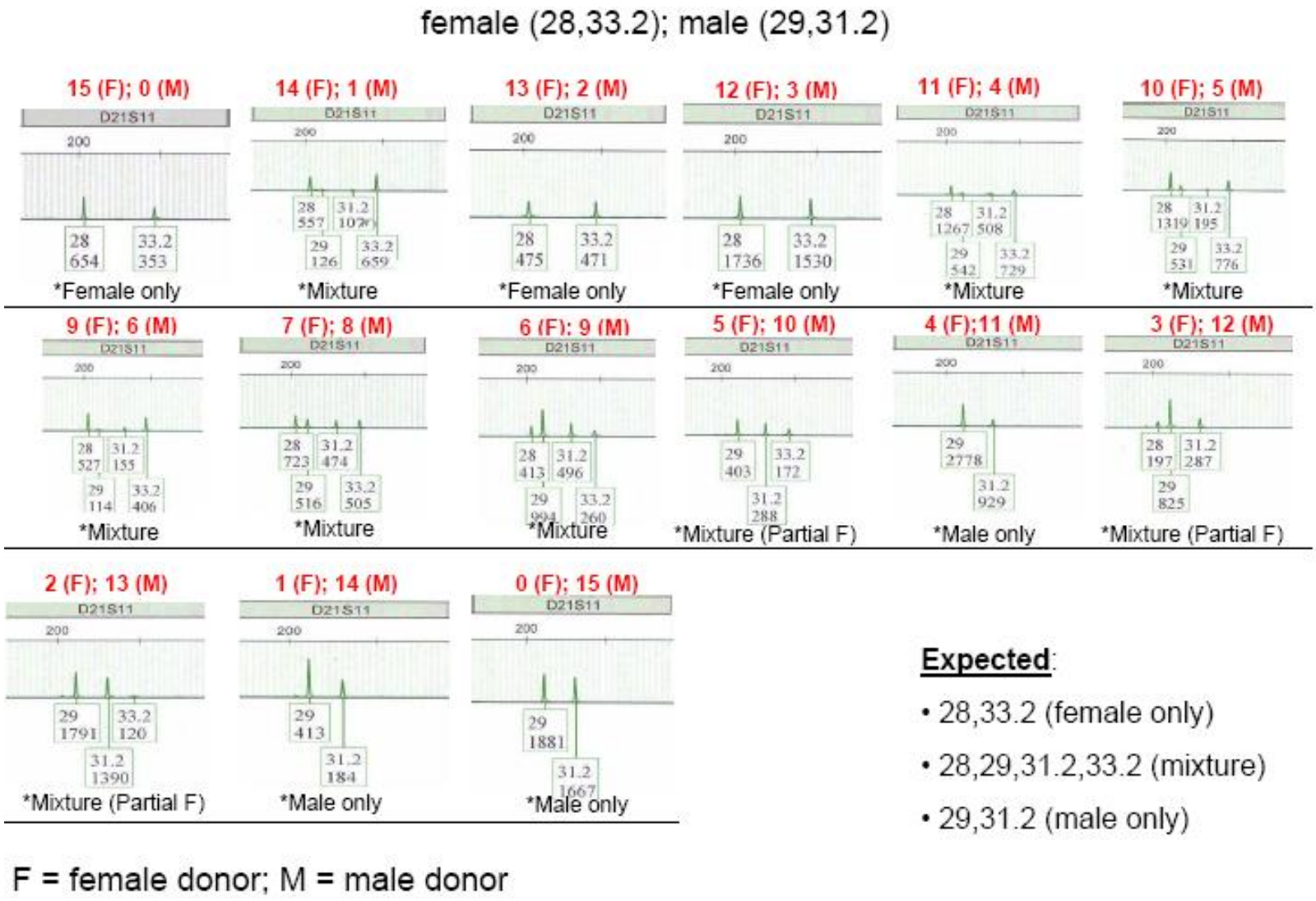


Figure 5. Genotypes Recovered from a Four-Allele Locus in a Two-Donor 15-Cell Mixture

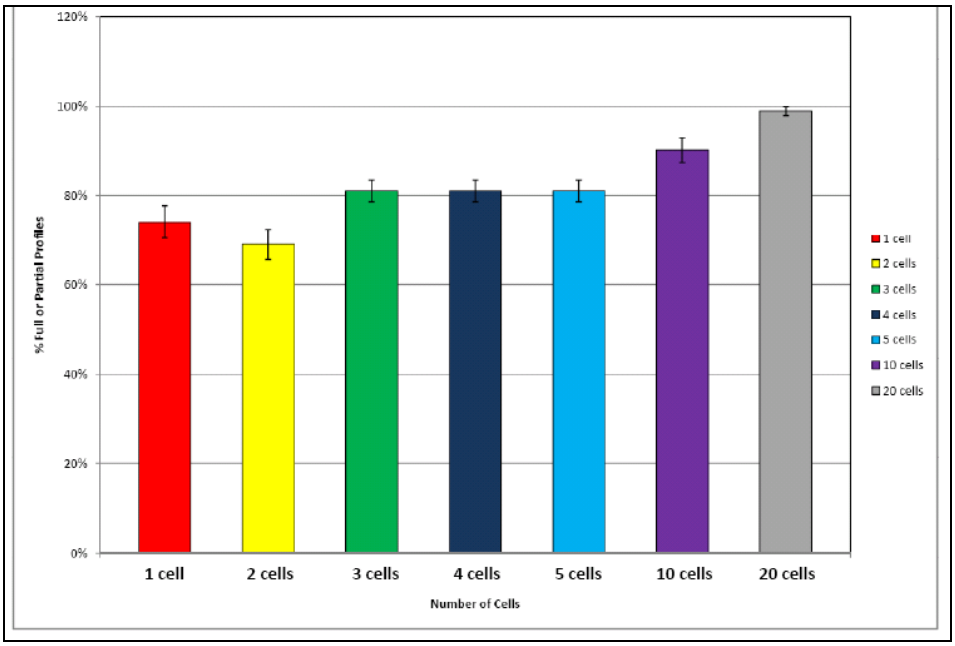


Figure 6. Average Success Rate – Partial or Full Profiles (Identifiler, 34 cycles) From Single Source Buccal Cells (1-20 cells, 10 replicates each – average of 10 donors)

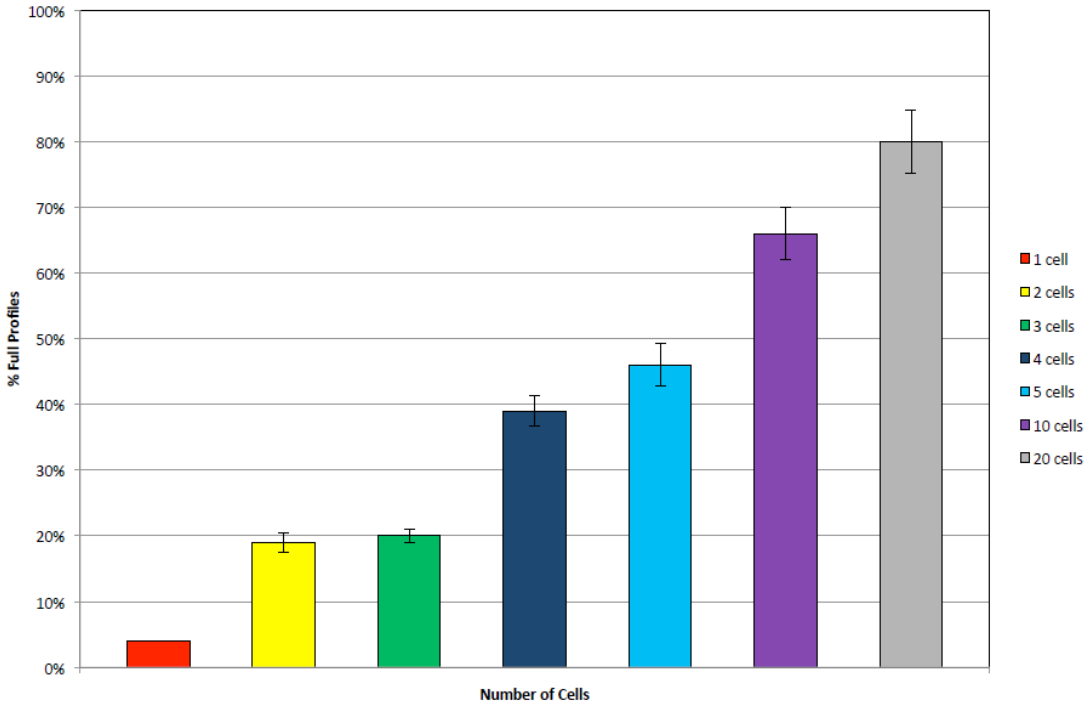


Figure 7. Average Success Rate – Full Profiles (Identifiler, 34 cycles) From Single Source Buccal Cells (1-20 cells, 10 replicates each – average of 10 donors)

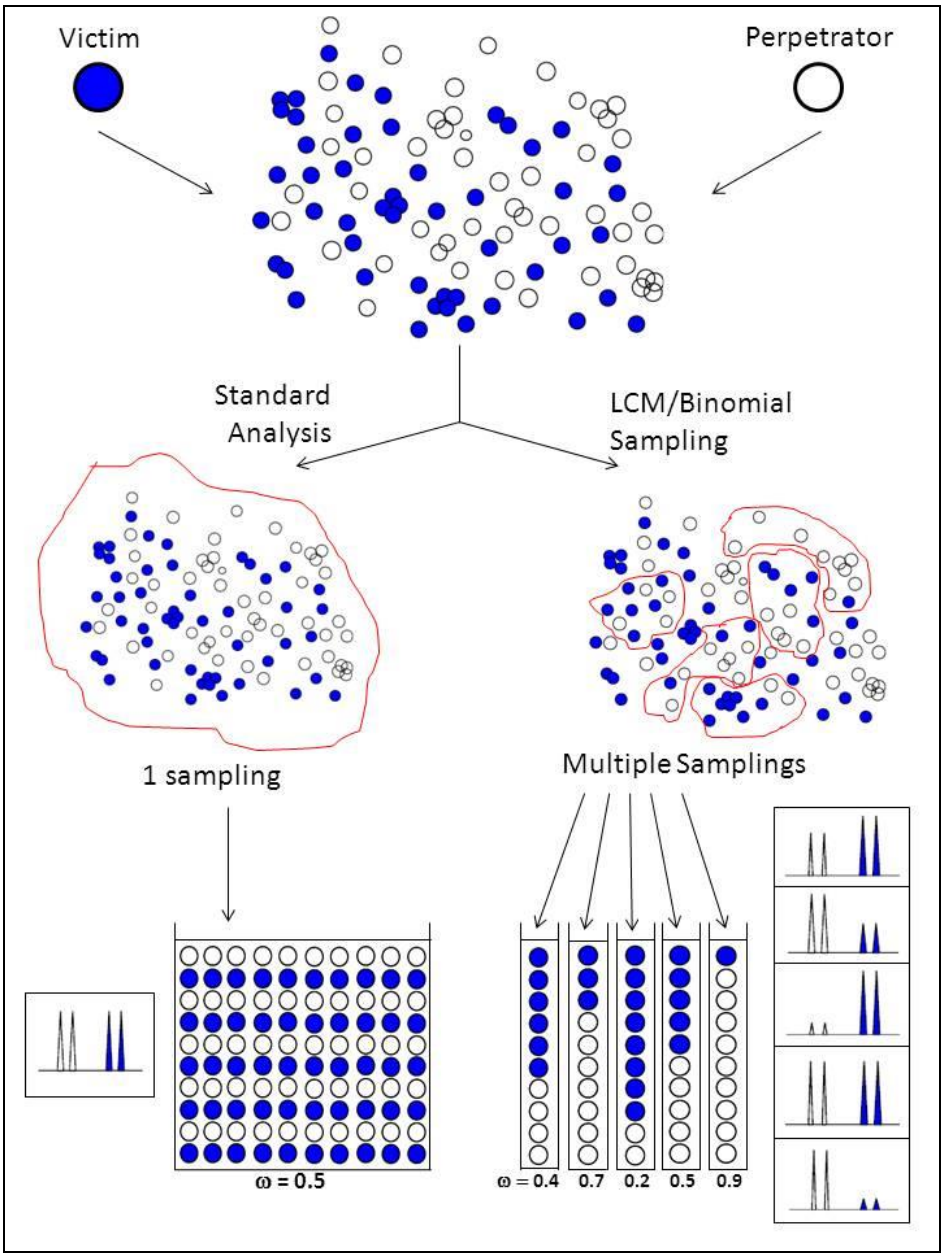


Figure 8. Schema of Standard Versus Binomial Sampling of Mixtures. Admixed biological fluid stains will contain varying amounts of material from each donor (1:1 mixture shown). Cells from the victim are shown in blue and cells from the perpetrator are shown in white. With standard sampling, all of the cellular material would be collected and analyzed thereby resulting in the recovery of a 50:50 mixture profile ($\omega = 0.5$) with the genotypes of each donor not clearly distinguishable (left side). Binomial sampling of the same admixed stain includes sub-sampling of a smaller number of cells within the mixture thereby resulting in sub-samples containing different numbers of cells from each contributor (different ω). The resulting genotypes are more easily deconvoluted into their constituent genotypes (right side).

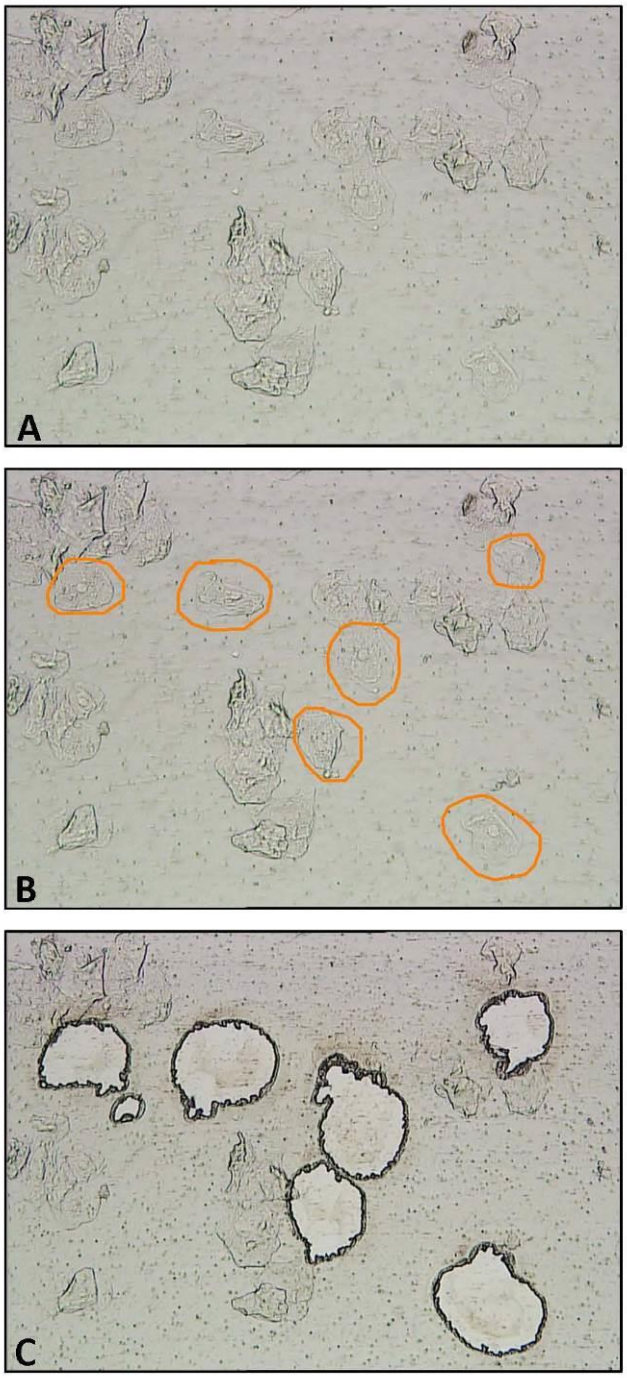


Figure 9. Cell Capture from Epithelial Cell Mounts using the Leica AS LMD System. Epithelial cell suspensions are prepared and mounted on Leica PEN film slides (A). Cells of interest visualized under a stereomicroscope (400 X) are marked for subsequent laser-mediated capture using operator-drawn circles (shown in orange, B). The laser then cuts along the marked lines allowing the cells to drop by gravity into the collection tube cap directly below the microscope stage. After cutting, removal of the the marked cells is visually verified (C).

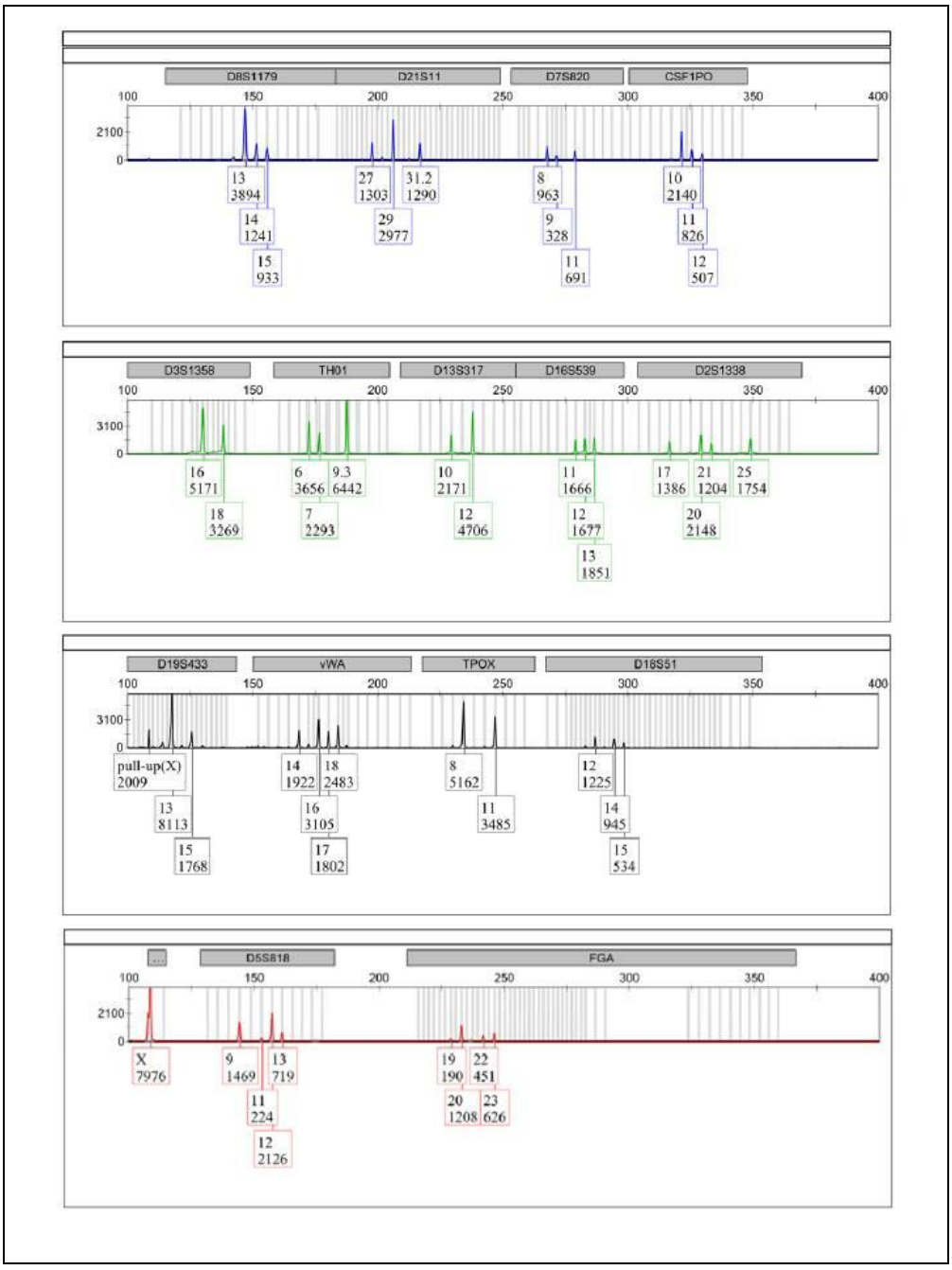


Figure 10. Full Admixed DNA Profile from a Single 20-Cell Sub-Sampling from a 50:50 Mixture. Admixed autosomal STR profiles were obtained after collecting ten 20-cell samples from a 50:50 buccal epithelial mixture from two females using laser capture micro-dissection. A representative electropherogram from one of the ten sub-samples is shown with STR alleles from both donors detected at every one of the 15 STR loci. The x-axis represents fragment size (bp) and the y-axis represents signal intensity (relative fluorescence intensity, RFU).

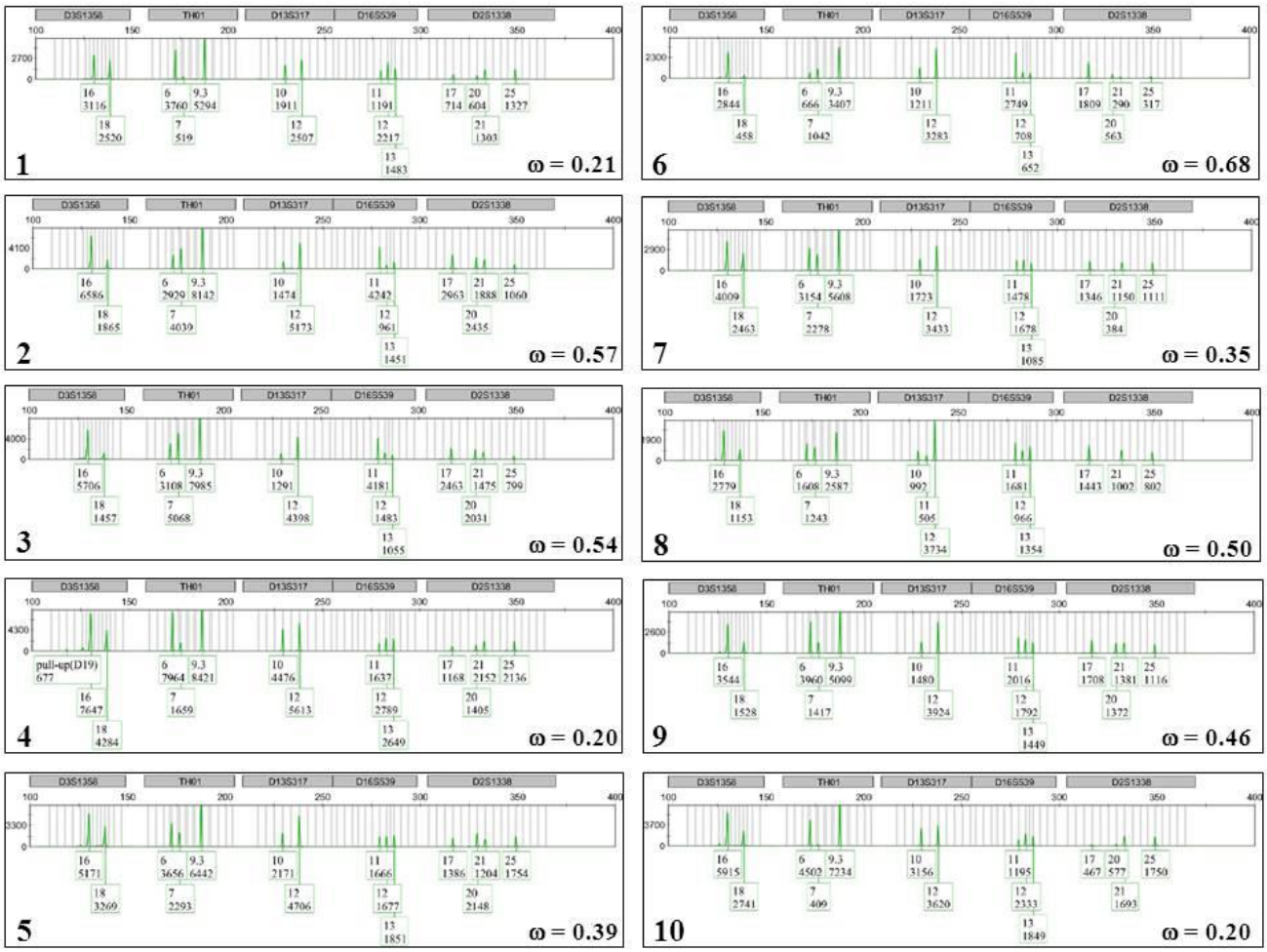


Figure 11. Binomial Sampling from a 50:50 Mixture Yields DNA Profiles with Different Proportions of Each Component. Ten separate 20-cell sub-samples were collected from a 50:50 buccal epithelial cell mixture (two female donors) and autosomal STR profiles obtained. The five locus green channel (VIC) from each of the 10 sub-samples is shown (1-10). The x-axis indicates size in base pairs and the y-axis indicates relative fluorescence units. Allele designations at each locus are indicated. The sub-samples exhibit varying proportions of the two components as evidenced by the varying peak height ratios. The quantitative computer-determined constituent weight ratio (ω) for each sub-sample is listed.

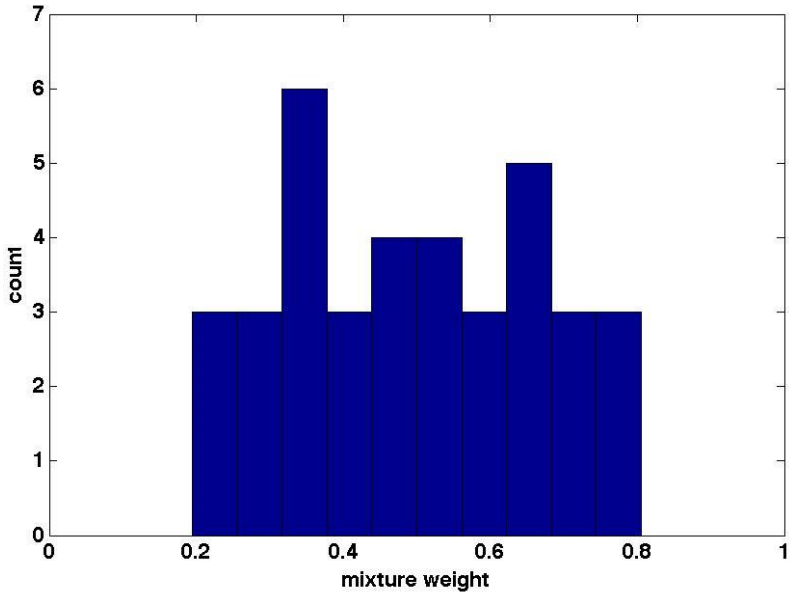


Figure 12. Histogram of Mixture Weight Distribution for Informative Genotypes. Mixture weights are shown for the inferred genotypes used in this study. The histogram's x-axis is the contributor's computer-inferred mixture weight, while the y-axis indicates the number of counts at that weight.

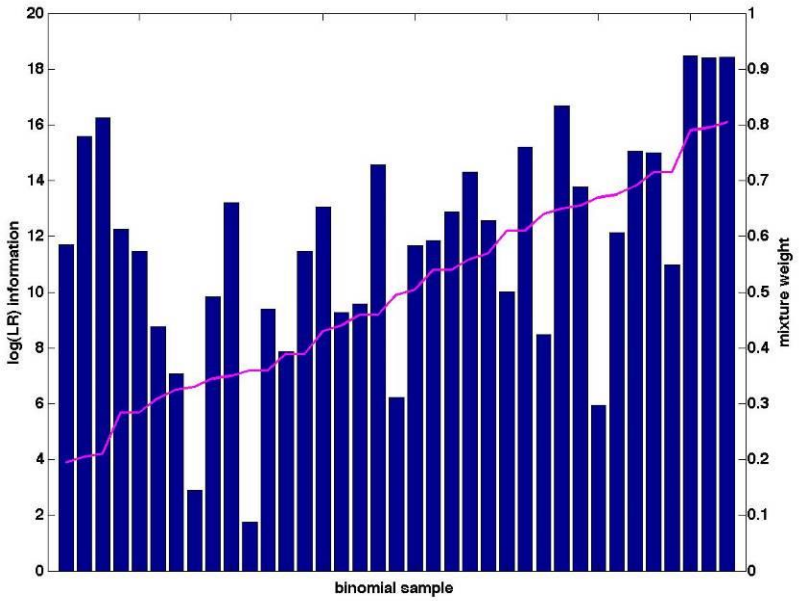


Figure 13. Bar Chart of Inferred Genotype Identification Information. The inferred genotype identification information is shown along its mixture weight. The information is ordered by increasing mixture weight (purple line). The genotype's computer-inferred match information is given in log(LR) ban units (left axis), together with its mixture weight (right axis).

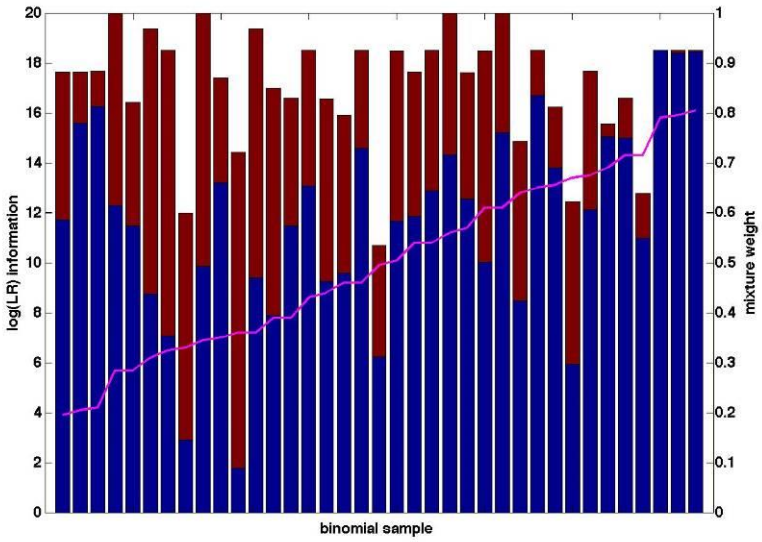


Figure 14. Bar Chart of Match Information for One and Two Samples. TrueAllele® interpreted the LCM LCN EPG data both for one sample and for all possible sample pairings. The identification information is expressed in log(LR) ban units (blue bar, left axis) for ascending mixture weight (purple line, right axis). Shown are the single sample analysis (blue bar), the incremental information added by a joint interpretation with the maximally informative second sample (red bar), and the total match information resulting from this maximal joint inference of two samples (stack height).

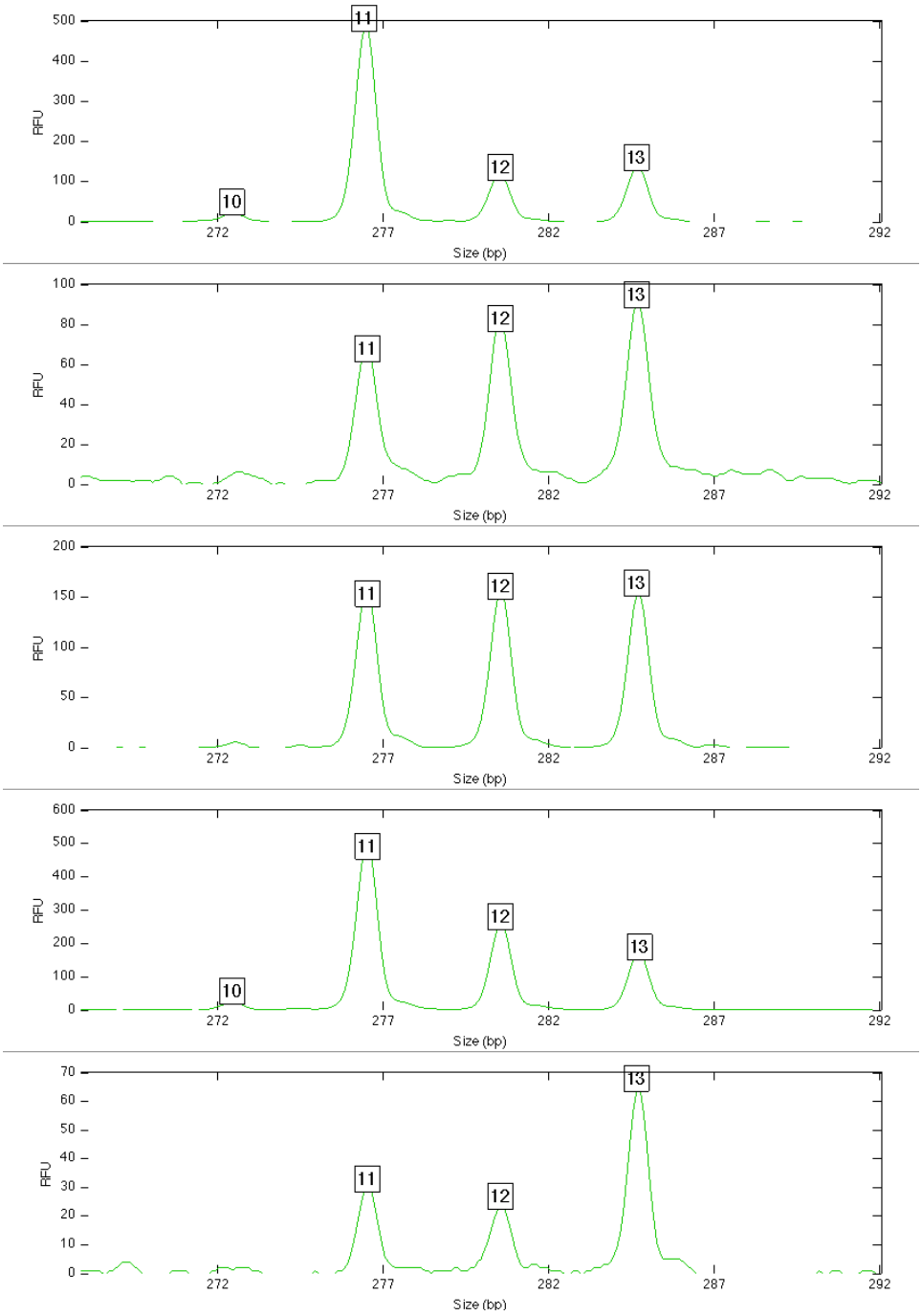


Figure 15. **Electropherogram Data at Locus D16S539**. STR data show peak height and pattern variation when measured in different PCR amplification experiments. This natural variation can be exploited by a computer using a quantitative JLF to conduct a more informative genotype inference. The data shown are from five different PCR amplifications of a 20 cell LCM binomial sample of mixture pair F2:F3, run at the standard 28 cycles.

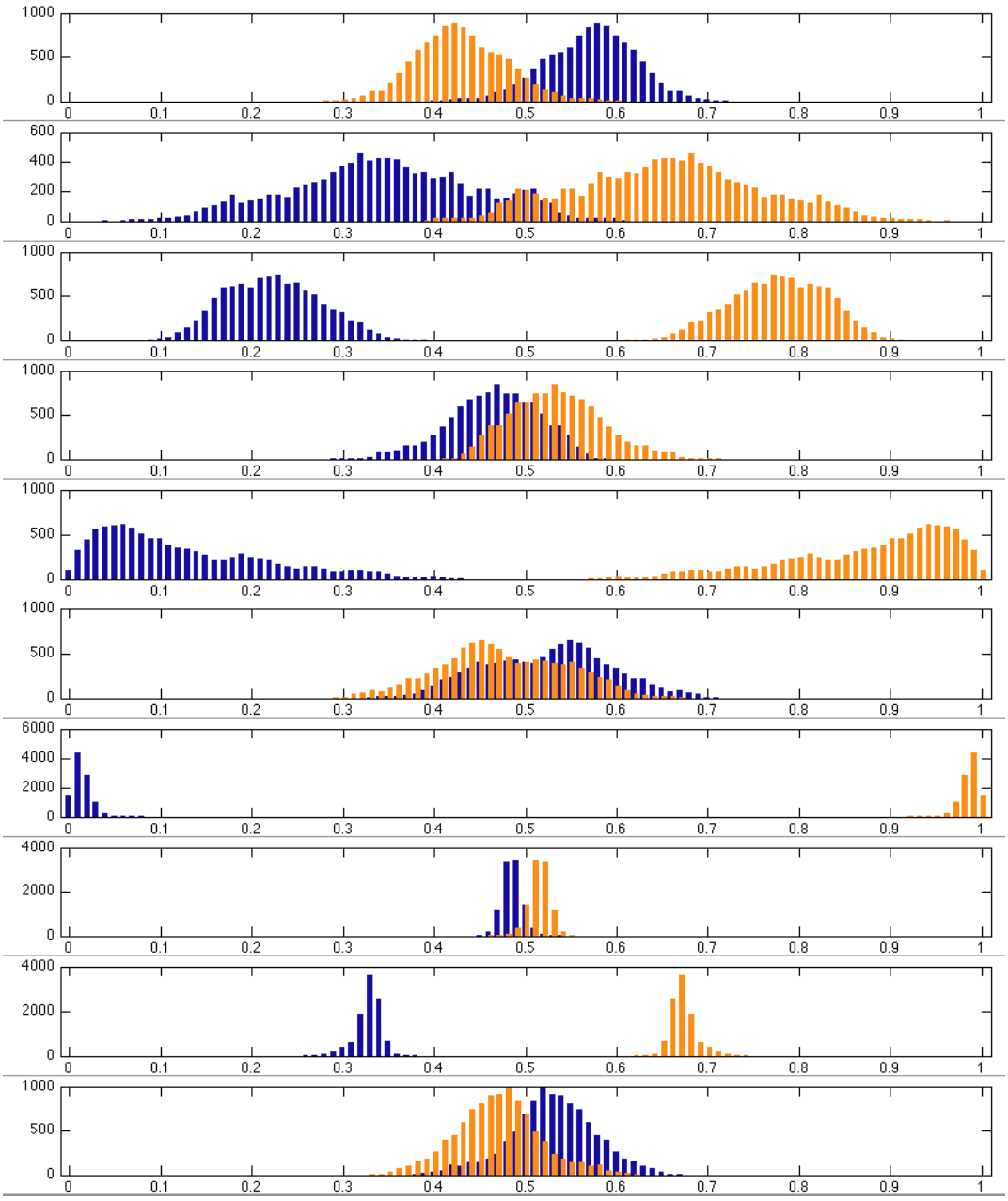


Figure 16. Ten Histograms of Mixture Weight. Binomial sampling separates a 50:50 mixture into DNA samples having a weight generally other than 50:50. TrueAllele® determines the mixture weights in a binomial sample, representing weight uncertainty as a probability distribution, summarized here visually in a histogram. Each histogram shows the mixture weight probability distribution for the two contributors (blue and orange) in one sample. TrueAllele® inferred these distributions in single computer run that jointly interpreted ten binomial samples simultaneously, each sample a 28 cycle amplification of mixture pair F2:F3.

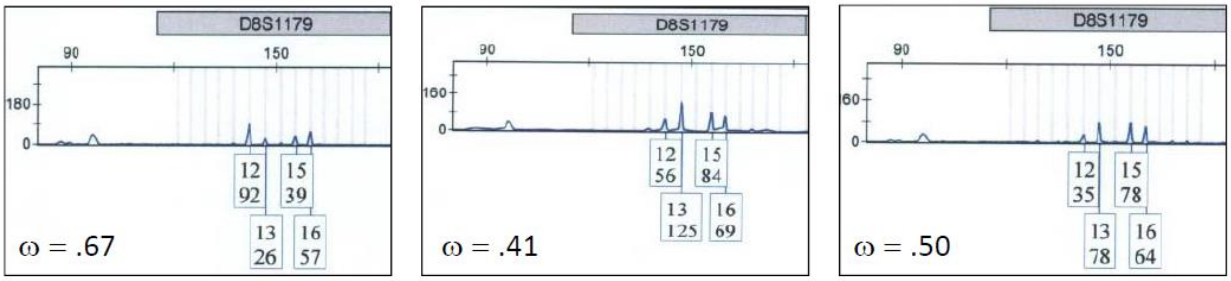


Figure 17. The use of limiting dilutions from a 1:1 admixture to obtain varying donor weight ratios.

The D8S1179 locus (FAM channel, Identifiler® Plus) from three of the twenty 50 pg replicates from a 1:1 male-female mixture is shown. These examples represent the mimimum and maximum weight ratios obtained for the male donor, as well as a comparison to a replicate in which a 0.50 (equal proportions of both donors) was obtained. The x-axis indicates size in base pairs and the y-axis indicates relative fluorescence units. Allele designations at each locus are indicated. The sub-samples exhibit varying proportions of the two components as evidenced by the varying peak height ratios. The weight ratio (ω) for each is listed.

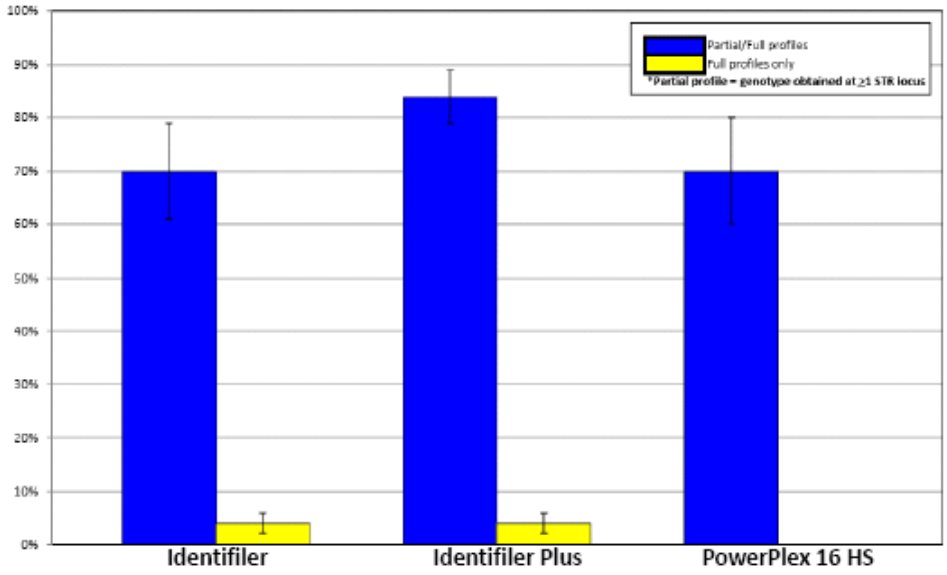


Figure 18. Comparison of the Average Success Rate of Single Cell Profile Recovery Using Three STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).

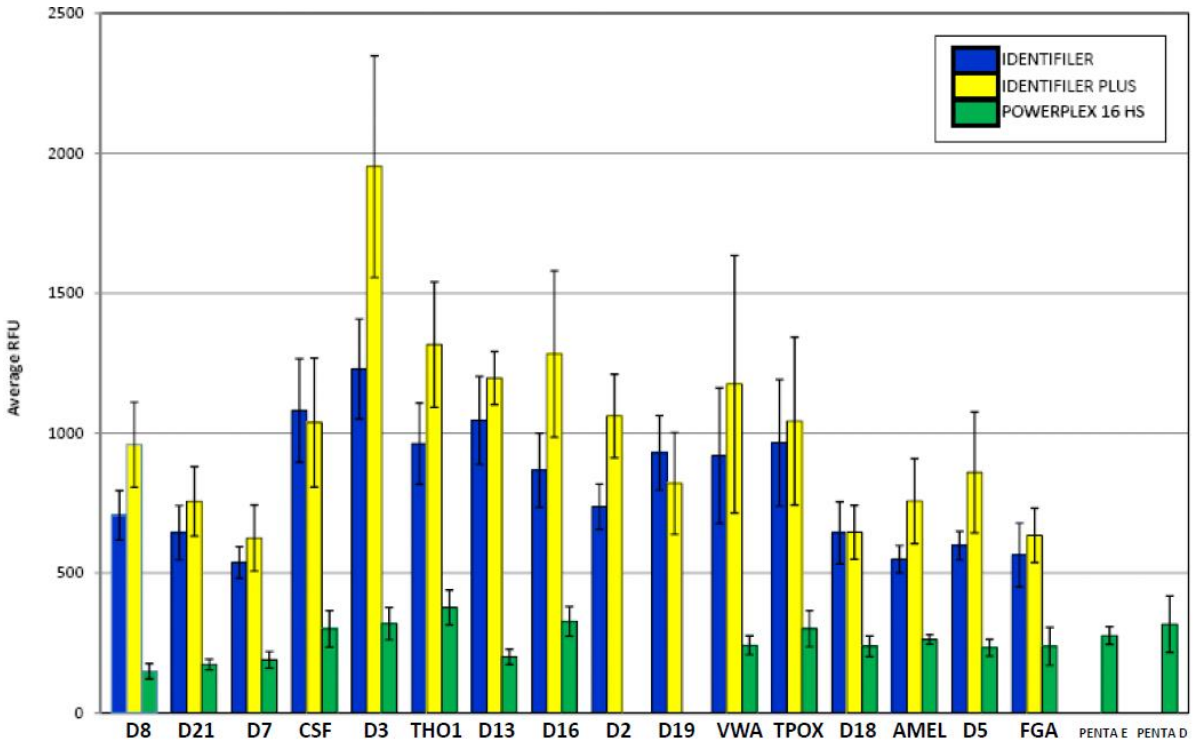


Figure 19. Comparison of the Average RFU Values for Single Cell Profiles Obtained Using Three STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).

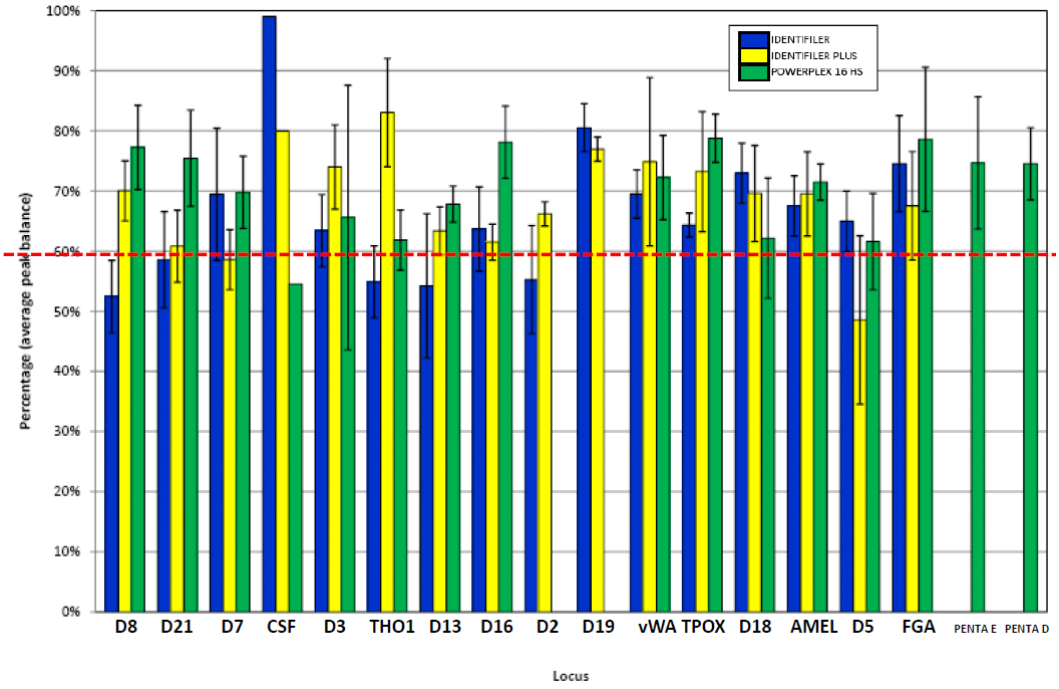


Figure 20. Comparison of the Average Heterozygous Peak Balance for Single Cell Profiles Using Three STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).

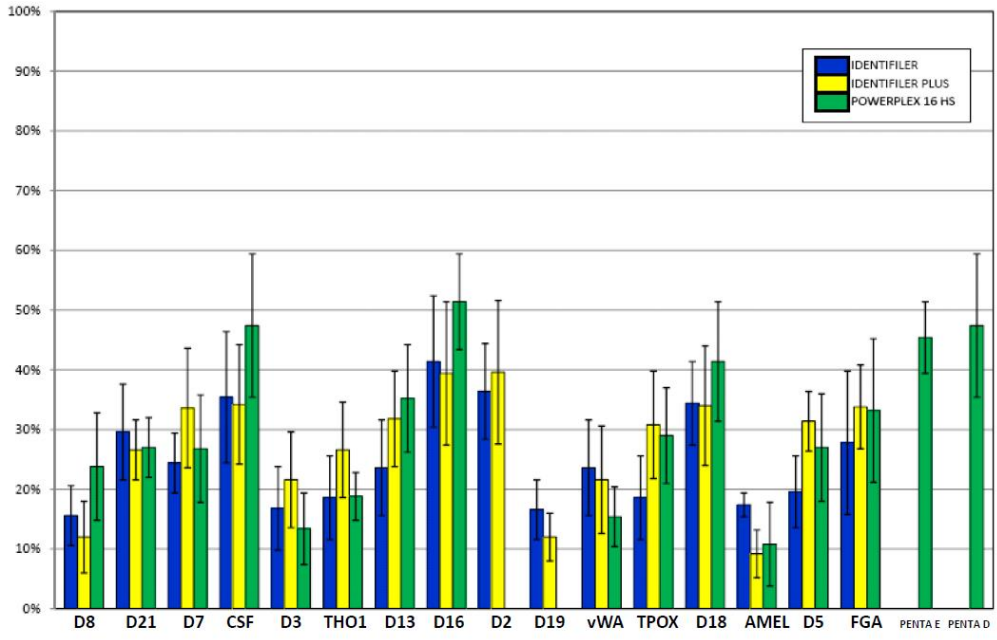


Figure 21. Comparison of the Average Locus Drop Out Rates from Single Cell Profiles Using Three STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).

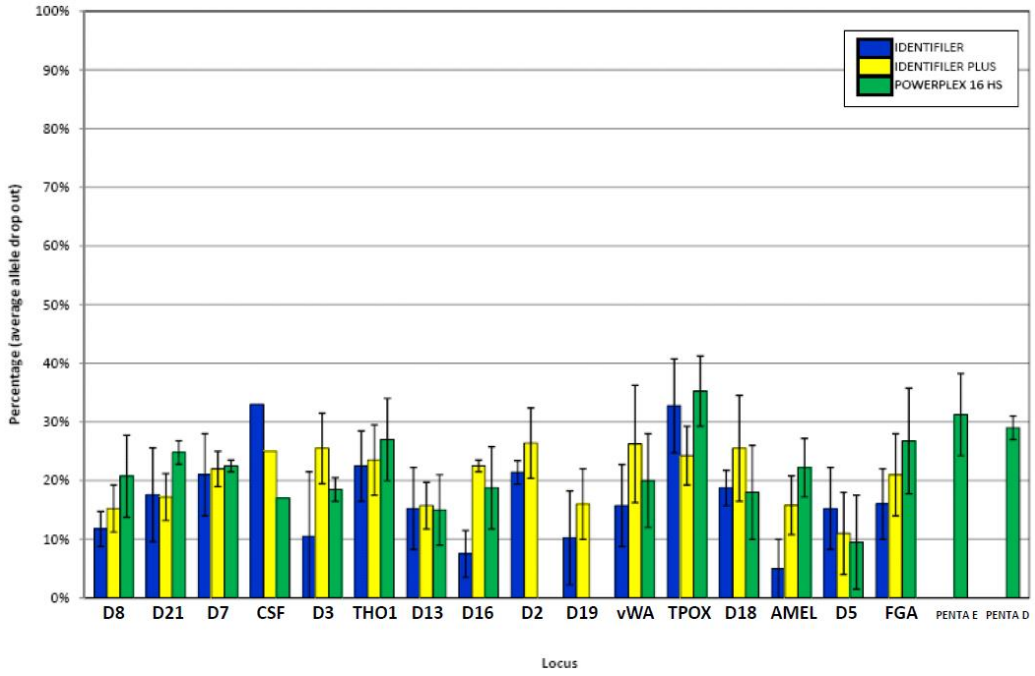


Figure 22. Comparison of the Average Allele Drop Out Rates from Single Cell Profiles Using Three STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).

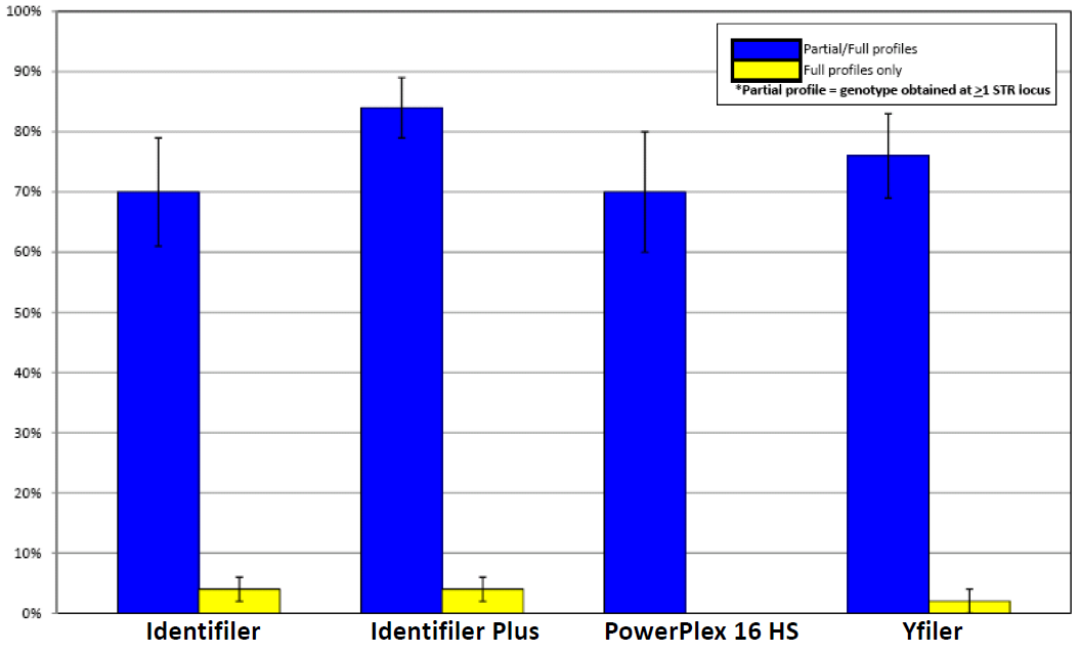
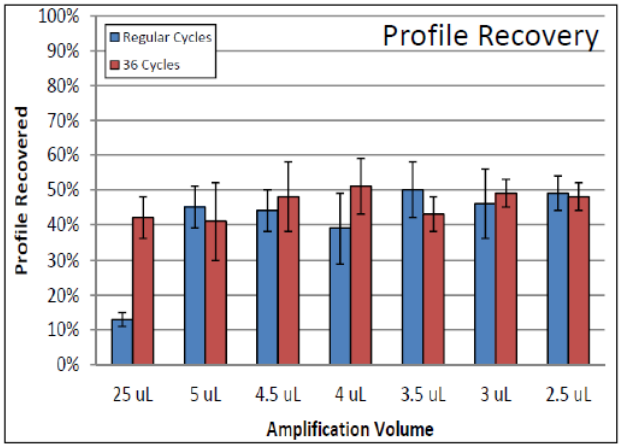
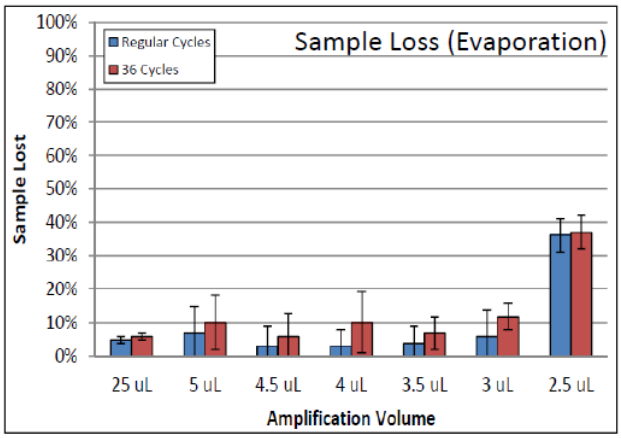
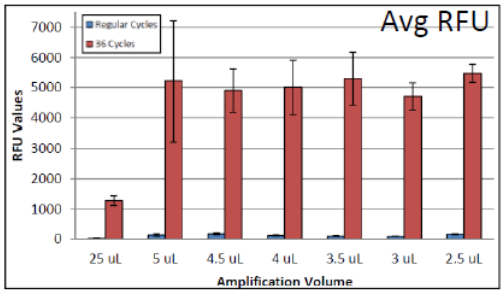


Figure 23. Comparison of the Average Success Rate of Single Cell Profile Recovery Using Four STR Amplification kits. Averages shown were calculated from five donors (average of 10 replicates per donor).







N=1; 5 reps (3-5µl); 20 reps (2.5µl, 25µl)

Figure 24. Comparison of profile recovery, evaporation rates and average RFU values for 5pg DNA inputs using micro-volume PCR reactions.

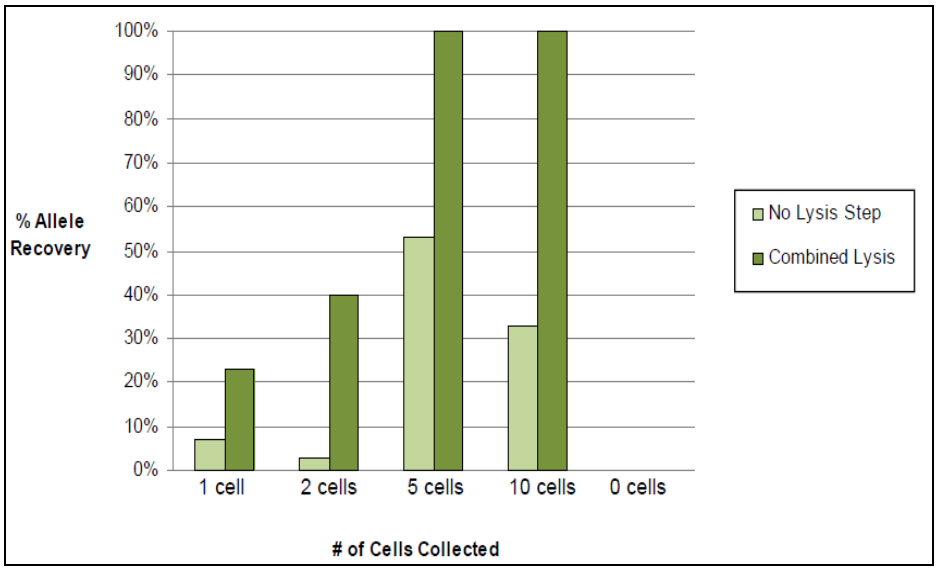


Figure 25. Comparison of profile recovery from micro-dissected buccal epithelial cells using a direct PCR amplfiication and a combined lysis/amplification reaction (standard reaction volume)

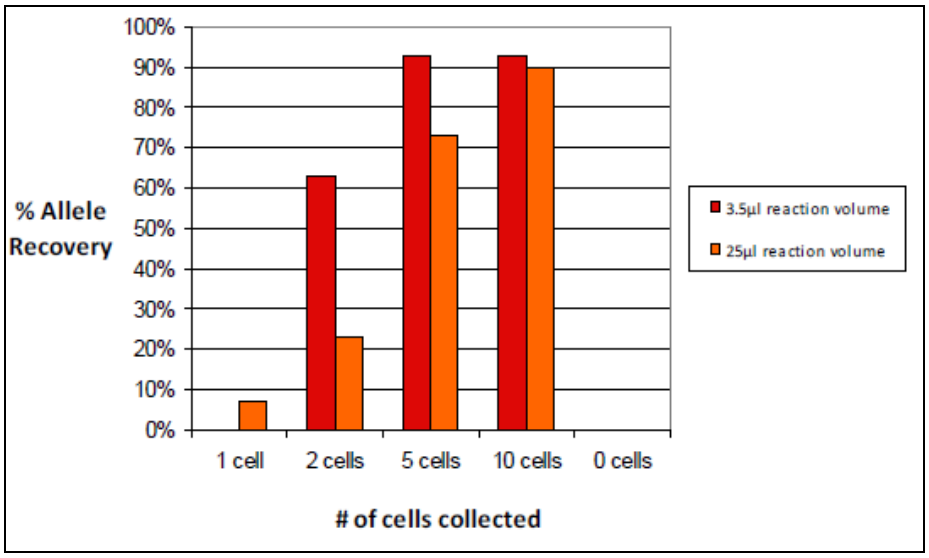


Figure 26. Comparison of profile recovery of micro-dissected buccal epithelial cells using a reduced and standrd volume combined lysis/amplification reaction

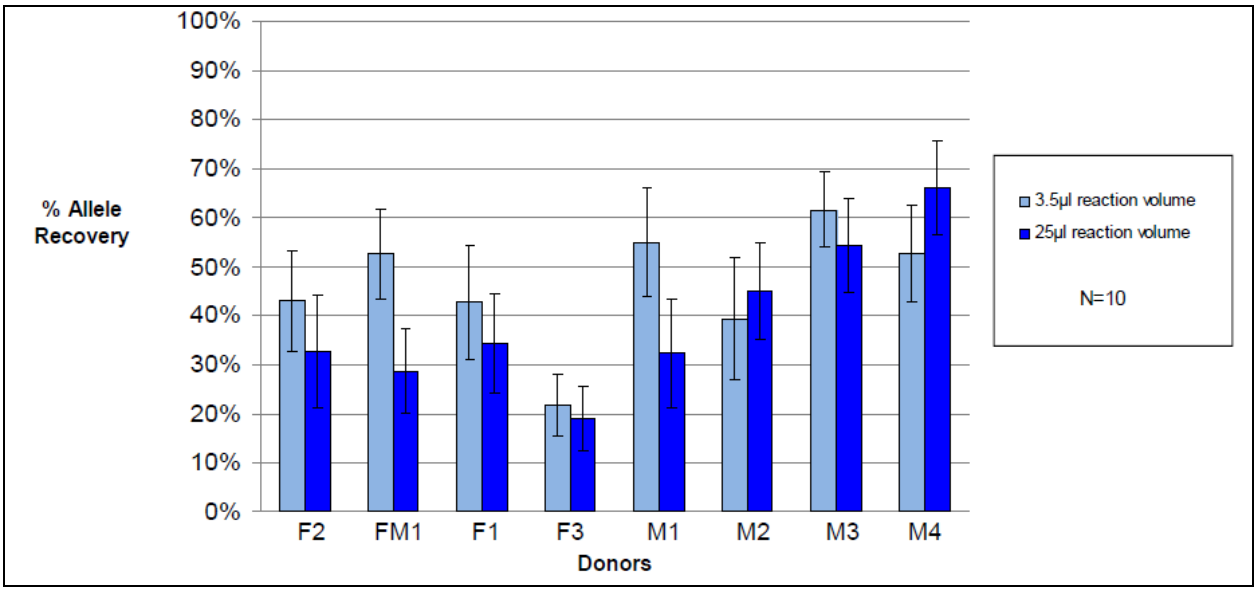


Figure 27. Single Buccal Cell Profiling Success Using a 3.5µl Combined Lysis/Amplifications reaction. Averages shown were calculated from 10 replicates per donor.

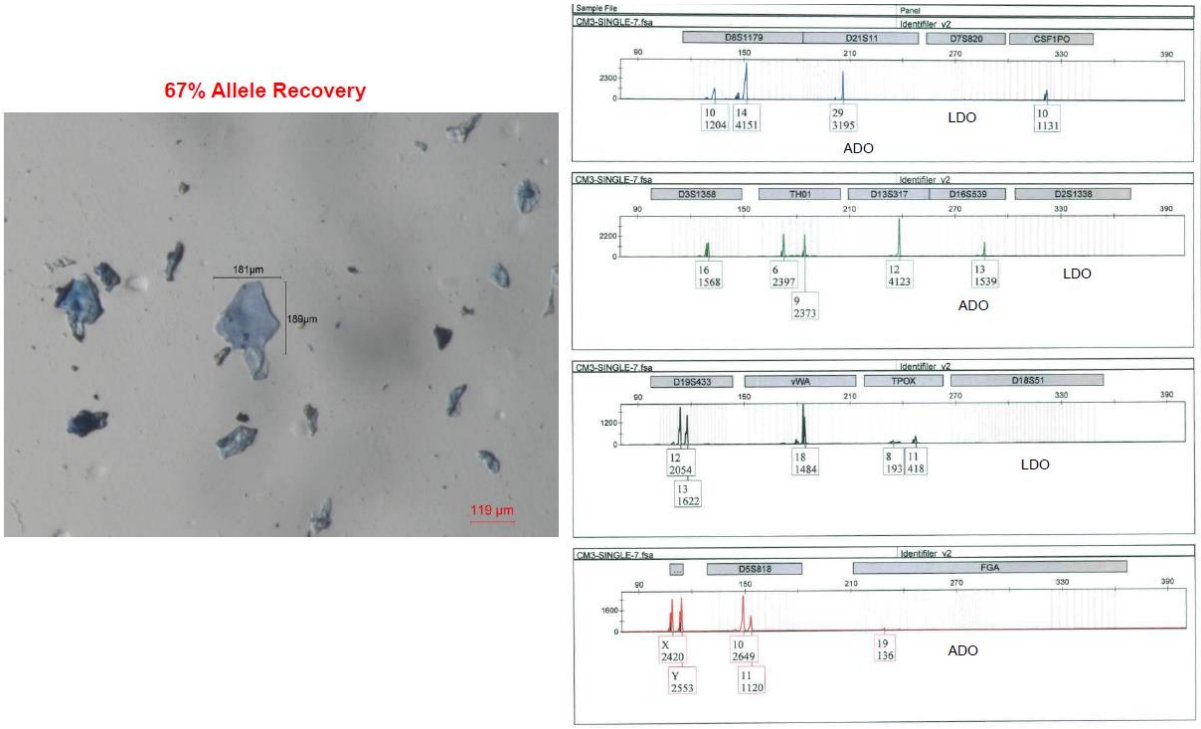


Figure 28. Profile Recovery from Individual Skin Bio-particles Using a 3.5µl Combined Lysis/Amplifications reaction. A single bio-particle (shown on left) was collected from a shirt collar sample using micro-manipulation and analyzed using the developed 3.5µl combined lysis/amplification reaction. The STR profile (partial profile, 67% allele recovery) obtained from this bio-particle is shown (right).

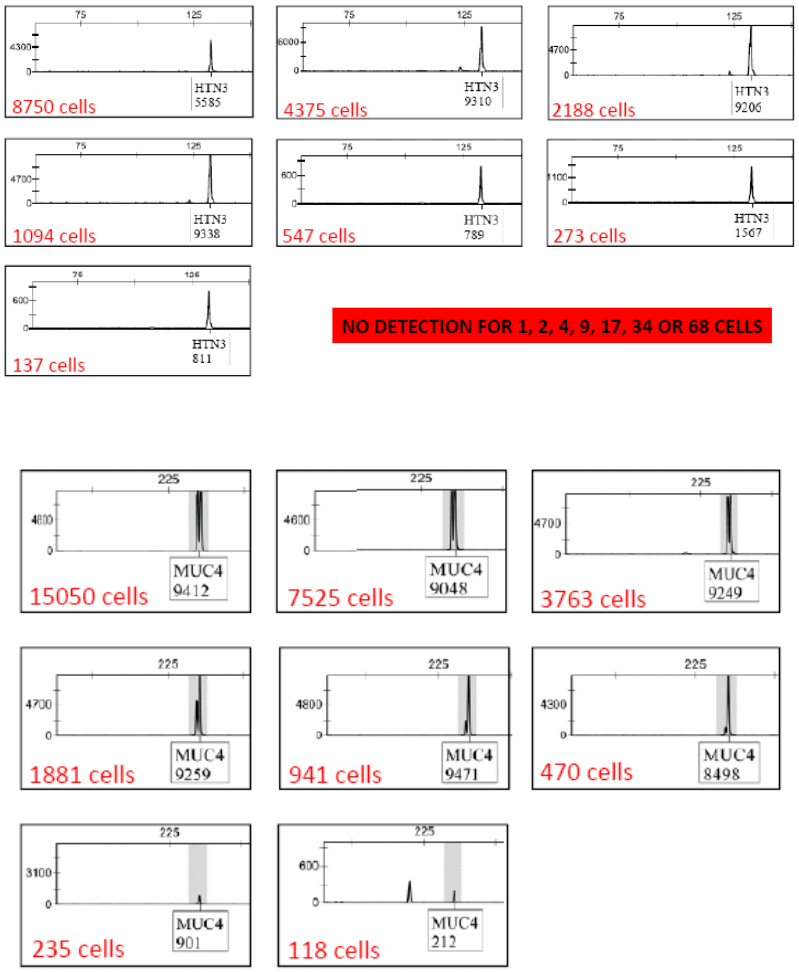


Figure 29. Detection of Saliva Specific mRNAs in Cell Dilution Series. RT-PCR reactions were performed using total RNA extracted from buccal (top panel) and vaginal secretions (bottom panel) epithelial cell dilution series. Each sample was evaluated for the presence of the saliva (HTN3) or vaginal secretions (MUC4) biomarkers. Representative electropherograms for detected biomarkers are shown for each sample. The RFU values for each detected biomarker are provided below the gene name.

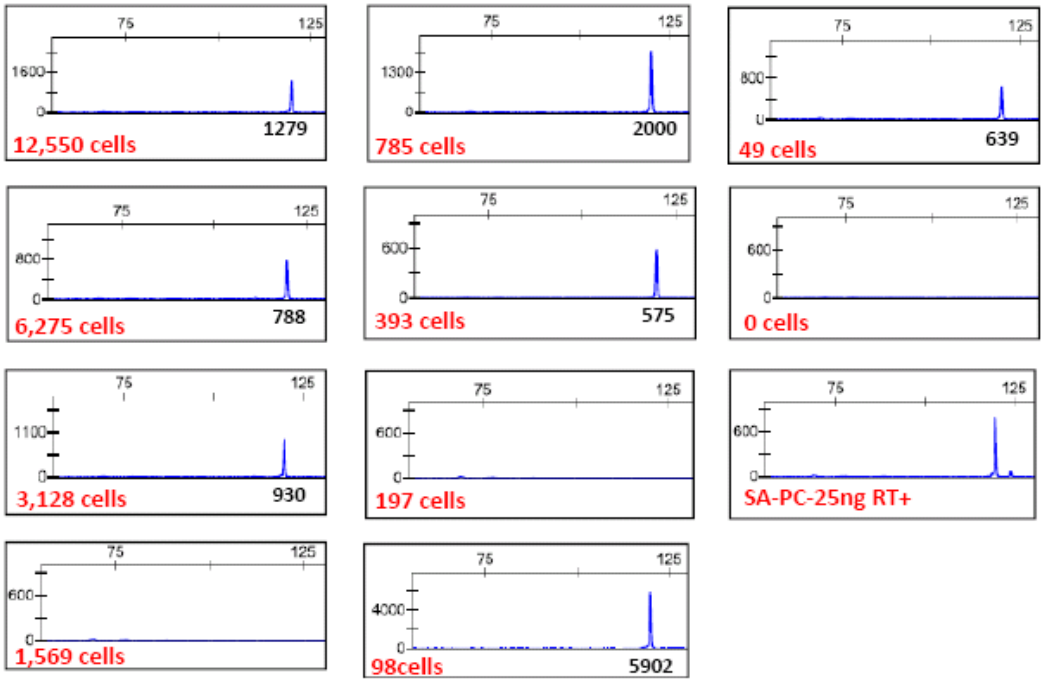


Figure 30. Detection of a housekeeping gene (B2M) in RNA isolated from cell lysates using a *RNA*GEM direct lysis. RT-PCR reactions were performed using total RNA extracted from a buccal epithelial cell dilution series. Electropherograms for each sample in the cell dilution series are shown. The RFU values for each detected biomarker are provided below the gene name. A positive control comprising 25ng of total RNA isolated from a buccal swab is shown (SA-PC25ng RT+) as well as a 0 cell negative control (no detection).

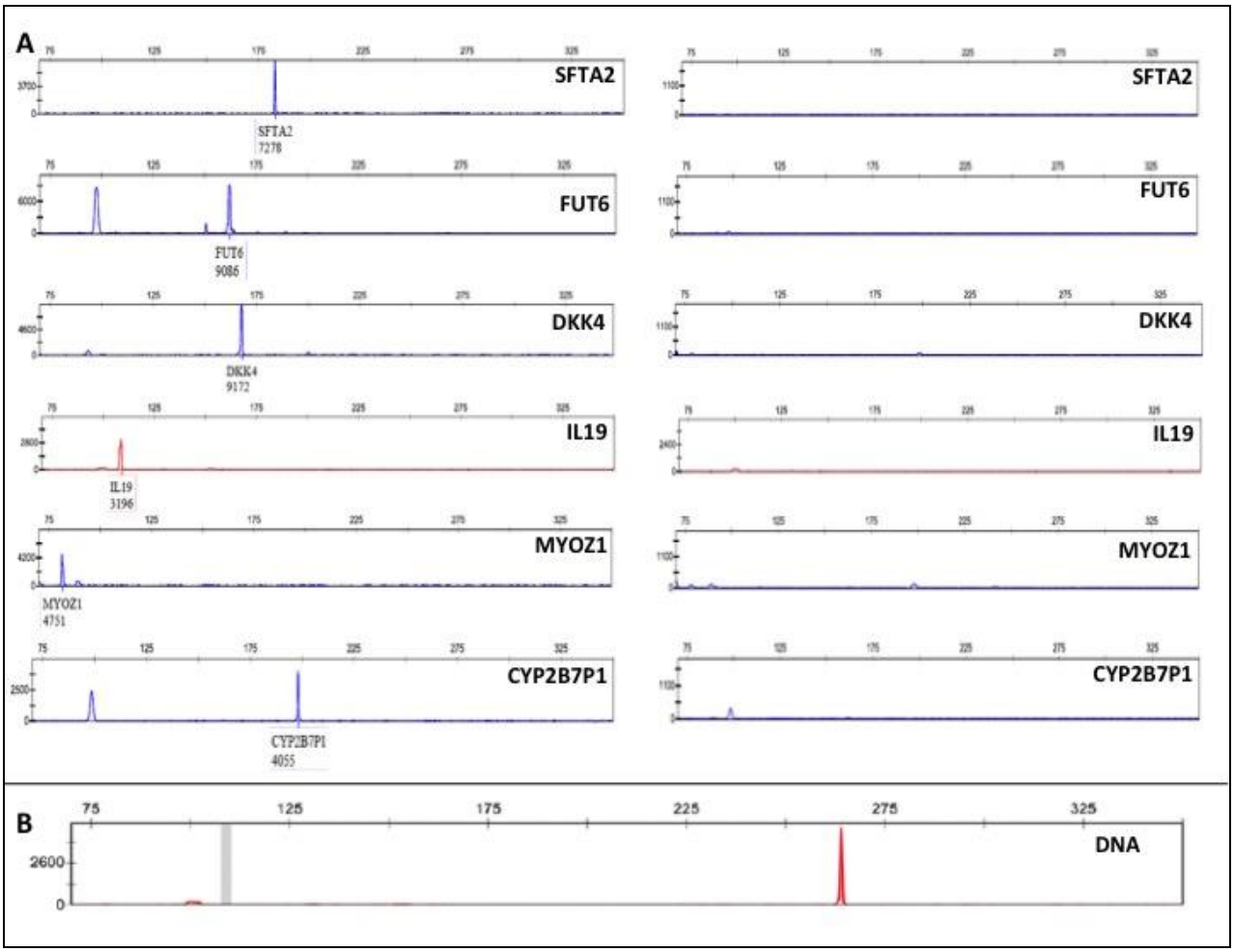


Figure 31. Capillary Electrophoresis Analysis of Putative VS mRNA Biomarkers. RT-PCR products from RNA extracted from vaginal secretion swabs were amplified in singleplex reactions using fluorescent dye-labeled primers. Capillary-electrophoresis separated amplimer products are shown (A) (panel 1 – SFTA2, 183bp; panel 2 – FUT6, 162bp; panel 3 DKK4, 168bp; panel 4 - IL19, 109bp; panel 5 – MYOZ1, 81bp; panel 6 – CYP2B7P1, 198bp). The x-axis indicates size in base pairs and the y-axis indicates relative fluorescence units (Note: The y-axis scales are different for most panels due to varying signal intensities of the observed products). Controls without RT (RT-, A – right panels) were run in parallel with the RT reactions (RT+, B – left) panels. When genomic DNA was tested with the same singleplex primers a product (260bp) was only observed for IL19 (B). The latter was of the expected size, which is different from the mRNA transcript.

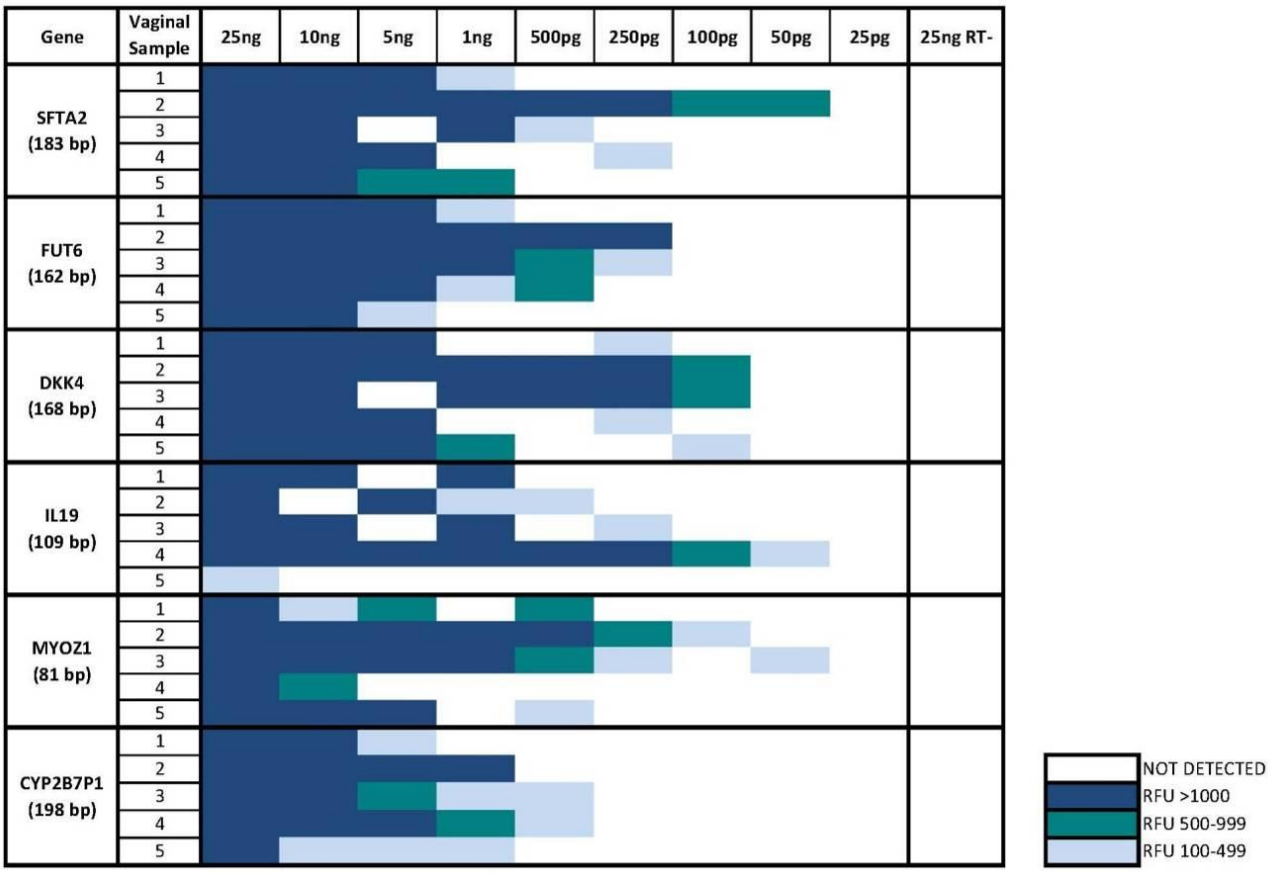


Figure 32. Sensitivity Heat Map for the Six Putative VS Specific Gene Transcripts. RT-PCR products for six putative vaginal specific gene transcripts using different quantities of input RNA are shown (25ng – 25pg). Sensitivity was evaluated with five individual female donors (1-5). Levels of detection of the vaginal gene transcripts are indicated by a colored square (dark blue >1000 RFUs; teal blue 500-999 RFUs; light blue 100-499 RFUs; white – no detection). No products were detected in the RT negative controls (RT-) which employed 25ng of input total RNA.

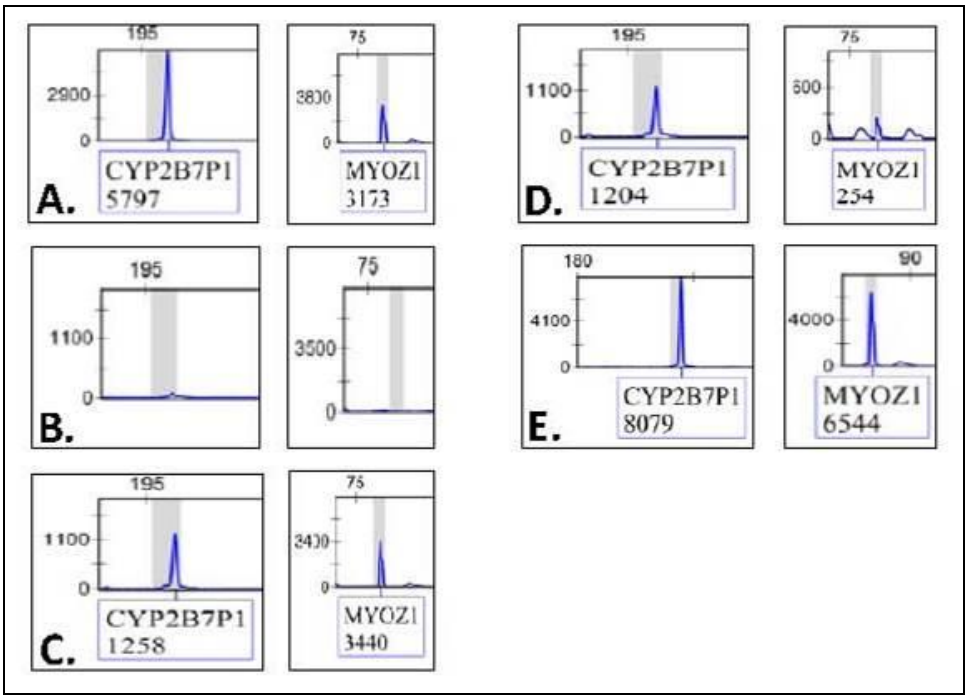


Figure 33. Detection of VS mRNA Markers in Mock Sexual Assault Casework Samples. RT-PCT products using RNA extracted from a male finger after digital penetration (A), a penile swab taken prior to (B) and after (C) sexual intercourse, a swab of the inside front of male underwear worn for 3 hours after sexual intercourse (D), and a surface swab of a foreign object after insertion into the vaginal canal (E) are shown after singleplex amplification with two vaginal specific gene transcripts (CYP2B7P1 (198bp), left panel; MYOZ1 (81bp), right panel). Relative fluorescence units (RFU) for obtained products are provided below the gene name label. Expected amplimers were detected in all samples, except the pre-intercourse penile swab that was included to demonstrate the absence of vaginal secretions prior to the collection of the post-intercourse penile swab.

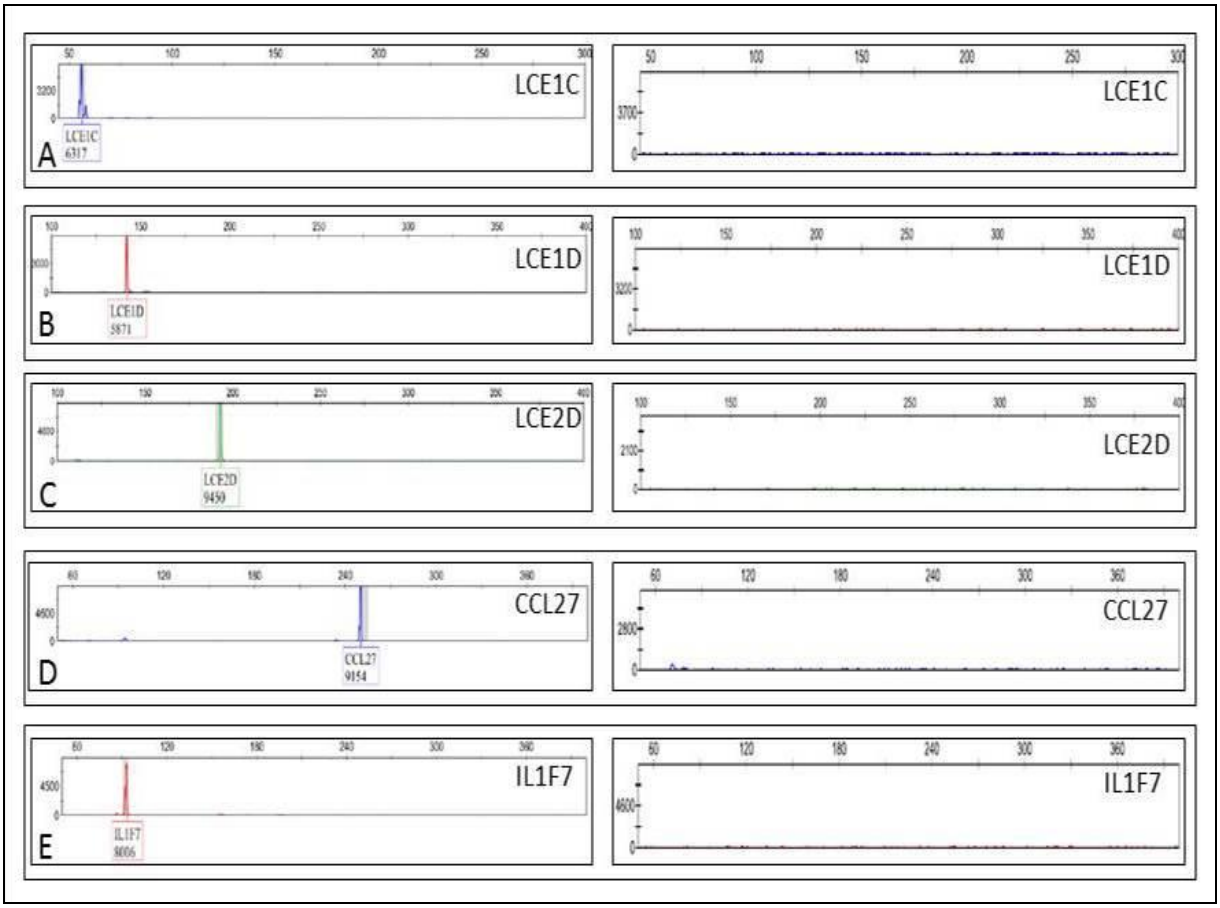


Figure 34. Singleplex CE assays for skin biomarkers. RT-PCR reactions were performed using 25 ng of total RNA isolated using a manual organic extraction, with the exception of LCE1C where 100 pg of total RNA was used. Singleplex amplification products are shown (LCE1C – 56/58 bp (A); LCE1D – 142 bp (B); LCE2D – 193 bp (C); CCL27 – 254 bp (D); IL1F7 – 92 bp (E)). Representative electropherograms are shown. Controls without RT (RT-, right panel) were run in parallel to the RT reaction (RT+, left panel).

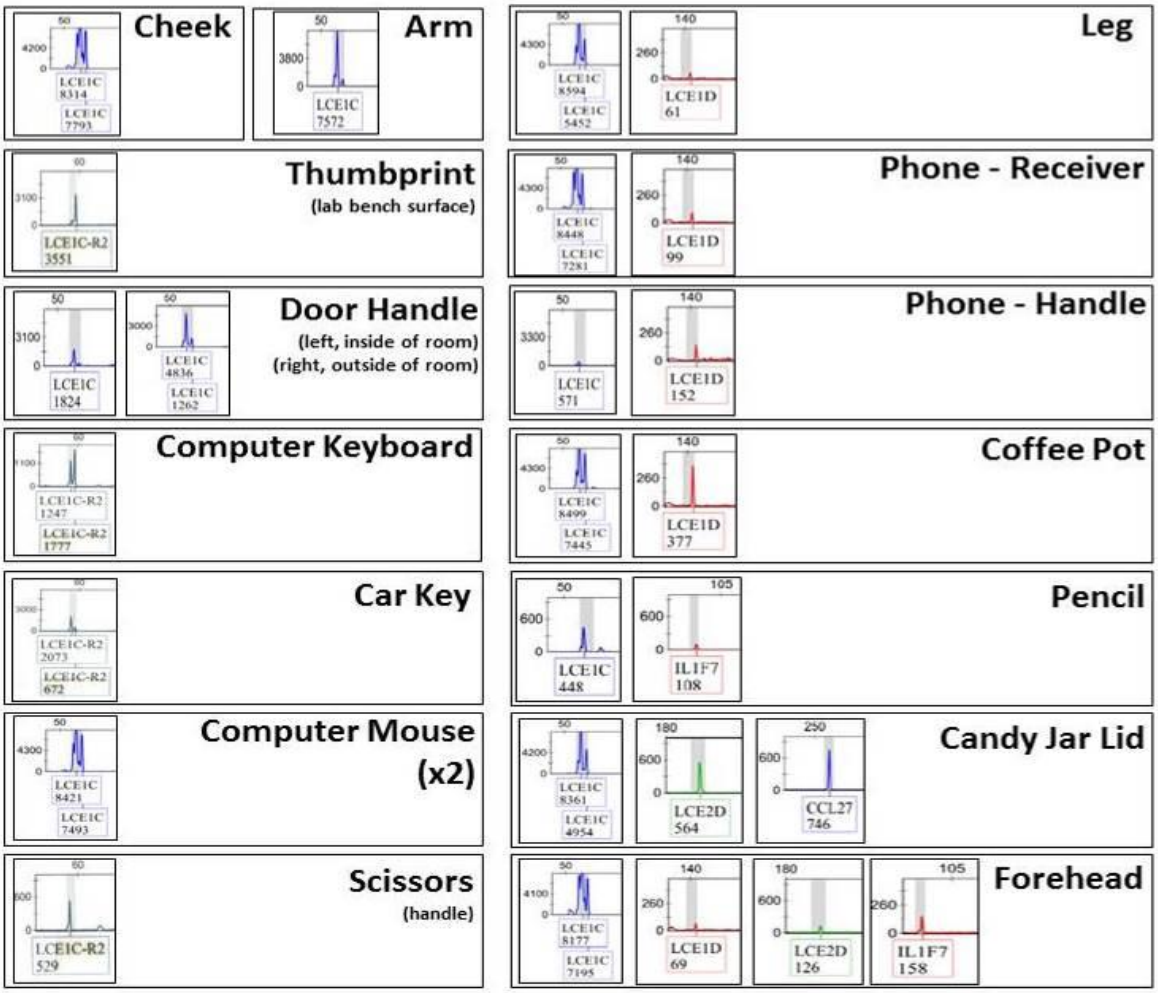


Figure 35. Skin biomarker detection in human skin and touched objects. RT-PCR reactions were performed using total RNA extracted from surface swabs of human skin and various touched objects. Each sample was evaluated for the presence of the five skin biomarkers. Representative electropherograms for detected biomarkers are shown for each sample. The 'x 2' designation indicates that positive results for two different computer mouse samples were obtained but only one representative electropherogram is shown.

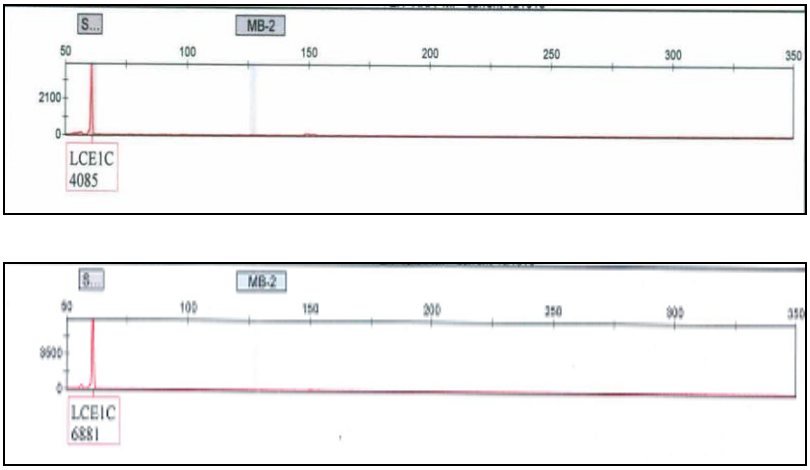
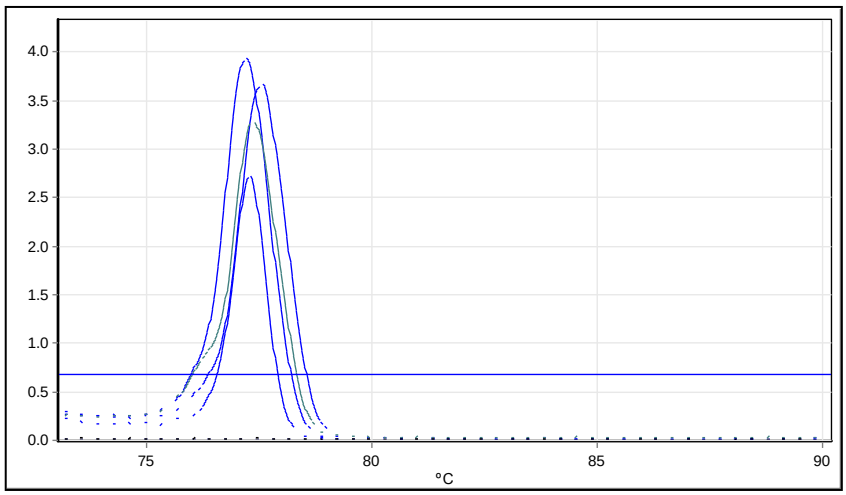
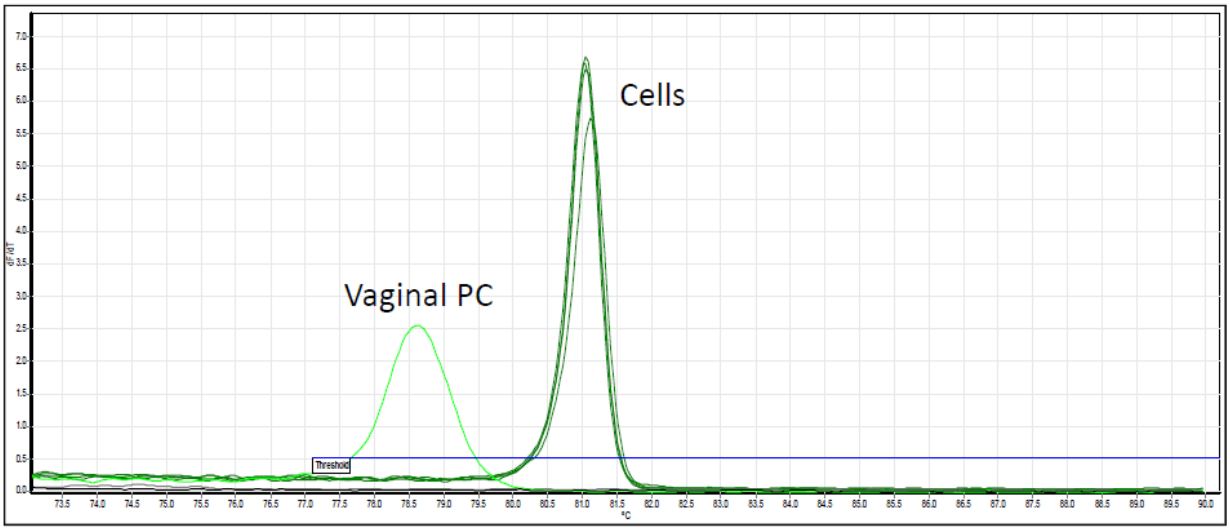


Figure 36. Skin biomarker detection in isolated bio-particles from worn clothing items using RNA pre-amplification strategies. Bio-particles from a shirt back and lab coat were collected using micro-manipulation. RNA was isolated from the recovered bio-particles using a reduced volume direct RT reaction, followed by singleplex LCE1C pre-amplification. Pre-amplifications products were then used in a subsequent mRNA body fluid identification multiplex (CE-based detection). Representative electropherograms from the shirt back (10 bio-particles) and the lab coat (25 bio-particles) are shown.



No.	Name	Tm	Saliva?
13	SA-10c		no
14	SA-25c	77.55	yes
15	SA-50c	77.28	yes
16	SA-100c	77.2	yes
17	SA-0c		no
18	Pre-amp-BL		no
19	SA49(+)PC	77.35	yes
20	Blank		no

Figure 37. Detection of the Saliva/Buccal Cell Origin of Isolated Buccal Epithelial Cells Using a Singleplex (HTN3) HRM Detection Assay. Ten, twenty-five, fifty and one hundred buccal cells were collected using micro-manipulation. A reduced volume combined lysis/RT reaction was performed with the resulting cDNA pre-amplified using a singleplex HTN3 pre-amplification reaction. A singleplex HTN3 HRM assay was used for biomarker detection. A positive result is indicated by the presence of a peak at the expected Tm value for HTN3 (~77ºC)



No.	Name	Peak 1
9	VS-10C(+)I	81.1
10	VS-25C(+)I	81.0
11	VS-50C(+)I	81.1
12	VS-100C(+)I	81.1
14	PA-Blank-I	
15	VS-FM38(+)PC-I	78.6
16	Blank-I	

Figure 38. Detection of the Vaginal Secretions Origin of Isolated Vaginal Epithelial Cells Using a Singleplex (IL19) HRM Detection Assay. Ten, twenty-five, fifty and one hundred buccal cells were collected using micro-manipulation. A reduced volume combined lysis/RT reaction was performed with the resulting cDNA pre-amplified using a singleplex IL19 pre-amplification reaction. A singleplex IL19 HRM assay was used for biomarker detection. A positive result is indicated by the presence of a peak at the expected Tm value for IL19 (~79^{o}C). A slight increase in the expected Tm was observed for the vaginal cell samples.

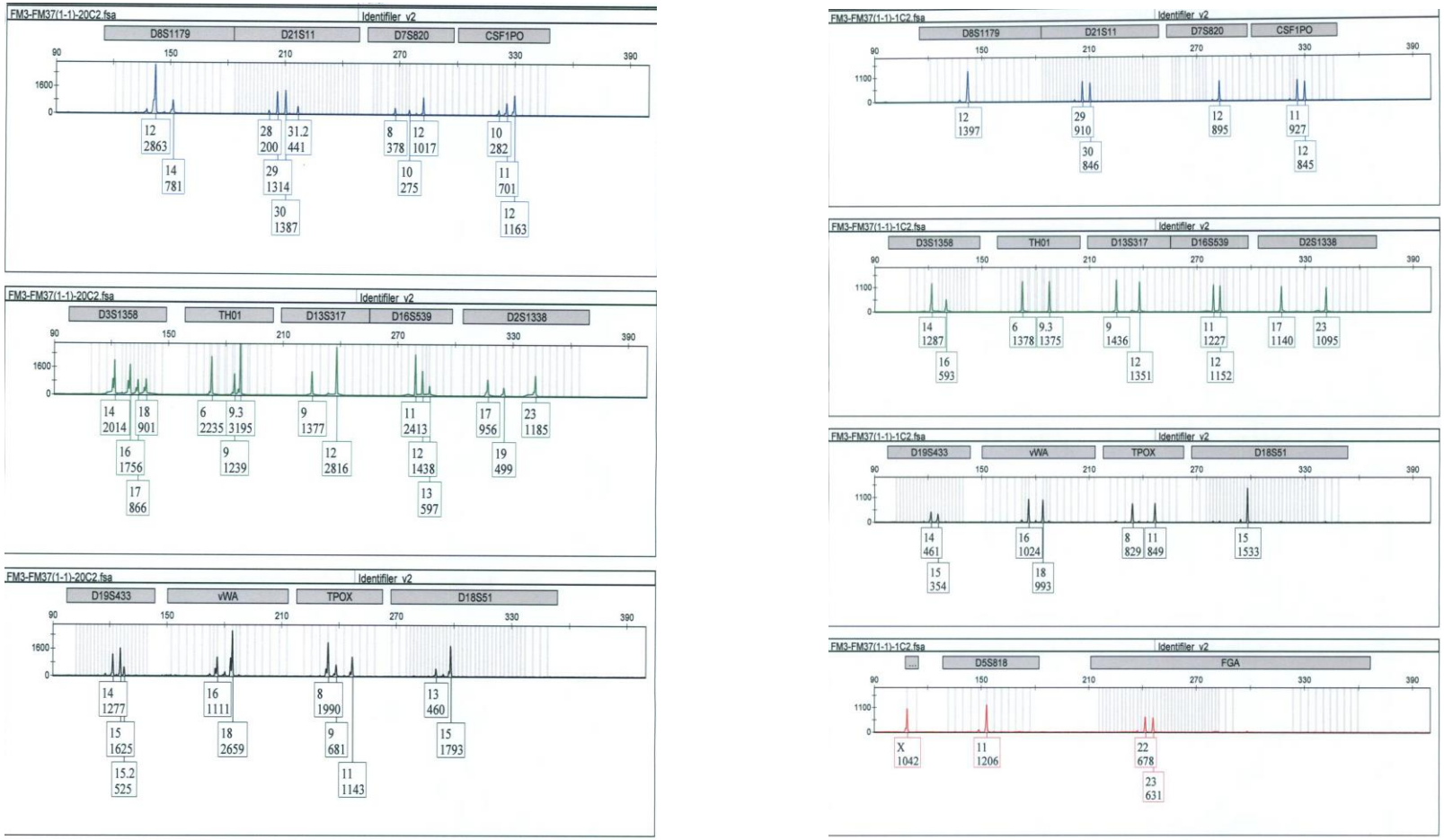


Figure 39. De-convolution of Epithelial Cell Admixtures (Left) Using Single Cell Profiling (Right). An admixed autosomal STR profile (Identifiler) was obtained after collecting a 20-cell sample from a 50:50 buccal epithelial mixture from two females using laser capture micro-dissection (left panel). A single source autosomal STR profile of one of the donors was obtained after collecting and typing a 1-cell sample from the same 50:50 buccal epithelial mixture. The x-axis represents fragment size (bp) and the y-axis represents signal intensity (relative fluorescence intensity, RFU).

IV. CONCLUSIONS

A. Discussion of findings

The goal of the current work was to develop novel strategies for the de-convolution of body fluid mixtures containing non-distinguishable cell types (e.g. epithelial cells) through an isolation of individual cells using laser capture micro-dissection and recovery of genetic material (RNA and DNA). Using this type of approach, isolation and profile recovery from a single cell would by definition eliminate the presence of an admixed profile. Therefore, a significant focus of the current study was to develop robust and sensitive DNA profiling methods for the analysis of single epithelial cells. While the use of single cell profiling may raise serious concerns for operational laboratories due to the extreme low template nature of these samples, the goal in the current work was to begin to evaluate such a strategy to demonstrate possible future use in selected cases in which admixed profiles couldn't be resolved by conventional means.

In this study, we demonstrate the ability to isolate single and few epithelial cells isolated from admixed samples using either laser capture micro-dissection or micro-manipulation and recover genetic profiles from these samples using enhanced typing strategies such as micro-volume PCR reactions and single-tube one-step lysis and amplification reactions. Figure 39 provides an example of the potential use of single cell analysis. If a two person admixed profile was initially obtained from a non-distinguishable cell type mixture (Figure 39 left panel, 20 cell sample) in which the profiles could not be resolved easily by conventional means, profiles recovered from single cells collected from the mixed sample (Figure 39 right panel, 1 cell sample) could be used as "reference profiles" for one of the individual donors in order to aid in de-convolution of the admixed sample by simple profile subtraction. The quality of the single cell example shown in Figure 39, at this time, will not be obtained in every single cell sample collected. However, replicate analysis of single cells (10-20 replicates) can result in the recovery of suitable single source profiles potentially from one or both donors separately, or can be used collectively to generate a composite reference profile if necessary.

While the above example demonstrates the proof of concept ability to de-convolute an epithelial cell mixture using single cell analysis, we aknowledge that such an approach could not be implemented into routine casework at this time. Therefore, rather than relying solely on the use of single cell profiles to aid in mixture de-convolution (since further validation and optimization is required), we also explored the use of the quantitative computer-based TrueAllele® interpretation system to aid in mixture de-convolution (collaboration with Mark Perlin, CyberGenetics). We hypothesised that LCM-mediated isolation of multiple groups of cells ('binomial sampling') from the admixture would create separate cell sub-populations with differing constituent weight ratios. Furthermore we predicted that interpreting the resulting DNA profiling data by the TrueAllele® interpretation system would result in an efficient recovery of the constituent genotypes due to newfound abilities to compute a maximum LR from sub-samples with skewed weight ratios, and to jointly interpret all possible pairings of sub-samples using a joint likelihood function. The data confirmed that, as predicted, binomial sampling of two person (1:1) epithelial cell admixtures by LCM produces sub-samples with differing weight ratios. The recovery of different sub-samples containing different proportions of each constituent necessarily and demonstrably facilitates DNA mixture genotype inference. We showed that quantitative computer interpretation using the TrueAllele® system efficiently recovered the full genotype information from the original sample using the sub-samples' genotype data. Conceptually, then, it is possible using the binomial sampling method to

construct more informative data from the same evidence, compared to standard biological sampling strategies.

The efficient ('smart') recovery and STR genotyping of individual cells by LCM described in detail in this report is achieved by (i) lysing the cells using a thermostable protease, (ii) performing the lysis and amplification in a single tube and (iii) using an increased PCR cycle number (34 cycles). The latter is strictly not essential since 28 cycles also yields comprehensive genotype information with TrueAllele®. This LCM approach is not a high throughput method, but it could be used 'as is' for lab-determined high priority mixture cases with a heightened requirement to maximize the genotype information recovered.

Joint examination of multiple data samples was seen to preserve more identification information than conducting individual analyses on each sample in isolation. This result was predicted by statistical theory [29], since a joint likelihood function assigns higher probability to those genotype values that better explain all of the observed quantitative data patterns. Each additional data element imposes further constraints on the genotype possibilities, potentially sharpening the probability distribution. Indeed, probability theory mandates a joint interpretation, requiring that every data element that is affected by a (genotype, say) variable be included in that variable's likelihood function [75,76].

We used TrueAllele® in this study as a measurement tool that accurately and reproducibly determined identification information and mixture weight. Previous TrueAllele® papers were validation studies that centered on demonstrating the improvement of computer interpretation of quantitative DNA evidence relative to threshold-based human review [28,42]. We moved beyond those foundational studies that had established TrueAllele®'s reliability and other advantages. Instead, we employed the validated TrueAllele® system as an analytic laboratory instrument that could accurately quantify more precise scientific information about our data.

The power of above described techniques could be further enhanced if the tissue source of the sampled cells could be determined by RNA profiling. Therefore another aim of this study was not only to recover individual DNA profiles, but to also directly characterize the tissue source of each of the profiles (e.g. that a particular DNA profile originates from buccal epithelia, while the other from vaginal epithelia). While RNA profiling of individual cells proved to be rather challenging, we present here preliminary results demonstrating the ability to determine the cell type of origin of directly sampled epithelial cells. We performed extensive and comprehensive studies to identify novel sensitive and highly specific mRNA biomarkers for the identification of vaginal secretions and skin, and were succesful in this quest. We have developed a preliminary single-step micro-volume cell lysis/reverse transcription reaction for RNA isolation and cDNA synthesis from isolated epithelial cells. We additionally developed simple gene-specific pre-amplification reactions in order to improve the sensitivity of subsequent detection assays. We next evaluated the use of an alternative RNA detection method that we have also developed in our laboratory involving high-resolution melt (HRM) analysis. This approach utilizes differing melt temperatures of the mRNA biomarker amplification products, each with their own unique melt temperature. This method is more rapid and inexpensive than traditional CE detection (~2 hours to complete analysis after reverse transcription and involves the use of unlabeled primers for detection). Using a combination of the developed methods, we were able to successfully detect body-fluid specific biomarkers from as few as 10 isolated epithelial cells. We also began an evaluation of several approaches to combined DNA/RNA profiling strategies.

In summary, while some of the developed approaches require additional optimization and validation, they successfully demonstrate our ability to 'get more (genetic information) from less' (a fundamental aim of the current work).

B. Implications for policy and practice

Interpretation of admixed STR profiles from multiple donors present a complex challenge for forensic laboratories. Few studies have been carried out to aid the scientific community in the physical de-convolution of non-distinguishable cell type mixtures. The proposed work sought to help the scientific community in this regard by developing a novel strategy to isolate individual epithelial or leukocyte cells from admixed samples using laser capture micro-dissection. This will permit the both the positive identification of the cell type by mRNA expression profiling and the individual by STR profiling. Statistical analysis will be simplified and strengthened by requiring only single source profile estimates instead of likelihood ratio or combined probability of exclusion calculations. Such an approach, should the testing laboratory choose to employ it in a particular case, is expected to yield more probative information from evidentiary items from cases that contain biological material from more than one person.

C. Implications for further research

Additional studies need to be carried out to determine the efficacy of alternative binomial sampling strategies to optimize information recovery while providing flexibility when processing cases that may have different numbers of available cells for sub-sampling. We can specify the binomial sampling as an m x n matrix where m = the number of sub-samples removed and n = the size of each sub-sample. Thus our described strategy of removing ten sub-samples of size 20 would be described as a 10 x 20 binomial sampling. Although this proof of concept paper used n = 20, other sample sizes (e.g. 5-15, >20, including a range of differing sub-sample sizes from the same stain) as well as different numbers of sub-samples (e.g. m = 5-9) should be tested using the same LCM recovery and typing strategy. Optimal sampling parameters (m, n) should be determined for DNA mixtures incorporating weight ratios different from the (albeit worse-case scenario) 1:1 studied here, as well as those with more than 2 donors. Lastly, the number of jointly examined sub-samples necessary to compute an optimal joint likelihood function should be empirically determined for different values of the m and n parameters.

We believe that the preliminary work described here has implications for forensic practice that go beyond the specialized use of binomial sampling/LCM for mixture genotype inference. The binomial strategy relies on the deliberate generation of variation in the proportion of the two different cell constituents comprising the mixture. The same effect can be obtained and exploited in routine casework much more easily than with LCM. There are two different ways this could be done in the routine setting.

Firstly, dried physiological stain mixtures are not homogenous due to (i) imperfect mixing of the constituents during the physical act that caused the formation of the mixture and (ii) wicking and other physical effects due to the drying process. Thus multiple micro-geographical samplings of such stains could result in sub-samples with different component proportions. We are aware that this practice represents a different paradigm to that currently

used, but it may be worth considering for single discrete stains known (or strongly suspected to be) mixed. Initial testing of the efficacy of the micro-geographical sampling hypothesis will be described below.

A second alternative approach to generating variation in cell composition could involve a standard DNA extraction of a mixed stain, followed by preparation of sub-samples by limiting dilution (an very preliminary evaluation provided in this report). These 'Poisson' random aliquot samples would be expected to contain varying proportions of the starting constituents, just as in binomial sampling. However, whereas the binomial approach uses costly LCM to dissect out a fixed number of cells, Poisson sampling would employ inexpensive pipetting to sample a variable (Poisson distributed) number of cells in each aliquot.

Additionally, while our initial results demonstrate the ability to determine the cell type of origin of isolated epithelial cells using RNA profiling, further optimization and validation of these techniques will be needed in order to improve sensitivity. However, we are hopeful that with continued development of our novel HRM RNA body fluid identification assays will result in a sensitive, robust and cost-effective for cell type identification of single or few epithelial cells to aid in the individualization of individual cells in admixed samples.

V. REFERENCES

[1] M. Bill, P. Gill, J. Curran, T. Clayton, R. Pinchin, M. Healy, J. Buckleton, PENDULUM--a guideline-based approach to the interpretation of STR mixtures, Forensic Sci Int. 148 (2005) 181-189.

[2] I.W. Evett, C. Buffery, G. Willott, D. Stoney, A guide to interpreting single locus profiles of DNA mixtures in forensic cases, J Forensic Sci Soc. 31 (1991) 41-47.

[3] I.W. Evett, P.D. Gill, J.A. Lambert, Taking account of peak areas when interpreting mixed DNA profiles, J Forensic Sci. 43 (1998) 62-69.

[4] P. Gill, R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker, J. Buckleton, Interpreting simple STR mixtures using allele peak areas, Forensic Sci Int. 91 (1998) 41-53.

[5] C. Ladd, H.C. Lee, N. Yang, F.R. Bieber, Interpretation of complex forensic DNA mixtures, Croat.Med J. 42 (2001) 244-246.

[6] B.S. Weir, C.M. Triggs, L. Starling, L.I. Stowell, K.A. Walsh, J. Buckleton, Interpreting DNA mixtures, J Forensic Sci. 42 (1997) 213-222.

[7] P. Gill, A.J. Jeffreys, D.J. Werrett, Forensic application of DNA 'fingerprints', Nature. 318 (1985) 577-579.

[8] J. Chen, L. Kobilinsky, D. Wolosin, R. Shaler, H. Baum, A physical method for separating spermatozoa from epithelial cells in sexual assault evidence, J Forensic Sci. 43 (1998) 114-118.

[9] Y. Fujita, S. Kubo, Application of FTA technology to extraction of sperm DNA from mixed body fluids containing semen, Leg.Med.(Tokyo). 8 (2006) 43-47.

[10] A.M. Garvin, Filtration based DNA preparation for sexual assault cases, J Forensic Sci. 48 (2003) 1084-1087.

[11] D. Di Martino, G. Giuffre, N. Staiti, A. Simone, G. Sippello, G. Tuccari, L. Saravo, LMD as a Forensic Tool in a Sexual Assault Casework: LCN DNA Typing to Identify the Responsible, Progress in Forensic Genetics. 11 (2006) 571-573.

[12] K. Elliott, D.S. Hill, C. Lambert, T.R. Burroughes, P. Gill, Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides, Forensic Sci Int. 137 (2003) 28-36.

[13] C.T. Sanders, N. Sanchez, J. Ballantyne, D.A. Peterson, Laser microdissection separation of pure spermatozoa from epithelial cells for short tandem repeat analysis, J Forensic Sci. 51 (2006) 748-757.

[14] T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Sci Int. 91 (1998) 55-70.

[15] Y. Torres, I. Flores, V. Prieto, M. Lopez-Soto, M.J. Farfan, A. Carracedo, P. Sanz, DNA mixtures in forensic casework: a 4-year retrospective study, Forensic Sci Int. 134 (2003) 180-186.

[16] D.J. Lu, H.L. Lu, Y.C. Chen, [Interpretation of DNA profiles for mixed stains], Fa.Yi.Xue.Za Zhi. 18 (2002) 185-188.

[17] M.W. Perlin, B. Szabady, Linear mixture analysis: a mathematical approach to resolving mixed DNA samples, J Forensic Sci. 46 (2001) 1372-1378.

[18] S. Shrestha, S.A. Strathdee, K.W. Broman, M.W. Smith, Unknown biological mixtures evaluation using STR analytical quantification, Electrophoresis. 27 (2006) 409-415.
[19] T. Wang, N. Xue, J.D. Birdwell, Least-square deconvolution: a framework for interpreting short tandem repeat mixtures, J Forensic Sci. 51 (2006) 1284-1297.
[20] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures, Forensic Sci Int. 160 (2006) 90-101.
[21] C.S. Tomsey, M. Kurtz, B. Flowers, J. Fumea, B. Giles, S. Kucherer, Case work guidelines and interpretation of short tandem repeat complex mixture analysis, Croat.Med.J. 42 (2001) 276-280.
[22] A. Edwards, A. Civitello, H.A. Hammond, C.T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats, Am.J Hum.Genet. 49 (1991) 746-756.
[23] G. Stolovitzky, G. Cecchi, Efficiency of DNA replication in the polymerase chain reaction, Proc.Natl.Acad.Sci U.S.A. 93 (1996) 12947-12952.
[24] M.W. Perlin, G. Lancia, S.K. Ng, Toward fully automated genotyping: genotyping microsatellite markers by deconvolution, Am.J Hum.Genet. 57 (1995) 1199-1210.
[25] A.W.F. Edwards, Likelihood,Expanded Ed., Johns Hopkins University, Baltimore, 1992.
[26] J.B. Kadane, Principles of Uncertainty, Chapman & Hall, Boca Raton, FL, 2011.
[27] R.G. Cowell, S.L. Lauritzen, J. Mortera, A gamma bayesian network for DNA mixture analysis, Bayesian Anal. 2 (2007) 333-348.
[28] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PLoS.One. 4 (2009) e8327.
[29] B.W. Lindgren, Statistical Theory,Fourth ed., Chapman & Hall, New York, 1993.
[30] W. Feller, An Introduction to Probability Theory and Its Applications,Third ed., John Wiley & Sons, New York, 1968.
[31] D.V. Lindley, A.F.M. Smith, Bayes estimates for the linear model., Journal of the Royal Statistical Society, Series B. 34 (1972) 1-41.
[32] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika. 57 (1970) 97-109.
[33] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J Chem Phys. 21 (1952) 1087-1092.
[34] C.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists,Second ed., John Wiley & Sons, Chicester, UK, 2004.
[35] I.J. Good, Probability and the Weighing of Evidence, Griffin, London, 1950.
[36] M.W. Perlin, Explaining the likelihood ratio in DNA mixture interpretation, 21st International Symposium on Human Identification (Promega).2010).
[37] E. Essen-Moller, Die Biesweiskraft der Ahnlichkeit im Vater Schaftsnachweis, Mitteilungen der anthropologischen Gesellschaft in Wien. 68 (1938) 9-53.
[38] I.J. Good, Studies in the history of probability and statistics. XXXVII. A.M. Turing's statistical work in World War II., Biometrika. 66 (1979) 393-396.
[39] S. Singh, The Code Book: The science of Secrecy from Ancient Egypt to Quantum Cryptography, Doubleday, New York, 1999.

[40] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci Int. 64 (1994) 125-140.
[41] A.E. Gelfand, Smith AFM, Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association.(1990) 125-140.
[42] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, B.W. Duceman, Validating TrueAllele((R)) DNA Mixture Interpretation, J Forensic Sci. 56(6) (2011) 1430-1447.
[43] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, Anal.Biochem. 162 (1987) 156-159.
[44] C. Haas, B. Klesser, C. Maake, W. Bar, A. Kratzer, mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR, Forensic Sci Int Genet. 3 (2009) 80-88.
[45] J. Juusola, J. Ballantyne, Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification, Forensic Sci Int. 135 (2003) 85-96.
[46] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci Int. 152 (2005) 1-12.
[47] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, J Forensic Sci. 52 (2007) 1252-1262.
[48] S. Rozen, H.J. Skaletsky, Bioinformatics Methods and Protocols: Methods in Molecular Biology, Humana Press, Totowa, NJ, 2000.
[49] C. Haas, E. Hanson, A. Kratzer, W. Bar, J. Ballantyne, Selection of highly specific and sensitive mRNA biomarkers for the identification of blood, Forensic Sci.Int.Genet. 5 (2011) 449-458.
[50] C. Haas, E. Hanson, M.J. Anjos, R. Banemann, A. Berti, E. Borges, A. Carracedo, M. Carvalho, C. Courts, C.G. De, M. Dotsch, S. Flynn, I. Gomes, C. Hollard, B. Hjort, P. Hoff-Olsen, K. Hribikova, A. Lindenbergh, B. Ludes, O. Maronas, N. McCallum, D. Moore, N. Morling, H. Niederstatter, F. Noel, W. Parson, C. Popielarz, C. Rapone, A.D. Roeder, Y. Ruiz, E. Sauer, P.M. Schneider, T. Sijen, Court DS, B. Sviezena, M. Turanska, A. Vidaki, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from human saliva and semen stains--results of a third collaborative EDNAP exercise, Forensic Sci Int Genet. 7 (2013) 230-239.
[51] E.K. Hanson, J. Ballantyne, Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations, Sci Justice. 53 (2013) 14-22.
[52] A.D. Roeder, C. Haas, mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification, Int J Legal Med. 127 (2013) 707-721.
[53] E. Hanson, C. Haas, R. Jucker, J. Ballantyne, Identification of skin in touch/contact forensic samples by messenger RNA profiling, Forensic Sci Int Genet.Suppl Series. 3 (2011) e305-e306.
[54] E. Hanson, C. Haas, R. Jucker, J. Ballantyne, Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence, Forensic Sci Int Genet. 6 (2012) 548-558.
[55] J. Ballantyne, E.K. Hanson, M.W. Perlin, DNA mixture genotyping by probabilistic computer interpretation of binomially-sampled laser captured cell populations: combining quantitative data for greater identification information, Sci Justice. 53 (2013) 103-114.

[56] M. Perlin, Combining DNA evidence for greater match ifnormation., Forensic Sci Int Genet Supp Ser. 3 (2011) e510-e511.
[57] M. Alvarez, J. Juusola, J. Ballantyne, An mRNA and DNA co-isolation method for forensic casework samples, Anal.Biochem. 335 (2004) 289-298.
[58] M. Setzer, J. Juusola, J. Ballantyne, Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains, J Forensic Sci. 53 (2008) 296-305.
[59] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat.Rev.Genet. 10 (2009) 57-63.
[60] J.A. Bright, S.F. Petricevic, Recovery of trace DNA and its application to DNA profiling of shoe insoles, Forensic Sci Int. 145 (2004) 7-12.
[61] M. Goray, E. Eken, R.J. Mitchell, R.A. van Oorschot, Secondary DNA transfer of biological substances under varying test conditions, Forensic Sci Int Genet. 4 (2010) 62-67.
[62] D. Hall, M. Fairley, A single approach to the recovery of DNA and firearm discharge residue evidence, Sci Justice. 44 (2004) 15-19.
[63] B.C. Pang, B.K. Cheung, Double swab technique for collecting touched evidence, Leg.Med.(Tokyo). 9 (2007) 181-184.
[64] S.F. Petricevic, J.A. Bright, S.L. Cockerton, DNA profiling of trace DNA recovered from bedding, Forensic Sci Int. 159 (2006) 21-26.
[65] R.A. Wickenheiser, Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact, J Forensic Sci. 47 (2002) 442-450.
[66] C. Haas, E. Hanson, W. Bar, R. Banemann, A.M. Bento, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, A. Choma, M. Dotsch, M. Duriancikova, P. Hoff-Olsen, C. Hohoff, P. Johansen, P.A. Lindenbergh, B. Loddenkotter, B. Ludes, O. Maronas, N. Morling, H. Niederstatter, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezena, L. Zatkalikova, J. Ballantyne, mRNA profiling for the identification of blood--results of a collaborative EDNAP exercise, Forensic Sci.Int.Genet. 5 (2011) 21-26.
[67] C. Haas, E. Hanson, M.J. Anjos, W. Bar, R. Banemann, A. Berti, E. Borges, C. Bouakaze, A. Carracedo, M. Carvalho, V. Castella, A. Choma, C.G. De, M. Dotsch, P. Hoff-Olsen, P. Johansen, F. Kohlmeier, P.A. Lindenbergh, B. Ludes, O. Maronas, D. Moore, M.L. Morerod, N. Morling, H. Niederstatter, F. Noel, W. Parson, G. Patel, C. Popielarz, E. Salata, P.M. Schneider, T. Sijen, B. Sviezena, M. Turanska, L. Zatkalikova, J. Ballantyne, RNA/DNA co-analysis from blood stains--results of a second collaborative EDNAP exercise, Forensic Sci Int Genet. 6 (2012) 70-80.
[68] B. Jackson, C.M. Tilli, M.J. Hardman, A.A. Avilion, M.C. MacLeod, G.S. Ashcroft, C. Byrne, Late cornified envelope family in differentiating epithelia--response to calcium and ultraviolet irradiation, J Invest Dermatol. 124 (2005) 1062-1070.
[69] D. Marshall, M.J. Hardman, K.M. Nield, C. Byrne, Differentially expressed late constituents of the epidermal cornified envelope, Proc.Natl.Acad.Sci U.S.A. 98 (2001) 13031-13036.
[70] S. Fujimoto, H. Uratsuji, H. Saeki, S. Kagami, Y. Tsunemi, M. Komine, K. Tamaki, CCR4 and CCR10 are expressed on epidermal keratinocytes and are involved in cutaneous immune reaction, Cytokine. 44 (2008) 172-178.

[71] J. Morales, B. Homey, A.P. Vicari, S. Hudak, E. Oldham, J. Hedrick, R. Orozco, N.G. Copeland, N.A. Jenkins, L.M. McEvoy, A. Zlotnik, CTACK, a skin-associated chemokine that preferentially attracts skin-homing memory T cells, Proc.Natl.Acad.Sci U.S.A. 96 (1999) 14470-14475.
[72] S.J. Busfield, C.A. Comrack, G. Yu, T.W. Chickering, J.S. Smutko, H. Zhou, K.R. Leiby, L.M. Holmgren, D.P. Gearing, Y. Pan, Identification and gene organization of three novel members of the IL-1 family on human chromosome 2, Genomics. 66 (2000) 213-216.
[73] M.F. Nold, C.A. Nold-Petry, J.A. Zepp, B.E. Palmer, P. Bufler, C.A. Dinarello, IL-37 is a fundamental inhibitor of innate immunity, Nat.Immunol. 11 (2010) 1014-1022.
[74] S.L. Taylor, B.R. Renshaw, K.E. Garka, D.E. Smith, J.E. Sims, Genomic organization of the interleukin-1 locus, Genomics. 79 (2002) 726-733.
[75] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems (with discussion), J R Statist Soc B. 50 (1988) 157-224.
[76] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,Second, Morgan Kaufmann, San Francisco, 1988.

VI. DISSEMINATION OF RESEARCH FINDINGS

A. Publications

1. Ballantyne, J., Hanson, E.K. and Perlin, M.W. DNA mixture genotype by probabilistic computer interpretation of binomially-sampled laser captured cell populations: Combining quantitative data for greater identification information. Science and Justice 53 (2013), p. 103-114.
2. Hanson, E., Haas, C. and Ballantyne, J. "Identification of Skin in Touch/Contact Forensic Samples by Messenger RNA Profiling." Forensic Science International: Genetics Supplement Series 3 (2011), e305-306.
3. Hanson, E., Haas, C., Jucker, R. and Ballantyne, J. Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. Forensic Science International: Genetics 6(5) (2012), p. 548-558.
4. Hanson, E.K. and Ballantyne, J. Highly specific mRNA biomarkers for the identification of vaginal secretions in sexual assault investigations. Science and Justice 53(2) (2013), p. 14-22.

Manuscripts in preparation

5. *Subject:* Micro-volume PCR reactions for DNA/RNA profiling of micro-dissected cells isolated from epithelial cell mixtures

Related publications (related to Aim 3 – mRNA profiling)

1. Hanson, E., Ballantyne, J. "RNA Profiling for the Identification of the Tissue Origin of Dried Stains in Forensic Biology." Forensic Science Review 22(2) (2010), p. 145-157.
2. Haas, C., Hanson, E. and Ballantyne, J. mRNA and MicroRNA for Body Fluid Identification. In: Siegel JA and Saukko PJ (eds). *Encyclopedia of Forensic Sciences, Second Edition*., vol. 1, pp. 402-408. Waltham: Academic Press. doi: 10.1016/B978-0-12-382165-2.00069-6. (2013).
3. Hanson, E., Ballantyne, J. "Chapter 5. RNA Profiling for the Identification of the Tissue Origin of Dried Stains in Forensic Biology." In Topics on Forensic DNA Analysis – Current Practices and Emerging Technologies, Taylor & Francis, ISBN: 978-1-4665-71266. In Press. (2012).

B. Presentations

1. Research at the National Center for Forensic Science. Ballantyne, J. ESR Forensic Group, Auckland, New Zealand. 2008.
2. Research at the National Center for Forensic science. Ballantyne, J. ESR Headquarters, Wellington, New Zealand. 2008.

3. De-convolution of Body Fluid Mixtures: Cell Type Identification and Single Source Genetic Profiling of Micro-dissected Cells. Ballantyne, J and Hanson, E. The Annual NIJ Conference, Washington, DC. 2009.
4. Forensic Biology Research at the National Center for Forensic Science. EDNAP (European DNA Profiling Group) Meeting. Ballantyne, J. Kiev, Ukraine. 2010.
5. Evolution of DNA Mixture Interpretation. Ballantyne, J. DNA Mixture Interpretation Workshop organized by the National Forensic Science Technology Center. Clearwater Beach, FL 2011.
6. Forensic Biology Research at the National Center for Forensic Science. Ballantyne, J. EDNAP (European DNA Profiling Group) Meeting, Brussels, Belgium.
7. DNA Mixture De-convolution by Binomial Sampling of Individual Cells. Ballantyne, J. and Perlin, M. International Conference on Forensic Inference and Statistics (ICFIS). Seattle, WA. 2011.
8. Advanced Topics in Forensic DNA Analysis. Ballantyne, J. Forensic Science Training for Capital Defense Attorneys: Advanced Issues. Bureau of Justice Assistance and the National Clearing House for Science, Technology and the Law, Las Vegas, NC. 2012.
9. DNA Mixture Genotyping by Probabilistic Computer Interpretation of Binomially-Sampled Laser Captured Cell Populations: Combining Quantitative Data for Greater Identification Information. Ballantyne, J., Hanson, E.K. and Perlin, M.W. NIJ Conference, Crystal City, VA. 2012.
10. Enhanced Genetic Analysis of Bio-Particles Isolated from Single- and Multi-Source Touch DNA Evidence Using Micro-Volume DNA/RNA Profiling Strategies. 24th International Symposium on Human Identification (Promega), Atlanta, GA October 2013.