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

Identification of Forensically Relevant Fluids and Tissues by Small RNA Profiling

FINAL REPORT

December 16, 2013

Department of Justice, National Institute of justice
Award Number: 2009-DN-BX-K255
(1 November 2009 – 31 December 2013)

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

The serology-based methods routinely used in forensic casework for the identification of biological fluids are costly in terms of time and sample and have varying degrees of sensitivity and specificity. The use of messenger RNA (mRNA) profiling (a molecular genetics based approach) has been proposed to supplant conventional methods for body fluid identification. However, the size of the amplification products used in these mRNA assays (~100-250 nt) may not be ideal for use with degraded or compromised samples frequently encountered in forensic casework. There has been an explosion of interest in a class of small non-coding RNAs, microRNAs (~20-25 bases in length), with numerous published studies reporting that some miRNAs are expressed in a tissue specific manner. Additionally, another novel class of small regulatory RNAs, piwi-interacting RNAs (~26-30 bases in length), were initially identified only in germ-line cells (including sperm-producing cells) in mammals. We therefore theorized that it may be possible that this class of small RNAs could be ideally suited for the identification of semen.

Through the current study, we sought to: 1) determine whether novel body fluid and tissue specific small RNAs, such as microRNAs or piRNAs, can be identified that would permit the identification of forensically relevant biological fluids and tissues (i.e. blood, semen, saliva, vaginal secretions, menstrual blood and skin), 2) demonstrate the suitability of the small RNA profiling assays for use with forensic casework (performance with low template samples, admixed body fluid samples, and also mock casework samples), and 3) compare other body fluid identification strategies (mainly mRNA profiling) in order to determine the most suitable strategy for the analysis of environmentally impacted or degraded samples frequently encountered in forensic casework.

We performed an extensive comprehensive screening of ~1,200 microRNAs (through miRBase v.16) and have identified a panel of 18 differentially expressed miRNAs that permit the identification of all forensically relevant body fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin). An initial validation of assays incorporating these miRNA biomarkers was performed which included an evaluation of specificity, sensitivity, and performance with admixed body fluid samples. We compared the ability to accurately identify the body fluid of origin of compromised or environmentally impacted body fluid samples using miRNA and mRNA profiling. The analysis (i.e. determination of the body fluid of origin) of the validation samples was aided by a novel statistical approach we developed for the analysis of normalized miRNA express data involving the use of logistic regression models. This permits an objective “prediction” of the presence of the above-mentioned body fluids and tissues using a multi-candidate approach rather than the two-candidate approach previously used. While at this time it cannot be conclusively determined whether miRNA profiling is a superior to mRNA profiling for the analysis of environmentally impacted samples, the results of our work demonstrate the ability to successful use a miRNA profiling approach for the identification of the body fluid or tissue origin of forensic biological evidence with picogram-level sensitivity.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

1. The serology-based methods routinely used in forensic casework for the identification of biological fluids are costly in terms of time and sample and have varying degrees of sensitivity and specificity. Additionally, none of the routinely used serological or immunological tests available to forensic crime laboratories permit an identification of some forensically relevant biological fluids and tissues such as menstrual blood, vaginal secretions and skin.

2. The inability to positively confirm the presence of certain biological fluids, the consumption of valuable samples and the time and labor required has resulted in a trend to bypass conventional body fluid identification methods and proceed straightaway to the analysis of DNA present in forensic samples. However, a disadvantage of the disuse of body fluid identification methods is that the identification of the biological material present might be crucial to the investigation and prosecution of the case

3. The routine use of body fluid identification methods prior to DNA analysis awaits the development of suitable molecular genetics based methods that are fully compatible with the current DNA analysis pipeline. In order for any new body fluid assay to be suitable for forensic casework it must demonstrate a high degree of specificity for each body fluid, permit parallel analysis of the different biological fluids, be completed in a timely and labor efficient manner and must be sufficiently sensitive.

4. One such molecular-based approach is the use of mRNA profiling. Terminally differentiated cells, whether they comprise blood monocytes or lymphocytes, ejaculated spermatozoa, epithelial cells lining the oral cavity or epidermal cells from the skin become such during a developmentally regulated program in which certain genes are turned off (i.e. are transcriptionally silent) and others are turned on (i.e. are actively transcribed and translated into protein). This produces a pattern of gene expression that is unique to each cell type not only evidenced by the specific mRNAs present but also their relative abundance. Thus if the type and abundance of mRNAs in a stain or tissue sample recovered at the crime scene could be determined, then a definitive identification the tissue or body fluids present could be made. Such an approach would offer a number of advantages over conventional methods for body fluid identification including: (i) the ability to perform parallel tests for numerous markers of a single body fluid in a single assay format, (ii) the ability to perform parallel tests for different body fluids in a single assay format, (iii) a definitive identification of body fluids for which presently no specific tests exist.

5. Although mRNA profiling assays appear to be suitable for use with forensic samples, the size of the amplification products used in these assays (~200-300 nt) may not be ideal for use with degraded or compromised samples frequently encountered in forensic casework.

6. There has been an explosion of interest in a class of small non-coding RNAs, microRNAs, whose regulatory functions in various developmental and biological processes have been identified. Several studies have examined the relative abundance of miRNAs in human tissue with numerous miRNAs reported to be tissue-specific.

7. We hypothesized it should be possible to identify miRNA species that, due to differential tissue expression, could be used to identify the body fluid origin of forensic biological stains with a high degree of sensitivity and specificity. Additionally, the use of miRNA profiling assays may improve the analysis of degraded or compromised forensic samples due to their significantly smaller size (~20-30 nt).

8. The proposed study sought to develop a novel strategy for the identification of the body fluid or tissue origin of dried forensic stains using small RNA profiling. This study firstly sought to identify candidate small RNAs, including microRNAs and piRNAs, to permit the identification of forensically relevant biological fluids (blood, semen, saliva, vaginal secretions, menstrual blood, and skin). Secondly the study sought to evaluate the suitability of the small RNA profiling assays for use with forensic casework through an examination of their sensitivity, their ability to identify multiple fluids in an admixed sample and their ability to identify the body fluid or origin of mock and *bona fide* casework samples. Lastly, an initial direct comparison of existing body fluid identification strategies (mainly small RNA and mRNA profiling) was conducted in order to provide for the first time a comprehensive assessment of the most optimal body fluid identification strategy for use in forensic casework.

9. In preliminary previous work, we examined the expression of 452 miRNAs (~67% of the known miRNA*ome* at that time) in dried, forensically relevant biological fluids (blood, semen, saliva, vaginal secretions, and menstrual blood) in an attempt to find body-fluid specific miRNA candidates. MiRNA profiling assays for the identification of blood, semen, saliva, vaginal secretions and menstrual blood were developed using as little as 50 – 500 pg for detection (blood - miR451, miR16; semen - miR10b, miR135b; saliva - miR658, miR205; vaginal secretions - miR124a, miR372; menstrual blood - miR451, miR412). Using the two-candidate scatterplot assays, distinct clustering of miRNA expression data (U6b-normalized) from each body fluid of interest was observed. The degree of separation between the body fluid of interest and the other body fluids varied between the assays and was dependent upon the strength of the selected miRNA candidates. Therefore, while these assays demonstrated a degree of specificity for their respective body fluids and little cross-reactivity with twenty-one other human tissue samples, additional work was required in order to improve the specificity of each of the body fluid assays.

10. The preliminary screening work involved only miRNAs available at that time (miRBase v 1.0). However, with subsequent releases of miRBase, hundreds of novel miRNAs were deposited. We therefore performed additional extensive screening of available miRNAs through miRBase v.16 (364 novel candidates through v.11; 30 novel candidates through v.13-v.14; 352 novel candidates through v.15-v.16 – total of 746). With the additional

746 miRNAs, we in total evaluated the expression 1,198 miRNAs in forensically relevant biological fluids: blood, semen, saliva, vaginal secretions and menstrual blood.

11. With the increasing frequency of touch DNA evidence, there is a critical need to definitively identify the tissue source of origin of biological material in touch DNA evidence (i.e. skin). We therefore also included an evaluation of miRNA expression in human skin in addition to the forensically relevant biological fluids. The expression of all 1,198 miRNAs was evaluated in human skin samples.

12. All miRNA analysis was performed using the miScript system (QIAGEN). During reverse transcription, miRNAs are simultaneously poly-adenylated since they do not naturally contain a poly(A) tail. Reverse transcription is performed using an oligo-dT primer as well as random primers. The oligo-dT primer has a universal tag sequence at the 5'end that serves as a universal primer binding site for subsequent real time PCR assays using the miScript system. The miScript system provides the ability to analyze all RNA species in total RNA samples, as they will all be reverse transcribed. Additionally unlike other miRNA assays, which involve the use of miRNA specific RT reactions (one miRNA per reaction), numerous miRNAs can be profiled from a single miScript RT reaction.

13. After an extensive evaluation of the expression data for the 1,198 miRNAs as well as extensive assay optimization, we successfully identified miRNA biomarkers for each of the forensically relevant body fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin) and developed suitable two-candidate profiling assays for the identification of each body fluid and tissue.

14. The previous nine miRNA body fluid identification panel was expanded to a 19 miRNA panel (18 miRNAs plus one "housekeeping-like" miRNA) which included five of the original miRNAs as well as 13 novel miRNAs. The miRNAs in the new body fluid identification panel included the following: miR451, miR16 (blood); miR891b, miR892a (semen); miR658, miR205, miR124* (saliva); miR124a, miR1280, miR4286 (vaginal secretions); miR144, miR144*, miR185, miR142-3p (menstrual blood); miR139, miR494, miR455-3p, miR3169 (skin). Two-candidate scatterplots were constructed for each body fluid using normalized expression data (miR940) from the identified candidates which resulted in distinct clustering of the body fluids, with the body fluid of interest separated from the other body fluid types. For body fluids with more than two identified biomarkers, several 2D scatterplot assays were developed: saliva – miR658/miR205, miR658/miR124a*; vaginal secretions – miR124a/miR1280, miR124a/miR4286; menstrual blood - ; skin – miR139/miR494, miR455-3p/miR3169.

15. The initial development of the miRNA profiling assays involved the use of various 2D (two-candidate) scatterplot assays. Body fluid identification using these assays was somewhat subjective with a visual determination of whether a data point was within the cluster of known body fluid samples. The limits of the body fluid cluster were not clearly defined making interpretation of results quite challenging. Therefore, more refined analysis interpretation metrics were needed in order to provide a robust objective

statistical approach for declaring a sample to be positive or negative with respect to the presence of a particular body fluid. In an attempt to provide such interpretation metrics, we developed novel logistic regression (LogR) statistical models that provide accurate predictions of the presence of forensically relevant fluids. Additionally, the developed logR models permit the use of multi-candidate models instead of the more limited two-candidate scatter plots which adds multi-dimensional capabilities that can further improve the specificity of each assay. During model development it was determined that two miRNAs (miR494 (skin) and 142-3p (menstrual blood)) did not contribute significantly to the specificity of the models and were therefore removed from the body fluid identification panel (reduced to 16 body fluid miRNA candidates and miR940 for normalization).

16. The performance of all models was evaluated for the selected cases (samples that were used to construct the logistic models) and the unselected cases (samples that were not used to construct the logistic models. Among the unselected cases, the success rate of body fluid identification was 100% for all models (i.e. all body fluids). Significantly the rate of false positives (i.e. the wrongful identification of a body fluid) for all the models was 0% thus demonstrating the high degree of specificity of the miRNA profiling assays.

17. The miRNA profiling assays (including results interpretation using the developed LogR models) were tested and evaluated to assess sensitivity and performance with admixed body fluid samples and environmentally impacted and compromised samples (blood, semen, saliva and vaginal secretions only). A comparison of the profiling success for the environmentally impacted samples using both miRNA and mRNA profiling assays was performed.

18. The overall results of the sensitivity study demonstrate the high sensitivity of the developed miRNA assays. For each of the developed miRNA assays picogram-level sensitivity was obtained (blood 10-50pg; semen 12.5-250pg; saliva – 2.5-250pg; vaginal – 2.5-250pg; menstrual blood – 2.5-250pg; skin - 2.5-12.5pg). This high level of sensitivity, which is one to two orders of magnitude high than standard CE-based mRNA assays, could be advantageous in forensic casework analysis where a limited amount of starting material is often available for analysis.

19. The ability to identify the presence of body fluids in admixed samples using the miRNA panels was evaluated. Twenty-four mixture samples were tested: blood-semen, blood-saliva, saliva-semen, vaginal secretions-semen, menstrual blood-semen, and vaginal secretions-semen. For each, the first body fluid listed was a dried stain to which the second body fluid in liquid form was placed on the dried stain thus creating admixed body fluid samples. For the dried stains, blood samples were 50µl bloodstains and for all other body fluids ½ swabs were used. For the liquid body fluid added, we evaluated four different volumes (50µl, 25µl, 10µl and 5µl). Therefore for each body fluid mixture (6 types), 4 samples each with decreasing amounts of the liquid body fluid were evaluated. While there were some false negatives, the results of the mixture study indicate that the presence of multiple body fluids can be identified using the developed miRNA assays. Significantly, no false positive results were observed.

20. A hypothesis tested was that miRNA profiling assays would be more suitable for use with environmentally compromised and degraded samples compared to mRNA profiling due to the much smaller size of miRNAs. We evaluated a significant number of challenging or environmentally compromised samples for blood, semen, saliva and vaginal secretions using both our developed miRNA assays and our current gold standard CE mRNA body fluid identification multiplex (12 biomarker system; 2 biomarkers per body fluid; see Methods for list of included biomarkers). Body fluid samples were exposed to various conditions including the following: room temperature storage (1 – 2 yrs), exposure to 37ºC and 56ºC temperatures (1 – 2 yrs), outside environments (heat, light, humidity, direct and indirect sunlight, rain), and exposure to UV light (> 1yr). For blood and vaginal secretions, similar overall success rates were observed for both the miRNA and mRNA profiling assays (blood – 77% for miRNA, 73% mRNA; vaginal secretions – 32% for both miRNA and mRNA assays). For semen, there was an increase in profiling success using mRNA profiling (86%) compared to miRNA profiling (55%). However, for saliva, a significant improvement in profiling success was obtained using miRNA profiling (83%) compared to mRNA profiling (52%). As with most environmental studies, body fluids were often not identified in the samples exposed to outside conditions (heat, sunlight, humidity and rain).

21. For the purpose of the environmentally compromised study described above, we classified a mRNA result as positive if one or both biomarkers were detected. However, it is not clear whether this will indeed be an accurate assessment of profiling success. Currently our mRNA multiplex system contains two RNA biomarkers for each body fluid and it has not yet been determined if a positive result would require detection of both biomarkers. If in fact the presence of both biomarkers is needed for a definitive result, then the miRNA assays would result in successful body fluid identification at a much higher rate compared to the mRNA assays (results for semen are similar even with two marker mRNA detection, but still a slight improvement using miRNA assays).

22. As a result of the above described evaluation of environmentally compromised samples, at this time it cannot be conclusively determined whether miRNA profiling is a superior to mRNA profiling for the analysis of environmentally impacted samples as truly environmentally impacted samples (exposed to outside conditions – heat, light, humidity and rain) may be largely affected by other factors such as sample loss (e.g. rain, insect activity) rather than just degradation of the biological material present. Future work will include and evaluation of artificially degraded RNA samples (RNase digestions) in order to assess the effects of only RNA degradation to determine if miRNA profiling will be more successful for the analysis of these samples. The results of the environmentally impacted samples provide support to the routine use of RNA (either miRNA or mRNA) profiling assays in forensic casework.

23. We also evaluated an additional novel class of small regulatory RNAs, piwi-interacting RNAs (piRNAs) as potential semen markers. piRNAs are involved in germline development, silencing of transposable DNA elements and in maintaining germline DNA integrity. Piwi-interacting RNAs initially had been reported to be present only in sperm-producing cells in mammals and therefore could be extremely specific to semen. We

evaluated the expression of a number of piRNAs (~35 piRNAs, 17 of which should only be transcribed from the Y chromosome) in forensically relevant body fluids as well as prostate and testes tissues. While we did not observe semen-specific expression of the small number of piRNAs tested, a more comprehensive large-scale screening of known piRNAs may permit an identification of suitable candidates for the identification of semen.

24. In conclusion, we have successfully developed novel highly sensitive and specific miRNA profiling assays for the identification of blood, semen, saliva, vaginal secretions, menstrual blood and skin. We have demonstrated their suitability for use with forensic samples through their performance with admixed body fluid samples as well as environmentally compromised samples. We have developed novel statistical methods (logistic regression models) for improved data analysis and interpretation of results. The results of this work support the potential use of miRNA profiling for the accurate and definitive identification of the body fluid or tissue origin of forensic samples.

I. INTRODUCTION

A. Statement of the Problem

The serology-based methods routinely used in forensic casework for the identification of biological fluids are costly in terms of time and sample and have varying degrees of sensitivity and specificity. Additionally, none of the routinely used serological or immunological tests available to forensic crime laboratories permit an identification of some forensically relevant biological fluids and tissues such as menstrual blood, vaginal secretions and skin. Recently the use of a molecular genetics based approach using messenger RNA (mRNA) profiling has been proposed to supplant conventional methods for body fluid identification. However, the size of the amplification products used in these mRNA assays (~100-250 nt) may not be ideal for use with degraded or compromised samples frequently encountered in forensic casework. Therefore, through the current study, we sought to determine whether small RNAs (miRNAs/piRNAs) profiling would be a suitable approach for the identification of all forensically relevant biological fluids and tissues and whether these assays would offer an advantage over mRNA profiling for the analysis of environmentally impacted body fluid samples.

B. Literature Review

In the past, standard practice in forensic casework analysis typically included a preliminary screening of evidentiary items recovered during the investigation of criminal offenses in order to identify the presence, and possible tissue origin, of biological material [1]. The presence of biological material such as blood, semen and saliva stains can indicate the location of potential sources of DNA that, once recovered, could be used to identify the donor of the biological material. Typically, conventional methods for body fluid stain analysis are carried out in a serial manner, with a portion of the stain being tested for only one body fluid at a time [1]. Frequently multiple tests are required to first presumptively identify the presence of biological fluids followed by additional testing in order to confirm the presence of the fluid or identify the species of origin. Therefore these methods are costly not only in the time and labor required for their completion, but in some instances also in terms of the amount of sample consumed during the performance of each assay. While these conventional methods can confirm the presence of human blood and semen, none of the routinely used serological and immunological tests can definitely identify the presence of human saliva or vaginal secretions. With the large volume of cases that operational crime laboratories are faced with processing every year, a significant amount of the total time spent on an individual case can be devoted solely to the screening of evidentiary items for the presence of biological materials. The inability to positively confirm the presence of certain biological fluids, the consumption of valuable samples and the time and labor required has resulted in a trend to bypass conventional body fluid identification methods and proceed straightaway to the analysis of DNA present in forensic samples. Proponents of this approach argue that the recovery of human DNA from evidentiary items directly indicates the presence of human biological material and thereby eliminates the need for conventional body fluid testing.

There are several disadvantages to bypassing the body fluid identification step during bio-molecular forensic analysis [2]. First, the analytical methods used to analyze DNA are

considerably more expensive than basic serological testing. Therefore the use of DNA analysis as a means to identify the presence of human biological material might be difficult to justify from a budgetary standpoint. Additionally, a smaller number of samplings than otherwise might be prudent from an individual piece of evidence containing multiple stains may be collected in an attempt to reduce the associated cost of analysis. Critical evidence may be missed using this type of approach that might have been identified using a larger preliminary screen with basic serological methods. A second disadvantage of the disuse of body fluid identification methods is that the identification of the biological material present might be crucial to the investigation and prosecution of the case [2]. For example, consider a sexual assault of a child with a step-father suspect where the step-father's DNA profile was recovered from samples taken from the child's underwear and bedding. The step-father could argue that the source of the DNA was from his skin cells deposited from casual and frequent contact with the child's clothing and bedding. However, the finding that his DNA originated from a semen stain and not skin cells would be more problematic for him to explain away and would more strongly support the allegation of a sexual assault. Another example of a case demonstrating the importance of identifying the body fluid could be that DNA from a sexual assault victim is found in a suspect's vehicle and the suspect claims it was present due to casual contact since the victim had ridden in his car numerous times. However, the significance of this evidence would increase if the source of the DNA could be shown instead to originate from the victim's vaginal secretions, a circumstance which would be more difficult to attribute to casual as opposed to sexual contact. The routine use of body fluid identification methods prior to DNA analysis awaits the development of suitable molecular genetics based methods that are fully compatible with the current DNA analysis pipeline. In order for any new body fluid assay to be suitable for forensic casework it must demonstrate a high degree of specificity for each body fluid, permit parallel analysis of the different biological fluids, be completed in a timely and labor efficient manner and must be sufficiently sensitive.

Possible alternatives to the currently used body fluid identification methods have been proposed: Raman spectroscopy and messenger RNA (mRNA) profiling [2-8]. The use of Raman spectroscopy, while currently used in forensic laboratories for the identification of trace evidence materials (drugs, fibers, paint, ink) [9-20], has not been extensively investigated for use with biological samples [8]. While it shows promise for future use as it is a nondestructive method, there needs to be additional studies performed to demonstrate the reproducibility of the method (i.e. inter-sample and inter-individual variation) and the ability to detect and correctly identify multiple biological fluids in an admixed sample. Messenger RNA profiling of tissue specific gene transcripts for the identification of dried body fluid stains has been reported and appears to satisfy most of the above mentioned criteria for an integrated, current-technology-compatible body fluid identification method [2,5,6].

The mRNA in aged and compromised dried stains appears to be sufficiently stable for forensic analysis [21]. However, as with DNA, heat and humidity is detrimental to RNA stability and results in a time dependent fragmentation of the polynucleotide chain [21]. Typically forensic assays employ some biomarkers whose amplicon sizes are $\geq$ 250 bases which results in amplification failure when highly degraded samples are encountered [22]. Thus reduced size amplicons for STR and mitochondrial DNA profiling methods are being increasingly used for the analysis of degraded samples [23-32]. Similarly smaller amplicons could be designed for use in mRNA based forensic assays although they may present additional technical assay design challenges because of the need to ensure that contaminating genomic

DNA does not confound the analyses. In theory, another way to reduce the amplicon size would be to employ short RNA biomarkers instead of mRNA.

There has been an explosion of interest in a class of small non-coding RNAs, microRNAs, whose regulatory functions in various developmental and biological processes have been identified [33-50]. The role of miRNAs in various cancers and diseases are also being revealed [51-62]. Several studies have examined the relative abundance of miRNAs in human tissue with numerous miRNAs reported to be tissue-specific [63-73]. We hypothesize it should be possible to identify miRNA species that, due to differential tissue expression, could be used to identify the body fluid origin of forensic biological stains with a high degree of sensitivity and specificity.

The proposed study sought to develop a novel strategy for the identification of the body fluid or tissue origin of dried forensic stains using small RNA profiling. This study firstly sought to identify candidate small RNAs, including microRNAs and piRNAs, to permit the identification of forensically relevant biological fluids (blood, semen, saliva, vaginal secretions, menstrual blood, and skin). Secondly the study sought to evaluate the suitability of the small RNA profiling assays for use with forensic casework through an examination of their sensitivity, their ability to identify multiple fluids in an admixed sample and their ability to identify the body fluid or origin of mock and *bona fide* casework samples. Lastly, an initial direct comparison of existing body fluid identification strategies (mainly small RNA and mRNA profiling) was conducted in order to provide for the first time a comprehensive assessment of the most optimal body fluid identification strategy for use in forensic casework.

C. Statement of Hypothesis or Rationale for the Research

The currently available serological and immunological tests used in forensic casework for body fluid identification are costly not only in terms of the time and labor required for their completion, but also in the amount of samples consumed. Additionally, none of these routinely used serological and immunological tests can definitively identify the presence of human saliva, vaginal secretions, menstrual blood and skin. Additionally, the samples encountered in forensic casework will include degraded and environmentally impacted samples. The analysis of these compromised and challenging samples may be difficult due to the presence of degraded nucleic acids or the limited amount of biological material present after environmental insults (e.g. humidity and rain). It is a responsibility of the forensic science community to provide suitable methods for the analysis of all biological evidence in order to assist in the resolution of criminal investigations. Therefore, the goal of this project was to determine if small RNA profiling assays could be developed in order to provide highly specific and sensitive assays to identify the presence of all forensically relevant fluids and that may be used for the analysis of environmentally impacted body fluid samples.

II. METHODS

Preparation of Body Fluid Stains

Body fluids 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. 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 they were dried onto sterile cotton swabs. Saliva samples were provided in sealed plastic tubes and stored frozen until they were dried onto sterile cotton swabs. Buccal samples 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. Skin samples consisted of total RNA from human skin obtained from commercial sources (Biochem, Hayward, CA; Zenbio, Research Park Triangle, NC; Strategene/Agilent, Santa Clara, CA; Zyagen, San Diego, CA) and surface swabs of various touched objects, surfaces and human skin (swabbed using a sterile swab moistened with nuclease free water). Prostates and testes total RNA (three-donor pooled samples, certified to contain small RNAs) preparations were obtained from the FirstChoice Human Total RNA Survey Panel (Ambion/Life Technologies, Grand Island, NY). All samples were stored at -20^{o}C until needed (except for tissue total RNA (skin, prostate, testes) which was stored at -47^{o}C). A 50µl stain or a single cotton swab was used for RNA isolation unless otherwise stated.

Any modifications to the preparation of, or further treatment of, these standard single source samples (e.g. mixtures, environmentally compromised samples) will be described in detail in Section IIIA (Statement of Results).

RNA Isolation

Total RNA was extracted from blood, semen, saliva, vaginal secretions and menstrual blood and skin with guanidine isothiocyanate-phenol:chloroform and precipitated with isopropanol [74]. Briefly, 500 µl of pre-heated (56^{o}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^{o}C for 30 minutes. The swab or stain pieces were then placed into a DNA IQ^{TM} spin basket (Promega, Madison, WI), 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 600µl acid phenol:chloroform (5:1), pH 4.5 (Ambion/Life Technologies, Grand Island, NY). The samples were placed at 4^{o}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 1.5ml microcentrifuge tube, to which 2µl of GlycoBlue^{TM} glycogen carrier (Applied Biosystems/Ambion) and 500 µl of isopropanol were added. RNA was precipitated for 1 hour at -20^{o}C. 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/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 each extraction as a negative control.

DNase I digestion

DNase digestion was performed using the Turbo DNA-*free*™ kit (Ambion/Life Technologies) according to the manufacturer's protocol. Briefly, 2U units of TURBO™ DNase (2U/µl) and 2 µl (0.1 volume) of Turbo DNase I Buffer (10X) were added to each RNA extract (20 µl) and incubated at 37ºC for 30 minutes. The DNase was inactivated by the addition of 2.3 µl (0.1 volume) of the supplied DNase Inactivation Reagent. The samples were incubated at room temperature (occasional mixing) for 5 minutes. The samples were centrifuged at 10,000 x *g* for 1.5 minutes and then the supernatant was transferred to a new 1.5mL tube. The samples were used immediately or stored at -20ºC until needed.

RNA quantitation

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

cDNA Synthesis (Reverse Transcription, RT)

miScript Reverse Transcription Kit (QIAGEN, Valencia, CA) – for miRNA. The 20µl reaction consisted of: 4µl 5x miScript RT buffer, 1µl miScript Reverse Transcriptase mix, and 15 µl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: incubation at 37ºC for 60 min, followed by inactivation of reverse transcriptase at 95ºC for 5 min. All reactions were performed on an ABI 9700 thermal cycler. The standard RT input for the miRNA profiling assays was 1-5ng of total RNA. A reverse transcription blank (containing reverse transcriptase enzyme but no sample) was included for each RT reaction. Reverse transcription negatives (sample but no reverse transcriptase enzyme) were included as needed. For RT reactions involving non-standard samples (i.e. environmentally compromised samples or touch DNA samples), a positive control was included in the RT reaction, which consisted of a relevant known body fluid total RNA sample known to give consistent and reproducible RNA profiling results.

miScript Reverse Transcription II Kit (QIAGEN) – for miRNA. The 20µl reaction consisted of: 4µl 5x miScript buffer (HiSpec or HiFlex), 2µl 10x miScript Nucleics mix, 2µl miScript Reverse Transcriptase mix, and 12 µl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: incubation at 37ºC for 60 min, followed by inactivation of reverse transcriptase at 95ºC for 5

min. All reactions were performed on an ABI 9700 thermal cycler. The standard RT input for the miRNA profiling assays was 5ng of total RNA for HiSpec-RT reactions and 1ng for of total RNA HiFlex-RT reactions. A reverse transcription blank (containing reverse transcriptase enzyme but no sample) was included for each RT reaction. Reverse transcription negatives (sample but no reverse transcriptase enzyme) were included as needed. For RT reactions involving non-standard samples (i.e. environmentally compromised samples or touch DNA samples), a positive control was included in the RT reaction, which consisted of a relevant known body fluid total RNA sample known to give consistent and reproducible RNA profiling results.

High Capacity cDNA Reverse Transcription kit (Life Technologies) – for mRNA. The 20µl reaction consisted of: 2µl 10x RT buffer, 0.8µl 25x dNTPs (100mM), 2µl 10X RT random primers, 1µl of MultiScribe™ Reverse transcriptase, 14.2 µl of sample and water (volume of water adjusted accordingly depending on the amount of sample required). The RT reaction was as follows: 25ºC for 10 min, 37ºC for 120 min, and 85ºC for 5 min. The standard RT input for the miRNA profiling assays was 25ng of total RNA. A reverse transcription blank (containing reverse transcriptase enzyme but no sample) was included for each RT reaction. All RT products were stored at -20ºC until needed. A positive control was included in the RT reactions, which consisted of a relevant (i.e. body fluid of interest for the particular experiment) known body fluid total RNA sample (previously demonstrated to give consistent and reproducible RNA profiling results).

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: 95ºC 15 min; 33 cycles of 94ºC 30 sec, 55ºC 90 sec (+0.2ºC/cycle), 72ºC 45 sec; 72ºC 30 min; 4ºC hold. The multiplex system used included 12 mRNA biomarkers (2 biomarkers per body fluid or tissue): blood – ALAS2, ANK1 [6,75]; semen - PRM2, TGM4 [5,76]; saliva – HTN3, STATH [2,5,76]; vaginal secretions – CYP2B7P1 [77], VAG2 (novel unpublished vaginal secretions biomarker so gene name not provided at this time); menstrual blood – MMP10, LEFTY2 [5,78]; skin – LCE1C, CCL27 [79].

Product Detection – Real Time PCR (miRNA/piRNA)

Real-time PCR was performed using the Relative Quantitation protocol on an ABI Prism 7000 or 7500 Sequence Detection System (Life Technologies). One microliter of cDNA was amplified using the miScript SYBR® Green PCR kit (QIAGEN) (25µl reaction: 12.5µl 2x QuantiTect SYBR Green PCR master mix, 2µl 10x miScript universal primer, 2µl 10x miScript primer assay), according to the manufacturer's protocols with one minor change (cycle number modified to 45-50 for initial screening and assay development in order to evaluate possible low abundance candidates). All miScript primer assays were obtained from QIAGEN. Custom primer assays for piRNA candidates were designed using QIAGEN miDesign (www.qiagen.com/miDesign). The cycling program consisted of the following: 95ºC 15 min, 40-

45 cycles of 94ºC 15 sec, 55ºC 30 sec, 70ºC 34 sec. The time for the 72ºC extension step was increased to 34 sec (from 30 sec) since the 7500 real instrument requires a minimum of 34 sec (detection step). Initially, a 30 sec extension step was used on the 7000 instrument. However, for consistency in programs between instruments 34 sec was eventually used for all instruments. A positive control was included for all real time PCR reactions, which consisted of a relevant (i.e. body fluid of interest for the particular experiment) RT product that had been tested previously (positive result obtained). An assay blank, consisting of all reaction mix components except that nuclease free water was used in place of sample, was included for each real time PCR reaction.

PCR Product Detection - Capillary Electrophoresis (mRNA)

All amplified fragments were detected with the ABI Prism 3130 Genetic Analyzer capillary electrophoresis system (AB by Life Technologies). A 1.0 µL aliquot of the amplified product was added to 9.7 µL of Hi-Di™ formamide (AB by Life Technologies) and 0.3 µL of GeneScan™ 500 LIZ® (G5 dye set) (AB by Life Technologies). The electrophoretic conditions used were as follows: 16 sec injection time, 1.2 kV injection voltage, 15 kV run voltage, 60°C, 20 min run time, dye set G5. All samples were analyzed with GeneMapper® Software v4.0 (peak detection thresholds of 50 RFUs).

III. RESULTS

A. Statement of Results

Preliminary Work – miRNA expression profiling

In preliminary previous work, we examined the expression of 452 miRNAs (~67% of the known miRNA*ome* at that time) in dried, forensically relevant biological fluids (blood, semen, saliva, vaginal secretions, and menstrual blood) in an attempt to find body-fluid specific miRNA candidates [80,81]. The body fluid samples consisted of five-donor pooled samples. One nanogram of input total RNA was utilized in the reverse transcription reaction. For our initial studies, the reverse transcription method we employed allowed for the reverse transcription of all RNA species in a total RNA extract thus permitting, if needed, the simultaneous analysis of multiple miRNAs and mRNAs (miScript system, QIAGEN). Detection was then performed using a SYBR-Green based qPCR assay with a miRNA-specific primer (miScript system, QIAGEN). Most other miRNA assay systems utilize miRNA-specific reverse transcription reactions that permit the analysis of only one miRNA per reaction. The SYBR-Green based qPCR assays allowed for the assessment of the expression levels of each miRNA in the pooled body fluid samples. The expression data, in the form of Ct values, was plotted for each miRNA in order to visualize the expression differences between the body fluids tested. After analysis of the 452 miRNAs, no truly fluid-specific miRNAs were identified in the sense that they were only present in single body fluid. However, numerous candidates were identified that were differentially expressed to such a degree as to potentially permit the identification of the body fluid of origin of forensic biological stains. Initial miRNA based body fluid identification assays were developed that comprised, for each body fluid, the expression of two miRNAs normalized against an internal small RNA positive control (i.e. U6b) displayed as scatterplots. MiRNA biomarkers for the identification of blood, semen, saliva, vaginal secretions and menstrual blood were developed using this approach that required as little as 50 – 500 pg for detection [81] (Figure 1). Once these assays were constructed, individual rather than pooled samples were evaluated in order to confirm the observed specificity and examine any potential sample variation issues. As can be seen from the preliminary assays that were developed (Figure 1; blood-miR451/miR16; semen-miR10b/miR135b; saliva-miR658/miR205; vaginal secretions-miR124a/miR372; menstrual blood-miR451/miR412), a distinct clustering of miRNA expression data from each body fluid of interest was observed. The degree of separation between the body fluid of interest and the other body fluids varied between the assays and was dependent upon the strength of the selected miRNA candidates. We originally designated a miRNA as a ‘strong’ candidate if its expression was greater than that of the U6b positive control (normalizer). Therefore in assays with two strong candidates the body fluid cluster would be located in the upper right hand quadrant such as that observed for blood (Figure 1). The inclusion of two strong saliva candidates is also indicated by the presence of the body fluid of interest cluster in the upper right hand quadrant (Figure 1). As can be seen for the vaginal secretions and menstrual blood assays, only one strong candidate was identified for each (Figure 1). For semen, two ‘weak’ candidates were identified. Therefore, while these assays demonstrated a degree of specificity for their respective body fluids and little cross-reactivity with twenty-one other human

tissue samples, additional work was required in order to improve the specificity of each of the body fluid assays.

Novel miRNA Candidate Identification

The initial screening described above included an evaluation of expression data from 452 miRNAs which was ~67% of the known miRNA*ome* at that time (miRBase v1.0, www.mirbase.com). However with each update to miRBase, novel human miRNAs were deposited. It was therefore possible that novel highly abundant and specific miRNA candidates could be identified. We therefore performed additional extensive screening of available miRNAs through miRBase v.16 (364 novel candidates through v.11; 30 novel candidates through v.13-v.14; 352 novel candidates through v.15-v.16). With the additional 746 miRNAs, we thus evaluated the expression of a total of 1,198 miRNAs in forensically relevant biological fluids: blood, semen, saliva, vaginal secretions and menstrual blood. We evaluated both reverse transcription positive (RT+) and negative samples (RT-) for each body fluid in order to ensure that any obtained Ct values were in fact from RNA and not from contaminating DNA.

During the new candidate screening studies, it was also evident that the development of a skin identification assay would be beneficial to the forensic community due the analysis of "touch DNA" evidence. Touch DNA evidence is generally perceived to be the result of DNA obtained from shed skin cells yet never confirmed with scientific evidence. This is largely due to the perception that it is not possible to ascertain the tissue source of origin of the biological material in touch DNA evidence. The uncertainty with regard to the source of trace biological material is now being exploited in some criminal proceedings in an attempt to diminish the significance of trace biological evidence. In separate work, we were performing novel biomarker discovery experiments for the identification of mRNA markers to identify the presence of skin. However, due to the high sensitivity of the original miRNA body fluid identification assays, we wanted to evaluate the possible use of miRNA profiling as well for an identification of the skin tissue source of origin of touch DNA evidence. Since skin was not included in the original screening of 452 miRNAs, the expression of all 1,198 miRNAs was evaluated in skin as well.

For the new candidate testing, we utilized single-donor samples for blood, semen, saliva, vaginal secretions, menstrual blood and skin. For this initial comprehensive screening for skin, we utilized skin tissue total RNA rather than total RNA from touch samples. Touch DNA samples will vary significantly in the amount of RNA (if any) present. With often undetectable amounts of total RNA in these samples, hundreds of touch DNA samples would have required in order to complete the screening of 1,198 miRNAs. Therefore, commercially-available skin tissue total RNA was used since it was present in sufficient quantity and quality for analysis and provided consistency throughout the candidate screening. Any developed skin miRNA assays would then need to be evaluated for use with actual touch DNA samples.

All body fluid samples were reverse transcribed using the miScript system (QIAGEN). The miScript reverse transcription kit (Figure 2) includes the miScript reverse transcriptase mix and miScript RT buffer. The miScript reverse transcriptase mix is an optimized blend of enzymes comprising a poly(A) polymerase and a reverse transcriptase (miScript system handbook, QIAGEN). During reverse transcription, miRNAs are simultaneously poly-adenylated since they do not naturally contain a poly(A) tail. Reverse transcription is performed using an oligo-dT primer as well as random primers. The oligo-dT primer has a universal tag sequence at

the 5'end that serves as a universal primer binding site for subsequent real time PCR assays using the miScript system. The miScript system provides the ability to analyze all RNA species in total RNA samples, as they will all be reverse transcribed. Additionally unlike other miRNA assays, which involve the use of miRNA specific RT reactions (one miRNA per reaction), numerous miRNAs can be profiled from a single miScript RT reaction. Therefore while multiplex detection assays are not possible due to the use of SYBR green for detection, the miScript system still permits the ability to analyze numerous miRNAs from the same sample. The cDNA samples are then used in subsequent real-time PCR assays using a miRNA miScript primer assay (QIAGEN), the universal primer supplied in the kit (for the universal tag sequence added during the RT reaction) and the miScript SYBR Green PCR kit.

Raw Ct values were plotted for each miRNA and the expression data was evaluated in order to identify potential candidates for each body fluid. A potential candidate was identified by higher expression (i.e. low Ct value) in a particular body fluid compared to all others. Figure 3 is an example of the expression data plots from which potential candidates were identified. All initial screening experiments were performed using 1ng of input total RNA into the RT reaction (5% used for detection which is equivalent to 50pg for detection) in order to identify highly abundant and specific candidates. From the 1,198 candidates, 73 potential candidates were identified amongst the various body fluids using 1ng of input total RNA (blood – 17, semen - 14, saliva – 4, vaginal secretions – 4, menstrual blood – 7 and skin – 27). Since there were only a few candidates identified for some of the body fluids, we wanted to ensure that we were not over-looking possible low or moderately abundant candidates because of the low amount of input total RNA that was used for this initial screening. There were many miRNAs in which Ct values of >30-35 were observed for all body fluids and tissues. Therefore, we further evaluated these candidates using increased input amounts (5ng total RNA – 250pg (i.e. 5%) for detection). Using this increased input amount, we identified an additional 144 miRNA candidates. Overall, 217 potential miRNA candidates (~18% of all miRNAs evaluated) were identified. The number of candidates from each miRBase version tested is provided in Table 1. While it appears that the largest number of candidates was identified for skin, the number of candidates for skin is higher since it includes candidates from the v.1 set (original miRNA screening set). The number of candidates identified for each body fluid using this v.1 set is not included since that work was performed previously.

miRNA Assay Development

Initial Assay Development

With the successful identification of a large number of potential candidates for a majority of the body fluids and tissues, we next further evaluated each candidate in order to determine if suitable miRNA assays could be developed. We continued to use 2D scatterplots using normalized expression data from two putative body fluid specific candidates. Our original qPCR miRNA assays for body fluid identification included the use of U6b for normalization of the miRNA expression data. The use of U6b in these initial assays was sufficient and 2D scatter plot assays were successful using the U6b-normalized expression data. For initial assay development, we also continued to utilize U6b.

Numerous candidates were identified for blood. However since our original blood identification assay was sufficiently sensitive and specific, we did not focus our efforts on

screening these additional blood candidates. Numerous candidates were identified for semen, vaginal secretions and menstrual blood (19, 26, and 22 respectively). For saliva, there were fewer potential candidates with only 8 additional candidates identified. Despite the low number of potential candidates for some of the body fluids, candidates were in fact identified for each fluid and we were therefore hopeful that novel candidates could be identified that would allow for an improvement in the specificity of the miRNA body fluid identification assays. With further evaluation of potential candidates (additional sample testing (specificity) and development of 2D scatterplots), we had developed prototype assays for the identification blood, semen, vaginal secretions and skin. However through this additional testing, it became evident that U6b may not be the ideal choice for normalization due to large expression variance amongst all body fluids. Therefore, we evaluated the use of alternative normalizers.

Normalization Strategy

Our previous work had shown that none of the other small RNAs (e.g. U29, U44, U26, U27, U30, U31, U90) were suitable for universal normalization across all of the forensically relevant biological fluids. However, a miScript PCR control set was available from QIAGEN, which consisted of small RNAs (snRNA and snoRNA) that we had not previously tested (RNU1A, RNU5A, SNORD25, SCARNA17, SNORA73A) as well as U6b. We tested each of the new normalizers, as well as U6b, in each of the forensically relevant biological fluids. SCARNA17 was not detected in any of the samples and was therefore rejected. RNU5A demonstrated significant differences between body fluids (Cts ranging from 24 to 34) and was therefore not considered to be an ideal normalizer. SNORA73A andSNORD25 were not highly abundant in any of the body fluids except for blood and was therefore rejected. RNU1A appeared to be highly abundant in most body fluids (Cts ranging from 20-27). Differential expression levels were observed, but RNU1A appeared to be more abundant in all fluids compared to U6b. Since the highest Ct (lowest abundance) was observed in saliva, additional samples were tested in order to determine the stability of RNU1A in saliva. An additional 10 samples were run and the Ct values ranged from 23 to 30 with expression undetectable in one sample. Due to the inconsistencies in saliva samples, it was decided that RNU1A may not be the optimal choice.

As a result of the inability to find a suitable small RNA normalizer, we evaluated the use of a “housekeeping” miRNA. Similar to mRNA housekeeping genes, it may be possible that a particular miRNA species could demonstrate similar and stable expression amongst the forensically relevant biological fluids. Since we had screening data available from the over 1100 miRNAs, we were able to evaluate the data in order to determine whether putative “housekeeping” miRNAs could be identified. From the data, we were able to identify 18 potential housekeeping miRNAs: miR-602, miR-1234, miR-1274b, miR-1181, miR-3195, miR-4270, miR-4274, miR-940, miR-448, miR-1233, miR-1238, miR-665, miR-138-1*, miR-313, miR-3180-3p, miR-3190, miR-3675 and miR-3295). Since the screening data from which these miRNAs were identified involved the use of only one donor per body fluid, we needed to evaluate each candidate with additional samples in order to determine if the expression levels would be independent of cell type. Expression levels of the set of 18 mRNAs were evaluated in three additional donors per body fluid. Based on abundance and consistency of expression, 12 miRNAs were identified as potential normalizers (miR-1234, miR-1181, miR-4270, miR-940, miR-448, miR-1233, miR-1238, miR-665, miR-138-1*, miR3180-3p, miR-3190). The

expression data from these candidates is provided in Table 2. The miRNAs were first ranked based on their expression levels (i.e. Ct values). There were six highly abundant candidates (top section in the table), three moderately abundant candidates (middle section of the table) and then three lower abundance candidates (bottom section of the table). Depending on the expression data being normalized in the various assays, it may be aapropriate to utilize the high or moderately abundant miRNAs. Within each classification, the miRNAs were then ranked based on expression levels and amount of variation between biological fluids. The top miRNA candidate (miR-940) was highly abundant and demonstrated little variation between samples.

We evaluated the use of miR-940 as the normalizer in our existing assays and it appeared that the miR-940 would be utilized in the blood, vaginal, semen and skin assays. However, it did not work as well in the saliva assay. Using miR-940 normalized data, there was no longer a clear separation of the saliva samples in a distinct cluster. Therefore additional saliva candidates needed to be identified.

Final First Generation Assay Development

The challenge with identification of saliva candidates is that screening data revealed few potential candidates. Therefore, we had to re-evaluate the screening data to identify candidates that we may not have initially considered since they were not "strong" candidates with significant differential expression compared to the other body fluids. Using this less stringent screening process, we were able to identify numerous potential candidates. Each of these candidates was initially screened using a small sample set. We also had to perform additional testing using more samples as many of the candidates were found in low abundance. After this extensive screening and optimization, we were able to identify eight potential saliva candidates. After further testing with additional samples, only one of the candidates demonstrated suitable and stable expression levels (miR-124*). Since a second candidate was not identified, we constructed a novel 2D scatter plot using one of the previous saliva miRNA candidates (miR-658). The combination of these two miRNAs, normalized with the housekeeping miRNA, resulted in a distinct saliva cluster.

With the development of the new saliva assay, we had successful assays for blood (original candidates miR-451 and miR-16), semen (miR-892a/miR8-91a), saliva (miR-658 miR-miR124*) vaginal secretions (two assays: miR124a/miR1280 and miR124a/4286) and skin (two assays: miR-455-3p/miR-3169, miR-1393/miR-494). At this stage, a suitable menstrual blood assay was not developed since it is a more challenging and "special case" body fluid.

Final (i.e. second generation) miRNA Assay Development

As the initial first generation assays were developed for blood, semen, saliva, vaginal secretions and skin, we discovered that the existing miScript RT system (QIAGEN) utilized for the miRNA assays had been modified and the original RT kit we had been using was no longer available. The miScript II RT kit was available and now contained two buffers: miScript HiSpec (Figure 4) and miScript HiFlex (Figure 5). According to the manufacturer "miScript HiSpec Buffer is specifically formulated to provide optimal results in mature miRNA profiling and quantification experiments using miScript miRNA PCR Arrays and miScript Primer Assays. miScript HiFlex Buffer is highly suited to low-throughput miRNA quantification experiments, in

addition to experiments where different RNA species, such as miRNA, mRNA, and precursor miRNA, are quantified from the same sample" (QIAGEN). We were informed that the original miScript RT kit essentially contained the "HiFlex" buffer. We were therefore hopeful that the developed assays would not be significantly impacted by the RT kit change. We were also hopeful, however, that the use of the new "HiSpec" buffer could potentially result in improved specificity via a better body fluid "cluster" separation.

In order to evaluate the new buffers, the same samples were reverse transcribed using both buffers and the resulting expression data was plotted on the 2D scatter plots. The results of the use of the two different buffers for the blood miRNA assay (miR-451/miR-16) are provided in Figure 6. As can be seen from these 2D scatter plots, a large degree of separation between the blood "cluster" and the other body fluids was observed using either buffer. However, there was a difference in the location of the clusters depending on the buffer used. The location of the positive blood cluster was better suited for our purposes using the HiSpec buffer (i.e. positive results in the upper right quadrant, negative results in lower left quadrant) and therefore the HiSpec buffer was selected for the blood miRNA assay.

We evaluated the other body fluid assays with both HiSpec and HiFlex buffers and it was determined that the HiSpec buffer was more suitable for some assays (blood, semen and skin) and the HiFlex buffer was more suitable for the remaining assays (saliva and vaginal secretions). The use of the HiFlex buffer for the saliva also allowed us to utilize a previously rejected saliva candidate, miR-205. Therefore, two suitable assays for saliva were developed (miR-658/miR-205 and miR-658/miR-124*). The developed assays utilized either 1ng or 5ng input amounts. However, with the use of the two different buffers, this would have required four different RT sets for the same sample in order to profile the sample using all developed assays. This was not feasible and therefore, after some optimization, we were able to use 5ng inputs for all HiSpec assays (blood, semen, skin) and 1ng inputs for all HiFlex assays. For some of the 5ng HiSpec assays (blood, skin miR139/miR494), a 1:5 dilution of the RT product was used in the subsequent real time PCR assays.

From the buffer selection and input amount experiments (described above), we realized that there were differences between the new RT buffers that were advantageous for some assays. We therefore decided to revisit our menstrual blood candidates and make a final attempt to develop a suitable menstrual blood assay. We re-evaluated potential menstrual blood miRNA assays using the 5ng HiSpec and 1ng HiFlex RTs. We evaluated expression of twenty miRNAs to identify possible menstrual blood candidates. From this initial testing, we identified four potential candidates (miR-185, miR142-3p, miR144 and miR-144*) using the 1ng HiFlex RTs that were utilized in various 2D scatterplots. We were hopeful that with additional validation (see below) the most suitable assay (two candidate combination) would be determined.

The developed miRNA profiling assays for each body fluid are provided in Figures 7-12 (for body fluids with multiple 2D assays only one of the assays is shown). An overview of the workflow for the miRNA profiling assays is provided in Figure 13. Total RNA isolated from suspected body fluid stains using a suitable RNA extraction method (we only utilized our standard manual organic extraction in this study since our previous work demonstrated no significant advantage or improvement with the use of miRNA specific extraction kits or enrichment protocols). The quantity of total RNA is then determined (we utilized the Quant-iT™ RiboGreen® RNA kit). The miScript II RT kit is then used to reverse transcribe 5ng of total input RNA with the HiSpec buffer and 1ng of total input RNA with the HiFlex buffer (two RTs are made from an individual sample). One microliter of the 5ng HiSpec RT reaction is then utilized

in subsequent real time detection assays: undiluted RT product for miR-891b, miR-892a, miR-455-3p and miR-3169 (semen and skin) as well as miR-940; 1:5 diluted RT product for miR-451, miR-16 and miR-139 (blood and skin) as well as miR-940. One microliter of the 1ng HiFlex buffer RTs are then also utilized in subsequent real time PCR assays: undiluted RT product for miR-940, miR-658, miR-205, miR-124*, miR-124a, miR-1280, miR-4286, miR-185, miR-144* and miR-144 (saliva, vaginal secretions and menstrual blood). Overall for each sample, 19 separate reactions are required in order to evaluate the sample for the presence of all six body fluids and tissues (**note: miR-139 and miR-142-3p are not shown in this workflow as they were subsequently removed after statistical model development as described below). However, all of these reactions can be performed from the single HiSpec and HiFlex RTs from each sample. Ct values for each of the miRNAs are then obtained. Ct value thresholds are first applied to eliminate any invalid data (differentiating a true weak positive value from background signal). We have done this by evaluating various RT negative (reverse transcriptase negative samples; average of ten samples per body fluid (five biological replicates, two technical replicates of each donor)) samples. The average value for the RT negative samples is then used as the assay threshold. For a majority of the assays, an undetected result was obtained and therefore any detected Ct could be considered in analysis. For the miRNAs where the threshold was below a 40 Ct value, the values ranged from 32.6 – 39.2 Ct. For these assays, anything above this threshold would be considered an invalid result: miR-940 (HiSpec, undil) – 38.5, miR-940 (HiSpec, 1:5) – 34.7, miR-891b – 37.5, miR-892a – 39.2, miR-455-3p – 38.7, miR-3169 – 32.6, miR-124a – 39.3). The acceptable raw Ct values are then normalized with miR-940 and the ΔCt values obtained are subjected to LogR analysis (see below).

Logistic Regression Modeling for the Analysis of miRNA Expression Data

The initial development of the miRNA profiling assays involved the use of various 2D (two-candidate) scatterplot assays. Body fluid identification using these assays was somewhat subjective with a visual determination of whether a data point was within the cluster of known body fluid samples. The limits of the body fluid cluster were not clearly defined making interpretation of results quite challenging. The use of the 95% discriminant analysis confidence ellipses (using Origin 8.1 software) provided some objective way to determine positive results. However, more refined analysis interpretation metrics were needed in order to provide a robust objective statistical approach for declaring a sample to be positive or negative with respect to the presence of a particular body fluid. Additionally, for skin, saliva, vaginal secretions and menstrual blood, multiple 2D miRNA assays were developed. However, for saliva and vaginal secretions, one candidate was shared between the two assays. While both assays provided suitable specificity (i.e cluster separation), we were hopeful that advanced statistical approaches would permit the use of a multi-candidate analysis where the specificity from all candidates for a specific body fluid could be utilized in combination to further strengthen the specificity of the overall assay.

In order to develop a robust statistical approach we began a collaboration with Dr Kamel Rekab, a statistician at the University of Missouri Kansas City (UMKC), to develop logistic regression (LogR) statistical models to permit accurate predictions of the presence of forensically relevant fluids. Importantly, the use of the logR models permit the use of multi-candidate models instead of the more limited two-candidate scatter plots. This was significant since it adds multi-

dimensional capabilities that can further improve the specificity of each assay. We had initial success (100% accuracy) with LogR in predicting the body fluid of origin using our original U6b 2D miRNA data (data not shown) used as the basis for our original 2009 miRNA publication [81]. Due to the success of this approach with our original expression data, we continued to utilize the logistic regression model approach with our newly developed miRNA assays (second generation (i.e. final) assays described above). Using this data set, we were able to develop statistical models that resulted in 100% accuracy in determining the body fluid of interest. This analysis included a set of 'unknown' samples to test each model after it was developed. For each body fluid, 100% accuracy was observed with the statistical models.

The models developed for the new miRNA assays were based on individual data sets specific to each body fluid. However, for unknown samples in casework, each sample would be profiled with all assays in order to determine which body fluids or tissues were present. Therefore, we combined all of the separate data sets into one large single data set so that the normalized expression data for all 18 miRNAs could be viewed for each individual donor. When this data was compiled, we realized that the same donors were not always used for the testing of each individual body fluid. We therefore had gaps in the data where expression data for all assays was not available for all samples. Once the data was assembled in this manner, we could easily see what assays needed to be run for these samples and were able to generate expression data for the complete 18-miRNA set for all samples. In addition to completing the data for existing donors, we were also able to generate expression data from numerous additional donors: blood (15 donors); semen (4 donors); saliva (14 donors); vaginal secretions (14 donors); and menstrual blood (5 donors). For each of the new donors, we generated data for all of the miRNAs in the current body fluid identification panel (19 biomarkers: 18 body fluid specific miRNAs and a 'housekeeping-gene-like' miR-940) (52 samples x 19 assays = 988 data points). Once this large data set was complete, further statistical analysis was conducted to refine the logistic regression models. Below we have provided a full detailed description of how the LogR model for the identification of semen as developed and then a briefer description of the other models developed for each of the other body fluids and tissues. The models will be described in more detail in upcoming publications (see section VI).

Details of the LogR statistical model for the identification of semen:

All data analyses were performed using the statistical software SPSS. The semen data set consisted of 115 donors. There were 15 semen donors and 100 non-semen donors. A subset of 103 cases (cases = donors) was selected (i.e. cases that were used to construct the logistic models). It consisted of 13 semen cases and 90 non-semen cases. There were 12 unselected cases (cases that were not used to construct the logistic models) that consisted of 2 semen cases and 10 non-semen cases.

The goal of an analysis using logistic regression is the same as any modeling technique used in statistics: To find the best fitted model to describe the relationship between an outcome (response) and a set of particular variables. What distinguishes a logistic regression model from the linear regression model is that the outcome is binary. The log distribution (or logistic transformation of p) is also called the logit of p or logit(p). It is the log (to the base e) of the odds ratio or likelihood ratio that the outcome is semen. In symbols it is defined as: $\text{logit}(p) = \ln[p/(1-p)]$, where p denotes the probability of identifying semen correctly. The estimated logit based on 103 selected cases is given by the following expression: $\ln[p/(1-p)] = -4.016*\text{miR891b-}$

3.101*miR892a+0.276*miR892a*miR891b+34.414. The two main effects miR891b and miR892a and the two-factor interaction miR891b with miR892a are very significant (Sig < .05). The standard errors (S.E.) of all three terms are less than 2, and the Wald statistics are significantly large.

To validate the model, in our model the chi square has 3 degrees of freedom, a value of 64.7333 and a probability of $p < 0.000$. This indicates miR891b, miR892a and their (two-factor) interaction have a very significant effect. So we need to look closely at the predictors and determine which markers are significant in predicting semen. The Nagelkerke R square indicates that 87.8% of the variation in the semen data is explained by the logistic model. This shows that there is a strong relationship of 87.8% between all three predictors miR891b, miR892a, the two-factor interaction and the prediction equation. Furthermore, Hosmer and Lemeshow (H-L) test shows that the model prediction does not significantly differ from the observed. A probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. Since the H-L goodness-of-fit test statistic is 0.405, which is greater than .05 which we want for well fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the model's estimates fit the data at an acceptable level.

The receiver-operating characteristic (ROC) Curve is very useful for evaluating the predictive accuracy of a chosen model in logistic regression. We obtain the curve by plotting sensitivity (True Positive, TP) against 1 – specificity (False Positive, FP). A perfect classification with 100% true positive and 0% false positive would have an area equal to 1. The area under the curve is 99.2% with a 95% confidence interval (97.6%, 100%). The optimal threshold was determined as the threshold corresponding to the nearest (FP, TP) to (0,1). In this analysis the optimal threshold is 0.5.

Once the logit is estimated, then the probability that the outcome is semen is given by: $P = \exp(\ln(O))/[1+\exp(\ln(O)]$, where the odd ratio $O = p/(1-p)$. Since the optimal threshold is 0.5, then $P > 0.5$ will be classified as semen and $P \leq 0.5$ will be classified as non-semen. Among the 103 selected cases, 90 non-semen cases were predicted 100% correctly, and 12 semen cases out of 13 were predicted correctly (one false negative). Among the 12 unselected cases, 10 non-semen cases were predicted 100% correctly and two semen cases were also predicted correctly.

Similar analysis was performed for the remaining body fluids and tissues. Table 3 contains the full LogR models (logit functions) for each of the body fluids. Additionally, summary statistics for the body fluid models are provided in Table 4. A brief summary of each of the remaining body fluids is also provided below:

For *saliva (miR658/miR205/miR124*)* using the developed model, among the 104 selected cases, 80 non-saliva cases were predicted 100% correctly, and 22 saliva cases of out 24 were predicted correctly (two false negatives). Among the 12 unselected cases, 10 non-saliva cases were predicted 100% correctly, and two saliva cases were also predicted correctly.

For *vaginal secretions (miR124a/miR1280/miR4286)*, along the 96 selected cases, 72 non-vaginal secretions cases were predicted 100% correctly, and 18 vaginal secretions cases out of 24 were predicted correctly (6 false negatives). Among the 10 unselected

cases, 8 non-vaginal secretions cases were predicted 100% correctly, and two vaginal secretions cases were also predicted correctly. The higher number of false negatives is not entirely surprising for a body fluid such as vaginal secretions. The RNA quantitation method used is not human specific and therefore for any fluids where endogenous bacteria is present may give artificially higher quantitation results. Therefore, less human specific total RNA may be added to the reverse transcription reaction than would be estimated based on the RNA quantitation. While the occurrence of false negatives is not ideal, the significance of these models is that no false positive results are observed. A false positive result would present a far more significant challenge for operational crime laboratories. The occurrence of false negative results might possibly be reduced if a human specific quantitation method were to be developed.

For *blood (miR451/miR16)*, among the 103 selected cases, 78 non-blood cases were predicted 100% correctly, and 24 blood cases out of 25 were predicted correctly (one false negative). Among the 12 unselected cases, 10 non-blood cases were predicted 100% correctly, and two blood cases were predicted correctly.

For *skin (miR139/miR455-3p/miR3169)*, among the 103 selected cases, 94 non-skin cases were predicted 100% correctly, and 8 out of 9 skin cases were predicted correctly (one false negative). Among the 12 unselected cases, 10 non-skin cases were predicted 100% correctly, and two skin cases were predicted correctly. During the development of the skin model, it was determined that only 3 of the 4 miRNAs used in initial assay development were required (miR-494 was not included in the developed model). Therefore miR-494 will no longer be included in the miRNA panel.

For *menstrual blood (miR185/miR144/miR144*),* amongst the 103 selected cases, 95 non-menstrual blood cases were predicted 100% correctly, and 7 out of 8 menstrual blood cases were predicted correctly (one false negative). Among the 12 unselected cases, 10 non-menstrual blood cases were predicted 100% correctly, and two menstrual blood samples were predicted correctly. During the development of the menstrual blood model, it was determined that only 3 of the 4 miRNAs used in initial assay development were required (miR-142-3p was not included in the developed model). Therefore the fourth menstrual blood miRNA candidate will no longer be included in the miRNA panel.

As these results demonstrate, we have developed logistic regression models that enable an accurate prediction of the presence of forensically relevant body fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin). The performance of all models was evaluated for the selected cases (cases that were used to construct the logistic models) and the unselected cases (cases that were not used to construct the logistic models). Significantly the rate of false positives (i.e. the wrongful identification of a body fluid) for all the models was 0%.

miRNA Assay Validation

Sensitivity

The results of the LogR model development demonstrate the high degree of specificity of each of the developed miRNA assays. To validate the developed miRNA assays (including the use of the LogR models for data interpretation), we evaluated assay sensitivity, and performance with admixed body fluid samples and mock casework samples.

For the sensitivity, experiments, a range of total input RNA was tested using multiple donors for each body fluid. The standard input for the HiSpec assays is 5ng and therefore the input range test included 5ng, 1ng, 500pg, 250pg, 100pg and 50pg inputs. The standard input for the HiFlex assays is 1ng and therefore the input range tested included 1ng, 500pg, 250pg, 100pg and 50pg inputs. The developed logistic regression models were used in the analysis of the sensitivity data.

For ***blood*** (5ng HiSpec assays), the input amounts listed above were used in the reverse transcription reactions. However, a dilution of the RT product (1:5) is made prior to analysis using the real time miScript assays. Therefore, $1/100^{th}$ of the actual amount added to the RT is used for detection. For two of the three blood donors tested, blood was accurately predicted only using the 5ng input (50pg for detection). A positive result for the third blood donor was obtained using 1ng and 5ng inputs (10 and 50pg for detection, respectively).

For ***semen*** (5ng HiSpec assays), the input amounts listed above were used in the reverse transcription reactions. For the semen assays $1/20^{th}$ of the input amount is actually used for detection in the real time PCR assays. For two of the three donors, a positive result was only obtained when 5ng input was used (250pg for detection). For the third donor, a positive result was observed for 250pg – 5ng of total RNA input into the RT reaction (12.5 to 250pg for detection).

For ***saliva*** (1ng HiFlex assays), the input amounts listed above were used in the reverse transcription reactions. For the saliva assays, $1/20^{th}$ of the input amount is actually used for detection in the real time PCR assays. The presence of saliva was detected with 250pg – 1ng input total RNA (12.5pg to 250pg for detection) for one of the two donors tested. For the second donor, saliva was detected in all input amounts, 50pg to 1ng (2.5 to 250pg for detection).

For ***vaginal secretions*** (1ng HiFlex assays), the input amounts listed above were used in the reverse transcription reactions. For the vaginal secretions assays, $1/20^{th}$ of the input amount is actually used for detection in the real time PCR assays. For both donors, the presence of vaginal secretions was detected in all inputs amounts tested, 50pg to 1ng (2.5pg to 250pg for detection).

For ***menstrual blood*** (1ng HiFlex assays), the input amounts listed above were used in the reverse transcription reactions. For the menstrual blood assays, $1/20^{th}$ of the input amount is actually used for detection in the real time PCR assays. For menstrual blood, the presence of menstrual blood was detected in two of three donors for all input amounts, 50pg to 1ng (2.5pg to 250pg for detection). For the third donor, a positive result was only obtained for the 500pg input amount (25pg for detection). It is unclear why the presence of menstrual blood was not detected in the remaining input amounts.

For ***skin*** (5ng HiSpec Assays), one of the three assays utilizing the 1:5 diluted RT product ($1/100^{th}$ of the total input used for detection), the other two utilizing the undiluted RT product ($1/20^{th}$ of the total input used for detection)), the input amounts listed above were used in the reverse transcription reactions. Skin was successfully detected using as little as 250pg input for both donors tested (2.5pg to 12.5pg for detection depending on the skin miRNA used (1:5 diluted product or undiluted product, respectively).

The overall results of this study demonstrate the high sensitivity of the developed miRNA assays. This could be advantageous in forensic casework analysis where a limited amount of

starting material is often available for analysis. For each of the developed miRNA assays picogram-level sensitivity was obtained (2.5 – 250pg amongst the various assays). Currently, the standard total RNA input for our mRNA CE-based multiplex is 10-25ng (input amount into the RT reaction, with 1-2.5ng used for detection). Therefore, the miRNA assays appear to offer improved sensitivity over mRNA profiling assays.

Mixtures

The ability to detect multiple body fluids in the same admixed sample using the developed miRNA assays was evaluated using the developed logistic regression models to analyze the expression data. All mixture samples were treated as "unknowns", meaning that expression data for the full miRNA panel (16 miRNAs and miR-940, after removal of the one menstrual blood and one skin miRNA candidate removed during logistic regression model development) was obtained for each sample and therefore each mixture sampled was evaluated for the presence of all body fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin). Twenty-four mixture samples were tested and included the following body fluid mixtures: blood-semen, blood-saliva, saliva-semen, vaginal secretions-semen, menstrual blood-semen, and vaginal secretions-semen. For each of the above-mentioned mixtures, the first body fluid listed was a dried stain to which the second body fluid in liquid form was placed on the dried stain thus creating admixed body fluid samples. For the dried stains, blood samples were 50µl bloodstains and for all other body fluids ½ swabs were used. For the liquid body fluid added, we evaluated four different volumes (50µl, 25µl, 10µl and 5µl). Therefore for each body fluid mixture (6 types), 4 samples each with decreasing amounts of the liquid body fluid were evaluated. The color-coded Table 5 shows the expected outcome for the admixed body fluid samples as well as the observed outcome. The results for each mixture type are described in more detail below.

For the *blood-semen* mixture (50µl, 25µl, 10µl and 5µl semen added to dried bloodstains), blood was correctly identified in three of the four mixtures. The one mixture in which it was not predicted to be present was the 50µl semen-blood mixture. This data point appears to be within the known blood sample 2D scatter plot cluster and therefore a small modification to the blood logR may be necessary. Semen was also successfully detected in the 25µl, 10µl and 5µl semen –blood mixtures. Semen was not detected in the same 50µl semen-blood mixture in which blood was not identified. All four blood-semen mixture samples were negative for saliva, vaginal secretions, menstrual blood and skin.

The next mixture type evaluated was *blood-saliva*, with saliva added to the dried bloodstains in 50µl, 25µl, 10µl and 5µl volumes. For these mixtures, blood was detected in the blood-saliva mixtures with 50µl and 10µl saliva volumes. Again, the 2D scatter plot data points for the blood using 25µl and 5µl of saliva appear to cluster with the known blood data and again warrant re-evaluation of the blood logistic regression model. Saliva was successfully detected in all four blood-saliva mixtures. All four blood-saliva mixture samples were negative for semen, vaginal secretions, menstrual blood and skin.

The next mixture was *saliva-semen*, with semen added to the dried saliva sample in 50µl, 25µl, 10µl and 5µl volumes. For these mixtures, semen was only detected in the 25µl saliva-semen mixture. Therefore, we were less able to detect semen in this mixture type. Saliva was detected in all of the saliva-semen mixtures except for the 50µl sample.

All four saliva-semen mixture samples were negative for blood, vaginal secretions, menstrual blood and skin.

We tested *vaginal secretions-semen mixtures*, with semen added to the dried semen sample in 50µl, 25µl, 10µl and 5µl volumes. For these mixtures, vaginal secretions was successfully detected in all four samples while semen was not detected in any of the samples. These mixture samples were diluted to the appropriate assay input amount (1ng or 5ng total RNA) and therefore due to the high amount of total RNA typically obtained from vaginal secretions samples, it is possible that the smaller amount of semen RNA was diluted out because of the high overall quantitation value. All four vaginal secretions-semen mixtures were negative for blood, saliva, menstrual blood and skin.

Menstrual blood-semen mixtures were analyzed, with semen added to the dried menstrual blood sample in 50µl, 25µl, 10µl and 5µl volumes. For these mixtures, menstrual blood was successfully detected in all mixture samples except for the mixture in which 5µl of semen was added. Semen was successful detected in the mixtures where semen was present in 50µl and 25µl volumes. All four menstrual blood-semen mixtures were negative for the presence of blood, saliva, vaginal secretions and skin.

The final mixture type evaluated was *vaginal secretions-saliva*, with saliva added to the dried vaginal secretions samples in 50µl, 25µl, 10µl and 5µl volumes. Vaginal secretions and saliva were successfully detected in all four mixture samples. All four of the vaginal secretions-saliva mixtures were negative for the presence of blood, semen, menstrual blood and skin.

In summary, while there were some false negatives, these results indicate that the presence of multiple body fluids can be, but not always, identified using the developed miRNA assays. Significantly, no false positive results were observed.

Additional validation studies

We intend to continue our validation efforts for the developed miRNA assays beyond the validation studies completed in the current work. We will evaluate the stability of the vaginal secretions and menstrual blood assays over a full 28-day female cycle for both menstruating and post-menopausal female donors. Additionally, we will evaluate *bona fide* admixed body fluid samples such as sexual assault evidence (semen/vaginal secretions). For the skin assay, we will also evaluate touch DNA samples in order to provide a determination of the skin source of origin of these samples. We are also currently evaluating all initially identified skin miRNA candidates with touch DNA samples in order to determine if touch DNA-specific miRNA candidates can be identified. It is expected that gene expression differences will be observed between surface skin layers (dead or dying keratinocytes) compared to deeper skin layers and therefore it may be possible to identify miRNAs specific to surface skin layers and more suitable for the analysis of touch DNA samples.

Comparison of miRNA and mRNA Profiling Success of Environmentally Impacted Body Fluid Samples

One of the potential significant advantages originally hypothesized for miRNA profiling was an improved ability to identify the body fluid or tissue source of origin in degraded samples due to the significant smaller size of miRNAs. Akin to utilizing miniSTRs for improved STR analysis of degraded DNA samples, miRNAs represent a "mini" RNA molecule that may be well suited for use with degraded samples. We therefore evaluated a significant number of challenging or environmentally compromised samples for blood, semen, saliva and vaginal secretions using our developed miRNA assays as well as our current CE mRNA body fluid identification multiplex (12 biomarker system; 2 biomarkers per body fluid; see Methods for list of included biomarkers). Environmentally impacted samples were not available for menstrual blood and skin.

For blood (Table 6), bloodstains exposed to the following conditions were evaluated: room temperature for one and two years (two donors: 1 male, 1 female), 37ºC for one and two years (two donors: 1 male, 1 female), 56ºC for one and two years (two donors: 1 male, 1 female), outside on a patio for 1 – 4 months (exposed to heat, humidity, some light, largely protected from rain; 1 female donor), outside in direct sunlight for 1 week, 1 month and 2 months (exposed to heat, humidity, light and rain; 1 female donor), outside in the shade for 1 month, 2 months and 6 months (exposed to heat, humidity, rain, some light but not direct exposure; 1 female donor), the back seat of a car for 1 week (two donors: 1 male, 1 female) and inside a car trunk for 2 weeks (two donors: 1 male and 1 female). Fifty-microliter bloodstains were used for analysis. The optimal total RNA input was used for each assay (5ng HiSpec for the miRNA assays with 1:5 dilution of the RT product; 25ng input for the mRNA multiplex system). As previously described with the mixture study, we observed some false negative results using the developed LogR models that appear to be within the known blood sample cluster on the 2D scatter plot. We will seek to refine the LogR model in order to reduce the number of false negative calls. Blood was not detected using either the miRNA or mRNA assays for the bloodstains exposed to sun and shade (outside: heat, light, humidity, rain). It is certainly possible that no biological material remained on these samples after exposure to rain. Blood was accurately identified for the samples stored in the car back seat and car trunk (both donors) using both the miRNA and mRNA assays (both mRNA biomarkers detected). Blood was also accurately identified in the room temperature (1 and 2 year samples; two donors), 37ºC (1 year samples (male donor only) and 2 year sample (both donors), 56ºC (1 and 2 year samples; two donors) and the patio samples (1, 2, 3 and 4 months) using both the miRNA and mRNA assays. It is important to note that for a majority of the mRNA results, the positive result is based on the detection of only one of the two biomarkers (ALAS2). ANK1 was frequently not detected (only present in the car trunk and back seat samples; male room temperature 1 and 2 year samples; and the patio 1 month sample). For the 37ºC 1 year sample, ALAS2 was present but only at 42 RFUs which is below our 50 RFU threshold and therefore was considered a negative result for both markers (blood successfully detected using the miRNA assay for this sample). We considered the presence of at least one of the two mRNA biomarkers a positive result. Overall, blood was successfully detected in a majority of the samples using both RNA profiling assays. However, perhaps somewhat surprisingly, no improvement in analysis success rate was observed for miRNA compared to mRNA.

For semen (Table 7), similar samples were evaluated: room temperature for one and two years (two male donors), 37ºC for one and two years (two male donors), 56ºC for one and two years (two male), outside covered (exposed to heat, light, humidity but protected from rain) for 1 day (one male donor), 3 days (two male donors) and 1 week (one male donor), UV light for more

than 1 year (one male donor), outside on a patio for 1 day, 2 days, 12 days and 1 month (one male donor; exposed to heat, humidity, some light, largely protected from rain; 1 female donor), outside in direct sunlight for 1 day, 2 days, 12 days and 1 month (1 male donor; exposed to heat, humidity, light and rain; 1 female donor), and outside in the shade for 1 day, 2 days, 12 days and 1 month (1 male donor; exposed to heat, humidity, rain, some light but not direct exposure; 1 female donor). Fifty-microliter semen stains were used for analysis. The optimal total RNA input was used for each assay (5ng HiSpec for the miRNA assay; 25ng input for the mRNA multiplex system with the exception of the room temp 2 yr male 2 sample from which only 15ng could be added due to a low quant value). For the room temperature, 37ºC and 56ºC, semen was successfully detected in all samples from both donors using both the miRNA and mRNA assays (both mRNA biomarkers detected in 9/12 samples). For the outside covered samples, semen was detected only in the sample exposed for 1 day. However, semen was detected with mRNA analysis in all samples except for one of the 3- day samples. Therefore, mRNA profiling was superior to miRNA profiling for the outside covered samples. For the semen samples stored on a patio, semen was successfully detected using both assays in the 1 day, 2 day and 12 day samples. Semen was detected in the 1 month sample using mRNA profiling only. For the samples stored in the sun, semen was not detected in any of the samples using the miRNA assay, but was detected in the 1 and 2 day samples using the mRNA multiplex system. For the samples stored in the shade, semen was not detected in any of the samples using the miRNA assay, but was detected in the 1 day, 2 day and 12 day samples using the mRNA multiplex system. Therefore, for semen, it appeared that mRNA profiling was more successful that miRNA for the analysis of the environmentally compromised samples.

For saliva (Table 8), 50µl stains (liquid saliva) were exposed to the following conditions: room temperature for one and two years (two donors: 1 male, 1 female), 37ºC for one and two years (two donors: 1 male, 1 female), 56ºC for one and two years (two donors: 1 male, 1 female), outside covered (exposed to heat, light, humidity but protected from rain) for 1 day (two donors, 1 male and 1 female), outside uncovered (exposed to heat, light, humidity and rain) for 1 and 3 days (one male donor), outside in direct sunlight for 1 day, 2 days, 1 week and 1 month (exposed to heat, humidity, light and rain; 1 female donor), outside in the shade for 1 day, 2 days and 1 month (exposed to heat, humidity, rain, some light but not direct exposure; 1 female donor), outside on a patio for 1 day, 2 days, 1 week and 1 month (exposed to heat, humidity, some light, largely protected from rain; 1 female donor), and saliva on denim exposed to 56ºC for 2 and 3 weeks (one donor each). Whole stains were used for analysis. The optimal total RNA input was used for each assay (1ng HiFlex for the miRNA assays; 25ng input for the mRNA multiplex where possible). Less than 25ng input was available for reverse transcription for the following samples due to a low quant value (mRNA assay only): room temp 2 yr – male (20ng), outside covered – 3 day – male (7ng) and sun – 1 day (18ng). For the saliva samples, saliva was successfully detected in more samples using the miRNA assays compared to the mRNA multiplex system. For the miRNA assays, saliva was successfully detected in all samples except for sun-1 day, shade-1 month, and patio-1week (83% success rate). For the mRNA multiplex system, the presence of saliva was only detected in the room temperature 1 year and 2 year samples (both donors), the 37ºC 1 year (both donors) and 2 year samples (male donor only), the 56ºC 1 year samples (both donors), the shade 1 and 2 day samples, and the patio 1 day, 2 day and 1 week samples (52% success rate with any of the markers, 31% with both saliva biomarkers detected). This is a significant improvement in saliva detection using the miRNA assays compared to mRNA profiling.

The final body fluid evaluated was vaginal secretions (Table 9): room temperature for one and two years (two female donors), 37°C for one and two years (two female donors), 56°C for one and two years (two female donors), outside covered (exposed to heat, light, humidity but protected from rain) for 1 day (two female donors), 3 days, 1 week and 1 month (one female donor), and outside uncovered (exposed to heat, light, humidity and rain) for 1 day and 1 month (one female donor). A half-swab to full swab was used for analysis. The optimal input was used for each assay (1ng HiFlex for the miRNA assays; 25ng input for the mRNA multiplex system). The success rate for detection of vaginal secretions was 32% for both the miRNA and mRNA assays (only 16% success of 2-marker detection for the mRNA assays). Vaginal secretions was successfully detected using both assays for the room temperature 1 and 2 year samples (one donor only) and the 37°C 1 year and 2 year samples (one donor only). Vaginal secretions was detected using the miRNA assays for the 56°C 1 and 2 year samples, whereas vaginal secretions was weakly detected (83 – 184 RFUs) using the mRNA multiplex system for the outside covered 1 day samples (both donors).

Overall the results of the environmentally compromised system are promising for the use of RNA profiling in forensic casework. However there was no apparent advantage or improvement in using miRNA profiling for the analysis of these samples, with the exception of saliva. Additionally, the poor results obtained from selected environmentally compromised samples (in particular the samples exposed to outside conditions) included in this study may be due to the absence of biological material rather than extensive degradation of the biological material itself. In the future we will conduct controlled RNA degradation experiments in order to further assess whether miRNA offer any advantages over mRNA analysis with *bona fide* degraded samples.

As mentioned above, for the purpose of this experiment we classified a mRNA result as positive if one or both biomarkers were detected. Currently our mRNA multiplex system contains two RNA biomarkers for each body fluid and it has not yet been determined whether declaration of a positive identification would require detection of both biomarkers. If in fact the presence of both biomarkers is needed for a definitive result, then the miRNA assays would result in successful body fluid identification at a much higher rate compared to the mRNA assays (results for semen are similar even with two marker mRNA detection, but still a slight improvement using miRNA assays) (Figure 14).

Low Density miRNA Arrays

During the initial candidate screening (see the 'novel miRNA candidate identification' section above), we also evaluated alternative miRNA analysis and detection methods in order to determine if they would provide improved sensitivity and specificity compared to the developed miScript assays or if they would result in the identification of alternative body fluid specific miRNA candidates.

The alternative strategy that we were able to evaluate in this study involved the use of qNPA (quantitative nuclease protection assay) technology to evaluate gene expression (HT Genomics, Tucson, AZ). Like any technology it has its advantages and disadvantages. Advantages of this technology are that no RNA amplification is needed, it is automated and provides an accurate assessment of the expression levels of the miRNAs present in the sample. However, at the time we evaluated this technology, it required a significant amount of RNA

input (~500ng), which may not be ideally suited for use in forensic casework. However, due to the advantages of this approach, we felt that it is worth an initial evaluation of the qNPA assays for body fluid identification.

In qNPA, gene-specific oligonucleotides are added to the sample and hybridize to the RNA present in solution. The oligonucleotides are added in excess to ensure that every molecule RNA capable of hybridizing does so. S1 nuclease is added to the hybridized sample buffer. The S1 nucleases will degrade any non-hybridized nucleic acid. This removes the non-hybridized portion of the targeted RNA, all of the non-targeted RNA and any excess oligonucleotides. The S1 nuclease is then inactivated. The RNA:DNA heteroduplexes are then treated to remove the RNA portion of the duplex, leaving only the previously protected oligonucleotide probes. The resulting DNA oligos are essentially a library of the original RNA sample. Measurement of miRNAs is then performed using a qNPA ArrayPlate. The ArrayPlate spots are programmed using the appropriate gene-specific linkers that provide a specific capture sequence for the targeted miRNA. The qNPA protection oligos are added to the ArrayPlate wells and allowed to hybridize. The short, biotinylated miRNA protection oligos quantitatively hybridize to the programmed array spots. The array-bound biotinylated miRNA qNPA protection oligos are labeled with avidin-horseradish peroxidase (HRP) conjugate. The addition of a chemiluminescent substrate produces light that is then quantitatively measured.

Our original work with HT Genomics was to develop a custom plate including the nine miRNAs originally utilized in our body fluid identification panel in order to determine if this approach would result in the same specificity. Five body fluid samples (one blood, semen, saliva, vaginal secretions and menstrual blood) were provided for analysis. The nine miRNAs as well as U6b, GAPDH and B2M were included in order to determine if suitable normalizers could be identified. The results of initial analysis using the nine miRNA candidates can be found in Figure 15. The data is displayed by body fluid to easily visualize any unique body fluid "signatures" that were obtained with the miRNA panel. For the blood sample, expression of miR451 and miR16 was observed. This is consistent with the results we obtained using our original miRNA assays. However, unexpectedly miR16 was also present in high abundance in semen. This has never been observed with the QIAGEN miScript assays. For the semen sample, our original candidate miR10b was very weakly expression and miR135b was not detectible at all. For the saliva sample, miR205 was highly expressed which is consistent with our original data. miR658 was not detectible indicating that this miRNA may be weakly abundant, but is detectible using our original assays since it is a PCR-based system. For the vaginal secretions samples, miR124a showed the highest expression, which is also consistent with our original data. However, miR124a was also present in the saliva sample although with a slightly lower expression than obtained in the vaginal sample. miR124a is also highly expressed in the menstrual blood samples. This is somewhat expected as menstrual blood samples will contain significant amounts of vaginal secretions. The overall results of this initial work demonstrated that there are some similarities in expression between the two approaches. However, it was evident that additional miRNA candidates would need to be identified for use in the qNPA assays in order to develop more unique "signatures" for each body fluid.

Since additional miRNA candidates were needed in order to develop unique body fluid "signatures" using this new technology, we decided that whole transcriptome profiling would be performed in order to identify miRNA candidates for each body fluid. HT Genomics had the capability of performing this analysis using the miRNAs current included in the miRBase database (human, mouse, and rat). Two samples for six biological fluids and tissues (blood,

semen, saliva, vaginal secretions, menstrual blood and skin) were provided. The obtained whole transcriptome data was evaluated in order to identify miRNA candidates with detectible expression in a single body fluid or with significantly higher abundance in one particular body fluid. Additionally, we also examined the data for the presence of miRNAs demonstrating consistent expression levels between the different body fluids. These miRNAs could potentially be used as normalizers for candidate expression data. Several potential "housekeeping" miRNAs were identified in the whole transcriptome profiling data (miR-1181, -1826, -638, -923, -197).

Numerous candidates were identified for blood. Figure 16 shows the expression levels of 7 miRNAs (miR-25, miR-425, miR-324-3p, miR-320c, miR-15b, miR-182, miR-106b) that were identified as potential blood candidates (each body fluid represented in a different color). As can be seen from this data, the signal intensity (i.e. expression level) of each of the included miRNAs is significantly higher in blood than any other body fluid. Attempts were made to normalize the expression data using a Δsignal approach (similar to ΔCt) using one of the identified potential "housekeeping" miRNAs. The results from the normalization (using miR-1181 for normalization) of the blood miRNA candidate expression data are provided in Figure 17. Normalization of this data using the single potential "housekeeping" miRNA resulted in positive Δsignal values to be obtained for blood and negative values to be obtained for all other body fluids. Therefore, it might be feasible to utilize the normalized expression data for these 7 miRNAs for the identification of blood.

The miRNA whole transcriptome data was examined for the identification of candidates for the remaining body fluids and numerous candidates were identified. The smallest number of potential candidates was observed for saliva. The reason for this is unknown but it is possible that the quantity of total human RNA in the saliva samples was affected by the presence of contaminating bacteria. Therefore, less human total RNA would be added to the assay based on an inflated quantitation value. Similarly vaginal samples might be affected as well. However, numerous candidates for vaginal secretions were identified.

The next step in the qNPA analysis was to examine the expression levels of any identified body fluid candidates using a larger number of samples of each body fluid. This was done using a qSelect plate, which allowed for the inclusion of up to 47 miRNAs and up to 96 samples. Based on the evaluation of the whole transcriptome data, we selected 7 candidates for blood (miR-25, -425, -15b, -106b, -182, -324, and -320c), semen (miR-891a, -602, -483-5p, -654-5p, -375, -886-5p, and rno-750-5p), vaginal secretions (miR-612, -720, -124, -1251, -1267, -663 and -1826) and menstrual blood (miR-223, -486-5p, -210, -144, -92a*, -425* and -142-5p). Only 4 potential candidates (miR-606, -205, -1308 and mmu-885) were identified for saliva and therefore all four were selected for inclusion on the qSelect plate. Numerous candidates were identified for skin and therefore we selected 10 miRNAs that were found to be in higher abundance in skin (miR-145, -125b, -195, -214*, -143, -193b, -23a, -100, -99b, -768-5p). We also included the five potential "housekeeping" miRNAs. We were able to analyze up to 96 samples, so it was decided that we would analyze 48 samples with two technical replicates for each sample. The 48 samples included the following: 8 blood, 8 semen, 9 saliva, 8 vaginal secretions, 3 menstrual blood, 6 skin, 1 blank and 1 of each of the following tissues – adipose, brain, heart, muscle, and prostate (thought to be more forensically relevant than other available tissues). All body fluid samples were extracted from dried stains. However, the skin and tissue samples were available as total RNA extracts from commercially available sources and were certified to contain small RNAs. Initially, the plate was run using the same input amount as previous experiments (~500ng/well). When the data was reviewed using this input level, very

few of the candidates demonstrated the same specificity that was observed with the whole transcriptome profiling. However better correlation of the specificity results between the two methods were obtained with less input total RNA.

Four of the seven potential blood miRNA candidates demonstrated concordant specificity between the two array types (custom plate and whole transcriptome)(miR-25, -425, -15b and -106b). Figure 18 shows an example of one of the identified candidates (miR-25). As can be seen from this bar graph, the signal intensity observed in the blood samples is significantly higher than in the other body fluids. Only two of the seven semen candidates tested (miR-891a and -602) were confirmed as potential semen markers. Two additional miRNAs (miR-482-5p and -654-5p) demonstrated a possible specificity for semen, but the expression was not consistent amongst all semen samples. Only one of the four saliva candidates (miR-606) demonstrated sufficient specificity for saliva. Higher expression levels for three of the seven vaginal secretions candidates (miR-612, -720 and -124) were observed. An additional two miRNAs (miR-1251 and -1267) demonstrated possible specificity but had very low signals and therefore might require additional input. Similarly, two of the seven potential menstrual blood candidates (miR-223 and -486-5p) demonstrated specificity for menstrual blood with another two potential candidates that could be potential candidates with further work (miR-210 and -144). Five of ten potential skin candidates demonstrated sufficient specificity for skin (miR-145, -125b, -195, -214* and -143).

With the identification of potential candidates, we next evaluated suitable normalization strategies for the expression data. At this stage in the project, it did not appear that the qNPA assays provided any advantage over our qPCR assays for blood, saliva and menstrual blood and therefore our focus on further qNPA assay development was focused on semen, vaginal secretions and skin (suitable miScript miRNA assays had not been developed at the time, although they were later successfully developed as described previously).

Since the qNPA assays are not PCR based, it was unclear what the proper strategy for normalization should be. Unfortunately, we did not have success in the use of U6b or a "housekeeping" miRNA for normalization (data not shown). As a further attempt to identify potential normalizers, we considered the use of a housekeeping gene (mRNA). A qNPA assay was available that permitted the evaluation of 44 housekeeping genes. Sufficient sample was remaining for a majority of the body fluid samples included in the qSelect plate for candidate identification and therefore these same samples were used for evaluation of the housekeeping gene panel. If a suitable candidate was identified, direct normalization of the miRNA expression data could be performed since the same samples were used. Unfortunately, there was significant variability between the body fluids and a "universal" housekeeping gene was not clearly identified.

Despite the initial promising results with the qNPA assays there were too many difficult challenges to be overcome including the cost, the determination of normalization strategies, the irreproducibility for some body fluids as well as considerable variability between analysis methods (candidates identified using qNPA could not be confirmed using the miScript assays). We were also not able to perform this analysis in-house and therefore further optimization and validation work would have been particularly difficult and costly challenging and costly. Despite this it is still possible that non-PCR based qNPA assays could play a role in miRNA profiling given sufficient effort and resources to optimize the system for forensic use. .

Evaluation of piRNA Expression for the Identification of Semen

Our preliminary screening of the 452 miRNAs, indicated that miRNAs were less abundant in semen samples compared to other biological fluids. Few semen candidates were originally identified and the differentiation of semen from the other body fluids using available semen miRNA candidates was not as good as that obtained with the other body fluid assays. However, as described above, with the screening of 746 novel miRNAs, suitable potential candidates for semen were identified and a sensitive and specific miRNA profiling assay for semen has been developed. However, semen was not successfully detected in all admixed body fluid samples or environmentally compromised semen samples and therefore it may be prudent to continue to seek additional biomarkers for the identification of semen using a small RNA profiling approach. We therefore evaluated as potential semen markers an additional novel class of small regulatory RNAs, piwi-interacting RNAs (piRNAs). piRNAs are slightly larger than miRNAs and are generally ~24-30 nt in length [82]. Studies indicate that piRNAs are involved in germline development, silencing of transposable DNA elements, and in maintaining germline DNA integrity [82]. Piwi-interacting RNAs initially had been reported to be present only in sperm-producing cells in mammals and therefore could be extremely specific to semen [82-85]. This would ideally eliminate any potential cross-reactivity with vaginal secretions and menstrual blood, or expression from any body fluids from female donors.

There are possibly hundreds of thousands of known piRNAs [86]. However, unlike our miRNAs studies, we could not perform a large scale survey of piRNA expression in forensically relevant biological fluids. We intended to only perform a small scale evaluation of piRNAs expression in order to provide preliminary support for a possible improvement in small RNA profiling assays for semen identification (if in fact they were specific to or enriched in germ-line cells). Few published studies have involved an evaluation of piRNAs in human tissues. However, we found a small number of published articles that identified piRNAs in human tissues [87-90]. While these studies typically evaluated piRNA expression in cancer cells and cancerous tissue, we decided to use three of the identified piRNAs as an initial test set since they were identified in human tissues. We employed the same QIAGEN miScript system for the analysis of piRNAs that was used for our miRNA assays since ideally we wanted all developed small RNA assays to be compatible with one another and not require different reagents or assays conditions. We submitted the three piRNA sequences for custom primer assay development (QIAGEN) and were able to obtain suitable primer assays for two of the three piRNAs (hsa-piR-015520 and hsa-piR-00651). The third sequence was too long for the miScript system and therefore a suitable primer assay could not be developed. To include other piRNAs in this study, we chose a few additional miRNAs from the piRNA database (http://pirnabank.ibab.ac.in/) (hsa-piR-000001 through hsa-piR-0000010, with hsa indicating human). We also evaluated an additional five piRNAs that were potentially specific to the human Y chromosome (hsa-piR-017183, -011188, -006465, 013745, and -016792). If these piRNAs were specifically transcribed from the Y chromosome in tissue, then no cross-reactivity should be observed with single source vaginal secretions or menstrual blood since they would originate from females and therefore not contain a Y chromosome. The same would be expected for the other body fluids as well if originating from a female donor.

We evaluated expression of each of the piRNAs in all forensically relevant body fluids (semen (5 donors); blood (2 donors); saliva (2 donors); vaginal secretions (2 donors); menstrual blood (1 donor)). Although skin was excluded from the initial testing, it would be subsequently

analyzed if suitable assays were developed. In addition to the body fluids, a testes and prostate tissue sample (3 donor pools; obtained from commercial sources) were included as positive controls. Expression of each of the piRNAs was evaluated using both the HiSpec and HiFlex reverse transcription buffers in the miScript II RT kit to determine the most suitable buffer for analysis. After initial screening, the HiFlex buffer appeared the more suitable for use with the piRNA candidates since high Ct values were obtained for most candidates using the HiSpec buffer (data not shown). None of the piRNA candidates demonstrated either (i) a significantly higher abundance in semen or (ii) expression only in semen (or the prostate and testes tissues). Expression was observed in all body fluids. After analysis of the expression data, eleven of the piRNAs were rejected with no further testing warranted. These candidates were rejected either due to poor abundance (high Ct values in all samples) or no identifiable expression trends (similar Ct values in all samples). The remaining six candidates demonstrated a degree of body fluid specificity and therefore were selected for further analysis. These six candidates included one piRNA identified from the literature searches and the five Y chromosome piRNAs. For normalization of the expression data, we utilized U6b. The normalized expression data for each of the piRNA candidates was used in the development of 2D scatter plots in order to determine if a possible trend could be identified (i.e. distinctive clustering of the various body fluids and tissues). All possible two-candidate 2D plots were constructed. Several of the two-candidate combinations demonstrated a clustered separation of semen samples (with one semen donor consistently appearing as a false negative (not within the cluster of the other four semen donors)). However, the menstrual blood sample also consistently appeared near or within the semen sample cluster. An example of one of the piRNA 2D expression plots is provided in Figure 19.

Since a majority of these piRNAs were supposedly Y chromosome specific, it was quite surprising for the menstrual blood sample to be detected and near/within semen data points. While this was only one sample, if the piRNAs were specific to the Y chromosome then no detection should be observed in menstrual blood. This would be the same for vaginal secretions. However, expression (i.e. measurable Ct values) was also detected in vaginal secretions samples. We searched the piRNA database (http://pirnabank.ibab.ac.in/) in order to confirm that these piRNAs were specific to the Y chromosome. While these piRNAs are found on the Y chromosome (sometimes in numerous places throughout the Y chromosome (duplications)), they were also all found on numerous other chromosomes throughout the genome and often numerous times within the same chromosome as well. We previously had only evaluated the piRNAs listed on the Y chromosome without realizing that they were present throughout the genome. We therefore performed a more stringent search of the piRNAs listed as being present on the Y chromosome in order to identify piRNAs only on the Y chromosome. This was a very exhaustive search since over 5000 “hits” were present when the entire Y chromosome sequence was used in the search. This did not mean than 5000 unique piRNAs were present on the Y chromosome. This represented all of the piRNA sequences, with duplications of the same piRNA sequence counted as separate “hits”. Upon analysis of each piRNA within this database search, we were able to determine if each piRNA was identified on other chromosomes. A majority of the piRNAs identified on the Y chromosome were also present on other chromosomes, including the five we previously tested which likely explains the expression levels in menstrual blood and vaginal samples. However, we were able to identify 18 piRNAs that appeared to only be present on the Y chromosome (with up to four different locations (duplications) along the Y chromosome). These piRNAs included the following: piRNA-001255, -001841, -007022, -

009292, -012679, -016319, -016887, -004506, -007238, -007992, -008170, -009405, -009424, -014810, -019938, -013912, -019192 and -022169. Using miDesigner (QIAGEN), we were able to design custom miScript primer assays. However, there is a limitation of 30 bases for the length of the sequence. Seventeen of the eighteen piRNAs identified on the Y chromosome had sequence lengths of 30 or less and therefore we were able to design custom miScript primer assays for these piRNAs (piR-019192 was 32 nt in length). For the analysis of these piRNAs, we used miR-940 for normalization as we had done with our miRNA assays. We evaluated expression of the new 17 miRNA candidates using the following set of samples: semen (two donors); blood, saliva, vaginal secretions, menstrual blood and skin (one donor each), prostate and testes tissue; and appropriate negative controls (semen reverse transcription negative control, assay blank) and DNA (semen). Nine of the 17 piRNAs were initially rejected due to a failure to detect expression in any of the body fluid samples. For 6 of the remaining 8 piRNAs, expression was detected but with considerably high Ct values (>34 in most cases) and for most, if not, all body fluids. Therefore, these candidates were considered relatively low abundance and not ideal for the analysis of semen (at least in terms of demonstrating improvement in sensitivity compared to the existing piRNA assays). For the remaining two piRNAs (piR-009294 and piR-004506), lower Ct values were observed (~15-27 Ct), but these same high levels of expression were obtained in all body fluids and tissues, including vaginal secretions and menstrual blood. Overall, based on the results of this additional set of piRNAs, we were unable to identify a suitable piRNA candidate for semen identification in this limited preliminary screening.

As described above, initial reports of piRNA expression indicated a presence only in sperm-producing cells in mammals and therefore might be ideal candidates for the identification of semen. However, our results do not support this finding as we observed piRNA expression in all body fluids. This was also observed for those piRNAs that we thought were present only on the Y chromosome (i.e. male specific). Recent studies suggest that the originally perceived restricted expression (germ-line only cells) of piRNAs may not be the case and piRNA expression, although possibly enriched in germ-line cells, may also be present throughout the genome [86]. Even if there is considerable lower abundance in somatic tissues, our initial data suggests that we can detect such expression. Our initial studies only involved the analysis of a small number of piRNAs with limited sample sets and therefore it is possible that with further large-scale screening of all known piRNAs that suitable, and highly germ-line cell specific, piRNAs could be identified. Such an undertaking was beyond the scope of the current work.

Initial attempts to evaluate piRNA expression in semen were performed using the miScript system and primer assays specific for individual piRNAs. However, the miScript system is based upon the addition of a polyA tail to the end of the RNA molecules during reverse transcription to allow for subsequent binding of the universal primer in the detection reaction. However, there is evidence that the 3' end of the piRNA molecule is modified by methylation [91-93] and therefore could prevent the addition of the polyA tail. Therefore, it is possible that technical issues could have resulted in the failure to detect piRNA expression in any of the body fluids or issues that was observed for numerous piRNAs evaluated in our study. PiRNA expression was observed for other piRNA candidates in this study and therefore does not appear to be the principal explanation for any challenges experienced with piRNA expression, but it is a factor to consider in possible future piRNA studies.

B. TABLES

Table 1. miRNA candidate identification summary.
The number of potential miRNA candidates for each body fluid from each "batch" of miRNAs (miRBase v.1, v.11, v.13-4, v.15-16 updates; 1,198 total) is listed for both 1ng and 5ng input total RNA levels. NA = not applicable (body fluids not tested at that input).

Body Fluid/Tissue	Input	v.1	v.11	v.13-14	v.15-16	TOTAL
Blood	1ng	NA	7	10	NA	47
	5ng	NA	20	NA	10	
Semen	1ng	NA	13	1	NA	19
	5ng	NA	3	NA	2	
Saliva	1ng	NA	2	2	NA	8
	5ng	NA	2	NA	2	
Vaginal	1ng	NA	3	1	NA	26
	5ng	NA	8	NA	14	
Menstrual Blood	1ng	NA	2	5	NA	22
	5ng	NA	6	NA	9	
Skin	1ng	27	0	0	NA	95
	5ng	27	10	NA	31	
						217

Table 2. Expression Data From Potential "Housekeeping" miRNAs.
The average Ct and standard deviations are shown in each of the forensically relevant body fluids and tissues for the identified potential "housekeeping" miRNAs.

miRNA	All		Blood		Semen		Saliva		Vaginal		Menstrual		Skin	
	Avg Ct	SD	Avg Ct	SD	Avg Ct	SD	Avg Ct	SD	Avg Ct	SD	Avg Ct	SD	Avg Ct	SD
940	20.5	0.4	21.1	0.6	20.5	0.2	20.4	0.3	20.3	0.2	20.2	0.2	20.3	0.1
665	24.7	0.3	25.1	0.5	24.5	0.1	24.7	0.2	24.6	0.1	24.6	0.1	24.6	0.0
1234	24.6	0.5	25.0	0.6	24.5	0.4	24.8	0.5	24.8	0.5	24.3	0.2	24.4	0.0
4270	26.8	0.3	26.8	0.3	26.9	0.5	26.7	0.3	26.4	0.2	26.7	0.2	27.0	0.1
138-1*	27.6	0.3	27.9	0.5	27.2	0.1	27.6	0.2	27.5	0.2	27.7	0.1	27.6	0.2
3190	28.3	0.3	28.5	0.6	28.1	0.1	28.3	0.4	28.3	0.2	28.2	0.2	28.2	0.1
4274	25.8	0.7	26.3	1.5	25.2	0.2	25.8	0.3	25.7	0.3	26.2	0.6	25.6	0.2
3180-3p	27.9	0.6	28.0	0.3	27.5	0.1	27.8	0.3	27.7	0.1	27.8	0.1	28.4	1.5
1233	29.0	0.9	29.2	0.5	28.3	0.1	30.5	0.2	28.9	1.1	28.6	0.3	28.5	0.4
448	32.7	0.2	33.0	0.5	32.7	0.2	32.6	0.2	32.5	0.1	32.6	0.2	32.7	0.1
1238	32.8	0.4	33.1	0.1	32.2	0.8	33.0	0.4	32.8	0.2	32.7	0.1	32.6	0.2
1181	30.7	1.7	30.2	0.8	29.8	2.0	33.8	0.5	30.4	1.4	29.6	0.8	30.4	0.7

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 3. Logistic Regression (LogR) Models for All Body Fluids and Tissues

Body Fluid/Tissue	Logistic Model
Blood	ln [p/(1-p)] = (0.680 miR16*miR451) -(0.452 miR451^2) – 11.95
Semen	ln [p/(1-p)] = (-4.016 * miR891b) – (3.101 * miR892a) + (0.276 * miR892a*miR891b) + 34.414
Saliva	ln [p/(1-p)] = (-1.170 * miR658) – (0.020 * (miR124*)* miR205) + 19.863
Vaginal Secretions	ln [p/(1-p)] =(0.19 * miR124a*miR4286)– (1.695*miR1280*miR4286) + (0.124*miR124a*miR1280*miR4286) + (1.386 miR1280^2) + (0.328 miR4286^2) – (0.14 miR4286* miR124^2) -(0.109 miR124a*miR4286) - (0.029 miR124*miR4286^2)+ 2.730
Menstrual Blood	ln [p/(1-p)] = (-0.005*miR185*(miR144*)*miR144) + (3.718*miR185) - 32.017
Skin	ln [p/(1-p)] = (3.146*miR3169)-(0.060*miR139*miR455-3p*miR3169)-8.823

*the miRNAs listed in each model refer to the miR940 normalized DCt values for each miRNA

Table 4. Summary Statistics for Logistic Regression (LogR) Models for All Body Fluids and Tissues

		Semen	Saliva	Vaginal Secretion	Blood	Menstrual Blood	Skin
Model Validation	Model Chi-square	64.73(p<.001)	95.47(p<.001)	69.23(p<.001)	109.25(p<.001)	46.43(p<.001)	51.13(p<.001)
	Nagelkerke R-square	87.8%	90.9%	76.1%	97.6%	86.2%	87.5%
	Hosmer-Lemsehow Test	.41	1.00	.69	1.00	.90	1.00
Area Under Receiver Operating Curve	Estimated Area under Operating Curve	99.2%	99.5%	95.0%	99.9%	99.4%	99.7%
	95 Percent Confidence Interval	97.6%-100%	98.7%-100%	90.0%-100%	99.7%-100%	98.2%-100%	99.1%-100%
	Optimal Threshold	0.5	0.63	.60	0.5	0.50	0.54
Classification: Selected Cases	True Positive Rate	92.3%	91.7%	79.2%	96%	87.5%	88.9%
	False Positive Rate	0%	0%	0%	0%	0%	0%
	Correct Classification Rate	99%	98.1%	94.8%	99%	99.0%	99.0%
Classification: Unselected Cases	True Positive Rate	100%	100%	100%	100%	100%	100%
	False Positive Rate	0%	0%	0%	0%	0%	0%

Table 5. Expected and Observed Body Fluid Identification Results for Two-Fluid Admixed Body Fluid Samples.

The colors under the dried and liquid component columns represent the expected results for the admixed body fluid samples. The observed results are indicated by colored squares under each body fluid heading (red – blood positive; orange – possible blood; yellow – semen positive; blue – saliva positive; pink – menstrual blood positive; white – no detection).

Mixture	Dried Component	Liquid Component	Volume of liquid fluid	Blood	Semen	Saliva	Vaginal	Mesntrual	Skin
Blood-Semen	Blood	Semen	50ul						
			25ul						
			10ul						
			5ul						
Blood-Saliva	Blood	Saliva	50ul						
			25ul						
			10ul						
			5ul						
Saliva-Semen	Saliva	Semen	50ul						
			25ul						
			10ul						
			5ul						
Vaginal-Semen	Vaginal Secretions	Semen	50ul						
			25ul						
			10ul						
			5ul						
Menstrual-Semen	Menstrual Blood	Semen	50ul						
			25ul						
			10ul						
			5ul						
Vaginal-Saliva	Vaginal Secretions	Saliva	50ul						
			25ul						
			10ul						
			5ul						

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 6. Comparison of miRNA and mRNA Profiling Results for Environmentally Impacted Blood Samples.

50µl bloodstains (cotton cloth) were exposed to various temperatures and outdoor conditions (conditions listed in 'sample' column). For miRNA profiling (columns 2 and 3), the logit value (between 0 and 1) obtained using the logR model is listed. A positive result was determined if it was above the designated 0.5 threshold. The (*) indicates samples that appear positive on the 2D scatterplot but values were below the threshold. For mRNA profiling (columns 4, 5 and 6), the RFU values for the blood biomarkers ANK1 and ALAS2 (if observed) are recorded. A 50 RFU threshold was used for detection.

Sample	Value	Result (threshold 0.5)	ANK1	ALAS2	Result
Room Temp 1yr - Male	1.0	yes	503	8336	$yes_{(2/2)}$
Room Temp 2yr - Male	1.0	yes	175	4112	$yes_{(2/2)}$
37oC 1 yr - Male	1.0	yes		1134	$yes_{(1/2)}$
37oC 2 yr - Male	1.0	yes		456	$yes_{(1/2)}$
56oC 1 yr - Male	1.0	yes		414	$yes_{(1/2)}$
56oC 2 yr - Male	0.4	yes*		100	$yes_{(1/2)}$
Room Temp 1 yr - Female	1.0	yes		8833	$yes_{(1/2)}$
Room Temp 2 yr - Female	0.0	yes*		138	$yes_{(1/2)}$
37oC 1 yr - Female	0.0	yes*		42	?
37oC 2 yr - Female	0.7	yes		404	$yes_{(1/2)}$
56oC 1 yr - Female	0.0	yes		2365	$yes_{(1/2)}$
56oC 2 yr - Female	0.0	yes		1446	$yes_{(1/2)}$
Patio - 1 month	1.0	yes	791	9088	$yes_{(2/2)}$
Patio - 2 months	0.2	yes*		1894	$yes_{(1/2)}$
Patio - 3 months	0.0	yes*		841	$yes_{(1/2)}$
Patio - 4 months	0.5	yes		1484	$yes_{(1/2)}$
Sun - 1 week	0.0	no			no
Sun - 1 month	0.0	no			no
Sun - 2 months	0.0	no			no
Shade - 1 month	0.0	no			no
Shade - 2 months	0.0	no			no
Shade - 6 months	0.0	no			no
Car Back Seat - 1 week - Male	1.0	yes	2868	9051	$yes_{(2/2)}$
Car Back Seat - 1 week - Female	0.9	yes	3035	8991	$yes_{(2/2)}$
Car Trunk - 2 weeks - Male	1.0	yes	1304	9115	$yes_{(2/2)}$
Car Trunk - 2 weeks - Female	1.0	yes	1388	9110	$yes_{(2/2)}$
		20/26 (77%)			19/26 at least one marker (73%) 7/26 both markers (27%)

Table 7. Comparison of miRNA and mRNA Profiling Results for Environmentally Impacted Semen Samples.

50µl semen stains (cotton cloth) were exposed to various temperatures and outdoor conditions (conditions listed in 'sample' column). For miRNA profiling (columns 2 and 3), the logit value (between 0 and 1) obtained using the logR model is listed. A positive result was determined if it was above the designated 0.5 threshold. For mRNA profiling (columns 4, 5 and 6), the RFU values for the semen biomarkers PRM2 and TGM4 (if observed) are recorded. A 50 RFU threshold was used for detection.

Sample	Value	Result Threshold 0.5	PRM2	TGM4	Result
Room temp - 1 yr - Male 1	1.0	yes	9411	9338	$yes_{(2/2)}$
Room temp - 2 yr - Male 1	1.0	yes	9441	9381	$yes_{(2/2)}$
37oC - 1 yr - Male 1	1.0	yes	9296	9334	$yes_{(2/2)}$
37oC - 2 yr - Male 1	1.0	yes	7840	2443	$yes_{(2/2)}$
56oC - 1 yr - Male 1	1.0	yes	517	293	$yes_{(2/2)}$
56oC - 2 yr - Male 1	0.9	yes	1047	1288	$yes_{(2/2)}$
Room temp - 1 yr - Male 2	1.0	yes	2459		$yes_{(1/2)}$
Room temp - 2 yr - Male 2	1.0	yes	415		$yes_{(1/2)}$
37oC - 1 yr - Male 2	1.0	yes	597	91	$yes_{(2/2)}$
37oC - 2 yr - Male 2	1.0	yes	115		$yes_{(1/2)}$
56oC - 1 yr - Male 2	1.0	yes	415	245	$yes_{(2/2)}$
56oC - 2 yr - Male 2	1.0	yes	1308	217	$yes_{(2/2)}$
Outside covered - 1 day - M2	0.5	yes	3121	81	$yes_{(2/2)}$
Outside covered - 3 days - M2	0.0	no			no
Outside covered - 1 week -M2	0.0	no	1919		$yes_{(1/2)}$
Outside covered - 3 days -M1	0.0	no	676	449	$yes_{(2/2)}$
UV >1 yr	0.1	no	595		$yes_{(1/2)}$
Sun - 1 day	0.0	no	330		$yes_{(1/2)}$
Sun - 2 days	0.1	no	4647		$yes_{(1/2)}$
Sun - 12 days	0.0	no			no
Sun - 1 month	0.0	no			no
Patio - 1 day	1.0	yes	7650	5464	$yes_{(2/2)}$
Patio - 2 day	1.0	yes	3040	621	$yes_{(2/2)}$
Patio - 12 days	0.5	yes	1044	75	$yes_{(2/2)}$
Patio - 1 month	0.0	no	6102	924	$yes_{(2/2)}$
Shade - 1 day	0.0	no	3874		$yes_{(1/2)}$
Shade - 2 days	0.0	no	4311		$yes_{(1/2)}$
Shade - 12 days	0.0	no	940		$yes_{(1/2)}$
Shade - 1 month	0.0	no			no
		16/29 (55%)			25/29 at least one marker (86%) 15/29 both markers (52%)

Table 8. Comparison of miRNA and mRNA Profiling Results for Environmentally Impacted Saliva Samples.

50µl saliva stains (cotton cloth) were exposed to various temperatures and outdoor conditions (conditions listed in 'sample' column). For miRNA profiling (columns 2 and 3), the logit value (between 0 and 1) obtained using the logR model is listed. A positive result was determined if it was above the designated 0.63 threshold. For mRNA profiling (columns 4, 5 and 6), the RFU values for the saliva biomarkers HTN3/HTN1 and STATH (if observed) are recorded. A 50 RFU threshold was used for detection.

Sample	Value	Result Threshold 0.63	HTN3	HTN1	STATH	Result
Room temp 1 yr - Male	0.86	yes	3060	528	280	$yes_{(2/2)}$
Room temp 2 yr - Male	0.95	yes	1678	272	170	$yes_{(2/2)}$
37oC 1 yr - Male	0.97	yes	569	155		$yes_{(1/2)}$
37oC 2 yr - Male	0.94	yes	297			$yes_{(1/2)}$
56oC 1 yr - Male	1.00	yes	311			$yes_{(1/2)}$
56oC 2 yr - Male	0.99	yes				no
Room temp 1 yr - Female	0.97	yes	548	478	292	$yes_{(2/2)}$
Room temp 2 yr - Female	0.99	yes	734	524		$yes_{(1/2)}$
37oC 1 yr - Female	0.98	yes	1022		77	$yes_{(2/2)}$
37oC 2 yr - Female	0.98	yes				no
56oC 1 yr - Female	0.97	yes	159			$yes_{(1/2)}$
56oC 2 yr - Female	0.99	yes				no
Outside covered - 1 day - M2	1.00	yes				no
Outside uncovered - 1 day - M2	0.98	yes				no
Outside uncovered - 3 days - M2	1.00	yes				no
Outside covered - 1 day - F2	1.00	yes				no
Sun - 1 day	0.22	no				no
Sun - 2 days	0.88	yes				no
Sun - 1 week	0.06	no				no
Sun - 1 month	0.02	no				no
Shade - 1 day	0.96	yes	9607	4880	2053	$yes_{(2/2)}$
Shade - 2 days	0.93	yes	8958	3550	1712	$yes_{(2/2)}$
Shade - 1 month	0.14	no				no
Patio -1 day	0.97	yes	960	190	142	$yes_{(2/2)}$
Patio - 2 days	0.88	yes	9469	3542	1851	$yes_{(1/2)}$
Patio - 1 week	0.00	no	9579	4297	2224	$yes_{(2/2)}$
Patio - 1 month	0.84	yes				no
56oC on denim 2 weeks - S2	0.99	yes				no
56oC on denim 2 weeks - S3	1.00	yes				no
		24/29 (83%)	15/29 at least one marker (52%) 9/29 both markers (31%)			

Table 9. Comparison of miRNA and mRNA Profiling Results for Environmentally Impacted Vaginal Secretions Samples.

Vaginal secretions samples (cotton swabs, ½ swabs used for analysis) were exposed to various temperatures and outdoor conditions (conditions listed in 'sample' column). For miRNA profiling (columns 2 and 3), the logit value (between 0 and 1) obtained using the logR model is listed. A positive result was determined if it was above the designated 0.6 threshold. For mRNA profiling (columns 4, 5 and 6), the RFU values for the vaginal secretions biomarkers CYP2B7P1 and VAG2 (unpublished biomarker) (if observed) are recorded. A 50 RFU threshold was used for detection.

Sample	Value	Result Threshold 0.60	CYP2B7P1	VAG2	Result
Room temp 1 yr - Female 1	0.85	yes	2493	164	$yes_{(2/2)}$
Room temp 2 yr - Female 1	0.64	yes	124		$yes_{(1/2)}$
37oC 1 yr - Female 1	0.76	yes	1294		$yes_{(1/2)}$
37oC 2 yr - Female 1	0.91	yes	742	480	$yes_{(2/2)}$
56oC 1 yr - Female 1	0.82	yes			no
56oC 2 yr - Female 1	0.86	yes			no
Room temp 1 yr - Female 2	0.37	no			no
Room temp 2 yr - Female 2	0.53	no			no
37oC 1 yr - Female 2	0.47	no			no
37oC 2 yr - Female 2	0.45	no			no
56oC 1 yr - Female 2	0.52	no			no
56oC 2 yr - Female 2	0.39	no			no
Outside covered - 1 day - F1	0.14	no	137		$yes_{(1/2)}$
Outside covered - 1 day - F2	0.02	no	184	83	$yes_{(2/2)}$
Outside covered - 2 dys - F2	0.19	no			no
Outside covered - 1 week - F2	0.11	no			no
Outside covered - 1 month - F2	0.00	no			no
Outside uncovered - 1 day - F2	0.06	no			no
Outside uncovered - 1 month - F2	0.00	no			no
		6/19 (32%)			6/19 at least one marker (32%) 3/19 both markers (16%)

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 10. Comparison of miRNA Candidates From Published Studies

miRNA candidates by group	Hanson et al		Zubakov et al		Courts		Wang			van der Meer et al	Uchimoto et al
	Qiagen miScript (SYBR)		Microarray	TaqMan (LifeTech)	Biochips	SYBR green	qPCR array	TaqMan (LifeTech)	TaqMan (LifeTech	Stem-loop/CE	TaqMan (Life Tech)
	Original	New									
Blood	451	451			451	451				451	451
	16	16					16	16	16		
			20a	20a							
			106a	106a							
			185	185							
				144							
					126	126					
					150	15					
							486	486			
Semen	135b										
	10b										
		891b									
		892a									
			943	943							
			135a	135a							
			10a	10a							
			507	507							
				891a			891a	891a			
							888	888			
Saliva	658	658							658		
	205	205			205	205			205	205	205
		124*									
			583								
			518c*								
			208b								
					200c	200c					
					203	203					
							138-2				
Vaginal Secretions	124a	124a					124a				
	372										
		1280									
		4286									
			617								
			891a								
Menstrual Blood	451										
	412										
		185									
		144*									
		144	144								
			185*								
							214	214			
Skin		494									
		455-3p									
		3169									

C. FIGURES

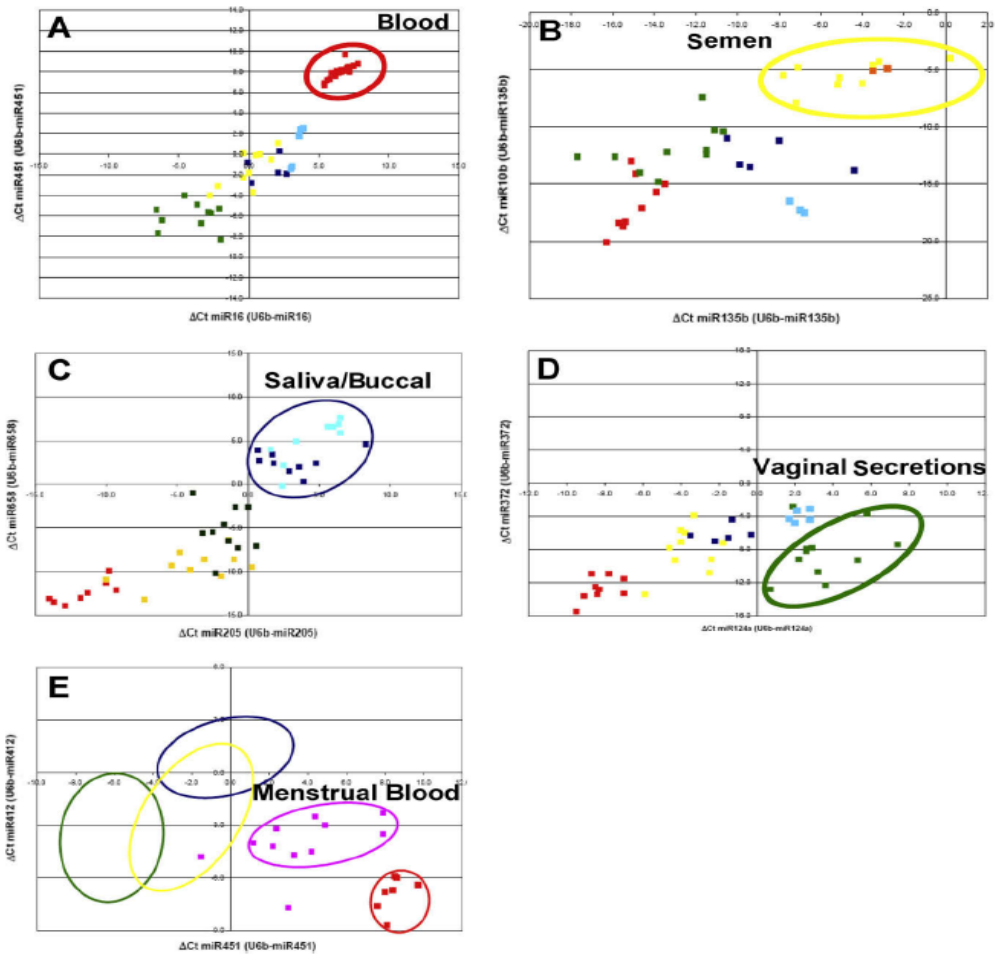

Figure 1. Original Preliminary miRNA Assays for the Identification of Forensically Relevant Biological Fluids. (A) Blood assay (miR451/miR16); (B) semen assay (miR10b/miR135b), (C) saliva assay (miR658/miR205), (D) vaginal secretions assays (miR124a/miR372), (E) menstrual blood assay (miR451/miR412). Individual body fluid data points are represented by colored squares: red, blood; yellow, semen; orange, semen from vasectomized males; dark blue, saliva; light blue, buccal swabs; green, vaginal secretions; pink, menstrual blood. Each body fluid of interest cluster is circled.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

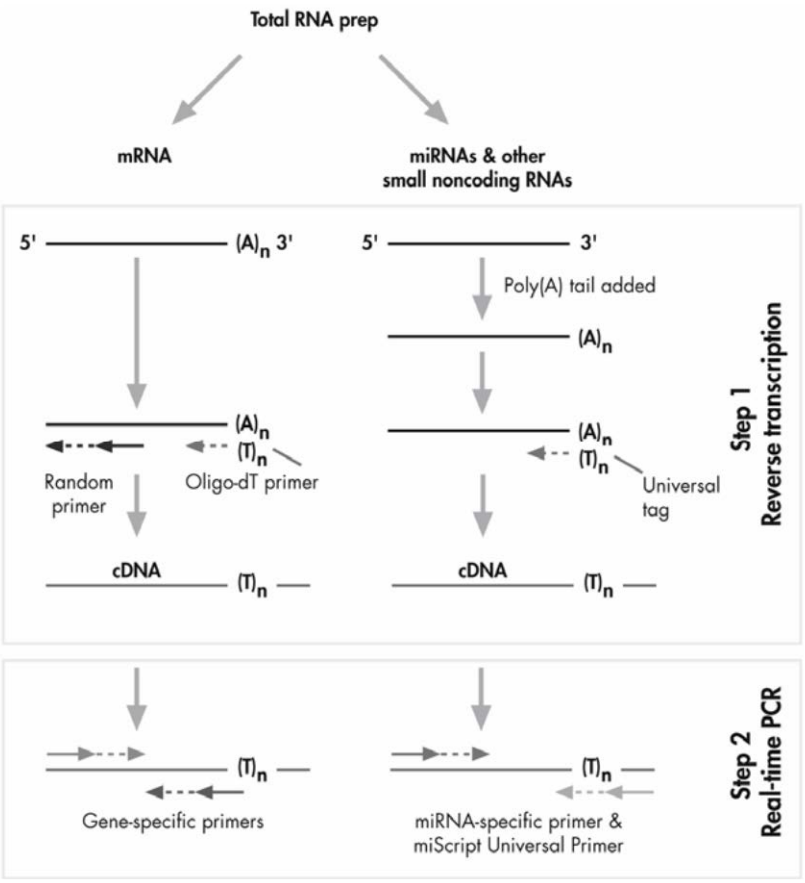

Figure 2. miScript principle. miRNAs and other noncoding RNAs are poly-adenylated by poly(A) polymerase and subsequently converted into cDNA by reverse transcriptase with oligo-dT priming. mRNAs are converted into cDNA by reverse transcriptase using both oligo-dT and random priming. The cDNA is then used for real-time PCR quantification of any miRNA (using the miScript Primer Assay and the miScript Universal Primer), noncoding RNA (using a target-specific primer and the miScript Universal Primer), or mRNA (using gene-specific primers).

**Figure and description are directly from the miScript system handbook (QIAGEN, version 04/2007)

Figure 3. miRNA candidate screening (miScript System). X-axis: miRNA, Y-axis: raw Ct values (miScript SYBR Green real time PCR assays). Expression data from each of the forensically relevant body fluids and tissues is depicted by a colored line: blood (red), semen (yellow), saliva (blue), vaginal secretions (green), menstrual blood (pink) and skin (peach). The dashed horizontal line is placed at the 40 Ct level since that is the recommended cycle number for the miScript real time PCR assays (50 cycles used in initial candidate screening in order to identify potential low abundance candidates). Colored vertical rectangles indicate potential candidates for a particular body fluid (menstrual blood – pink, saliva – blue, semen – yellow.

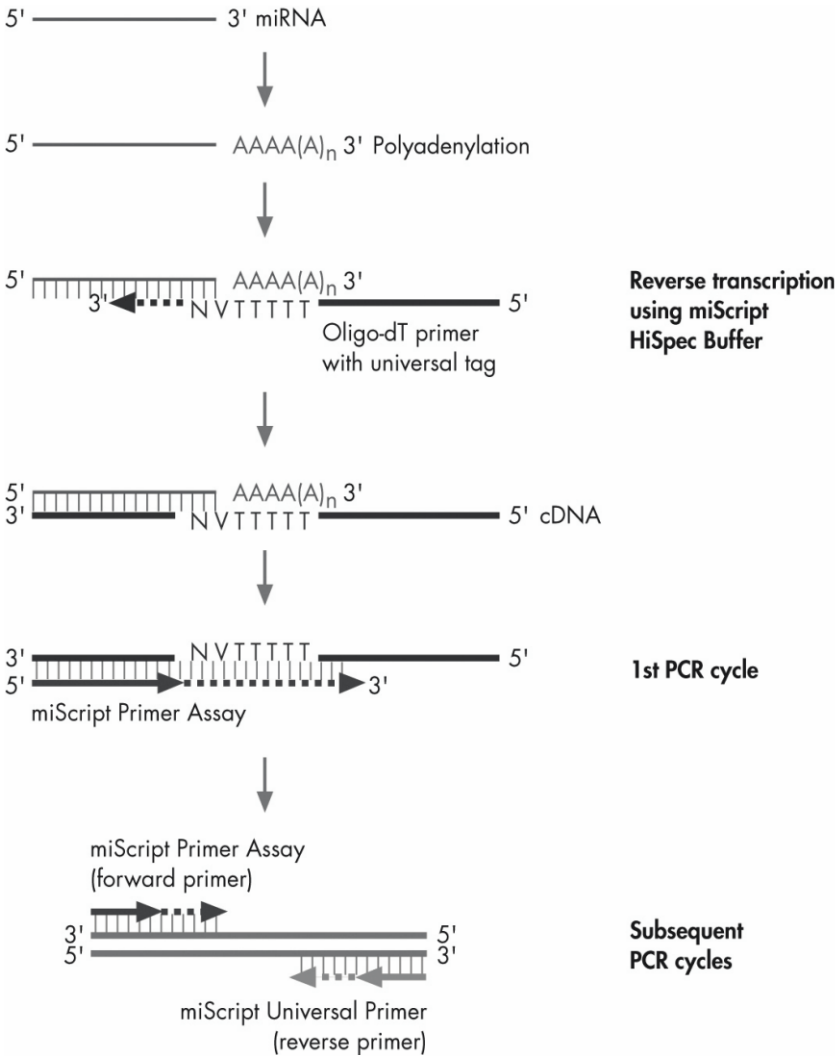

Figure 4. Selective conversion of mature miRNAs into cDNA in miScript HiSpec Buffer. In a reverse transcription reaction with miScript HiSpec Buffer, mature miRNAs are polyadenylated by poly(A) polymerase and converted into cDNA by reverse transcriptase with oligo-dT priming. The cDNA is then used for real time PCR quantification of mature miRNA expression.

**Figure and description are directly from the miScript PCR system handbook (QIAGEN, version 05/2012)

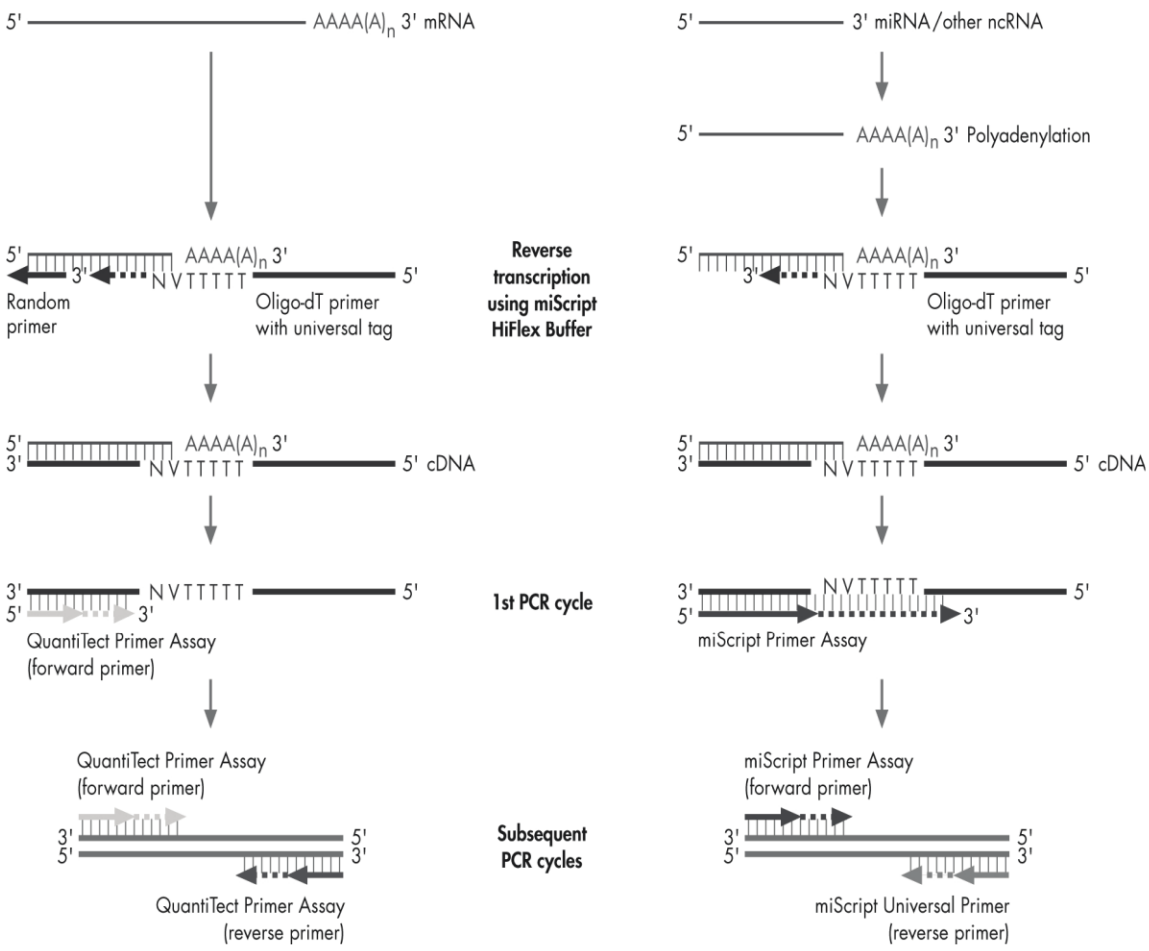

Figure 5. Simultaneous conversion of all RNA species into cDNA in miScript HiFlex Buffer. In a reverse transcription reaction with miScript HiFlex Buffer, miRNAs and other noncoding RNAs (ncRNAs) are polyadenylated by poly(A) polymerase and converted into cDNA by reverse transcriptase with oligo-dT priming. mRNAs are converted into cDNA by reverse transcriptase using both oligo-dT and random priming. Detection of mature miRNA, precursor miRNA, other ncRNA, and mRNA can be performed using the appropriate assay.

**Figure and description are directly from the miScript PCR system handbook (QIAGEN, version 05/2012)

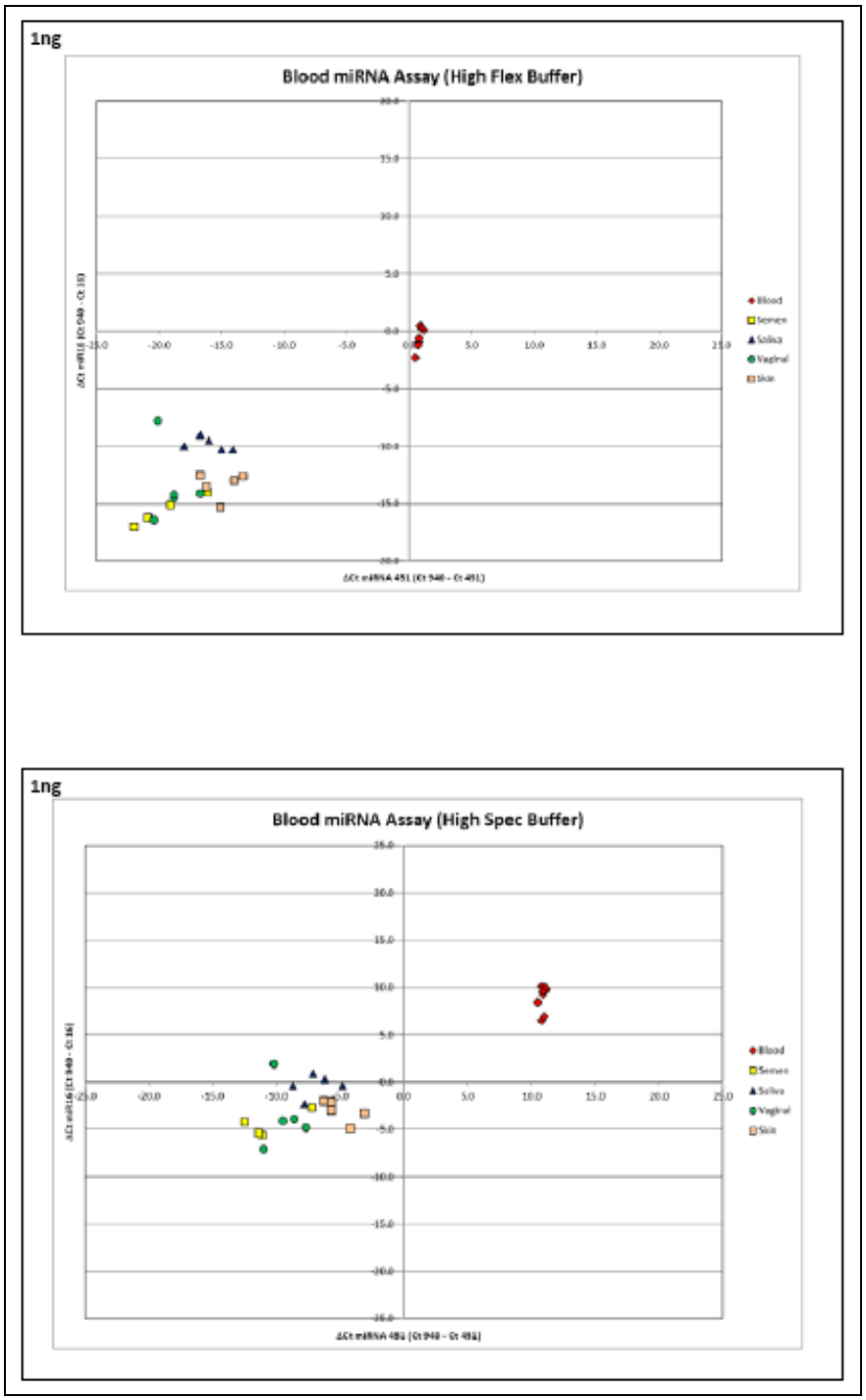

Figure 6. Comparison of the Blood (miR-451/miR-16) Assay Using the HiSpec and HiFlex buffers in the miScript II RT kit. X-axis: miR-451 ΔCt (miR940 – miR451), Y-axis: miR-16 ΔCt (miR940 – miR16). Known body fluid samples are represented by colors and shape: blood – red diamonds, semen – yellow squares, saliva – blue triangles, vaginal secretions – green circles, skin - peach squares.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

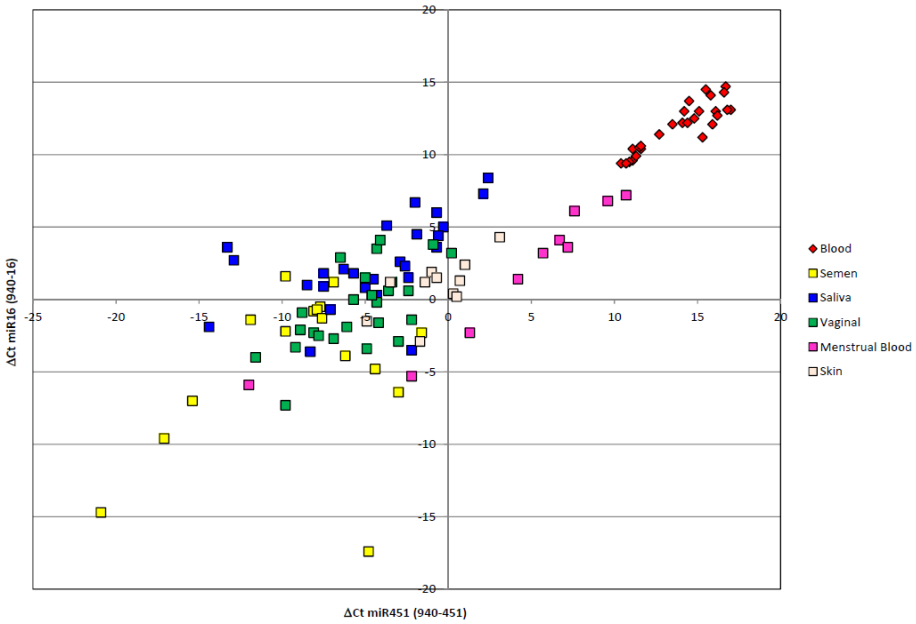

Figure 7. Final Blood miRNA Assay. X-axis: miR-451 ΔCt (miR940 – miR451), Y-axis: miR-16 ΔCt (miR940 – miR16). Known body fluid samples are represented by colors and shape: blood – red diamonds, semen – yellow squares, saliva – blue squares, vaginal secretions – green squares, menstrual blood – pink squares and skin - peach squares.

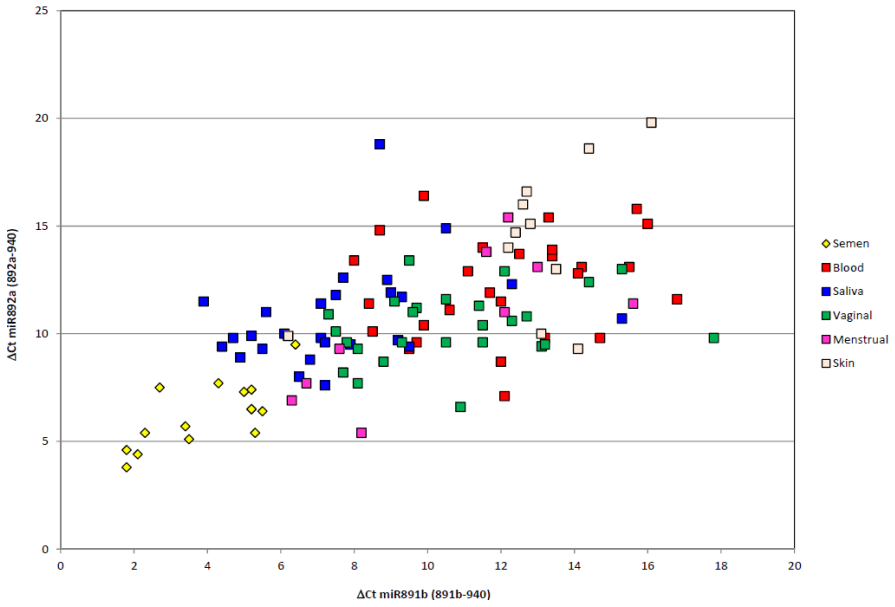

Figure 8. Final Semen miRNA Assay. X-axis: miR-891b ΔCt (miR891b – miR940), Y-axis: miR-892a ΔCt (miR892a – miR940). Known body fluid samples are represented by colors and shape: semen – yellow diamonds, blood – red squares, saliva – blue squares, vaginal secretions – green squares, menstrual blood – pink squares and skin - peach squares.

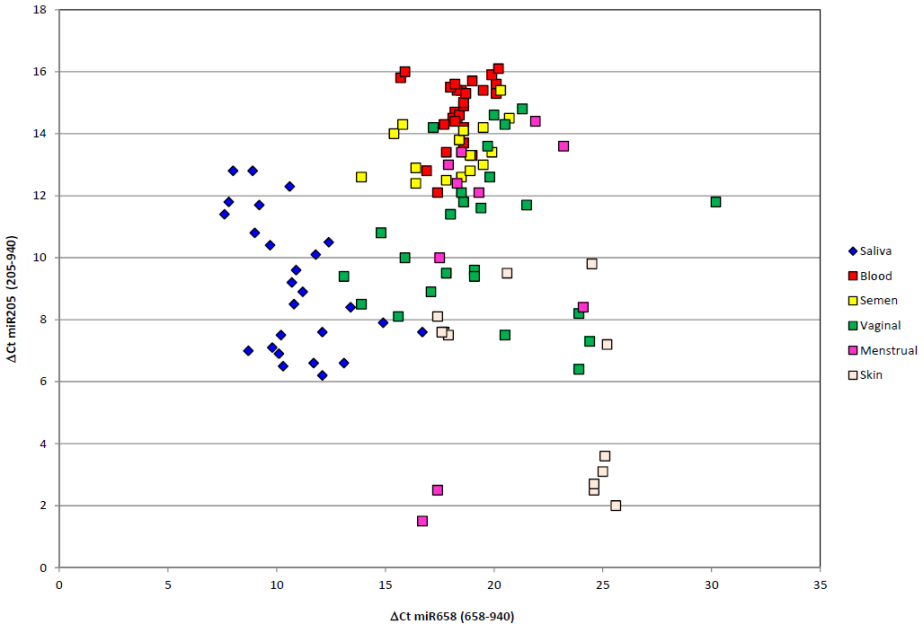

Figure 9. Final Saliva miRNA Assay. X-axis: miR658 ΔCt (miR658 – miR940), Y-axis: miR-205 ΔCt (miR205 – miR940). Known body fluid samples are represented by colors and shape: saliva – blue diamonds, semen – yellow squares, blood – red squares, vaginal secretions – green squares, menstrual blood – pink squares and skin - peach squares.

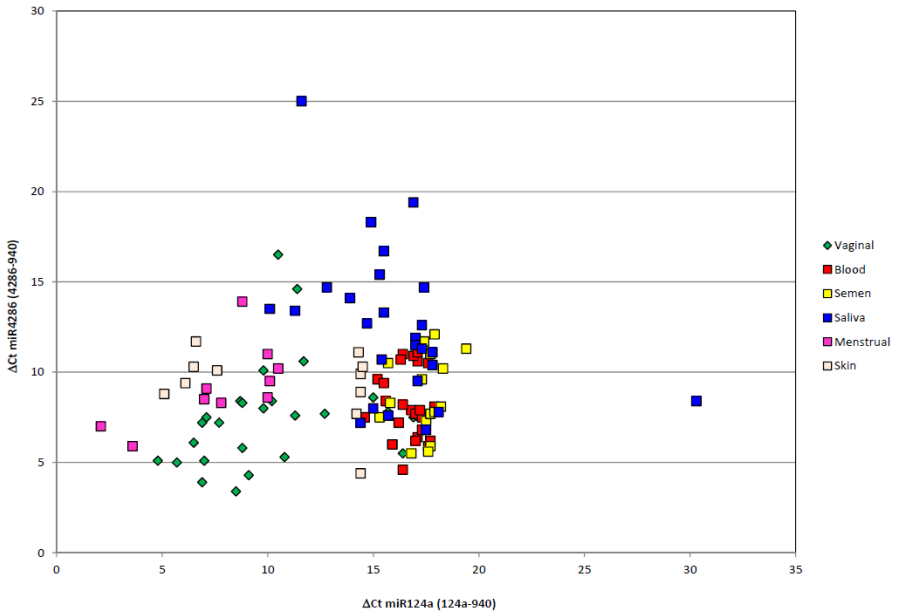

Figure 10. Final Vaginal Secretions miRNA Assay. X-axis: miR124a ΔCt (miR124a – miR940), Y-axis: miR-4286 ΔCt (miR4286 – miR940). Known body fluid samples are represented by colors and shape: vaginal secretions – green diamonds, saliva – blue squares, semen – yellow squares, blood – red squares, menstrual blood – pink squares and skin - peach squares.

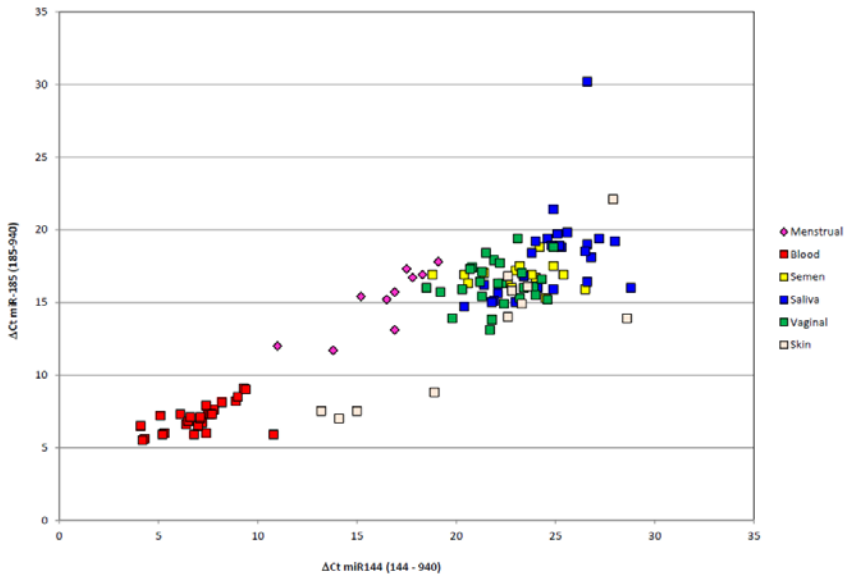

Figure 11. Final Menstrual Blood miRNA Assay. X-axis: miR144 ΔCt (miR144 – miR940), Y-axis: miR-185 ΔCt (miR185 – miR940). Known body fluid samples are represented by colors and shape: menstrual blood – pink diamonds, vaginal secretions – green squares, saliva – blue squares, semen – yellow squares, blood – red squares and skin - peach squares.

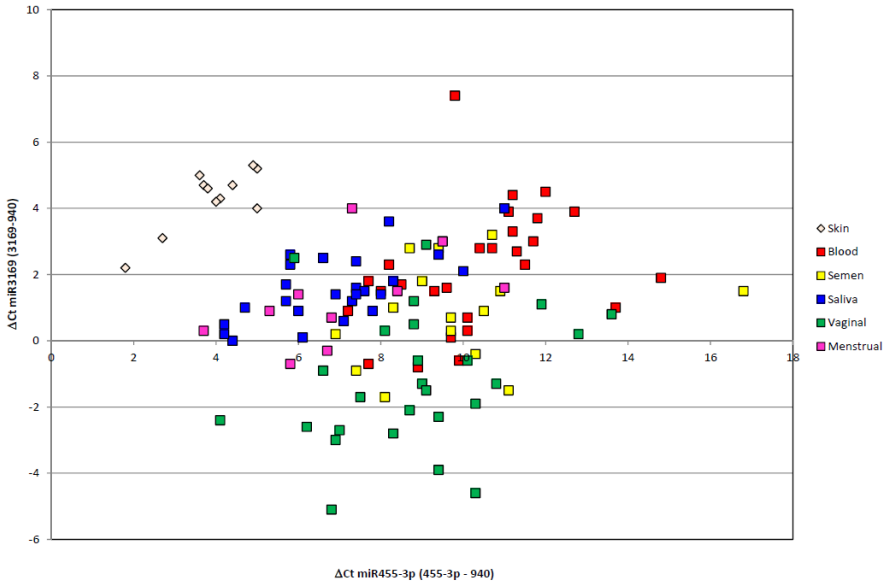

Figure 12. Final Skin miRNA Assay. X-axis: miR455-3p ΔCt (miR455-3p – miR940), Y-axis: miR-3169 ΔCt (miR3169 – miR940). Known body fluid samples are represented by colors and shape: skin - peach diamonds, vaginal secretions – green squares, saliva – blue squares, semen – yellow squares, blood – red squares and menstrual blood – pink squares.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

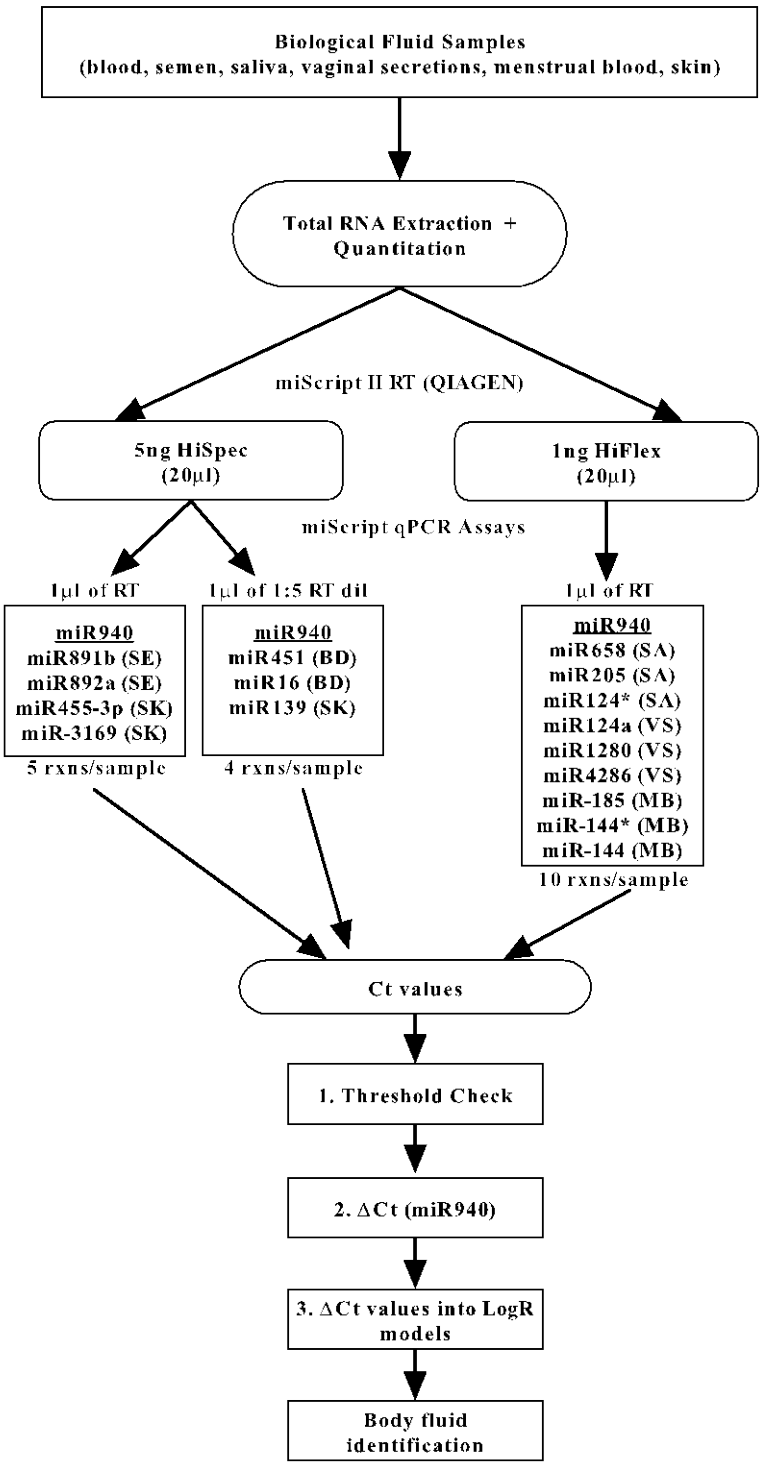

Figure 13. Experimental Schema – Second Generation miRNA Profiling Assays

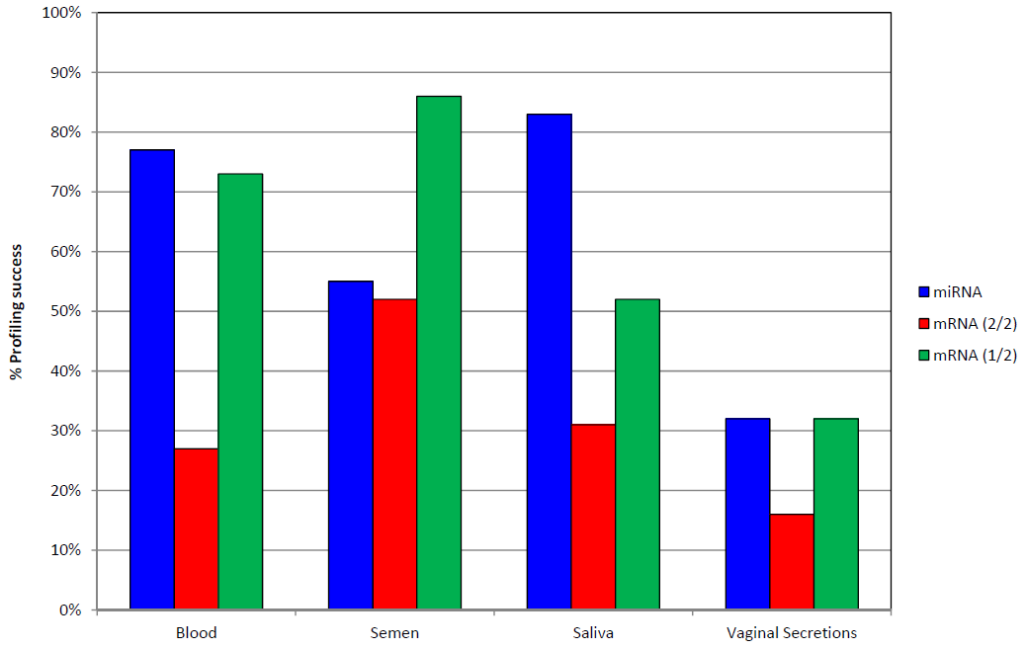

Figure 14. Comparison of miRNA and mRNA Profiling Success Rates for Environmentally Impacted Body Fluid Samples. X-axis: body fluid, Y-axis: Percent profiling success (number of positive samples out of total number tested).

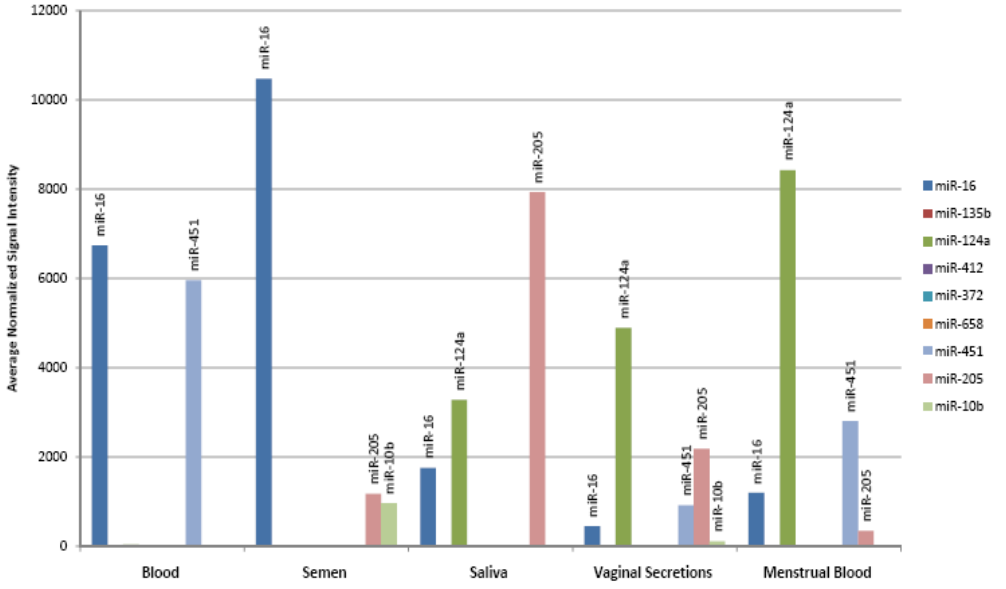

Figure 15. qPNA miRNA Assay For Body Fluid Identification. X-body fluid; Y-axis – signal intensity. Expression of each of the nine miRNAs from the original body fluid identification panel was evaluated using qNPA assays (HT Genomics). Each bar represents detection of a miRNA in the various body fluids.

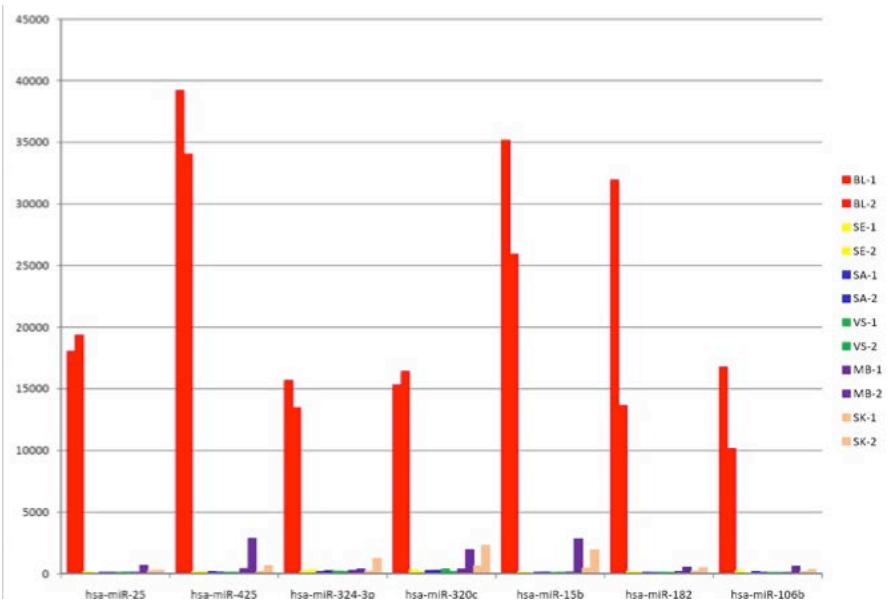

Figure 16. Identification of Blood miRNA Candidates Using qNPA Whole Transcriptome Profiling. X-axis – miRNA; Y-axis – signal intensity. Expression of each miRNA in six forensically relevant body fluids and tissues is depicted by colored bars (blood-red; semen-yellow; saliva-blue; vaginal secretions-green; menstrual blood-purple; skin-peak).

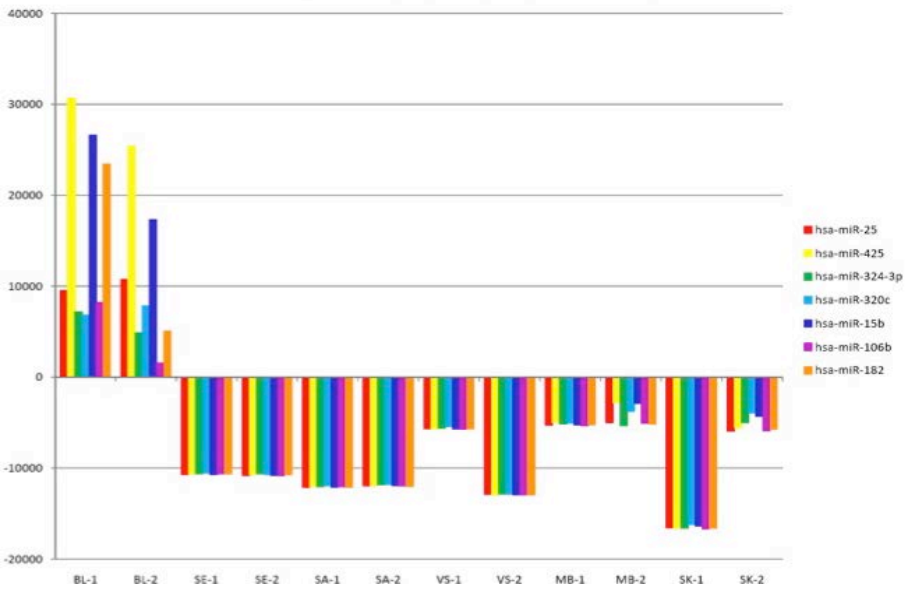

Figure 17. Normalized Expression Data for qNPA Blood miRNA Candidates. X-axis – sample (BL=blood, SE=semen, SA=saliva, VS=-vaginal secretions, MB=menstrual blood, SK=skin); Y-axis – normalized signal intensity (Δsignal using miR-1181). Normalization of the obtained signals resulted in positive values for all miRNA candidates in the blood samples (BL-1 and BL2) and negative values for all other body fluids.

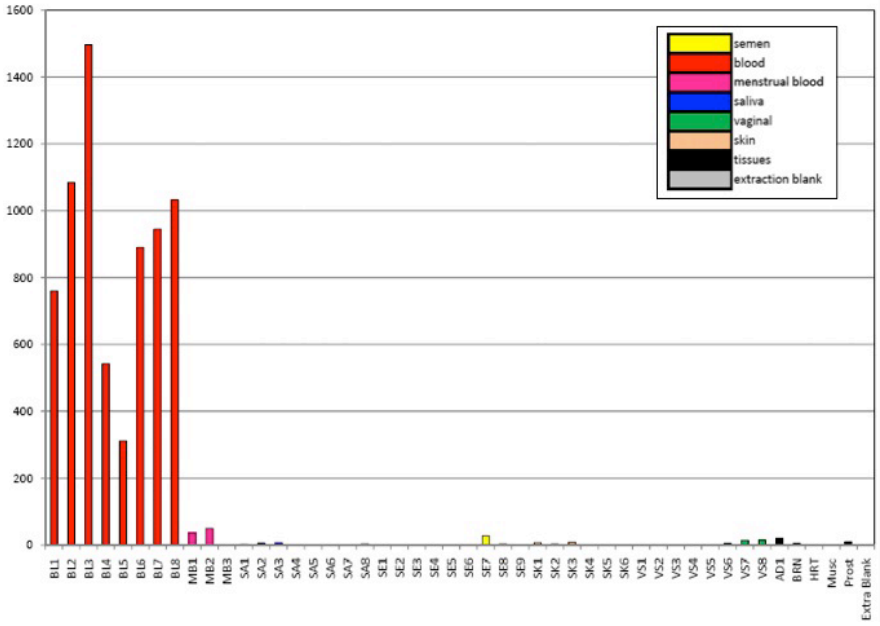

Figure 18. Additional sample testing for qNPA Identified Blood miRNA candidate miR-25. X-axis – sample (BL=blood, SE=semen, SA=saliva, VS=-vaginal secretions, MB=menstrual blood, SK=skin, ADI =adipose, BRN=brain, HRT=heart, MUSC=muscle, PROST=prostate); Y-axis –signal intensity. Expression of miR-25 in the various body fluids is represented by color bars, with higher expression observed for all blood samples (red bars).

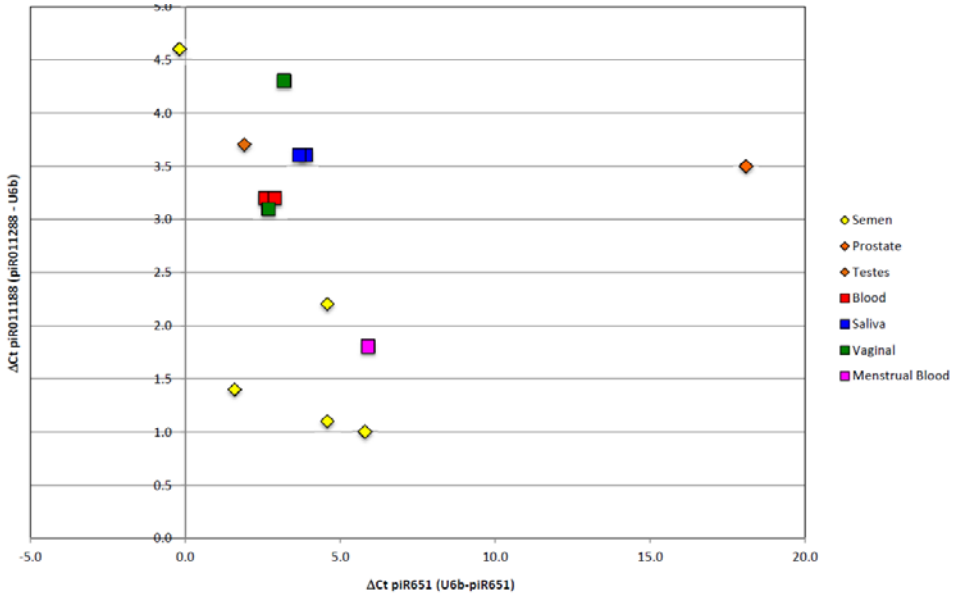

Figure 19. piRNA Expression Assay (U6b-normailzed data), 2D scatterplot. X-axis: normalized piR651 (U6b-piR651), Y-axis: normalized piR-011188 (piR-011188-U6b). Individual body fluid data points are represented by colored symbols: red square, blood; yellow diamond, semen; orange diamond, prostate; brown diamond, testes; blue square, saliva; green square, vaginal secretions; pink square, menstrual blood.

IV. CONCLUSIONS

A. Discussion of findings

Previously, we reported the first comprehensive evaluation of miRNA expression in dried, forensically relevant biological fluids (blood, semen, saliva, vaginal secretions and menstrual blood) which resulted in the development of a panel of nine miRNAs (miR451, miR16, miR135b, miR10b, miR658, miR205, miR124a, miR372, miR412) that permitted the identification of the body fluid of origin of forensic biological stains using as little as 50pg of total RNA [80,81]. However, this previous work involved an evaluation of only 452 miRNAs. Since this initial study, hundreds to thousands of novel miRNAs have been deposited into miRBase. In the current work, we evaluated an additional 746 miRNAs (1,198 total) in order to identify additional highly specific and sensitive miRNA biomarkers for the identification of forensically relevant body fluids. Additionally, we also performed the first comprehensive evaluation of miRNA expression in skin for potential future use with touch DNA samples. After the evaluation of the additional 746 miRNAs, we have expanded our original body fluid identification panel from nine miRNAs to 16 miRNAs (five of the original nine miRNAs and 11 novel miRNAs) which permits an identification of all forensically relevant biological fluids and tissues: blood, semen, saliva, vaginal secretions, menstrual blood and skin. All develop assays demonstrated picogram-level sensitivities, performed well with admixed body fluid samples (with the exception of the semen assay) and were suitable for use with environmentally compromised body fluid samples. A hypothesis preliminarily tested in the current work was that miRNA profiling assays would be more suitable for use with environmentally compromised and degraded samples compared to mRNA profiling due to the much smaller size of miRNAs. However, at this time it cannot be conclusively determined whether miRNA profiling is superior to mRNA profiling for the analysis of environmentally impacted samples as truly environmentally impacted samples (exposed to outside conditions – heat, light, humidity and rain) may be largely affected by other factors such as sample loss (e.g. rain, insect activity) rather than just degradation of the biological material present. Future work will include and evaluation of artificially degraded RNA samples (RNase digestions) in order to assess the effects of only RNA degradation to determine if miRNA profiling will be more successful for the analysis of these samples. The results of the environmentally impacted samples provide support for the potential routine use of RNA (miRNA and/or mRNA) profiling assays in forensic casework.

Uniquely, in the current work we developed a novel statistical approach for the analysis of normalized miRNA express data involving the use of logistic regression models. This permits an objective “prediction” of the presence of the above-mentioned body fluids and tissues using a multi-candidate approach rather than the two-candidate approach previously used. These models provide the necessary refined analysis interpretation metrics that were lacking in our original miRNA assays in which a subjective determination of the presence of a particular body fluid was made by comparison to known body fluid sample “2D expression clusters”. There was no previous definitive metric by which to determine a positive or negative result. The logistic regression models provide an objective and definitive means by which to interpret the miRNA expression data from body fluid stains.

Since the development of our original miRNA profiling assays, numerous other studies have reported the use of miRNA profiling for body fluid identification [94-98] demonstrating the continued interest in this type of approach for body fluid identification. It is encouraging to see

with additional studies and validations that several of our miRNA candidates are consistently used for particular body fluids using various assays and technologies (blood – miR451, miR16; saliva – miR205; vaginal – miR124a) (Table 10). It has been previously demonstrated that miRNA candidates cannot always be confirmed when different assays and detection technologies are used [97,98]. This presents a challenge when large-scale efforts for screening are performed using a different technology (micro-arrays, whole transcriptome analysis) than will be used for development of specific body fluid identification assays. It is unclear at this time why such a discrepancy between assay types or detection technologies exists. However, the confirmed use of several miRNA markers amongst different studies and assay platforms is encouraging for the development of a universal panel of miRNA biomarkers for body fluid identification.

B. Implications for policy and practice

The serology-based methods routinely used in forensic casework for the identification of biological fluids are costly in terms of time and sample and have varying degrees of sensitivity and specificity. The routine use of body fluid identification methods prior to DNA analysis awaits the development of suitable molecular genetics based methods that are fully compatible with the current DNA analysis pipeline. The proposed work sought to develop a novel molecular genetics-based strategy for the identification of the body fluid origin of forensic biological evidence by miRNA and piRNA profiling. Such an approach might be expected to result in significant improvements in the specificity and sensitivity of body fluid identification methods and provide a strategy better suited for the analysis of environmentally impacted or degraded samples. The ability to definitively identify a specific body fluid and, in some instances, implicate it as being the source of a genetic profile will aid in the prosecution and resolution of some cases. The results of the current work fully support the potential use of miRNA profiling as a suitable strategy for a determination of the body fluid or tissue source of origin of forensic samples, including environmentally impacted samples. The pictogram-level sensitivity of the developed miRNA assays also demonstrate their suitability of use with forensic samples that contain only trace amounts of biological material.

C. Implications for further research

We have successfully developed miRNA profiling assays as well as a novel statistical approach (logistic regression modeling) for the identification of all forensically relevant body fluids and tissues (blood, semen, saliva, vaginal secretions, menstrual blood and skin). While our initial evaluation of environmentally impacted samples did not provide a clear indication of whether miRNA profiling was more advantageous for the analysis of these samples compared to mRNA profiling, the results further support the use of RNA profiling strategies for use in forensic casework. Further studies using artificially degraded RNA samples may provide the necessary additional clarification regarding which approach is more suitable for these types of samples. Our developed logistic regression modeling provide an exciting new opportunity to perform objective and definitive analysis of normalized expression data which will have applications far beyond just the miRNA profiling assays. Such an approach can be utilized for any study involving the analysis of gene expression data. We also began to evaluate the use of other classes of small RNA, such as piRNAs, for the identification of body fluids (semen in

particular with piRNA analysis). While we did not observe semen-specific expression of the small number of piRNAs tested, a more comprehensive large-scale screening of known piRNAs may permit an identification of suitable candidates for the identification of semen.

V. REFERENCES

[1] J. Ballantyne Serology Overview. Encyclopedia of Forensic Science, Academic Press, London, 2000.

[2] 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.

[3] M. Bauer, D. Patzelt, Evaluation of mRNA markers for the identification of menstrual blood, J Forensic Sci. 47 (2002) 1278-82.

[4] M. Bauer, D. Patzelt, Protamine mRNA as molecular marker for spermatozoa in semen stains, Int J Legal Med. 117 (2003) 175-9.

[5] J. Juusola, J. Ballantyne, Multiplex mRNA profiling for the identification of body fluids, Forensic Sci Int. 152 (2005) 1-12.

[6] J. Juusola, J. Ballantyne, mRNA profiling for body fluid identification by multiplex quantitative RT-PCR, J Forensic Sci. 52 (2007) 1252-1262.

[7] C. Nussbaumer, E. Gharehbaghi-Schnell, I. Korschineck, Messenger RNA profiling: a novel method for body fluid identification by real-time PCR, Forensic Sci Int. 157 (2006) 181-186.

[8] K. Virkler, I.K. Lednev, Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids, Forensic Sci Int.2008).

[9] E.M. Ali, H.G. Edwards, M.D. Hargreaves, I.J. Scowen, In-situ detection of drugs-of-abuse on clothing using confocal Raman microscopy, Anal.Chim.Acta. 615 (2008) 63-72.

[10] E.M. Ali, H.G. Edwards, M.D. Hargreaves, I.J. Scowen, Raman spectroscopic investigation of cocaine hydrochloride on human nail in a forensic context, Anal.Bioanal.Chem. 390 (2008) 1159-1166.

[11] V.C. Bass, The identification of sympathomimetic amines by Raman spectroscopy, Forensic Sci. 11 (1978) 57-65.

[12] S.E. Bell, L.A. Fido, S.J. Speers, W.J. Armstrong, S. Spratt, Forensic analysis of architectural finishes using fourier transform infrared and Raman spectroscopy, part II: white paint, Appl.Spectrosc. 59 (2005) 1340-1346.

[13] S.E. Bell, L.A. Fido, S.J. Speers, W.J. Armstrong, S. Spratt, Forensic analysis of architectural finishes using fourier transform infrared and Raman spectroscopy, part I: the resin bases, Appl.Spectrosc. 59 (2005) 1333-1339.

[14] M. Claybourn, M. Ansell, Using Raman spectroscopy to solve crime: inks, questioned documents and fraud, Sci Justice. 40 (2000) 261-271.

[15] J.S. Day, H.G. Edwards, S.A. Dobrowski, A.M. Voice, The detection of drugs of abuse in fingerprints using Raman spectroscopy I: latent fingerprints, Spectrochim.Acta A Mol Biomol.Spectrosc. 60 (2004) 563-568.

[16] J.S. Day, H.G. Edwards, S.A. Dobrowski, A.M. Voice, The detection of drugs of abuse in fingerprints using Raman spectroscopy II: cyanoacrylate-fumed fingerprints, Spectrochim.Acta A Mol Biomol.Spectrosc. 60 (2004) 1725-1730.

[17] L. Lepot, W.K. De, F. Gason, B. Gilbert, Application of Raman spectroscopy to forensic fibre cases, Sci Justice. 48 (2008) 109-117.

[18] N.A. Macleod, P. Matousek, Emerging non-invasive Raman methods in process control and forensic applications, Pharm.Res. 25 (2008) 2205-2215.

[19] W.D. Mazzella, A. Khanmy-Vital, A study to investigate the evidential value of blue gel pen inks, J Forensic Sci. 48 (2003) 419-424.
[20] W.D. Mazzella, P. Buzzini, Raman spectroscopy of blue gel pen inks, Forensic Sci Int. 152 (2005) 241-247.
[21] 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.
[22] P.M. Schneider, K. Bender, W.R. Mayr, W. Parson, B. Hoste, R. Decorte, J. Cordonnier, D. Vanek, N. Morling, M. Karjalainen, C.C. Marie-Paule, M. Sabatier, C. Hohoff, H. Schmitter, W. Pflug, R. Wenzel, D. Patzelt, R. Lessig, P. Dobrowolski, G. O'Donnell, L. Garafano, M. Dobosz, K.P. de, B. Mevag, R. Pawlowski, L. Gusmao, V.M. Conceicao, A.A. Alonso, F.O. Garcia, N.P. Sanz, A. Kihlgreen, W. Bar, V. Meier, A. Teyssier, R. Coquoz, C. Brandt, U. Germann, P. Gill, J. Hallett, M. Greenhalgh, STR analysis of artificially degraded DNA-results of a collaborative European exercise, Forensic Sci.Int. 139 (2004) 123-134.
[23] H. Asamura, S. Fujimori, M. Ota, H. Fukushima, MiniSTR multiplex systems based on non-CODIS loci for analysis of degraded DNA samples, Forensic Sci Int. 173 (2007) 7-15.
[24] H. Asamura, H. Sakai, M. Ota, H. Fukushima, MiniY-STR quadruplex systems with short amplicon lengths for analysis of degraded DNA samples, Forensic Science International: Genetics. 1 (2007) 56-61.
[25] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, J Forensic Sci. 48 (2003) 1054-1064.
[26] M.D. Coble, J.M. Butler, Characterization of new miniSTR loci to aid analysis of degraded DNA, J Forensic Sci. 50 (2005) 43-53.
[27] L.A. Dixon, A.E. Dobbins, H.K. Pulker, J.M. Butler, P.M. Vallone, M.D. Coble, W. Parson, B. Berger, P. Grubwieser, H.S. Mogensen, N. Morling, K. Nielsen, J.J. Sanchez, E. Petkovski, A. Carracedo, P. Sanchez-Diz, E. Ramos-Luis, M. Brion, J.A. Irwin, R.S. Just, O. Loreille, T.J. Parsons, Syndercombe-Court, H. Schmitter, B. Stradmann-Bellinghausen, K. Bender, P. Gill, Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise, Forensic Sci.Int. 164 (2006) 33-44.
[28] C. Eichmann, W. Parson, 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples, Int J Legal Med.2008).
[29] C.R. Hill, M.C. Kline, M.D. Coble, J.M. Butler, Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, J Forensic Sci. 53 (2008) 73-80.
[30] K.L. Opel, D.T. Chung, J. Drabek, J.M. Butler, B.R. McCord, Developmental validation of reduced-size STR Miniplex primer sets, J Forensic Sci. 52 (2007) 1263-1271.
[31] M.J. Park, H.Y. Lee, U. Chung, S.C. Kang, K.J. Shin, Y-STR analysis of degraded DNA using reduced-size amplicons, Int J Legal Med. 121 (2007) 152-157.
[32] K. Tsukada, K. Takayanagi, H. Asamura, M. Ota, K. Kobayashi, H. Fukushima, Simultaneous PCR of Eight Loci for Very Short Y-STR Fragment Size, Progress in Forensic Genetics. 10 (2004) 322-324.
[33] I.A. Babar, F.J. Slack, J.B. Weidhaas, miRNA modulation of the cellular stress response, Future.Oncol. 4 (2008) 289-298.

[34] E.H. Baehrecke, miRNAs: micro managers of programmed cell death, Curr.Biol. 13 (2003) R473-R475.
[35] T.V. Bagnyukova, I.P. Pogribny, V.F. Chekhun, MicroRNAs in normal and cancer cells: a new class of gene expression regulators, Exp.Oncol. 28 (2006) 263-269.
[36] J.C. Carrington, V. Ambros, Role of microRNAs in plant and animal development, Science. 301 (2003) 336-338.
[37] R.W. Carthew, Gene regulation by microRNAs, Curr.Opin.Genet.Dev. 16 (2006) 203-208.
[38] C.Z. Chen, H.F. Lodish, MicroRNAs as regulators of mammalian hematopoiesis, Semin.Immunol. 17 (2005) 155-165.
[39] H.Y. Cheng, K. Obrietan, Revealing a role of microRNAs in the regulation of the biological clock, Cell Cycle. 6 (2007) 3034-3035.
[40] L.G. Di, G.A. Calin, C.M. Croce, MicroRNAs: fundamental facts and involvement in human diseases, Birth Defects Res.C.Embryo.Today. 78 (2006) 180-189.
[41] M. Jovanovic, M.O. Hengartner, miRNAs and apoptosis: RNAs to die for, Oncogene. 25 (2006) 6176-6187.
[42] M.E. Klein, S. Impey, R.H. Goodman, Role reversal: the regulation of neuronal gene expression by microRNAs, Curr.Opin.Neurobiol. 15 (2005) 507-513.
[43] W.P. Kloosterman, R.H. Plasterk, The diverse functions of microRNAs in animal development and disease, Dev.Cell. 11 (2006) 441-450.
[44] B. Kusenda, M. Mraz, J. Mayer, S. Pospisilova, MicroRNA biogenesis, functionality and cancer relevance, Biomed.Pap.Med Fac.Univ Palacky.Olomouc.Czech.Repub. 150 (2006) 205-215.
[45] C.T. Lee, T. Risom, W.M. Strauss, MicroRNAs in mammalian development, Birth Defects Res.C.Embryo.Today. 78 (2006) 129-139.
[46] H.F. Lodish, B. Zhou, G. Liu, C.Z. Chen, Micromanagement of the immune system by microRNAs, Nat.Rev.Immunol. 8 (2008) 120-130.
[47] J.T. Mendell, MicroRNAs: critical regulators of development, cellular physiology and malignancy, Cell Cycle. 4 (2005) 1179-1184.
[48] L. O'Driscoll, The emerging world of microRNAs, Anticancer Res. 26 (2006) 4271-4278.
[49] B. Zhang, Q. Wang, X. Pan, MicroRNAs and their regulatory roles in animals and plants, J Cell Physiol. 210 (2007) 279-289.
[50] Y. Zhao, D. Srivastava, A developmental view of microRNA function, Trends Biochem.Sci. 32 (2007) 189-197.
[51] G.A. Calin, A. Cimmino, M. Fabbri, M. Ferracin, S.E. Wojcik, M. Shimizu, C. Taccioli, N. Zanesi, R. Garzon, R.I. Aqeilan, H. Alder, S. Volinia, L. Rassenti, X. Liu, C.G. Liu, T.J. Kipps, M. Negrini, C.M. Croce, MiR-15a and miR-16-1 cluster functions in human leukemia, Proc.Natl.Acad.Sci U.S.A. 105 (2008) 5166-5171.
[52] A. Feber, L. Xi, J.D. Luketich, A. Pennathur, R.J. Landreneau, M. Wu, S.J. Swanson, T.E. Godfrey, V.R. Litle, MicroRNA expression profiles of esophageal cancer, J Thorac.Cardiovasc.Surg. 135 (2008) 255-260.
[53] F. Gottardo, C.G. Liu, M. Ferracin, G.A. Calin, M. Fassan, P. Bassi, C. Sevignani, D. Byrne, M. Negrini, F. Pagano, L.G. Gomella, C.M. Croce, R. Baffa, Micro-RNA profiling in kidney and bladder cancers, Urol.Oncol. 25 (2007) 387-392.

[54] P.A. Gregory, A.G. Bert, E.L. Paterson, S.C. Barry, A. Tsykin, G. Farshid, M.A. Vadas, Y. Khew-Goodall, G.J. Goodall, The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1, Nat.Cell Biol.2008).
[55] M.V. Iorio, R. Visone, L.G. Di, V. Donati, F. Petrocca, P. Casalini, C. Taccioli, S. Volinia, C.G. Liu, H. Alder, G.A. Calin, S. Menard, C.M. Croce, MicroRNA signatures in human ovarian cancer, Cancer Res. 67 (2007) 8699-8707.
[56] Y. Ladeiro, G. Couchy, C. Balabaud, P. Bioulac-Sage, L. Pelletier, S. Rebouissou, J. Zucman-Rossi, MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations, Hepatology.2008).
[57] W.J. Lukiw, Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus, Neuroreport. 18 (2007) 297-300.
[58] L. Ma, J. Teruya-Feldstein, R.A. Weinberg, Tumour invasion and metastasis initiated by microRNA-10b in breast cancer, Nature. 449 (2007) 682-688.
[59] M. Negrini, G.A. Calin, Breast cancer metastasis: a microRNA story, Breast Cancer Res. 10 (2008) 203.
[60] N. Rosenfeld, R. Aharonov, E. Meiri, S. Rosenwald, Y. Spector, M. Zepeniuk, H. Benjamin, N. Shabes, S. Tabak, A. Levy, D. Lebanony, Y. Goren, E. Silberschein, N. Targan, A. Ben-Ari, S. Gilad, N. Sion-Vardy, A. Tobar, M. Feinmesser, O. Kharenko, O. Nativ, D. Nass, M. Perelman, A. Yosepovich, B. Shalmon, S. Polak-Charcon, E. Fridman, A. Avniel, I. Bentwich, Z. Bentwich, D. Cohen, A. Chajut, I. Barshack, MicroRNAs accurately identify cancer tissue origin, Nat.Biotechnol. 26 (2008) 462-469.
[61] S. Subramanian, W.O. Lui, C.H. Lee, I. Espinosa, T.O. Nielsen, M.C. Heinrich, C.L. Corless, A.Z. Fire, R.M. van de, MicroRNA expression signature of human sarcomas, Oncogene. 27 (2008) 2015-2026.
[62] N. Tran, T. McLean, X. Zhang, C.J. Zhao, J.M. Thomson, C. O'Brien, B. Rose, MicroRNA expression profiles in head and neck cancer cell lines, Biochem.Biophys Res.Commun. 358 (2007) 12-17.
[63] N. Baroukh, M.A. Ravier, M.K. Loder, E.V. Hill, A. Bounacer, R. Scharfmann, G.A. Rutter, O.E. Van, MicroRNA-124a regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines, J Biol Chem. 282 (2007) 19575-19588.
[64] I. Beuvink, F.A. Kolb, W. Budach, A. Garnier, J. Lange, F. Natt, U. Dengler, J. Hall, W. Filipowicz, J. Weiler, A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs, Nucleic Acids Res. 35 (2007) e52.
[65] H. Bruchova, D. Yoon, A.M. Agarwal, J. Mendell, J.T. Prchal, Regulated expression of microRNAs in normal and polycythemia vera erythropoiesis, Exp.Hematol. 35 (2007) 1657-1667.
[66] C. Conaco, S. Otto, J.J. Han, G. Mandel, Reciprocal actions of REST and a microRNA promote neuronal identity, Proc.Natl.Acad.Sci U.S.A. 103 (2006) 2422-2427.
[67] R. Garzon, F. Pichiorri, T. Palumbo, R. Iuliano, A. Cimmino, R. Aqeilan, S. Volinia, D. Bhatt, H. Alder, G. Marcucci, G.A. Calin, C.G. Liu, C.D. Bloomfield, M. Andreeff, C.M. Croce, MicroRNA fingerprints during human megakaryocytopoiesis, Proc.Natl.Acad.Sci U.S.A. 103 (2006) 5078-5083.
[68] A.M. Krichevsky, K.C. Sonntag, O. Isacson, K.S. Kosik, Specific microRNAs modulate embryonic stem cell-derived neurogenesis, Stem Cells. 24 (2006) 857-864.
[69] P. Landgraf, M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A.O. Kamphorst, M. Landthaler, C. Lin, N.D. Socci, L. Hermida, V. Fulci, S. Chiaretti,

R. Foa, J. Schliwka, U. Fuchs, A. Novosel, R.U. Muller, B. Schermer, U. Bissels, J. Inman, Q. Phan, M. Chien, D.B. Weir, R. Choksi, V.G. De, D. Frezzetti, H.I. Trompeter, V. Hornung, G. Teng, G. Hartmann, M. Palkovits, L.R. Di, P. Wernet, G. Macino, C.E. Rogler, J.W. Nagle, J. Ju, F.N. Papavasiliou, T. Benzing, P. Lichter, W. Tam, M.J. Brownstein, A. Bosio, A. Borkhardt, J.J. Russo, C. Sander, M. Zavolan, T. Tuschl, A mammalian microRNA expression atlas based on small RNA library sequencing, Cell. 129 (2007) 1401-1414.

[70] Y. Liang, D. Ridzon, L. Wong, C. Chen, Characterization of microRNA expression profiles in normal human tissues, BMC.Genomics. 8 (2007) 166.

[71] S. Masaki, R. Ohtsuka, Y. Abe, K. Muta, T. Umemura, Expression patterns of microRNAs 155 and 451 during normal human erythropoiesis, Biochem.Biophys Res.Commun. 364 (2007) 509-514.

[72] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, Identification of novel genes coding for small expressed RNAs, Science. 294 (2001) 853-858.

[73] M. Lagos-Quintana, R. Rauhut, J. Meyer, A. Borkhardt, T. Tuschl, New microRNAs from mouse and human, RNA. 9 (2003) 175-179.

[74] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, Anal.Biochem. 162 (1987) 156-159.

[75] 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.

[76] 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.

[77] 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.

[78] 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.

[79] 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.

[80] E. Hanson, H. Lubenow, J. Ballantyne, Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs, Forensic Sci Int Genet.Suppl.Ser. 2 (2009) 503-504.

[81] E.K. Hanson, H. Lubenow, J. Ballantyne, Identification of Forensically Relevant Body Fluids Using a Panel of Differentially Expressed microRNAs, Anal.Biochem. 387 (2009) 303-314.

[82] C. Klattenhoff, W. Theurkauf, Biogenesis and germline functions of piRNAs, Development. 135 (2008) 3-9.

[83] Y.H. Guo, L. Liu, R. Cai, C. Qian, [piRNA: a novel member of small RNA family], Yi.Chuan. 30 (2008) 28-34.
[84] L.S. Sai, S. Agrawal, piRNABank: a web resource on classified and clustered Piwi-interacting RNAs, Nucleic Acids Res. 36 (2008) D173-D177.
[85] F. Tang, K. Hayashi, M. Kaneda, K. Lao, M.A. Surani, A sensitive multiplex assay for piRNA expression, Biochem.Biophys Res.Commun. 369 (2008) 1190-1194.
[86] J.C. Peng, H. Lin, Beyond transposons: the epigenetic and somatic functions of the Piwi-piRNA mechanism, Curr.Opin.Cell Biol. 25 (2013) 190-194.
[87] J. Cheng, J.M. Guo, B.X. Xiao, Y. Miao, Z. Jiang, H. Zhou, Q.N. Li, piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells, Clin.Chim.Acta. 412 (2011) 1621-1625.
[88] J. Cheng, H. Deng, B. Xiao, H. Zhou, F. Zhou, Z. Shen, J. Guo, piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells, Cancer Lett. 315 (2012) 12-17.
[89] L. Cui, Y. Lou, X. Zhang, H. Zhou, H. Deng, H. Song, X. Yu, B. Xiao, W. Wang, J. Guo, Detection of circulating tumor cells in peripheral blood from patients with gastric cancer using piRNAs as markers, Clin.Biochem. 44 (2011) 1050-1057.
[90] T. Esposito, S. Magliocca, D. Formicola, F. Gianfrancesco, piR_015520 belongs to Piwi-associated RNAs regulates expression of the human melatonin receptor 1A gene, PLoS.ONE. 6 (2011) e22727.
[91] Y. Kirino, Z. Mourelatos, Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini, Nat.Struct.Mol Biol. 14 (2007) 347-348.
[92] Y. Kirino, Z. Mourelatos, 2'-O-methyl modification in mouse piRNAs and its methylase, Nucleic Acids Symp.Ser.(Oxf).2007) 417-418.
[93] T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated, Nat.Struct.Mol Biol. 14 (2007) 349-350.
[94] C. Courts, B. Madea, Specific micro-RNA signatures for the detection of saliva and blood in forensic body-fluid identification, J Forensic Sci. 56 (2011) 1464-1470.
[95] D. van der Meer, M.L. Uchimoto, G. Williams, Simultaneous Analysis of Micro-RNA and DNA for Determining the Body Fluid Origin of DNA Profiles, J Forensic Sci. 58 (2013) 967-971.
[96] Z. Wang, H. Luo, X. Pan, M. Liao, Y. Hou, A model for data analysis of microRNA expression in forensic body fluid identification, Forensic Sci Int Genet. 6 (2012) 419-423.
[97] Z. Wang, J. Zhang, H. Luo, Y. Ye, J. Yan, Y. Hou, Screening and confirmation of microRNA markers for forensic body fluid identification, Forensic Sci Int Genet. 7 (2013) 116-123.
[98] D. Zubakov, M. Kokshoorn, A. Kloosterman, M. Kayser, New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains, Int J Legal Med. 123 (2009) 71-74.

VI. DISSEMINATION OF RESEARCH FINDINGS

A. Publications

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. mRNA and microRNA for Body Fluid Identification. C. Haas, E. Hanson, and J. Ballantyne. Encyclopedia of Forensic Sciences, Second Edition. Vol. 1, pp. 402-408 (2013). Elsevier. doi: 10.1016/B978-0-12-382165-2.00069-6.
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. (2013).
4. Hanson, E.K., Ballantyne, J. "Chapter 18: Circulating MicroRNA for the Identification of Forensically Relevant Body Fluids." In Circulating MicroRNAs: Methods and Protocols, Methods in Molecular Biology, vol 1024, Springer Science + Business Media, New York (Humana Press). Vol 1024 (2013). Doi: 10.1007/978-1-62703-453-1_18.
5. Hanson, E.K., Rekab, K. and Ballantyne, J. Binary logistic regression models enable miRNA profiling to provide accurate identification of forensically relevant body fluids and tissues. For Sci International Genetics Supp Series 4 (2013), p. e127-e128.

Manuscripts in preparation

6. *Subject:* The development and testing of logistic regression modeling in miRNA profiling assays: commonly encountered body fluids (blood, semen, saliva and vaginal secretions).
7. *Subject:* The use of logistic regression modeling in miRNA profiling assays for "special case" body fluids and tissues (menstrual blood and skin)
8. *Subject:* Validation of the miRNA profiling assays including a comparison of mRNA/miRNA profiling.

Related publications (related to Aim 3 – mRNA profiling)

1. Hanson, E., Haas, C., Jucker, J., and Ballantyne, J. "Specific and Sensitive mRNA Biomarkers for the Identification of Skin in 'Touch DNA' Evidence. Forensic Sci Int Genetics (2012) 6(5): p. 548-558.
2. Hanson, E. and Ballantyne, J. "Highly Specific mRNA Biomarkers for the Identification of Vaginal Secretions in Sexual Assault Investigations". Science & Justice: 53 (2013) p. 14-22. Doi: 10.1016.j.scijus.2012.03.007
3. "Identification of Skin in Touch/Contact Forensic Samples by Messenger RNA Profiling." Erin Hanson, PhD, Cordula Haas, PhD, Ronaldo Jucker and Jack Ballantyne, PhD. Forensic Science International: Genetics Supplement Series 3 (2011), e305-306.

B. Presentations

1. Body Fluid Identification by RNA Expression Profiling. Ballantyne, J. BrightTALK Forensic Science Community webcast, DNA Forensic Summit. 2009
2. Identification of Forensically Relevant Body Fluids Using a Panel of Differentially Expressed microRNAs. Hanson, E. and Ballantyne, J. 20th International Symposium on Human Identification (Promega), Las Vega, NV. 2009
3. MicroRNA Expression Profiling for the Identification of Forensically relevant Biological Fluids. Ballantyne, J. Cambridge Health Institute's 17th International Molecular Medicine Tri-Conference, San Francisco, CA. 2010.
4. Not your CSI Profiling: RNA Applications in Forensic Genetics. Ballantyne, J. 12th annual Ancient DNA Training Program, Lakehead University Paleo-DNA Laboratory, Thunder Bay, Ontario, Canada. 2010.
5. Identification of Forensically Relevant Body Fluids by Small RNA Profiling. Hanson, E. and Jack Ballantyne. NIJ Conference, Crystal City, VA. 2010.
6. Body Fluid Identification by RNA Profiling. Webinar, California Criminalistics Institute, Jan Bashinski DNA Laboratory, Richmond, CA. 2010.
7. Identification of Highly Specific RNA Biomarkers for the Identification of Vaginal Secretions in Forensic Caseowrk. Hanson, Ballatyne, J. 22nd Annual Symposium on Human Identification (Promega), National Harbor, MD. 2011.
8. Identification of Forensically Relevant Body Fluids and Tissues by RNA Profiling. Presented at the 'Upstream Screenings for Downstream Savings at the Front End of a Forensic Investigation' Session. Ballantyne J and Hanson EK. NIJ Conference, Crystal City, VA. 2012.
9. Identification of Skin in Touch/Contact Forensic Samples by Messenger RNA Profiling. Ballantyne, J. and Hanson, E. Association of Forensic DNA Administrators and Analysts (AFDAA), San Antonio, TX. 2012.
10. Beyond DNA: Other Forensic Biology Applications to Legal Medicine. Visionen und Zukundft der Forensik, University of Zurich, Institute for Legal Medicine, Zurich, Switzerland. 2012.
11. "Not your CSI DNA Profiling: RNA Applications in Forensic Genetics." Cold Spring Harbor Genomics Workshop. August 2013.
12. The Use of Logistic Regression Modeling in microRNA Profiling Assays to Provide a Definitive Identification of Forensically Relevant Body Fluids and Tissues. International Society of Forensic. Melbourne, Australia. September 2013.