Deep Sequencing for Identification and Characterization of MicroRNAs in Forensically Important Biological Fluids

Award 2012-DN-BX-K017

Final Technical Report

Submission Date: 05/31/2016

National Institute of Justice

FY 2012: Basic Scientific Research to Support Forensic Science for Criminal Justice Purposes

Principal Investigator: Zendra E. Zehner, PhD
Professor, Department of Biochemistry and Molecular Biology
Virginia Commonwealth University
zezehner@vcu.edu

Co-Investigator: Sarah Seashols-Williams, PhD
Assistant Professor, Department of Forensic Science
sseashols@vcu.edu

Susan E. Robb
Assoc VP for Research Administration and Compliance
Office of Research
Virginia Commonwealth University
(804) 828-6772
dirospa@vcu.edu

Virginia Commonwealth University
P.O Box 980568
800 East Leigh Street, Suite 113
Richmond, VA 23298-0568

Project/Grant Period: 1/1/2013 – 1/31/2016

Abstract

MicroRNAs (miRNAs) are small non-coding RNAs that are 18-22 nucleotides in length and have previously been identified as potential markers for the identification of forensically relevant body fluids. *In-vivo*, miRNAs suppress protein expression through binding to messenger RNA (mRNA) in the cytosol, and as such, can be tissue-specific. There are minimal postprocessing modifications, and thus miRNAs are simpler, and potentially less problematic for detection than proteins and mRNAs. There is significant interest in the use of miRNAs for forensic casework because their short length and high resistance to degradation, potentially allowing for robust detection in highly degraded samples.

The purpose of this work was to utilize high-throughput sequencing (HTS) of eight forensically relevant biological fluids to identify candidate miRNAs that could be as a molecular tool for body fluid identification, and subsequently perform developmental validations for those identified miRNAs. Samples of feces, urine, peripheral blood, menstrual blood, vaginal secretions, semen, saliva, and perspiration were collected from 20-50 donors following approved human subjects research protocol, and small RNA isolation methods identified and optimized for each biological fluid. HTS using Illumina Hi-Seq® was performed on 4-5 individual donations of each body fluid, and the data analysis identified several candidate miRNAs for each body fluid with potential body fluid specificity.

miRNAs let-7g and let-7i were identified and validated for use as endogenous reference controls, as they were shown to be expressed with consistent levels in the majority of body fluids, and consistently among tested donors. A standard curve using a synthetic miRNA of known quantity was developed and applied to gain a more accurate quantitation of miRNA expression and limit of detection. Reverse-transcription quantitative PCR (RT-qPCR) evaluations revealed that while no evaluated miRNA was absolutely body-fluid specific in that it was only expressed in that tissue, a panel of 6 miRNAs were identified as providing significantly different relative expression levels alongside 2 miRNAs used for internal control and normalization purposes. This panel is able to identify blood, semen, urine, saliva, feces, and menstrual secretions, and provide some information regarding vaginal secretions and perspiration. Each of the candidate miRNAs was evaluated using classic developmental validation methods including species specificity, limit of detection, abundance within the population, and abundance within an individual over a biological time period or cycle, depending on the secretion under assessment.

miRNA stability was assessed in blood, urine, semen and saliva, and several miRNAs were analyzed were found to behave in a consistent manner. Treatment resistance was found to be dependent on the body fluid under evaluation, but for the majority of treated samples, miRNA expression was detectable and comparable in signal to untreated control expression. The comprehensive evaluation of miRNA expression in forensically relevant biological fluids, and consequent development of candidate miRNAs for further research that we completed in this body of work was a vital first step in proceeding towards an eventual commercial assay for body fluid ID that is robust and reliable in the hands of practitioners. Given additional developmental research, this panel could rapidly revolutionize forensic body fluid identification, resulting in quantifiable confidence in the body fluid or fluids present in the sample.

TABLE OF CONTENTS

Executive Summary

While forensic DNA analysis has reached a level of maturity in the Forensic Science field with regards to the sophistication of the techniques and confidence in the results, the equally important question of body fluid identification has lagged behind, and could still be considered to be in a primitive state. Current crime scene and in-laboratory methods utilize detection methods that exploit the properties of each biological fluid (e.g. Phenolphthalin or TMB testing for blood, amylase detection for saliva, and urease tests for urine), but validated identifying techniques are largely limited to microscopic methods (i.e. identification of spermatozoa) or immunological methods, as seen in the widely used immunochromatographic commercial tests for blood, semen, and other biological fluids. Thus, while there is widespread confidence in the DNA profile generated, there is often significantly less assurance in the identity of the body fluid that the DNA profile was developed from. It is common during trials for attorneys to categorically accept the STR analysis, but probe the forensic scientist on the source of the DNA that generated the profile. Because of this dichotomy, significant efforts have been made over the past ten years in order to develop forensic serological techniques of a more discriminatory nature.

Of late, there has been some work in the forensic science field in regards to exploring microRNAs (miRNAs) for a molecular-based, forensic body fluid identification method. miRNAs are small non-coding RNAs that are 18-22 nucleotides in length and have previously been identified as potential markers for the identification of forensically relevant body fluids. *In-vivo*, miRNAs suppress protein expression through binding to messenger RNA (mRNA) in the cytosol, and as such, can be tissue-specific. There are minimal postprocessing modifications, and thus miRNAs are simpler, and potentially less problematic for detection than proteins and mRNAs. There is significant interest in the use of miRNAs for forensic casework because their short length and high resistance to degradation, potentially allowing for robust detection in highly degraded samples.

The body of forensic literature characterizing candidate miRNAs for body fluid identification has grown dramatically over the last 5 years. However, exploration of forensic tissue specificity until very recently has evaluated miRNAs using microarray and RT-qPCR analysis of panels to identify tissue-specific candidates. Differences in platforms, detection

chemistries, and normalization to different endogenous reference miRNAs have resulted in inconsistencies between forensic studies. Additionally, feces, urine, and perspiration have not been evaluated for characteristic miRNAs.

The purpose of this work was to utilize high-throughput sequencing (HTS) of eight forensically relevant biological fluids to identify candidate miRNAs that could be as a molecular tool for body fluid identification, and subsequently perform initial characterization for those identified miRNAs (Figure 1).

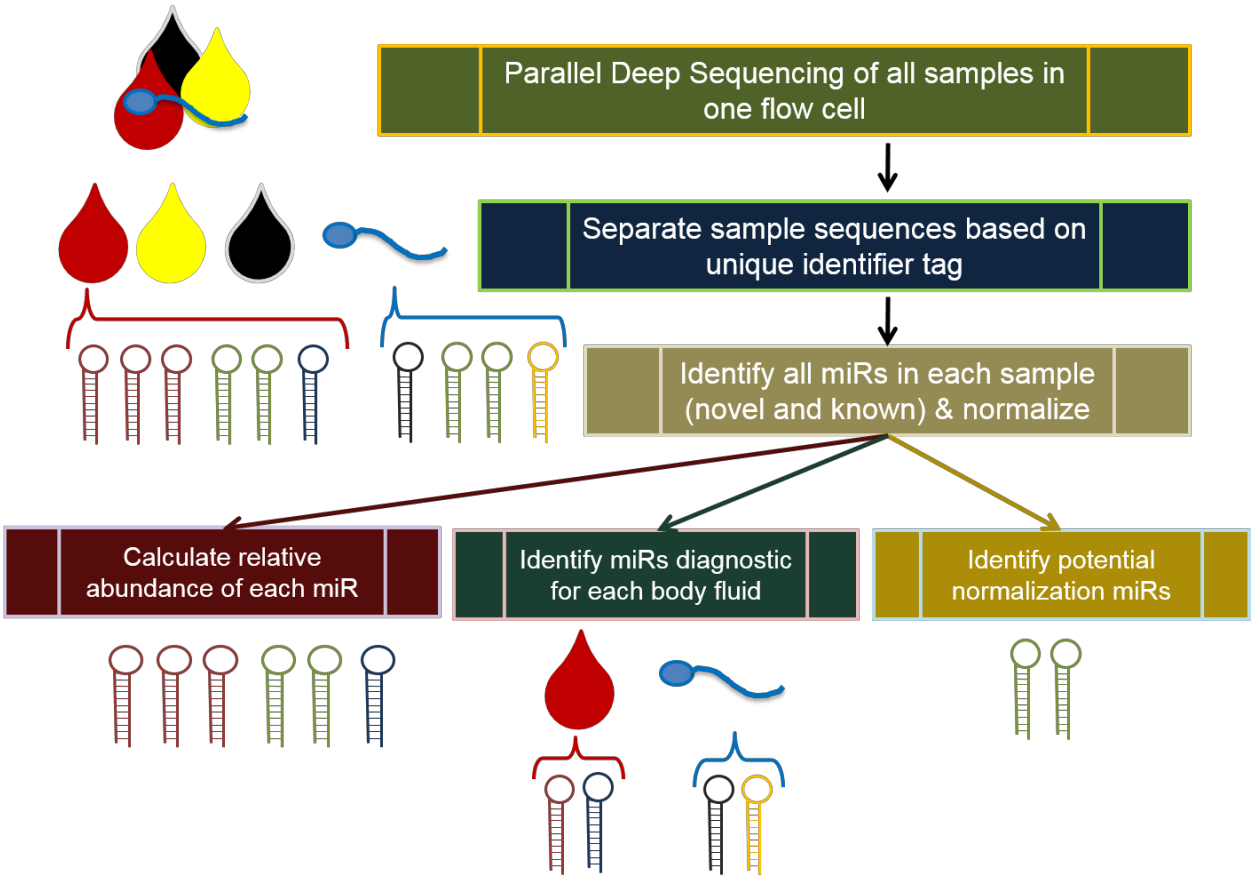


Figure 1: Workflow for miRnome sequencing data analysis using high-throughput sequencing. "Barcoding" of cDNA samples prior sequencing allowed for combining all samples onto a single Illumina® HiSeq flow cell. Data analysis was conducted using Partek® Flow®, and sequence reads were separated using known barcode sequences attached during library preparation, and sequences aligned to the known miRnome and sequences annotated. Abundance of miRNAs identified was calculated based on the number of reads/total annotated reads. Sequencing data was compared between and among biological sources for identification of diagnostic and normalization/endogenous reference miRNAs.

Samples of feces, urine, peripheral blood, menstrual blood, vaginal secretions, semen, saliva, and perspiration were collected from 20-50 donors following approved human subjects research protocol, and small RNA isolation methods identified and optimized for each biological fluid for maximum amplifiable miRNA quantity, finding that all tested methods were suitable for all biological fluids, with the exception of feces, which requires a specialized RNA isolation method. The MoBio Power Microbiome Fecal RNA kit (MoBio Laboratories) was optimized for this project. HTS using Illumina Hi-Seq® was performed on 4-5 individual donations of each body fluid. The resultant sequence data is a compelling finding, as we specifically chose to extract RNA and take forward samples consistent with the small volume found in evidence samples, with the exception of urine. High-throughput sequencing for miRNA expression characterization has, to our knowledge, never been performed on sample sizes this small, and the success of the analysis warrants publication not only for application to forensic HTS analysis of small volume evidentiary samples in the future.

The number of unique miRNAs identified in each biological fluid was found to correlate with both the body fluid type and number of raw sequencing reads obtained. Coverage of those fluid types with high bacterial loads or low RNA quantities yielded fewer identified miRNAs (Figure 2). Interestingly, comparisons of the miRNAs expressed in a particular biological fluid among the donors showed a high degree of diversity, with only a minority of common miRNAs expressed among the donors. Those miRNAs commonly observed in all biological samples tested were identified as constitutive miRNAs that had the potential to be utilized for normalization purposes, and those miRNAs expressed in all donors of a particular biological fluid were identified for further evaluation as body fluid specific candidate miRNAs.

The high-throughput sequencing was markedly compromised by the bacterial small RNAs in many biological fluids, resulting in competition for reverse transcription and sequencing reagents. This did not dramatically impact the number of sequencing reads collected from the instrument, but rather the percent of sequencing reads that were annotated as human miRNAs and consequently the number of miRNAs that could be evaluated was very limited for some of these fluids, particularly feces and perspiration.

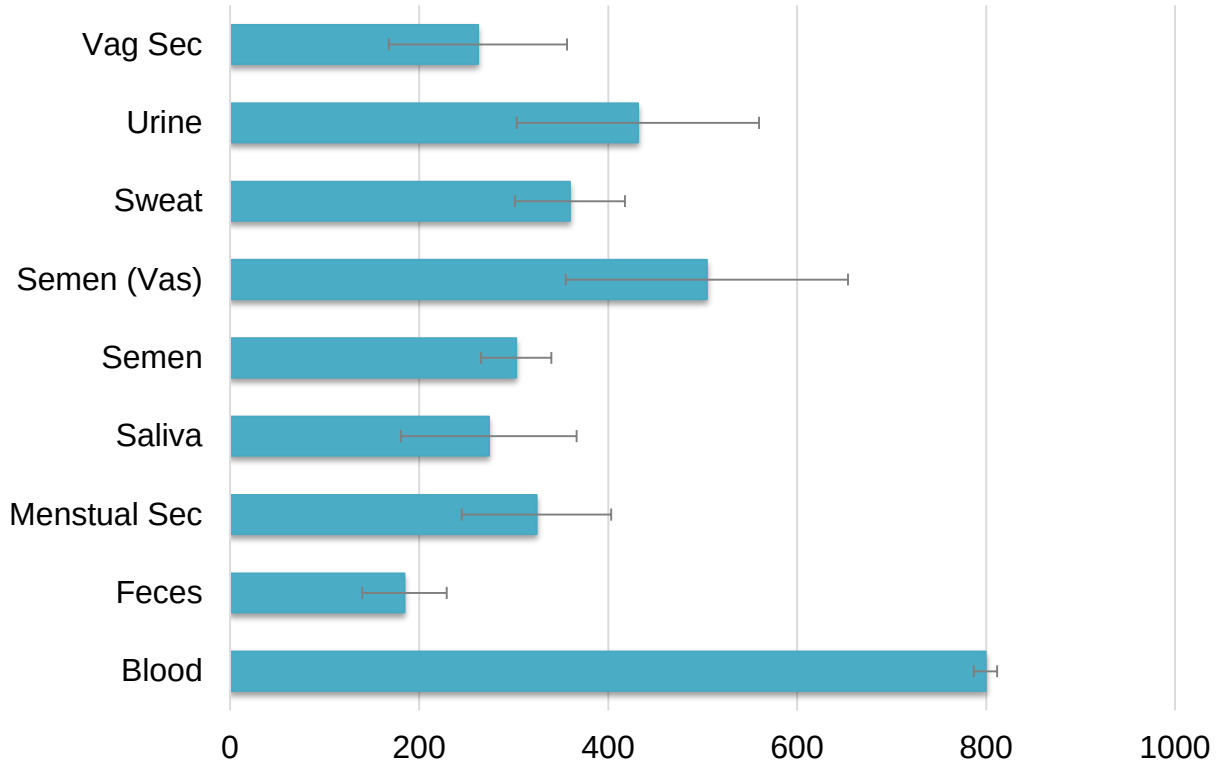


Figure 2: Unique annotated miRNAs identified in each biological fluid by high throughput sequencing. Data is the average number of annotated miRNAs from 3-5 individual donors.

We first combed the sequence data for miRNAs observed in all 33 samples, regardless of biological fluid. We identified four miRNAs – let-7g, let-7i, miR-451a, and miR-21, as expressed in all samples. RT-qPCR evaluation and correlation analysis using Bestkeeper® software eliminated miRs-451 and 21, but lets-7g and i showed high correlation to measured miRNA quantity, resulting in r values above 0.9 for both miRNAs. Confirmation with an additional set of 24 additional samples yielded an even stronger coefficient of correlation of 0.971 and 0.947 for let-7g and i, respectively. This resulted in the development of a normalization method that can be used for all eight biological fluids using the average expression of both lets-7g and i. This method is a remarkable advance, as it allows for elimination of RNA quantity evaluation prior to reverse transcription, thus streamlining analysis.

Once our normalization method was validated, evaluation of the HTS data resulted in a list of potentially body fluid-specific miRNAs that appeared to be either exclusively or differentially expressed in a particular body fluid. The candidate miRNA findings from the HTS

data were not supported by following RT-qPCR analysis, which showed that sensitivity of the HTS platform was markedly lacking and expression of the candidate miRNAs was neither in proportion to that expected, nor reduced in other biological fluids. Regardless of these disappointing findings, we were able to identify 7 miRNAs for initial validation of body fluid specificity. Additional population studies resulted in 6 of those miRNAs successfully showing differential expression ($p<.05$) for one or more biological fluids (Figure 3). The six miRNAs (along with lets-7g and i) were also evaluated in and most found to be expressed in a panel of 20 human organs and tissues, along with 33 different animal body fluid RNA samples, corroborating the well-known conservation of miRNA sequence over evolutionary time.

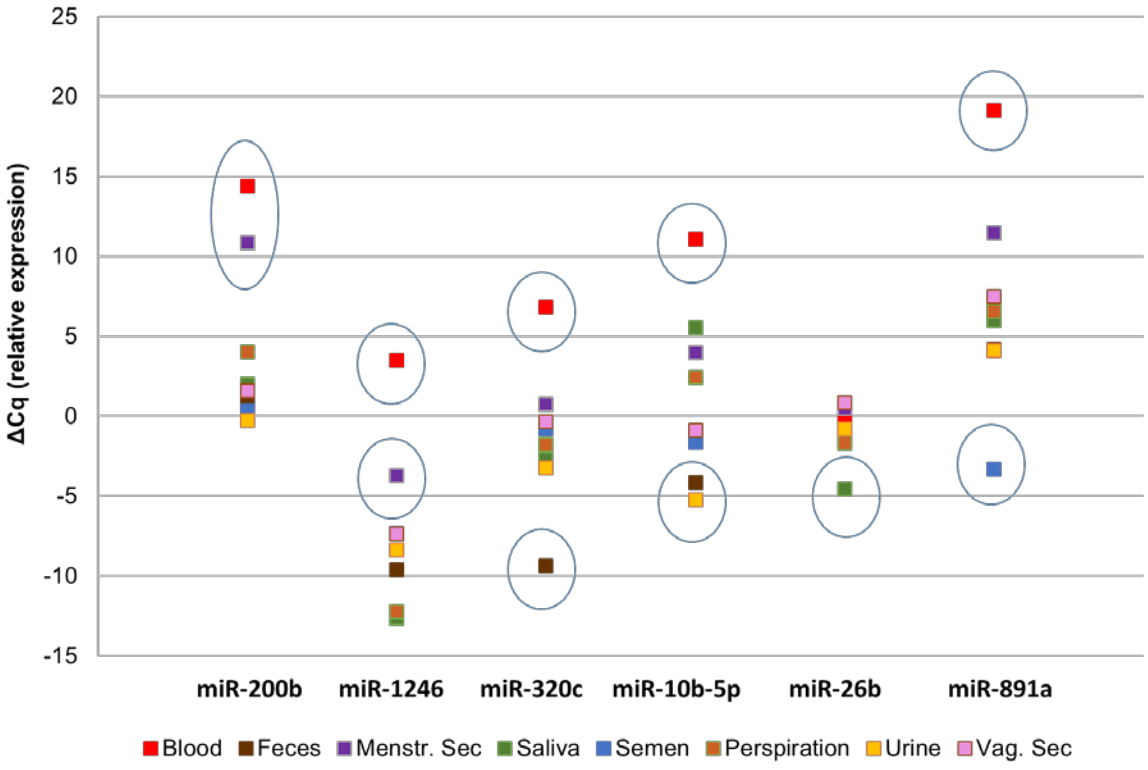


Figure 3: A miRNA panel for forensic body fluid identification using differential expression. Blood, feces, menstrual secretions, saliva, semen, and urine can be differentiated based on expression patterns of these six miRNAs, normalized with let-7g and i as described. Circled body fluids are differentiated in expression (n=20 tested population sample for each circled fluid, n=5-20 for the other 5-6 fluids tested against that miRNA. $p<0.05$).

The decision matrix for a biological fluid using the differential expression is in some places convoluted: in two cases, an additional marker is necessary to differentiate the two

fluids. miR-200b distinguishes blood and menstrual blood from all other body fluids, but not from each other. However, miR-1246 differentiates blood and menstrual secretions, as do the rest of the miRNAs. Likewise, miR-10b-5p distinguishes urine and feces from the rest of the body fluids, and miR-320c differentiates feces from urine. Thus, the identification of urine is reductive. Perspiration *may* be differentiated from the other body fluids using miR-200b, but the sufficient population samples have not been tested to evaluate this possibility. We were not able to identify a candidate miRNA for vaginal secretions, but further analysis of candidate miRNAs could likely develop a potential marker.

We also sought to evaluate miRNA stability in compromised body fluid samples. To achieve this, blood, urine, semen, and saliva were exposed to moderate and high heat conditions for varying time points, ultraviolet exposure, glacial acetic acid and dish detergent treatment, along with exposure to 1:10 and full-strength household bleach. The samples were then isolated for RNA and miRNA expression differences from the untreated control were analyzed. We evaluated several miRNAs, and found them to behave in a consistent manner. Treatment resistance was found to be dependent on the body fluid under evaluation, but for the majority of treated samples, miRNA expression was detectable and comparable in signal to untreated control expression (Table 1).

Table 1: Successful let-7g detection rates in body fluids after compromising treatments.

	UV	10% Bleach	100% Bleach	Dish Detergent	Glacial Acetic Acid
Blood	100%	100%	100%	100%	100%
Urine	77.8%	33.3%	55.6%	66.7%	100%
Semen	100%	100%	100%	66.7%	100%
Saliva	100%	44.4%	0%	77.8%	100%

The miRNAs present in blood were both more abundant and robust than those of the other biological fluids, and were detectable throughout all treatments. Even though degradation in the form of reduced amplification was observed, sufficient miRNA levels remained for positive analysis. Semen was more susceptible to treatment, resulting in

amplification failures when treated with dish detergent, but was otherwise largely detectable in the majority of treatment scenarios. The better detectability of blood and semen are likely reflections of the high quantities of total RNA isolated from those samples, but if that were the only impacting factor, saliva would also have been more successful. As could be expected, semen miRNA levels were not negatively impacted by glacial acetic acid treatment; given that semen contains the necessary buffering capacity and basic pH that has co-evolved to neutralize and survive in the acidic vaginal vault. Fortunately, miRNA degradation did appear to occur consistently across multiple miRNAs, and thus normalization to endogenous reference miRNAs can be confidently utilized for differential expression analyses, as the majority of the work in forensic miRNA expression utilize a normalization methodology. Given that the miRNAs tested were readily and consistently detectable in samples consistent with forensic evidence, it is highly possible that stability is even better than seen in these results.

Implications for policy and practice

The comprehensive evaluation of miRNA expression in forensically relevant biological fluids, and consequent development of candidate miRNAs for further research that we completed in this body of work was a vital first step in proceeding towards an eventual commercial assay for body fluid ID that is robust and reliable in the hands of practitioners. Our findings regarding the stability of miRNAs upon compromise further enhanced the desirability of the development of a body fluid identification platform using miRNAs. Given the known existence of amplifiable miRNAs in DNA extracts, it is entirely possible that body fluid identification using the miRNA panel described in this project could be performed using DNA extracts from forensic samples, which would significantly reduce the resistance for a novel body fluid identification method that requires second or additional RNA isolation procedures. The compilation of miRNAs in a robust multiplex developed from DNA extracts could rapidly revolutionize forensic body fluid identification, resulting in quantifiable confidence in the body fluid or fluids present in the sample. The resultant increased efficiency will reduce or eliminate analyst time on microscopic analysis and serological assays that yield poor confidence in the body fluid reported (excepting microscopic identification of sperm cells). Given that serological testing consumes a large proportion of analyst time, implementation of a multiplexed miRNA-based body fluid identification panel could result in a modest reduction of backlogs, and this

assay could very quickly and easily be integrated into the forensic workflow, utilizing the instrumentation already present.

In conclusion, we have fully addressed the proposed Aims of the funded project, and have developed a panel of 8 miRNAs that can be used to differentiate the six major biological fluids commonly seen in forensic evidence, and provide some differential information regarding vaginal secretions and perspiration.

Final Technical Report

Introduction

While forensic DNA analysis has reached a level of maturity in the Forensic Science field with regards to the sophistication of the techniques and confidence in the results, the equally important question of body fluid identification has lagged behind, and could still be considered to be in a primitive state. Current crime scene and in-laboratory methods utilize detection methods that exploit the properties of each biological fluid (e.g. Phenolphthalin or TMB testing for blood, amylase detection for saliva, and urease tests for urine), but validated identifying techniques are largely limited to microscopic methods (i.e. identification of spermatozoa) or immunological methods, as seen in the widely used immunochromatographic commercial tests for blood, semen, and other biological fluids. Thus, while there is widespread confidence in the DNA profile generated, there is often significantly less assurance in the identity of the body fluid that the DNA profile was developed from. It is common during trials for attorneys to categorically accept the STR analysis, but probe the forensic scientist on the source of the DNA that generated the profile. Because of this dichotomy, significant efforts have been made over the past ten years in order to develop forensic serological techniques of a more discriminatory nature.

There are three main areas that can be exploited for molecular methods for body fluid identification: the genome, the transcriptome, and the proteome, and each has been subject to a great deal of recent research. The genome is, for the most part eliminated, as DNA is the same from cell to cell, with minor differences in methylation status and mutational states[1-3]. The proteome consists of the proteins that are found in a particular tissue. For forensic body fluid identification, proteins uniquely found in a particular biological fluid could be exploited, but significant complications in the identification of proteins hamper this approach[4-6]. Proteins undergo extensive posttranslational modification and tend to be low in abundance in fluids such as urine[7]. Secondarily, proteins degrade and denature easily; given the compromised nature of forensic evidence, they cannot be considered ideal candidates for an identification matrix. Finally, researchers have evaluated messenger RNA (mRNA), and have found many suitable candidates in body fluid identification. mRNA transcripts unique to body fluids have been described and multiplex identification techniques have been developed[8-16]; however,

mRNA methods have not gained widespread acceptance due to the well-known challenges in working with mRNA. Namely, mRNA is designed to be a temporary working copy of the coding regions of genes, and contains specific components that can vary degradation rates. Factors that affect degradation of mRNA include the poly-A tail and varying sequences in the 5'- and 3'-UTRs. Additionally, mRNAs are commonly alternatively spliced, which can present problems in designing appropriate detection systems.

Of late, there has been some work in the forensic science field in regards to exploring microRNAs (miRNAs) for a molecular-based, forensic body fluid identification method. miRNAs are small structures that are transcribed as larger precursors (60-100 nts) that immediately form a stem-loop structure with incomplete base-pairing and flanking nucleotides[17]. Processing of the immature miR is accomplished through excision of the loop, resulting in a mature miRNA of 19-23 nucleotides long. There are minimal postprocessing modifications, and thus miRNAs are simpler, and potentially less problematic for detection than proteins and mRNAs. MiRNAs show distinct promise for forensic body fluid identification on several grounds. There is significant literature that some miRNAs are differentially expressed, and in fact are involved in embryonic stem cell development and tissue differentiation[7,18,19]. They are found in extracellular fluids[7], and thus the application of unique miRNAs for forensically relevant body fluids is a distinct possibility.

miRNAs regulate cellular processes through interactions with mRNA[7,20]. They can regulate gene expression in three fashions: they can negatively regulate gene expression by: 1) incomplete binding, usually to the 3'-UTR of mRNA, causing translational suppression or 2) cleavage of a targeted mRNA[21]. Less frequently, miRNAs act to enhance translation by binding to the 5'-terminal oligopyrimidine tract (TOP) and releasing a cis-element in the 5'-UTR[22]. The complete functional role of miRNAs has yet to be fully determined, though they are highly conserved among organisms, indicating their importance in regulating biological processes[23]. As such, some miRNAs can be consistently expressed in all human tissues, and others can be tissue-specific[7]. Because of the potential for tissue specificity, their small size and consequent inherent stability, miRNAs have been the subject of recent research interest as a potential forensic body fluid identification technique[24-28].

Because of their small size and lack of a poly-A tail, miRNAs are inherently less susceptible to degradation than mRNA. miRNAs have been shown to be remarkably stable in

clinical samples, notably because of their small size, but also due to the fact that when secreted from the cell, they are either encased in a single-stranded protective lipid vesicle or encapsulated in a protein (Argonaute) or cholesterol matrix[29,30]. This results in a longer half-life than mRNAs[31] and makes miRNAs more resistant to degradation than naked nucleic acids, including exposure to nucleases, detergents, and harsh conditions[26].

The body of forensic literature characterizing candidate miRNAs for body fluid identification has grown dramatically over the last 5 years. However, exploration of forensic tissue specificity until very recently has evaluated miRNAs using microarray and RT-qPCR analysis of panels to identify tissue-specific candidates. Differences in platforms, detection chemistries, and normalization to different endogenous reference miRNAs have resulted in inconsistencies between forensic studies (Table 1). Additionally, feces, urine, and perspiration have not been evaluated for characteristic miRNAs.

Table 1: miRNAs evaluated in the forensic literature for body fluid identification purposes

Menstrual secretions	Venous Blood	Semen	Saliva	Vaginal secretions
		miR-943		
	miR-20a	miR-135a/b	miR-583	
miR-185	miR-106a	miR-10a/b	miR-518c	
miR-451	miR-185	miR-507	miR-208b	miR-617
miR-412	miR-451	miR-16	miR-205	miR-124a
miR-16	miR-16	miR-891a	miR-658	miR-16
miR-214	miR-126	miR-943	miR-16	miR-372
miR-486	miR-150	miR-2392	miR-200c	
	miR-14	mIR-3197	miR-203	
		miR-888		
From Zubakov et al, Hanson et al, Courts et al, Park et al, Wang et al[24,27,28,32,33]				

A major limitation to the exploratory studies that derived the above studied miRNAs is that they were conducted using microarrays or RT-qPCR panels containing only the most commonly catalogued miRNAs in the human body, not the entire miRnome. All except one of the exploratory miRNA studies for forensic body fluid identification were based on panels containing only 452-800 human miRNAs[19,24,28,33], when the miRnome is composed of 2,588

distinct mature miRNAs[34-37]. Thus, a significant number of miRNAs were left out of the original analyses. Wang et al evaluated a 1733-miRNA array in 2012[27], and has very recently described high-throughput small RNA sequencing of blood and saliva[38].

To further complicate the research, results between the studies did not establish concordance: with the exception of 2 markers for semen, the groups found differential expressions of the majority of miRNAs evaluated. This can be attributed to a variety of factors, including choice of normalization control and probe and assay design. Furthermore, replication of the Hanson et al. method was attempted by the Zubakov group, but was for the most part unsuccessful. Those differences were explored further in a manuscript by Wang et al, but the study only evaluated three miRNAs in several body fluids, and found differing levels of a commonly used reference RNA as well, indicating that perhaps a different reference RNA should be evaluated[27]. Additionally, a recent study in Clinical Chemistry using microarrays found completely different miRNAs in the commonly tested body fluids, with no overlap to the Zubakov or Hanson studies[7]. Regardless, a panel that can discriminate between blood, semen, saliva, vaginal secretions, and menstrual blood has been described by Hanson et al[24,39]. In regards to sensitivity, miRNAs far exceed the detection limits of most mRNA markers; Zubakov et al found that miRNAs could be detected using qPCR from just picograms of total RNA, below the generally detectable limit of mRNAs[28]. Recent reports have also shown that miRNAs are detectable and coextracted in silica-column based DNA extracts at a similar level to RNA extracts[40,41]. Because of the preliminary results reported from these authors, it is clear that miRNAs are detectable in forensic samples, and that it is an area that should be evaluated thoroughly.

The clinical biomedical research fields have shown some interest in the biological fluids in and of themselves, but studies in this area have been primarily focused on biomarkers for medical conditions such as cancer[42-45]. There have been NO studies searching for novel miRNA expression in vaginal secretions or perspiration, and the data on urine and saliva is minimal, and has not attempted to identify novel miRNAs in those fluids[7]. miRNAs have been shown in feces as well in some studies, but we were unable to find an evaluation of feces for novel miRNAs[46,47]. Likewise, while living endometrium has been widely examined for miRNAs, menstrual blood itself has not been evaluated, and because of the cellular changes that take place, novel candidate miRNAs may in fact be present[43,44,48]. Blood and semen are more widely described,

but many of the miRNAs discussed and tested by the authors for forensic use are associated with disease states, and were evaluated and described in concert with affected individuals or cell lines, and may not reflect the healthy individual miRNA transcriptome in any particular body fluid. Thus, the miRNAs present in the majority of forensically-relevant body fluids have not been catalogued thoroughly, and the studies were hampered by the limits of the available data on biological fluids.

The methods dependent on screening the expression of only known miRNAs are limited as they survey only pre-existing miRNAs, favor identification of the highly abundant miRNAs, and are not able to quantify absolute miRNA levels. Thus, the authors of the current literature applying miRNAs to forensic applications admittedly were working with an incomplete picture of the miRNA transcriptome, and by the time of publication, such studies were already obsolete.

In this body of work, we utilized a high-throughput sequencing (HTS) approach to sequence the entire transcriptome of small RNAs within the range of miRNAs (18-25 nt) in eight forensically relevant biological fluids. HTS analysis alleviates many concerns noted in previous studies, as it permits not only the identification of all small RNA transcripts in a forensic sample, but allows quantification and a sensitivity of detection can be several orders of magnitude greater than that obtained by previous screening methods[49]. In an early high-throughput sequencing study reported on serum for the purpose of detecting miRNAs that could be diagnostic for cancer, it was concluded that HTS data was reproducible and consistent among individuals, supporting the potential use of miRNAs as diagnostic probes not only in serum, but more importantly in other biological fluids as well[50]. Since 2010, HTS has become an industry standard, and has been applied to both clinical and forensic sample sets of thousands of patients/donors.

This study is the first of its kind applied to the full range of biological samples of forensic interest, with only one very recent report using HTS for identification of miRNAs indicative of blood and saliva[38]. Thus, it is timely, relevant, and highly justified. High-throughput sequencing to describe the miRNA transcriptomes within each of the eight body fluids will provide, in some cases, the first ever described miRNA transcriptome, and for other body fluids, will hopefully clarify the inconsistencies in miRNA expression seen in other studies.

HTS can sequence the entirety of a cDNA library in parallel, resulting in all miRNAs in the sample to be sequenced in proportion to their relative abundances[51]. This approach starts by the creation of a cDNA library representative of all of the transcribed small RNAs present in the sample. Briefly, specific 5'- and 3'-adaptor molecules are ligated on each small RNA and the resulting template reverse-transcribed followed by PCR amplification (≈ 15 cycles). Different forensic samples are tagged with a different 4-base sequence. The library is loaded onto an Illumina cluster station where up to forty-eight individually tagged, different samples can by analyzed on a single flow cell. Individual DNA molecules anneal to the high-density universal adaptors embedded in the flow cell and are amplified by cluster generation, yielding hundreds to thousands of duplicate copies. The resulting DNA clusters are sequenced via sequencing-by-synthesis technology where data is captured based on the recording of the fluorescence excitation of the specific base incorporated during each cycle. Transcriptome analysis of miRNAs using next generation sequencing, and in particular, the Illumina platform, has been successfully performed and reported by hundreds of investigators, and both clinical and forensic panels for human identity and phenotypic markers are now commercially available.

Research Rationale and Hypothesis

Specific Aim 1

In this study, 3-5 samples of eight forensically relevant biological fluids or secretions were subjected to high throughput sequencing on the Illumina® HiSeq platform. The biological samples evaluated were blood, seminal fluid (3 normal and 2 vasectomized individuals), perspiration, vaginal secretions, urine, feces, saliva, and menstrual secretions. Evaluation of normal seminal fluid and vasectomized seminal fluid was considered particularly important and has not been evaluated to our knowledge, and allows for an evaluation of the miRNA contribution (if any) from the testes and the ability to separate it out from the contributions of the downstream glands contributing to the ejaculate. Using these data, we identified not only miRNAs consistent in and specific to each biological fluid, but we also evaluated potential reference miRNAs for normalization purposes (Fig 1).

Specific Aim 1 Objectives:

- 4 individual samples of 8 different biological fluids will be collected according to University and US IRB accepted procedures (blood, seminal fluid (normal and vasectomized individuals), perspiration, vaginal secretions, urine, feces, saliva, and menstrual blood).
- The 36 body fluid samples will be preferentially isolated for small RNAs, cDNA generated, and subjected to deep sequencing on the Illumina platform.
- The sequencing data from the 36 samples will be separated by identity, and analyzed for both novel and known miRNAs using a bioinformatics approach.
- Novel and known miRNAs unique to a specific body fluid will be identified for further examination. This will include a comparison of presence and abundance levels of miRNAs between the body fluids, and among the 4 individuals tested, so that only miRNAs that are seen in all 4 individuals at detectable levels in the body fluid in question are tested further.

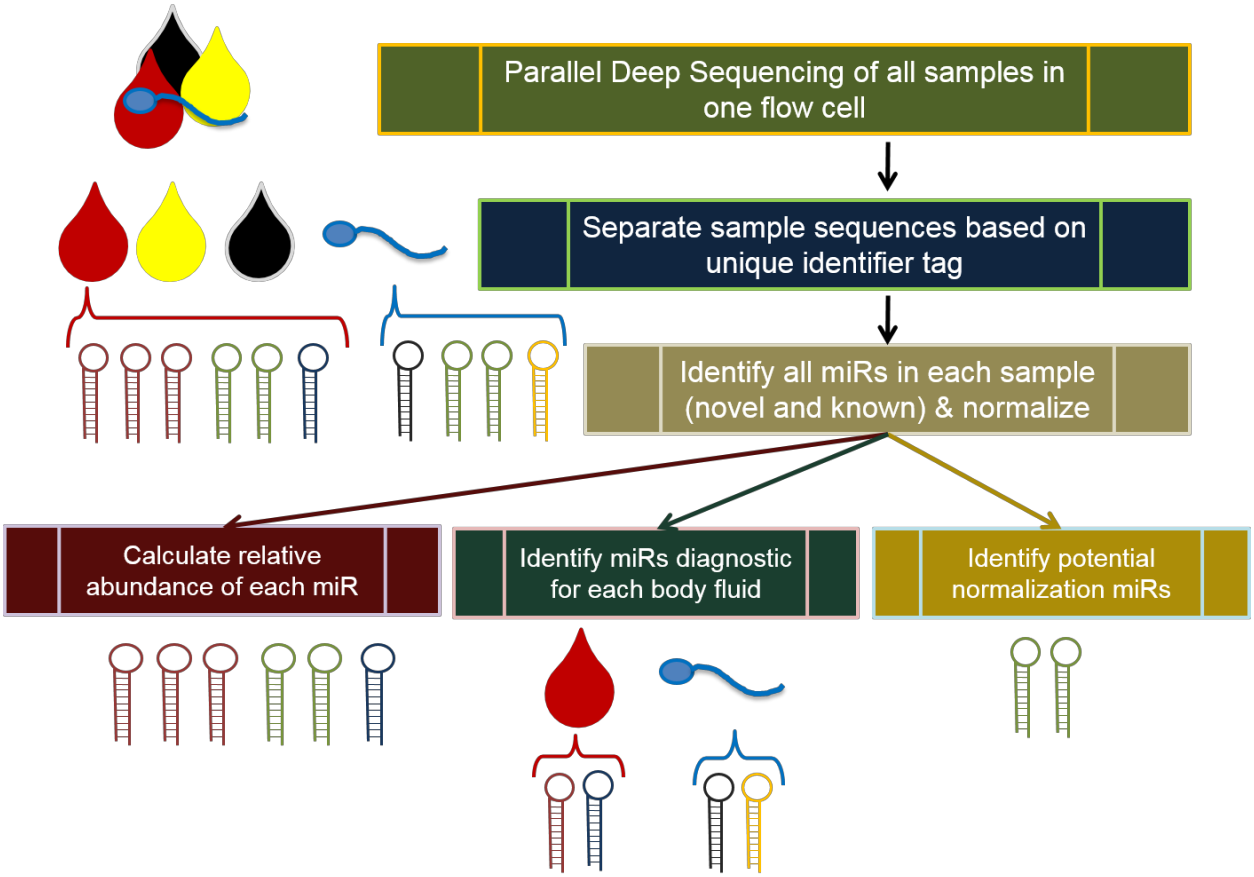


Figure 1: Workflow for miRnome sequencing data analysis using high-throughput sequencing. "Barcoding" of cDNA samples prior sequencing allowed for combining all samples onto a single Illumina® HiSeq flow cell. Data analysis was conducted using Partek® Flow®, and sequence reads were separated using known barcode sequences attached during library preparation, and sequences aligned to the known miRnome and sequences annotated. Abundance of miRNAs identified was calculated based on the number of reads/total annotated reads. Sequencing data was compared between and among biological sources for identification of diagnostic and normalization/endogenous reference miRNAs.

Specific Aim 2

Following high-throughput sequencing, we evaluated candidate miRNAs identified through HTS for each biological fluid both for diagnostic potential, also miRNAs that we observed in all samples as potential endogenous reference miRNAs. We evaluated each candidate miRNA using a tiered population study in order to conserve reagents and evaluate many candidates. Those body-fluid specific miRNAs that continued to show consistent expression within the population while also remaining specific for a given body fluids were further evaluated for species and organ specificity, and for sensitivity using a limit of detection study.

Specific Aim 2 Objectives

- Confirmation of novel miRNAs through qRT-PCR
- Review deep sequencing data of all samples tested for use as normalization miRNAs. Criteria will require that miRNAs chosen are observed in all samples, and at similar levels.
- Choose 1-2 miRNAs for each body fluid with high abundance and consistency between the four individuals for further characterization.
- Characterization studies – using qPCR assays for selected body-fluid specific and control miRNAs:
- purchase and validate probes for qRT-PCR analysis of novel and known miRNAs; optimize all miRNA qPCR probes
- Confirmation of body fluid specificity by testing miRNAs against a sampling of each of the 8 body fluids, plus an additional panel of 20 organs & tissues (commercially supplied)
- Evaluation of species specificity by comparison against animal body fluids. A minimum of 20 species will be evaluated using all relevant biological fluids. (commercially supplied)
- Evaluation of specificity, presence, and relative abundance among the population (minimum of 20 individuals of varying ethnicities, gender, and age/life stage)
- Abundance of the miRNAs within the same individual over short periods of time (i.e. vaginal secretions throughout the 28-day menstrual cycle, menstrual blood from days 1-7, urine in various stages of hydration, saliva and feces throughout the day and varying dietary changes)
- Limit of Detection. Evaluate detection of candidate miRNAs and reference miRNAs over a range of RNA concentration.

II. Methods & Materials

HTS and RT-qPCR validation

Sample Preparation

Samples of each body fluid were collected from 3-5 volunteers using a VCU-IRB approved human subjects research protocol. Every effort was made to collect samples from

individuals of varying age, ethnicity (self-described) and gender (when appropriate) (Table 1). Venous blood was collected into a Vacutainer® containing EDTA (Beckton, Dickinson & Company, Franklin Lakes, NJ), inverted for 15 seconds, and 200 µL aliquotted for RNA isolation. Urine, semen, and saliva were deposited into sterile collection cups and 200 µL (semen, saliva, urine-all assays except HTS) or 30 mL (urine - HTS) were aliquotted for RNA isolation. Vaginal secretions, menstrual secretions and feces were collected on swabs by the donors and returned in swab boxes. Perspiration was collected by gentle rubbing with swabs after moderate exercise from lotion and makeup free locations (outer bicep, small of back), and two swabs from the same location were combined for analysis.

Table 2: small RNA high-throughput sequencing - donor ages, ethnicities, and gender

	Donor 1	***Donor 2***	***Donor 3***	***Donor 4***
Blood	7309 – 23 yo AP ♂	7311 – 45 yo C ♂	7319 – 23 yo C ♀	7318 – 23 yo AA ♀
Saliva	7319 – 23 yo C ♀	7321 – 35 yo C ♀	7322 – 26 yo C ♂	8425 – 55 yo C ♂
Urine	Z16 - 66 yo C ♀	Z17 – 24 yo C ♀	Z18 -- 24 yo C ♂	Z19 -- 71 yo C ♂
Feces	7318 – 23 yo AA ♀	8401 – 21 yo AA ♂	8407 – 25 yo H ♂	7319 – 23 yo C ♀
Perspiration	7318 – 23 yo AA ♀	8413 – 23 yo AA ♀	8419 – 21 yo C ♀	9625 – 22 yo AA ♀
Vaginal Secretions	7314 – 66 yo C ♀	7315 – 23 yo H ♀	7319 – 23 yo C ♀	7321 – 35 yo C ♀
Menstrual Secretions	7315 – 23 yo H ♀	7318 – 23 yo AA ♀	7329 – 22 yo H ♀	7319 – 23 yo C ♀
Seminal Fluid	7322 – 26 yo C ♂	7520 – 20 yo C ♂	9602 – 24 yo C ♂	
Seminal Fluid (Vas)	7311 – 45 yo C ♂	9624 – 53 yo C ♂		

yo: year old AA: African American AP: Asian/Pacific Islander C: Caucasian H: Hispanic/Latin

Optimization of RNA isolation methods

Several RNA isolation methods were evaluated in order to identify the method(s) that would yield the highest level of RNA, which was necessary for high-throughput sequencing. All body fluids were evaluated using the following extraction methods: miRNeasy mini (Qiagen™ N.V., Venlo, The Netherlands), miRNeasy Serum/Plasma (Qiagen™), mirVana™ miRNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA, USA), miRCURY™ RNA Isolation Kit (Exiqon, Vedbaek, Denmark) and PureLink™ miRNA Isolation Kit (Thermo Fisher Scientific) Isolation methods. Additionally, urine RNA isolation was evaluated using the Urine Exosome RNA Isolation Kit

(Norgen Biotek, Ontario, Canada), and fecal samples were evaluated using the PowerMicrobiome RNA Isolation method (Mo-Bio, Carlsbad, CA, USA). All isolations were performed according to the manufacturer's protocol, with initial minor adjustments for centrifugation of lysis buffer from swabs using spin columns. For ease of comparison between isolation methods, elutions were standardized at 50 µL.

RNA Isolation and quantity evaluation

Based on evaluation of the RNA isolation results, RNA isolation was performed using the Qiagen™ miRNeasy mini (semen, menstrual secretions, saliva) or Serum/Plasma kit (blood, vaginal secretions, perspiration, urine), according to the manufacturer's instructions. After initial HTS analysis using RNA isolated from 200 µL of urine resulted in poor read depth and quality, a second analysis was undertaken using RNA was isolated from 30 mL of urine using the Norgen Biotek Urine RNA Concentration, Preservation and Isolation Kit, and Fecal RNA samples were isolated using the MoBio PowerMicrobiome RNA Isolation method, both according to the manufacturer's instructions. Protocols were modified for those samples on swabs (menstrual and vaginal secretions, perspiration and feces); after lysis, the swab was placed in a DNA IQ spin basket™ (Promega Corporation, Madison, WI, USA) and centrifuged for 1 minute at 13,000xg to allow flow of residual liquid back to the lysate.

RNA Integrity and quantitation analysis was performed using the 2100 Bioanalyzer and the RNA Pico quantitation method (Agilent Technologies Inc, Santa Clara, CA, USA) according to the manufacturer's recommendations for high-throughput sequencing preparation and endogenous reference evaluations. For sensitivity and copy number analysis, the Small RNA quantitation method (Agilent Technologies Inc) was performed according to the manufacturer's recommendations. RNA samples were also quantitated on a NanoDropTM 2000 UV-Vis Spectrophotometer (ThermoFisher Scientific, Inc., Waltham, MA), and select samples were quantitated for endogenous reference evaluations using the Qubit miRNA Assay Kit standard protocol on the Qubit 2.0 Fluorometer (ThermoFisher Scientific, Inc.).

High-Throughput Sequencing

Small RNA library preparation was conducted using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (Set 1) (New England Biolabs, Ipswich, MA, USA). High throughput

sequencing was conducted using the HiSeq 2500 (1x150) (Illumina®, San Diego, CA, USA). Adapter trimming and sequence analysis was conducted using Flow, v3.0 (Partek® Incorporated, St. Louis, Missouri, USA) using Bowtie 2 (v2.1.0) and miRbase v20[34,35,37] for alignment and annotation. miRNA sequencing reads were normalized using the reads per million (RPM) formula: (read counts of an individual miRNA/sum of read counts of all mappable miRNAs) multiplied by $1x10^6$.

RT-qPCR analysis for evaluation and validation of candidate miRNAs

Quantitative reverse transcription was carried out via the qScript™ miRNA Quantification System (Quanta Biosciences Inc., Gaithersburg, MD). Reverse transcription was carried out according to the manufacturer's protocol using RNA extract (7 µL of RNA extract for candidate miRNA evaluation and validation studies, miRNA stability studies. 10 ng for endogenous reference miRNA and varying quantities for LOD studies). qPCR reactions were prepared in triplicate for each sample using a modified protocol: 6.25 µL of PerfeCTa SYBR Green SuperMix (2X), 0.25 uL (2.5 µM) PerfeCTa miRNA Assay Primer, 2 µL of cDNA reaction, and 4 µL of nuclease-free water. Thermal cycling was conducted on the Life Technologies Prism 7500 instrument (Life Technologies, Foster City, CA) using the following parameters: 95ºC for 2 minutes, followed by 40 cycles of 95ºC for 5 seconds, 60ºC for 15 seconds, and 70ºC for 30 seconds (data collection). qPCR analysis was conducted using SDS software, v1.3.1 (Life Technologies). Negative amplification controls were included on each plate and DNase-treated controls for each evaluated miRNA were included and did not impact analysis. All treatments were performed on a minimum of three treated samples, each from a different donor, with 3 technical (qPCR) replicates for each treated sample.

Standard Curve Preparation

5'-phosphorylated oligonucleotides corresponding to the sequences of miR-144-3p (5'-uacaguauagaugauguacu) and miR-200b (5'- uaauacugccugguaaugauga) were synthesized at the 50 nm scale by Life Technologies, Inc. and purified using HPLC to ensure accurate sequence and copy number. The lyophilized oligonucleotides were resuspended in sterile, nuclease-free water to a stock solution of 100 µM. The concentration was confirmed via analysis using the NanoDrop™ 2000 UV-Vis Spectrophotometer (ThermoFisher Scientific, Inc.) and serially diluted

from 10^{13} to 10^{5} copies/µL. The standards were then reverse transcribed and subjected to qPCR in a minimum of duplicate standards as previously described using the qScript™ miRNA Quantification System (Quanta Biosciences Inc.).

Human Organ Panel Analysis

In order to evaluate miRNA specificity in a variety of human organs and tissues, 10 µg of each sample from the FirstChoice® Human Total RNA Survey Panel (Ambion, Life Technologies, Inc.) was reverse transcribed and subjected to qPCR as previously described using the qScript™ miRNA Quantification System (Quanta Biosciences Inc.). Samples evaluated included: bladder, brain (cerebellum), brain (whole), bone marrow, spinal cord, uterus, adrenal gland, colon, kidney, liver, lung, fetal liver, placenta, prostate, skeletal muscle, small intestine, spleen, testis, thymus, and salivary gland.

Species Specificity

Blood, urine, saliva, and fecal samples were collected from animals expected to have interactions with humans and thus likely to be involved with evidence (pets, livestock, local wildlife). RNA was isolated using the Qiagen™ miRNeasy mini kit (Qiagen, Inc.) according to the manufacturer's recommended protocol. RNA quantity and quality was assessed using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Inc.). 7 µL of each RNA extract was reverse transcribed and subjected to qPCR as previously described using the qScript™ miRNA Quantification System (Quanta Biosciences Inc.).

Data Analysis

Bestkeeper Normalization Analysis

Bestkeeper is an Excel-based algorithm software program that analyzes the cycle threshold values of all pairs of candidate normalization miRNAs in a pair-wise correlation analysis to determine the most stable miRNA for endogenous reference control purposes[52]. The Bestkeeper software was downloaded and utilized in Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) according to protocol. Average Cq values from Qubit and Bioanalyzer-quantitated RNA samples subjected to RT-qPCR analysis were entered into the software. The correlation coefficient (r) values were assessed, defined as a measure of the

strength of linear relationship between the test samples and the Bestkeeper index, which is defined as the geometric mean of the test samples[45]. The r value for each candidate miRNA provides a measure of the correlation of miRNA expression for any given single miRNA among all eight body fluids, given constant RNA input quantity.

RT-qPCR Data Analysis

The cycle threshold values (Cq) for the triplicate wells of each sample were averaged and the delta cycle threshold values calculated. The delta Cq (dCq) was calculated by averaging the Cqs of Let-7i and Let-7g in each sample and then subtracting the average Let Cq from the average Cq for the body fluid specific miRNA used. Pairings of body-fluid specific miRNAs with normalization miRNA probes was performed using the same reverse-transcription reaction, in an effort to minimize variation.

Microsoft-Excel and Past software (University of Oslo, Oslo, Norway) was used to analyze significant differences between the body fluids[53]. Using multivariate and univariate statistics where applicable, a students t-test or Analysis of Variance (ANOVA) was conducted to determine significant differences. When ANOVA was performed, a Tukey's pairwise comparison was then used to determine which body fluids were causing the significant differences.

miRNA Stability in compromised samples

Sample Collection & Treatment

Blood, urine, semen, and saliva were collected from volunteers under the human subjects research protocol approved by Virginia Commonwealth University's Institutional Research Board. Urine, semen, and saliva were deposited into sterile collection cups and 50 µL (semen, saliva) or 100 µL (urine) was applied to cotton swabs or cloth. Blood was collected into a Vacutainer® containing EDTA (Beckton, Dickinson & Company, Franklin Lakes, NJ), inverted for 15 seconds, and 50 µL immediately applied to cotton swabs or cloth. All samples were allowed to dry for 24 hours, swabs or stains cut and placed into 1.5 mL microcentrifuge tubes, and stored at -20ºC before and after subjection to treatment.

Samples undergoing irradiation were placed on a UVP High-Performance ultraviolet transilluminator (UVP, Upland, CA, USA) at 302 nm for 4 hours. Swabs were placed directly onto the sanitized transilluminator surface, and replaced into the 1.5 mL microcentrifuge tube after treatment. Samples undergoing heat treatment were incubated at 55ºC or 95ºC for .5, 1,

2, 4, or 24 hours. For samples exposed to chemical conditions, 100 µL of 87 mM or 870 mM sodium hypochlorite (1:10 dilution and full-strength household bleach, respectively), glacial acetic acid, or household dishwashing detergent were applied to the samples. The microcentrifuge tubes containing the treated samples were left open to air dry for 72 hours. The samples were then stored at -20ºC until isolation.

RNA Isolation and Analysis

RNA isolation was conducted on all samples using the Qiagen miRNeasy mini kit (Qiagen N.V., Venlo, The Netherlands). Briefly, the entire swab or fabric cutting was placed in 700 µL of QIAzol lysis reagent, and incubated for 5 minutes at room temperature, vortexing every minute. The swab or cutting was then placed in a DNA IQ™ Spin Basket (Promega, Madison, WI, USA), and centrifuged at 13000xg for 3 minutes. The resultant lysate was then processed according to protocol and total RNA eluted in 30 µL of RNase-free water. RNA was quantitated using the NanoDrop ND-2000 UV Spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA).

Quantitative reverse transcription was carried out via the qScript™ miRNA Quantification System (Quanta Biosciences Inc., Gaithersburg, MD) as previously described. All treatments were performed on a minimum of three treated samples, each from a different donor, with 3 technical (qPCR) replicates for each treated sample. Let-7g was chosen as a representative miRNA due to well-established clinical use as an endogenous and high-expression reference miRNA[54,55], as well as its shown use as an endogenous reference miRNA in this study. In order to consider additional miRNAs and demonstrate consistent modulation upon sample compromise, indicating relatively consistent degradation of the miRnome in general, miRs-16, 21, and 24 were also evaluated in one set of treated and untreated samples.

Data Analysis

One body fluid donation each from 3 different donors was used for all sample treatments. Use of identical RNA extraction, reverse transcription, and qPCR parameters provided a consistent comparison of miRNA quality from treated sample to control.

In order to determine relative changes in sample integrity, the average Cq (of technical triplicate qPCR wells) of the treated sample was compared to the average Cq of the paired untreated control. Samples in which one or two of the 3 replicate wells failed to amplify were

averaged with a Cq of 40 for the failed well, and are indicated in the figure legends. The average and standard deviation of the positive untreated control was determined from the average Cq data of the untreated controls of the three separate donors, and thus is relatively large.

With all conditions identical except for treatment, comparison to the untreated control was deemed the best method for analysis. As endogenous reference miRNAs for forensically relevant body fluids would be expected to degrade similarly to body-fluid specific miRNAs, we felt this to be the most discriminating measure rather than a dCq analysis. A one-way ANOVA with a Tukey's Honest Significance Difference test was applied to determine significance between treated and untreated samples.

III. Results

Evaluation of RNA Isolation Methods

RNA isolation methods were evaluated for maximum RNA yield and amplifiable miRNA levels. We chose to evaluate isolation method efficiency through RT-qPCR analysis, as it is well known that UV spectrophotometry and even Bioanalyzer small RNA chips cannot precisely predict miRNA concentrations in the low concentrations observed in biological fluids, and qPCR quantitation has been shown to be a more precise method[56]. Consistent volumes of biological secretions were used, and elution volumes were normalized to 50 µL to allow for ease of comparison. With the exception of fecal samples, all RNA methods evaluated yielded similar RNA yields and detectable miRNA through RT-qPCR analysis (Figures 2A and B). Choice of the isolation method for each biological fluid was based not only on RNA yield and high relative levels of the ubiquitous and abundant miRs-16 and -21, but also on ease of use due to our anticipated analysis of high numbers of population samples

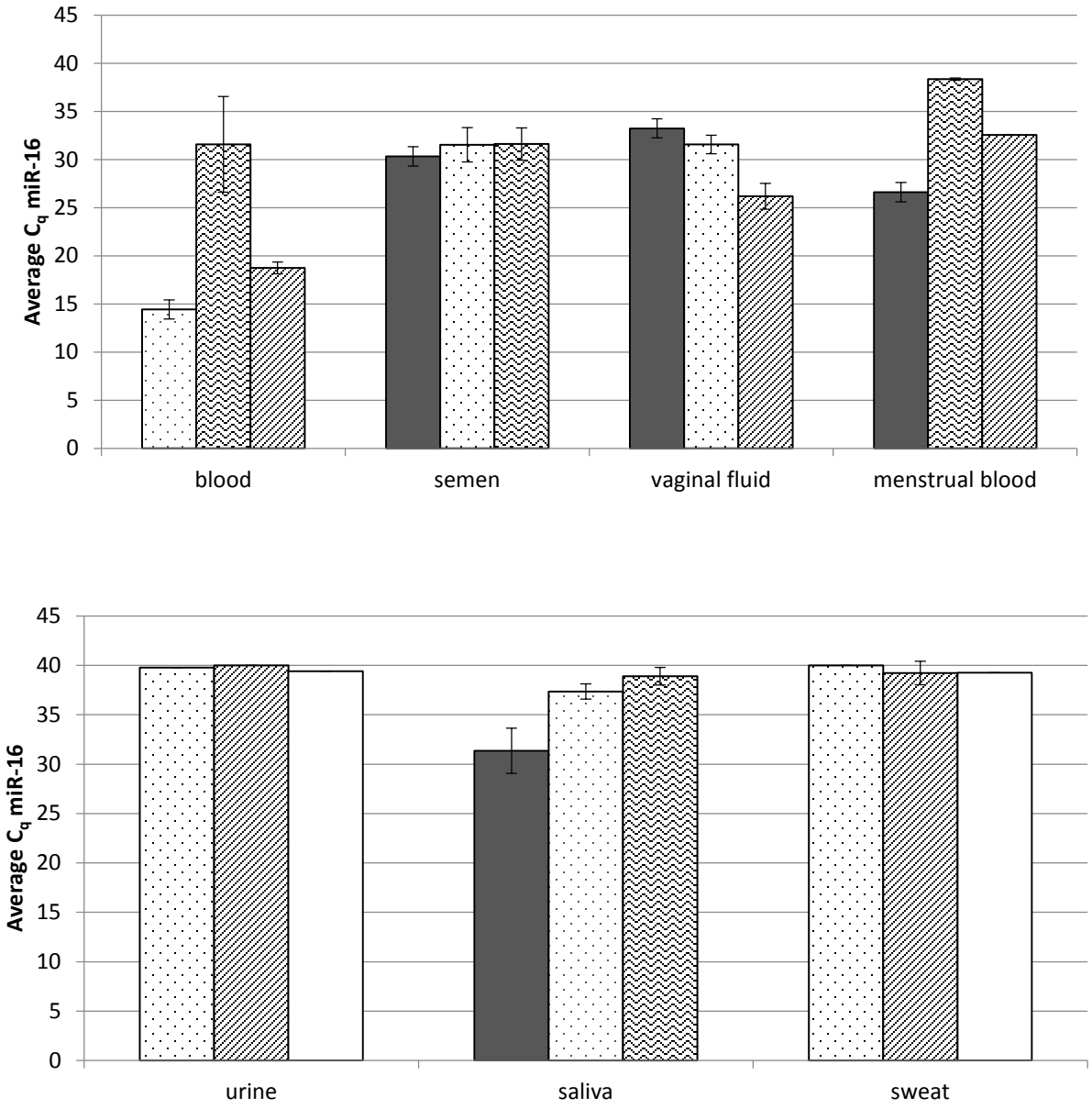


Figure 2: Evaluation of miR-16 abundance for the top three isolation methods for each body fluid shows no major differences in efficiency. RT-qPCR analysis of miR-16 relative abundance in order to assess isolation efficiency. Each sample was analyzed in triplicate and the average C_q value was calculated. **Top:** Blood, semen, vaginal secretions and menstrual secretions. **Bottom:** Urine, saliva, and perspiration.

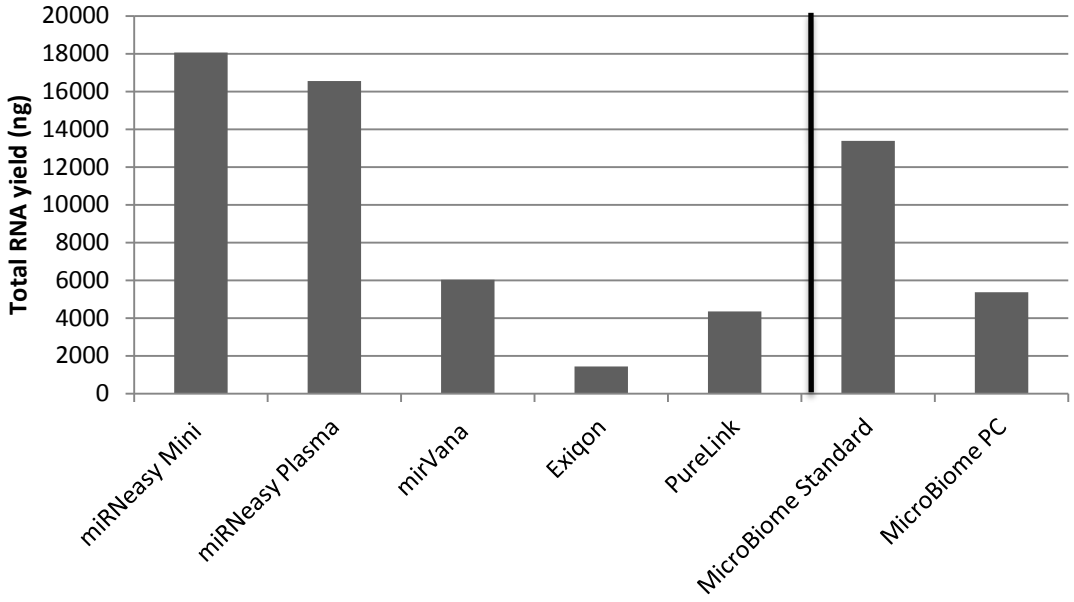
Initial evaluation of one fecal sample revealed high RNA yield and acceptable amplifiable miRNA levels; however, subsequent additional sample analyses resulted in three out of four

samples returning undetected amplification when subjected to RT-qPCR (data not shown). Further exploration with additional purification steps and evaluation of other methods, including TriZOL reagent (Thermo Fisher Scientific, Inc.), resulted in consistent sample amplification failures, likely due to the inhibitors inherently present in fecal samples. Both standard and phenol:chloroform versions of the PowerMicroBiome™(Mo-Bio) resulted in consistently high levels of RNA and amplifiable miRNAs (Figures 3A and B). There was virtually no difference in amplified miR-16 levels between the Standard method and the Phenol:Chloroform method of the PowerMicroBiome™ method, but total RNA yields differed, likely due to differences in bacterial RNA coextraction.

For the remainder of the reported work, each biological fluid was extracted with the optimal RNA isolation method (Table 3), with the exception of the miRNA stability in compromised samples sub-project. The urine samples were first subjected to high-throughput sequencing using the optimized method, but low read number and quality required reanalysis using the Norgen Biotek Urine RNA Concentration, Preservation and Isolation Kit, which was capable of concentrating RNA from 30 mL of urine and has been shown to yield maximal miRNA yields in other work[57]. While we chose to take the approach of different isolation methods for different biological fluids for this study, that would not be our recommendation moving forward due to the potential confusion and real likelihood of mixed samples. With the exception of feces, all of the RNA methods resulted in similar amplifiable levels of miRNAs, and it would be standard for a caseworking laboratory to choose one method for all biological fluids excepting fecal samples.

Table 3: Optimal RNA isolation methods for each biological fluid, used in this study

Blood, Urine, Perspiration, Vaginal Secretions	Qiagen™ miRNeasy Serum/Plasma
Saliva, Menstrual Secretions, Seminal Fluid	Qiagen™ miRNeasy mini
Feces	MoBio PowerMicroBiome™ Standard
Urine (HTS)	Norgen Biotek Urine RNA Concentration, Preservation and Isolation Kit



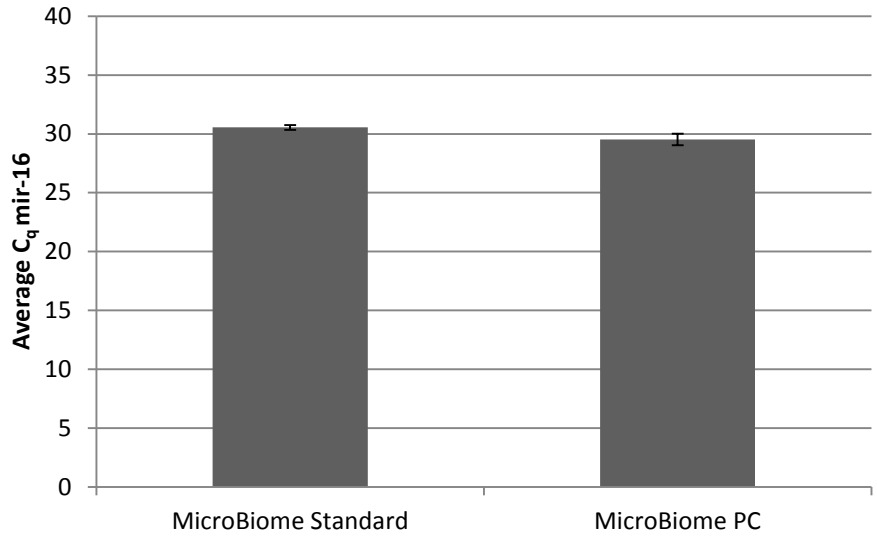


Figure 3: RNA isolation from fecal samples requires a feces-specific RNA isolation method. A (Top): Total RNA yields among five original methods tested were similar, but were not consistently detectable with RT-qPCR (>75% amplification failures). The PowerMicroBiome™ Standard method produced comparable yields of total RNA with **B (Bottom):** successful RT-qPCR analysis of miR-16 relative abundance. Each sample was analyzed in triplicate and average C_q value calculated.

High Throughput Sequence Analysis

High-throughput sequencing reads from the sampled body fluids via Illumina Hi-Seq® using the NEBNext® Multiplex Small RNA Library Prep Set resulted in sequence reads of all small RNAs within a particular sample (Table 4) . While the number of raw sequencing reads was similar in most cases compared to other studies[38,57-59], raw sequencing reads in vasectomized semen were significantly higher than those found in whole semen ($p<0.05$). As RNA was isolated from 200 µL of each seminal fluid contribution, the small RNAs isolated from vasectomized samples were taken from a larger relative volume of seminal plasma, as the contribution of spermatozoa has been removed, thus increasing the "concentration" of the downstream glands contributing the majority of the seminal fluid. This could represent a significant bias, and we were able to show parallel observations using RT-qPCR analysis (see Candidate miR validation and Limit of Detection studies).

Table 4: High-Throughput Sequencing Read Data (data averaged among the tested samples)

Sample	Raw Sequencing Reads	PHRED Score	Alignment %
Blood (n=3)	1,903,060 ± 146,235	39.2	60.1 ± 2.8%
Saliva (n=4)	1,480,782 ± 385,984	39.4	1.5 ± 2.3%
Urine (n=4)	3,642,441 ± 1,749,066	38.0	3.3 ± 2.5%
Feces (n=4)	1,765,452 ± 488,850	39.2	0.2 ± .04%
Mens Bld (n=4)	1,056,133 ± 652,312	38.5	4.7 ± 4.9%
Vag Fluid (n=4)	1,276,136 ± 1,490,639	39.4	1.4 ± 0.4%
Semen (n=3)	404,903 ± 66,913	39.6	6.4 ± 1.2%
Semen (Vas) (n=2)	2,165,497 ± 1,407,297	38.9	9.6 ± 0.1%
Perspiration (n=4)	3,190,111 ± 415,386	37.7	0.9 ± 0.6%

PHRED Quality of raw sequencing reads varied between 37 and 39.89, with one outlier at 34.7 (Table 4). A quality score of greater than 32 indicates high confidence in the sequence[60]. While the number of raw sequencing reads was consistent among all secretions between 1 and $5x10^6$ (with the exception of whole semen), the percentage of annotated sequences aligned to the raw sequences tended to cluster by secretion. Reduced annotated

reads (aligned to known miRNAs) were observed in all fecal samples, and 3 out of 4 donations of saliva and vaginal secretions. However, the relationship between high raw sequencing reads and high % annotation did not always correlate. Many of the biological fluids with annotated sequencing reads are those known to harbor large microbiota loads (saliva, contribution from vaginal secretions, feces). Depth of annotated miRNA coverage varied by secretion type, with blood and vasectomized semen with highest annotated reads, menstrual secretions, saliva, whole (non-vasectomized) semen, perspiration, urine and vaginal secretions ranging from 3000-10,500 reads, and fecal samples averaging 880 average annotated reads (Figure 4). The disparity between annotated vasectomized and non-vasectomized semen could be random variation, or could be associated with the high volume of piwiRNAs found in sperm cells, which can compete for sequencing reagents and thus reduce miRNA coverage[61].

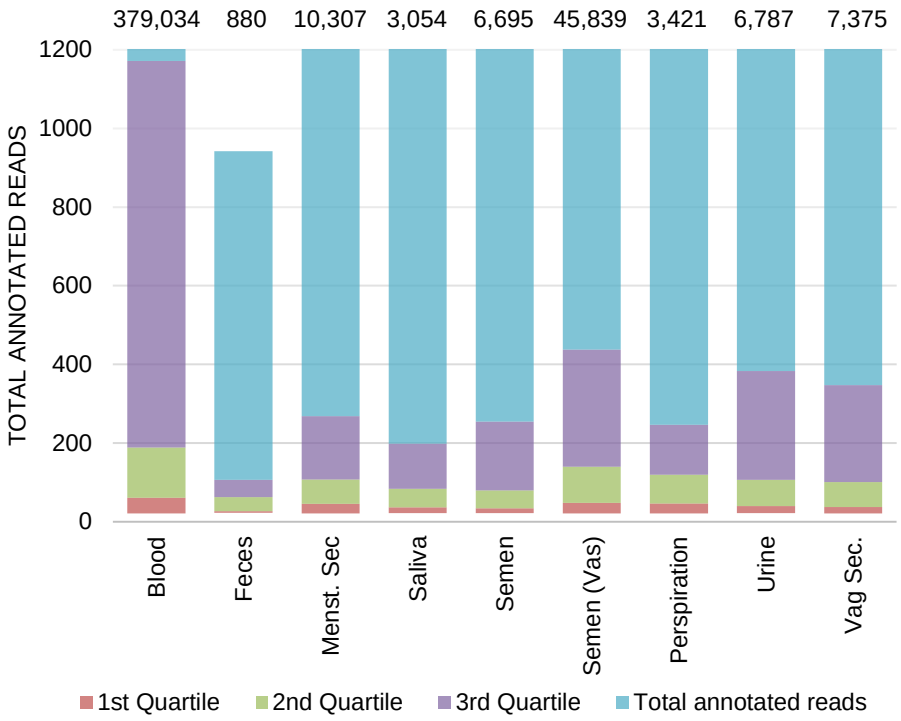


Figure 4: Depth of annotated miRNA coverage. Sequencing reads were aligned to mirBase (v20) and total read counts of annotated miRNAs (with read count >20) calculated. Data is the average of samples sequenced for each biological fluid type, and organized in quartiles to indicate relative abundances. Top labels indicate total number of annotated sequencing reads.

The number of unique miRNAs annotated and identified in each biological fluid was found to be similar among donors of a given body fluid, as were the quantity of raw sequencing

reads obtained. Again, coverage of those fluid types with high bacterial loads or low RNA quantities yielded fewer identified miRNAs (Figure 5). Interestingly, comparisons of the miRNAs expressed in a particular biological fluid among the donors showed a high degree of diversity, with only a minority of common miRNAs expressed among the donors (annotated HTS data in Supplemental File). Those miRNAs commonly observed in all biological samples tested were identified as constitutive miRNAs that had the potential to be utilized for normalization purposes. miRNAs expressed in all donors of a particular biological fluid, and either found solely in that fluid or with a significant expression pattern in that fluid were identified for further evaluation as body fluid specific candidate miRNAs (Table 5). Over 70 miRNAs were identified as exclusively detected in all blood samples; thus, we chose a subset of the most abundant miRNAs to take forward for further analysis.

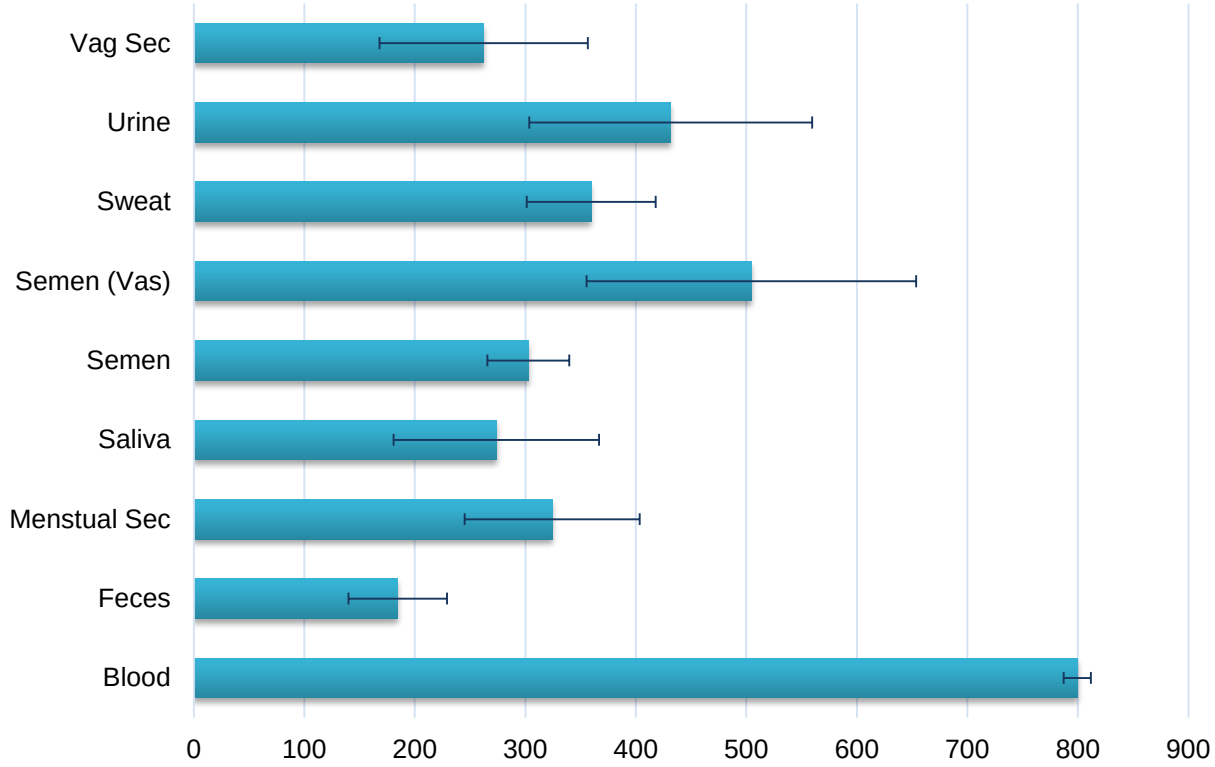


Figure 5: Unique annotated miRNAs identified in each biological fluid by high throughput sequencing. Data is the average number of annotated miRNAs from 3-5 individual donors.

Table 5: miRNAs identified through HTS as body-fluid specific candidates

Venous Blood	mir-128-2-3p mir-3615-3p mir-15a-5p mir-194-1-5p mir-19a-3p	**Saliva**	miR-1246-5p miR-23a-3p miR-223-3p
Menstrual Secretions	miR-200b miR-26b	**Urine**	miR-4497-5p miR-4488-5p
Vaginal Secretions	miR-210	**Feces**	miR-101-3p miR-144-3p
Semen	miR-10b-5p miR-10a-5p miR-183 miR-100-5p miR-200a-3p	**Perspiration**	miR-203b-5p miR-1290-3p miR-320b-3p miR-320c-3p

In evaluating the high-throughput sequencing data generated in this study and comparing to those miRNAs identified in previous forensic research as indicative for a particular body fluid, we found a number of discrepancies in which miRNAs identified by previous work were not observed in our HTS data. Our conclusions from this and further validation work are that for body-fluid specific miRNA identification using an amplification method, we found that HTS data of small RNAs from forensic-sized biological samples is of low depth, and only those miRNAs with high abundance were identified. Large sample volumes, such as those used in the clinical setting, or enrichment for human miRNAs prior to HTS analysis would perhaps alleviate this shortcoming, but would not be as relevant for analysis of forensic evidence. Subsequent RT-qPCR analysis has yielded positive amplification results for miRNAs not identified to be present in a particular body fluid based on the HTS data (candidate miRNA validation results follow and demonstrate this point). This corresponds to other work in the clinical arena, in which RT-qPCR analysis was found to be vastly more sensitive in identifying miRNAs of interest[62,63].

Evaluation of potential endogenous reference miRNAs for normalization purposes

miRs 451a-5p, 21-5p, and let-7 g and i were identified as the **only** four miRNAs present in all donors and all biological fluid sample types, and were evaluated for potential as miRNAs that could be used for normalization purposes (Table 6). Optimally, one or multiple endogenous miRNAs would be found in all relevant biological fluids, and at similar expression levels both between individuals and between body fluids. In order to investigate the identified miRNAs for this purpose, 10 ng of RNA (calculated using Bioanalyzer RNA pico method) from three RNA samples from each biological fluid type were subjected to RT-qPCR analysis and evaluated for consistency of expression both within and among the biological fluids. Analysis using BestKeeper®[52] determined that let-7g and i were most consistently expressed, resulting in r values above 0.9. Confirmation with an additional set of 24 additional samples (10 ng, quantitated with the Qubit miRNA quantitation method) yielded an even stronger coefficient of correlation of 0.971 and 0.947 for let-7g and i, respectively. miRs 451a-5p and 21 were eliminated as potential endogenous miRNAs both based on their inconsistency of expression (r=.734 and .763, respectively) as based on RNA quantity as well as their known dysregulation in a variety of carcinomas and other conditions[64]. Based on the data, we found that averaging the C_Ts of let-7g and i allow for a more consistent analysis among all eight body fluid types sampled, similar to the normalization proposed by Chen et al[55], and analysis of expression of the two miRNAs in all body fluids showed a remarkably low variation across and within samples (Figure 6). Consequently, based on the poor quantitation results observed with low-quantity miRNA samples, we chose to use this normalization method for all miRNA studies, utilizing a constant RNA volume input for RT-qPCR instead of depending on inaccurate quantitation methods. This approach has been consistently used by other researchers as well, and would streamline analysis if miRNA analysis were to be implemented by caseworking laboratories.

Table 6. Average relative expression (RPM) of candidate endogenous reference miRNAs.

	Let-7g	Let-7i	miR-21-5p	miR-451a-5p
SEMEN (N=5)	0.019	0.009	0.105	0.009
SALIVA (N=4)	0.020	0.019	0.120	0.106
PERSPIRATION (N=4)	0.006	0.011	0.018	0.076
URINE (N=4)	0.015	0.026	0.024	0.087
BLOOD (N=4)	0.039	0.035	0.023	0.291
MENST. SEC. (N=3)	0.015	0.027	0.275	0.076
VAG SEC (N=4)	0.008	0.008	0.227	0.039
FECES (N=4)	0.022	0.025	0.134	0.340
STANDARD DEVIATION	±10143	±9939	±96239	±120146

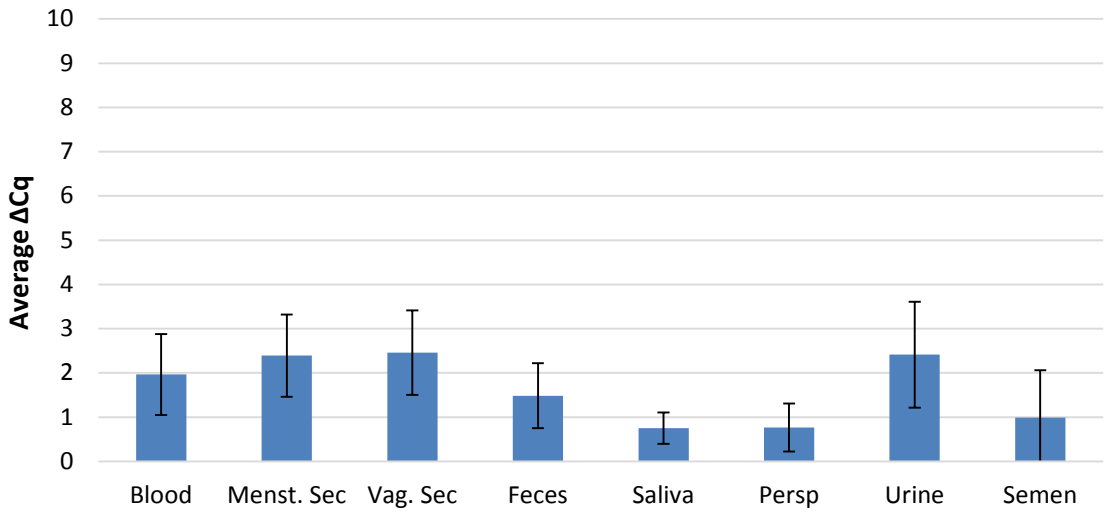


Figure 6: Assessment of average ΔCq demonstrates low variation between let-7g and let-7i both within and among body fluids. 10 ng of RNA extract was subjected to RT-qPCR analysis and ΔCq calculated. ΔCq: $Cq_{(let\text{-}7g)} - Cq_{(let\text{-}7i)}$. n=5 separate donors for each biological fluid, average of technical triplicate wells for each miRNA.

Optimization of a synthetic standard curve for accurate quantitation of miRNAs

Given our lack of confidence in the available miRNA quantitation methods for low-quantity samples such as perspiration and urine, and the overinflated RNA quantities derived from bacterial small RNA contributions to fecal, vaginal, and saliva samples, we chose to couple

the demonstrated endogenous reference normalization using an average of the Cqs of lets-7g and i with an exact analysis of miRNA copy number in a given sample. In order to do so, we confirmed reported concentrations of the oligonucleotides with triplicate UV spectrophotometry readings, and prepared serial dilutions of two synthetic miRNAs, miR-144-3p and miR-200b and evaluated precision and accuracy of the synthetic standard curves. We found that the dynamic range for amplifiable miRNAs was 10^5-10^{12} copies/well. Given that with our analysis system, 7 µL of RNA extract is used for a 20 µL reverse transcription reaction, and 2 µL is pipetted into the qPCR reaction, this represents an original copy number range of $1x10^5$-$1x10^{12}$ original copies/µL of RNA extract, assuming similar processivity rates of reverse transcriptase in the synthetic and human samples.

R^2 values for the replicates consistently surpassed 0.99 (Figure 7); the standards were reproducible over multiple wells within the same plate, across plates and days, and reverse transcription reactions (Table 7). We found that separate serial dilution preparations cause minor differences in Cq levels when amplifying the same quantities, but have addressed the variability by including carrier DNA in the serial dilution preparations; nonetheless, both miR-144 and miR-200b demonstrated high precision and reproducibility.

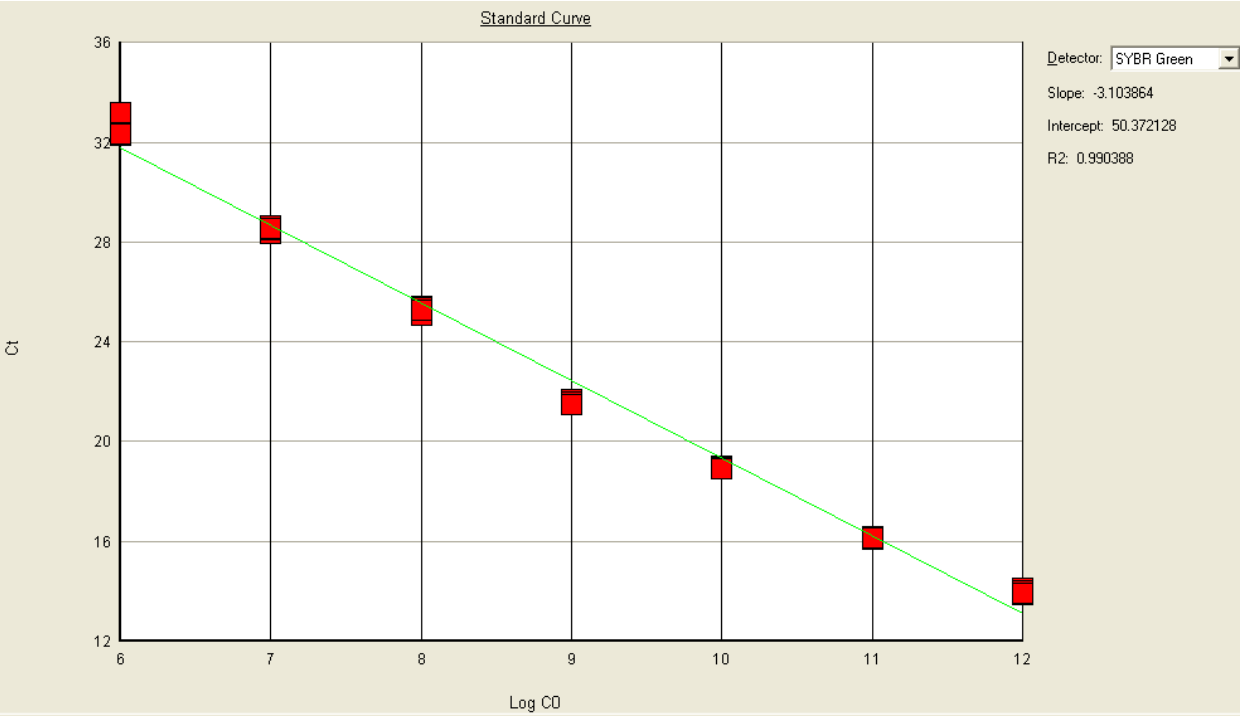


Figure 7: miR-144-3p standard curve. n=5 replicate wells demonstrates high precision (representative of 9 independent experiments). Analysis performed by SDS software, version 1.3.1 (Thermo Fisher Scientific, Inc.)

Table 7: Precision and reproducibility of synthetic miR-144-3p

		STANDARD DEVIATION	
Copies/rxn	**Avg Cq**	**replicate**	**RT**
10^{12}	13.35	0.14	0.70
10^{11}	15.70	0.10	0.37
10^{10}	18.67	0.13	0.06
10^{9}	21.44	0.07	0.26
10^{8}	25.10	0.09	0.38
10^{7}	28.65	0.16	0.92
10^{6}	33.10	0.33	1.42

Selection of candidate body-fluid specific miRNAs based on HTS data

In order to be identified as a body-fluid specific miRNA initially according to the HTS data, a candidate miRNA must have been found in all 3-5 donors of the given biological fluid,

and not detected in the HTS data for any other biological fluid sample. 72 miRNAs were identified through the HTS data as expressed only in blood, and common to the four donors (Supplemental Table 1). Unfortunately, the other biological fluids had neither the abundance of annotated miRNAs, nor such a large number of "unique" miRNAs to blood. Urine was the only other body fluid that demonstrated expression of miRNAs (2 miRNAs) found in no other samples. For saliva, semen, menstrual and vaginal secretions, and perspiration, miRNAs common to all donors of the body fluid in question were evaluated for minimum expression in other samples, resulting in a differential analysis based on expression levels along with number and source of expression (Table 8). As the HTS data indicated, the read depth of the fecal samples was dramatically lower than the other samples, resulting in very few annotated miRNAs for all four samples (11-33 total annotated miRNAs). This is likely due to the massive competition for sequencing components by the small RNAs from the bacteria present in fecal samples. Because of the low depth of coverage, it was not possible to identify miRNAs common to all fecal donors that were not expressed in the other body fluid samples. Nine unique miRNAs were found to be common to all four fecal samples; four of those were the endogenous reference candidate miRNAs, and thus were eliminated. Of the five remaining miRNAs, miR-101-3p and miR-144-3p were chosen for further evaluation, and were admittedly poor choices, given the expression profiles in the HTS data, and high known expression levels of miR-144 in venous blood[65,66].

Table 8: Expression profiles for initial candidate body-fluid specific miRNA

		Urine	MB	VF	Saliva	Perspiration	Feces	Blood	Semen
Semen (n=5)	10b-5p	0.75	1.0	0.5	-	-	-	-	1.0
	10a-5p	0.75	0.75	0.5	-	-	-	1.0	1.0
	183	0.5	0.75	-	0.25	0.25	-	1.0	1.0
	100-5p	0.75	0.25	-	-	-	-	1.0	1.0
	200a-3p	0.75	1.0	0.75	0.5	0.25	-	-	1.0
Saliva (n=4)	1246-5p	0.75	0.75	-	1.0	0.25	0.25	-	0.4
	23a-3p	0.25	-	0.5	1.0	0.75	-	-	0.2
	223-3p	-	1.0	0.66	1.0	-	-	1.0	-
Perspiration (n=4)	203b-5p	0.75	-	0.75	-	1.0	-	-	-
	1290-3p	0.75	-	-	0.25	1.0	-	-	-
	320b-3p	0.5	-	-	-	1.0	-	1.0	0.2
	320c-3p	0.25	-	-	-	1.0	-	1.0	-
Urine (n=4)	4497-5p	1.0	-	-	-	-	-	-	-
	4488-5p	1.0	-	-	-	-	-	-	-
Vag Sec (n=4)	210	-	0.25	1.00	-	-	-	1.00	-
	200b	-	1.00	1.00	-	-	-	-	-
Menst Sec (n=4)	26b	-	1.00	0.25	-	-	-	1.00	-
Feces (n=4)	101-3p	0.75	1.00	1.00	1.00	0.25	1.00	1.00	1.00
	144-3p	-	0.75	0.50	0.75	-	1.00	1.00	0.40
Blood (n=3)	128-2	-	-	-	-	-	-	1.00	-

Initial evaluation of candidate body-fluid specific miRNAs

In an effort to conserve reagents and sample, we performed an initial evaluation of each candidate miRNA by evaluating expression in 5 population samples for the target body fluid, and 2 each of the other 7 biological fluids. While a small set of samples, we were by no means positive that RT-qPCR analysis would be reflective of the sequencing data, and this stepwise approach would enable us to quickly eliminate candidate miRNAs that did not exhibit differences in expression. Relative expression was calculated by comparison of candidate

miRNA expression as compared to the endogenous reference miRNAs let-7g and let-7i (ΔCq: – $Cq_{(target)}$-$Cq_{(avg\ of\ let\text{-}7g\ and\ let\text{-}7i)}$). Our reservations regarding correlation between HTS and qPCR data were found to be relevant: not a SINGLE candidate miRNA identified as potentially body fluid specific through HTS was found to be discriminatory for the body fluid in question (Figure 8, Supplemental Figure 2). This includes those miRNAs that were found through the HTS data to be solely expressed in the body fluid of interest, with NO sequencing reads in the other 28-30 samples.

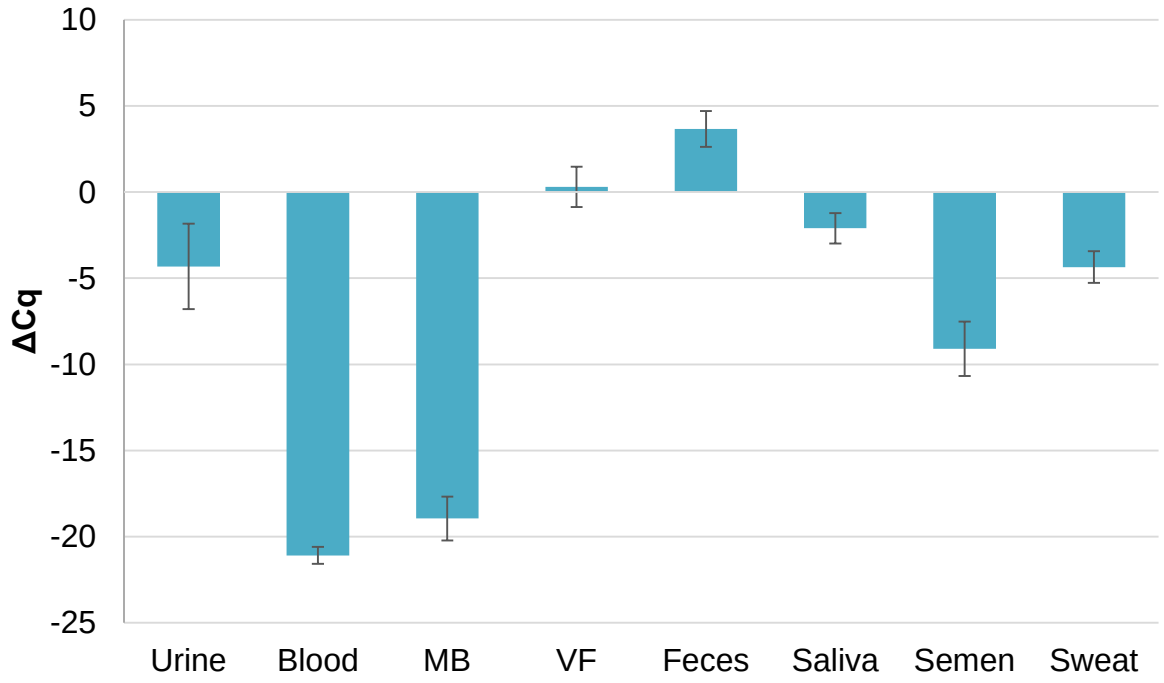


Figure 8: Representative Data for candidate miRNA expression as identified through HTS analysis. miRNAs identified as potentially body-fluid specific through high-throughput sequencing analysis were evaluated using RT-qPCR analysis. None of the miRNAs evaluated showed a correlative expression analysis pattern among body fluid samples as indicated through the HTS data. RT-qPCR analysis: n=5 of target body fluid (Above: miR-4488-5p for urine), and n=2 population samples for the other seven body fluids assessed. RT-qPCR data for other candidate miRNA initial evaluations are found in Supplemental Figure 2.

There are several variables that could have contributed to this observed discontinuity. Firstly, given the low numbers of annotated miRNAs in some biological fluids (likely based on

competing microbial small RNAs), the HTS analysis performed is not truly indicative of the full miRnome, but is rather a measure of the top expressed miRNAs within a given sample. This is specifically illustrated by the very high number and percent annotation of miRNAs in blood, which as a relatively sterile body fluid, would have minimal competing small RNAs for the limited reverse transcription and sequencing reagents. Thus, even though we did not observe sequence data for a given miRNA in the four donors of a specific body fluid, the miRNA very likely was present, but at levels undetectable by the HTS method without enrichment for human or exclusion of bacterial small RNAs, which is not possible at the present time at a miRnome-wide level. qPCR methods are well known to be the most sensitive methods, capable of detecting very low quantities of target sequences – certainly what was observed in the case of our study.

An additional variable contributing to our results is the discontinuity of expression analysis. For HTS analysis, we normalized expression to read counts per million. This method is absolute, while expression analysis for the candidate miRNAs was analyzed through expression relative to lets-7g and i. As discussed previously, lets-7g and i were found to correlate very closely with miRNA quantity, and thus this method is more reflective of the miRnome expression, and accommodating for variability in input RNA quantity of any given sample.

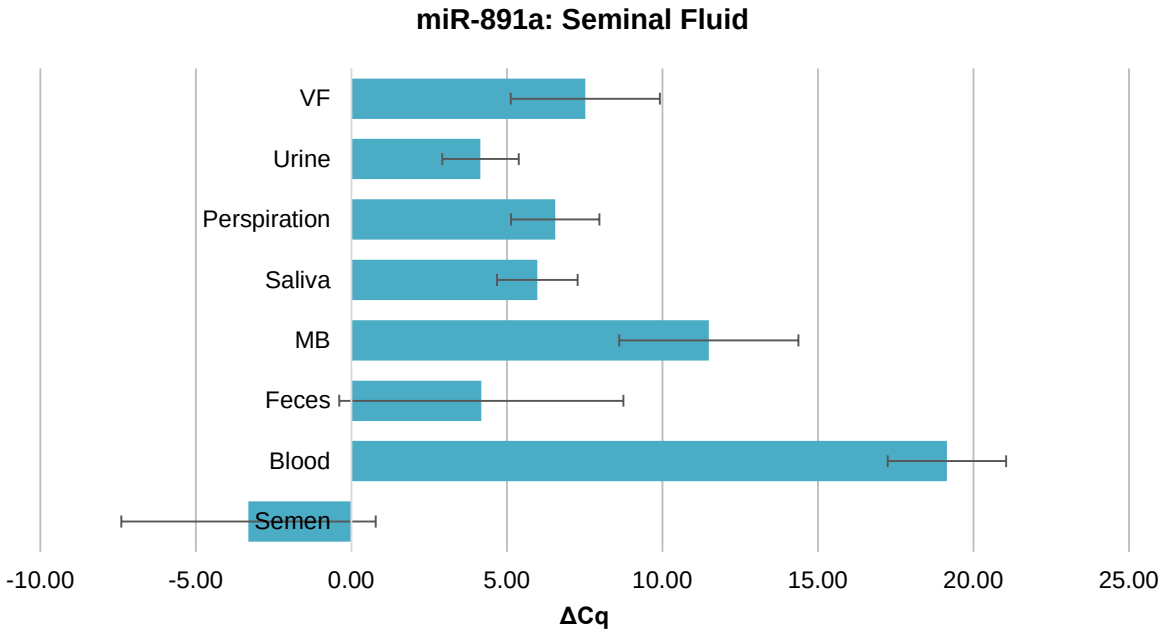
While we were quite disappointed in the failure of the HTS analysis to correlate with miRNA expression, we found that several of the miRNAs analyzed showed potential for differential expression in body fluids **other** than those expected. Consequently, based on these data, we chose to take the following candidates forward for a larger population analysis (Table 9). We were not able to identify candidate miRNAs for all eight body fluids at this point. miR-891a was chosen for evaluation based on the forensic literature, though it was not found at a significant level in the HTS data for the five semen samples tested.

Table 9: Narrowed field of body-fluid specific candidate miRNAs

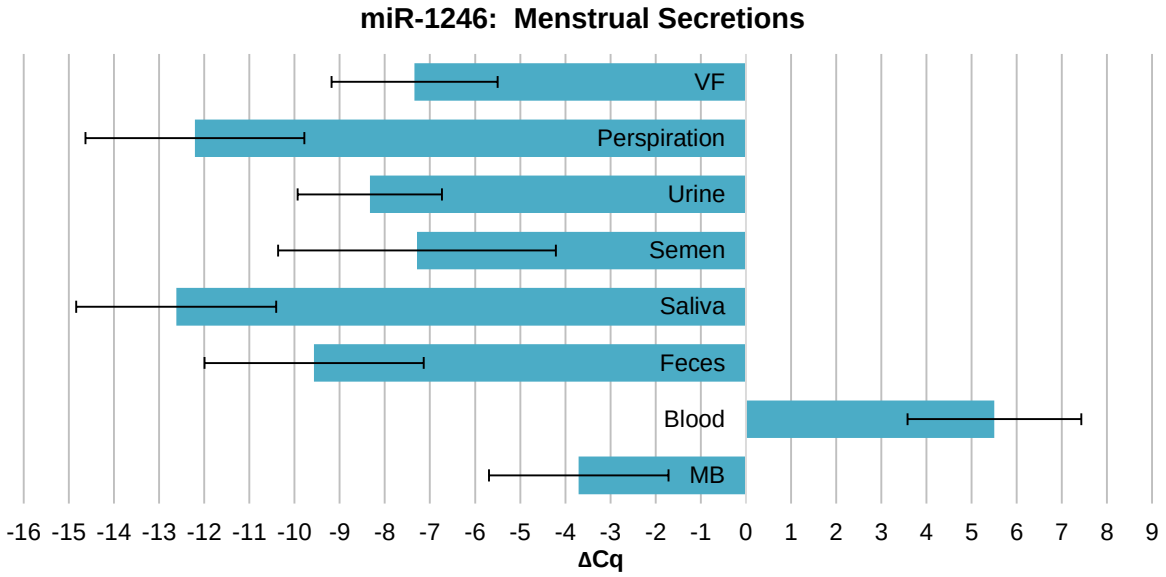
Target	miRNA(s) for evaluation
Blood	miR-200b
Semen	miR-26b, miR-891a
Vaginal Secretions	-----
Menstrual Secretions	miR-1246
Saliva	miR-23a-3p
Urine	miR-10b-5p
Feces	miR-320c
Perspiration	---------

Expanded evaluation of body-fluid specific candidate miRNAs

For the expanded evaluation of a narrowed field of candidate miRNAs, we performed an initial evaluation of each candidate miRNA by evaluating relative expression as described above in 20 population samples for the target body fluid, and 5 each of the other 7 biological fluids. Results were concordant with the initial evaluation RT-qPCR analyses. miRs-23a-3p, 26b, and 10b-5p were eliminated at this stage due to lack of statistical variation between their target and the other biological fluids. The remaining miRNAs were confirmed to demonstrate differential expression for the above targeted body fluids ($p<0.05$) (Figures 9, 10, and 11). Based on these data, we felt confident in moving forward with the six finalized miRNAs for further characterization.

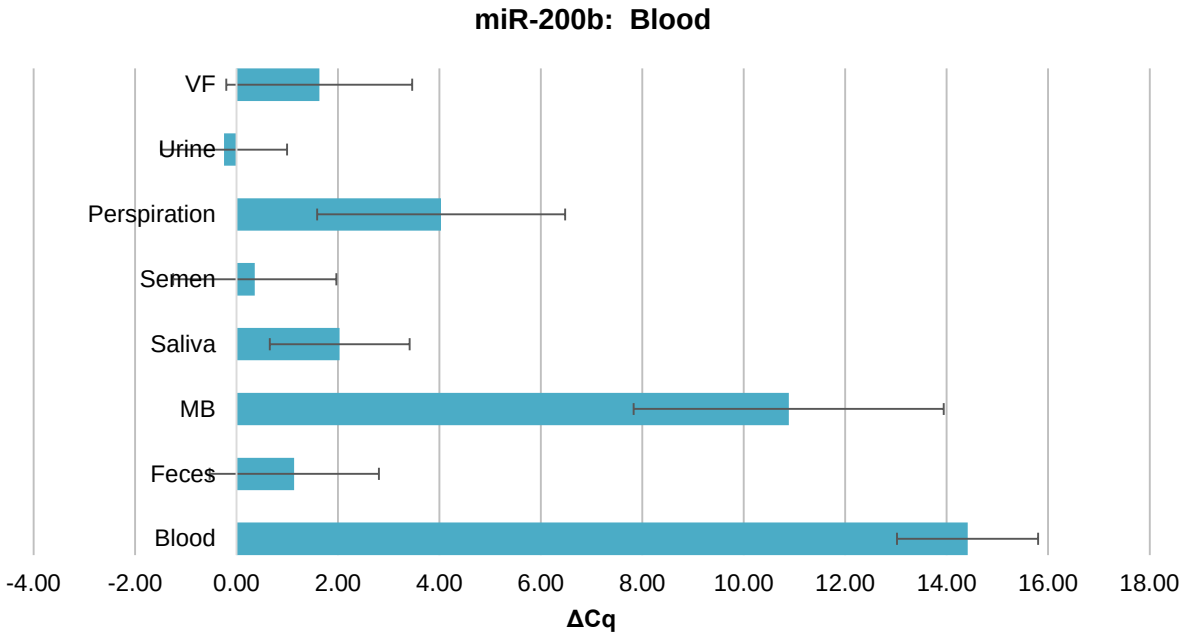


1

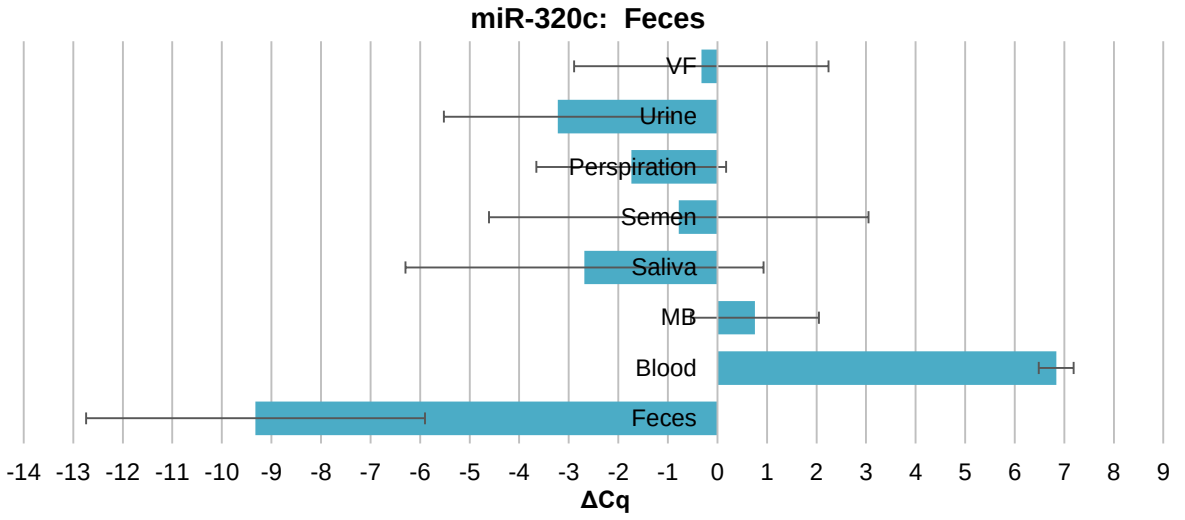


2

3 **Figure 9: Relative Expression of miRNAs shows specificity for the target biological fluid.** RT-
4 qPCR analysis: n=20 of target body fluid, and n=5 population samples for the other seven body
5 fluids assessed. **Top:** miR-891a for semen **Bottom:** miR-1246 for menstrual secretions



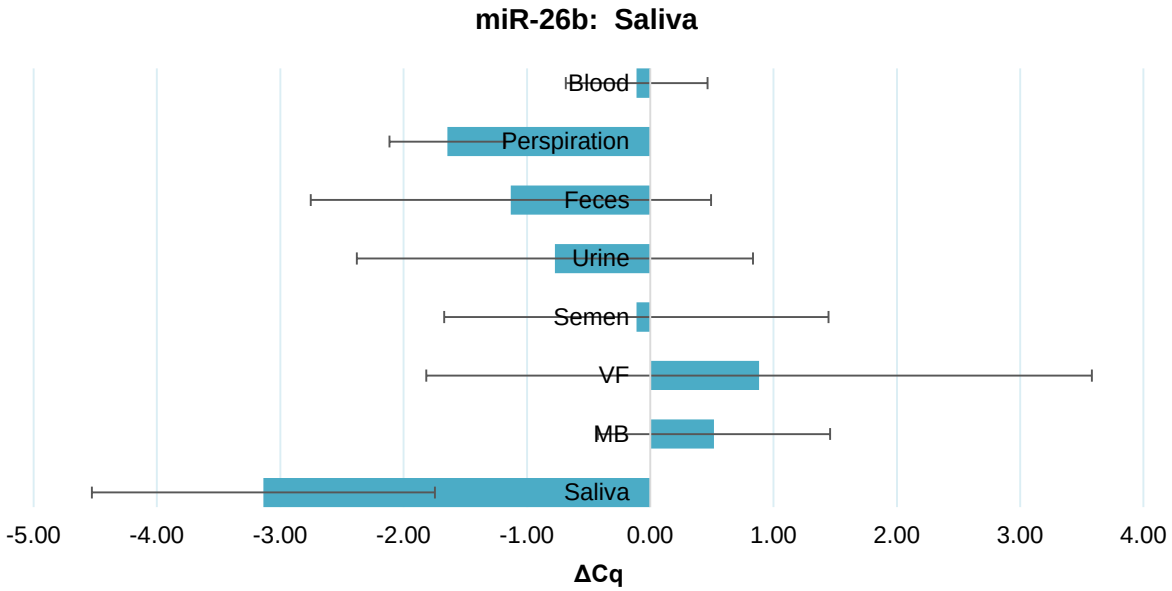
1



2

3 **Figure 10: Relative Expression of miRNAs shows specificity for the target biological fluid.** RT-
4 qPCR analysis: n=20 of target body fluid and n=5 population samples for the other seven body
5 fluids assessed. **Top:** miR-200b for blood **Bottom:** miR-320c for feces.

6



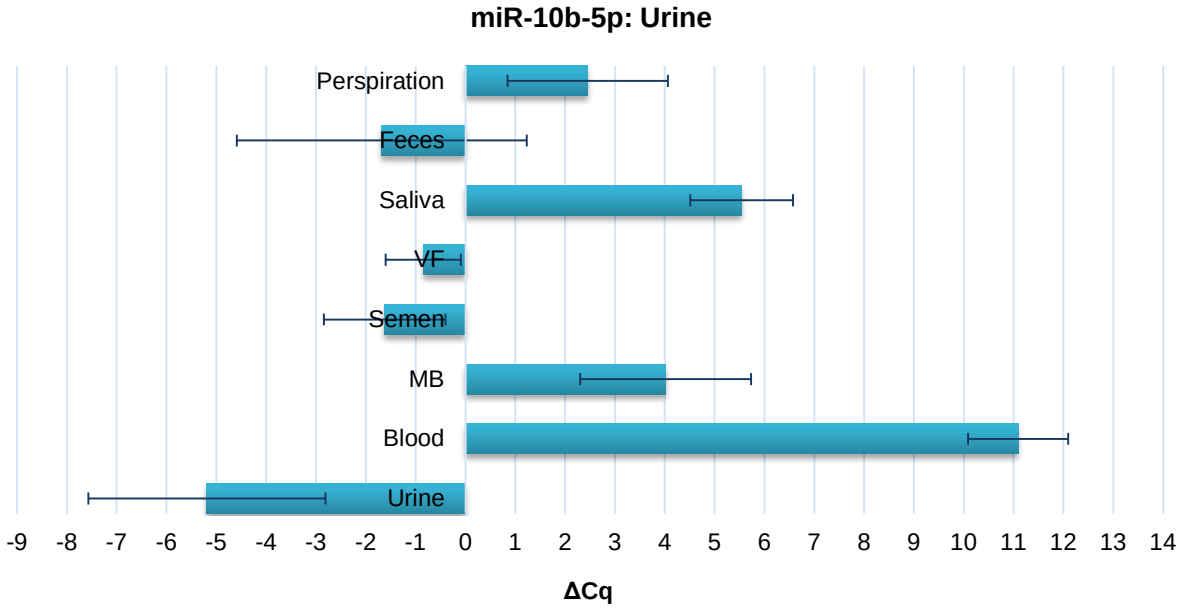


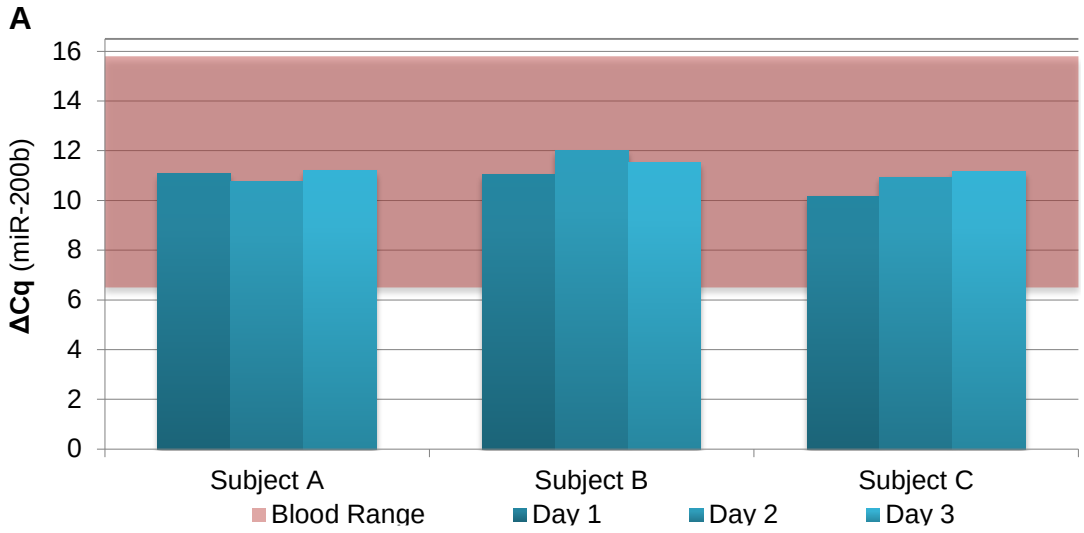
Figure 11: Relative Expression of miRNAs shows specificity for one or more biological fluids. RT-qPCR analysis: n=20 of target body fluid, and n=5 population samples for the other seven body fluids assessed. **Top:** miR-26b was evaluated for semen based on previous data, but shows expression differentiation for saliva. **Bottom:** miR-10b-5p was originally evaluated for feces, but found to distinguish urine/feces from other six fluids, and blood from the other seven fluids.

miRNA expression over biological time

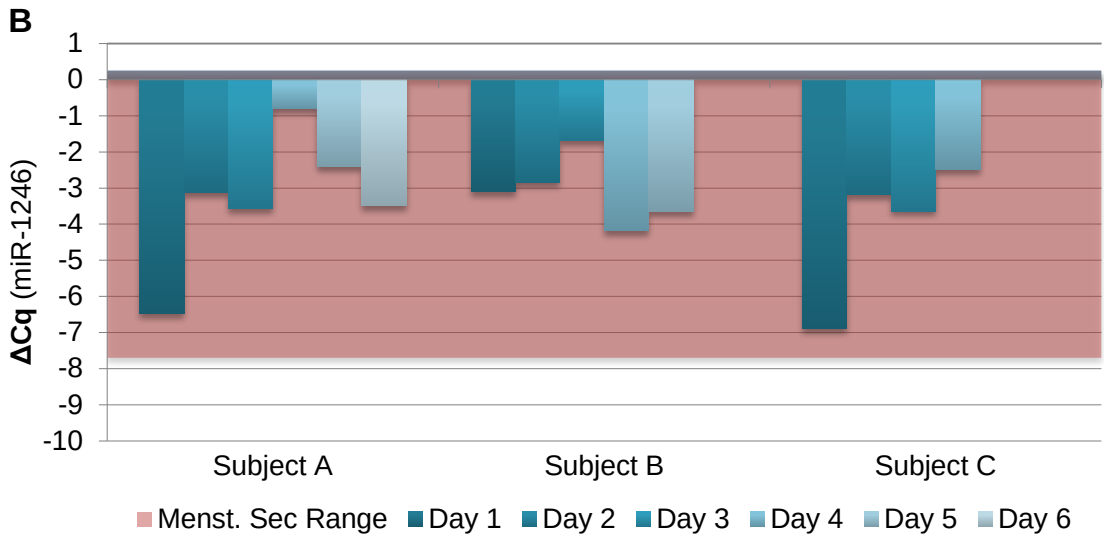
The final candidate miRNAs were then evaluated using sample sets of each biological fluid corresponding to cycles or extended biological time within a given person. Three donors donated samples corresponding to time periods relevant to each biological fluid (Table 10), and RNA was isolated and analyzed for differential expression in the target miRNAs as previously described. As seen in Figure 12(A-F), all miRNAs showed variations in differential expression, but all within the stastically significant ΔCq ranges determined for each biological fluid.

Table 10: Biological time samples for testing

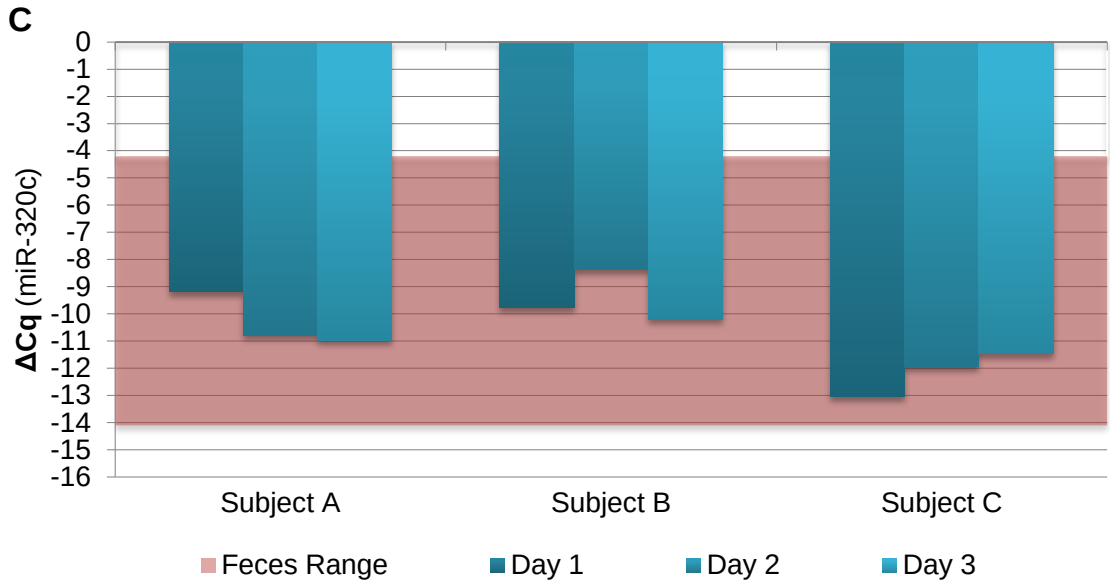
Body Fluid	Samples collected
Blood	3 donations within a 7-day period
Saliva	3 donations/day for 3 days: upon waking, before a meal, after a meal
Semen	3 donations within a 30-day period (>3 days postcoital activity)
Vaginal Secretions	21 donations (average time between menstrual cycles)
Menstrual Secretions	3-7 day donations (dependent on donor menstrual cycle)
Urine	6 donations over a 3-day period: upon waking and afternoon
Feces	3 donations within a 7-day period
Perspiration	6 donations: 3 each active (exercise) and passive perspiration



1

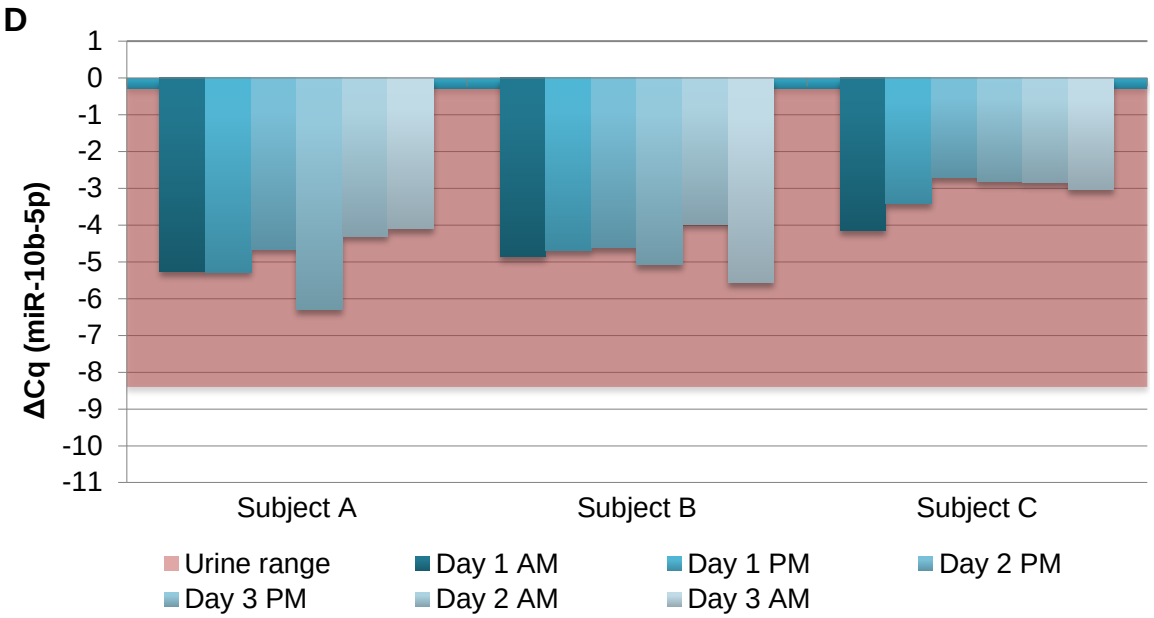


2

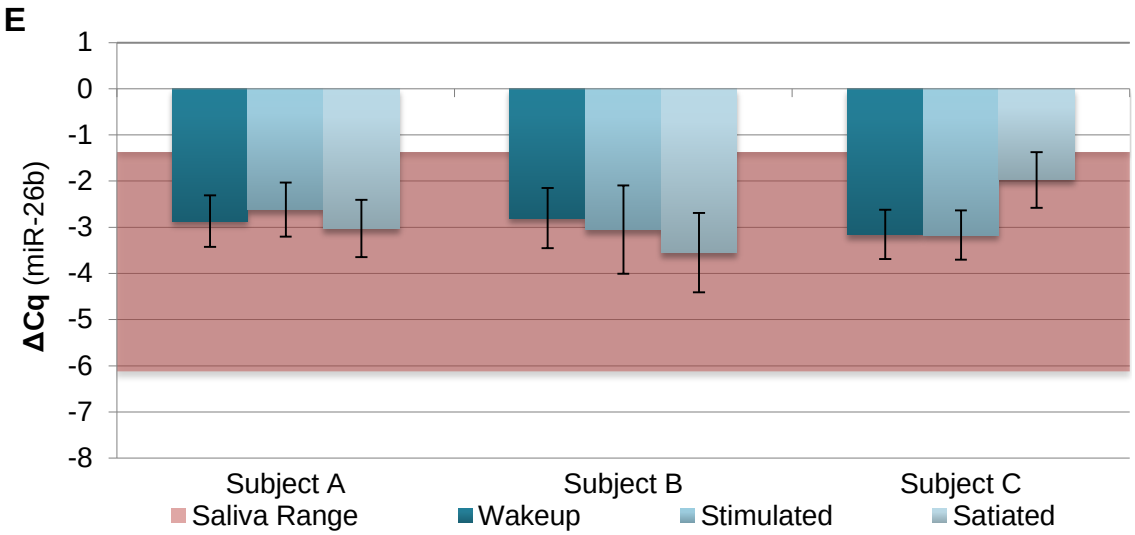


3

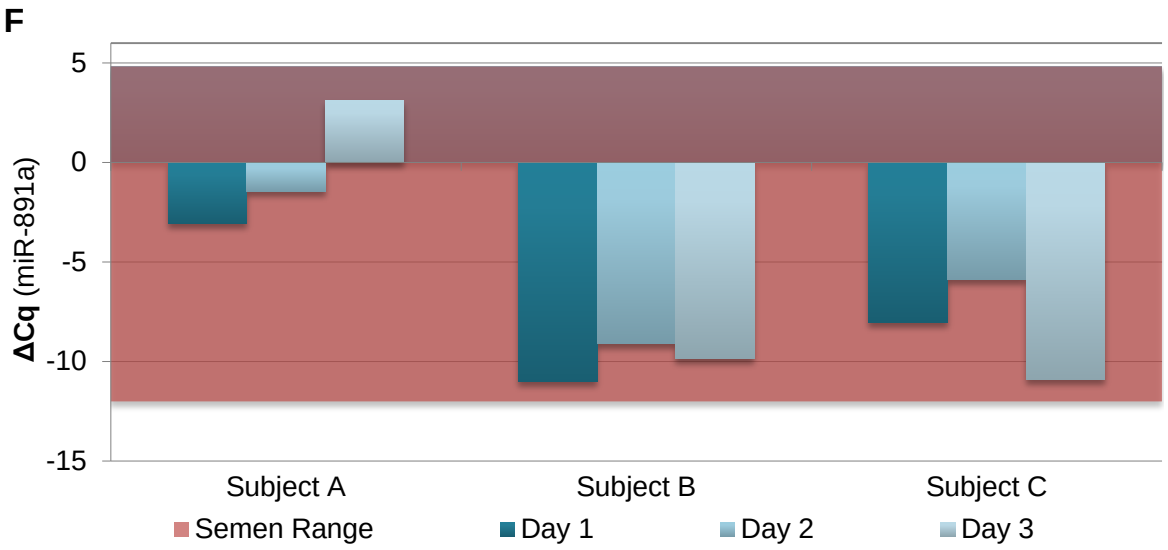
4



1



2



3

Figure 12: Variation within donors over biological time remains within the range identified in the population. miRs 200b (A: blood), 1246 (B: menstrual secretions), 320c (C: feces), 10b-5p (D: urine), 26b (E: saliva), and 891a (F: semen) were evaluated against samples taken over biological time appropriate for each biological fluid. Differential expression (ΔCq) of each miRNA was evaluated in their respective target body fluid(s) as previously described across 3 donors. Each sample was analyzed in triplicate. Data is the average of the technical triplicate wells, with the exception of E:, which is the average ΔCq of each sampling type across 3 days.

Human Organ Panel Analysis

In order to evaluate miRNA specificity in a variety of human organs and tissues, 10 µg of each sample from the FirstChoice® Human Total RNA Survey Panel (Ambion, Life Technologies, Inc.) was reverse transcribed and analyzed for relative expression as described above. This evaluation was strictly intended to be informative, for purposes of developmental validation of the candidate miRNAs (Table 11). As expected, the majority of the miRNAs assessed were expressed in these samples; however, the cell and RNA content is high and thus detection is not unexpected when compared to the low abundance of miRNAs comparatively in the biological fluids tested. Differential expression in the range of the target biological fluid was encountered more frequently for miRs-10b-5p, 26b, and 891a, markers for urine, saliva, and semen, respectively.

Table 11: Expression of candidate miRNAs in organ tissues. Yellow shading indicates differential expression within the range of the target biological fluid.

	miR-1246 ΔCq	miR-200B ΔCq	miR-320c ΔCq	miR-10b-5p ΔCq	miR-26b ΔCq	miR-891a ΔCq
Fetal Brain	-13.1	0.8	1.7	3.7	-2.9	1.4
Liver	-12.3	-4.7	5.0	-6.3	-11.4	14.0
Spinal Cord	-10.9	0.0	-2.4	-6.7	-7.5	8.6
Adrenal Gland	-13.1	1.0	0.8	-7.8	-7.1	1.9
Cerebellum	-9.6	5.4	-3.6	6.1	-0.7	6.3
Lung	-13.7	-8.0	-1.9	-5.4	-6.2	2.8
Kidney	-10.8	-9.4	8.7	-13.4	-10.9	5.1
Prostate	-5.3	0.6	14.7	4.8	3.1	9.4
Testis	-9.1	-3.6	4.4	-6.6	-6.5	-4.1
Colon	15.9	-8.9	-0.4	-8.1	-9.2	6.2
Placenta	-10.0	0.2	2.6	2.4	-2.2	5.6
Uterus	-8.6	2.4	-2.8	-1.7	-1.5	7.8
Skeletal Muscle	-13.9	0.1	10.5	-3.7	-6.3	0.8
Small Intestine	-12.2	-6.3	-0.8	-4.0	-4.6	4.0
Brain	-18.2	-9.6	-2.5	NT	-11.0	0.1
Spleen	-9.3	3.2	15.2	-2.7	-3.1	9.0
Thymus	-9.7	-4.3	2.4	-1.3	-2.2	0.7
Salivary Gland	-11.4	-7.4	2.3	-1.4	-4.8	5.7
Fetal Liver	-17.4	-11.1	-0.7	-10.3	-14.9	9.5
Bone Marrow	-10.9	-9.2	-5.3	-22.5	-24.4	-2.8

Species specificity

Blood, saliva, and fecal samples were collected from animals expected to have interactions with humans and thus likely to be involved with evidence (pets, livestock, local wildlife). Semen and menstrual secretions (where relevant) were not collected due to the unlikely nature of encountering them and the difficulty in acquiring them. This experiment was conducted using the same parameters used for all other miRNA evaluation analyses in duplicate technical replicates, and used the "human" miRNA primers evaluated above. Species specificity studies are required for purposes of developmental validation, it being important to understand the limitations of any new marker or method[67]. In particular, we intended this

study as a cautionary note for body fluid identification using miRNAs. miRNA conservation among all species, but particularly the higher eukaryotes, has been thoroughly documented[23,68-70]. Analysis of the candidate miRNAs revealed significant cross-reactivity and amplification success as expected, and relatively irrespective of species or biological fluid tested (Table 12). We continue to evaluate the saliva samples that were collected and isolated, and are pursuing obtaining canine heat secretions to compare to venous and human menstrual secretions.

Table 12: miRNA conservation results in amplification of human candidate miRNAs in other species

Sample Name	let-7g	let-7i	miR-200b	miR-320c
Bobcat blood	+	+	+	+
Red fox blood	+	+	+	+
Sheep blood	+	+	+	+
Pig blood	+	+	+	+
Cat blood	+	+	+	+
Dog blood	+	+	+	+
Goose blood	+	+	+	+
Chicken feces	+	+	+	+
Goat feces	+	+	+	+
Pig feces	+	+	+	+
Ferret feces	Neg	Neg	+	+
Alligator feces	+	+	+	+
Bearded Dragon feces	+	+	+	+
Bison feces	+	+	+	+
Dog feces	+	+	+	+
Otter feces	+	+	+	+
Horse feces	+	+/Neg	+	+

+: detected Neg: undetected at 40 cycles amplification +/Neg: duplicate wells at detection limit - one well >39, second well undetected

Limit of Detection Studies

Our optimized method using synthetic miRNAs to develop a standard curve (described above) allows for absolute quantitation of miRNAs in all samples. This provides a precise

measure of the limit of detection, rather than a range of total RNA concentration, which is in no way indicative of miRNA quantity[71]. This set of standards also serves to improve the comparison between miRNA levels in different body fluid samples and thus more precisely identify body fluids. We quantitated samples of each biological fluid using the Bioanalyzer 2100 (Agilent) small RNA method, which targets the low quantity miRNAs in a sample specifically. We then performed RT-qPCR analysis using reverse transcription reactions of fixed miRNA quantities (variable for the range detected in each biological fluid) of 4-fold serial dilutions alongside the synthetic standard curve. We found that limit of detection was correlated with RNA quantity, but only within a given body fluid. The absolute lower limit of detection was detected for multiple miRNAs as approximately $1x10^4$ copies/µL of RNA extract (corresponding to a Cq of approximately 38), and stochastic variation reduced confidence below that range.

We found that the concentration of RNA that achieved that lower limit varies by miRNA expression as well as body fluid type, and as would be hoped, let-7g and i expression tightly correlates with copy number for each RNA quantity range (individual to the body fluid). Again, regarding the body fluid type, even miRNA quantity as measured by a Bioanalyzer trace is not exclusively human; bacterial small RNAs within the size range (18-25 nt) would be expected in most samples (barring blood and perhaps semen), thus overestimating sample concentration (Table 13).

Table 13: Approximate LOD of each biological fluid

Biological fluid	LOD (ng RNA)
Blood	0.05
Feces	2.0
Saliva	<2.0*
Semen	<2.0*
Perspiration	0.10
Urine	0.10
Vaginal Secretions	<30*
Menstrual Secretions	<2000*

**: detection limit not reached*

Results: miRNA stability in compromised samples

An additional area of inquiry that we were interested in centered around the stability of miRNAs when compromised, as would be seen in a forensic sample. miRNAs have been successfully recovered from highly compromised samples including formalin-fixed paraffin embedded (FFPE) tissue[72]. In serum, miRNAs have been shown to survive harsh conditions such as boiling, low or high pH, cycles of freeze-thaw, and extended storage[50]. In the clinical arena, miRNAs have been shown in urine, and potential markers for bladder cancer have been evaluated[73,74]. A recent study evaluated liquid semen, kept at room temperature for up to seven days, and frozen and thawed up to eight times, and found minimal differences in levels of three miRNAs evaluated[73]. Urine stability was evaluated in up to four consecutive freeze-thaw cycles, and no significant difference was found in miR-1 and miR-16 levels[75]. There is also considerable evidence that when secreted from the cell, the majority of miRNAs are encapsulated in an particle, which, depending on the miRNA and the secretion process, could be membrane-, HDL, or protein-based[76,77]. Depending on the type of encapsulation, the miRNAs can be RNase or detergent resistant and the encapsulated miRNAs still detectable after treatment[50]. miRNAs have not been methodically treated and tested for stability, beyond a small aged sample analysis performed by Courts et al[19].

RNA Isolations

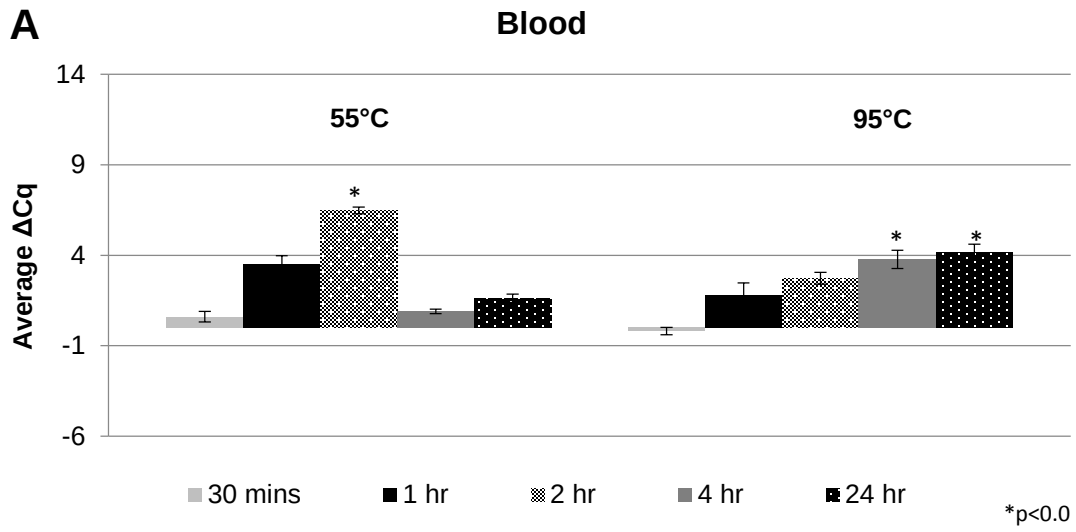
Total RNA yield as measured by the Nanodrop UV spectrophotometer was variable, but with no significant differences between untreated controls and treated samples (Supplemental Fig 1). All treatments resulted in high total RNA yields (>100 ng). Even though UV spectrophotometric data of RNA samples is not a true indication of RNA quality, and particularly of miRNA quantity, these results are consistent with a previous report by Setzer et al[78], in which mRNA stability in environmentally challenged samples was evaluated.

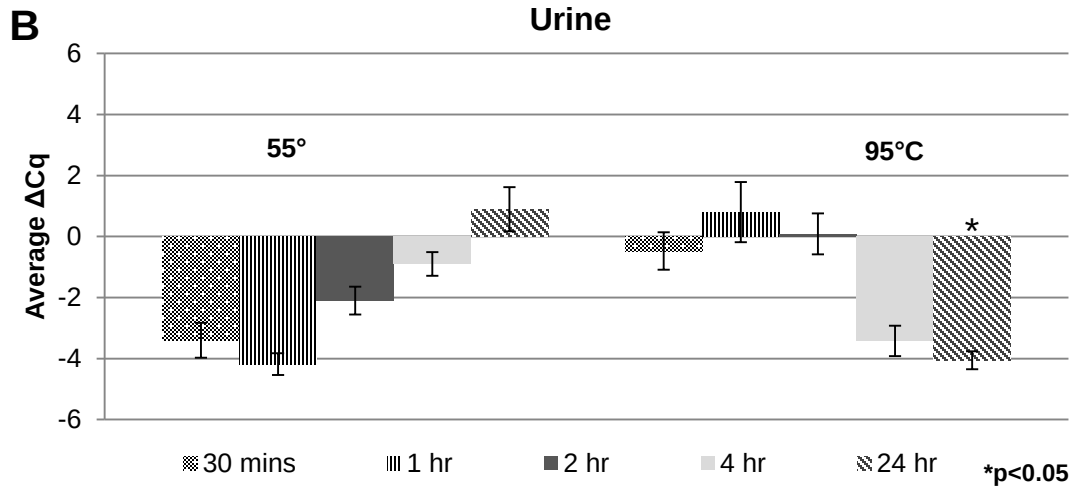
Heat Treatment

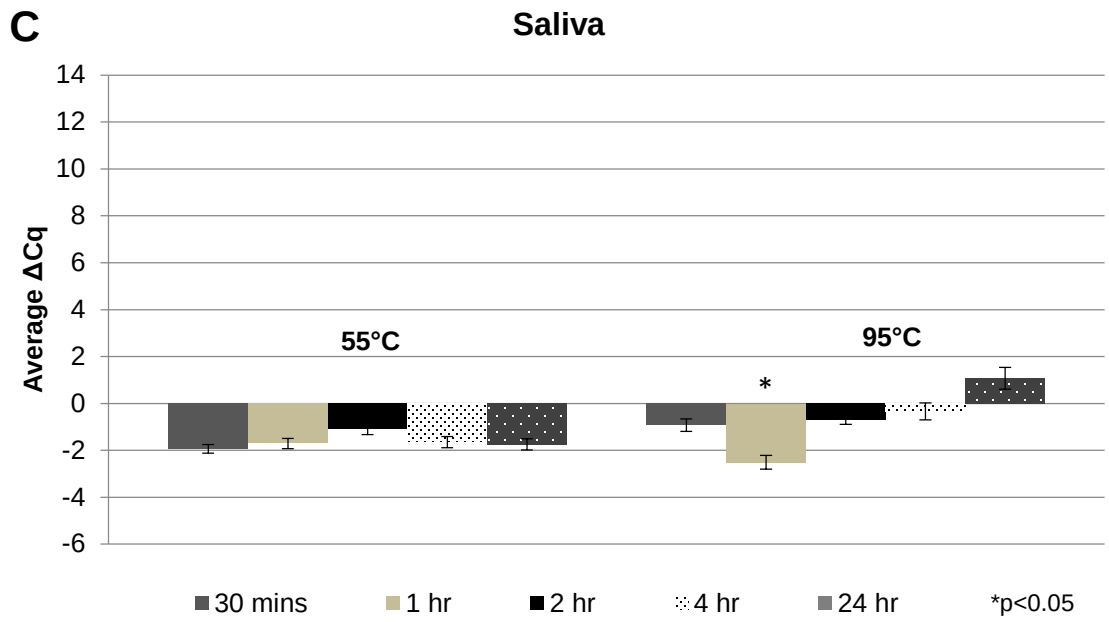
Blood, treated for any length of time at both 55 and 95°C, was highly resistant to degradation, with no failed reactions (Fig 13A). Let-7g levels are high in blood (untreated controls Avg Cq: 15.2 ± 1.1), which resulted in low C_T values regardless of treatment. Urine, as a high-volume, low cell-content fluid, understandably has low levels of miRNAs to begin with

(untreated controls Avg Cq: 34.1 ± 1.9), and any degradation can quickly result in loss of signal. Consequently, sample failure rates were expected, but were only observed in one replicate well at 55°C for 24 hours. What was unexpected, however, was improved detection of let-7g in urine, significantly so when treated for 24 hours at 95°C (Fig 13B). As these results were unexpected, several biological and technical replicates were repeated and results corresponded with initial observations. miR-16 levels of a fourth donor under heat treatment showed similar patterns in urine and saliva over time at 55°C, with significantly elevated levels after heat treatment (Supplemental Table 2).

Saliva was remarkably stable under heat treatment at 55°C and 95°C, and always detectable (untreated controls Avg Cq: 28.5 ± .25) (Fig 13C). Let-7g levels in semen showed the most typical degradation pattern at 95°C, but only one time point was statistically significant (Fig 13D).







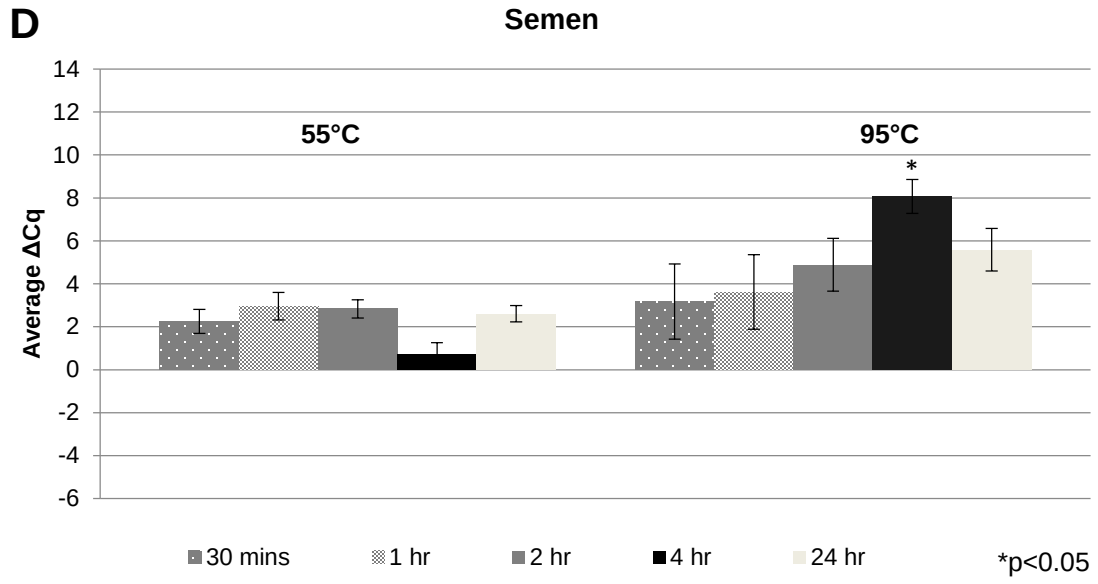


Figure 13: miRNA stability under heat conditions over time. Relative let-7g levels as determined by the difference in Cq from treated sample to paired untreated control (n=3 donors, data is the average Cq of the technical triplicate wells for paired treated and untreated samples). **A**: Blood heat treatment. Let-7g levels in blood were significantly depleted after 2 hrs at 55°C and 4 and 24 hrs at 95°C ($p<.05$). **B:** In urine, detectable let-7g levels were improved for many timepoints, significantly so for 24 hrs at 95°C. **C:** Heat treatment did not dramatically impact saliva miRNA detection levels, but **D.** semen was markedly impacted by 95°C treatment over time.

Ultraviolet Treatment

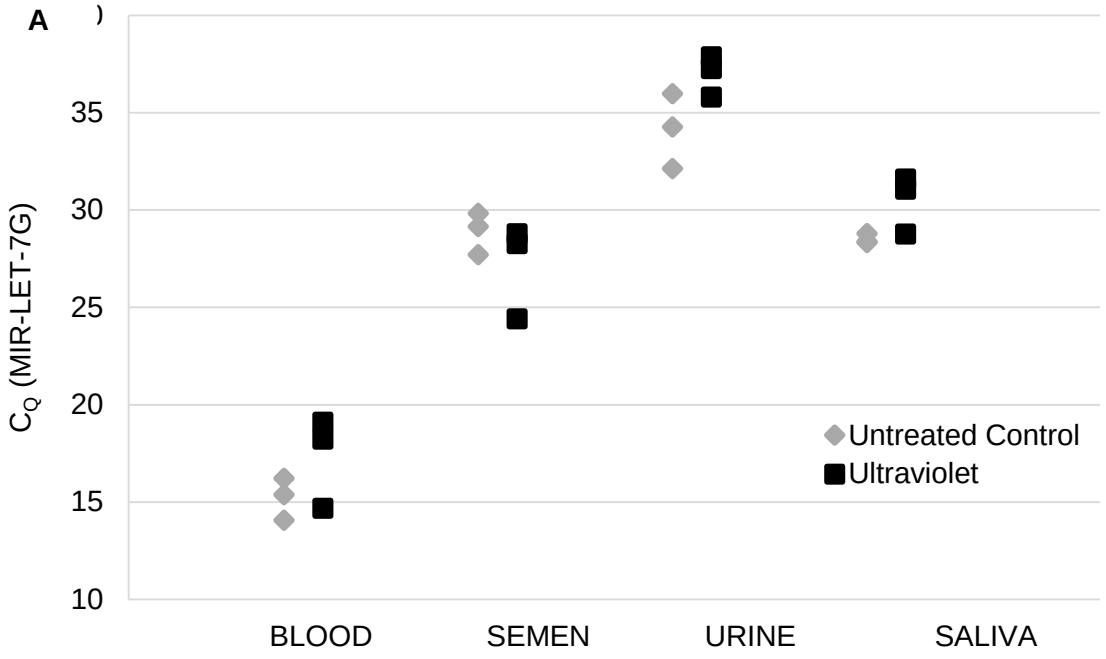
All body fluids tested were highly resistant to ultraviolet treatment, yielding detectable miRNA levels with no significant reduction in detection (Fig 14A). While there is very little research in the area of UV damage to RNAs, and none for small RNAs, the primary mode of action for ultraviolet damage is the fusion of pyrimidine doublets, and ultraviolet exposure has been shown to impact RNA in pure extracted form[79]. However, this is not relevant to the matrix of a dried body fluid sample; ultraviolet light has been shown to affect mRNA levels only after 90 and 180 days of exposure in mock forensic samples[78]. Certainly UV transillumination is not a perfect comparison to normal daylight conditions; outside environmental studies are the only real method to assess UV degradation; however, they come with their own set of difficult variables to quantify (i.e. humidity, precipitation, temperature, wind speed, etc.).

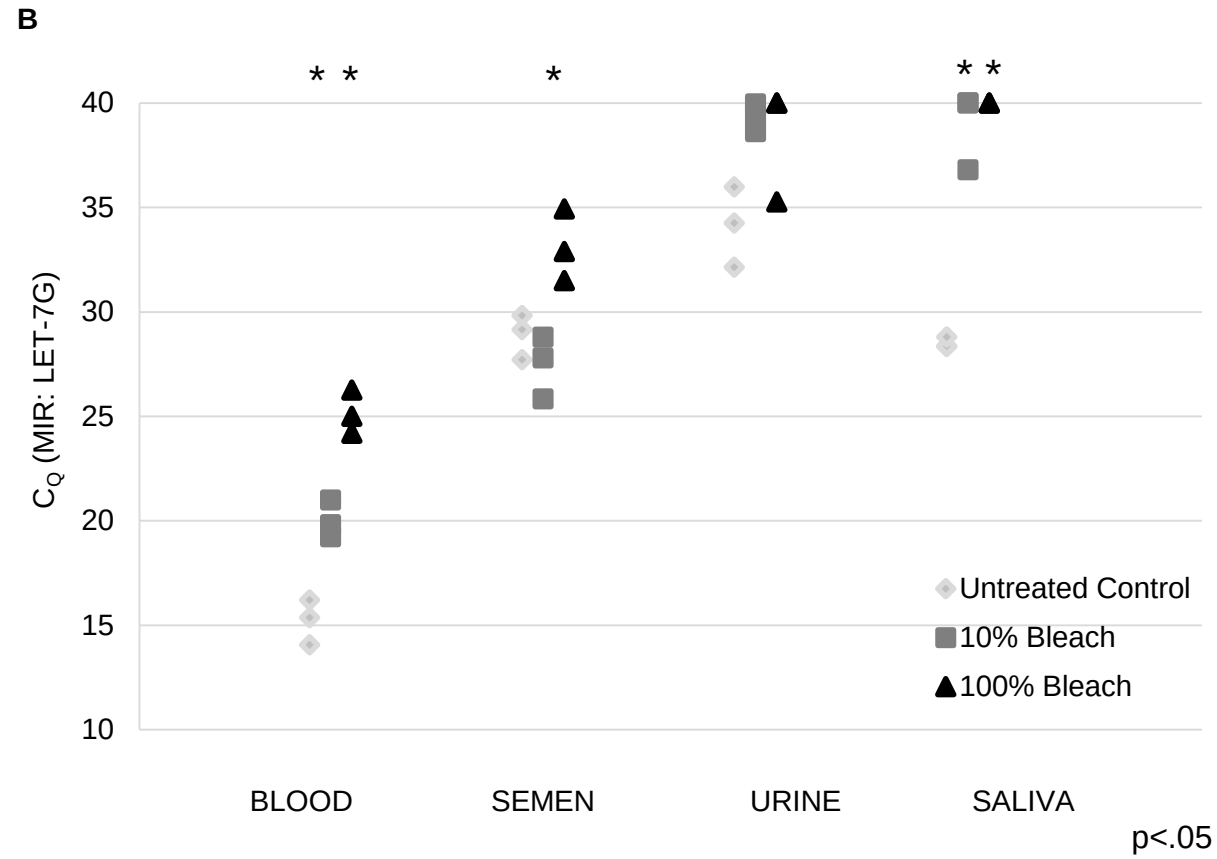
Chemical Treatment

Bleach treatment resulted in two widely variant results. While blood and semen were consistently detectable after diluted and household strength bleach application, semen and saliva proved vulnerable, resulting in >50% technical replicate failures at 10%, and 100% failure in saliva at full-strength bleach application (Fig 14B). This is enigmatic, as application with glacial acetic acid (pH ~2.4) yielded minimal difference from the control for all body fluids, with semen yielding significantly better detection levels across all three donors (Fig 14C). Given the fact that RNA is more stable under slightly acidic conditions than DNA, it would be interesting to evaluate samples that had been exposed to similar conditions for DNA quantity and STR profile analysis.

Application of dish detergent was performed for multiple reasons. First, it is a canonical treatment variable that should be evaluated for forensic samples, given the fact that crime scene cleanup by perpetrators are typically attempted using common and convenient household chemicals. Secondly, it stands to reason that because some secreted miRNAs are encapsulated within a microvesicle, which contains large amounts of cell membrane lipid components, or co-transported with HDL particles[50,77,80,81], stability could be affected by disrupting such particles and exposing the miRNAs. Interestingly, only urine detection was not affected by detergent application, with blood, semen, and saliva showing significantly reduced

detection levels, with sample amplification failures in semen and saliva.





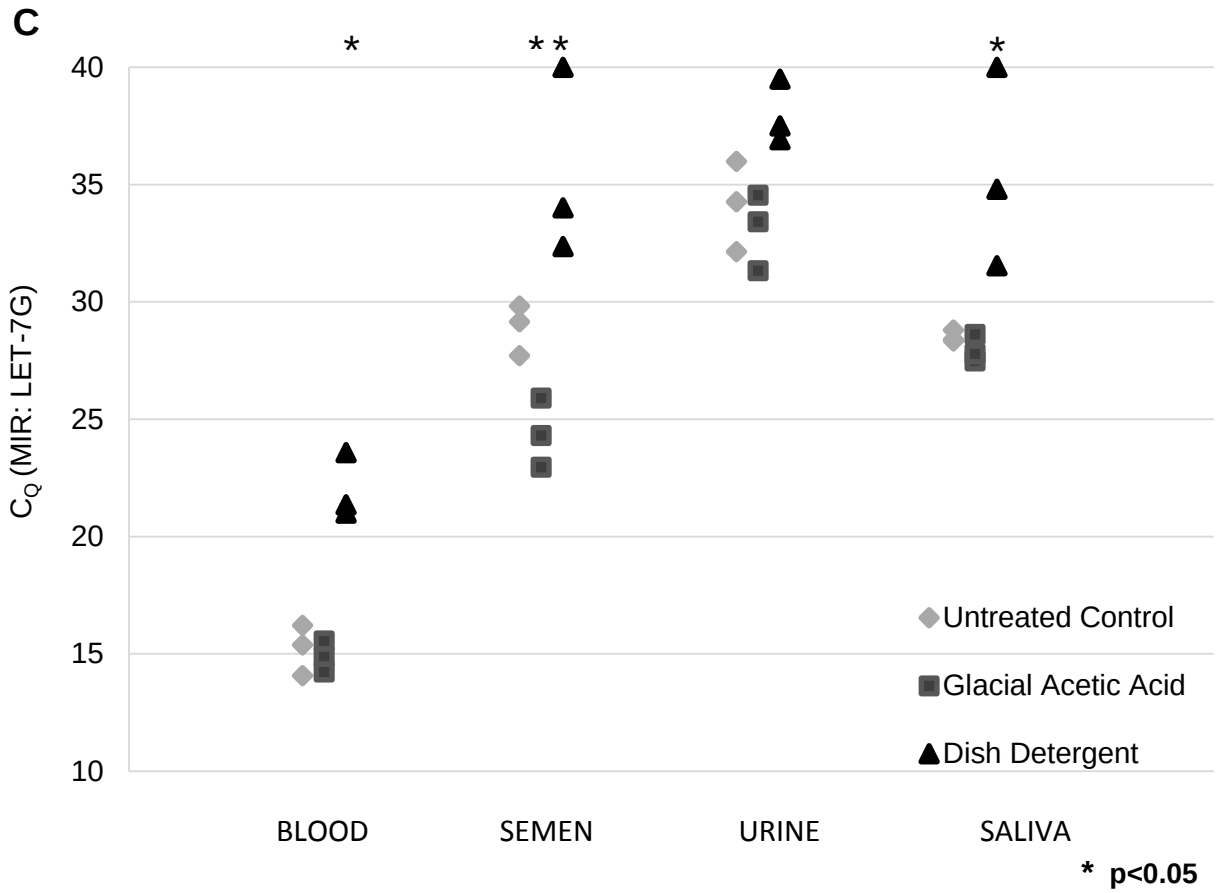


Figure 14: miRNA stability under various conditions. Let-7g levels in treated and untreated controls (n=3 donors, data is the average Cq of the technical triplicate wells). **A**: Ultraviolet exposure showed minimal impacts on let-7g levels, but **B:** 1:10 sodium hypochlorite significantly reduced let-7g levels in blood and saliva. Application of full-strength household sodium hypochlorite significantly reduced levels in blood, semen, and saliva. **C:** Blood, semen, and saliva were susceptible to dish detergent treatment, but only semen was susceptible to exposure to acetic acid.

Stability across miRNAs

In addition to let-7g, three additional candidate miRNAs (miR-16, 21, and 24) were evaluated in treated and untreated samples. These miRNAs are commonly detected in tissues and body fluids, and found to be expressed at relatively high levels. This additional analysis was performed in order to show that the reported stability and degradation patterns of let-7g are representative of the given population of miRNAs in a sample. An evaluation of the average difference in Cq from treated sample to untreated control across 12 treatments in blood, urine,

semen and saliva showed that degradation of the four tested miRNAs correlated closely with the average of the four tested miRNAs, with no dramatic differences in degradation levels (Figure 15, Supplemental Table 2) These data indicate that degradation observed in this study is likely to be representative of the integrity levels of the miRNA population in the body fluids assessed.

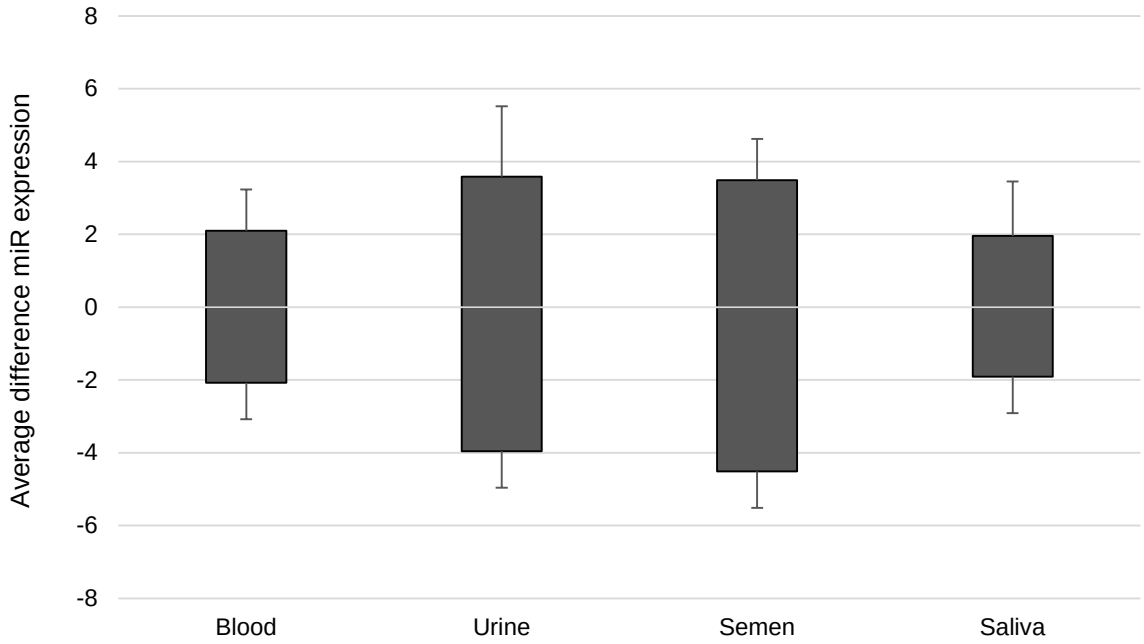


Figure 15: Sample degradation does not result in wide variations in detection levels among tested miRNAs. miRs 16, 21, 24, and let-7g levels were evaluated in 12 treatment groups from blood, urine, semen and saliva. Data shown is the minimum and maximum differences between the Average dCqs (difference in Cq of treated vs untreated control across 2-4 miRNAs) of 9-12 treatments per biological fluid. Error bars show standard deviation of the average minimum and maximum differences. See Supplemental Table 2 for data details.

Overall, this study sought to characterize the stability of miRNAs for forensic samples. Given the explosion in recent interest in miRNAs, it is important to understand the limitations to the forensic analysis of any species of nucleic acid under consideration for implementation into casework. Overall, the miRNAs assayed, as an indicator of the species as a whole, were remarkably stable, with detection in the majority of treatments (Table 14).

Table 14: Successful let-7g detection rates in body fluids after compromising treatments.

	UV	10% Bleach	100% Bleach	Dish Detergent	Glacial Acetic Acid
Blood	100%	100%	100%	100%	100%
Urine	77.8%	33.3%	55.6%	66.7%	100%
Semen	100%	100%	100%	66.7%	100%
Saliva	100%	44.4%	0%	77.8%	100%

IV. Conclusions

High-throughput sequencing for the identification of body-fluid specific miRNAs

In conclusion, we have fully addressed the proposed Aims of the funded project, and have developed a panel of 8 miRNAs that can be used to differentiate the six major biological fluids commonly seen in forensic evidene, and provide some differential information regarding vaginal secretions and perspiration.

To achieve this goal, we first optimized RNA isolation conditions for maximum amplifiable miRNA quantity, finding that all tested methods were suitable for all biological fluids, with the exception of feces, which requires a specialized RNA isolation method. We then subjected 33 RNA samples to high-throughput sequencing, resulting in successful high-throughput sequencing of miRNA from forensic samples. This in itself is a compelling finding, as we specifically chose to extract RNA and take forward samples consistent with the small volume found in evidence samples, with the exception of urine. High-throughput sequencing for miRNA expression characterization has, to our knowledge, never been performed on sample sizes this small, and the success of the analysis warrants publication not only for application to forensic HTS analysis of small volume evidentiary samples in the future, but also clinical sampling of small volumes of samples. Our choice to use small-volume samples was also a limitation of the study; if we were to have utilized larger sample sizes, greater depth of sequence may have resulted.

The high-throughput sequencing was markedly compromised by the bacterial small RNAs in many biological fluids, resulting in competition for reverse transcription and

sequencing reagents. This did not dramatically impact the number of sequencing reads collected from the instrument, but rather the percent of sequencing reads that were annotated as human miRNAs and consequently the number of miRNAs that could be evaluated was very limited for some of these fluids, particularly feces and perspiration.

We first combed the sequence data for miRNAs observed in all 33 samples, regardless of biological fluid. We identified four miRNAs – let-7g, let-7i, miR-451a, and miR-21, as expressed in all samples. RT-qPCR evaluation and correlation to RNA quantity eliminated miRs-451 and 21, but lets-7g and i showed high correlation to measured miRNA quantity. This resulted in the development of a normalization method that can be used for all eight biological fluids using the average expression of both lets-7g and i. This method is a remarkable advance, as it allows for elimination of RNA quantity evaluation prior to reverse transcription, thus streamlining analysis.

Once our normalization method was validated, evaluation of the HTS data resulted in a list of potentially body fluid-specific miRNAs that appeared to be either exclusively or differentially expressed in a particular body fluid. The candidate miRNA findings from the HTS data were not supported by following RT-qPCR analysis, which showed that sensitivity of the HTS platform was markedly lacking and expression of the candidate miRNAs was neither in proportion to that expected, nor reduced in other biological fluids. Regardless of these disappointing findings, we were able to identify 7 miRNAs for initial validation of body fluid specificity. Additional population studies resulted in 6 of those miRNAs successfully showing differential expression ($p<.05$) for one or more biological fluids (Figure 16). The decision matrix for biological fluid identification using the differential expression method is in some places convoluted, as in two cases, an additional marker is necessary to differentiate the two fluids (Figure 17). miR-200b distinguishes blood and menstrual blood from all other body fluids, but not from each other. However, miR-1246 differentiates blood and menstrual secretions, as do the rest of the miRNAs. Likewise, miR-10b-5p distinguishes urine and feces from the rest of the body fluids, and miR-320c differentiates feces from urine. Thus, the identification of urine is reductive. Perspiration *may* be differentiated from the other body fluids using miR-200b, but the sufficient population samples have not been tested to evaluate this possibility. We were not able to identify a candidate miRNA for vaginal secretions, but further analysis of candidate miRNAs could likely develop a potential marker.

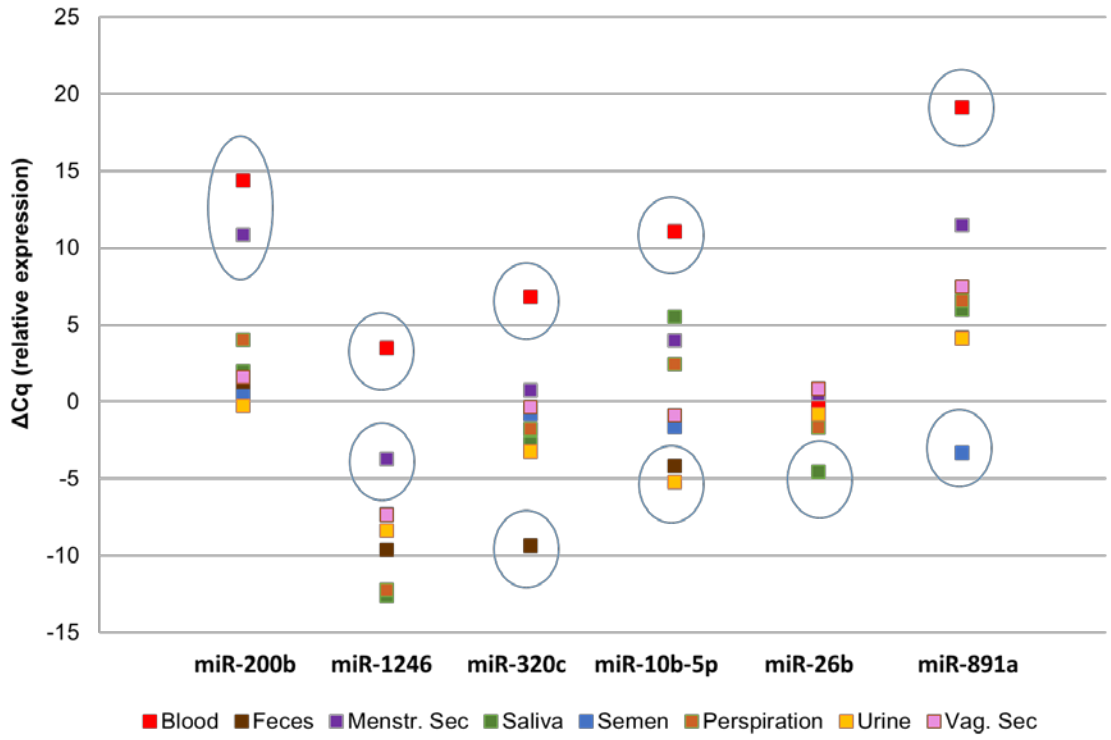


Figure 16: A miRNA panel for forensic body fluid identification using differential expression. Blood, feces, menstrual secretions, saliva, semen, and urine can be differentiated based on expression patterns of these six miRNAs, normalized with let-7g and i as described. Circled body fluids are differentiated in expression (n=20 tested population sample for each circled fluid, n=5-20 for the other 5-6 fluids tested against that miRNA. $p<0.05$).

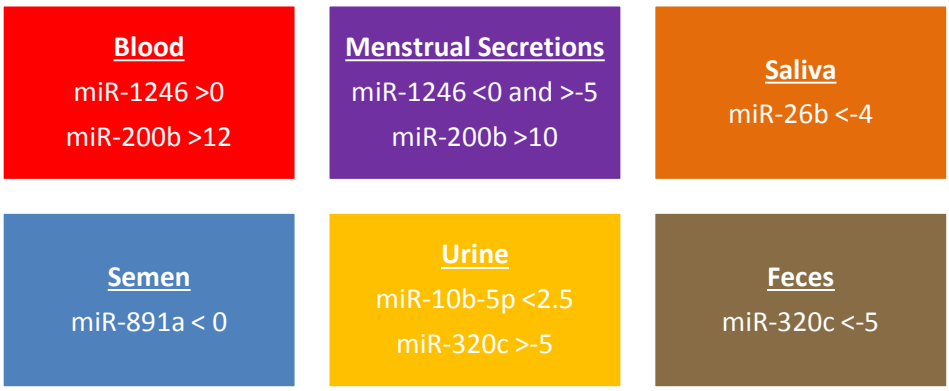


Figure 17: Body fluid decision matrix using differential expression. Each of the candidate miRNAs show a range of expression in the population, and samples falling within that range are identified as originating from that biological fluid type.

miRNA Stability in compromised samples

The miRNAs present in blood were both more abundant and robust than those of the other biological fluids, and were detectable throughout all treatments. Even though degradation in the form of reduced amplification was observed, sufficient miRNA levels remained for positive analysis. Semen was more susceptible to treatment, resulting in amplification failures when treated with dish detergent, but was otherwise largely detectable in the majority of treatment scenarios. The better detectability of blood and semen are likely reflections of the high quantities of total RNA isolated from those samples, but if that were the only impacting factor, saliva would also have been more successful (Supplemental Fig 1). It is possible that the high protein content of blood and semen offers some protection. As could be expected, semen miRNA levels were not negatively impacted by glacial acetic acid treatment; given that semen contains the necessary buffering capacity and basic pH that has co-evolved to neutralize and survive in the acidic vaginal vault[82,83]. It is important to note that the lower the expression level, the more likely that detection of the miRNA will fail upon sample compromise. Fortunately, miRNA degradation does appear to occur consistently across multiple miRNAs, and thus normalization to endogenous reference miRNAs can be confidently utilized for differential expression analyses, as the majority of the work in forensic miRNA expression utilize a normalization methodology[19,24,26,28,39,84,85] . Given that the miRNAs tested were readily and consistently detectable in samples consistent with forensic evidence, it is highly possible that stability is even better than seen in these results, once a final panel of both abundant and source-indicative miRNAs are identified for each body fluid.

Implications for policy and practice

The comprehensive evaluation of miRNA expression in forensically relevant biological fluids, and consequent development of candidate miRNAs for further research that we completed in this body of work was a vital first step in proceeding towards an eventual commercial assay for body fluid ID that is robust and reliable in the hands of practitioners. Our findings regarding the stability of miRNAs upon compromise further enhanced the desirability of the development of a body fluid identification platform using miRNAs. Given the known existence of amplifiable miRNAs in DNA extracts[41] (also unpublished work in our laboratory), it is entirely possible that body fluid identification using the miRNA panel described in this project

could be performed using DNA extracts from forensic samples, which would significantly reduce the resistance for a novel body fluid identification method that requires second or additional RNA isolation procedures. The compilation of miRNAs in a robust multiplex developed from DNA extracts could rapidly revolutionize forensic body fluid identification, resulting in quantifiable confidence in the body fluid or fluids present in the sample. The resultant increased efficiency will reduce or eliminate analyst time on microscopic analysis and serological assays that yield poor confidence in the body fluid reported (excepting microscopic identification of sperm cells). Given that serological testing consumes a large proportion of analyst time, implementation of a multiplexed miRNA-based body fluid identification panel could result in a modest reduction of backlogs, and this assay could very quickly and easily be integrated into the forensic workflow, utilizing the instrumentation already present.

Implications for further research

The logical next step for this research is to complete additional developmental validation of the miRNA panel. This developmental validation should include first of all an expansion on the population study, including additional ethnicities and a more expanded set of ages for each sample type, as well as an analysis of mixed body fluid samples. The mixed body fluid sample question is of particular importance, particularly for sexual assault cases. After successful evaluation of mixed and expanded population samples, a multiplex method should be developed and optimized, along with an analysis decision workflow. To confirm performance as observed in single miRNA analysis, the multiplexed panel should be evaluated for efficiency and performance using population, mixed, low quantity, and compromised samples, along with concordance and reproducibility studies, as set forth in the Developmental Validation guidelines for new methodologies (Federal Bureau of Investigation (FBI) Quality Assurance Standards)[86]. We have not yet identified if the ideal platform for the miRNA multiplex is a qPCR method, but the panel once described and characterized could be easily adapted for the microarray, capillary electrophoresis or even the HTS platform alongside a panel of STR or phenotypic SNP markers.

V. Literature cited

1. Antunes J, Silva DS, Balamurugan K, Duncan G, Alho CS, McCord B. High-resolution melt analysis of DNA methylation to discriminate semen in biological stains. *Anal Biochem*. 2016;494:40-45.

2. Bartling CM, Hester ME, Bartz J, Heizer E,Jr, Faith SA. Next-generation sequencing approach to epigenetic-based tissue source attribution. *Electrophoresis*. 2014;35(21-22):3096-3101.

3. Lee HY, An JH, Jung SE, et al. Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. *Forensic Sci Int Genet*. 2015;17:17-24.

4. Legg KM, Powell R, Reisdorph N, Reisdorph R, Danielson PB. Discovery of highly specific protein markers for the identification of biological stains. *Electrophoresis*. 2014;35(21-22):3069-3078.

5. Van Steendam K, De Ceuleneer M, Dhaenens M, Van Hoofstat D, Deforce D. Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science. *Int J Legal Med*. 2013;127(2):287-298.

6. Yang H, Zhou B, Prinz M, Siegel D. Proteomic analysis of menstrual blood. *Mol Cell Proteomics*. 2012;11(10):1024-1035.

7. Weber JA, Baxter DH, Zhang S, et al. The microRNA spectrum in 12 body fluids. *Clin Chem*. 2010;56(11):1733-1741.

8. Bauer M, Patzelt D. A method for simultaneous RNA and DNA isolation from dried blood and semen stains. *Forensic Sci Int*. 2003;136(1-3):76-78.

9. Nussbaumer C, Gharehbaghi-Schnell E, Korschineck I. Messenger RNA profiling: A novel method for body fluid identification by real-time PCR. *Forensic Sci Int*. 2006;157(2-3):181-186.

10. Juusola J, Ballantyne J. Messenger RNA profiling: A prototype method to supplant conventional methods for body fluid identification. *Forensic Sci Int*. 2003;135(2):85-96.

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

12. Alvarez M, Juusola J, Ballantyne J. An mRNA and DNA co-isolation method for forensic casework samples. *Anal Biochem*. 2004;335(2):289-298.

13. Haas C, Klesser B, Maake C, Bar W, Kratzer A. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Sci Int Genet*. 2009;3(2):80-88.

14. Bauer M, Patzelt D. Identification of menstrual blood by real time RT-PCR: Technical improvements and the practical value of negative test results. *Forensic Sci Int*. 2008;174(1):55-59.

15. Fleming RI, Harbison S. The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Sci Int Genet*. 2010;4(4):244-256.

16. Hanson EK, Ballantyne J. RNA profiling for the identification of the tissue origin of dried stains in forensic biology. *Forensic Sci Rev*. 2010;22(2):145-157.

17. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*. 2009;136(4):642-655.

18. Zhang L, Stokes N, Polak L, Fuchs E. Specific microRNAs are preferentially expressed by skin stem cells to balance self-renewal and early lineage commitment. *Cell Stem Cell*. 2011;8(3):294-308.

19. Courts C, Madea B. Specific micro-RNA signatures for the detection of saliva and blood in forensic body-fluid identification. *J Forensic Sci*. 2011;56(6):1464-1470.

20. Budd WT. *Combinatorial analysis of tumorigenic microRNAs driving prostate cancer.* [dissertation]. Virginia Commonwealth University; 2012.

21. Qavi AJ, Kindt JT, Bailey RC. Sizing up the future of microRNA analysis. *Anal Bioanal Chem*. 2010;398(6):2535-2549.

22. Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell*. 2008;30(4):460-471.

23. Koufaris C. Human and primate-specific microRNAs in cancer: Evolution, and significance in comparison with more distantly-related research models: The great potential of evolutionary young microRNA in cancer research. *Bioessays*. 2016.

24. Hanson EK, Lubenow H, Ballantyne J. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Anal Biochem*. 2009;387(2):303-314.

25. Li Y, Zhang J, Wei W, Wang Z, Prinz M, Hou Y. A strategy for co-analysis of microRNAs and DNA. *Forensic Sci Int Genet*. 2014;12C:24-29.

26. Wang Z, Luo H, Pan X, Liao M, Hou Y. A model for data analysis of microRNA expression in forensic body fluid identification. *Forensic Sci Int Genet*. 2012;6(3):419-423.

27. Wang Z, Zhang J, Luo H, Ye Y, Yan J, Hou Y. Screening and confirmation of microRNA markers for forensic body fluid identification. *Forensic Sci Int Genet*. 2013;7(1):116-123.

28. Zubakov D, Boersma AW, Choi Y, van Kuijk PF, Wiemer EA, Kayser M. MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *Int J Legal Med*. 2010;124(3):217-226.

29. Zhuang G, Wu X, Jiang Z, et al. Tumour-secreted miR-9 promotes endothelial cell migration and angiogenesis by activating the JAK-STAT pathway. *EMBO J*. 2012;31(17):3513-3523.

30. Hunter MP, Ismail N, Zhang X, et al. Detection of microRNA expression in human peripheral blood microvesicles. *PLoS One*. 2008;3(11):e3694.

31. Balzano F, Deiana M, Dei Giudici S, et al. miRNA stability in frozen plasma samples. *Molecules*. 2015;20(10):19030-19040.

32. Courts C, Madea B. Micro-RNA - A potential for forensic science? *Forensic Sci Int*. 2010;203(1-3):106-111.

33. Park JL, Park SM, Kwon OH, et al. Microarray screening and qRT-PCR evaluation of microRNA markers for forensic body fluid identification. *Electrophoresis*. 2014;35(21-22):3062-3068.

34. Griffiths-Jones S. miRBase: microRNA sequences and annotation. *Curr Protoc Bioinformatics*. 2010;Chapter 12:Unit 12.9.1-10.

35. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34(Database issue):D140-4.

36. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: Tools for microRNA genomics. *Nucleic Acids Res*. 2008;36(Database issue):D154-8.

37. Griffiths-Jones S. The microRNA registry. *Nucleic Acids Res*. 2004;32(Database issue):D109-11.

38. Wang Z, Zhou D, Cao Y, et al. Characterization of microRNA expression profiles in blood and saliva using the ion personal genome machine((R)) system (ion PGM system). *Forensic Sci Int Genet*. 2016;20:140-146.

39. Hanson EK, Ballantyne J. Circulating microRNA for the identification of forensically relevant body fluids. *Methods Mol Biol*. 2013;1024:221-234.

40. Omelia EJ, Uchimoto ML, Williams G. Quantitative PCR analysis of blood- and saliva-specific microRNA markers following solid-phase DNA extraction. *Anal Biochem*. 2013;435(2):120-122.

41. van der Meer D, Uchimoto ML, Williams G. Simultaneous analysis of micro-RNA and DNA for determining the body fluid origin of DNA profiles. *J Forensic Sci*. 2013;58(4):967-971.

42. Hou J, Lin L, Zhou W, et al. Identification of miRNomes in human liver and hepatocellular carcinoma reveals miR-199a/b-3p as therapeutic target for hepatocellular carcinoma. *Cancer Cell*. 2011;19(2):232-243.

43. Kuokkanen S, Chen B, Ojalvo L, Benard L, Santoro N, Pollard JW. Genomic profiling of microRNAs and messenger RNAs reveals hormonal regulation in microRNA expression in human endometrium. *Biol Reprod*. 2010;82(4):791-801.

44. Pan Q, Chegini N. MicroRNA signature and regulatory functions in the endometrium during normal and disease states. *Semin Reprod Med*. 2008;26(6):479-493.

45. Wu S, Huang S, Ding J, et al. Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene*. 2010;29(15):2302-2308.

46. Wu WK, Law PT, Lee CW, et al. MicroRNA in colorectal cancer: From benchtop to bedside. *Carcinogenesis*. 2011;32(3):247-253.

47. Link A, Balaguer F, Shen Y, et al. Fecal MicroRNAs as novel biomarkers for colon cancer screening. *Cancer Epidemiol Biomarkers Prev*. 2010;19(7):1766-1774.

48. Creighton CJ, Benham AL, Zhu H, et al. Discovery of novel microRNAs in female reproductive tract using next generation sequencing. *PLoS One*. 2010;5(3):e9637.

49. Jima DD, Zhang J, Jacobs C, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood*. 2010;116(23):e118-27.

50. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*. 2008;18(10):997-1006.

51. Friedlander MR, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*. 2008;26(4):407-415.

52. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--excel-based tool using pair-wise correlations. *Biotechnol Lett*. 2004;26(6):509-515.

53. Hammer O, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2001;4(1):1-9.

54. Floris I, Billard H, Boquien CY, et al. MiRNA analysis by quantitative PCR in preterm human breast milk reveals daily fluctuations of hsa-miR-16-5p. *PLoS One*. 2015;10(10):e0140488.

55. Chen X, Liang H, Guan D, et al. A combination of let-7d, let-7g and let-7i serves as a stable reference for normalization of serum microRNAs. *PLoS One*. 2013;8(11):e79652.

56. Wong W, Farr R, Joglekar M, Januszewski A, Hardikar A. Probe-based real-time PCR approaches for quantitative measurement of microRNAs. *J Vis Exp*. 2015;(98). doi(98):10.3791/52586.

57. Cheng L, Sun X, Scicluna BJ, Coleman BM, Hill AF. Characterization and deep sequencing analysis of exosomal and non-exosomal miRNA in human urine. *Kidney Int*. 2014;86(2):433-444.

58. Azzouzi I, Moest H, Wollscheid B, Schmugge M, Eekels JJ, Speer O. Deep sequencing and proteomic analysis of the microRNA-induced silencing complex in human red blood cells. *Exp Hematol*. 2015;43(5):382-392.

59. Ogawa Y, Taketomi Y, Murakami M, Tsujimoto M, Yanoshita R. Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. *Biol Pharm Bull*. 2013;36(1):66-75.

60. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. error probabilities. *Genome Res*. 1998;8(3):186-194.

61. Hu L, Wu C, Guo C, Li H, Xiong C. Identification of microRNAs predominately derived from testis and epididymis in human seminal plasma. *Clin Biochem*. 2014;47(10-11):967-972.

62. Knutsen E, Fiskaa T, Ursvik A, et al. Performance comparison of digital microRNA profiling technologies applied on human breast cancer cell lines. *PLoS One*. 2013;8(10):e75813.

63. Ma J, Mannoor K, Gao L, et al. Characterization of microRNA transcriptome in lung cancer by next-generation deep sequencing. *Mol Oncol*. 2014;8(7):1208-1219.

64. Yan LX, Huang XF, Shao Q, et al. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA*. 2008;14(11):2348-2360.

65. Keller A, Leidinger P, Vogel B, et al. miRNAs can be generally associated with human pathologies as exemplified for miR-144. *BMC Med*. 2014;12:224-014-0224-0.

66. Chang CW, Wu HC, Terry MB, Santella RM. microRNA expression in prospectively collected blood as a potential biomarker of breast cancer risk in the BCFR. *Anticancer Res*. 2015;35(7):3969-3977.

67. Federal Bureau of Investigation. Quality assurance standards for forensic DNA testing laboratories. *Forensic Sci Comm*. 2000;2(3).

68. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol Cell*. 2007;27(1):91-105.

69. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009;19(1):92-105.

70. Wang X, El Naqa IM. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*. 2008;24(3):325-332.

71. Crossland RE, Norden J, Bibby LA, Davis J, Dickinson AM. Evaluation of optimal extracellular vesicle small RNA isolation and qRT-PCR normalisation for serum and urine. *J Immunol Methods*. 2016;429:39-49.

72. Zhang X, Ladd A, Dragoescu E, Budd WT, Ware JL, Zehner ZE. MicroRNA-17-3p is a prostate tumor suppressor in vitro and in vivo, and is decreased in high grade prostate tumors analyzed by laser capture microdissection. *Clin Exp Metastasis*. 2009;26(8):965-979.

73. Wang C, Yang C, Chen X, et al. Altered profile of seminal plasma microRNAs in the molecular diagnosis of male infertility. *Clin Chem*. 2011;57(12):1722-1731.

74. Hanke M, Hoefig K, Merz H, et al. A robust methodology to study urine microRNA as tumor marker: microRNA-126 and microRNA-182 are related to urinary bladder cancer. *Urol Oncol*. 2010;28(6):655-661.

75. Gidlof O, Andersson P, van der Pals J, Gotberg M, Erlinge D. Cardiospecific microRNA plasma levels correlate with troponin and cardiac function in patients with ST elevation myocardial infarction, are selectively dependent on renal elimination, and can be detected in urine samples. *Cardiology*. 2011;118(4):217-226.

76. Wang K, Zhang S, Weber J, Baxter D, Galas DJ. Export of microRNAs and microRNA-protective protein by mammalian cells. *Nucleic Acids Res*. 2010;38(20):7248-7259.

77. Lasser C, Alikhani VS, Ekstrom K, et al. Human saliva, plasma and breast milk exosomes contain RNA: Uptake by macrophages. *J Transl Med*. 2011;9:9-5876-9-9.

78. Setzer M, Juusola J, Ballantyne J. Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. *J Forensic Sci*. 2008;53(2):296-305.

79. Kladwang W, Hum J, Das R. Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. *Sci Rep*. 2012;2:517.

80. Etheridge A, Lee I, Hood L, Galas D, Wang K. Extracellular microRNA: A new source of biomarkers. *Mutat Res*. 2011;717(1-2):85-90.

81. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. 2007;9(6):654-659.

82. Fox CA, Meldrum SJ, Watson BW. Continuous measurement by radio-telemetry of vaginal pH during human coitus. *J Reprod Fertil*. 1973;33(1):69-75.

83. Tevi-Benissan C, Belec L, Levy M, et al. In vivo semen-associated pH neutralization of cervicovaginal secretions. *Clin Diagn Lab Immunol*. 1997;4(3):367-374.

84. Sauer E, Babion I, Madea B, Courts C. An evidence based strategy for normalization of quantitative PCR data from miRNA expression analysis in forensic organ tissue identification. *Forensic Sci Int Genet*. 2014;13:217-223.

85. Wang Z, Zhang J, Wei W, et al. Identification of saliva using MicroRNA biomarkers for forensic purpose. *J Forensic Sci*. 2015;60(3):702-706.

86. United States Department of Justice Federal Bureau of Investigations. Quality assurance standards for forensic DNA testing laboratories. *Forensic Sci Comm*. 2000;2.

Dissemination of Research Findings & Research Products

Submitted Manuscripts:

1. SJ Seashols-Williams, C Calloway, N Peace, A Priola, Q Wu, S Fleming, A Albornoz, C Hayes, ZE Zehner. High-Throughput Sequencing of the miRnome of eight forensically relevant biological fluids. Manuscript submitted to special forensic science edition of Electrophoresis, May 2016.

Manuscripts in Preparation or Revision:

2. SJ Seashols-Williams, F Nogales, C Hayes, A Albornoz, S Fleming, ZE Zehner. Stability of microRNAs in forensically relevant biological fluids. Manuscript under revision, to be submitted summer 2016.
3. SJ Seashols-Williams, C Lewis, C Calloway, N Peace, A Priola, Q Wu, S Fleming, A Albornoz, C Hayes, ZE Zehner. Development of a miRNA panel to distinguish forensically relevant biological fluids. Manuscript in preparation, to be submitted to Journal of Forensic Sciences in July 2016.

Presentations:

1. C Lewis, C Calloway, N Peace, A Albornoz, S Fleming, C Hayes, Z Zehner, S-Seashols Williams. Developmental Validation of a miRNA panel for the Identification of Six Forensically Relevant Body Fluids. Talk, Mid-Atlantic Association of Forensic Scientists Annual Meeting, May 2016. Federal Support Acknowledged.
2. C Lewis, J Gentry, C Calloway, N Peace, A Albornoz, S Fleming, C Hayes, Z Zehner, SJ Seashols-Williams. Developmental Validation of microRNAs for Body Fluid Identification. Talk, presented at the American Academy of Forensic Sciences Annual Meeting, February 2016. Federal Support Acknowledged.
3. C Calloway, SJ Seashols-Williams, N Peace, C Hayes, A Albornoz, S Fleming, T Layne, J Gentry, K Sharma, Q Wu, and ZE Zehner. Investigation of Next Generation Sequencing Data for Constitutive and Body Fluid Specific MicroRNAs in forensically relevant body fluids. Talk, Mid-Atlantic Association of Forensic Scientists Annual Meeting, May 2015. Federal Support Acknowledged.
4. T Layne, ZE Zehner, SJ Seashols-Williams. Stability and Variation of microRNAs for Body Fluid Identification. Talk in the Young Forensic Scientists "Bring your own slides" Forum, American Academy of Forensic Sciences Annual Meeting, February 2015. Federal Support Acknowledged.
5. S Seashols-Williams, A Albornoz, C Hayes, S Fleming, C Calloway, N Peace, Q Wu, K Sharma, J Gentry, and ZE Zehner. The use of high throughput sequencing to identify potential normalization microRNAs found in eight forensically relevant body fluids. Poster, "Bring Your Own Poster" Session, American Academy of Forensic Sciences Annual Meeting, February 2015. Federal Support Acknowledged.
6. SJ Seashols-Williams, A Albornoz, C Hayes, S Fleming, Q Wu, K Sharma, J Gentry, ZE Zehner. High-throughput miRNA sequencing of 8 forensically relevant biological fluids. Poster presentation, International Symposium of Human Identification, September 2014. Federal Support Acknowledged.

7. SR Fleming, A Albornoz, CR Hayes, ZE Zehner, SJ Seashols. Optimized Methods for Isolation of microRNAs from Forensically Relevant Body Fluids. Talk, Mid-Atlantic Association of Forensic Scientists Annual Meeting, May 2014. Federal Support Acknowledged.
8. A Albornoz, SR Fleming, CR Hayes, ZE Zehner, SJ Seashols. Variation of microRNA expression in Blood, and Menstrual Blood, and Vaginal Fluid over biological time. Talk, Mid-Atlantic Association of Forensic Scientists Annual Meeting, May 2014. Federal Support Acknowledged.
9. SR Fleming, A Albornoz, CR Hayes, ZE Zehner, SJ Seashols. Optimized Methods for Isolation of microRNAs from Forensically Relevant Body Fluids. Talk, American Academy of Forensic Sciences Annual Meeting, February 2014. Federal Support Acknowledged.
10. CR Hayes, SR Fleming, A Albornoz, SJ Seashols, ZE Zehner. microRNA variation between individuals and stability in forensically relevant body fluids. Talk, National Institute of Justice Grantees meeting at the American Academy of Forensic Sciences Annual Meeting, February 2014. Federal Support Acknowledged.
11. SJ Seashols, F Nogales, C Hayes, A Albornoz, S Fleming, W Budd, Z Zehner. An Evaluation of microRNA Stability and Internal Standard Selection for Forensic Body Fluid Identification. Talk, Mid-Atlantic Association of Forensic Scientists Annual Meeting, May 2013. Federal Support Acknowledged.
12. SJ Seashols, W Budd, Z Zehner. An Evaluation of microRNA Stability and Internal Standard Selection for Forensic Body Fluid Identification. Poster, American Academy of Forensic Sciences Annual Meeting, February 2013. Federal Support Acknowledged.

Website(s) or other Internet site(s): Nothing to Report.

Technologies or techniques:

1. miRNA panel for forensic body fluid identification. US Patent 62/288,788, filed January 29, 2016.

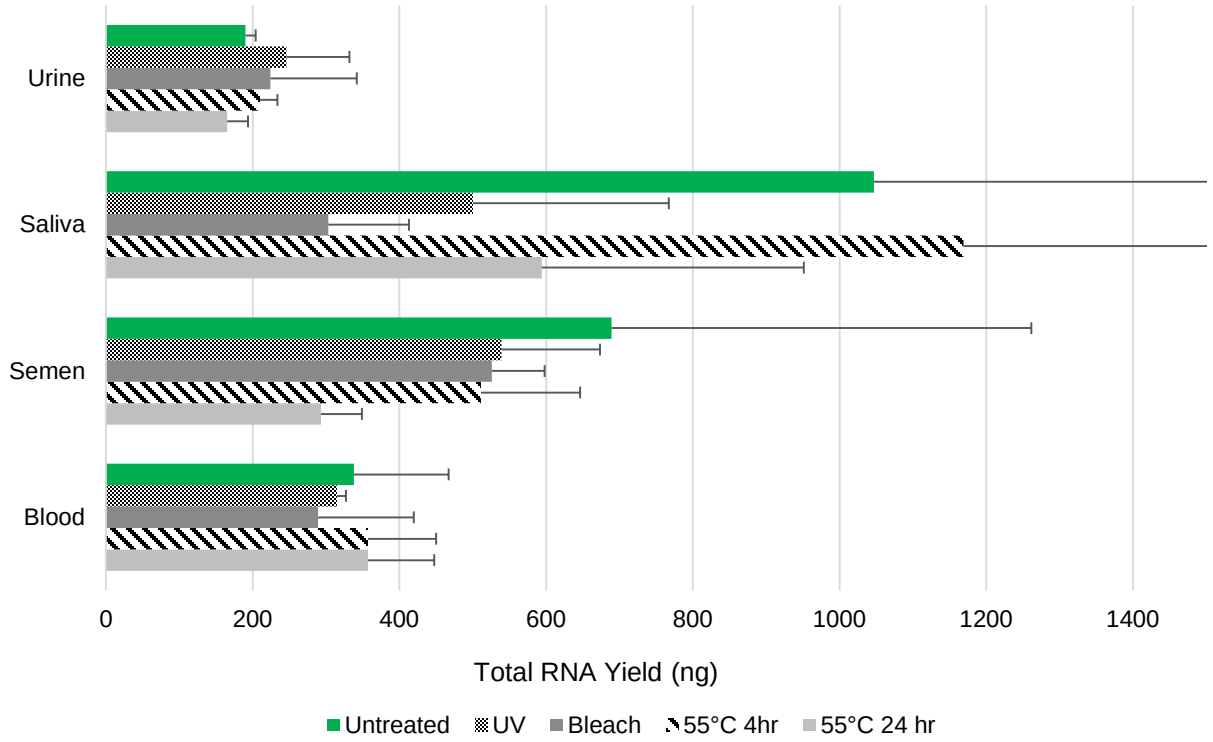
Other products:

1. Physical collection of body fluid samples. 90 volunteers, over 600 biological samples from eight forensically relevant biological fluids donated, preserved and catalogued according to approved Human Subjects Protection protocols.

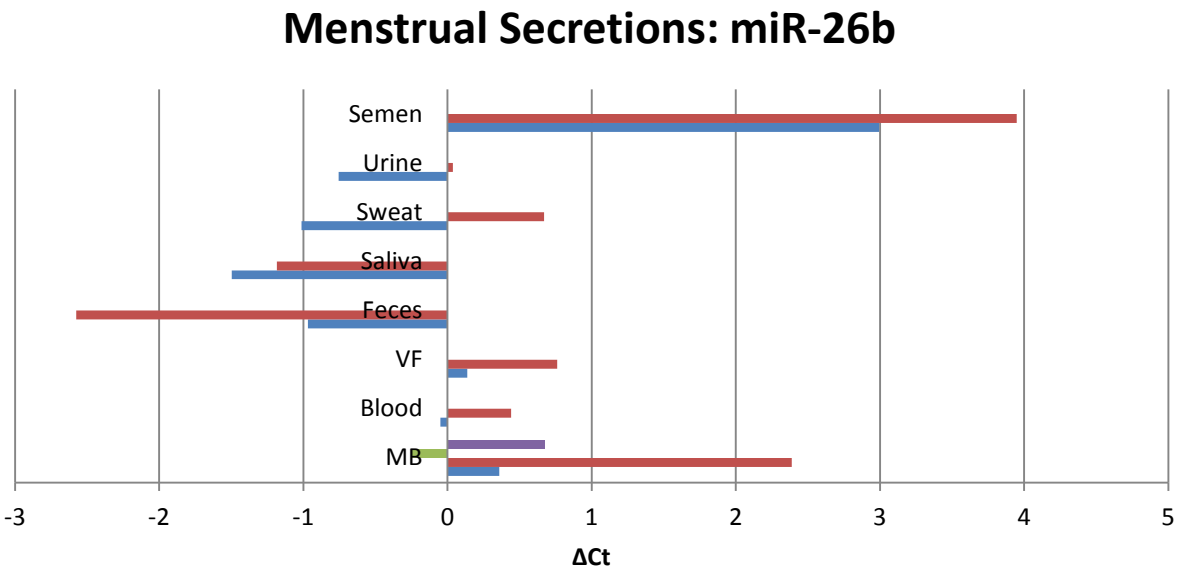
Supplemental Data

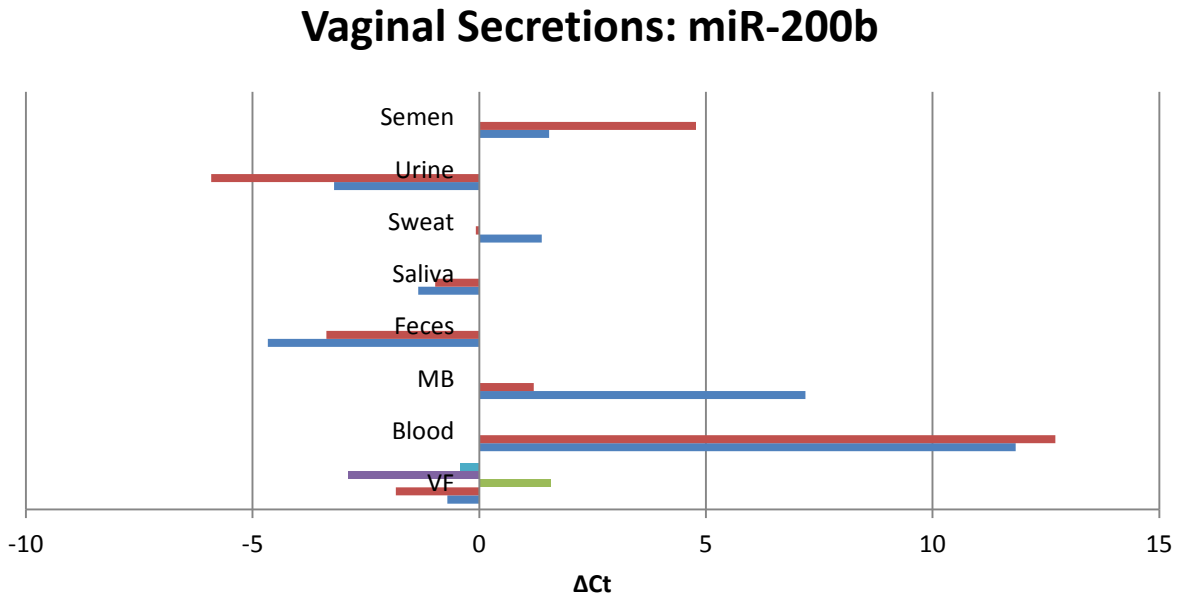
Supplemental Table 1: 72 miRNAs identified through HTS analysis as expressed only in blood

Blood-indicative miRNAs	
mir-15a 5'	mir-576 5'
mir-17 3'	mir-590 3'
mir-18a 5'	mir-598 3'
mir-18a 3'	mir-624 5'
mir-19a 3'	mir-651 5'
mir-22 5'	mir-320c-1 3'
mir-93 3'	mir-454 5'
mir-29b-1 3'	mir-942 5'
mir-29b-2 5'	mir-548e 3'
mir-29b-2 3'	mir-1285-1 3'
mir-107 3'	mir-1285-2 3'
mir-16-2 3'	mir-548k 5'
mir-30d 3'	mir-1294 5'
mir-181a-2 3'	mir-548o 3'
mir-181b-1 5'	mir-320c-2 3'
mir-199a-2 3'	mir-1976 3'
mir-215 5'	mir-2110 5'
mir-221 5'	mir-3158-1 3'
mir-130a 3'	mir-3158-2 3'
mir-185 3'	mir-3200 3'
mir-190a 5'	mir-4306 3'
mir-194-1 5'	mir-3613 5'
mir-181b-2 5'	mir-3615 3'
mir-128-2 3'	mir-3688-1 3'
mir-340 3'	mir-3912 3'
mir-331 5'	mir-548o-2 3'
mir-331 3'	mir-4454 5'
mir-324 3'	mir-4508 5'
mir-339 5'	mir-4732 5'
mir-335 5'	mir-4732 3'
mir-335 3'	mir-3688-2 3'
mir-424 5'	mir-5010 3'
mir-484 5'	mir-7641-1 5'
mir-502 3'	mir-7641-2 5'
mir-450a-2 5'	mir-7976 5'
mir-503 5'	let-7i 3'



Supplemental Figure 1: RNA Yield is not significantly impacted by environmental or chemical exposure. RNA was isolated from samples treated as described and quantitated using the NanoDrop 2000 UV Spectrophotometer (n=3 donors).





Supplemental Figure 2: Relative Expression of miRNAs did not follow expected patterns suggested by HTS data. While expression patterns were not significant for the body fluid in question, several of the tested miRNAs showed potential for differential expression in other boy fluid types. RT-qPCR analysis: n=5 of target body fluid (miRNA and target fluid as identified in HTS analysis in chart title), and n=2 population samples for the other seven body fluids assessed. RT-qPCR data for other candidate miRNA initial evaluations are found in Supplemental Figure 2.

Supplemental Table 2: ΔCqs of tested miRNAs under 12 treatments in 4 biological fluids.
(NT= not tested)

		miR-16		let-7g		miR-21		miR-24	
		ΔCq	stdev	ΔCq	stdev	ΔCq	stdev	ΔCq	stdev
	Blood	4.18	1.77	2.11	0.21	NT	---	0.75	0.27
	Urine	4.62	1.39	2.85	1.90	3.66	0.27	4.00	0.48
UV	Semen	4.79	1.79	-1.75	0.29	7.38	0.95	3.98	0.10
	Saliva	-1.33	2.55	1.96	0.20	NT	---	0.34	0.28
	Blood	10.16	2.46	9.93	0.21	NT	---	6.33	0.27
	Urine	6.49	0.27	3.75	1.58	8.35	1.26	0.75	0.21
FS Bleach	Semen	12.00	0.00	4.21	0.75	9.93	2.66	8.63	0.57
	Saliva	12.00	0.00	11.49	0.00	NT	---	17.64	1.66
	Blood	0.04	2.43	0.60	0.29	NT	---	0.33	0.14
55C 30'	Urine	4.65	2.30	-3.40	0.57	3.19	0.84	2.41	0.05
	Semen	8.50	0.37	2.25	0.56	10.17	2.25	7.90	0.71
	Saliva	-0.44	5.01	-1.93	0.18	NT	---	-0.15	0.04
	Blood	0.16	1.89	3.51	0.46			-1.39	0.11
55C 60'	Urine	6.51	1.62	-4.18	0.36	7.13	0.97	6.80	0.33
	Semen	6.42	2.68	2.95	0.64	-0.75	1.07	0.90	0.13
	Saliva	-3.07	2.21	-1.69	0.20			0.14	0.05
	Blood	3.25	0.90	6.48	0.18	NT	---	-1.20	0.14
55C 120'	Urine	6.39	2.12	-2.10	0.45	2.52	1.06	0.72	0.12
	Semen	4.29	2.15	2.83	0.43	0.90	0.37	0.00	0.23
	Saliva	-4.30	0.53	-1.08	0.24	NT	---	-0.46	0.04
	Blood	5.95	1.66	1.63	0.23	NT	---	0.23	0.21
55C 24 hr	Urine	4.88	1.53	0.89	0.73	4.65	0.70	4.43	0.32
	Semen	5.15	1.67	2.61	0.38	4.71	0.42	0.96	0.47
	Saliva	-4.59	2.48	-1.75	0.24	NT	---	0.72	0.21
	Blood	8.51	2.73	1.80	0.68	NT	---	NT	---
95C 1 hr	Urine	2.59	1.77	0.80	0.99	10.39	1.83	NT	---
	Semen	NT	---	NT	---	NT	---	NT	---
	Saliva	NT	---	NT	---	NT	---	NT	---
	Blood	7.88	3.47	2.73	0.33	NT	---	NT	---
95C 2 hr	Urine	3.99	1.31	0.08	0.67	10.73	1.43	NT	---
	Semen	NT	---	NT	---	NT	---	NT	---
	Saliva	NT	---	NT	---	NT	---	NT	---
95C 24 hr	Blood	10.11	1.07	4.17	0.43	NT	---	NT	---

	Urine	2.73	0.54	-4.05	0.29	9.98	1.13	NT	---
	Semen	NT	---	NT	---	NT	---	NT	---
	Saliva	NT	---	NT	---	NT	---	NT	---
10% Bleach									
	Blood	6.95	0.66	4.79	0.35	NT	---	NT	---
	Urine	1.75	1.26	5.19	1.17	3.46	0.28	NT	---
	Semen	12.00	0.00	-1.42	0.78	8.96	2.84	NT	---
	Saliva	12.00	0.00	10.42	2.00	NT	---	NT	---
DD	Blood	4.58	0.26	6.75	0.44	NT	---	NT	---
	Urine	3.12	1.14	3.84	2.51	9.15	2.76	NT	---
	Semen	7.60	1.28	4.01	0.61	10.58	1.28	NT	---
	Saliva	-2.31	0.06	6.94	0.23	NT	---	NT	---
Acetic Acid									
	Blood	1.51	0.53	-0.34	0.36	NT	---	NT	---
	Urine	4.52	1.99	-1.04	0.85	NT	---	NT	---
	Semen	4.39	0.77	-4.96	0.46	6.59	2.74	NT	---
	Saliva	-1.01	0.70	-0.56	0.48	NT	---	NT	---

1
2