RNA-SEQ OF BIOLOGICAL FLUIDS FOR THE EVALUATION OF MRNA DEGRADATION IN RELATION TO SAMPLE AGE

By

KATE WEINBRECHT

Bachelor of Science in Biology
Pacific Lutheran University
Tacoma, Washington
2009

Master of Science in Forensic Science
Oklahoma State University
Tulsa, Oklahoma
2011

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2014

RNA-SEQ OF BIOLOGICAL FLUIDS FOR THE EVALUATION OF MRNA DEGRADATION IN RELATION TO SAMPLE AGE

Dissertation Approved:

Dr. Robert Allen

Dissertation Adviser

Dr. Gerwald Koehler

Dr. Jarrad Wagner

Dr. Mark Payton

Name: KATE WEINBRECHT

Date of Degree: JULY, 2014

Title of Study: RNA-SEQ OF BIOLOGICAL FLUIDS FOR THE EVALUATION OF MRNA DEGRADATION IN RELATION TO SAMPLE AGE

Major Field: BIOMEDICAL SCIENCES

Abstract: Research on the forensic applications of RNA analysis has increased greatly in the last decade. Defined uses of RNA in forensic analysis include the use of RNA to identify tissue type, determine sample age, and play a role in molecular autopsies. Although recent research has indicated many possible forensic applications of RNA analysis, many questions remain concerning the behavior of RNA in degraded and limited samples. Specifically, there remains to be a thorough understanding of the differing patterns and rates of RNA degradation in post-mortem and deposited samples. Thus, choosing suitable RNA markers for evaluating the approximate age of a forensic sample can be problematic. Development of a reliable and accurate molecular assay for the determination of sample age (time-since deposition of a biological sample and/or post-mortem interval) will play a critical role in helping investigators establish the timeline of events that surround a crime. The purpose of this research is to evaluate mRNA degradation in forensically relevant biological sample types (blood, saliva, semen, and vaginal fluid) in order to establish tissue-specific transcriptome (total mRNA) degradation profiles and patterns that may correlate with the age of a sample. Transcriptome sequencing of mRNA isolated from fresh and aged samples (0 days to 360 days old) was performed to evaluate the patterns of mRNA degradation in relation to sample age. Sequencing data was used to determine the pattern and rate of degradation for each individual mRNA transcript in each sample type. Sequencing data indicates that the mRNA population and transcript degradation rates appear to be tissue-specific. The mRNA degradation profiles obtained from this study can be used to determine the transcripts in each sample type that have degradation patterns and rates correlated with sample age.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Historically, deoxyribonucleic acid (DNA) analysis has played a dominant role in forensic investigations, while use of ribonucleic acid (RNA) analysis has been limited. For several decades RNA was thought to be both too labile and too susceptible to degradation for use with most forensically relevant samples. However, several studies over the past decade have demonstrated that RNA may be much more stable in *ex vivo* samples than was once believed (Fordyce, Kampmann, Doorn, & Gilbert, 2013). With a greater number of studies being performed on RNA in a forensic context, researchers have begun investigating practical applications of RNA analysis in forensic science. Research has demonstrated the possible use of RNA analysis in tissue identification, estimation of time since deposition and post mortem interval, and determination of disease state, drug use, and mechanism of death (S. Anderson, Howard, Hobbs, & Bishop, 2005; M Bauer, 2007; Lindenbergh et al.., 2012; Vennemann & Koppelkamm, 2010a). While the analysis of DNA can provide investigators with human identity, the analysis of RNA from forensic samples may potentially provide a wealth of information concerning when and how a crime occurred.

A heightened interest in forensic RNA research can in part be attributed to several studies that have demonstrated long-term stability of RNA molecules in *ex* vivo samples. Forensically relevant biological samples are often degraded and/or minimally available, thus the demonstrated presence of RNA in these sample types was required before further forensic applications of RNA analysis could be explored. In research performed by Kohlmeier and Schneider, RNA was successfully isolated and profiled from a 23-year old blood stain (Kohlmeier & Schneider, 2012). Similar results were achieved in a study by Bauer et al.. with a 15 year old blood stain and in a study by Zubakov et al. with a 16 year old blood stain (Martin Bauer, Polzin, & Patzelt, 2003; Zubakov, Kokshoorn, Kloosterman, & Kayser, 2009). In addition to blood samples, studies of saliva, semen, seminal fluid, vaginal secretion, sweat, and skin demonstrate that RNA can be isolated in a variety of biological samples that are several years old (Haas, Muheim, Kratzer, Bär, & Maake, 2009; Sakurada, Akutsu, Fukushima, Watanabe, & Yoshino, 2010; Sakurada, Akutsu, Watanabe, Fujinami, & Yoshino, 2011; Visser, Zubakov, Ballantyne, & Kayser, 2011). Studies such as these demonstrate the possibility of isolating and analyzing RNA from forensically relevant samples. While past studies have demonstrated the stability of RNA in aged biological material, the next line of research is determining how the presence of RNA can be used to learn more about a sample.

Over the past decade, there has been initial research on monitoring RNA degradation as a time-clock for sample age estimation. However, past research is limited in both the number of RNA markers evaluated and the types of biological samples included in analysis. Research evaluating sample age by monitoring RNA degradation has mainly focused on ribosomal RNA (rRNA), housekeeping mRNA transcripts, and

tissue-specific mRNA transcripts. These studies have utilized both end-point PCR paired with capillary electrophoresis and real-time reverse transcriptase PCR (RT qPCR) to monitor degradation rates in a few select RNA species. Early work focused on the degradation rates of housekeeping mRNA transcripts and rRNA, as these species have a known presence in all tissue types. In work performed by Bower et al., analysis of 106 bloodstains, aged up to 15 years, revealed that the abundance of β-actin and cyclophilin transcripts decreased in relation to sample age (Martin Bauer, Gramlich, Polzin, & Patzelt, 2003). Anderson et al.. expanded research on β-actin mRNA degradation by demonstrating that the approximate age of a bloodstain can be predicted by determining the ratio between β-actin mRNA and 18S rRNA (S. Anderson et al.., 2005). The 18S rRNA product is stable and remains at a steady level of abundance in aged stains compared to β-actin mRNA, which decreases in abundance over time. This work expanded to include the evaluation of different amplicon sizes of both β-actin mRNA and 18S rRNA, with older bloodstains having a reduced presence of longer amplicons than fresh samples (S. E. Anderson, Hobbs, & Bishop, 2011). Anderson et al.. found that the most robust estimation of age came from a multivariate analysis that takes into account multiple amplicons (of varying sizes) on multiple genes (S. E. Anderson et al.., 2011). While these initial studies of RNA degradation in aged bloodstains have been limited to only examining a few select RNA transcripts, the results do indicate a correlation between sample age and RNA degradation rates.

The literature clearly indicates initial promise for using RNA degradation as a time-clock for sample age estimation. However, while researchers have identified a few possible RNA markers for determining the approximate age of a biological sample, the

research is limited and does not include any evaluation of whole transcriptome (total RNA in a sample) degradation patterns. The reduced specificity in predicting the actual age of a sample in many of these studies may be attributed to the lack of consideration of more accurate RNA markers available in the transcriptome of a sample type (S. E. Anderson et al.., 2011; Martin Bauer, Gramlich, et al.., 2003; Vass, Fleming, Harbison, Curran, & Williams, 2013; Young, Wells, Hobbs, & Bishop, 2013). Up until this point, all past studies have worked under the assumption that their selected RNA transcripts were accurate enough to measure the age of a sample. However, in total, past studies on estimating sample age through RNA degradation have evaluated less than 20 RNA species. This limited number of evaluated transcripts is problematic because the transcriptome of any given tissue contains thousands of possible RNA targets that may have degradation patterns more closely tied to predicting accurate sample age. While past studies have chosen their RNA targets based on known transcript availability (i.e. rRNA, housekeeping mRNA transcripts, or tissue-specific mRNA transcripts), these targets may not be the most accurate predictor of sample age. No study has ever monitored whole transcriptome degradation in biological samples over an extended period (several months to years). Thus, researchers have no way of choosing the most accurate RNA markers for establishing sample age in a specific sample type.

In addition to the limited number of RNA markers evaluated, past research on RNA degradation in deposited biological fluids has focused largely on blood, with no major studies having been performed on other forensically relevant sample types (such as semen, saliva, and vaginal fluid). Evaluation of biological fluid types other than blood is critical since the cell types, cellular environments, and transcriptomes vary considerably

with each fluid type. Thus, RNA degradation patterns and rates may likely be different in each sample type. If investigators are going to be able to evaluate time since deposition in a variety of sample types, it is critical to study RNA degradation patterns and rates in forensically relevant biological fluid types other than blood.

Research Purpose

The main purpose of this study is evaluation of total mRNA degradation in deposited biological fluid samples in an effort to identify specific mRNA markers that correlate with sample age. This research aims to increase the body of knowledge on how mRNA behaves in *ex vivo* samples (specifically, deposited blood, saliva, vaginal fluid, and semen; and human teeth), aged up to one year. All past research using RNA to establish the age of a sample (time since deposition or PMI) has relied upon a minimal number of housekeeping mRNA transcripts, tissue-specific mRNA transcripts, and 18S rRNA. While these studies have demonstrated a clear relationship between RNA degradation and sample age, previous studies have not identified RNA markers that are accurately correlated with long-term sample age (samples aged up to one year or longer). While the RNA markers examined in past studies were presumably chosen because of their known presence in biological tissues, these markers are not necessarily the RNA species whose abundance most closely correlates with sample age. This study will take a different approach than any past study of RNA degradation by evaluating the total mRNA of fresh and aged samples through use of next-generation RNA sequencing (RNA-seq). The broad knowledge gained from this study on RNA degradation will

facilitate selection of specific mRNA markers for establishing approximate sample age in each individual sample type.

Research Questions

Throughout the execution of this study, the following research questions will provide guidance and focus to the research.

1. Is there an observable pattern or profile of total mRNA degradation in deposited biological fluid samples (blood, saliva, semen, and vaginal fluid)?
2. Do different mRNA transcripts degrade at different rates?
3. Do different biological fluid-types (blood, saliva, semen, and vaginal fluid) have different patterns and/or rates of RNA degradation?
4. Does RNA degradation correlate with approximate sample age for each of the sample types?

Hypothesis

It is established that the transcriptome of a biological sample does degrade once it is deposited outside of the body or upon death. Therefore, RNA degradation patterns should be observable by sequencing the transcriptomes of biological samples that have been aged under known conditions for controlled amounts of time. RNA transcripts have varying size and complexity and will most likely have distinct rates of degradation. Therefore, different RNA transcripts may produce unique degradation profiles. Each of the different fluid types that are tested will contain unique RNA transcripts, thus each

sample type will also produce a unique transcriptome degradation profile. If these degradation profiles correlate with time since deposition of the biological sample, then the degradation profiles of specific identified mRNA transcripts within a sample could help predict the approximate age of the sample.

Overview of Methodology

The methods utilized in this study aim to provide a comprehensive snapshot of mRNA degradation over a specified amount of time (6 months or 1 year) in four forensically relevant sample types (blood, semen, saliva, and vaginal fluid). The methodology can be broken into four main components.

In the first component of this study, a comprehensive literature search was performed to identify mRNA markers for blood, semen, saliva, and vaginal fluid (Appendix A). Tissue- and fluid-specific gene products are well established in the literature, and several mRNA markers for specific sample types have been validated for specificity and sensitivity. In this study tissue-specific RNA transcripts are utilized to establish sample-specific mRNA degradation patterns. The databases utilized in the literature search include Google Scholar, NCBI PubMed, and ScienceDirect. The search terms included “RNA markers for tissue identification”, “RNA markers for biological fluid identification”, “RNA used to identify tissues and fluids”, “forensic identification of fluids and tissues using RNA”, and “mRNA markers for biological tissues and fluids”. The identified tissue-specific RNA transcripts will be analyzed in the RNA-seq data to determine if tissue-specific RNA degradation patterns are present in RNA markers already published for forensic applications.

In the second component of this study, RNA-seq library preparation was optimized for low input and degraded biological samples. Protocols for production of cDNA and sequencing libraries were evaluated for reduced sequencing bias and successful use with minimally available and degraded samples. The selected protocol for cDNA generation was the NuGEN Ovation® Kit for cDNA (NuGEN Technologies, San Carlos, CA). Following cDNA production, samples were fragmented and libraries were constructed using the Ion Plus Fragment Library Kit (Ion Torrent™, Life Technologies, Carlsbad, CA). All libraries were constructed using the same protocol, independent of sample type.

In the third component of this study, deposited body fluid samples were aged and mRNA was isolated and sequenced at periodic intervals up to six months (saliva, semen, and vaginal fluid) or one year (blood). Biological samples were collected and stored at room temperature in the dark. Two replicates of each sample type were analyzed at each of the time-course sampling intervals. An RNA/DNA co-isolation procedure was used to isolate RNA and cDNA generation and library production were carried out using the optimized protocol chosen in the second component of the study. Once constructed, libraries underwent template preparation on the OneTouch™ 2 (OT2™) and prepared templates were sequenced on the Ion Torrent™ Personal Genome Machine®, referred to subsequently as Ion PGM™.

In the fourth component of this study, all RNA-seq data were analyzed. All raw RNA-seq data sets were trimmed for quality and aligned to the human genome (HG19, GrCH37). RNA-seq abundance values were first normalized by calculating the Reads per Kilobase per Million (RPKM) value for each gene. RPKM values were then normalized

to a standard curve reflecting abundance levels for a set of ERCC RNA external controls that were spiked in to each individual RNA sample prior to cDNA synthesis. The standard curve prepared from the normalized abundance values of the ERCC spike-in controls (Ambion®) was used to quantify mRNA levels of specific templates between time points both within and between sample types to determine the presence of mRNA degradation profiles and patterns

CHAPTER II

REVIEW OF THE LITERATURE

A review of the literature reveals that while there is a large body of knowledge on the types, roles, and degradation mechanisms of RNA in an *in vivo* context, much remains to be understood about the *ex vivo* behavior of RNA. The lack of knowledge concerning RNA in *ex vivo* samples is beginning to be remedied by studies surrounding the use of RNA analysis of post-mortem and deposited biological samples. The results of these studies indicate that RNA has the potential to offer a substantial amount of information in a forensic context. However, if RNA analysis is going to be fully utilized in forensic analysis, it is critical that investigators obtain a more comprehensive understanding of the *ex vivo* behavior of RNA in different sample types.

Properties of RNA

RNA is a class of biological macromolecules responsible for a wide variety of functions within the mammalian cell. In the human body RNA is responsible for vital tasks including, coding, decoding, facilitating translation, monitoring protein expression, and catalyzing reaction within the cell. In order for RNA to perform a large variety of vital functions, many different types of RNA exist within a single cell. Major classes of

RNA include messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), and micro RNA (miRNA). Each different class of RNA is uniquely adapted to perform a specific function within the cell. The major population of RNA in any given human cell is rRNA (80%). The 28S, 5S, 5.8S, and 18S rRNAs form the two ribosomal subunits that help catalyze protein synthesis during translation. The tRNA, which is the next largest population of RNA within a cell (15%), also aids in the process of translation by moving the correct amino acids to the ribosome. The mRNA, which carries transcribed genetic information, constitutes a small percentage of the overall RNA population with in a cell (3-5%). All other classes of RNA, including both snRNA and miRNA, constitute a very small percentage of the total RNA population (<2%) (Lodish et al.., 2000; Vennemann & Koppelkamm, 2010a).

RNA molecules, like DNA molecules, are composed of nucleotides. However, unlike DNA, RNA is single stranded (mRNA). Additionally, RNA can be highly complexed with proteins that help define the structure and function of the molecule (rRNA and tRNA). The single stranded nature of RNA lends itself to fast production, molecular instability, and rapid degradation, properties that are important to the role of RNA as an intermediate molecule. The RNA transcript must be both rapidly produced and rapidly degraded in order for the cell to tightly regulate protein production. While the single stranded structure of RNA is necessary to maintain a high turnover rate within the living cell, this quality makes RNA much less stable and much more vulnerable to degradation then DNA. The double stranded structure of DNA lends stability and durability to the molecule, qualities that have allowed scientists access to the genetic code even in very old samples; whereas RNA can be degraded and possibly absent from aged

samples. While DNA has offered reliability in the lab, the highly unstable and labile nature of RNA has caused concern for scientists in the past when considering degraded or limited biological samples (Vennemann & Koppelkamm, 2010b).

The entire set of RNA molecules expressed within a given cell or tissue type is called the transcriptome. While the genome of a given person is conserved throughout every cell of the body, a property that has been exploited by investigators for means of personal identification, the transcriptome is different in each cell type. Every different cell type within the body will express a unique set of RNA molecules, providing for different cell types to have a distinct structure and function. Although the entire genome is found in every cell of the body, only specific genes get expressed in each different cell type. The transcriptome of a given cell type or tissue is fluctuating and is influenced by many factors including cell life cycle and cell environment. Additionally, RNA transcript levels can vary within a given individual as well as between individuals, a fact that must be recognized when performing analysis of RNA samples (Vennemann & Koppelkamm, 2010a, 2010b). While analysis of the genome can provide information on human identity, analysis of the transcriptome can provide deeper biological insight. Due to the unique RNA profile of each tissue type and the responsive nature of RNA production, analysis of the transcriptome can offer scientists information regarding tissue identity and biological conditions (drug use, health status, activity level, etc.) at the point in time when a tissue sample is collected (Bauer, 2007).

Mechanisms of RNA Degradation: *In vivo* vs. *Ex vivo*

Mechanisms of *in vivo* RNA degradation are well elucidated; with the three major types of RNA degradation being deadynelation-mediated mRNA decay (the most common type of RNA degradation), non-sense mediated decay, and AU-rich element mediated mRNA decay. While researchers know that multiple methods of *in vivo* RNA degradation take place, the most common type of mRNA degradation is deadynelation-mediated decay. During deadynelation-mediated mRNA decay, the poly-A tail is gradually decreased by deadenylating nucleases to the point that the PABP1 molecules can no longer bind to the e1F4E and e1F4G molecules (bound to the 5' cap), thus exposing the 5' cap. Once exposed, the 5' cap is removed by decapping enzymes and the unprotected mRNA is degraded by 5' to 3' exonucleases and exosomes containing 3' to 5' exonucleases. RNA degradation in the cell is a rapid process, with most RNAs having a half-life of several hours (Sharova et al.., 2009). Efficient turn-over of RNA is essential to for a cell to be able to adapt to its environment and monitor cellular function.

The mechanisms of *ex vivo* RNA degradation are not as well understood as *in vivo* degradation mechanisms. This is witnessed by the fact that the first paper on the mechanism of RNA degradation in deposited and post-mortem samples was not published until April, 2013 (Fordyce et al.., 2013). The degradation of RNA in *ex vivo* samples depends largely on sample type and sample condition. RNA degradation in fresh post-mortem samples that are not preserved or dried is driven by cellular RNases that remain active in moist cellular material. However, in samples that are dried (such as dried blood stains) or preserved (such as FFPE tissue samples), Rnases are largely inactivated, resulting in RNA degradation that is driven mostly by physical and chemical factors, such

as sunlight or pH. Degradation in *ex vivo* samples is also driven by the molecular structure of RNA. Due to the structure of RNA, in particular the 2'OH group, RNA molecules are more susceptible to spontaneous hydrolysis than DNA molecules. Specifically, the 2'OH group can attack the phosphodiester bond and cleave the backbone of RNA. While RNA is more prone to spontaneous hydrolysis than DNA, it is important to note that RNA is less prone to depurination or depyrimidination than DNA. This is because RNA forms stronger N-glycosidic bonds than DNA. This quality of RNA actually increases its *ex vivo* stability.

Although RNA is certainly more prone to degradation than DNA in post-mortem or deposited samples, RNA is often times more stable *ex vivo* then it is *in vivo.* This quality is due to the inactivation of Rnases in many *ex vivo* samples (such as those that have been dried, frozen, or preserved) (Fordyce et al.., 2013). Thus, RNA is much more stable in biological samples than was once assumed. However, the *ex vivo* RNA degradation rate is measurable over days and weeks and can be exploited to provide information about sample deposition time (S. E. Anderson et al.., 2011).

Stability and Variability of RNA

In order for RNA to be successfully utilized in a forensic context, critical issues including the questioned stability of the transcriptome and variable quantity of RNA in degraded or minimally available biological samples must be addressed. Originally, it was believed that RNA would be too difficult or even impossible to access in degraded samples due to its fragile, single stranded structure. However, research has proven that with enhanced molecular materials and methods, accessing RNA in aged, degraded, and

minimally available forensic samples is possible. It has also become clear that RNA may not be as unstable as was once believed, with several studies showing the successful isolation and analysis of RNA in decades old samples. Kohlmeier and Schneider successfully isolated and profiled mRNA from a 23-year old blood stain (Kohlmeier & Schneider, 2012). Bauer, et al. and Zubakov, et al. both demonstrated successful isolation of RNA from 15 and 16 year old blood stains, respectively. (Martin Bauer & Patzelt, 2008; Zubakov et al.., 2009). Similar results have been achieved with other aged biological sample types, including; saliva, semen, seminal fluid, vaginal secretion, sweat. (Haas, Muheim, et al.., 2009; Sakurada et al.., 2010, 2011; Visser et al.., 2011). Studies such as these have laid the ground work use of RNA analysis in forensic science, by demonstrating that RNA is much more stable in aged samples than was once believed.

Aside from the presence of RNA in aged and degraded samples, a second issue that must be considered is the known variability of RNA expression levels. The expression levels of RNA transcripts are not constant; they are known to vary between tissues within the same individual and between donors. The pool of mRNA in a given tissue is labile, reactive, and fluctuating due to constant environmental and biological influence. Multiple factors are known to effect RNA expression, including, gender, age, health status, weight, activity level, medications, amount of water intake, stress level, and drug and alcohol use, among several other factors (Vennemann & Koppelkamm, 2010b). A study performed by Koppelkamm, et al.. showed that RNA integrity and degradation pattern fluctuate depending on tissue type, cause of death, duration of agony, and body mass index (BMI) of the donor. For example, brain tissue appears to have reduced RNA integrity compared to cardiac and skeletal muscle and RNA from donors with an

increased BMI (>25) has a lower integrity then RNA isolated from normal weight donors (Koppelkamm, Vennemann, Lutz-Bonengel, Fracasso, & Vennemann, 2011). In order to lessen the effects of transcriptome variability between tissues and between donors, internal standards should be used to normalize data and degradation profiles should be obtained for individual RNA transcripts that are included in analysis. While sample to sample transcriptome variation will never be completely eliminated, by recognizing that variability does exist, steps can be taken to ensure that correct conclusions are drawn from the data produced.

The Role of RNA in Forensic Science

Over the past several decades the majority of forensic science research has focused on the use of DNA, as witnessed by the fact that up until 1994 there had only been two articles published that focused on the forensic application of RNA analysis (Oehmichen & Zilles, 1984; Phang, Shi, Chia, & Ong, 1994). In the past decade, research focusing on the use of RNA in forensic science has heightened due to improved technology to better support RNA analysis and an increased understanding that RNA is more stable in biologic samples than once believed (Vennemann & Koppelkamm, 2010a). With these improvements, researchers have begun to consider RNA analysis as a possible forensic investigative tool, used to enhance the knowledge already obtainable through traditional DNA analysis.

While the human genome can offer valuable information concerning human identification, DNA analysis does not offer insight into the events that surround a crime. Expanding forensic molecular analysis to include RNA will increase the amount of

information that can be gained from each individual biological sample. Some of the most abundant research on the applications of RNA in forensic science has focused on the identification of biomarkers for sample identification. Several investigators have focused their research efforts on the unique mRNA profile present in each biological sample type. The unique profile of both mRNA and miRNA in each tissue and fluid allows investigators to specifically identify a biological sample based on its RNA expression pattern (Liang, Ridzon, Wong, & Chen, 2007; Lindenbergh et al.., 2012; Park et al.., 2012; Richard et al.., 2012; Zubakov et al.., 2010). While tissue and fluid identify can be established with RNA biomarkers, research on the use of RNA biomarkers has recently expanded to provide a wider amount of information about a given biological sample.

The responsive nature of the transcriptome to biological conditions allows researchers to determine many aspects of the biological state of a sample upon deposition or death. By analyzing the RNA expression profile of a sample, researchers obtain a snapshot of the biological condition of the donor. For example, analyzing the RNA expression patterns of a tissue sample can provide researchers with information concerning the biological status of that tissue. In a study by Kagawa, et al., researchers identified seven genes that had differential expression patterns throughout the process of wound healing. By measuring the expression of these transcripts, researchers could successfully determine the approximate age of a wound (Kagawa et al.., 2009). Molecular autopsies can also be performed through assessing the expression pattern of gene products at time of death (Vennemann & Koppelkamm, 2010a). Studies have indicated gene products that are viable markers for methamphetamine related deaths, hypoxia related deaths, and mechanical asphyxiation (Ikematsu, Takahashi, Kondo,

Tsuda, & Nakasono, 2008; Matsuo, Ikematsu, & Nakasono, 2009; Zhao et al.., 2008). In one study by Ikematsu et al., researchers were able to successfully identify four candidate biomarkers of strangulation (Ikematsu, Tsuda, & Nakasono, 2006). Although much more work is needed in the field of molecular autopsies, recent research does indicate that monitoring RNA expression in post-mortem tissue may assist analysts in determining cause of death. RNA analysis as a whole has the potential to provide investigators with additional information about a biological sample that will complement information that is already available through traditional molecular analysis, offering insight on questions that can simply not be answered by DNA. However, successful application of any biomarker will require having a thorough understanding of mRNA in an *ex vivo* context is critical to insure proper interpretation of RNA analysis results.

Identification of Body Fluids and Tissues

Body fluids and tissues, including blood, semen, seminal fluid, vaginal secretion, saliva, and skin are regularly encountered in forensic casework. Often times, it is critical to an investigation to positively identify what type of tissue(s) and/or fluid(s) DNA was isolated from. Serological techniques are regularly employed in crime labs to identify what type of biological material is present on forensic samples. Current presumptive tests are most frequently enzymatic or immunologically based, and are at risk for inaccuracy. Current presumptive testing methods also lack the ability to identify all tissues and fluids in a mixed sample if certain fluid or tissue types are present in only minor quantities. Common presumptive tests for blood include the Kastle–Meyer phenolphthalein test, which relies on the peroxidase-like activity of hemoglobin and can give false-positive test

results with other peroxidases commonly found in plant material. Other confirmatory tests, such as the HemaTrace® Card, rely on the detection of hemoglobin in blood and therefore cannot determine if the source of the blood is menstrual or venous (Fleming & Harbison, 2010). Most presumptive tests for saliva rely on the detection of salivary amylase, which can be present in variable amounts in donors, thus sensitivity level and false-negatives are consistent issues. Presumptive tests for semen often rely on the detection of Prostate Specific Antigen (PSA), which can also be detected in male urine. Current presumptive testing methods require a separate test for the identification of each different body fluid (blood, semen, and saliva), leading to increased sample consumption and analysis time. Presumptive testing methods are also limited to the type of biological sample that they can test for. For example, there is no commonly used presumptive test for vaginal secretions or skin, both of which are considered to be common biological samples in forensic casework (Haas, Klesser, Maake, Bär, & Kratzer, 2009). Due to the risk of inaccuracy, high levels of sample consumption, and lack of comprehensive tests for fluid and tissue identification, researchers have sought out the use of molecular markers for sample identification.

The presence of a unique transcriptome in each body tissue and fluid allows biological samples to be identified based on the presence of specific mRNA and miRNA products. By co-isolating RNA and DNA, sample identification and human identity testing can be streamlined into a single molecular work flow. Additionally, with the creation of molecular panels of RNA markers, a single assay could potentially identify several different kinds of tissues and fluids in both single and mixed samples. Researchers have identified mRNA and miRNA markers for every different kind of

forensically common tissue and fluid, including venous blood, menstrual blood, semen, seminal fluid, vaginal secretion, saliva, and skin (Appendix A). Many of these markers have been evaluated for specificity and sensitivity and optimized for identification of forensic samples. In this study, tissue-specific RNA transcripts will be monitored in aged fluid samples to assess tissue-specific RNA degradation patterns and to determine how published RNA biomarkers hold up in aged samples.

Evaluation of RNA Degradation

While RNA can be successfully obtained from aged and minimally available biological samples, the isolated RNA is often degraded due to the inherent instability of the transcriptome. Unlike DNA, which can remain stable in biological samples for decades, RNA begins to degrade almost immediately after sample deposition or death (Martin Bauer, Polzin, et al.., 2003). Due to the immediate onset of degradation, changes in the transcriptome of a sample can be observed minutes, hours, days, months, and years after a biological sample is deposited or death occurs. By monitoring specific RNA products in biological samples that are aged in a controlled environment, degradation profiles can be developed to help analysts establish a timeline of events and approximate sample age (Bauer, Polzin, & Patzelt, 2003).

Assessment of the state of RNA degradation in a sample can help determine age of a biological sample, order of sample deposition, and post-mortem interval (PMI). In research performed by Anderson, et al., blood stains were aged under controlled conditions for 150 days. The results of this research demonstrate a linear relationship between sample age and the ratio of two RNA products, β-actin mRNA and 18S rRNA

(S. E. Anderson et al.., 2011; S. Anderson et al.., 2005). A separate study by Bauer, et al. looked at blood stains that had been stored for up to 15 years and found a correlation between sample age and capillary electrophoretic peak area quotients of two housekeeping gene products, β-actin and cyclophilin (M. Bauer, Polzin, & Patzelt, 2003). While both of these methods need refining before being employed in a crime lab, they provide initial evidence for a correlation between RNA degradation profiles and sample age. In addition to determining the time since deposition of a sample, monitoring RNA degradation of multiple samples from the same crime scene can provide an investigator with an order of sample deposition. If an analyst can determine the approximate age of multiple samples, an order of deposition can be established (S. E. Anderson et al.., 2011; Martin Bauer, Polzin, et al.., 2003). Past studies have provided solid evidence that a correlation does exist between the state of RNA degradation and sample age. However, more research is needed to look at more biological sample types and to further identify RNA markers with degradation patterns that most closely correlate with sample age.

Previous studies on the correlation between RNA degradation and PMI have not been as conclusive as studies that have examined time-since deposition of bloodstains. Some studies have found that RNA degradation does correlate with the PMI(Martin Bauer, Gramlich, et al.., 2003; Catts et al.., 2005; Inoue, Kimura, & Tuji, 2002; Kimura, Ishida, Hayashi, Nosaka, & Kondo, 2011), while other studies show no correlation (Heinrich, Matt, Lutz-Bonengel, & Schmidt, 2007; Partemi et al.., 2010; Preece & Cairns, 2003). Specifically, one study used qPCR analysis of 11 transcripts (both housekeeping and tissue-specific) and found a correlation between RNA degradation and the PMI in tissue from the femoral quadriceps and liver, but found no correlation in skin, spleen,

pancreas, stomach, and lung tissue (Sampaio-Silva, Magalhães, Carvalho, Dinis-Oliveira, & Silvestre, 2013). The studies that found no correlation used a variety of tissues including, brain, heart, muscle, liver, kidney, and spleen (Heinrich et al.., 2007; Partemi et al.., 2010; Preece & Cairns, 2003). While there does not appear to be a distinct trend in which tissues do show a correlation and which do not, the stability of RNA does appear to vary by tissue type(Heinrich et al.., 2007; Inoue et al.., 2002). Most of the studies that have evaluated RNA degradation as a means for estimating PMI have analyzed tissues for a very short time after death, ranging from 1 to 11 hours with the shortest study and 7 days with one of the longest (Inoue et al.., 2002; Sampaio-Silva et al.., 2013).

While most of the studies that evaluate RNA degradation in relation to PMI examine samples aged for a short time (less than 7 days), there has been initial research done on RNA stability in tissues over an extended PMI (up to several months). Studies by Vass et al. and Young et al. examine the stability of RNA over 120 days and 140 days, respectively (Vass et al.., 2013; Young et al.., 2013). The results of these studies support the possible use of RNA degradation as an estimator of PMI over an extended time. Young et al. performed the only study to date on the behavior of RNA in post-mortem teeth. Researchers buried eight pig heads in the ground, routinely sampled teeth over 140 days and performed qPCR to analyze the abundance of β-actin mRNA (Young et al.., 2013). The PCR assay targeted two separate, non-overlapping regions of the β-actin mRNA transcript, one small amplicon and one large amplicon. Investigators analyzed the differential expression of these segments for 140 days postmortem. However, the large amplicon dropped below the level of detection at 84 days post-mortem. With increasing PMI, larger amplicons generally degrade faster than smaller amplicons, due to

the random nature of RNA degradation. On day 21, a sudden increase in the degradation of the small amplicon and a slowing in the degradation of the large amplicon interrupt the observed linear degradation pattern otherwise seen throughout the study. Thus, this particular assay provides an unreliable estimate of the PMI between days 14 and 28 (Young et al.., 2013). Despite this limitation, the concept of estimating the PMI using RNA degradation within dental pulp still has potential. Vass et al.. performed a similar study, examining nails and rib bones instead of teeth (Vass et al.., 2013). Investigators developed a multiplex PCR assay that evaluates degradation of keratin mRNA, 18S rRNA, and keratin DNA to monitor nucleic acid degradation in post-mortem nail and bone samples up to 120 days. Researchers did observe mRNA degradation in the aged nail samples with the larger keratin mRNA amplicon disappearing faster in the older samples than the smaller mRNA amplicon. However, the observed correlation between mRNA degradation and PMI was only slight, with an R^2 value of 0.21 for the longest keratin mRNA amplicon (i.e. only 21% of the variation in the data is attributable to age of the sample) (Vass et al.., 2013). This study also revealed that environment does have an impact on the rate of RNA degradation, with the larger amplicons disappearing at a faster rate in nail samples stored in soil and water as opposed to those stored in the air (Vass et al.., 2013). While these studies do indicate a possible use of RNA degradation as a predictor of PMI, more research is clearly needed to find the most accurate RNA targets for establishing both short and long-term sample age estimation.

While previous studies provide initial evidence for a direct correlation between RNA degradation and sample age, much of the research has focused on only a few sample types (blood being the major source for studies of deposited samples).

Additionally, the previous studies have been limited to only examining a few RNA markers, mainly focusing on housekeeping mRNA transcripts and 18S rRNA. Furthermore, no research has been performed to confirm that RNA degradation occurs at the same rate and patterns across multiple sample types. This research aims to provide a more in-depth study of RNA degradation, taking into account full transcriptome degradation in a variety of biological fluid types (blood, saliva, semen, and vaginal fluid) to identify the mRNA transcripts that have degradation patterns most closely related to sample age.

CHAPTER III

RNA-SEQ OF AGED BIOLOGICAL SAMPLES ON A NEXT-GENERATION SEQUENCING PLATFORM

Introduction

DNA analysis is routinely applied in both forensic and medical testing to provide information on human identify and genetic disease. Regular DNA testing performed by both forensic and medical personnel today includes fragment analysis (endpoint PCR paired with capillary electrophoresis, qPCR), targeted DNA sequencing, exome sequencing, and whole genome sequencing. While these technologies help investigators gain a wealth of genetic information, the knowledge obtainable through DNA testing can be bolstered by additional evaluation of the RNA present in a sample (Raghavachari et al.., 2012). The value of RNA analysis lies in the reactive and labile nature of the human transcriptome (total RNA in a sample) as opposed to the human genome. The genome of an individual is constant throughout all tissue and cell types, while the transcriptome is variable and unique (Vennemann & Koppelkamm, 2010a). No two cell types or tissues

within an individual will have the same transcriptome. Additionally, no two individuals are likely to have the exact same transcriptome because transcript expression levels fluctuate based on differing biological conditions (such as active disease state, drug use, activity level, trauma) (M Bauer, 2007). While the genome is constant, the transcriptome changes based on what proteins are biologically necessary in each cell and tissue type at any given time. It is this unique and labile nature of RNA that can be exploited to learn more about a sample. However, routine use of RNA analysis in forensic and medical investigations requires an in-depth knowledge of the transcriptome profile of different biological fluids and tissues.

Over the past several decades in the fields of molecular biology and genetics, a huge emphasis has been placed on sequencing the complete human genome. The first release of the sequenced human genome in 2001 has been followed by over a decade of re-sequencing and deep sequencing of the human genome, which still remains to be 100 percent complete. However, while thousands of human genomes have been sequenced in the years since the original human genome sequence was released; full human transcriptomes have not received nearly as much attention (Pertea, 2012). RNA analysis presents a challenge not encountered with the human genome in that there are hundreds of different transcriptomes in every individual. The human body is composed of four major tissue types, 13 organ systems, and 200 different kinds of cells, each with a unique transcriptome. Additionally, because RNA expression is reactive to biological conditions, transcriptomes can be variable between individuals (Pertea, 2012). Obtaining representative total RNA sequence data for every cell, tissue, and fluid type is a massive undertaking, still being pursued by research groups all over the world. While challenging

to obtain, full transcriptome data is a critical first step for investigators who are trying to identify significant RNA biomarkers for forensic or medical application.

Representative full transcriptome sequencing data are available for a large number of fresh biological fluids and tissues. Projects such as the Illumina Human BodyMap provide databases of full transcriptome next-generation sequence (RNA-seq) of a variety of human tissue types. Specifically, the Human BodyMap provides RNA-seq data for 16 human tissues (Thibaut, n.d.). While databases such as this provide an excellent starting point for full transcriptome data, they are far from comprehensive. Notably missing from RNA-seq datasets are transcriptomes for biological samples most commonly encountered in forensic casework. Blood, for instance, is regularly included in transcriptome databases, while semen, saliva, and vaginal fluid are noticeably absent. The lack of information concerning forensically relevant body fluids is made even more evident with a simple Pubmed search. Searching for articles on transcriptome sequencing reveals the following number of hits; 107 articles for transcriptome sequence of blood, ten articles for transcriptome sequence of semen, five articles for transcriptome sequence of saliva, and only one article for full transcriptome sequence of vaginal fluid. These results reveal a clear gap of knowledge concerning the population of RNA in forensically relevant sample types.

Increasing the amount of RNA-seq data for forensically relevant sample types (blood, semen, vaginal fluid, and saliva) is imperative if RNA analysis is going to be fully utilized in forensic analysis. Having representative total RNA-seq data for these sample types will allow investigators to choose the most applicable biomarkers for benefitting forensic investigation. Currently, researchers have identified forensically-

relevant RNA biomarkers for sample identity, time-since deposition/PMI estimation, disease state, and cause of death (molecular autopsy) (S. E. Anderson et al.., 2011; M Bauer, 2007; Vennemann & Koppelkamm, 2010b). Most of these biomarkers have been identified through microarray analysis or literature search paired with conformation with endpoint PCR and capillary electrophoresis or RT-qPCR. While these techniques are appropriate for evaluating a few select RNA targets, they provide a limited picture of the RNA present in a sample. RNA-seq data for forensically relevant sample types would provide investigators with a snapshot of most or all available mRNA species in a given sample type, allowing for the selection of more accurate, sensitive, and specific biomarkers for use with forensic investigation.

RNA-seq is most optimally performed on abundantly available, non-degraded RNA samples (Adiconis et al.., 2013). Unfortunately, samples that are usually of forensic relevance are often times low abundance and degraded. Thus, if RNA-seq is to be employed in forensic research, it is critical that sequencing methodologies be optimized for low input, low quality samples. We recently performed RNA sequencing of total mRNA isolated from fresh and aged biological fluid samples (blood, semen, saliva, and vaginal fluid) in order to monitor transcript degradation rates and patterns. This sequencing was performed in an effort to identify biomarkers for estimating sample age (RNA transcripts that have degradation rates that tightly correlate with sample age). In order to perform this study, an RNA-seq method for use with low input and degraded biological samples was developed. The selected methodology and representative RNA-seq results from aged samples are presented here.

Materials and Methods

Description of Samples

All sample collection, storage, and preparation methods described in this manuscript adhere to the OSU-CHS IRB approved protocol dated May 13, 2013 (See Appendix, “Documentation of IRB”).

This study utilized deposited human biological fluid samples, including, venous blood, saliva, vaginal fluid, and semen. All biological fluids were collected from adults, over the age of 18 with a college level education in science, who provided informed consent to having their samples sequenced. Each of the biological fluids (blood, saliva, vaginal fluid, and semen) was collected from study participants in a specific way. For blood collection, about 10 cc’s of blood was drawn from the participant’s arm. For saliva collection, the participant deposited their sample into a sterile tube provided by the investigator. Semen was obtained by providing the participant with a sterile container for deposition of the sample. For vaginal fluid, the participant received sterile swabs from the investigator for collection of the sample. Upon collection, samples were provided a 10-digit identification code that remained with the sample throughout the sample storage and analysis process. The 10-digit identification number consisted of the date of collection (mmddyy), a one letter symbol for the type of fluid or tissue the sample consists of (B=blood, S=saliva, E=semen, V=vaginal secretion), and the day (000-360) on which RNA was to be isolated from the sample. The date of collection is important to establish the real age of the sample, the one letter symbol is important to identify the true tissue or fluid type, and the day on which the sample RNA will be isolated is important to keep

track of time-course sampling. An example code for male blood that is collected on June 1st, 2013 and sampled after being aged for 30 days is 060113M030.

Sample Storage and Aging

Once collected and labeled, all samples were brought immediately to the lab for RNA extraction or storage under controlled and secure conditions (stored in the dark, in a closed lab cupboard). Blood, saliva, and semen were deposited onto paper cards in 50 µL aliquots and allowed to air dry for storage. Vaginal fluid was stored in the form that it was collected (cotton swabs).

Blood samples were aged for up to 360 days. RNA and DNA were isolated and sequenced at 0 (fresh), 30, 60, 120, 180, 270, and 360 days post-deposition. Saliva, vaginal fluid, and semen samples were aged for up to 180 days. RNA and DNA were isolated and sequenced at 0 (fresh), 60, 120, and 180 post-deposition. All samples were analyzed in duplicate at each sampling time-point. The sample aging and analysis time-course is presented in Table 1. Only blood had periodic sequencing out to 360 days, as opposed to 180 days, due to the expense of sequencing each sample. Evaluating the whole transcriptome of only one sample type out to a full 360 days aided in keeping costs controlled, while still evaluating total mRNA degradation over an extended time. Blood was chosen for extended aging due to its common presence in forensic casework. All samples were analyzed in duplicate, thus two RNA isolations were performed for every sample type at every time point.

Table 1. Aging and sampling time-course for biological samples. An "X" represents a sample that was extracted and sequenced.

Age (Days)	Blood	Saliva	Vaginal Fluid	Semen
0	XX	XX	XX	XX
30	XX			
60	XX	XX	XX	XX
90				
120	XX	XX	XX	XX
150				
180	XX	XX	XX	XX
270	XX			
360	XX			

Co-Isolation of RNA and DNA

After aging for the assigned amount of time, RNA was isolated from each of the sample types in duplicate. RNA isolation was performed under sterile conditions with all utilized equipment being treated with RNaseZap® (Life Technologies, Carlsbad, CA) prior to each extraction. The methodology utilized for nucleic acid extraction (TRI Reagent®, Sigma Aldrich) allows for the possible co-isolation of RNA and DNA. Isolation of both RNA and DNA is important for the feasibility of downstream forensic human identification using DNA.

For biological fluid stains including, blood, semen, and saliva, a cutting approximately 1 cm^2 in size was taken and placed in a 1.5 mL Eppendorf Tube®

(Eppendorf, Hauppauge, NY). For vaginal fluid swabs, the cotton swab was cut off of its stick and placed in a 1.5 mL Eppendorf Tube®.

RNA and DNA isolation was performed using TRI Reagent® (Sigma Aldrich, St. Louis, MO), following manufactures recommended protocol. After isolation with TRI Reagent®, the aqueous phase (containing the RNA) and the cloudy, middle phase (containing the DNA) of each sample were placed in two separate 1.5 mL tubes for nucleic acid clean-up. RNA clean-up was performed utilizing Zymo Research RNA Clean and Concentrator™ kit following manufacturer's instructions (Zymo Research, Irvine, CA). DNase Digestion was performed on each RNA sample using TURBO™ Dnase (Life Technologies, Carlsbad, CA) and following the manufacturers provided protocol. DNA clean-up was performed utilizing Zymo Research DNA Clean and Concentrator™ kit following manufacturer's instructions (Irvine, CA). Once eluted, all samples were quantitated using a Nanodrop ND-1000 microspectrophotometer (Thermo Scientific, Wilmington, DE).

Generation of cDNA

For library preparation, 20 ng of total RNA from each sample was mixed with 4 µl of ERCC spike-in mix1 at a dilution of 1:10,000 (Ambion®). The NuGEN Ovation® RNA-seq System v2 (NuGEN Technologies, San Carlos, CA) was used to generate cDNA from each total RNA sample containing the ERCC spike-in mix, following the manufactures instructions. Upon purification, each cDNA sample was checked for quality and quantity using the Nanodrop ND-1000 microspectrophotometer (Thermo Scientific, Wilmington, DE).

Fragment Library Preparation

All cDNA samples were fragmented with the Bioruptor® UCD 200 (Diagenode, Denville, NJ) using a sonication time of 30 minutes to an average fragment size of 200 bp. Once fragmented, cDNA libraries were generated using the Ion Plus Fragment Library Kit following the manufacturer's instructions (Life Technologies, Carlsbad, CA). Replicates for each sample type at each time point were barcoded using the Ion Xpress™ Barcode Adapters (Life Technologies, Carlsbad, CA) so they could be analyzed on the same Ion 318™v2 chip in downstream sequencing. Generated libraries were quantitated using the Ion Library Quantitation Kit following manufacturer's instructions (Life Technologies, Carlsbad, CA).

Template Preparation

After libraries were constructed and quantitated, template preparation was performed with each library. Template preparation is the process of amplifying individual RNA fragments onto Ion Sphere™ Particles (ISPs) and enriching the sample for template-positive ISPs that can be sequenced on the Ion Torrent PGM™. Template preparation of the cDNA libraries was performed using the OneTouch™ 2 (OT2) instrument and the Ion PGM™ Template OT2 200 kit, following the manufacturer's instructions (Life Technologies, Carlsbad, CA). The sample is enriched for template-positive ISPs with polyclonal and template-negative ISPs being washed away. The template-positive ISPs provide the sequencing template when loaded onto the Ion Torrent PGM™. The process for template preparation and enrichment was the same for all libraries created in this study, regardless of sample type.

Sequencing of the Prepared Samples

Sequencing of the cDNA library fragments was performed on the Ion Torrent PGM™, utilizing the Ion PGM™ Sequencing 200 kit v2 and following manufacturer's instructions (Life Technologies, Carlsbad, CA). The Ion PGM™ relies on semiconductor chip technology to sequence nucleic acid samples in a massively-parallel way. Each sample of template-positive, enriched ISPs was combined with buffer, primers, and enzyme and the total reaction was loaded onto an Ion 318™ v2Chip. The sequencing chip is composed of three layers; the top layer of micro-machined wells is where individual ISPs sit during sequencing (with each well large enough to hold a single ISP), the middle ion sensitive layer, and the bottom layer which consists of proprietary ion sensors. During sequencing, the Ion Chip is sequentially flooded with dNTPs that flow over individual ISPs that are deposited in the micro-machined wells. For this study, we will utilize Ion 318™ v2 chips (Life Technologies, Carlsbad, CA). The Ion 318™ v2 chip has the largest capacity for sequencing on the Ion PGM™. The large capacity is required for transcriptome sequencing. The sequencing methodology was the same for every library created in this study, regardless of sample type.

Analysis of RNA-Seq Data

All raw sequencing reads for a given sample were aligned to the human reference genome, Hg19 (GRCh38). After alignment, every sample had RNA expression levels calculated in the form of reads per kilobase per million (RPKM) using the following equation.

$$RPKM = (\frac{10^9 \times C}{N \times L})$$

RPKM values normalize expression levels by taking into account the total number of sequencing reads in a run (N), the exon length for a gene (L), and the number of sequencing reads that map to that gene (C). RPKM values are a more accurate assessment of expression level than raw sequencing reads as they adjust for fluctuating factors such as the total number of reads in a given run and the different sizes of genes in the genome (Mortazavi, Williams, McCue, Schaeffer, & Wold, 2008). Alignment and RPKM calculations were performed with CLC Bio Genomics Workbench software (Cambridge, MA).

After initial RPKM values are calculated, the RPKM values were normalized a second time to the ERCC spike-in standards (Ambion®), which form a standard curve (input quantity vs. RPKM). The final normalized value for each transcript is expressed in molecules of RNA. This molecule value represents the abundance of each RNA transcript that is present in a given tissue's transcriptome.

Results and Discussion

Utilization of an External Spike-in Standard

Quality assurance measures are critical when comparing RNA sequencing data in a number of sample types over a number of time-points. Variation in RNA expression patterns as measured by RNA-seq analysis can be attributed to differences in a number of factors, including the starting quantity of RNA, quality of RNA, techniques or instrumentation used, and the person performing the analysis. In order to perform comparison of sequence data generated from multiple samples, a control was incorporated into the RNA-seq procedure to normalize procedural variations and provide

a common baseline for data analysis. The External RNA Controls Consortium (ERCC) hosted by the National Institute of Standards and Technology (NIST) has created a set of internal RNA spike-in standards to help control for variation that is inherently present in RNA expression studies (Jiang et al.., 2011). The spike-in controls are a series of unlabeled, polyadenylated transcripts that are present in solution in varying, known molar concentrations. The ERCC control RNA (Ambion®) can be spiked in to RNA samples and be carried through the generation of cDNA, library building, template preparation, and sequencing analysis with RNA extracted from body fluid samples. The ERCC control RNA can be utilized to normalize comparisons of sequence results both within a single sample and between samples. The inclusion of the control RNA in every sample that was analyzed for this project allowed for the confident comparison of expression patterns in different samples. ERCC RNA Spike-in control mix was added to every RNA sample before conversion to cDNA to help ensure correct comparison and interpretation of downstream sequencing results.

Design of an RNA-seq Library Construction Protocol

The first step in sequencing RNA is the preparation of a sequencing library. The library preparation begins with RNA conversion to cDNA. The RNA that is to be converted into cDNA and sequenced must be free of rRNA. Removal of rRNA from the sample ensures that successful sequencing of the much less abundant mRNA population can be achieved. Elimination of rRNA from the total RNA sample is traditionally accomplished by either rRNA depletion or poly-A selection. In addition to elimination of rRNA and generation of cDNA, the library preparation methodology also includes

subsequent fragmentation of the cDNA into pieces of a known size (for this study, 200 bp libraries were created) by mechanical or enzymatic shearing. Following fragmentation, the cDNA fragments have adapter oligonucleotides ligated onto both ends. The adapter sequences are necessary for the library to undergo template preparation and sequencing. The sample is then size selected to insure that the final library contains only cDNA fragments of one consistent size (200 bp). While this described process includes the basic steps that must be present in every cDNA library preparation, there are multiple methods of pursuing each step and creating the final cDNA library.

For this study we evaluated two different library building methods that are compatible with the Ion Torrent™ System. Methods were compared on the basis of RNA input requirement, elimination of rRNA, sequencing bias, and time requirement. An optimized RNA library preparation method was imperative to the success of downstream sequencing and analysis of degraded samples, thus selection of cDNA conversion and library preparation methods was critical. Table 2 presents a comparison of the evaluated cDNA conversion and fragment library preparation methods.

Table 2. Comparison of cDNA Generation and Library Preparation Methods

	Kit	Input Requirement	rRNA Depletion	Poly-A Selection	Sequencing Bias Introduced	Time
cDNA Generation	Ion™ Total RNA-seq Kit v2	1 ng – 500 ng (Poly-A Selected or rRNA Depleted)	Yes (or poly-A selection)	Yes (or rRNA Depletion)	If poly-A selection: 3' bias	6 hrs
	NuGEN Ovation ® RNA-Seq System v2	500 pg – 100 ng (Total RNA)	No (SPIA Amplification)	No (SPIA Amplification)	No 3' Bias	4.5 hrs
Library Preparation	Ion™ Plus Fragment Library Prep	100 ng or 1 µg	N/A	N/A	N/A	2.5 hrs

The NuGEN Ovation® RNA-seq System for conversion of whole RNA into cDNA was found to consistently produce >5 µg of cDNA from an input of 20 ng of RNA (both from fresh samples and samples that had been aged up to one year). In addition to generating consistent amounts of cDNA from both fresh and aged RNA samples, the NuGEN Ovation® kit was also desirable because it required whole RNA for input, rather than rRNA depleted or poly-A selected samples. When dealing with degraded RNA samples, poly-A selection and rRNA depletion procedures are not ideal. Poly-A selection introduces distinct bias into a sample by only converting RNA fragments that contain a poly-A tail to cDNA. It is likely that in a degraded sample of RNA, many of the mRNA fragments will no longer be attached to a poly-A tail, thus during cDNA conversion much of the sample would be lost. Ribosomal RNA depletion is known to introduce degradation into an RNA sample. In a sample population that is already degraded due to

age, we did not want to further subject our samples to degradation via rRNA depletion. While the Ion™ Total RNA-Seq Kit v2 requires rRNA depletion or poly-A selection of RNA samples, the NuGEN Ovation® kit does not. Rather, the NuGEN Ovation® kit utilizes a SPIA™ amplification process to deplete the total RNA sample of rRNA. SPIA™ Amplification relies on a mix or poly-A primers and not-so-random random primers to selectively amplify mRNA in the cDNA conversion process, therefore depleting the sample of rRNA. The NuGEN Ovation® kit's low starting requirement of total RNA paired with the lack of an rRNA depletion or poly-A selection step made it ideal for use with the aged samples required in this study.

For library preparation, Life Technologies (Carlsbad, CA) Ion™ Plus gDNA Fragment Library kit, offered the most applicability for this study. The Ion™ Plus kit can be combined with the NuGEN Ovation® kit (our preferred way of producing cDNA) and can be utilized for production of cDNA or gDNA libraries, allowing our lab to streamline all library production into one workflow.

Optimization of a Fragmentation Protocol

Optimization of a 200 bp fragmentation protocol on the Bioruptor® UCD200 was performed by fragmenting 1µl aliquots of cDNA in 50 µl of low TE for varying amounts of time (10 minutes, 20 minutes, 30 minutes). The Bioruptor® was run on the low setting, 30 seconds on, 30 seconds off, with ice replenished every 10 minutes during the sonication. Once complete, samples were electrophoresed on an agarose gel to determine the fragment size range generated by each amount of sonication time. Results can be seen

in figure 1. The optimum sonication time that generated an average cDNA fragment size of 200 bp was found to be 30 minutes.

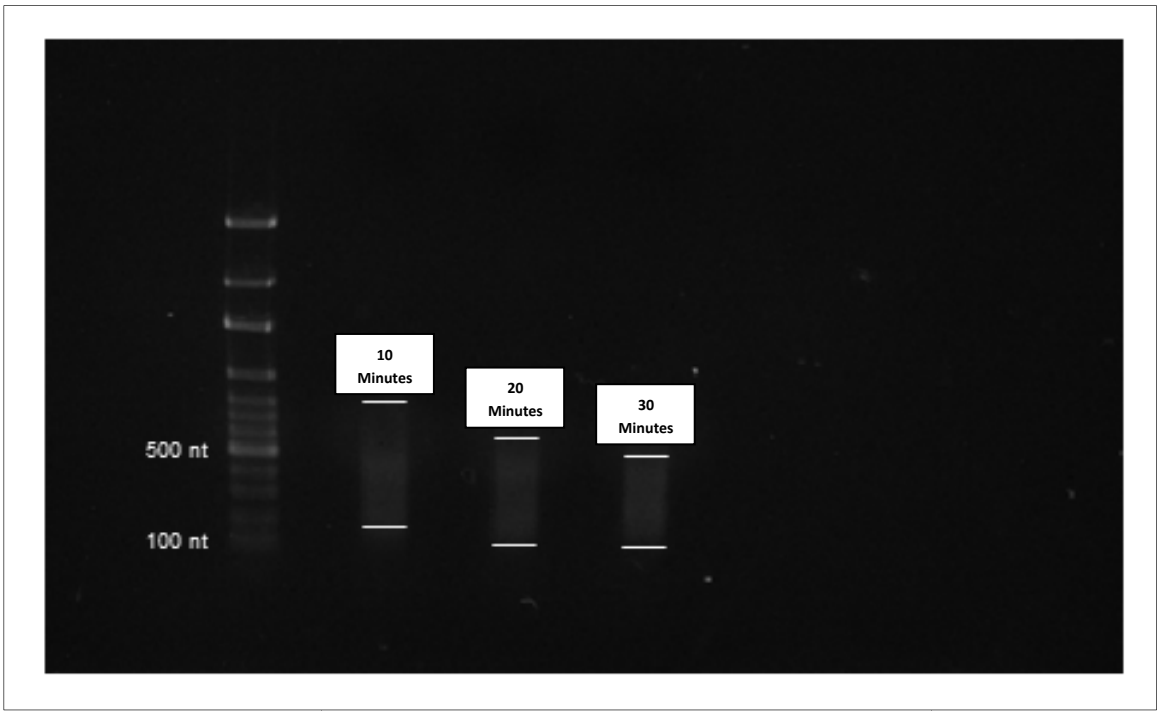


Figure 1. Image of Gel Demonstrating Fragmentation with Bioruptor. Aliquots of 50 µl of low TE containing 1 µg of cDNA were placed in the Bioruptor on the low setting for 10, 20, and 30 minutes. The optimum sonication time for producing an average cDNA fragment size of 200 bp was found to be 30 minutes.

RNA-seq on the Ion PGM™

The Ion Torrent™ PGM server provides a summary of run statistics for every sample that is sequenced. Reviewing run statistics provides information that reflects not only on the success of the OT2™ run and Ion PGM™ sequencing, but also on the quality of the input library samples. Ensuring that a library building protocol and subsequent emulsion PCR and sequencing are able to generate a high quality and usable amount of sequencing reads is imperative when evaluating a sequencing methodology.

The Ion PGM™ is not intended for whole transcriptome sequencing and does not offer the sequencing capacity to account for sequencing a whole human transcriptome. For this reason, it was imperative that we establish a sequence methodology on the Ion PGM™ that facilitated generation of as much mRNA sequencing data as was possible from each sample.

To determine the importance of output quantity of sequencing reads on the quality of sequencing data, the same cDNA library generated from RNA isolated from a fresh (Time 0) blood stain was sequenced on both an Ion 314™ sequencing chip and an Ion 318™ sequencing chip v2 (figure 2). The Ion 314™ chip has an average sequencing data output of 400 to 550 thousand reads per run, while the Ion 318™ chip has an average sequencing data output of 4 to 5.5 million reads per run. As can be seen in figure 2, the quality of each of the test sequencing runs that used the same blood cDNA library was high, with sufficient chip loading (over 80% in both chips), sufficient usable sequence, and the desired targeted fragment length of around 200 bp (202 bp for each of the reactions). While both runs were of usable quality, the main difference lies in the output of data from each chip type. The 314™ chip (figure 2A) produced 111M bases of data and 639,020 total sequencing reads. The 318™ chip (figure 2B) produced a substantially larger amount of data, with 1 G total bases and 5,630,598 total sequencing reads. These numbers indicate that sequencing the library on a 318 chip as opposed to a 314 chip produced approximately 10 times more data.

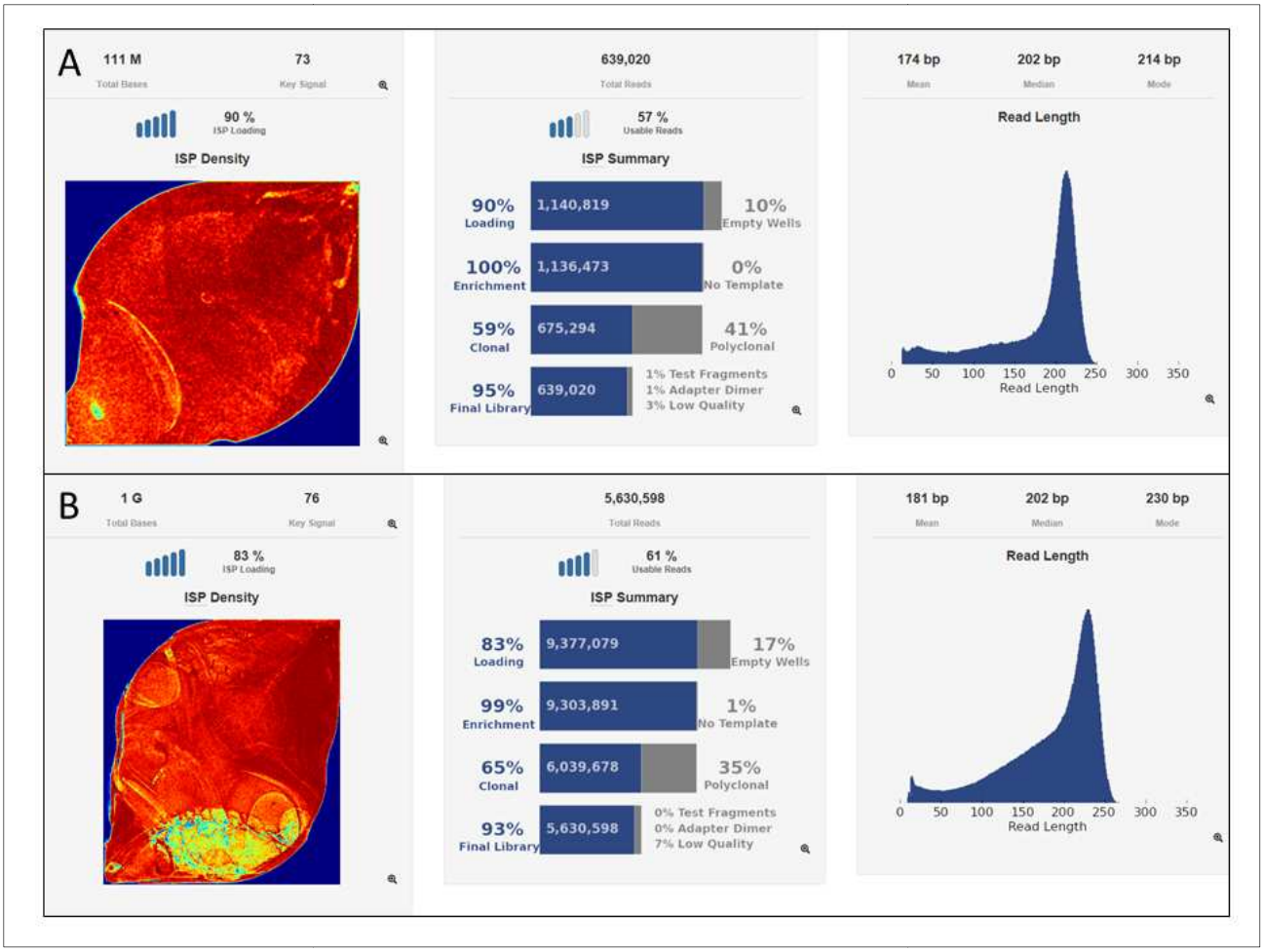


Figure 2. 314 and 318 chip comparison. **A.** Ion Server sequencing run report for a time 0 blood library sequenced on a 314 chip. **B**. Ion Server sequencing run report for a time 0 blood library sequenced on a 318 v2 chip. The Ion server run report includes a summary of chip loading, sequencing reads produced, and the average sequencing read length.

While these sequencing runs demonstrates that sequencing library and run quality appear equal with both the 314 chip and 318 chip, they also clearly show that there is a large increase in data output when sequencing on a 318 chip. For further evaluation of the effect of sequencing on a 314 chip vs. a 318 chip, the sequencing reads from each run were analyzed. The data obtained from both runs were trimmed for quality, aligned to the human genome (HG19), and RPKM values were calculated for each gene. Upon

comparison of these data, the affect that useable sequence output has on the analysis becomes clear (Table 3).

	Number of Sequencing Reads	% of Reads Aligned to HG19	Number of Genes Detected (RPKM >1)
000Blood Ion 314™ Chip	533,244	86%	8,137
000Blood Ion 318™ v2 chip	3,212,785	87%	12,116

Table 3. Summary of data generated by sequencing the 000Blood library on a 314 chip and a 318 v2 chip

As would be expected, the percentage of sequencing reads that aligned to HG19 was similar for each of the runs (86% for 314 and 87% for 38 v2). However, a vast difference is observed when the number of genes detected in each run is considered. The presence of about 10 times more sequencing reads from the Ion 318™ v2 chip as compared to the Ion 314™ chip, allowed for the detection of several thousand more genes (12,116 genes detected in the 318 v2 run, 8,137 genes detected in the 314 run). Specifically, there were 48.9% more genes detected with the sequencing reaction run on the 318 v2 chip when compared to the 314 chip sequencing reaction. The sequencing performed on the higher capacity 318 v2 chip is therefore a more complete representation of the blood transcriptome than the data obtained for the same library sequenced on the 314 chip. Thus, when attempting to get the most complete picture of the transcriptome while sequencing on the Ion PGM™, the 318 v2 chip should be used.

As the data generated from our developed library building procedure, OT2™ reactions, and chosen sequencing methodology will ultimately be used to evaluate and

compare the mRNA populations of multiple forensically relevant sample types, having the most complete picture of the transcriptome as possible is absolutely critical. All RNA-seq that aims to provide insight on the profiles and behaviors of sample transcriptomes should therefore be performed on 318 v2 chips in order to maximize the number of transcripts observed in the population.

RNA-seq Data Analysis Workflow

Once raw sequencing reads are obtained from the Ion PGM™, they undergo a multi-step data analysis process. In the first step of sequencing data analysis, all raw sequencing reads for a given sample are aligned to the human reference genome, Hg19 (GRCh38). After alignment, every sample will have RNA expression levels calculated in the form of reads per kilobase per million (RPKM). RPKM values normalize expression levels by taking into account the total number of sequencing reads in a run, the size of the gene, and the number of sequencing reads that map to that gene. RPKM values are a more accurate assessment of expression level than raw sequencing reads as they adjust for fluctuating factors such as the total number of reads in a given run and the different sizes of genes in the genome (Mortazavi et al.., 2008). Alignment and RPKM calculations are performed with CLC Bio Genomics Workbench software (Cambridge, MA). After initial RPKM values are calculated, the RPKM values will be normalized a second time to the ERCC spike-in standards, which form a standard curve (known input quantity of spike-in transcripts vs. RPKM). An example ERCC standard curve is displayed in figure 3. Normalization to the spike-in standard acts as a control for any variation that might have been introduced by sample preparation or user error, as the

ERCC standards are spiked-in to each sample individually at a known molar concentration (Jiang et al.., 2011). The ERCC spike-in standards form a standard curve that can be utilized to normalize the RPKM values of each individual transcript in a sample. The final normalized value for each transcript is expressed in molecules of RNA. This molecule value reflects the abundance of each RNA transcript that is present in a given tissues transcriptome.

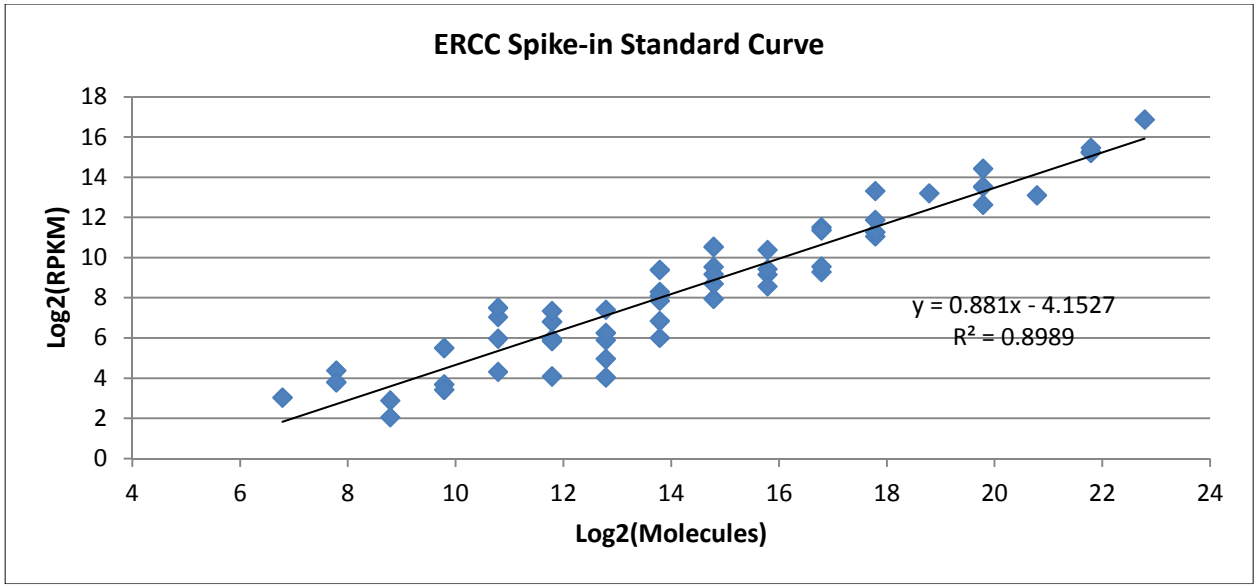


Figure 3. This graph displays the standard curve created by the external ERCC control that is spiked-in to each individual RNA sample prior to cDNA generation and library preparation. The ERCC control consists of 92 transcripts present in varying molar concentrations that, when sequenced, form a standard curve (input molecules vs RPKM).

Assessment of RNA-seq Data

Once sequencing reads have been aligned to the human reference genome and abundance values for each detected transcript have been normalized, the data can be further analyzed. In the development and optimization of this RNA-seq protocol for

analyzing total mRNA from aged biological fluid samples, the data were evaluated on multiple factors including alignment to the human genome, reproducibility between technical replicates, sequencing bias, and sequence capture of fresh and aged RNA samples.

Alignment of RNA-seq Data to a Reference

In a sequencing reaction that generates high quality sequencing reads, the majority of the reads should align to the established reference genome (if the sample is obtained from a single, known source). As all of the samples utilized in this study were single source human samples (blood, saliva, semen, and vaginal fluid), the majority of the reads were expected to align to the human reference genome. As expected, the majority of the reads for both blood and semen samples aligned to the human reference genome (HG19). On average, 87% of the sequencing reads for blood samples and 86% of the reads for semen samples aligned to HG19. However, unlike blood and semen, saliva and vaginal fluid did not align well to the human reference genome. On average, only 7% of the sequencing reads for saliva and 8% of the sequencing reads for vaginal fluid aligned to HG19. The drastic difference between the percent of reads aligning for blood and semen as opposed to saliva and vaginal fluid can be largely attributed to the significant amount of microbial RNA present in saliva and vaginal fluid samples. All unaligned reads from the saliva samples were assembled into contigs and aligned to the Human Oral Microbiome Database (HOMD) which consists of genomic sequences for over 400 oral microbial species. When aligned, on average over 90% of contigs (previously unaligned sequence reads) within a saliva sample mapped to over 340 microbial organisms (Figure

4). Similar results were achieved for the unaligned reads from the vaginal fluid samples. For vaginal fluid, the unaligned reads were assembled into contigs and aligned to the RefSeq genomic database. On average, over 90% of the vaginal fluid contigs (previously unaligned sequence reads) aligned to over 230 microbial organisms. Once aligned to the human genome and to the prospective sequence databases (HOMD and RefSeq) over 85% of reads in every sample were accounted for in an alignment.

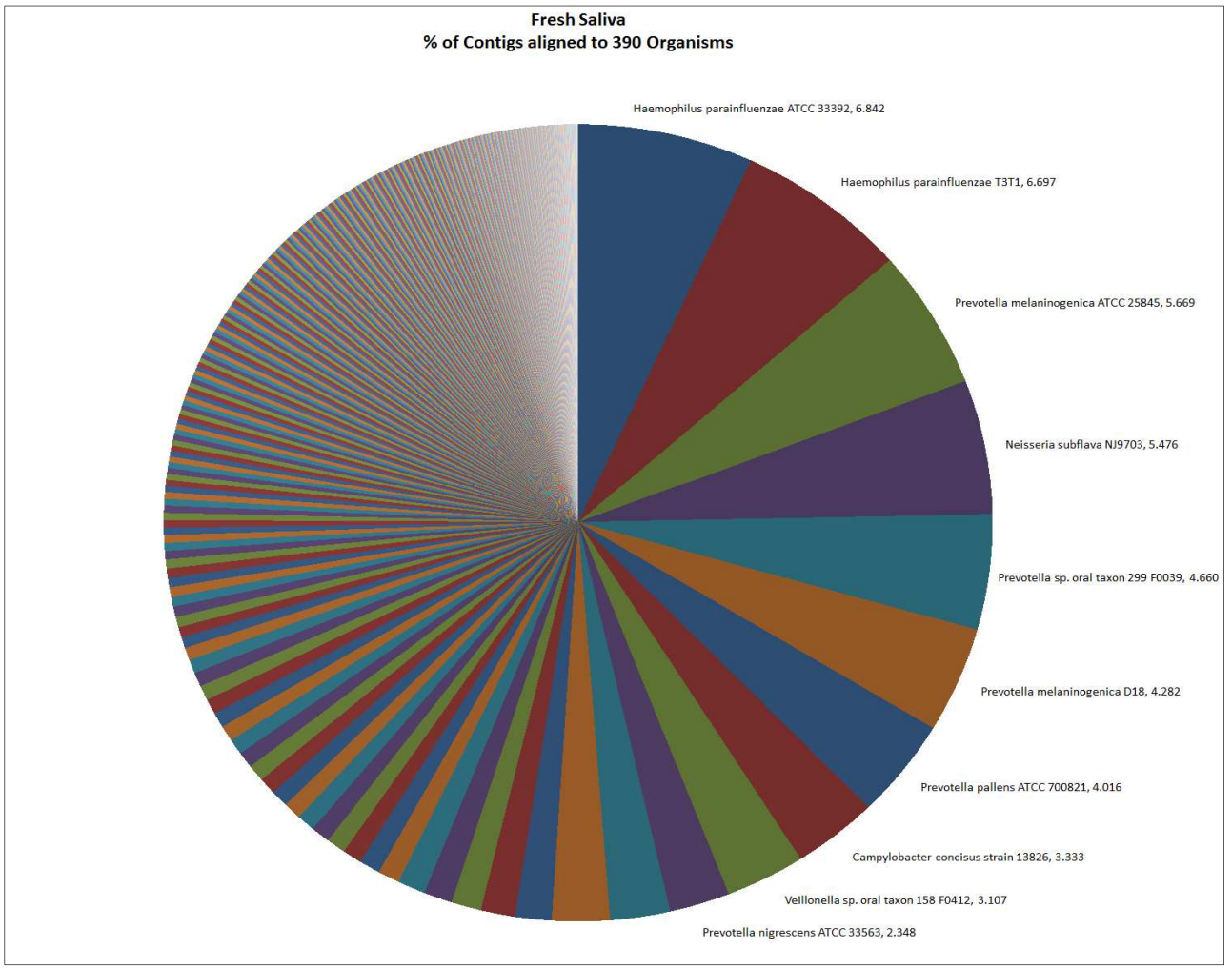


Figure 4. RNA data alignment shows presence of bacteria in saliva. This graph represents the percent of contigs aligned to 390 organisms in the Human Oral Microbiome Database (HOMD). The top ten organisms with aligned contigs are labeled. The other 380 organisms are represented by a wedge in the chart corresponding to their abundance level, but are not labeled.

Determining Reproducibility between Technical Replicates

Two technical replicates for each sample type at each of the time points were analyzed using the library building and sequencing methodologies described in this manuscript. Technical replicates are important for establishing the reproducibility of the methodology and increasing confidence in the sequencing results. Having a high degree of similarity between two technical replicates can increase the confidence of any conclusions drawn from the data. Reproducibility between technical replicates was of high importance for this study because whole transcriptome sequencing is not possible on the Ion Torrent PGM. Rather, with every sequencing reaction, a representative population of the mRNA in a sample library was being sequenced.

To determine if the representative populations of mRNA sequenced in each technical replicate were similar, comparisons were performed between $\log_2$RPKM values for each replicate in a sample pair (Figure 5 and Figure 6). For each sample pair, a graph was generated with a single plotted point for each gene represented in the sample population (with the x-value of that point corresponding to the abundance of that gene in replicate one and the y-value of that point corresponding corresponds to the abundance of that gene in replicate two). If a given gene has the same abundance (Log2RPKM) in both of the replicates, the point for that gene will have the same x- and y-values. When the data are plotted in this way, if two samples are perfect replicates of one another, you would generate a straight line with an R value of 1.00. The data generated in this study were highly reproducible, with replicates for each of the sample types generating an R value of more than 0.80 (Figure 5). Technical replicates of fresh blood, semen, and vaginal fluid all generated plots with an R value of 0.99. Technical replicates of fresh

saliva generated a slightly lower R value of 0.84. High reproducibility between technical replicates was also observed in samples that were aged. Figure 6 displays data for technical replicates of fresh blood and blood that was aged 60, 120, 180, 270, and 360 days. An R value of 0.99 is found in technical replicates of every age, demonstrating that degraded RNA does not affect the reproducibility of the sequence data.

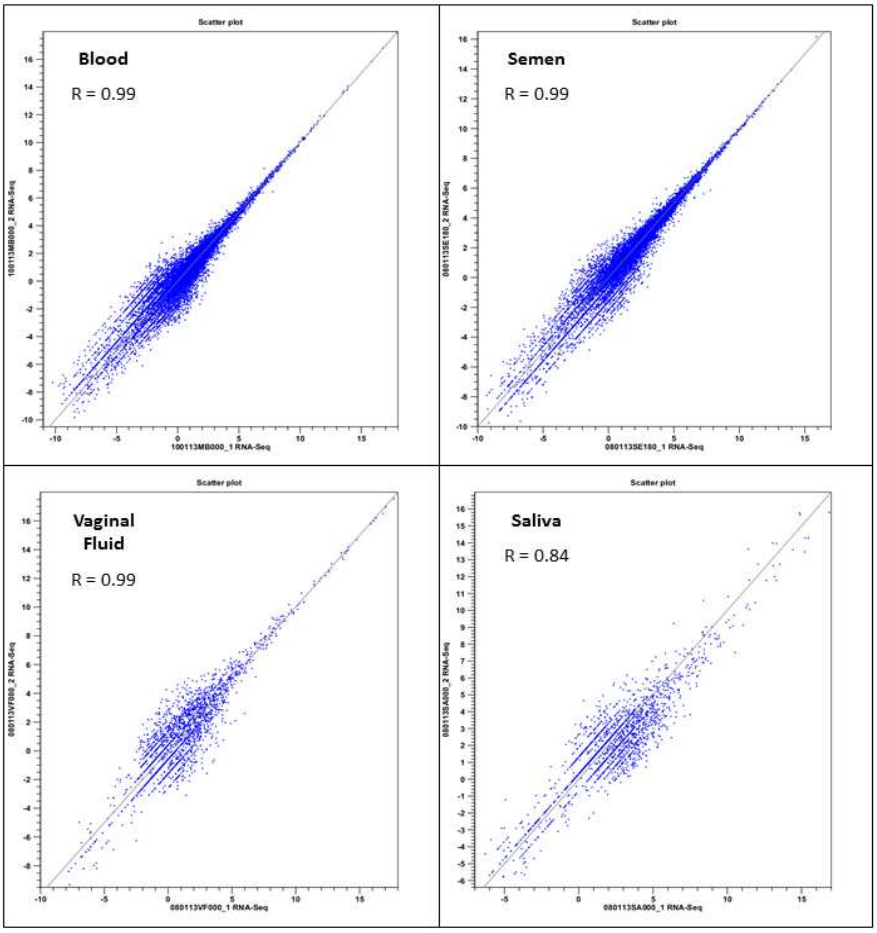


Figure 5. Replicate correlation in study sample types. Log_2 RPKM values for every gene in each of the time 0 (fresh) replicates for each body fluid were compared. Replicate 1 for each sample type is on the x-axis and replicate 2 is on the y-axis. If two replicates have the exact same abundance for a specific gene, the point for that gene would fall on the line. The R value for each of the sample types is displayed on the graph. The closer to 1 an R value is, the more tightly reproducible the replicates are.

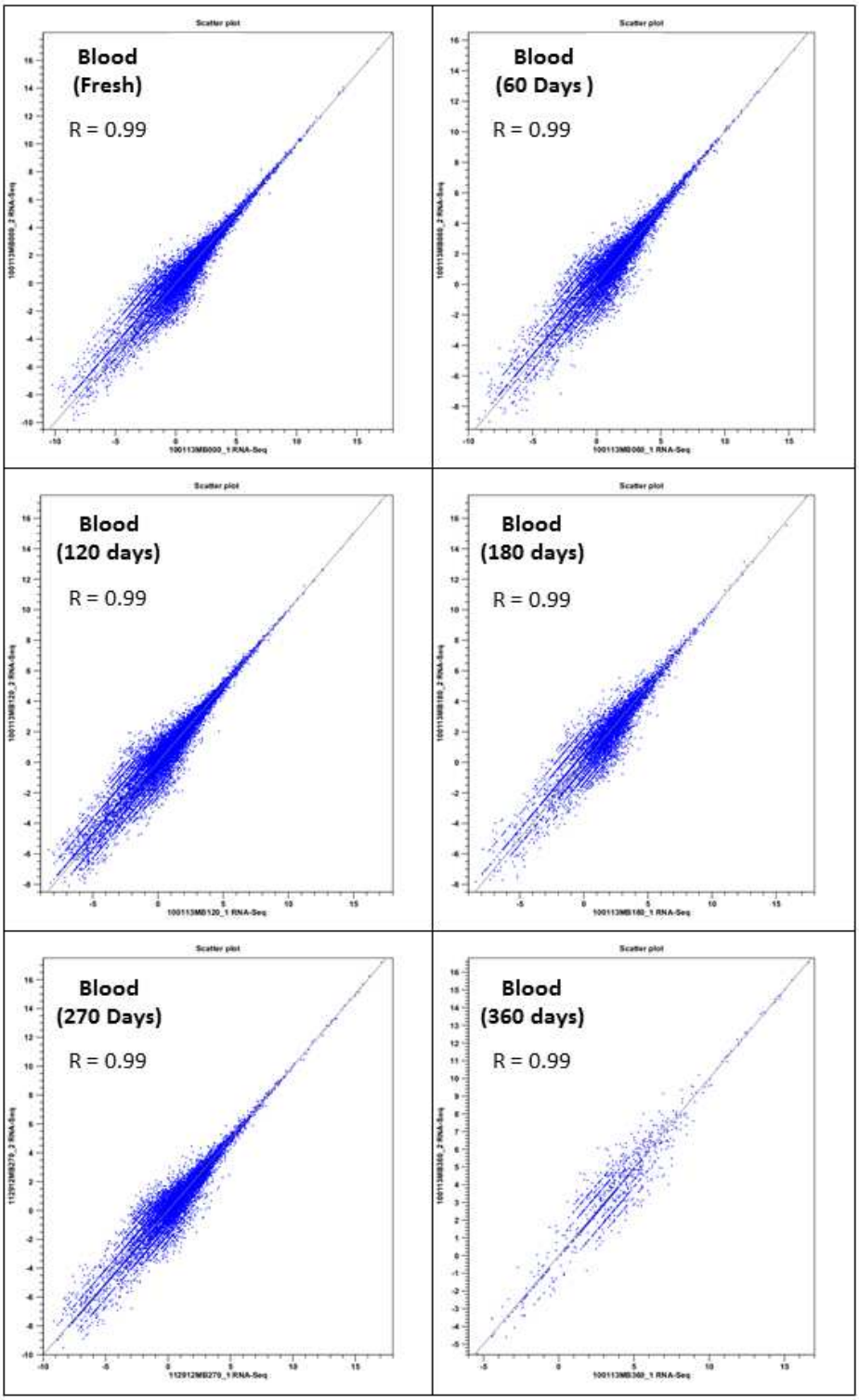


Figure 6. Replicate correlation in aged blood specimens. Log_2 RPKM values for every gene in each of the replicates for blood at every sampled age were compared. Replicate 1 for each sample is on the x-axis and replicate 2 is on the y-axis. The R value for each of the sample types is displayed on the graph. A tight correlation between replicates (R=0.99) is observed with every sample time point.

Assessing Sequencing Bias

With any RNA-seq method, bias is present. This is partly because bias is inherent in the preparation of all RNA sequencing libraries. While bias can never be completely eliminated, it can be reduced. Bias can potentially be introduced at a couple of points in the preparation of libraries, including cDNA generation and fragmentation. Bias introduced through the selection of mRNA from a population of total RNA was considered when outlining the library building protocol described in this manuscript. The NuGEN Ovation RNA-seq kit v2 was chosen, in part, because it utilizes SPIA technology to convert mRNA to cDNA using specialized amplification primer mixes. Alternative cDNA preparation methods include pre-treatment of the total RNA extract using rRNA depletion or poly-A selection. The SPIA primers (a mix of not-so-random random primers and oligo-dT primers) help alleviate the bias observed with poly-A selected RNA. Libraries generated from poly-A selected RNA generally produce data that have a 3' bias, as the cDNA is generated from the 3' end of the transcript. When only priming from the 3' end of a transcript for cDNA conversion, the whole transcript does not always get converted to cDNA. Thus, with the use of only poly-A primers you end up with a cDNA library that has favored only those transcripts containing a poly-A tail and is biased towards the 3' end of mRNA molecules. With SPIA primers, the addition of the random primers helps more evenly capture the entire mRNA population in the cDNA conversion, thus partially alleviating the issue of 3' bias. Additionally, the SPIA random primers were important so our libraries were not selective against mRNA molecules that do not contain a poly-A tail, as would be found more abundantly in aged or degraded samples.

Sequence data were evaluated for bias by surveying the depth of sequencing reads across the entire length of a transcript. If 3' bias is a large issue, sequencing reads will align to the most 3' exons, while few or no reads will be present on the exons that are more 5' in their location. Figure 7 displays sequencing reads for the hemoglobin B gene (HBB) aligned to the human genome. When observing sequencing reads aligned to the HBB reference gene, it is clear that there are sequencing reads aligning all the way across the reference, accounting for each of the exons. While sequencing depth does appear deeper on the most 3' exon when compared to the most 5' exon, sequencing depth for the sample is greatest on the center exon (exon 2). While there is minor 3' bias observed, sequencing reads are clearly spread across the entire transcript length with some clustering in the center of the transcript. The HBB transcript in figure 7 is representative of read depths observed for other transcripts.

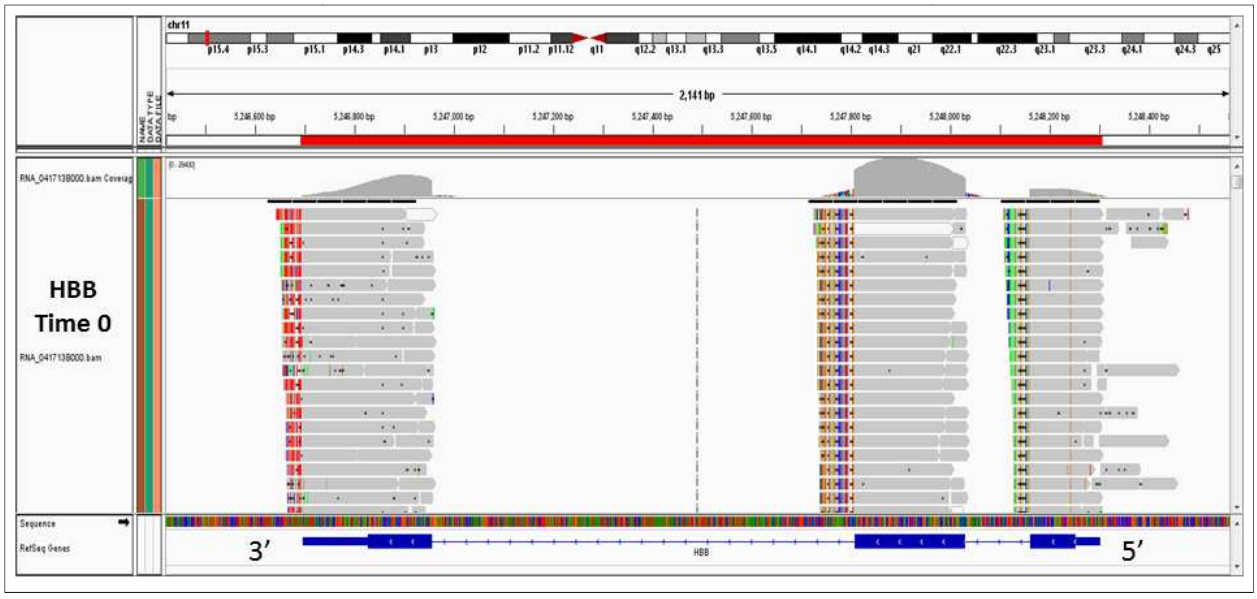


Figure 7. Alignment and Bias in Sequencing Reads. Sequencing reads for the fresh blood sample aligned to the hemoglobin B (HBB) gene. The sequencing reads align to all three exons in the transcript, with the most reads aligning on the center exon. The spread of reads across the entire length of the gene (all exons) is indicative of reduced 3' sequencing bias.

Sequence Obtained for Fresh and Aged Samples

Due to the fact that the sequencing methodology discussed in this manuscript was specifically developed for use with forensically relevant sample types, it is critical that these methods generate sequence data for both fresh and aged samples. The previously described methods were used to sequence both fresh and aged blood, saliva, vaginal fluid, and semen samples. Quality assurance steps were included throughout the library building procedure to ensure sufficient quantity and quality of the libraries being constructed from both fresh and aged samples. Upon isolation, all RNA samples were quantitated using the Qubit® RNA HS Assay following the manufacturer's protocol (Life Technologies, Carlsbad, CA). For all samples, 20 ng of RNA was converted to cDNA and all cDNA was checked for purity (A260/280>1.8) and quantity using the Nanodrop ND-1000 microspectrophotometer. Once cDNA libraries were constructed, all libraries were quantitated using the Ion Library Quantitation kit on the Applied Biosystems® ABI 7500 qPCR instrument to ensure ample quantity for sequencing (calculated dilution factor > 1.0).

Sequence data were generated for fresh samples and samples aged up to one year. Once sequencing for each sample was complete, data were evaluated for every sample based on total number of usable reads, alignment to the human genome, and the number of genes detected (Table 4). Every sample that was sequenced generated more than 1 million sequencing reads, except for a couple of the oldest samples (vaginal fluid 180 day sample and blood 360 day old sample). In the samples that received less than 1 million reads, the lower sequencing output could be due to lower library quality due to the degraded state of the RNA in those samples. All samples aligned to the human genome as

expected, with blood and semen aligning over 80% in all samples, saliva aligning on average less than 10%, and vaginal fluid aligning less that 15% (saliva and vaginal fluid have lower alignment to HG19 due to a large microbial presence in these sample types). It is interesting to note that, in every type of biological fluid examined, the older samples aligned less to the human genome than the fresh samples. This decrease of alignment over time could be due to reduced library quality with degraded RNA or could be attributed to an increasing microbial population as stains aged. Reduced alignment to the human reference genome and increased alignment to microbial genomes as stains age may be an interesting area for future investigation. In addition to an observed decrease with alignment to HG19 in aged samples, there also appears to be a decrease over time for each fluid in the number of genes detected in the sample. This observed decrease in the number of genes detected is most likely due to certain transcripts in the population dropping below the sequencing detection threshold by degrading into RNA fragments smaller than the 200 BP library size. If an mRNA molecule is cleaved in the degradation process and ends up with a fragment size smaller than 200 BP, that transcript will be under represented in the sequencing library. As samples age, it is likely that a greater number of transcripts have degraded to a point below the detection of this sequencing methodology, thus the sequencing reads align to a fewer number of genes in aged samples. Ultimately, this logic is the approach we have taken to examine mRNA degradation.

While Table 4 does demonstrate that there is an observable decrease in the percent of alignment to HG19 and in the number of genes detected in aged samples, the

RNA isolated from both fresh and aged samples was of sufficient quantity and quality to generate usable sequencing data for every sample that was analyzed.

Table 4. Evaluation of sequencing data for fresh and aged biological fluid samples

	Total Number of Sequencing Reads	% Alignment to HG19	Number of Genes Detected
Blood (Fresh)	3,212,785	87%	12,116
Blood (Aged 30 Days)	2,467,087	82%	11,831
Blood (Aged 60 Days)	1,830,68	82%	11,702
Blood (Aged 120 Days)	2,397,120	86%	10,397
Blood (Aged 180 Days)	1,195,493	82%	8,918
Blood (Aged 270 Days)	2,497,672	73%	8,838
Blood (Aged 360 Days)	510,998	75%	1,890
Saliva (Fresh)	3,132,727	7%	4,201
Saliva (Aged 60 Days)	3,241,140	6%	1,510
Saliva (Aged 120 Days)	2,577,010	6%	923
Saliva (Aged 180 Days)	1,341,420	4%	520
Vaginal Fluid (Fresh)	1,662,114	18%	4,070
Vaginal Fluid (Aged 60 Days)	2,824,657	11%	2,999
Vaginal Fluid (Aged 120 Days)	1,252,619	12%	2,112
Vaginal Fluid (Aged 180 Days)	884,564	10%	1,618
Semen (Fresh)	3,051,682	85%	12,300
Semen (Aged 60 Days)	1,536,882	86%	11,874
Semen (Aged 120 Days)	1,567,875	84%	10,937
Semen (Aged 180 Days)	2,338,402	81%	10,800

Conclusions

Appropriate library building protocols and sequencing procedures must be considered for analysis of low input, degraded samples, if RNA analysis is going to be explored more thoroughly for applications in forensic science. RNA analysis is being

considered more heavily than ever before for use with forensic investigation, as RNA biomarkers are being discovered for tissue identification, estimating sample age, and molecular autopsy purposes. If RNA analysis is going to be seriously investigated for use in routine casework, having an understanding of how total mRNA behaves in both fresh and aged samples is critical.

With this study, we have developed a methodology for transcriptome sequencing of RNA isolated from fresh and aged forensically relevant biological samples. Using this protocol we have generated the first base dataset of mRNA profiles for fresh and aged biological fluids (blood, saliva, semen, and vaginal fluid). With this methodology and this first dataset, investigators can begin to establish a broader understanding of the behavior of mRNA in deposited samples, allowing for the proper selection of biomarkers for investigative purposes.

CHAPTER IV

TIME-DEPENDENT LOSS OF TRANSCRIPTS IN AGED FORENSIC SAMPLES

Introduction

DNA analysis is routinely used in forensic casework to identify individuals who were at a crime scene or associated with evidence. However, while DNA can provide identify to an investigator, identification alone cannot always provide context to an investigation. While DNA has been the gold standard of forensic molecular investigation for several decades, recent advancements in RNA analysis may provide a role for RNA in forensic casework. Several recent studies have demonstrated the possibility that RNA analysis has an important role in body fluid identification, molecular autopsy, and also perhaps in suggesting a timeline for the deposition of a biological sample at a crime scene and/or the post-mortem interval (PMI) (S. E. Anderson et al.., 2011, 2011; Sampaio-Silva et al.., 2013). Based on this work, RNA analysis holds clear potential to contribute significantly to the investigation of forensic matters.

The stability of the transcriptome in degraded or minimally available biological samples has been of concern for the widespread use of RNA analysis in forensics. RNA was once considered difficult or even impossible to access in degraded or limited samples

due to its fragile, single stranded structure. However, research has proven that with enhanced analytical methods, accessing RNA in aged, degraded, and minimally available forensic samples is possible (Martin Bauer & Patzelt, 2008; Zubakov et al.., 2010). It has also become clear that RNA may not be as unstable as was once believed, with RNA being detected in forensic samples that are decades old (Kohlmeier & Schneider, 2012).

The first published investigation of *ex vivo* mechanisms of RNA degradation was published in 2013 (Fordyce et al.., 2013). In this paper, Fordyce et al.. discuss that the degradation of RNA in *ex vivo* samples depends largely on sample type and sample condition. Cellular ribonucleases (RNases) that remain active in moist cellular material drive RNA degradation in fresh post-mortem samples that are not preserved or dried. However, in samples that are dried (such as dried blood stains) or preserved (such as FFPE tissue samples), RNases are largely inactivated resulting in RNA degradation that is driven mostly by physical and chemical factors (such as sunlight or pH) (Fordyce et al.., 2013). Virtually all studies that have analyzed forensic samples for mRNA fragments have shown that RNA is much more stable in biological samples than was once assumed. A logical extension of this research has questioned if the degradation of mRNA molecules *ex vivo* occurs at a steady rate such that transcript decay could act as a "biological clock" (S. E. Anderson et al.., 2011; Vass et al.., 2013). Researchers evaluating sample age or PMI based upon RNA degradation have mainly focused their analyses on ribosomal RNA (rRNA), housekeeping mRNA transcripts, and tissue-specific mRNA transcripts (S. E. Anderson et al.., 2011; Martin Bauer, Gramlich, et al.., 2003; Vass et al.., 2013). Such studies have utilized both end-point PCR paired with

capillary electrophoresis and real-time reverse transcriptase PCR to monitor degradation rates in transcripts originating from a few selected genes.

Early work on the assessment of RNA degradation in relation to sample age focused on the degradation rates of housekeeping mRNA transcripts and rRNA, as these species have a known presence in all tissue types. In work performed by Bower et al., analysis of 106 bloodstains, aged up to 15 years, revealed that the abundance of β-actin and cyclophilin transcripts decreased in relation to sample age (Martin Bauer, Gramlich, et al.., 2003). Anderson et al. expanded research on β-actin mRNA degradation by demonstrating that the approximate age of a bloodstain can be predicted by determining the ratio between β-actin mRNA and 18S rRNA (S. Anderson et al.., 2005). Anderson, et al. further evaluated RNA degradation by examining different amplicon sizes of βactin mRNA and rRNA in aged samples (S. E. Anderson et al.., 2011). Results of this study indicated that large RNA amplicons disappear at a faster rate than small amplicons in aged samples. Although these initial studies on RNA degradation in aged bloodstains have been limited to examining a few selected RNA transcripts, the results do indicate a relationship between sample age and RNA degradation rates.

In addition to the limited number of RNA markers evaluated, past research on RNA degradation in deposited biological fluids has been narrowly focused on the number of sample types considered in degradation analysis. Past studies have focused mainly on blood, with no major research performed on other forensically relevant sample types (such as semen, saliva, and vaginal fluid). Evaluation of biological fluid types other than blood is critical as the cell types, cellular environments, and transcriptomes vary considerably with each fluid type. Thus, RNA degradation patterns and rates may be

different in each sample type as well. If investigators are going to be able to evaluate time since deposition in a variety of sample types, it is critical to study RNA degradation patterns and rates in other forensically relevant biological sample types.

In this research, we aimed to provide a more comprehensive study of *ex vivo* RNA degradation in dried body fluid stains (specifically, blood, saliva, vaginal fluid, and semen). In examining RNA degradation, this study took a different approach than previous studies by subjecting the total mRNA of fresh and aged samples to analysis through the use of next-generation sequencing. The results of RNA-seq provide the first ever comprehensive picture of mRNA presence in both fresh and aged biological fluid stains. These data facilitate evaluation of the changing profile of the mRNA population within a deposited sample over time. Based on these data, degradation rates and profiles for every individual transcript within the mRNA population of a sample can be determined. Furthermore, differences in mRNA degradation rates and profiles between sample types can be established. While these data provide an initial baseline for mRNA degradation, the comprehensive nature of the data allows for selection of the most appropriate mRNA markers for sample age estimation that should be considered for future evaluation with a larger sample set.

Materials and Methods

Description of Samples

All sample handling described in this methodology adheres to the OSU-CHS IRB approved protocol dated May 13, 2013.

Biological fluid samples including blood, saliva, semen, and vaginal fluid were collected from donors who are over the age of 18 and signed the informed consent form for having their sample sequenced. Blood samples were drawn in 10 cc aliquots by a medical technologist. For saliva collection, the donor deposited approximately 1.0 mL of saliva into a sterile collection tube. Vaginal fluid was collected by providing the donor cotton swaps for collection of the sample. Semen was collected by deposition of the sample into a sterile collection tube provided to the donor.

Upon collection blood, saliva, and semen samples were deposited on nuclease-free paper cards in 50 µl aliquots. Samples were labeled with a unique 10-digit code and all samples (cards containing blood, semen, and saliva and vaginal fluid cotton swabs) were stored in the dark, at room temperature. Samples were allowed to age for a specified amount of time (Table 1) before RNA extraction was performed.

Isolation of RNA

RNA isolation was performed with TRI Reagent® (Sigma Aldrich, St. Louis, MO), following the manufacturer's instructions. The aqueous phase of the TRI Reagent®, containing the isolated RNA, was transferred to a clean 1.5 mL Eppendorf tube. The RNA underwent further clean-up by Zymo Research Clean and Concentrator™ Kit, following the manufacturer's instructions (Zymo Research, Irvine, CA). RNA was eluted in 15 µl dH_2O and all samples underwent a DNase digestion using TURBO™ DNase (Life Technologies, Carlsbad, CA) following the manufacture's protocol. All samples were quantitated on the Qubit® using the RNA HS kit (Life Technologies, Carlsbad, CA).

cDNA Library Preparation

Samples containing 20 ng of isolated RNA were mixed with 4 µl of ERCC RNA Spike-in mix 1(Ambion®) at a 1:10000 dilution. RNA samples were converted to cDNA using the NuGEN Ovation® RNA-seq Kit v2 (NuGEN Technologies, San Carlos, CA), following the manufacture's protocol. All cDNA samples were checked for purity (A260/280 > 1.8) and quantity using the Nanodrop ND-1000 microspectrophotometer (Thermo Scientific, Wilmington, DE).

Sample aliquots of 30 µl of low TE containing 1 µgof cDNA were fragmented on the Bioruptor® UCD 200 (Diagenode, Denville, NJ) to an average fragment size of 200 bp. Once fragmentation was complete, cDNA libraries were constructed using the Ion Plus Fragment Library kit (Life Technologies, Carlsbad, CA) following the manufacturers protocol for 200 bp, 1 µg input library preparation. All libraries received barcoded adapters so that each pair of technical replicates (same sample type and time point) could be sequenced on the same Ion 318™ v2 chip. Ion Xpress™ barcode adapters were utilized for all barcoding. All libraries were quantitated using the Ion Library Quantitation kit on the ABI 7500 qPCR instrument following the manufacturer's protocol (Life Technologies, Carlsbad, CA). Based on quantitation results, dilution factors were calculated for each library.

Template Preparation

Template preparation was performed on the OneTouch2™ (OT2) instrument. Sample technical replicates (same sample type, same time point) were pooled together in an equal concentration of 26 pM and loaded on to the OT2™ following the

manufacture's protocol (Life Technologies, Carlsbad, CA). After emulsion PCR on the OT2™ was complete, samples were enriched on the Ion Torrent™ ES to remove non-templated and polyclonal ISPs.

Sequencing on the Ion Torrent PGM

Once template preparation was complete, enriched template positive ISP's were mixed with buffer, control ISPs, and enzyme and loaded onto an Ion 318™ v2 chip for sequencing on the Ion Torrent™ PGM. Default sequencing parameters for 200 bp libraries on a 318™ v2 chip were used for all sequencing reactions.

Data Analysis

Analysis of the sequence data obtained from duplicate fresh and aged biological fluid samples proceeded through a three-step process: In the first step of data analysis, raw sequence data for a given sample were aligned to the human reference genome, Hg19 (GRCh38). After alignment, every sample had RNA expression levels calculated in the form of reads per kilobase per million (RPKM). RPKM values normalize expression levels by taking into account the total number of sequencing reads in a run, the size of the gene, and the number of sequencing reads that map to that gene (Mortazavi et al.., 2008). Alignment and RPKM calculations were performed with CLC Bio Genomics Workbench software (Cambridge, MA). After initial RPKM values were calculated, they were normalized against the ERCC spike-in standards, from which a standard curve is created (input quantity of ERCC standard vs. RPKM) (Figure 3). Normalization to the spike-in standard acts as a control for any variation that might have been introduced by sample

preparation or user error, as the ERCC standards (Ambion®) are spiked-into each sequenced sample at a known molar concentration (Jiang et al.., 2011). The standard curve that is produced from the ERCC spike-in standards can be used to estimate abundance (molecules) of any sequenced transcript in the sample (Figure 3). The final normalized abundance value for each transcript is expressed in molecules of RNA. Once the data for each sample were normalized, datasets were compared within and between sample types to determine mRNA degradation profiles and patterns.

Results and Discussion

Determination of Transcript Abundance over Time

Once the transcriptome sequencing data were normalized for each sample, the data were compared across all of the analyzed time points for each tissue type to determine if transcript abundance changed with sample age. Abundance values for technical replicates of each sample type at each time point were averaged. The average abundance of each individual transcript was then compared across all of the time points for each sample type. By monitoring the change in individual transcript abundance over time, a degradation profile could be developed for each individual mRNA transcript in a given sample transcriptome. If a transcript is degrading over time, you will see a decrease in the transcripts abundance over time. Determination of individual transcript degradation profiles was accomplished through calculation of slope and R^2 for each transcript over all of the sequenced time points. The average sample has thousands of transcripts represented at time 0 (fresh). This approach examines transcriptome degradation, thus individual degradation profiles for thousands of genes in each sample type are revealed.

The data generated through sequencing total mRNA isolated from biological fluids aged up to one year indicate that there is observable change in the transcriptome of a sample over time. Figure 8 depicts the average abundance levels for every detected transcript in a given sample type (for each fluid and each time point). There are hundreds to thousands of points graphed above each time point for each sample. Each point on the graph represents a single gene and its average abundance at the given time-point (x-axis). When multiple comparisons analysis was performed on this data, all comparisons between each of the time points were significant at less than 0.0001. This is an indication that the abundance of detected genes within each sample type is decreasing over time.

While figure 8 provides a snapshot of the total mRNA abundance in each sample at each time point (plotting one point for each transcript detected at each time point), the data can be filtered to examine the change of individual transcripts over time. Having access to the degradation rate and pattern of each transcript in the transcriptome of a sample facilitates selection of mRNA transcripts that have a degradation rate that better correlates with sample age.

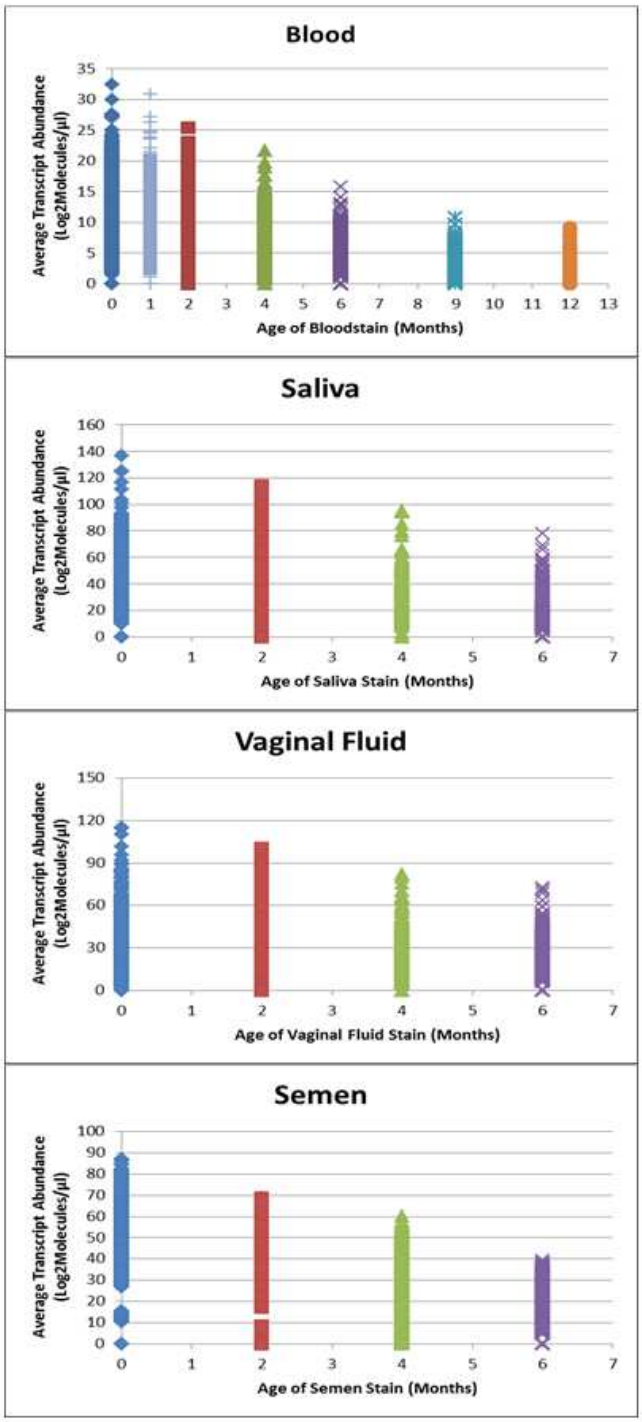


Figure 8. Determination of Transcript Abundance over Time. These charts depict the change in abundance of the transcripts in a sample over time. There is one point plotted for the average abundance of every gene detected at each sampled time point. Plotting one point for each detected gene at a given time point demonstrates that there is a clear decrease in transcriptome abundance over time for each tissue. Multiple comparisons analysis indicates a significant decrease between each of the time points ($p<0.0001$ for all comparisons).

Transcript Drop-Out Observed in Aged Samples

While there is an observable decrease in the abundance level of individual transcripts over time, there is also a decrease in the number of transcripts detected at each time point. Figure 9 depicts the number of transcripts detected at each sampled time point for each of the biological fluids. There is a decrease in the number of transcripts detected at each time point. This trend is observed for each of the biological fluid types. The point at which a transcript disappears from detection is called transcript drop-out. For example, in a given fluid type, if a transcript is present in the fresh sample and in the 60 day sample, but is no longer present in the 120 day sample, that transcript would have dropped out at 120 days. This trend of transcripts dropping-out of sequence data is represented by the decreasing number of transcripts detected over time in each of the sample types (figure 9).

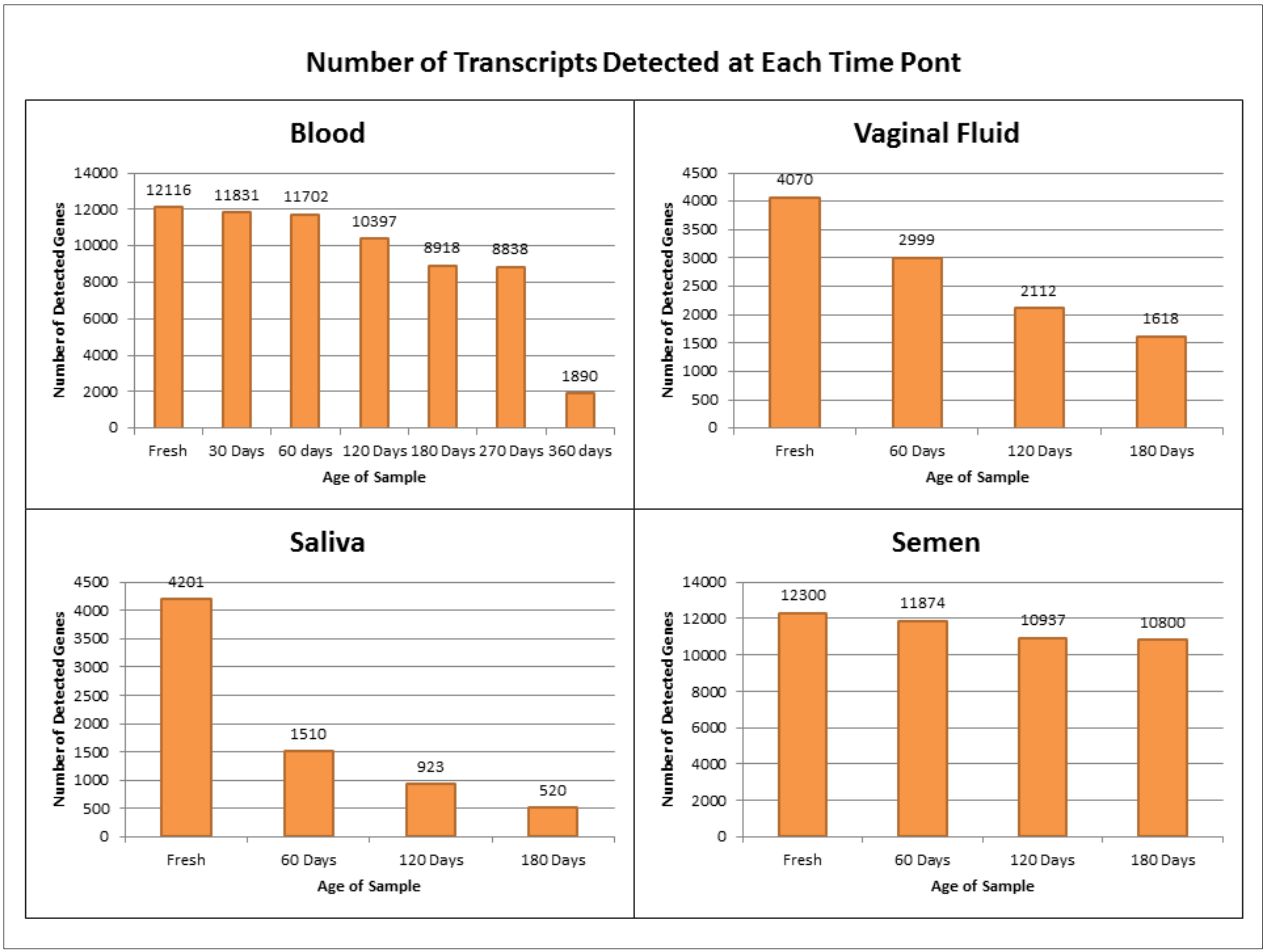


Figure 9. Transcript Abundance over Time. This figure depicts the number of transcripts detected at each time point for each fluid type. There are fewer transcripts detected as the samples age, indicating some transcripts are dropping below sequencing detection levels as the samples age.

Transcript drop-out is observed in all sample types as the samples age. Thus, transcript drop-out is a reflection of mRNA degradation within the sample. As transcripts degrade to a fragment size of less than 200 bp in size, they will no longer be captured in the sequencing library preparation. This gradual transcript degradation will be reflected by a reduction of transcript abundance in the sequencing data and eventual transcript drop-out from the sequencing data. It is important to note that just because a transcript has dropped-out of the sequencing data; it has not necessarily disappeared completely

from the sample. Fragments of a transcript may still be present in a sample, but simply not detectable by sequencing due to the requirement of 200 bp library fragments. Monitoring transcript drop-out in the sequencing data is critical to identifying transcript degradation profiles. Transcript drop-out time can provide insight into the degradation rate of individual transcripts. For instance, a transcript that drops-out of the sequencing data at 60 days has a much steeper degradation rate compared to a transcript that is still present at 180 days. Through evaluation of transcript drop-out, insight into mRNA degradation mechanisms can be achieved and appropriate short- and long-term markers of age can be selected.

The Effect of Starting Abundance on Degradation Rate

Through the evaluation of transcript drop-out time in aged samples, conclusions can be drawn about the effect of starting transcript abundance on degradation rate. Figure 10 displays the average abundance at time 0 (fresh sample) for the transcripts that drop out at each of the sampled time points. For instance, the average time 0 abundance for blood transcripts that have dropped-out by 30 days is 4.77 molecules per µl. The average time 0 abundance for blood transcripts that never drop out (i.e. “Drop-out Not Observed”, meaning these transcripts are detected at every time point, including 360 days) is 13.84 molecules per µl. Thus, the average starting abundance for transcripts that disappear by 30 days post-deposition is almost 3 times lower than transcripts that are still present at 360 days post-deposition.

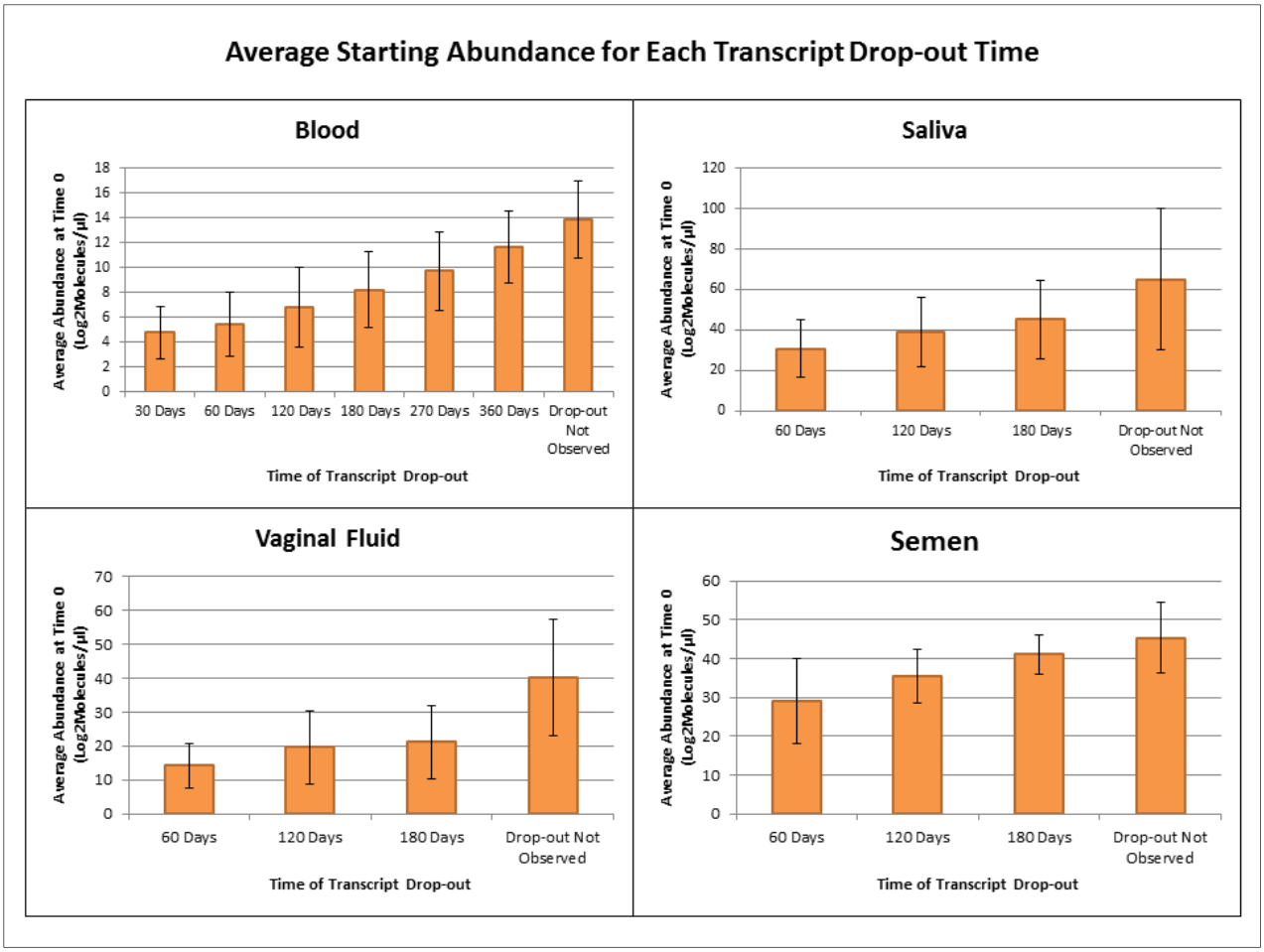


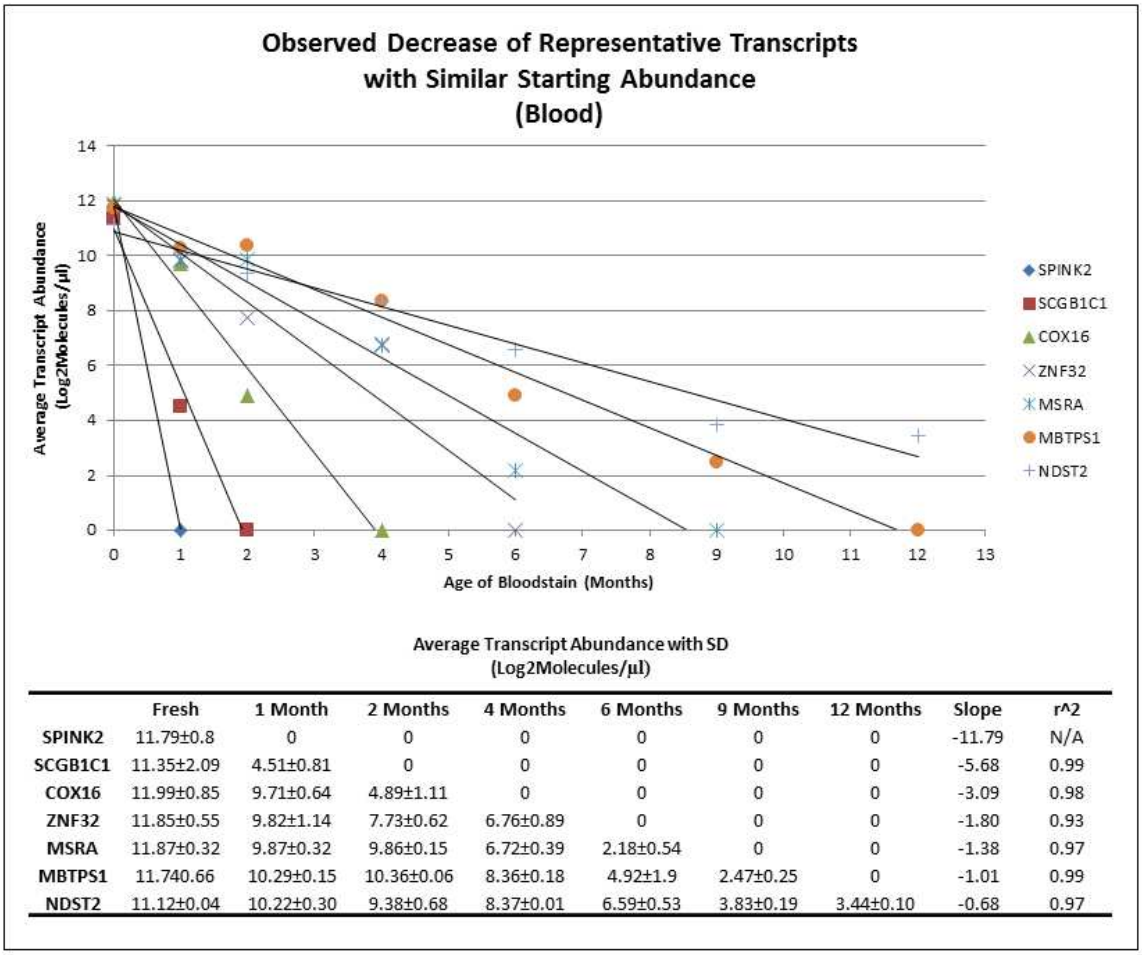
Figure 10. Starting Transcript Abundance by Sample type. These graphs depict the average starting abundance and standard deviation for transcripts that drop out at the specified times on the x-axis. There is an increase in the average starting abundance of transcripts that drop-out of detection at later time points. While transcripts that drop out in earlier time points (30 days, 60 days) have lower time 0 (fresh sample) abundance.

A high starting abundance for a transcript correlated well with that transcripts continued presence in older samples. This trend was observed in all of the sample types and is indicative of a correlation between transcript abundance in a fresh sample and the time it will take for a transcript to degrade to a point below sequencing detection. Therefore, the starting abundance of a transcript may have an effect on the transcript degradation rate. Transcripts with a higher abundance at time 0 (fresh sample) may

disappear more slowly, simply because a larger population of molecules will take longer to degrade. This property lends to abundant transcripts having an increased presence in aged samples compared to lowly abundant transcripts.

Other Factors Affect Transcript Degradation

While starting transcript abundance does appear to play a role in the rate at which a gene disappears from sequencing detection, the abundance level in the time 0 sample of a transcript is not the only factor that influences degradation rate. By examining the group of transcripts that drop-out at each sampled time point, genes with similar time 0 abundance values can be identified in each group (Figures 11, 12, 13, and 14). Therefore, the abundance level of a transcript in a fresh sample is not the only factor that affects that transcripts degradation rate. Figure 11 depicts this scenario in aged blood samples. This figure provides the average abundance levels of seven transcripts over seven time points (fresh, 1 month old, 2 month old, 4 months old, 6 months old, 9 months old, and 12 months old). The abundance of each transcript in the fresh sample is approximately 11 molecules per µl for each one of the transcripts (Figure 11). However, while all of the transcripts have a similar starting abundance, each transcript drops-out of detection at a different time point. For example, SPINK2 is present in the fresh blood sample at an abundance of 11.79 molecules/µl, but disappears by 1 month. This is in stark contrast to NDST2, which is present in the fresh blood sample at an abundance of 11.12 molecules/µl, but is also present at every sampled time point thereafter, including the 12 month sample.



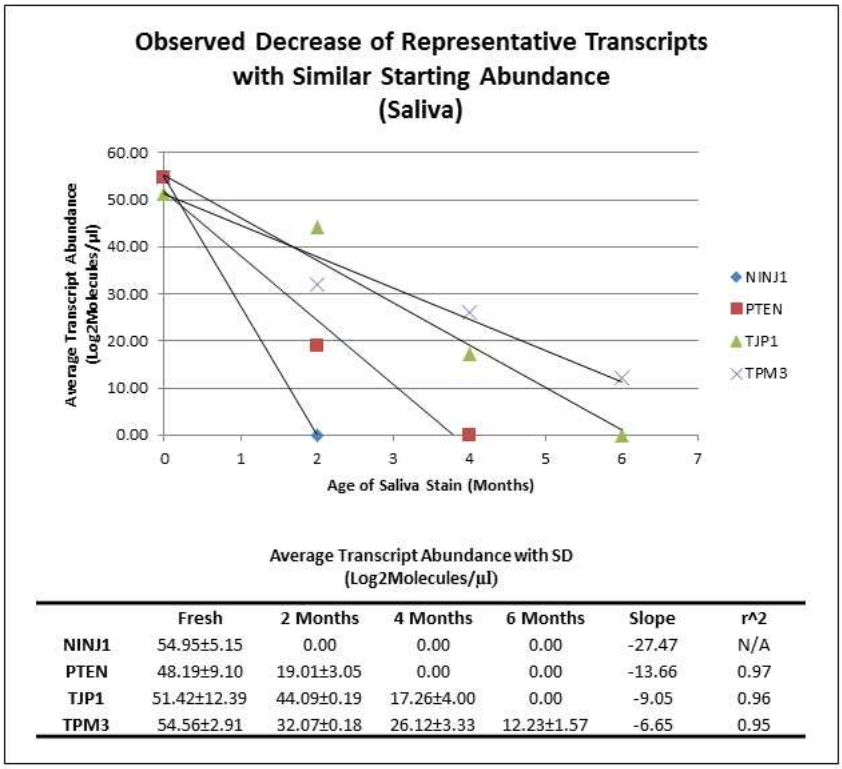
Average Transcript Abundance with SD
(Log2Molecules/µl)

	Fresh	1 Month	2 Months	4 Months	6 Months	9 Months	12 Months	Slope	r^2
SPINK2	11.79±0.8	0	0	0	0	0	0	-11.79	N/A
SCGB1C1	11.35±2.09	4.51±0.81	0	0	0	0	0	-5.68	0.99
COX16	11.99±0.85	9.71±0.64	4.89±1.11	0	0	0	0	-3.09	0.98
ZNF32	11.85±0.55	9.82±1.14	7.73±0.62	6.76±0.89	0	0	0	-1.80	0.93
MSRA	11.87±0.32	9.87±0.32	9.86±0.15	6.72±0.39	2.18±0.54	0	0	-1.38	0.97
MBTPS1	11.740.66	10.29±0.15	10.36±0.06	8.36±0.18	4.92±1.9	2.47±0.25	0	-1.01	0.99
NDST2	11.12±0.04	10.22±0.30	9.38±0.68	8.37±0.01	6.59±0.53	3.83±0.19	3.44±0.10	-0.68	0.97

Figure 11. Observed Decrease of Representative Transcripts in Blood. Seven genes with similar starting abundance in blood are displayed. While the seven transcripts have similar starting abundances, they disappear at different times over the 12 month time course, indicating that starting abundance is not the only factor in determining the degradation rate of a transcript.

This same trend is observed for every other biological fluid type evaluated. Figures 12, 13, and14 present data for saliva, vaginal fluid, and semen, respectively. Figure 12 provides data on four genes detected in saliva all with a starting abundance of about 50 molecules per µl (ranging from 48.19 molecules per µl to 54.95 molecules per µl). However, while these transcripts all have a similar starting abundance, they drop-out from detection at different times. NINJI is present in the fresh saliva sample, but is no longer detected as of the 2 month-old saliva sample. In contrast, TPM3 is present in the

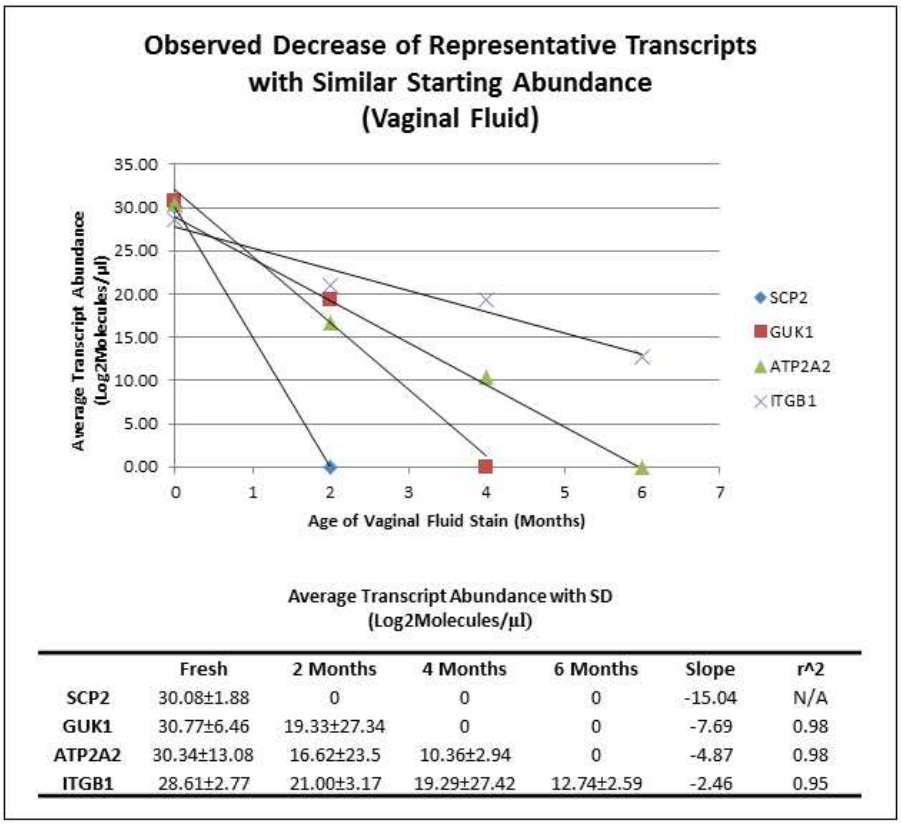
fresh saliva sample and detected at every other sampled time point thereafter, including the 6 month-old sample. Figure 13 presents four genes detected in vaginal fluid, all with a starting abundance of around 30 molecules per µl (ranging from 28.61 molecules per µl to 30.77 molecules per µl). Figure 14 presents four genes in semen, all with a starting abundance of around 35 molecules per µl. However, as with blood and saliva, the starting abundance in the genes presented for vaginal fluid and semen were also not predictive of drop-out time. In both of these fluids, genes are presented that have similar starting abundances, but drastically different drop-out times (ranging from dropping-out at 2 months, to no drop out observed).



Average Transcript Abundance with SD (Log2Molecules/µl)

	Fresh	2 Months	4 Months	6 Months	Slope	r^2
NINJ1	54.95±5.15	0.00	0.00	0.00	-27.47	N/A
PTEN	48.19±9.10	19.01±3.05	0.00	0.00	-13.66	0.97
TJP1	51.42±12.39	44.09±0.19	17.26±4.00	0.00	-9.05	0.96
TPM3	54.56±2.91	32.07±0.18	26.12±3.33	12.23±1.57	-6.65	0.95

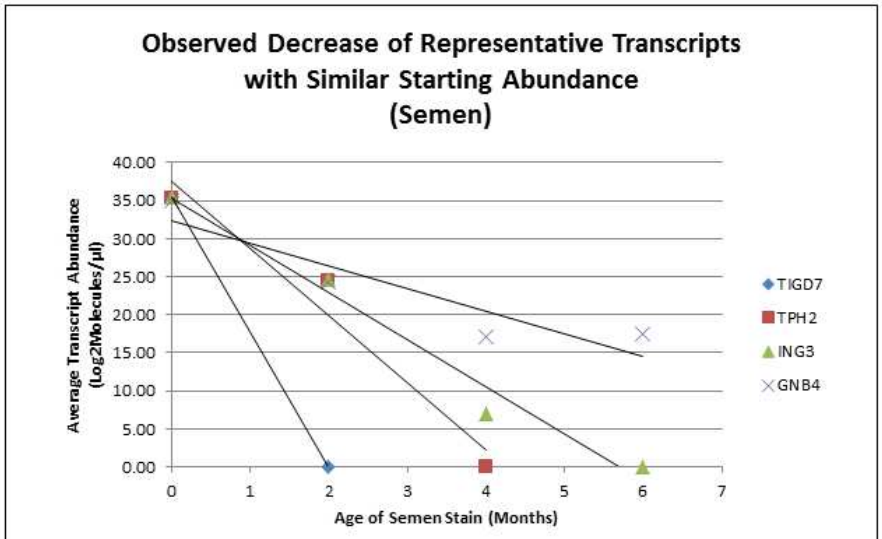
Figure 12. Observed Decrease of Representative Transcripts in Saliva. The abundance values for four transcripts in saliva are displayed. All of the transcripts have a similar starting abundance. However, they each drop-out at a different time point. These data indicates that factors other than time 0 abundance affect sample degradation rate.



Average Transcript Abundance with SD
(Log2Molecules/µl)

	Fresh	2 Months	4 Months	6 Months	Slope	r^2
SCP2	30.08±1.88	0	0	0	-15.04	N/A
GUK1	30.77±6.46	19.33±27.34	0	0	-7.69	0.98
ATP2A2	30.34±13.08	16.62±23.5	10.36±2.94	0	-4.87	0.98
ITGB1	28.61±2.77	21.00±3.17	19.29±27.42	12.74±2.59	-2.46	0.95

Figure 13. Observed Decrease of Representative Transcripts in Vaginal Fluid. The abundance values for four transcripts in vaginal fluid are displayed. All of the transcripts have a similar starting abundance. However, they each drop-out at a different time point. These data indicates that factors other than time 0 abundance affect sample degradation rate.



Average Transcript Abundance with SD
(Log2Molecules/µl)

	Fresh	2 Months	4 Months	6 Months	Slope	r^2
TIGD7	35.43±7.16	0.00	0.00	0.00	-17.7161	N/A
TPH2	35.23±7.12	24.39±1.13	0.00	0.00	-8.80683	0.952918
ING3	35.48±6.17	24.61±1.54	7.02±4.54	0.00	-6.20225	0.976272
GNB4	35.01±5.07	24.21±1.32	17.09±1.76	17.43±0.25	-2.99313	0.849442

Figure 14. Observed Decrease of Representative Transcripts in Semen. The abundance values for four transcripts in semen are displayed. All of the transcripts have a similar starting abundance. However, they each drop-out at a different time point. These data indicates that factors other than time 0 abundance affect sample degradation rate.

The presence of transcripts that have similar starting abundance in every fluid type that drop-out at different time points in the aging time course provides evidence that factors other than simple starting abundance affect the RNA degradation rate. While there does not appear to be a correlation between transcript specificity (tissue-specific vs. present in several sample types) or transcript function (housekeeping vs. cell signaling, etc.) and stability of the transcript in aged samples, more investigation is needed on this subject. In addition to transcript specificity and function, there does not appear to be a correlation between transcript size and degradation rate, however, more investigation is needed on this issue as well. In addition to the starting abundance of a given transcript, it

is possible that transcript structure and function contribute to the stability of the molecule as well.

Tissue-Specific Degradation Patterns

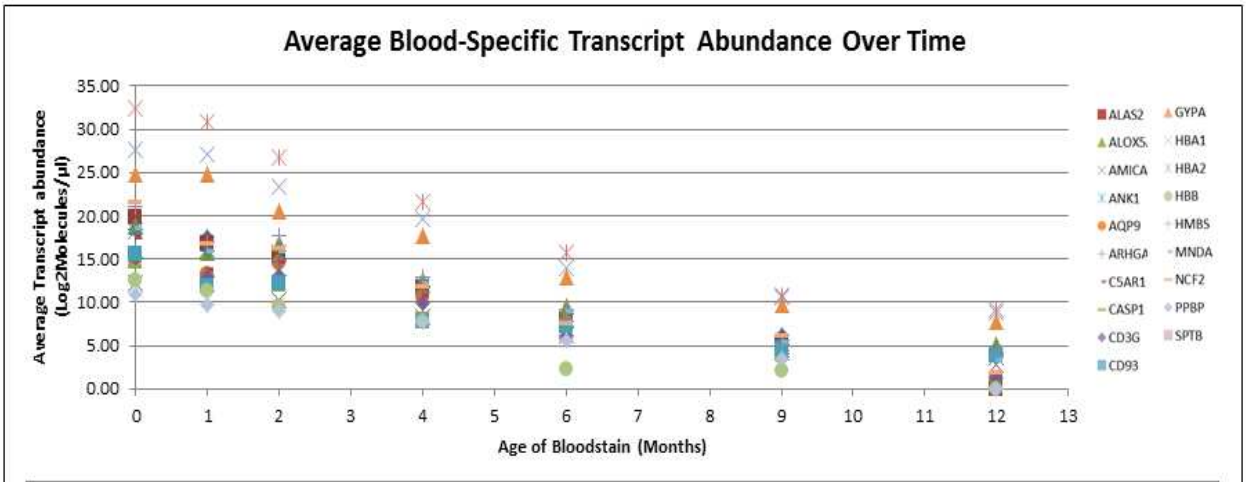
In addition to establishing the fact that the total mRNA of a sample is degrading over time, another goal of this study was to examine tissue-specific RNA degradation patterns and profiles. It was hypothesized that because each fluid (blood, saliva, vaginal fluid, and semen) is composed of a unique transcriptome, cell types, cell environments, and microbial populations, there would be differing rates and patterns of degradation in each fluid type. With just two technical replicates of every fluid at each time point represented in this sequencing data set, it is important to note that any observed tissue-specific differences are just initial observed trends. Further analysis of a larger sample size would be needed to draw any firm conclusions about tissue-specific RNA degradation rates.

Understanding tissue-specific differences in RNA degradation rates is important because the transcriptome of every biological tissue or fluid is unique and can therefore not be expected to act in a uniform manner. Forensic analysts deal with a variety of biological sample types. While it would be ideal to be able to streamline mRNA analysis into one universal assay, tissue-specific assays may be required for the proper analysis of mRNA degradation in relation to sample age. It is critical that an understanding of the transcriptome and its degradation rates and patterns are established for every forensically relevant sample type before conclusions are made concerning the creation of universal or sample type-specific analysis procedures.

Awareness of tissue-specific RNA degradation was first acquired by reviewing the published literature searching for forensic tissue-identification markers for each sample type. A comprehensive literature search was performed to identify mRNA markers for blood, semen, saliva, and vaginal fluid. Several tissue-tissue and fluid-specific mRNA biomarkers have been identified for use with forensic tissue identification, with the aim of replacing traditional serological techniques with molecular analysis. Many of these mRNA biomarkers have been validated for sensitivity and specificity. In this study, fluid-specific RNA transcripts were utilized to monitor sample type-specific mRNA degradation patterns. Google Scholar, NCBI PubMed, and ScienceDirect databases were all utilized in this literature search for sample identity biomarkers. Search terms included “RNA markers for tissue identification”, “RNA markers for biological fluid identification”, “RNA used to identify tissues and fluids”, “forensic identification of fluids and tissues using RNA”, and “mRNA markers for biological tissues and fluids”. Once identified, tissue-specific mRNA transcripts were placed in a database to be used in the analysis of the aged biological fluid samples (Appendix A).

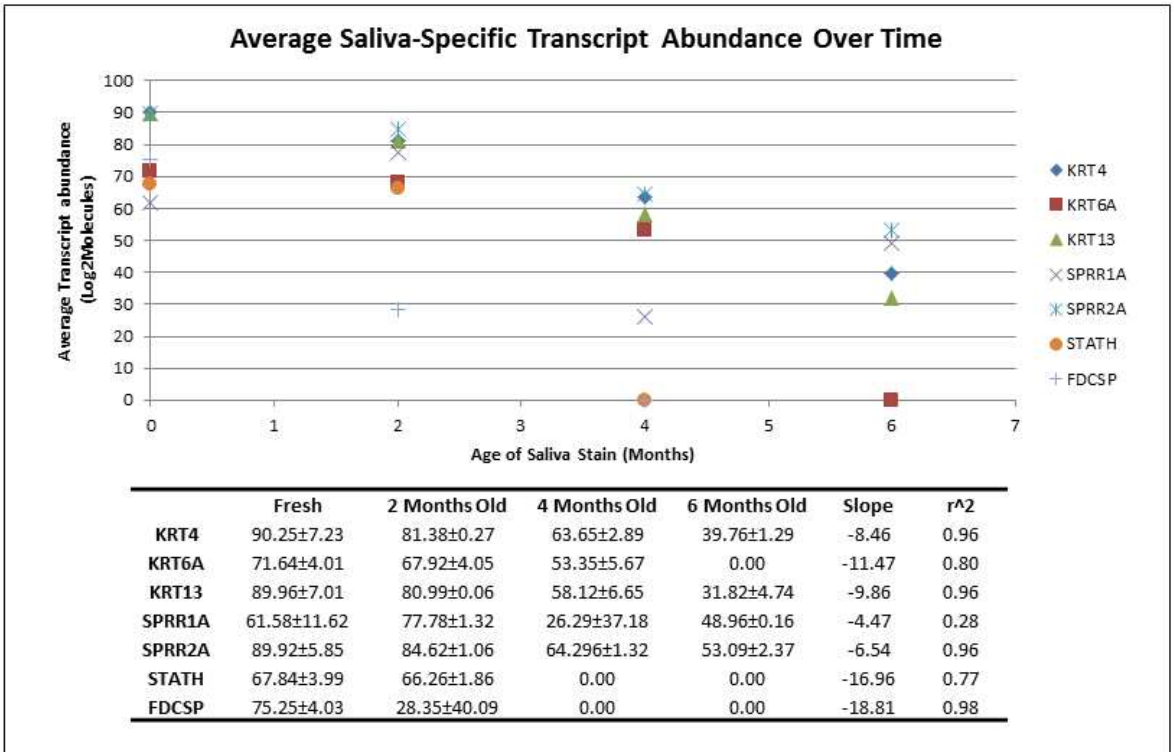
Tissue-specific transcripts were monitored in each of the sample types over the entire aging time course. Results for each of the fluid types and their specific mRNA markers are displayed in figures 15, 16, 17, and 18. Blood and semen specific markers appear to remain present in the sample for the longest amounts of time, with blood having no transcripts drop-out until 12 months and semen having no transcripts drop out in the observed time frame (Figures 15 and 18). Saliva and vaginal fluid do not appear to have as steady of a presence with their tissue-specific transcripts, with both sample types

having selected transcript drop-out by 4 months (Figures 16 and 17). This observed difference in stability of tissue-specific markers between the sample types (blood and semen vs. saliva and vaginal fluid) could be due to several factors including overall decrease in the percentage of genes detected in the sample types over time and the influence of microbial organisms on the samples.



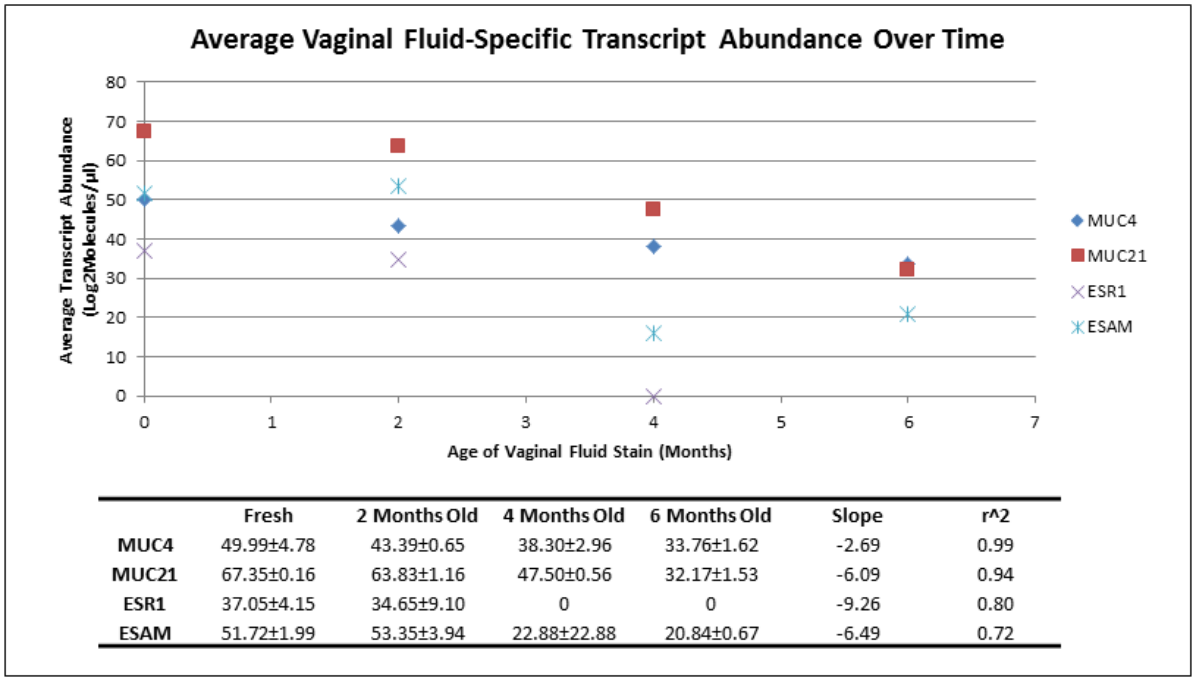
	Fresh	1 Month Old	2 Months Old	4 Months Old	6 Months Old	9 Months Old	12 Months Old	Slope	r^2
ALAS2	18.49±0.17	17.51±0.08	15.02±0.25	11.79±0.05	8.57±0.06	6.15±0.03	4.02±0.29	-1.25	0.96
ALOX5AP	19.86±0.40	16.82±0.08	15.22±0.27	11.65±0.08	7.91±0.13	5.13±0.05	0.00	-1.58	0.99
AMICA1	18.98±0.22	15.82±0.10	16.80±0.16	12.59±0.08	9.52±0.14	5.32±0.04	5.19±0.10	-1.22	0.94
ANK1	12.28±0.01	12.26±0.15	10.28±0.31	8.35±0.15	6.13±0.51	4.15±0.03	2.86±0.10	-0.85	0.97
AQP9	18.30±0.18	16.03±0.06	15.66±0.06	12.03±0.06	8.14±0.01	4.50±0.11	3.77±0.29	-1.29	0.96
ARHGAP26	14.78±0.09	13.26±0.03	14.44±0.33	10.59±0.03	7.76±0.16	3.64±0.05	1.20±0.69	-1.20	0.98
C5AR1	17.80±0.06	15.89±0.13	14.61±0.39	11.77±0.10	7.69±0.30	5.60±0.01	4.05±0.25	-1.19	0.96
CASP1	17.54±0.03	13.77±0.30	14.73±0.01	10.79±0.01	7.78±0.25	4.18±0.28	1.69±2.39	-1.30	0.97
CD3G	14.15±0.17	11.86±0.42	11.42±0.19	8.68±0.26	5.77±1.76	3.88±0.03	0.00	-1.13	0.99
CD93	15.26±0.12	12.87±0.10	13.08±0.26	9.83±0.12	6.11±0.59	4.09±0.06	1.26±0.79	-1.17	0.98
GYPA	24.84±0.08	11.85±0.09	12.21±0.21	7.73±0.23	7.16±0.31	4.56±0.07	3.79±0.11	-0.93	0.89
HBA1	27.57±0.21	24.78±0.12	20.46±0.08	17.61±0.10	12.87±0.93	9.70±0.09	7.81±0.01	-1.53	0.95
HBA2	27.57±0.45	27.13±0.03	23.30±0.06	19.71±0.12	13.97±0.98	10.65±0.04	8.81±0.01	-1.70	0.95
HBB	32.43±0.82	30.75±0.03	26.78±0.01	21.63±0.15	15.71±1.15	10.78±0.04	9.24±0.02	-2.08	0.95
HMBS	12.56±0.36	11.37±0.02	9.48±0.83	7.69±0.14	2.27±1.22	2.10±0.31	0.00	-1.09	0.92
MNDA	21.07±0.39	16.97±0.08	17.76±0.05	12.93±0.01	9.10±0.03	5.06±0.06	2.81±0.25	-1.46	0.95
NCF2	18.64±0.19	15.96±0.01	15.98±0.10	12.19±0.01	9.09±0.08	5.49±0.08	3.29±0.10	-1.30	0.98
PPBP	21.55±0.41	16.83±0.03	16.30±0.05	11.80±0.12	7.63±0.36	6.15±0.02	1.90±0.69	-1.53	0.95
SPTB	11.00±0.35	9.70±0.21	9.00±0.69	7.68±0.10	5.58±0.50	3.50±0.13	0.00	-0.88	0.99

Figure 15. Average Blood-Specific Transcript Abundance over Time. Blood-specific transcript degradation over 12 months is displayed. No transcripts drop-out until 12 months post-deposition, indicating a large degree of stability among blood-specific mRNA.



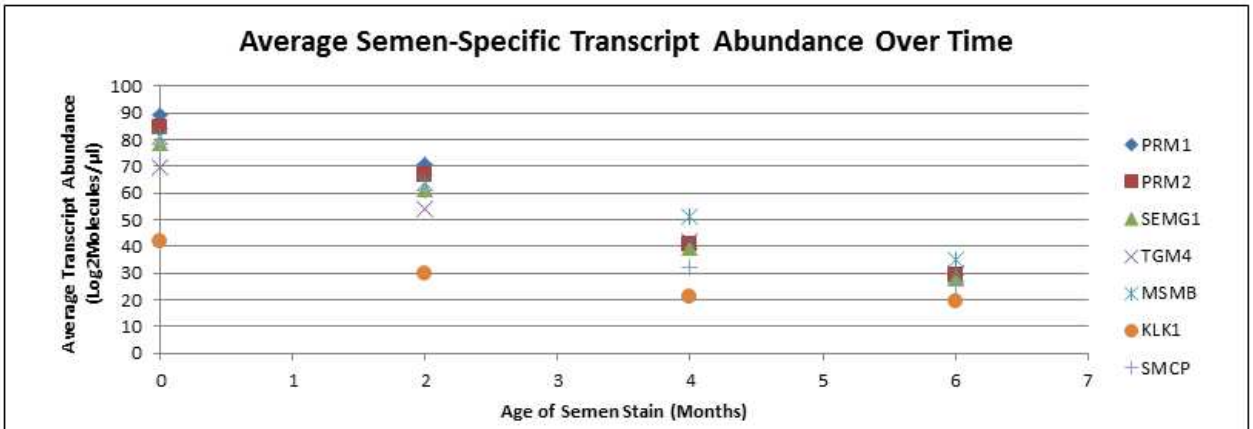
	Fresh	2 Months Old	4 Months Old	6 Months Old	Slope	r^2
KRT4	90.25±7.23	81.38±0.27	63.65±2.89	39.76±1.29	-8.46	0.96
KRT6A	71.64±4.01	67.92±4.05	53.35±5.67	0.00	-11.47	0.80
KRT13	89.96±7.01	80.99±0.06	58.12±6.65	31.82±4.74	-9.86	0.96
SPRR1A	61.58±11.62	77.78±1.32	26.29±37.18	48.96±0.16	-4.47	0.28
SPRR2A	89.92±5.85	84.62±1.06	64.296±1.32	53.09±2.37	-6.54	0.96
STATH	67.84±3.99	66.26±1.86	0.00	0.00	-16.96	0.77
FDCSP	75.25±4.03	28.35±40.09	0.00	0.00	-18.81	0.98

Figure 16. Average Saliva-Specific Transcript Abundance over Time. Saliva-specific transcript degradation is displayed. Transcript drop-out is observed at 4 months for two different genes, indicating a possible instability of RNA in saliva.



	Fresh	2 Months Old	4 Months Old	6 Months Old	Slope	r^2
MUC4	49.99±4.78	43.39±0.65	38.30±2.96	33.76±1.62	-2.69	0.99
MUC21	67.35±0.16	63.83±1.16	47.50±0.56	32.17±1.53	-6.09	0.94
ESR1	37.05±4.15	34.65±9.10	0	0	-9.26	0.80
ESAM	51.72±1.99	53.35±3.94	22.88±22.88	20.84±0.67	-6.49	0.72

Figure 17. Average Vaginal Fluid-Specific Transcript Abundance over Time. Vaginal fluid-specific transcript degradation is displayed. Transcript drop-out is observed at 4 months for one out of four genes, indicating a possible instability of RNA in saliva.



	Fresh	2 Months Old	4 Months Old	6 Months Old	Slope	r^2
PRM1	89.17±3.50	70.81±3.23	40.29±0.25	19.52±0.19	-10.51	0.97
PRM2	84.56±3.32	66.83±3.04	40.81±0.52	19.70±0.09	-9.55	0.98
SEMG1	78.61±3.09	61.72±2.82	39.42±0.28	19.23±0.02	-8.58	0.98
TGM4	69.65±2.74	53.99±2.46	41.71±0.39	20.01±0.05	-6.56	0.99
MSMB	81.03±3.18	63.80±2.91	51.16±0.30	23.23±0.15	-7.56	0.99
KLK1	41.74±1.64	29.96±1.37	21.12±1.52	12.99±0.14	-3.78	0.91
SMCP	78.02±3.06	61.20±2.79	32.29±0.07	16.79±0.19	-9.37	0.96

Figure 18. Average Semen-Specific Transcript Abundance over Time. Semen-specific transcript degradation is displayed. Transcript drop-out is not observed for any of the markers over the 6 month time course, indicating a stability of mRNA in semen.

If other data, such as the number of transcripts detected at each sampled time point (Figure 9) are taken in to account alongside the tissue-specific transcript data, some light may be shed on the apparent differences in transcript stability between the tissue types. The mRNA population in blood remains fairly stable over time, with over 70% of transcripts still present at 9 months post-deposition. A similar trend is observed in semen, with over 80% of the transcripts remaining at 6 months post-deposition. In contrast to the transcriptome stability observed in blood and semen, only 12% of transcripts in saliva and 39% of transcripts in vaginal fluid remain at 6 months post-deposition. The number of transcripts remaining in each of the tissues at 6 or 9 months post-deposition provides evidence that the transcriptomes of blood and semen appear much more stable over time than the transcriptomes of saliva and vaginal fluid. This fact could offer one reason for

the early drop-out of tissue-specific mRNA markers observed in saliva and vaginal fluid (Figures 16 and 17).

Deeper evaluation of the RNA population isolated from the different sample types offers some insight into the apparent differences in mRNA stability among the different fluid types. Read alignment statistics for each of the sample types provides further divide between blood and semen vs. saliva and vaginal fluid. As listed in Table 4, the average sequencing read alignment to HG19 was quite different among the four sample types. For blood, an average of 81% of the sequencing reads aligned to the human reference. Semen achieved similar results, with an average alignment of 84% of the sequencing reads. A large departure from this trend was seen with saliva samples, which, on average had only 6% of the sequencing reads aligned. The vaginal fluid samples were more in line with the saliva samples, with an average of 13% of reads aligning to the human reference. The unaligned reads for both saliva and vaginal fluid were blasted against HOMD and Refseq databases, respectively, and over 90% of the unaligned reads were found to align to microbial organisms. This alignment data demonstrates that while the majority of the RNA population was human for blood and semen, both saliva and vaginal fluid had a significant microbial presence. The heightened presence of microbial RNA in saliva and vaginal fluid, not seen in blood and semen, most likely had a large effect on the detected human mRNA population, possibly causing heightened degradation rates of human RNA or drowning out the population of human RNA that was detectable by sequencing on the PGM.

By examining known tissue-specific mRNA markers in combination with the number of transcripts detected at each time point for each sample type, some tissue-

specific differences in mRNA degradation patterns and rates become apparent. Most notable is the observed stability of mRNA in blood and semen compared to saliva and vaginal fluid. This observation must be kept in mind as these data are utilized for selecting appropriate mRNA markers for sample age, as universal mRNA transcripts present in all tissue types may have drastically different degradation rates based on what tissue type is being considered.

Transcript Populations of Biological Fluids

In addition to considering the different rates and profiles of transcript degradation in the different sample types, it is also important to account for the population similarities and differences of the different sample transcriptomes. While no two biological sample types will have an identical transcriptome, there is some degree of overlap between the transcripts expressed in all of the mRNA populations analyzed in this study. By establishing transcript population similarities and differences, transcripts can be categorized as universal (found in every sample type; can be used to monitor age in any sample type) or tissue-specific (found in only one sample type; can be used to monitor tissue-specific age).

Sequence data for the time 0 (fresh) samples were compared to identify overlap between transcriptome populations of the different biological fluids (blood, saliva, vaginal fluid, and semen). Figure 19 provides the number of genes found in each tissue, the number of tissue-specific genes, and the number of genes found in the transcriptomes of multiple sample types. While there is overlap between all of the different tissues, the transcript populations of particular importance are the tissue-specific transcripts (for

blood, semen, saliva, and vaginal fluid) and the transcripts found in all of the tissue types (universal transcripts).

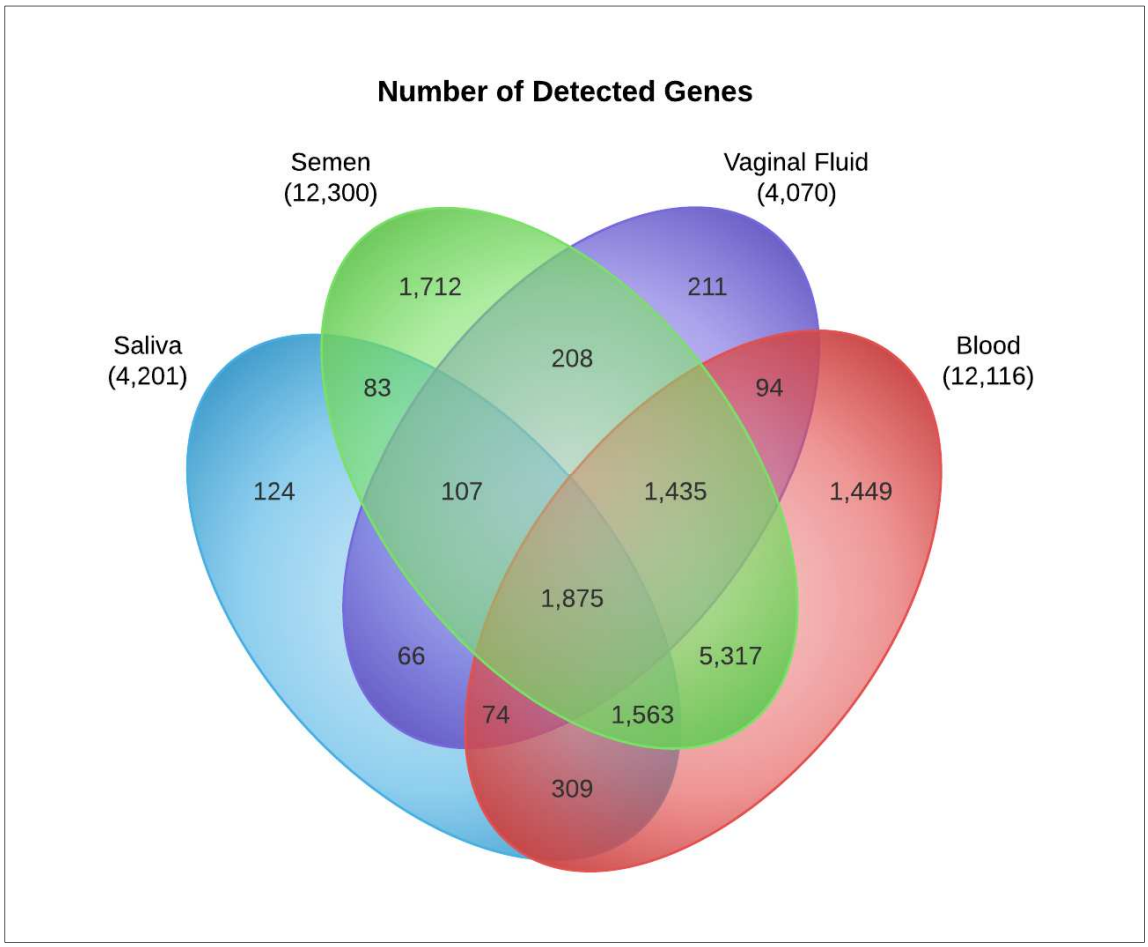


Figure 19. Tissue-Specific mRNA Profiles. This Venn diagram presents the number of detected genes in the transcriptome of each biological fluid. For every tissue type there are a number of tissue-specific transcripts. Additionally, there are 1,875 universal transcripts that are common to all of the sample transcriptomes.

For each of the tissues, there are a significant number of transcripts that are sample-type specific. Specifically, there are 1,449 blood specific transcripts,124 saliva-specific transcripts, 211 vaginal fluid-specific transcripts, and 1,712 semen-specific transcripts. These pools of tissue-specific transcripts provide a population for the selection of tissue-specific markers to estimate sample age. In addition to sample-specific

transcripts, there were also 1,875 transcripts common to all of the sample types. This pool of transcripts provides a population for the selection of universal markers of sample age that can be utilized with every sample type.

Identification of Markers for Sample Age Estimation

Having the degradation profile of the transcriptome for each fluid and tissue (consisting of thousands of transcripts per sample type) allows for the guided identification of mRNA transcripts that have degradation patterns and rates that most closely correlate with sample age. Different mRNA transcripts were identified as correlating to short, mid, or long-term sample age. Data analysis for mRNA candidate marker identification was performed in Microsoft Excel and Statistical Analysis Software (SAS). Short-term age mRNA markers disappear early in the degradation analysis (before 60 days). Candidate mRNA transcripts for short-term markers of sample age are present in the time 0 sample, but drop below detectable sequence levels by the first measured time point (1 month in blood, 2 months in saliva, semen, and vaginal fluid). Short term markers should have a large negative slope and a high r^2 value. Mid-term age mRNA markers are identified as having a steady linear degradation rate and drop below detectable sequence levels by middle time points (2, 4, 6, or 9 months in blood; 4 or 6 months in saliva, semen, and vaginal fluid). All markers selected for mid-range markers should have a measurable abundance at several time points and should have a strong negative slope and high r^2. All mid-term candidate mRNA markers have a clear linear decrease in abundance as samples age and drop below detectable levels before the final time-point (180 days or 360 days). Finally, candidates for long-term sample age

estimation are identified as transcripts that are present over all of the sequenced time points. Candidates for long-term sample age estimation have a negative slope and a high r^2.

Candidate markers for sample age estimation were identified with consideration for tissue specificity. Markers were identified that are unique to each of the tissues, to allow for tissue-specific estimation of sample age. Tissue-specific markers offer the benefit of being applicable in a mixed sample scenario, where each fluid in the mixture could have a unique estimate for sample age. Additionally, fluid-specific markers reflect fluid-specific mRNA degradation patterns and profiles. Markers were also identified that were found in every tissue that was analyzed. Universal mRNA markers for sample age have the benefit of being widely applicable. These markers could potentially be developed into an assay that would be applicable for use with a wide variety of sample types. Both types of markers (tissue-specific and tissue-nonspecific) were identified for future investigation.

Universal Markers of Sample Age

To identify universal transcripts that degrade predictably and could be useful to estimate sample age, the 1,875 transcripts found in every biological fluid type were evaluated based on degradation rate of each transcript across all of the samples. The slope and r^2 value for every transcript in the universal population was calculated for each tissue type. Transcripts were first separated based on when they disappeared from each of the sample types. Transcripts that disappeared by the 6 month time point in all sample types were placed into one group (short-term markers of sample age). Transcripts that never

dropped out, (i.e. were detected at every time point in every tissue type) were placed into a second group (long-term markers of sample age). Within each group, transcripts were further filtered based on r^2 value. All transcripts with a tight linear degradation rate ($r^2 > 0.8$) were kept for further consideration. After filtering based on a consistent degradation trend across all sample types, and on r^2 values, there were eight transcripts that qualified as potential mid-term mRNA markers for sample age and ten transcripts that qualified as long-term markers for sample age. Two example universal mid-term sample age markers, SERPINB2 and SPINT1 are displayed in Figure 20. All of the fluid types have SERPINB2 and SPINT1 present in the fresh sample (time 0), but both of these transcripts have dropped-out from all sample types by 6 months. While no mRNA markers were present in all samples at the same starting abundance, the presence or absence of these mRNA transcripts in a tissue could be an indicator of approximate sample age. Further validation is needed with a larger sample number.

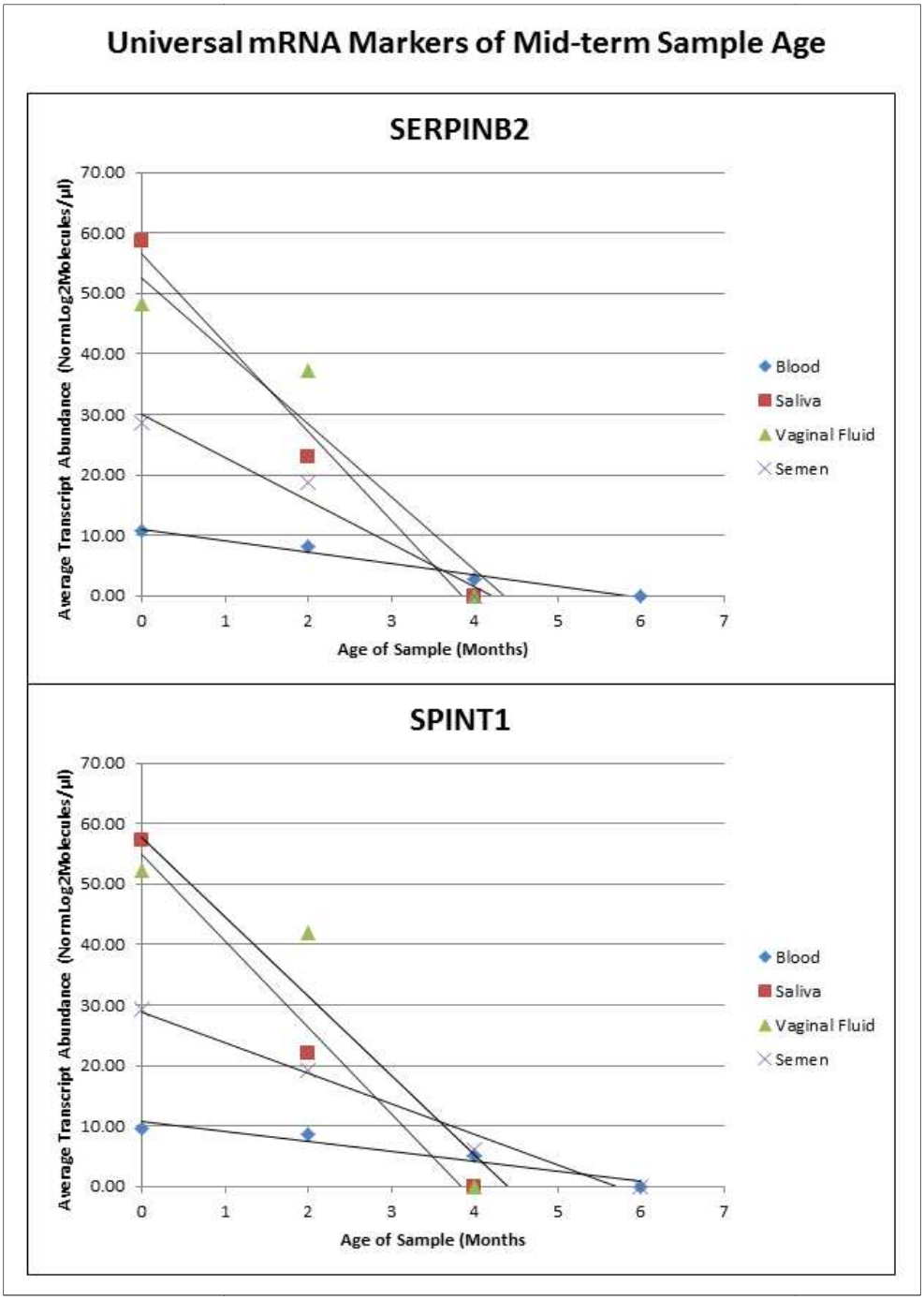


Figure 20. Universal mRNA Markers of Mid-term Sample Age. These graphs display two example universal mRNA markers of mid-range sample age, SERPINB2 and SPINT1. Both of these transcripts are present in all four analyzed tissues in fresh samples, but drop-out by 6 months in all sample types.

Two example universal long-term sample age markers, ACTB and and FTH1 are shown in Figure 21. Similar to the mid-range mRNA markers, there are no transcripts that have

the same starting abundance in all tissues and remain present throughout all time points. However, both ACTB and FTH1 are present at time 0 in all sample types and do not drop-out at any time. Both of these transcripts had measurable abundance in all of the sample types at the latest measured time point (6 months or 12 months).

While these transcripts on their own cannot be used to estimate the age of a sample, these data demonstrate that there are universal mRNA transcripts that appear to have unique degradation rates. As these transcripts are common to all tissues, a couple of them (for example, one mid-range marker and one long-range marker) could be combined with analysis of rRNA in a real time assay. Several studies have documented the stability of rRNA over an extended period of time. Thus, rRNA provides an excellent steady baseline and long-term marker of sample age. The universal mRNA markers documented in this study could be combined with rRNA in the development of a qPCR assay designed to measure transcript abundance. A qPCR assay would facilitate the rapid assessment of these RNA products in a large sample population. Sample age may be estimated from the presence or absence of different markers and their specific abundance level in the sample. Based on the sequencing data of aged samples, the mid-range mRNA marker would be expected to decrease with the fastest rate and disappear first from samples. The long-range mRNA marker would be expected to decrease at a slower rate and disappear at a later time point (after 6 months, according to our sequencing data).

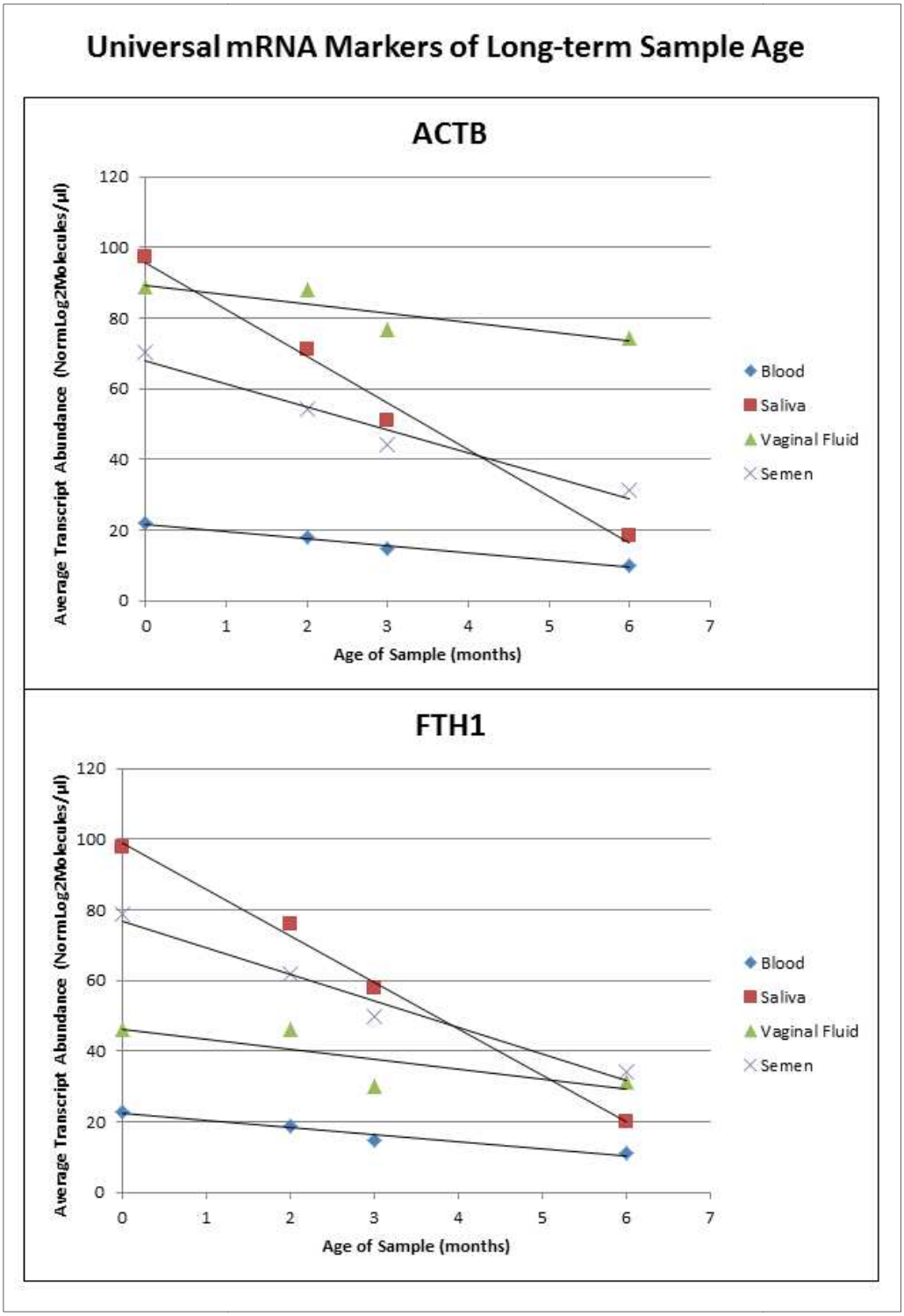


Figure 21. Universal mRNA Markers of Long-term Sample Age. These graphs display two example universal mRNA markers of long-range sample age, ACTB and FTH1. Both of these transcripts are present in all four analyzed tissues in fresh samples, and do not drop out of detection for the entire time course.

The idea of a universal mRNA marker of sample age is attractive, as the existence of universal markers would facilitate the development of a generic sample age estimation assay. However, a universal assessment of mRNA degradation would not necessarily produce the most accurate estimation of sample age for every sample type. Our transcriptome abundance data for each tissue type over a 12-month or 6-month time course demonstrate that each tissue has an individual transcriptome degradation rate. The presence of a unique transcriptome degradation profile for each sample type is evidenced by the examination of the number of transcripts detected at each time point for each tissue type (Figure 9), and by the presence of unique degradation rates among identified tissue-specific mRNA markers (Figures 15-18). For instance, as previously discussed, our data indicate that the transcriptomes of blood and semen appear to be degrading at a slower rate than the transcriptomes of saliva and vaginal fluid.

The notion that transcripts degrade at different rates in different sample types is further supported by consideration of the identified universal mRNA markers for long-term sample age estimation. Take for example, the β-actin transcript (ACTB) presented in figure 20. While ACTB is present at a measurable abundance level in all of the analyzed time points for each of the sample types, the starting abundance and degradation rate (slope) for this transcript is not consistent across all tissues (Figure 20). Thus, previous sample age estimation assays utilizing β-actin may not be equally applicable to all sample types (S. E. Anderson et al.., 2011). Due to a unique abundance and rate of degradation of ACTB in each sample type, simply monitoring the abundance of this transcript on its own or in relation to another transcript would not necessarily correlate with sample age. If a universal mRNA marker is going to be applied in the estimation of sample age, it is

imperative that the analyst know the identity of the sample they are working with, so data analysis and conclusions can be adjusted to fit the known transcript degradation profile for that specific sample type. Due to the presence of distinct RNA degradation rates in each sample type, evaluation of fluid-specific markers may be more indicative of actual sample age.

Tissue-specific Markers of Sample Age

Upon comparison of transcript populations between each of the different fluid types, there were 1,449 transcripts identified as being unique to blood (Figure 19). For further analysis, the population of blood-specific transcripts was sorted into groups based on when transcript drop-out occurred (1, 2, 4, 6, 9, or 12 months, or drop-out not observed). Those transcripts that dropped-out by 1 month or 2 months are considered good candidates for short term age estimation. The transcripts that dropped out by 4, 6, or 9 months are considered good candidates for mid-term sample age estimation. The transcripts that did not drop out until 12 months or were found to have abundance in all of the sampled time points are considered good candidates for long-term sample age estimation.

With several possible candidates in each of the distinguished groups, transcripts with a similar time 0 abundance, but different drop-out times could be selected (Figure 22). This approach for selecting possible candidate markers for sample age estimation in blood would allow for the determination of the approximate age of a sample based on the presence and specific abundance of a set of transcripts. If multiple transcripts all have the same starting abundance, but drop-out of sequencing detection at different time points

(spanning 1 to 12 months), those transcripts are degrading at different rates. Transcripts, identified as having distinct degradation rates, can be exploited for sample age estimation. If these transcripts (or a representative, short-range, mid-range, and long-range marker) were monitored in bloodstains, the approximate age of the bloodstain could be estimated based on the presence or absence of specific transcripts and the abundance of the detected transcripts.

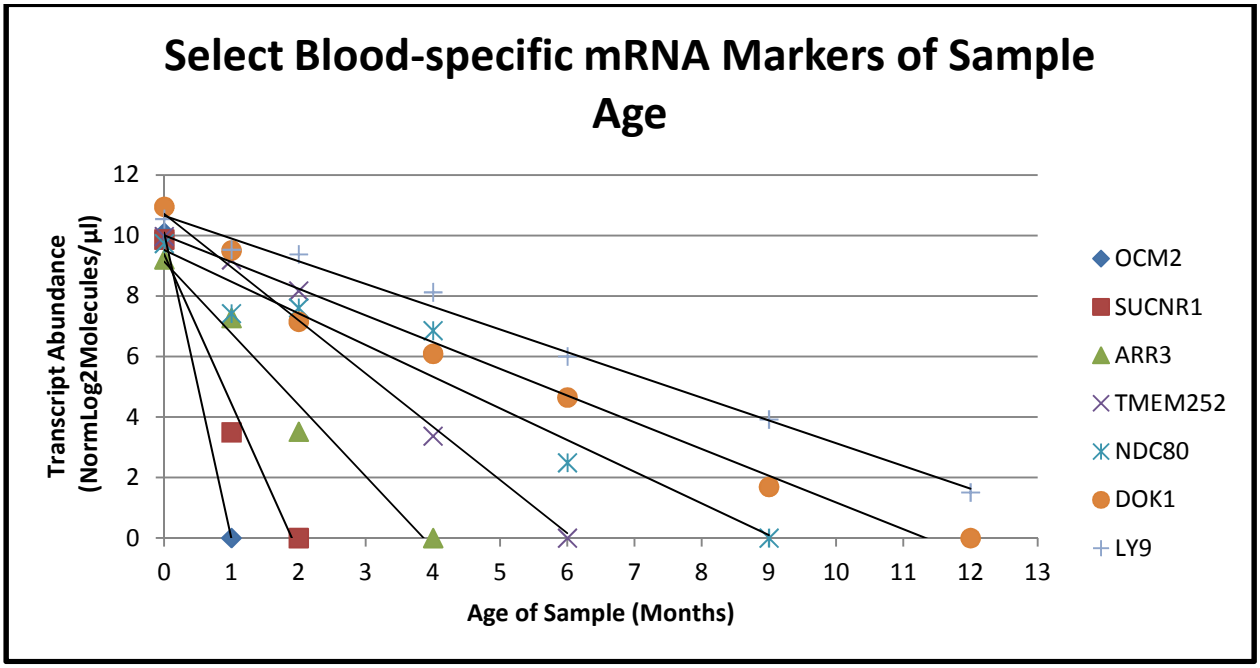


Figure 22. Select Blood-specific mRNA Markers of Sample Age. Select blood-specific mRNA markers for sample age estimation are displayed. One marker from each drop-out time (1, 2, 4, 6, 9, and 12 months and no-drop out observed) was chosen. The markers presented in this figure are representative of several possible markers present for each drop-out point. These markers were selected based on having similar time 0 abundance levels, which may facilitate their comparison when attempting to age a sample based on relative transcript abundance levels.

An example of how blood-specific transcript data may be used to estimate the approximate age of a sample can be found by reviewing figure 21. A sample of unknown

age could be profiled for the presence and abundance of these genes. The detection of OCM2, among other transcripts, would indicate that the sample was less than 30 days old. If, however, the only transcript detected in the sample was LY9, one might conclude that the sample was more than one year old. As this sequencing data are only composed of 2 samples, and has a required detection level of 200 bp fragments, these data are not definitive for estimating sample age and should only be used as a guide for the development of future assays that can be used to screen a larger number of samples. These data can provide an excellent roadmap for selection of transcripts that are likely to be successful biomarkers for short-, mid-, and long-range blood stain age estimation.

A similar data analysis pipeline was used with the other sample types to identify possible tissue-specific mRNA markers of sample age in saliva, vaginal fluid, and semen. The populations of sample-specific mRNA consisted of 124 transcripts specific to saliva, 211 transcripts specific to vaginal fluid, and 1,712 transcripts specific to semen. Transcripts specific for each tissue type were sorted based on when drop-out occurred (2 months, 4 months, 6 months, or drop-out not observed). Splitting the data into four groups based on transcript drop-out time allowed for the identification of possible tissue-specific markers of short-term age estimation (transcripts that disappeared by 2 months), mid-term age estimation (transcripts that disappeared by 4 or 6 months), and long-range age estimation (transcripts that have detectable abundance at all sampled time points). Select tissue-specific markers of sample age have been presented for saliva, vaginal fluid, and semen (Figure 23). As with blood, the selected markers for these fluids are representative of transcripts from each drop-out group (drop-out by 2, 4, and 6 months and drop-out not observed). Also similar to blood, the selected representative mRNA

markers of sample age in saliva, vaginal fluid, and semen were selected for their similar starting abundance values (abundance at time 0). It should be noted that the starting abundance values for the four vaginal fluid transcripts are more spread. The wide spread of starting abundances in vaginal fluid can be attributed to the absence of transcripts in each group that have similar starting abundances. Additionally, it should be noted that there is no saliva-specific transcript that drops-out of detection between 4 and 6 months.

As with blood, the identified markers for saliva, vaginal fluid, and semen may be indicative of sample age based on their presence or absence within a sample and their abundance. The selected markers are excellent candidates for further investigation with qPCR, allowing for a larger sample size to be evaluated.

The tissue-specific markers identified in this study may offer more power in estimating sample age then the universal markers that were identified. While more cumbersome for assay development, as one assay would potentially have to be developed per tissue, tissue-specific markers have the benefit of matching each specific biological fluid's unique degradation profile. While universal markers may be present in all sample types, these transcripts do not necessarily degrade at the same rate in all sample types. Tissue-specific markers allow for the selection of markers that have established degradation rates within a given sample type. This study has demonstrated that RNA degradation does not occur equally in all biological fluid types, thus sample –specific markers should be considered for the most accurate assessment of sample age.

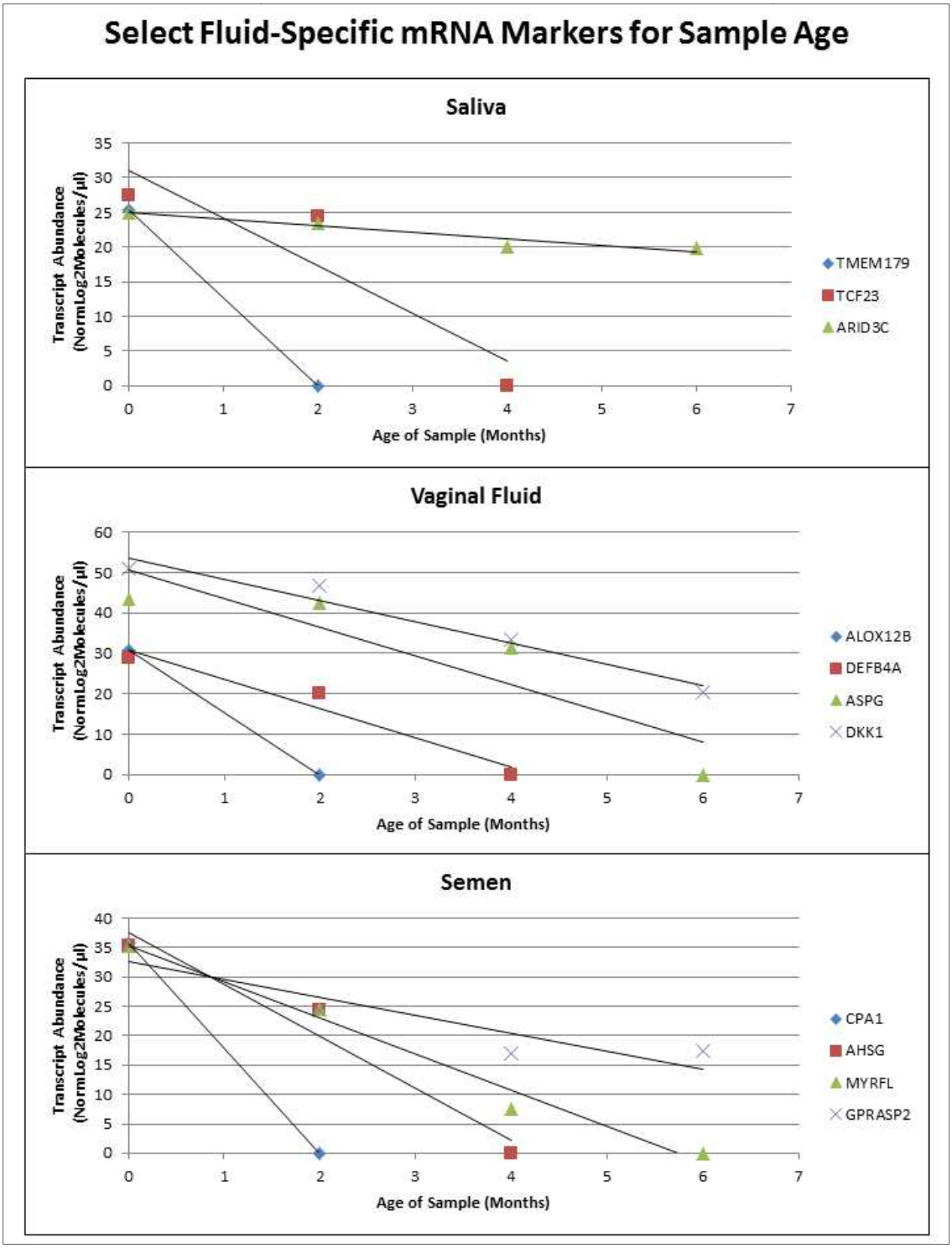


Figure 23. Select Tissue-specific mRNA Markers of Sample Age. Select fluid-specific mRNA markers for saliva, vaginal fluid, and semen are presented. One marker for each drop-out time (2, 4, and 6 months and no drop-out observed) is presented for each fluid. These markers are representative of larger groups of transcripts present for each drop-out time.

Conclusions

This study provides the first comprehensive analysis of *ex vivo* transcriptome degradation in four forensically relevant biological fluids. If RNA analysis is going to be adapted into routine forensic casework, it is critical that a more in-depth understanding of mRNA degradation patterns and profiles is established. In addition to providing baseline knowledge of the relationship between RNA degradation and sample age, establishing the behavior and profiles of *ex vivo* RNA samples is critical in other areas as well. If researchers and forensic personnel want to rely on the use of mRNA biomarkers for investigative purposes, including tissue identification and molecular autopsy findings, having a complete understanding of those transcripts in an *ex vivo* context is necessary. This study provides a database of mRNA transcripts present in fresh and aged samples at several time points spanning up to six months (saliva, vaginal fluid, and semen) or one year (blood). The transcriptome abundance data for each tissue type indicate that while mRNA is degrading in all sample types, specific transcript degradation rates can vary between different fluid types. The observed differences in transcript degradation rate can be attributed to a number of factors, including transcript starting abundance (level of abundance in a time 0, fresh sample), transcript environment (every fluid has a unique profile of cells and microbial organisms), and sample type.

The data generated in this study provide evidence that analyzing a single or a few generic transcripts (housekeeping, rRNA) may not be the most effective way to estimate sample age in a range of tissue and fluid types. Rather, the data generated in this study indicate that there are tissue-specific differences in RNA degradation rate that may affect the interpretation of RNA degradation data. One RNA assay and set of data analysis

guidelines may not be applicable to all sample types. Being aware that there are tissue-specific differences in RNA degradation will allow for the proper selection and analysis of tissue-specific markers for sample age estimation.

It is important to recognize that the age estimation markers identified for blood, saliva, vaginal fluid, and semen are selected from a sample size of two technical replicates at each time point for each fluid type. Due to the small sample size, these markers should not be considered absolute markers of sample age. Rather, these data should be used as guidance for selecting markers for further investigation with larger numbers of samples. The markers outlined in figures 21 and 22 are representative transcripts that have distinct degradation profiles in each of the tissue types. Thus, these markers warrant further investigation as possible markers for establishing sample age. Further investigation of markers identified from the full transcriptome degradation data for each sample type should be performed on lower cost, higher throughput technologies, such as RT qPCR. Simple qPCR assays could be designed to monitor identified transcripts for each fluid type and a larger number of samples could be screened to investigate the observed mRNA degradation trends for confirmation of their correlation to approximate sample age.

CHAPTER V

CONCLUSIONS

As research continues to evolve on the recovery and analysis of RNA from post-mortem tissues and forensic samples recovered from crime scenes, the possibility of RNA analysis playing a routine role in forensic casework increases. A critical point in the application of RNA analysis in forensic investigation is the development of a thorough understanding of RNA behavior in *ex vivo* samples. This study had two major goals concerning the analysis of mRNA in deposited samples. First, we developed an RNA-seq methodology and RNA-seq analysis pipeline for aged samples that consistently exceeded several stringent quality control measures that ensured that good sequencing data was obtained. The development of methods that accommodate degraded samples was imperative for the successful analysis of forensically relevant sample types. Additionally, we aimed to do all analysis on an RNA-seq platform as opposed to using the more traditional qPCR or capillary electrophoresis platforms, in order to gain a complete picture of the mRNA in aging samples. The second major goal of this study was the analysis of total mRNA in fresh and aged biological fluid samples of forensic relevance

(blood, saliva, vaginal fluid, and semen). The results of this study are the first comprehensive mRNA dataset in fresh and aged samples. The data produced in this study provide the first global look at how the transcriptome is fluctuating as deposited samples age. From these data, a greater understanding of specific transcript degradation rates and profiles can be gained for each sample type. In addition to outlining the degradation patterns of the transcriptomes of four different biological fluid stains, this study identified both universal and tissue-specific mRNA markers of sample age that warrant further investigation in a larger sample population. Specifically, the identification of these markers has the potential to facilitate assay development for assessing the age of deposited biological fluid samples.

Potential Impact

This research has potential impact not only in the field of forensic biology, but also in the field of medicolegal death investigation. This is the first organized study of full transcriptome degradation in human body fluids and tissues. While RNA degradation has been studied previously, only very few transcripts have ever been included in degradation analysis. The results of this study include total mRNA sequence data for a variety of sample types (blood, saliva, semen, and vaginal fluid) over several time points, spanning up to one year. This study has yielded the first comprehensive transcriptome sequencing dataset that includes fresh and aged biological samples over a period of several months to one year. These results provide a wealth of data, demonstrating mRNA degradation patterns and rates for each specific transcript present in each sample type. These transcriptome degradation data have the potential to aid investigators looking for

mRNA markers for post-mortem or forensic sample analysis. Use of RNA analysis is increasing in forensic investigations (including in the growing field of molecular autopsies). If investigators plan to use any specific mRNA transcripts in their investigations, it is critical that the *ex vivo* degradation pattern and rate for each transcript of interest be understood. The global data obtained from this study will provide an excellent starting point for investigators to determine the *ex vivo* degradation of individual transcripts of interest. Full transcriptome data from a fresh sample are valuable, but this dataset takes transcriptome sequencing further into the applicable realm of forensic science, providing data on aged and degraded samples as well. If RNA analysis is ever going to be successfully applied in forensic science, it is critical that investigators understand the behavior of transcripts in aged, as well as fresh samples.

Future Directions

While the results of this study are comprehensive in that they provide the first set of mRNA sequencing data for fresh and aged forensic samples, many questions were initiated by this research that deserve further investigation. In particular, more investigation is needed concerning the specific mechanism of *ex* vivo mRNA degradation. Additionally, deeper analysis of the microbiome mRNA population of saliva and vaginal fluid may be relevant to assessing the age of those sample types. Finally, the development of qPCR assays for further investigation of identified mRNA markers for sample age estimation will be performed in future work.

The results of this study indicate that the starting abundance of a transcript (the abundance of a transcript in the time 0, fresh sample) directly impact the degradation rate

of that sample. In general, transcripts that last longer in an *ex vivo* samples have higher starting abundances in the fresh sample (Figure 10). However, the results of this study also clearly indicate that the starting abundance of a transcript is not the only factor that affects transcript degradation rate. It is apparent that factors other than starting abundance are important when transcripts are easily identified in all tissue types that have similar starting abundance values but drop-out of sequencing detection at different time points (Figures 11, 12, 13 and 14). However, while it is apparent that other factors are influencing transcript degradation rate, more investigation is needed to determine what these specific factors are. Several factors should be evaluated in this effort, including transcript length and transcript secondary structure, both of which could have a direct influence on the stability of a molecule.

Two of the fluids surveyed for this study, saliva and vaginal fluid, have a large microbial population that influenced the amount of human mRNA that was recoverable from those sample types. While these samples did present a challenge due to the reduced amount of sequencing reads available for human transcripts, the mRNA of the microbiome of these samples does warrant further investigation for a possible correlation with sample age. On average, over 90% of the unaligned sequencing reads from the saliva and vaginal fluid samples mapped to either the HOMD or RefSeq databases. Initial assessment of the microbiome mRNA population has been performed for saliva (Figure 4, Appendix 2). The saliva mRNA data aligning to the HOMD database suggests that there is some fluctuation in the microbial populations of the saliva stains as they age. Evaluation of the microbiome of fresh samples in comparison to aged samples may reveal that there is a shift in the microbial population that corresponds with approximate

sample age. The vast majority of sequencing reads for both saliva and vaginal fluid aligned to microbial genomes (over 75% for both saliva and vaginal fluid), thus these sequencing reads warrant further investigation in future transcriptome analysis of these sample types.

The results of this study provide a baseline for the mRNA populations and degradation rates in individual biological fluid types. The data generated in this research will be used to develop qPCR assays for assessing the age of each sample type (blood, semen, saliva, and vaginal fluid) using data obtained from sequencing fresh and aged samples. The point of utilizing full transcriptome degradation data is so we can make an educated decision in selecting the most accurate markers for short-, mid-, and long-term age estimation for each individual sample type. Developed qPCR assays for sample age determination will be specifically designed to include the mRNA transcripts that have degradation rates and patterns most closely related to sample age, as reflected in the mRNA sequencing data for each sample type. The creation of qPCR assays allows for easy adaptation into high throughput, low cost sample analysis. Being able to determine the approximate age of a sample using a simple qPCR assay would greatly benefit both the fields of forensic biology and forensic pathology. Determining the time-since deposition of a sample would directly aid in providing a time-line of events surrounding a crime. While other qPCR assays have been developed to assess sample age, no assay is currently in regular use in forensic labs. By utilizing transcript degradation profiles generated from mRNA sequencing, the future development of qPCR assays will include transcript targets that are observed to degrade at specific rates in each tissue type, rather

than relying on the presence of housekeeping genes that are only assumed to degrade at a constant rate in every tissue type.

REFERENCES

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., … Levin, J. Z. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods*, *10*(7), 623–629. doi:10.1038/nmeth.2483

Anderson, S. E., Hobbs, G. R., & Bishop, C. P. (2011). Multivariate Analysis for Estimating the Age of a Bloodstain*. *Journal of Forensic Sciences*, *56*(1), 186–193. doi:10.1111/j.1556-4029.2010.01551.x

Anderson, S., Howard, B., Hobbs, G. R., & Bishop, C. P. (2005). A method for determining the age of a bloodstain. *Forensic Science International*, *148*(1), 37–45. doi:10.1016/j.forsciint.2004.04.071

Bauer, M. (2007). RNA in forensic science. *Forensic Science International. Genetics*, *1*(1), 69–74. doi:10.1016/j.fsigen.2006.11.002

Bauer, M., Gramlich, I., Polzin, S., & Patzelt, D. (2003). Quantification of mRNA degradation as possible indicator of postmortem interval--a pilot study. *Legal Medicine (Tokyo, Japan)*, *5*(4), 220–227.

Bauer, M., & Patzelt, D. (2008). Identification of menstrual blood by real time RT-PCR: Technical improvements and the practical value of negative test results. *Forensic Science International*, *174*(1), 55–59. doi:10.1016/j.forsciint.2007.03.016

Bauer, M., Polzin, S., & Patzelt, D. (2003). Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Science International*, *138*(1–3), 94–103. doi:10.1016/j.forsciint.2003.09.008

Catts, V. S., Catts, S. V., Fernandez, H. R., Taylor, J. M., Coulson, E. J., & Lutze-Mann, L. H. (2005). A microarray study of post-mortem mRNA degradation in mouse brain tissue. *Brain Research. Molecular Brain Research*, *138*(2), 164–177. doi:10.1016/j.molbrainres.2005.04.017

Fleming, R. I., & Harbison, S. (2010). The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Science International: Genetics*, *4*(4), 244–256. doi:10.1016/j.fsigen.2009.10.006

Fordyce, S. L., Kampmann, M.-L., Doorn, N. L. van, & Gilbert, M. T. P. (2013). Long-term RNA persistence in postmortem contexts. *Investigative Genetics*, *4*(1), 7. doi:10.1186/2041-2223-4-7

Haas, C., Klesser, B., Maake, C., Bär, W., & Kratzer, A. (2009). mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Science International: Genetics*, *3*(2), 80–88. doi:10.1016/j.fsigen.2008.11.003

Haas, C., Muheim, C., Kratzer, A., Bär, W., & Maake, C. (2009). mRNA profiling for the identification of sperm and seminal plasma. *Forensic Science International: Genetics Supplement Series*, *2*(1), 534–535. doi:10.1016/j.fsigss.2009.08.109

Heinrich, M., Matt, K., Lutz-Bonengel, S., & Schmidt, U. (2007). Successful RNA extraction from various human postmortem tissues. *International Journal of Legal Medicine*, *121*(2), 136–142. doi:10.1007/s00414-006-0131-9

Ikematsu, K., Takahashi, H., Kondo, T., Tsuda, R., & Nakasono, I. (2008). Temporal expression of immediate early gene mRNA during the supravital reaction in mouse brain and lung after mechanical asphyxiation. *Forensic Science International*, *179*(2-3), 152–156. doi:10.1016/j.forsciint.2008.05.007

Ikematsu, K., Tsuda, R., & Nakasono, I. (2006). Gene response of mouse skin to pressure injury in the neck region. *Legal Medicine (Tokyo, Japan)*, *8*(2), 128–131. doi:10.1016/j.legalmed.2005.08.012

Inoue, H., Kimura, A., & Tuji, T. (2002). Degradation profile of mRNA in a dead rat body: basic semi-quantification study. *Forensic Science International*, *130*(2–3), 127–132. doi:10.1016/S0379-0738(02)00352-3

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., … Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, *21*(9), 1543–1551. doi:10.1101/gr.121095.111

Kagawa, S., Matsuo, A., Yagi, Y., Ikematsu, K., Tsuda, R., & Nakasono, I. (2009). The time-course analysis of gene expression during wound healing in mouse skin. *Legal Medicine (Tokyo, Japan)*, *11*(2), 70–75. doi:10.1016/j.legalmed.2008.09.004

Kimura, A., Ishida, Y., Hayashi, T., Nosaka, M., & Kondo, T. (2011). Estimating time of death based on the biological clock. *International Journal of Legal Medicine, 125*(3), 385–391. doi:10.1007/s00414-010-0527-4

Kohlmeier, F., & Schneider, P. M. (2012). Successful mRNA profiling of 23 years old blood stains. *Forensic Science International: Genetics*, *6*(2), 274–276. doi:10.1016/j.fsigen.2011.04.007

Koppelkamm, A., Vennemann, B., Lutz-Bonengel, S., Fracasso, T., & Vennemann, M. (2011). RNA integrity in post-mortem samples: influencing parameters and implications on RT-qPCR assays. *International Journal of Legal Medicine*, *125*(4), 573–580. doi:10.1007/s00414-011-0578-1

Liang, Y., Ridzon, D., Wong, L., & Chen, C. (2007). Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, *8*(1), 166. doi:10.1186/1471-2164-8-166

Lindenbergh, A., de Pagter, M., Ramdayal, G., Visser, M., Zubakov, D., Kayser, M., & Sijen, T. (2012). A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Science International. Genetics*, *6*(5), 565–577. doi:10.1016/j.fsigen.2012.01.009

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Processing of rRNA and tRNA*. Text. Retrieved February 12, 2013, from http://www.ncbi.nlm.nih.gov/books/NBK21729/

Matsuo, A., Ikematsu, K., & Nakasono, I. (2009). C-fos, fos-B, c-jun and dusp-1 expression in the mouse heart after single and repeated methamphetamine administration. *Legal Medicine*, *11*(6), 285–290. doi:10.1016/j.legalmed.2009.09.002

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. doi:10.1038/nmeth.1226

Oehmichen, M., & Zilles, K. (1984). Postmortale DNS- und RNS-Synthese: Erste Untersuchungen an menschlichen Leichen. *Zeitschrift für Rechtsmedizin*, *91*(4), 287–294. doi:10.1007/BF02332322

Park, S.-M., Park, S.-Y., Kim, J.-H., Kang, T.-W., Park, J.-L., Woo, K.-M., … Lee, S.-H. (2012). Genome-wide mRNA profiling and multiplex quantitative RT-PCR for forensic body fluid identification. *Forensic Science International: Genetics*. doi:10.1016/j.fsigen.2012.09.001

Partemi, S., Berne, P. M., Batlle, M., Berruezo, A., Mont, L., Riuró, H., … Oliva, A. (2010). Analysis of mRNA from human heart tissue and putative applications in forensic molecular pathology. *Forensic Science International*, *203*(1-3), 99–105. doi:10.1016/j.forsciint.2010.07.005

Pertea, M. (2012). The Human Transcriptome: An Unfinished Story. *Genes*, *3*(4), 344–360. doi:10.3390/genes3030344

Phang, T., Shi, C., Chia, J., & Ong, C. (1994). Amplification of cDNA via RT-PCR using RNA extracted from postmortem tissues. *Journal of Forensic Sciences*, *39*(5), 1275–1279.

Preece, P., & Cairns, N. J. (2003). Quantifying mRNA in postmortem human brain: influence of gender, age at death, postmortem interval, brain pH, agonal state and inter-lobe mRNA variance. *Brain Research. Molecular Brain Research*, *118*(1-2), 60–71.

Raghavachari, N., Barb, J., Yang, Y., Liu, P., Woodhouse, K., Levy, D., … Kato, G. J. (2012). A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Medical Genomics*, *5*(1), 28. doi:10.1186/1755-8794-5-28

Richard, M. L. L., Harper, K. A., Craig, R. L., Onorato, A. J., Robertson, J. M., & Donfack, J. (2012). Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis. *Forensic Science International: Genetics*, *6*(4), 452–460. doi:10.1016/j.fsigen.2011.09.007

Sakurada, K., Akutsu, T., Fukushima, H., Watanabe, K., & Yoshino, M. (2010). Detection of dermcidin for sweat identification by real-time RT-PCR and ELISA. *Forensic Science International*, *194*(1–3), 80–84. doi:10.1016/j.forsciint.2009.10.015

Sakurada, K., Akutsu, T., Watanabe, K., Fujinami, Y., & Yoshino, M. (2011). Expression of statherin mRNA and protein in nasal and vaginal secretions. *Legal Medicine*, *13*(6), 309–313. doi:10.1016/j.legalmed.2011.07.002

Sampaio-Silva, F., Magalhães, T., Carvalho, F., Dinis-Oliveira, R. J., & Silvestre, R. (2013). Profiling of RNA Degradation for Estimation of Post Morterm Interval. *PLoS ONE*, *8*(2), e56507. doi:10.1371/journal.pone.0056507

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., & Ko, M. S. H. (2009). Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *16*(1), 45–58. doi:10.1093/dnares/dsn030

Thibaut. (n.d.). Human BodyMap 2.0 data from Illumina. *Ensembl Blog*. Retrieved from http://www.ensembl.info/blog/2011/05/24/human-bodymap-2-0-data-from-illumina/

Vass, A., Fleming, R., Harbison, S., Curran, J., & Williams, E. (2013, May). Evaluating the Use of DNA and RNA Degradation for Estimating the Post-Mortem Interval. NCJRS.

Vennemann, M., & Koppelkamm, A. (2010a). mRNA profiling in forensic genetics I: Possibilities and limitations. *Forensic Science International*, *203*(1-3), 71–75. doi:10.1016/j.forsciint.2010.07.006

Vennemann, M., & Koppelkamm, A. (2010b). Postmortem mRNA profiling II: Practical considerations. *Forensic Science International*, *203*(1-3), 76–82. doi:10.1016/j.forsciint.2010.07.007

Visser, M., Zubakov, D., Ballantyne, K., & Kayser, M. (2011). mRNA-based skin identification for forensic applications. *International Journal of Legal Medicine*, *125*(2), 253–263. doi:10.1007/s00414-010-0545-2

Young, S. T., Wells, J. D., Hobbs, G. R., & Bishop, C. P. (2013). Estimating postmortem interval using RNA degradation and morphological changes in tooth pulp.

Forensic Science International, *229*(1-3), 163.e1–6. doi:10.1016/j.forsciint.2013.03.035

Zhao, D., Ishikawa, T., Quan, L., Li, D.-R., Michiue, T., Yoshida, C., … Maeda, H. (2008). Tissue-specific differences in mRNA quantification of glucose transporter 1 and vascular endothelial growth factor with special regard to death investigations of fatal injuries. *Forensic Science International*, *177*(2–3), 176–183. doi:10.1016/j.forsciint.2007.12.004

Zubakov, D., Boersma, A., Choi, Y., van Kuijk, P., Wiemer, E., & Kayser, M. (2010). MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine*, *124*(3), 217–226. doi:10.1007/s00414-009-0402-3

Zubakov, D., Kokshoorn, M., Kloosterman, A., & Kayser, M. (2009). New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. *International Journal of Legal Medicine*, *123*(1), 71–74. doi:10.1007/s00414-008-0249-z

APPENDICES

APPENDIX A: Comprehensive Literature Search Results for Tissue-specific mRNA Markers

Venous Blood		
Marker	**Cross-Reactivity**	**Source**
CD93	menstrual blood	Lindenbergh (2012) Zubakov (2009) Zubakov (2008)
AMICA1	menstrual blood	Lindenbergh (2012) Zubakov (2009) Zubakov (2008)
HBB	menstrual blood	Lindenbergh (2012) Haas (March 2009, Jan 2011, Nov 2011) Wobst (2011)
ALAS2	semen	Richard (2012) Juusola (2007) Haas (Nov 2011)
PPBP		Park (2012)
Beta-Spectrin (SPTB)	vaginal secretion	Patel (2008) Haas (2008) Juusola (2005, 2007) Haas (March 2009, Nov 2011)
porphobilinogen deaminase (PBGD)		Juusola (2005) Patel (2008) Haas (2008, March 2009, Jan 2011, Nov 2011) Wobst (2011)

HBA		Nussbaumer (2006) Haas (Nov 2011)
ALOX5AP		Zubakov (2008, 2009)
AQP9		Zubakov (2008, 2009) Haas (Nov 2011)
ALOX5AP		Zubakov (2008, 2009)
AQP9		Zubakov (2008, 2009) Haas (Nov 2011)
ARHGAP26		Zubakov (2008, 2009)
C1QR1		Zubakov (2008, 2009)
C5R1		Zubakov (2008, 2009)
CASP1		Zubakov (2008, 2009)
MNDA		Zubakov (2008, 2009)
NCF2		Zubakov (2008, 2009)
Ankyrin 1 (ANK1)		Fang (2006) Haas (Nov 2011)
CD3G		Haas (Nov 2011)
Glycophorin A		Fleming (2010)

Saliva		
Marker	**Cross-Reactivity**	**Source**
KRT4	vaginal mucosa, menstrual secretion, and skin samples	Lindenbergh (2012) Zubakov (2009) Zubakov (2008)
KRT13	vaginal mucosa, menstrual secretion, and skin samples	Lindenbergh (2012) Zubakov (2009) Zubakov (2008)
SPRR2A	vaginal mucosa, menstrual secretion, and skin samples	Lindenbergh (2012) Zubakov (2009) Zubakov (2008)
STATH (Statherin)	nasal secretions	Lindenbergh (2012) Richard (2012) Sakurada (2009) Patel (2008) Juusola (2003, 2005, 2007) Haas (2008, March 2009) Wobst (2011) Fleming (2010)

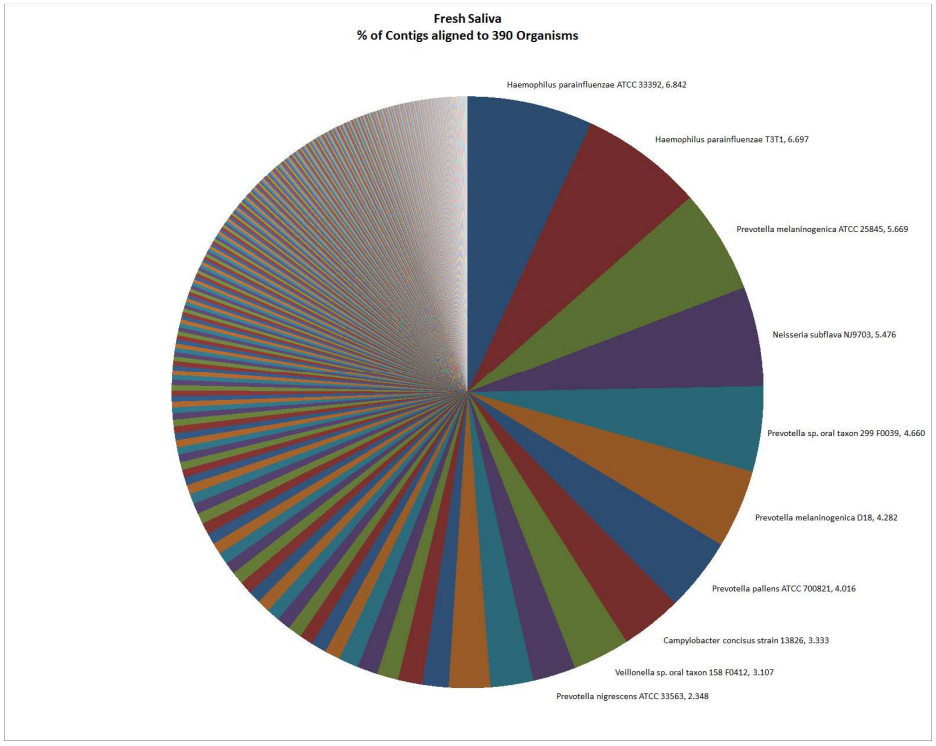
HTN3 (histatin 3)		Lindenbergh (2012) Richard (2012) Sakurada (2009) Patel (2008) Juusola (2003, 2005, 2007) Haas (2008, March 2009) Wobst (2011) Fleming (2010)
FDCSP		Park (2012)
PRB1		Juusola (2003)
PRB2		Juusola (2003)
PRB3		Juusola (2003)
PRB4		Fang (2006)
SPRR1A		Zubakov (2008, 2009)
KRT6A		Zubakov (2008, 2009)

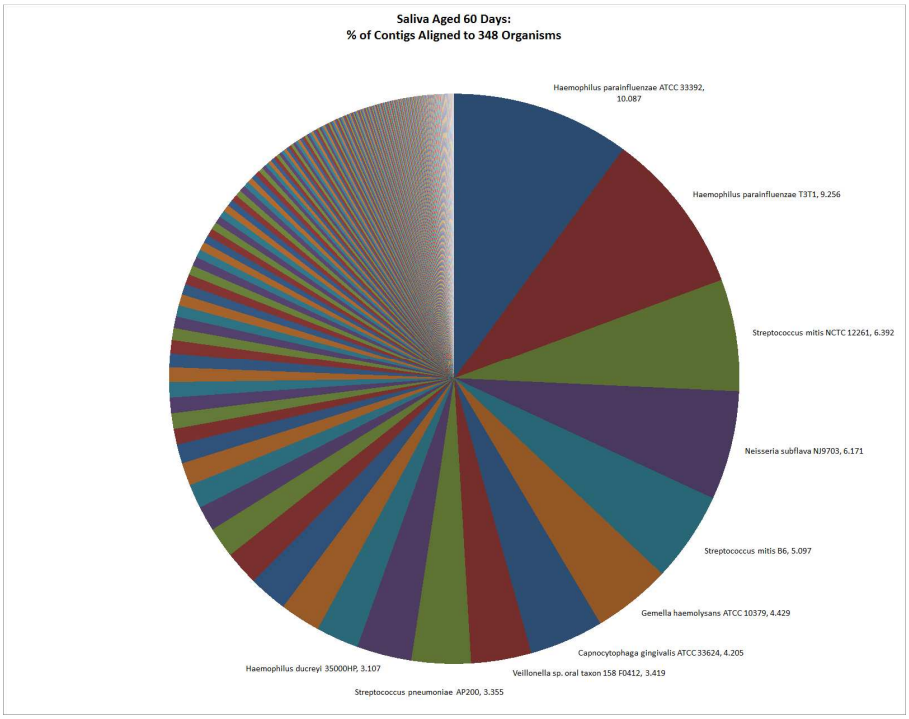
Semen		
Marker	**Cross-Reactivity**	**Source**
PRM1 (Marker for Sperm)		Lindenbergh (2012) Patel (2008) Haas (2008, March 2009, Dec 2009) Juusola (2005, 2007) Bauer (2003) Wobst (2011)
SEMG1		Lindenbergh (2012) Sakurada (2009) Haas (Dec 2009) Fang (2006)
PRM2		Richard (2012) Sakurada (2009) Patel (2008) Haas (2008, March 2009, Dec 2009) Juusola (2005, 2007) Bauer (2003) Fleming (2010)

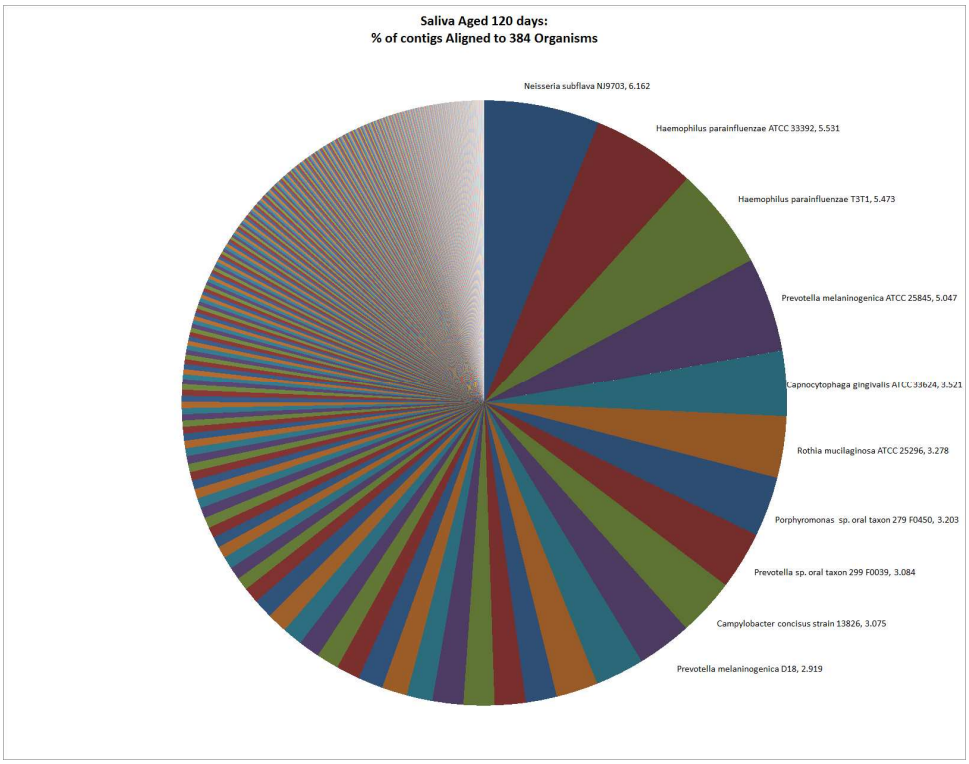
TGM4		Richard (2012) Fang (2006) Wobst (2011) Fleming (2010)
MSMB		Park (2012)
KLK (PSA)		Nussbaumer (2006) Haas (Dec 2009)
MCSP		Fang (2006)

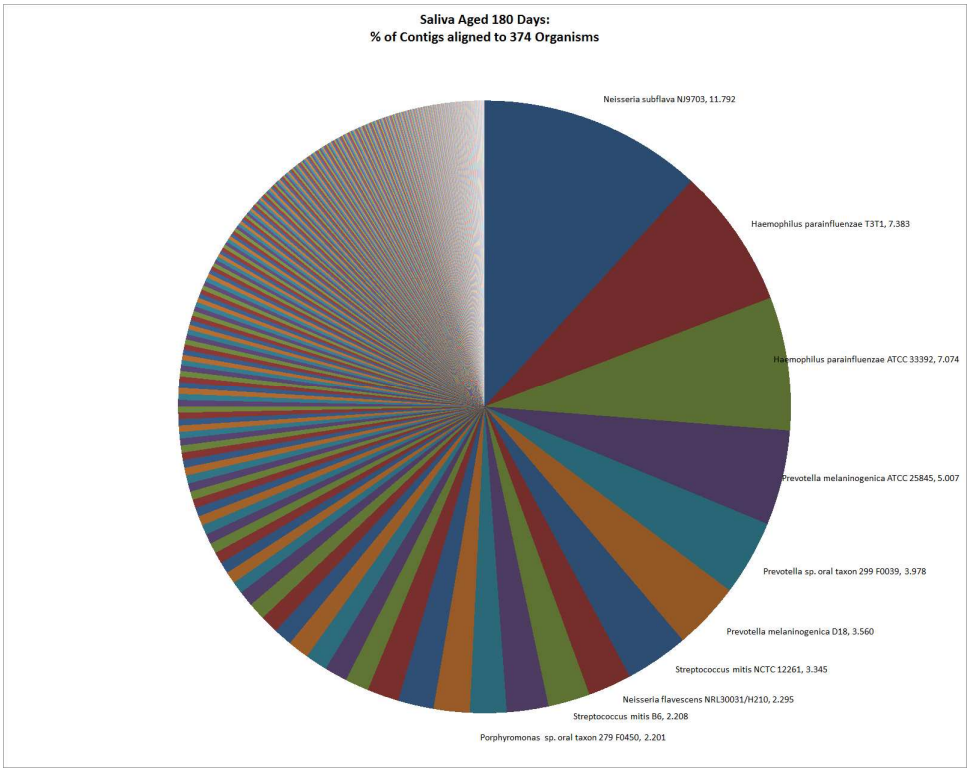
Vaginal Secretion (Mucosa)		
Marker	**Cross-Reactivity**	**Source**
MUC4	Saliva	Lindenbergh (2012) Richard (2012) Patel (2008) Haas (2008, March 2009) Nussbaumer (2006) Juusola (2005)
HBD1		Lindenbergh (2012) Patel (2008) Haas (2008, March 2009) Juusola (2005)
MSLN		Park (2012)
CYP2B7P1		Hanson (2012)
MYOZ1		Hanson (2012)
ESR1	Semen, Saliva	Fang (2006)
16S-23S rRNA intergenic spacer region for Lactobacillus gasseri (GASS)	menstrual blood	Wobst (2011)

Appendix B: Percent of Contigs Aligned to the HOMD Database for Saliva RNA-seq Samples






Saliva Aged 120 days:
% of contigs Aligned to 384 Organisms
Neisseria subflava NJ9703, 6.162
Haemophilus parainfluenzae ATCC 33392, 5.531
Haemophilus parainfluenzae T3T1, 5.473
Prevotella melaninogenica ATCC 25845, 5.047
Capnocytophaga gingivalis ATCC 33624, 3.521
Rothia mucilaginosa ATCC 25296, 3.278
Porphyromonas sp. oral taxon 279 F0450, 3.203
Prevotella sp. oral taxon 299 F0039, 3.084
Campylobacter concisus strain 13826, 3.075
Prevotella melaninogenica D18, 2.919


Saliva Aged 180 Days:
% of Contigs aligned to 374 Organisms
Neisseria subflava NJ9703, 11.792
Haemophilus parainfluenzae T3T1, 7.383
Haemophilus parainfluenzae ATCC 33392, 7.074
Prevotella melaninogenica ATCC 25845, 5.007
Prevotella sp. oral taxon 299 F0039, 3.978
Prevotella melaninogenica D18, 3.560
Streptococcus mitis NCTC 12261, 3.345
Neisseria flavescens NRL30031/H210, 2.295
Streptococcus mitis B6, 2.208
Porphyromonas sp. oral taxon 279 F0450, 2.201

VITA

Kate Weinbrecht

Candidate for the Degree of

Doctor of Philosophy

Thesis: RNA-SEQ OF BIOLOGICAL FLUIDS FOR THE EVALUATION OF MRNA DEGRADATION IN RELATION TO SAMPLE AGE

Major Field: Biomedical Sciences

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Biomedical Sciences at Oklahoma State University, Tulsa, Oklahoma in July, 2014.

Completed the requirements for the Master of Science in Forensic Sciences at Oklahoma State University, Tulsa, Oklahoma in July, 2011.

Completed the requirements for the Bachelor of Science in Biology at Pacific Lutheran University, Tacoma, Washington in May, 2009.

Experience:

Graduate Assistant, School of Forensic Sciences, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma (January, 2011 to June, 2011).

Undergraduate Genetics Lab, Adjunct Instructor, Northeastern State University Tulsa, Oklahoma (August, 2013 to December, 2013).

Graduate Assistant, School of Biomedical Sciences, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma (June, 2011 to July, 2014).