DETECTION OF RNA METHYLATION PATTERNS IN FORENSICALLY RELEVANT TRANSCRIPTS IN DRIED BLOODSTAINS

By

CHAELYNNE LOHR

Bachelor Science in Human Biology

Michigan State University

East Lansing, MI

2019

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2022

DETECTTION OF RNA METHYLATION PATTERNS

IN FORENSICALLY RELEVANT TRANSCRIPTS

IN DRIED BLOODSTAINS

Thesis Approved:

Dr. Robert W Allen

Thesis Adviser

Dr. Jun Fu

Dr. Gerwald Koehler

ACKNOWLEDGEMENTS

I would like to extend my gratitude to my research advisor, Dr. Robert Allen, for his continual guidance and support throughout not only this project, but my time at Oklahoma State University, Center for Health Sciences as he truly made me feel welcomed into the program. I would also like to sincerely thank Dr. Jun Fu for her dedication to my research and input with experimental methods and data interpretation. Her daily conversations and charisma made campus an enjoyable place to be every day. My gratitude also extends to Dr. Gerwald Koehler for his knowledge, input, and advice throughout my research. Without my committee members this project would not have come to fruition.

I would also like to thank the friends that I made throughout my time in the Forensic Science program. Their encouragement, listening ear, and advice did not go unnoticed, and I enjoyed exploring a new city with each of them. Elizabeth, Kayla, Ivette, Sadie, and Gwen, thank you for helping make Tulsa my home for the past year.

To my long-time mentor, boss, and friend Dr. Casey Droscha, thank you for your continued support while I left Michigan and completed my dream of earning a master's degree in Forensic Science. I truly accredit my success to your teaching, guidance, and constant pushing me beyond my capabilities.

Lastly, but certainly not least, I would like to thank my family and friends in Michigan. I felt your constant support from 900 miles away, and your frequent calls, facetimes, and visits made being away from home easier. To my significant other,

Jonathan August, thank you for the countless overnight drives, plane tickets and facetimes. Your constant support while I moved across the country made the transition that much easier, and I could not have done this without you by my side.

I dedicate this thesis to my late father, Dale Lohr. Although you are no longer here to provide your words of wisdom and guidance, I felt your presence and know that you are watching from above.

Name: Chaelynne Lohr

Date of Degree: July, 2022

Title of Study: DETECTION OF RNA METHYLATION PATTERNS IN FORENSICALLY RELEVANT TRANSCRIPTS IN DRIED BLOODSTAINS

Major Field: Forensic Sciences

Abstract: RNA degradation kinetics can be used to estimate the age of a biological sample found at a crime scene. RNA sequencing of transcripts from various tissue types shows that degradation occurs faster at the 5' end than the 3' end. This discovery led to the development of the 5'-3' assay, which quantifies and compares each end of a transcript to estimate sample age. This assay has been validated on dried bloodstains, however why the 5' end of the transcript degrades faster than the 3' end remains unknown. As this phenomenon is being observed in dried bloodstains, we hypothesize that chemical modifications of RNA molecules may be playing a role. Increasing research shows that methylation of RNA molecules can alter the kinetics of RNA degradation, leading to either increased or decreased transcript stability, dependent on the RNA binding proteins (RBPs) present. This study aims to investigate the methylation patterns in forensically relevant transcripts in dried bloodstains via next generation sequencing (NGS) using Oxford Nanopore Technologies' (ONT) native RNA sequencing protocol followed by bioinformatic analysis. We also developed a novel RNA enrichment technique that utilizes 120 nucleotide DNA probes designed to hybridize to the 5' or 3' end of a transcript for selected target enrichment. We detected possible modifications on twelve transcripts sequenced from dried bloodstains, five of which were identified on more than one sample. Although the results of this study are preliminary due to the lack of controls for statistical comparison, the duplicate modifications identified strengthens the overall findings. The results provided here can serve as the basis for future studies aiming to interrogate the epitranscriptome of forensically relevant samples to further optimize RNA-based assays for implementation into the field of forensic science.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Historically, the field of forensic biology has focused on DNA analysis techniques to aid in criminal investigations, mainly with its use in human identification. However, more recent research has expanded to RNA and its potential use in a forensic setting. Although a relatively new area of study, multiple uses of RNA in forensics have already been recognized. For example, the tissue of origin for an evidentiary sample can be identified using RNA's tissue specific expression pattern[1-4]. Furthermore, RNA has unique degradation kinetics that have proven useful in aging biological samples[5-9]. In both instances, RNA can be used to determine the relevance of a sample to the crime being investigated. Establishing sample relevance is crucial for preservation of laboratory resources, however much research needs to be done before this technology can be implemented into the field. More specifically, increased knowledge is needed surrounding RNA's complex structure to fully understand the molecule's unique degradation patterns and how they can be applied to the field of forensic science.

There are numerous differences between RNA and DNA that allow each molecule to contribute to criminal investigations in different ways, including but not limited to their chemical composition and molecular structure. RNA contains a ribose sugar, while the sugar on DNA is a deoxyribose, which lacks a hydroxyl group on the 2-prime carbon. Furthermore, RNA contains the base uracil in place of thymine found in DNA. Finally, in contrast to DNA's double-stranded, stable structure, RNA is single-stranded and is thus more susceptible to degradation. Due to the

enhanced degradation of RNA in crime scene samples, several investigators have suggested a correlation between the degradation state of RNA in a stain and the length of time the stain has existed at a crime scene, also known as the time since deposition (TSD). This can be critical information in establishing the relevance of a sample to a crime under investigation.

RNA Sequencing data from Oklahoma State University, School of Forensic Science, provides insights into how RNA behaves in aged bloodstains, revealing that the 5' end of numerous messenger RNA (mRNA) transcripts degrades at a faster rate than that of the 3' end[9]. This information was subsequently utilized for the development and validation of qPCR assay capable of aging dried bloodstains[5]. However, the underlying mechanism responsible for the observed differential degradation rates at the 5' and 3' ends of a transcript remains unknown. Understanding why mRNA transcripts are behaving in such a way is an imperative next step for validating assays for sample age estimation, as it will provide increased knowledge to further explain why RNA degrades as it does in dried bloodstains.

RNA degradation can be affected by external factors such as temperature and humidity[7] but may also be influenced by internal features such as chemical modifications to ribonucleotides[10-13]. Thus, studying RNA not only requires a basis of knowledge of how it behaves, but also a set of molecular tools that can aid in interrogating its molecular structure. As RNA research continues to grow, the number of molecular tools available for analyzing the molecule also expands. In recent years, the molecular biology field has gained access to resources that enable researchers to manipulate RNA to focus on transcripts of interest[14], as well as apply next generation sequencing (NGS) techniques to native RNA for epitranscriptomic interrogation[15,16], both of which are used in the present study to expand RNA research in a forensic setting. The infancy of these techniques being applied to RNA creates challenges and limitations, but also opportunity for an extensive expansion of knowledge.

The present study builds on Oklahoma State University, School of Forensic Science's knowledge of RNA degradation patterns. As the differential degradation patterns observed occur in dried stains, we believe the stability of each end of a transcript may be impacted by chemical modifications to ribonucleotides present in the RNA transcript, particularly methylation. In the study reported here, possible chemical modification(s) were explored within mRNA transcripts in dried bloodstains employing two experimental approaches. First, a method utilizing selective mRNA enrichment was developed to hybrid-select desired transcripts for downstream analyses via liquid chromatography coupled with mass spectrometry (LC-MS). In a second approach, direct sequencing of native RNA using Oxford Nanopore Technologies (ONT) was used to interrogate the epitranscriptome of all poly(A) tailed transcripts in dried bloodstains. Using previous research and novel techniques, the current study provides optimized protocols for RNA analysis, and preliminary results that start to unravel why RNA degrades the way it does when extracted from a crime scene sample.

CHAPTER II

REVIEW OF LITERATURE

RNA in Forensics

The use of RNA analysis to aid in forensic investigations is a relatively new tool with many potential applications identified since its introduction into the field, including body fluid and tissue identification[1-4], wound age determination[17], post-mortem interval (PMI) estimation[18], functionality of cells and organs[19,20], and aging biological stains[5-9]. Determining the age of a sample found at a crime scene is crucial for establishing evidence relevance to the crime being investigated. This additional information can be used to determine if further analyses of the sample are needed, helping to preserve time and laboratory resources. Traditionally, RNA degradation assays were designed to compare the degradation rates of either tissue specific markers of different stabilities[6-8,21], or housekeeping and sample specific markers[6,21,22]. Although each approach showed promise, more reliable methods were needed to predict the age of samples more accurately from forensically relevant tissues.

More recently, RNA sequencing results from Oklahoma State University, School of Forensic Science, revealed that the 5' end of several mRNA transcripts in blood degrades faster than the 3' end, based upon the sequencing read depth in stains as they aged[9]. These results led to the development of the 5'-3' qPCR assay[5]. Briefly, the 5'-3' assay is a quantitative polymerase chain reaction (qPCR) that uses two primer pairs to amplify regions at each end of an RNA transcript. Quantification cycle (Cq) values are then used to calculate a delta-Cq value, giving the

difference in abundance between the 5' and 3' ends[5]. Using this approach, researchers were able to accurately estimate sample age within 2-4 weeks for samples aged up to 6 months, and 4-6 weeks for samples aged 6 months to one-year old[5]. The 5'-3' assay confirmed that the 5' ends of the transcripts were in fact degrading faster than the 3' ends[5], however the assay does not address why this is happening. Literature suggests that the poly(A) tail, located at the 3' end of mRNA, works to stabilize the molecule and protect it from degradation[10]. However, further studies revealed that the link between the poly(A) tail and RNA stability is complex and that there are likely more factors playing a role in RNA degradation kinetics[10]. Since this phenomenon is being observed in dried bloodstains, the differential degradation rates may be due to chemical modifications present on the RNA molecules.

RNA Degradation

RNA degradation is a complex process; however, its understanding is crucial for insight into the various functions and forensic applications of RNA analysis. In living organisms, the degradation of RNA in living cells is essential for regulation of coding and non-coding genes, as well as eliminating faulty RNAs, both of which can lead to disease if not controlled[10]. RNA degradation occurs in all organisms- with several aspects the same or similar amongst eukaryotes and prokaryotes[23]. *Ex vivo* samples also undergo RNA degradation and could be relevant to a forensic investigation. Three mechanisms facilitating RNA degradation in post-mortem samples have been proposed and reflect enzymatic, physical, and chemical reactions[24]. Enzymatic degradation occurs naturally, especially in living tissues, and is mediated by a family of RNases present in tissues. In contrast, degradation facilitated by physical and/or chemical processes is affected by environmental conditions[24], as they often occur in body fluid stains or post-mortem tissues. Furthermore, some believe underlying chemical modifications of RNA may play a role in the molecule's degradation kinetics[10,24].

Chemical Modifications of RNA

The study of chemical modifications on RNA molecules is a relatively new concept that has been given the name epitranscriptomics[25]. Epitranscriptomics is a field similar to epigenetics, which studies chemical modifications of the deoxynucleotides of DNA; however, in epitranscriptomics the modifications are made to ribonucleotides. Perhaps importantly, there is more than 10 times the number of modifications reported on RNA compared to DNA[26]. Chemical modifications can influence RNA processing and stability, and thus influence the kinetics of its degradation. Furthermore, substantial evidence exists suggesting that the kinetics of RNA degradation correlates with the age of a forensic sample[5-8,21]. Consequently, any potential role of chemical modifications in promoting or protecting RNA degradation could further our ability to precisely estimate the age of a biological sample.

According to the MODOMICS database, which provides comprehensive information surrounding RNA modifications, around 144 RNA modifications have been identified in both coding and non-coding regions of the RNA molecule[27]. Additionally, chemical modifications have been identified on all types of RNA, including mRNA, transfer RNA (tRNA), and ribosomal RNA (rRNA)[28]. RNA modifications can be reversible or irreversible, and may affect transcription, mRNA processing, RNA export, translation, and/or degradation[11]. In living cells, all RNA modifications are controlled by three main effector proteins. Writer proteins are responsible for depositing the modification onto RNA by transferring a chemical group to the molecule[11]. The second group of proteins that recognize the modified nucleotides are reader proteins, also known as RNA binding proteins (RBPs)[11]. The final group of molecules involved in chemical modification are eraser proteins, which can remove the chemical group and return the RNA to its unmodified state[11]. Several possible chemical modifications of RNA have been identified. Among the most researched is the methylation of adenosines, oxidation of guanosines, isomerization of uridines, methylation of cytosines, and acetylation of cytosines[11].

N^6-Methyladenosine

Of the RNA chemical modifications cited in literature, N^6-methyladenosine (m6A) is the most abundant [11,29] and well-characterized[30] internal modification of mRNA. The modification m6A is relatively stable[31], reversible by demethylation[31], and located at either Gm^6AC or Am^6AC sites[11]. It is found on three to five nucleotides of mRNA on average, accounting for 0.1-0.4% of total adenosines in a typical mRNA[29]. The m6A modification is enriched in the 3'-UTR[32], near stop codons[32], and in the last coding exon[33], but is absent from poly(A) tails[32]. This modification is installed by the methyltransferase complex (the "writer"), composed mainly of methyltransferase like 3 (METTL3), methyltransferase like 14 (METTL14), WTAP, and KIAA1429[2,17,19]. Multiple m6A reader proteins have been identified, including but not limited to, YTHDF2, insulin-like growth factor 2 (IGF2BP), fragile X mental retardation protein (FMRP), and human antigen R (HuR)[11-13,30]. The m6A modification has two known eraser proteins. The α-ketoglutarate-dependent dioxygenase alkB homolog 5 protein (ALKBH5) is m6A specific and is the modification's primary demethylase[11]; however, the fat mass and obesity-associated protein (FTO) can also demethylate m6A modifications[11,30]. The primary functions of m6A are transcriptome turnover, translation regulation, cell differentiation, embryonic development, and stress response[30], with its best described function being the control of mRNA stability[12,29].

The effect of m6A on RNA stability in living cells depends largely on the RBPs present. Certain m6A RBPs signal pathways that destabilize and degrade an mRNA transcript thereby silencing gene expression, while others increase a transcripts stability through different molecular pathways[11]. There are two main hypotheses as to how this occurs. Research shows that different proteins can recognize and bind to m6A modifications, marking the molecule for downstream events in living cells relating to stability or degradation[12]. For example, the m6A reader protein YTHDF2 decreases mRNA transcript stability by recruiting the CNOT/CCR4/CNOT1 complex, which promotes RNA degradation[11,12,30]. Whereas other m6A RBPs such as IGF2BP, FMRP, and

HuR stabilize m6A modified transcripts[11]. Aside from differential signaling of downstream pathways, there is evidence that m6A alters the secondary structure of mRNA, influencing its overall stability[34]; however, the effect seen depends on the secondary structure being examined[34].

Although m6A does not interfere with base pairing selectivity, it does destabilize the interaction between adenosine and uracil in RNA duplexes[34], with the amount of destabilization proportional to the number of m6A modified bases[34]. A different effect is seen when looking at the interactions between different RNA helices. Base stacking is an important component of RNA stability, and unpaired adenosines at ends of RNA helices stack strongly with adjacent secondary structures[34]. Yet, when these adenosines have a m6A modification, they display significantly stronger stacking kinetics, stabilizing the structure[34]. This was seen at both ends of the RNA helix, with a greater impact observed at the 3' end[13]. This phenomenon introduces the idea of differential stability between the 5' and 3' ends of mRNA transcripts due to chemical modifications.

Historically, methods to detect chemical modifications on RNA have relied on antibody capture. However, antibody capture has shown to have poor specificity and resolution[35,36], most likely due to the cross-reactivity of antisera used with closely related chemical species[30]. More recent approaches include analysis via mass spectrometry (MS)[28, 37-40] and sequencing of native RNA using NGS techniques[15,16,41].

Mass Spectrometry for RNA Modification Detection

Mass spectrometry has been noted as one of the most useful tools to study epitranscriptomics[42] and is one of the few direct ways to study chemical modifications on RNA molecules to date[43]. This technique measures the mass of a molecule by converting the sample into a gaseous phase and separating the ions by their mass-to-charge ratios[44]. When using MS for detection of RNA methylation, liquid chromatography tandem mass spectrometry (LC-MS/MS)

is typically used[28,39] which includes the additional separation step of liquid chromatography. Analysis of RNA using LC-MS/MS employs both ribonucleotide and oligonucleotide analysis. Briefly, RNA is hydrolyzed into ribonucleotides and subjected to LC-MS/MS analysis to determine modified from unmodified bases[39]. In a separate reaction, the RNA is also hydrolyzed into oligonucleotides of different lengths for analysis via LC-MS/MS to map the location of the identified modifications[39]. For both analyses, reverse phase liquid chromatography (RP-LC) is used to separate the molecules, which is based on the hydrophobicity of the sample[39]. Modifications on an RNA molecule may alter its hydrophobicity and therefore change their retention time during RP-LC[39]. Once the molecules are separated, they are converted to their gaseous phase and identified using MS as described above[39]. Molecules with a chemical modification will likely display a shift in mass compared to its unmodified counterpart, displaying the attachment of a chemical group[31]. The use of LC-MS/MS as a tool to detect modifications was the initial strategy of this study. The concept was to enrich the 5' and 3' ends of our transcript markers, hydrolyze them, and subject the hydrolysate to LC-MS/MS. However, through discussions with experts with this technology it became apparent that the sensitivity needed to detect perhaps attograms of modified nucleotide at one end of a transcript or the other was lacking in LC-MS/MS.

Nanopore Sequencing of Native RNA

A more recent advancement in studying RNA modifications is sequencing native RNA on next generation platforms such as Oxford Nanopore Technologies (ONT). This method provides the potential for single base pair resolution, as well as analysis of full-length transcripts[41]. While this is a very new approach, native RNA Nanopore sequencing has proved capable of detecting modifications present on RNA[15,16,41]. Native RNA sequencing using ONT works by selecting poly(A) tailed transcripts with sequencing adapters during library preparation[45]; thus, only transcripts with a poly(A) tail are sequenced using this method. During

library preparation, an “enzyme motor” is also ligated to the RNA, which functions to pull the sample through the nanopore to be sequenced. An electric current pattern, called a “squiggle”, is produced as the sample passes through the pore and reflects the subtle changes in current flow as each ribonucleotide passes through the pore[16]. This information can then be used for base calling and downstream analyses such as modification detection. If a modification is present on the sample, the way the molecule passes through the pore will be affected, causing a disruption in the electric current produced[41]. Computational methods are then used to identify modifications by comparing the electrical signal from modified and unmodified samples[15,41].

Detection of modifications on native RNA is a relatively new application of Nanopore sequencing, thus ONT does not provide any downstream analytical tools. However, multiple bioinformatic pipelines have been built and made available to the public through resources such as GitHub. One such pipeline is Tombo. An advantage to the Tombo pipeline, is its option to run a model free analysis which does not require controls or comparison samples[41]. In most bioinformatic models developed to detect base modifications, two controls (one known modified sample and one known unmodified sample) are also sequenced and included for statistical comparison. A direct comparison between the modified and unmodified controls to the experimental sample will then reveal the modified base(s). An option with the Tombo application is to perform the analysis without these controls. While this approach can suggest modified bases exist in a transcript, it does not identify the exact location or specific modification present[41]. An additional downside to using the model free method of Tombo is the increased risk for false positives[41]. Nonetheless, bioinformatically analyzing NGS results for chemical modification detection is an acceptable way to identify potential modifications in a transcript as preliminary results. These results can then be utilized for subsequent studies to confirm the exact type and location of any modifications present.

CHAPTER III

METHODOLOGY

Preparation of Bloodstains

Bloodstains were prepared from 5–7mL of freshly drawn blood collected by venipuncture into vacutainer collection tubes. All collections were performed with consent and approved by the Institutional Review Board of the Center for Health Sciences, Oklahoma State University. 100µL aliquots were spotted onto Human ID Bloodstain Cards (BFC180) (Whatman, GE Healthcare Life Sciences) and labeled clearly with the collection date. Freshly spotted stains were allowed to dry completely and then stored at -80°C until RNA extraction for a fresh time 0 sample.

Targeted mRNA Enrichment

RNA Extraction

The initial strategy for the project was to enrich the 5' and 3' ends of an RNA transcript in bloodstains through hybrid selection and then to chemically analyze hydrolysates from each using LC-MS/MS. We therefore experimented with methods to enrich the ends of the transcripts through hybrid selection. Two samples underwent targeted enrichment, a fresh bloodstain, and a two-year-old aged bloodstain. To increase sample input, four 100µL stains were processed per sample. Each blood stain was cut into a minimum of five strips and placed in a 2mL tube. 1mL of TRIzol (Life Technologies, Carlsbad, CA) was added to each sample. Samples were then vortexed for 30 minutes at room temperature and spun down to remove any liquid from the lid.

200µL of chloroform was then added, followed by 15 seconds of shaking by hand and a 3-minute room temperature incubation. Samples were then centrifuged at 12,000xg for 15 minutes at 4°C and the upper aqueous phase, which contains the RNA, was removed into a clean 1.5mL microcentrifuge tube. Following the TRIzol procedure, samples were further processed using the RNA Clean and Concentrator kit from Zymo Research (Orange, CA). First, an equal volume (450-500µL) of 100% ethanol was added to each sample, mixed by pipette, and loaded onto the Zymo column; two TRIzol treated samples were combined onto one Zymo column, for a total of two columns per sample. 750µL was transferred to the spin column and centrifuged for 30 seconds. This was repeated until the entire sample was transferred to the column. The next steps of the procedure were as follows with all centrifugations at 16,000xg: 400µL RNA prep buffer with 30 second centrifugation, 400µL RNA wash buffer with a 30 second centrifugation and 400µL RNA wash buffer with a 2-minute centrifugation. The columns were then transferred to a clean, labeled 1.5mL tube. Samples were eluted in 22µL of nuclease-free H_2O. All samples were incubated for 1 minute and then centrifuged for 1 minute. The two eluates for each sample were combined for subsequent processing.

Samples were treated with ezDNase (ThermoFisher, Waltham MA) to digest possible contaminating genomic DNA using the following setup: 41µL RNA, 5µL of 10x ezDNase buffer and 4µL of ezDNase enzyme. The mixture vortexed, quickly spun down, incubated at 37°C for 2:30 minutes, and immediately placed on ice for 5 minutes. The samples were then processed with the OneStep PCR Inhibitor Removal kit from Zymo Research (Orange, CA). Each inhibitor column was prepped by adding 600µL of prep-solution and centrifuged for 3 minutes at 8,000xg. The column was then placed in a fresh labeled 1.5mL tube and 50µL of sample was added directly to the prepped column. Samples were centrifuged for 3 minutes at 16,000xg.

All samples were quantified using a Qubit fluorometer with the RNA quantification kit (ThermoFisher, Waltham, MA). Following the kits instructions, 1µL of sample was added to 199µL of buffer and 1µL of H.S. RNA dye. Samples were vortexed and spun down prior to being

measured. Following quantification, each sample was aliquoted into three separate tubes: extraction control, 5' hybridization, and 3' hybridization. Final sample concentrations and hybridization setup are noted in Table 1. The control samples were immediately placed at -80°C until further analysis the following day, while the hybridization samples were dried using a speed vacuum until only 1-2µL remained, about 40 minutes. Samples were immediately stored at -80°C until further analysis the following day.

Sample	Final (ng)	Extraction Control (ng)	5' Hyb. (ng)	3' Hyb. (ng)
Fresh	2,890	300	1,300	1,300
2 Year	2,210	300	955	955

Table 1. Hybrid Capture Input Concentrations

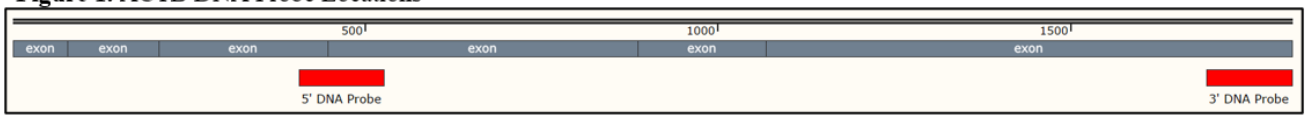
DNA Probe Hybridization

The 5' and 3' ends of beta-actin (*ACTB*) mRNA were enriched in separate reactions for each sample using the xGen Hybridization and Wash Kit from Integrated DNA Technologies (IDT) (Newark, New Jersey), optimized for our purpose based on previously described methods from Tan et al (2019)[14]. Two 120 nucleotide biotinylated DNA probes were designed to target each end of *ACTB* (accession NM_001101). Probes were designed to cover previously designed primer pairs used to quantify *ACTB* degradation during aging of blood stains (unpublished observations). Biotinylated DNA probe designs are shown in Figure 1 and Table 2.

Probe	Sequence	Start	Stop
5'	TACGGCCAGAGGCGTACAGGGATAGCACAGCCTGGATAGCAACGTAC ATGGCTGGGGTGTTGAAGGTCTCAAACATGATCTGGGTCATCTTCTCG CGGTTGGCCTTGGGGTTCAGGGGGG	406	526
3'	TCATTTTTAAGGTGTGCACTTTTATTCAACTGGTCTCAAGTCAGTGTAC AGGTAAGCCCTGGCTGCCTCCACCCACTCCCAGGGAGACCAAAAGCCT TCATACATCTCAAGTTGGGGGAC	1692	1812

Table 2. ACTB DNA Probe Sequences

Figure 1. ACTB DNA Probe Locations



For each hybridization reaction, the following components were added directly to the dried RNA: 8.5µL xGen 2x Hybridization Buffer, 2.7µL xGen Hybridization Buffer Enhancer, 4.1µL H_2O and 1.7µL 10µM DNA probe for a final probe concentration of 1µM. To evaluate the amount of probe left after final purification, the 5' and 3' probes alone (without any RNA sample input) also underwent the hybrid capture process. To ensure similar hybridization conditions, the same hybridization setup was used. Mixtures were gently vortexed, spun down, and incubated at room temperature for 5 minutes. Samples were then placed in the thermocycler set to the following program: 95°C for 30 seconds and 65°C for 2 hours. During this time a heat block was prewarmed to 65°C to ensure proper temperature for the next step. RNA:DNA probe hybrids (bearing a biotin at the 5' end of the probe) were "captured" from solution by binding to a streptavidin bead.

Bead Capture and Washes

xGen Hybridization Buffers and magnetic beads were freshly prepared during the two-hour hybridization. All xGen Buffers were diluted to 1x working stocks shown in Table 3.

	Volumes Per Reaction		
Reagent	Water (µL)	Buffer (µL)	Total (µL)
xGen 2x Bead Wash Buffer	160	160	320
xGen 10X Wash Buffer 1	252	28	280
xGen 10X Wash Buffer 2	144	16	160
xGen 10X Wash Buffer 3	144	16	160
xGen 10X Stringent Wash Buffer	288	32	320

Table 3. Dilutions for Hybrid Capture Buffers

Following the dilutions, 110µL per reaction of Wash Buffer 1 and 160µL per reaction of Stringent Wash Buffer were aliquoted into separate tubes for the heated wash steps. Bead resuspension mix was prepared in a low-bind tube with the following components per reaction:

8.5µL xGen Hybridization Buffer, 2.7µL Hybridization Buffer Enhancer and 5.8µL nuclease-free H_2O.

Streptavidin Beads (IDT, Newark, New Jersey) were equilibrated to room temperature for 30 minutes prior to washing. 50µL of streptavidin beads per reaction were aliquoted into a 1.7mL low-bind tube. Beads were washed with 100µL per reaction of bead wash buffer. Briefly, the mixture was gently pipetted 10 times, with special attention to not introduce any bubbles. Tubes were then placed on a magnet and the supernatant was allowed to separate from the beads, about 1 minute. The supernatant was then removed and discarded. The wash step was repeated twice more for a total of three washes. Following the washes, beads were resuspended in 17µL per reaction of previously prepared bead resuspension mix. Beads were then split into 17µL aliquots in 1.7mL low-bind tubes for each hybridization reaction.

Immediately following the two-hour probe hybridization, the 17µL of hybridization product in a 0.2mL PCR tube was transferred to the 17µL of beads in the 1.7mL low-bind tube, mixed by pipette, and placed on a heating block preheated to 65°C for 45 minutes. To ensure the beads stayed suspended in the solution, tubes were gently vortexed every 15 minutes throughout the bead capture incubation. The aliquot with 110µL of Bead Wash 1 and both Stringent Wash Buffer aliquots were placed on the heating block at the beginning of the incubation to ensure they were equilibrated to the correct temperature for the heated washes.

The heated wash step was started immediately following the bead capture. At each step, care was taken not to introduce bubbles when pipetting. The heated wash steps were as follows: 100µL of heated Wash Buffer 1 was added to each sample and pipetted 10 times. Samples were then placed on the magnet until the supernatant completely separated from the beads (~1 minute). The supernatant was removed and discarded. Beads were then resuspended in 150µL of heated Stringent Wash Buffer, pipetted 10 times and incubated at 65°C for 5 minutes. Following the

incubation, samples were placed back on the magnet until the supernatant was completely separated from beads (~1 minute). The supernatant was removed, and the beads were resuspended in 150µL of heated Stringent Wash Buffer, pipetted 10 times and incubated at 65°C for 5 minutes. Following the incubation, samples were placed on the magnet until the supernatant was completely separated from the beads. The supernatant was removed and discarded.

The room temperature washes immediately proceeded the heated wash steps, as follows: 150µL of room temperature Wash Buffer 1 was added to each sample and pipetted 10 times. Samples were then incubated at room temperature for 2 minutes while being vortexed periodically to ensure the beads remained suspended. Following the incubation, samples were placed on the magnet until the supernatant completely separated from the beads (~1 minute) and the supernatant was removed and discarded. This room temperature wash procedure was repeated with Wash Buffer 2 and Wash Buffer 3. Following the final wash, the beads were resuspended in 20µL of nuclease-free H_2O.

Each sample was then split into two 10µL aliquots. One aliquot underwent RNase A and DNase I digestion to increase specificity and only leave the 120bp region that is protected by the DNA probe (specific product). The second sample was only digested with DNase I to maintain additional *ACTB* regions extending from the probe bound region of the transcript and therefore also enriched (full-length product).

Post-Capture Processing

One aliquot from each sample was digested with RNase A to remove single stranded RNA not hybridized to the DNA probe. First, the samples were placed back on the magnet until the supernatant became clear. The supernatant was removed and discarded, and samples were taken off the magnet and resuspended in 9µL of 0.2M TNE. The RNase A digestion setup was as follows: 9µL of sample in 0.2M TNE, 1µL of 4M NaCL and 1µL of RNase A (ThermoFisher,

Waltham, MA). Samples were vortexed, spun down, and incubated at 37°C for 30 minutes. Samples were then digested with DNase I.

All samples were digested with TURBO DNase I (ThermoFisher, Waltham, MA) to remove the single-stranded DNA probe from the targeted RNA. The reaction setup was as follows: 2µL of 10x buffer, 4µL of TURBO DNase I (ThermoFisher, Waltham, MA) and 2µL of nuclease-free H_2O. Samples were vortexed, spun down, and incubated at 37°C for 30 minutes. Following the incubation, samples were placed back on the magnet. Once the supernatant was completely clear, it was removed and placed in a fresh tube.

The enriched RNA was purified using RNAClean XP Beads (New England Biolabs, Ipswich, MA). The beads were vortexed for a minimum of 15 seconds to ensure they were completely resuspended, and 36µL (1.8x volume) was added to each RNA sample and mixed by pipette at least 10 times. The solutions were then incubated on ice for 15 minutes. Following the incubation, samples were placed on a magnet until the supernatant became clear. The supernatant was then carefully removed, without disturbing the beads, and discarded. While still on the magnetic rack, 200µL of freshly prepared 80% ethanol was added and incubated for 30 seconds. The supernatant was then removed and discarded. This step was repeated twice more for a total of three washes. Any left-over ethanol was removed, and samples were air dried on the magnet with the lid open for up to 5 minutes. Attention was paid to not over dry the beads. Samples were then removed from the magnet and the beads were eluted in 15µL of nuclease-free H_2O, mixed thoroughly by pipette, briefly spun down, and incubated at room temperature for 2 minutes. The tubes were then placed back on the magnetic rack until the supernatant became completely clear. The supernatant was then removed and placed in a fresh 0.2mL tube.

Quantification

To evaluate the success of hybrid selection, the final selected mRNA was first reverse transcribed into complementary DNA (cDNA) using a reverse transcription (RT) reaction with random and oligo dT primers. Prior to RT, 3µL of the 5' and 3' hybridization products from both the RNase A digested and longer selection targets from fresh bloodstains were aliquoted out for direct qPCR analysis. Samples treated with RNase A would represent the 5' or 3' ends of the transcript that were protected from digestion by hybridization to the DNA probe whereas non-digested samples would represent that portion of the transcript hybridized to the DNA probe as well as any non-degraded transcript fragments (including full length transcript) present in the initial RNA extract. No-RT samples would serve as a control for possible genomic DNA contamination remaining in the final hybridization products. In addition, the probes used for hybrid-selection were also directly analyzed via qPCR and were not reverse transcribed. Enriched RNA was reverse transcribed using the SuperScript IV VILO Master Mix kit (ThermoFisher, Waltham, MA) by mixing 10µL of sample with 2µL SuperScript IV VILO Master Mix. Samples were briefly vortexed and spun down before being incubated according to manufacturer's instructions: 25°C for 10 minutes, 50°C for 10 minutes and 85°C for 5 minutes.

All experimental and control samples were analyzed via qPCR using SYBR green intercalating dye. Reaction setups were as follows: 5µL PowerUp SYBR Green Master Mix (ThermoFisher, Waltham, MA), 1µL of 10x primer pair at 3 or 5µM (primer dependent), 3µL H_2O and 1µL of 1:4 diluted sample. All samples were amplified using six *ACTB* "on-target" primer pairs along the transcript and two "off-target" transcripts, *HBB* (hemoglobin beta chain) and *B2M* (beta-2-microglobulin). Samples were placed in the thermocycler and incubated at 50°C for 2 minutes and 95°C for 2 minutes, followed by 40 cycles of 95°C for 15s and 60°C for 1 minute. Cq values were determined using a threshold of 0.2.

Nanopore Direct RNA Sequencing

RNA Extraction

A total of three samples were processed for native RNA sequencing using Nanopore Oxford Technologies (ONT); two fresh blood stain samples and one 11-week-old aged bloodstain sample. To ensure sufficient sample input for library prep, eight 100µL bloodstains were processed for each sample. Sample extraction followed the same procedure used for hybrid capture samples until post-processing procedures, with a few modifications. First, to ensure sufficient sample concentration and quality, the eight aged bloodstains were all processed on individual Zymo spin columns instead of combining two stains onto one. Additionally, final elution volumes were as follows, fresh sample one: 16µL, fresh sample two: 22µL and aged sample: 15µL.

Samples were treated with ezDNase (ThermoFisher, Waltham MA) using the following setup: fresh sample 1: 53µL RNA, 6.5µL of 10x ezDNase buffer and 6µL of ezDNase enzyme, fresh sample 2: 80µL RNA, 9.7µL ezDNase buffer and 7µL ezDNase enzyme, and aged sample: 98µL sample, 11.5µL ezDNase buffer and 6µL ezDNase enzyme. The mixture was mixed by vortex, quickly spun down, incubated at 37°C for 2:30 minutes, and immediately placed on ice for 5 minutes.

Sequencing samples were then processed using the Monarch RNA Clean-Up kit from NEB (New England Biolabs, Ipswich, MA). Briefly, 1.5 volume of Binding Buffer was added to the sample. An equal volume of 100% ethanol was added and mixed by pipette before the entire sample was loaded onto the spin column and centrifuged for 1 minute at 16,000xg, the flow through was discarded. 500µL of RNA Cleanup Wash Buffer was added and the sample was centrifuged for 1 minute at 16,000xg, this step was repeated for a total of two washes. The

column was then transferred to a clean 1.5mL tube and the sample was eluted in 16µL using nuclease-free H_2O.

Samples for Nanopore sequencing were quantified using a Qubit fluorometer with the RNA quantitation kit (ThermoFisher, Waltham, MA) as described above, and a Nanodrop Spectrometer (ThermoFisher, Waltham Massachusetts) set for RNA reading. For Nanodrop quantification, the machine was blanked with the elution buffer before the samples were read. Final concentrations and purities are shown in Table 4.

Sample	Qubit (ng/µL)	Nanodrop (ng/µL)	260/280	260/230	Library Prep. Input
1- Fresh	200	300	1.98	2.06	1.8-2.7ug
2- Fresh	273	372	2.01	2.32	2.4-3.3ug
Aged	256	281	2.04	2.13	2.3-2.5ug

Table 4. Final Concentrations and Purities for Nanopore Sequencing Samples

Library Preparation

RNA libraries were prepped using the Direct RNA Sequencing Kit (SQK-002) from Oxford Nanopore Technologies (ONT) (Oxford, United Kingdom) and the NEBNext Ultra II Ligation Module (New England Biolabs, Ipswich Massachusetts) according to the manufacturers' instructions, with one modification of using SuperScript IV Reverse Transcriptase (ThermoFisher, Waltham Massachusetts) instead of SuperScript III. Although direct RNA sequencing was utilized, a RT reaction was performed for stabilization of the RNA as it goes through the sequencing pore. In subsequent library preparation steps, the motor protein is ligated to the native RNA strand, and thus that is what is sequenced. Briefly, in a 0.2mL PCR tube the following reagents were mixed in the corresponding order: 3.0µL NEBNext Quick Ligation Reaction Buffer, 9µL total RNA, 0.5µL RNA CS (RCS), 1µL RT Adapter (RTA) and 1.5µL T4 DNA Ligase. Reagents were mixed by pipetting, spun down, and incubated for 10 minutes at room temperature. During the incubation, the RT master mix was created by mixing the following

reagents: 9.0μL nuclease-free H_2O, 2.0μL 10mM dNTPs, 8.0μL 5x SSIV buffer, and 4.0μL 0.1M DTT for a total volume of 23μL. The master mix was added to the 0.2mL PCR tubed, mixed by pipette. 2.0μL of SuperScript IV reverse transcriptase enzyme was then added and mixed by pipette. The tube was placed in a thermal cycler and incubated at 23°C for 10 minutes, 50°C for 10 minutes, 80°C for 10 minutes, and allowed to cool to 4°C before being removed.

Next, the sample was transferred to a clean 1.5mL Eppendorf DNA LoBind microfuge tube. Agencourt RNAClean XP beads (New England Biolabs, Ipswich, MA) were allowed to equilibrate to room temperature for at least 30 minutes and then vortexed until completely resuspended. 72μL of beads were added to the sample and mixed by pipette. The mixture was incubated at five minutes at room temperature while the tube was slowly rocked back and forth, imitating the movement of a HulaMixer. Following the incubation, 200μL of fresh 70% ethanol was prepared in nuclease-free H_2O. The sample was briefly centrifuged, and the beads were immobilized on a magnet. Keeping the tube on the magnet, the supernatant was removed and discarded, careful not to disturb the beads, and 150μL of fresh 70% ethanol was added. With the tube still on the magnet, the tube was rotated 180 degrees and incubated until the beads migrated towards the magnet. Once all of the beads reaggregated, the tube was rotated another 180 degrees to its original position until the beads moved to that position and were immobilized. With the tube still on the magnet, the 70% ethanol was removed and discarded. The tube was then removed from the magnet, briefly centrifuged, and placed back on the magnet in order to pipette off any remaining 70% ethanol. The tube was once again removed from the magnet, resuspended in 20μL of nuclease-free H_2O and incubated for 5 minutes at room temperature. Following incubation, the beads were immobilized on the magnet until the eluate was clear. The 20μL of eluate was removed and placed in a clean 1.5μL Eppendorf DNA LoBind tube.

The following reagents were then added to the sample tube in the corresponding order: 8μL NEBNext Quick Ligation Buffer, 6μL RNA Adapter (RMX), 3μL nuclease-free H_2O, 3μL

T4 DNA Ligase (New England Biolabs, Ipswich Massachusetts) for a total reaction volume of 40μL. The components were mixed by pipette and incubated at room temperature for 10 minutes. Agencourt RNAClean XP beads were once again resuspended by vortex, and 16μL were added to the ligation reaction and mixed by pipetting. The solution was then incubated for five minutes while being rocked in a manner that imitates a HulaMixer. The sample was then briefly centrifuged and then immobilized on a magnet. With the tube still on the magnet, the supernatant was removed and discarded. 150μL of the Wash Buffer (WSB) was added to the beads. With the lid closed, the beads were resuspended by flicking. Once fully suspended, the tube was placed back on the magnet until the beads were fully immobilized. The supernatant was removed, and the wash step was repeated for a total of two washes. After the second wash was complete, the tube was removed from the magnet and 21μL of Elution Buffer (ELB) was added. The sample was resuspended by gently flicking the tube. Once fully resuspended, the mixture was incubated at room temperature for 10 minutes. Following the incubation, the beads were immobilized on the magnet until the eluate was clear. The supernatant was then removed and put into a fresh 1.5mL Eppendorf DNA LoBind tube. The reverse-transcribed, adapted RNA was quantified using the Qubit fluorometer DNA HS assay following the manufacturer's instructions. Briefly, 1μL sample was mixed with 199μL buffer and 1μL of H.S. DNA dye. Samples were vortexed and centrifuged prior to being measured.

Sequencing

The first step in Nanopore sequencing is to check the flow cell where sequencing occurs to ensure an adequate number of sequencing pores are available. Briefly, a new SpotON Flow Cell MK 1 R9 was removed from its packaging and placed under the clip in the MinION Mk1C sequencer. On the MinKNOW software, the check flow cell procedure was chosen. Once the flow cell check was complete, priming and loading of the prepared RNA library onto the flow cell was done using the Flow Cell Priming Kit (EXP-FLP002) (Oxford, United Kingdom) according to the

manufacturer's instructions. The RNA Running Buffer (RRB), Flush Tether (FLT), and Flush Buffer (FB) were thawed at room temperature. Once completely thawed, the RRB, FB, and FLT were thoroughly vortexed and briefly centrifuged.

Prior to sample loading, the flow cell (R9.4.1) was primed. First, the priming port was carefully opened. Using a P1000 pipette set to 200µL, 20-30µL of fluid was removed by inserting the pipette tip into the priming port and turning the wheel until a small volume of buffer entered the tip. Flow cell priming mix was prepared by adding 30µL of thawed and mixed FLT to the full tube of thawed and mixed FB. The mixture was thoroughly vortexed. 800µL of the priming mix was then loaded into the flow cell using the priming port, being careful not to introduce air bubbles, and incubated for 5 minutes. During the incubation, 20µL of the prepared RNA library was mixed with 17.5µL of nuclease-free H_2O. In a fresh tube, the 37.5µL of RNA library was mixed with 37.5µL of thawed and vortexed RRB. When the 5-minute incubation was done, the flow cell priming was completed. The SpotON sample port cover was gently lifted. 200µL of the priming mix was loaded into the flow cell using the priming port, being careful not to introduce air bubbles. The RNA library was mixed by pipetting directly prior to loading onto the flow cell. The 75µL of prepared RNA library was loaded onto the SpotON sample port drop by drop, ensuring each one entered the port before adding the next. The SpotON sample port cover, priming port, and the MinION Mk1C lid were all closed, and sequencing was initiated.

Data Analysis

The sequencing run was set up using the MinKNOW software on the MinION Mk1C device with default parameters and live basecalling using Guppy. Sequencing output included FASTQ files, containing basecalling information, and Fast5 files, which contain the raw electrical signal data used for modification detection. Samples were sequenced for 19.5 hours. After sequencing was complete, FASTQ files were generated and then uploaded to Nanopore's

EPI2ME platform and analyzed using the Fastq Human Exome workflow, which maps generated sequencing reads against the human reference exome (GRCH38) and identifies all transcripts sequenced.

For modification detection, the bioinformatic package Tombo was used (available at https://nanoporetech.github.io/tombo/index.html) (v1.5.1). Briefly, the following steps were taken. First, Fast5 files were converted from multiple to single file format using multi_to_single_fast5 (v4.0.2) and then converted to gzip file format using compress_fast5 (v4.0.2). These two steps were important for preprocessing of Nanopore data as original output files are in multi-VBZ compressed format and downstream analyses require single-gzipped files. Next, reads were preprocessed using tombo preprocess annotate_raw_with_fastqs. This command performed basecalling on the Fast5 files using the FASTQ files generated from the sequencing run. Using tombo resquiggle, the electrical signal data stored in the Fast5 files were "resquiggled" to the latest human RNA reference GRCh38_latest_rna.fna from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/genome/guide/human/). This step evaluated the electrical signal from each base in the sequenced RNA fragment and collected information about current interruptions that may signify a modification. Next, potential modifications were detected using Tombo's model free option, *tombo detect_modifications de_novo*. This application was required as we did not chemically synthesize and sequence a control, non-methylated sample to compare electrical signal data with. We believe Tombo has a database of profiles from non-methylated validation samples that our sequences are matched against. The best data is obtained if one synthesizes a non-methylated copy of the transcript under study. Finally, tombo plot most_significant was used to visually display the results in two formats: raw electrical signal plots (squigglegrams) and box-and-whisker diagrams.

CHAPTER IV

FINDINGS

mRNA Targeted Enrichment

Results show that targeted enrichment of specific mRNA targets is possible using biotinylated DNA probes. *ACTB* enrichment was successful for the 3' (Figure 2A) and 5' (Figure 2B) ends from fresh bloodstains. Amplification was achieved along the entire *ACTB* transcript from full-length products for both the 3' (Figure 2A) and 5' (Figure2B) hybridization reactions, although less abundant compared to the control (Figure 2A and B). Off-target amplification of *HBB* was seen in both full-length products, whereas *B2M* was only amplified from the 3' hybridization (Figure 2A and B). However, off-target abundance decreased between 250,000 and 17,000,000-fold for *B2M* and *HBB* respectively. Looking at the off-target transcripts for the specific versus full-length targets, the specific approach did help to decrease off-target amplification.

Data from the specific fragment enrichment of fresh bloodstains shows non-specific amplification of *ACTB* regions not protected by the DNA probe (Figure 2A and B), although less abundant compared to the full-length products. For the 3' specific product, two *ACTB* targets not protected by the DNA probe were amplified, however both were low in abundance with Cq values ranging from 30-35 (Figure 2A). *HBB* was also amplified from the 3' specific product, but its abundance decreased by almost 250,000,000-fold. Additionally, two non-protected *ACTB* regions were also amplified from the 5' specific-fragment product (Figure 2B). One of these

regions, ACTB-1486, was also found in the 3' specific fragment enrichment, and the second, ACTB-145, was directly adjacent to the 5' probe location (Figure 2B). No off-target transcripts were amplified from the 5' specific product.

Figure 2. ACTB Enrichment in Fresh Bloodstains

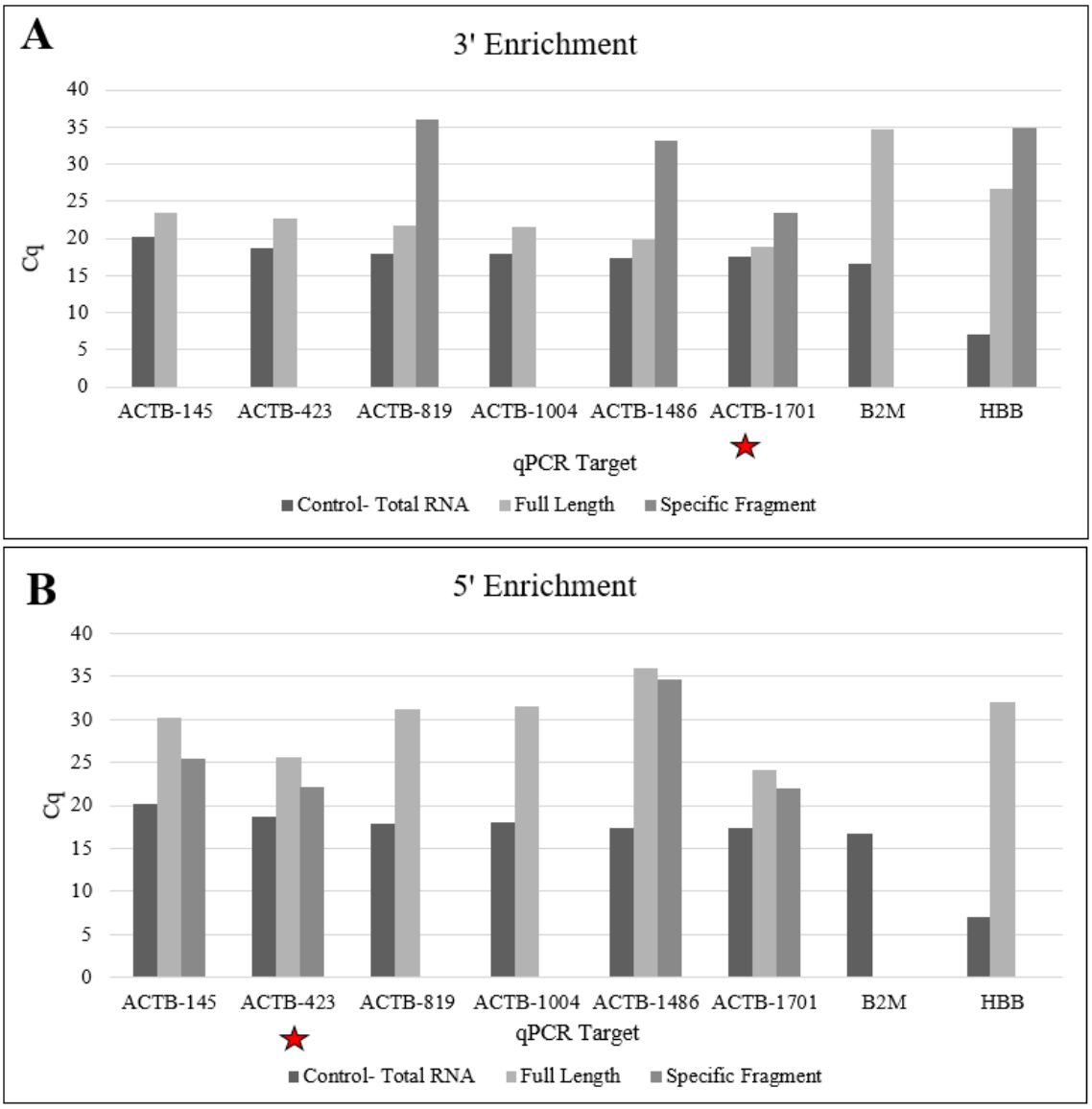


Figure 2A and B shows the comparison of DNase I only treated samples (Full length) and RNase A and DNase I treated samples (Specific fragment) for both 5' and 3' enrichment products. The X axis shows the transcript regions amplified during qPCR. ACTB numbers represent the location on the transcript with 145 and 1701 being the most 5' and 3' targets respectively. Stars represent DNA probe locations.

Targeted enrichment of *ACTB* in a two-year-old bloodstain was not successful with the 3' probe. Results show that no regions additional to the portion hybridized to the probe were

amplified from the full-length product (Figure 3A). However, no off-target transcripts were found in the sample either (Figure 3A). The 5' hybridization of aged-bloodstains was successful. The 5' full-length product shows amplification of regions along the *ACTB* transcript not hybridized to the DNA probe (Figure 3B), and *HBB* abundance was reduced by about 20,000-fold. For the specific product, three regions not protected by the probe were amplified, all of which were also amplified in the 5' specific product from fresh blood (Figure 2A and 3A).

Figure 2. ACTB Enrichment in Fresh Bloodstains

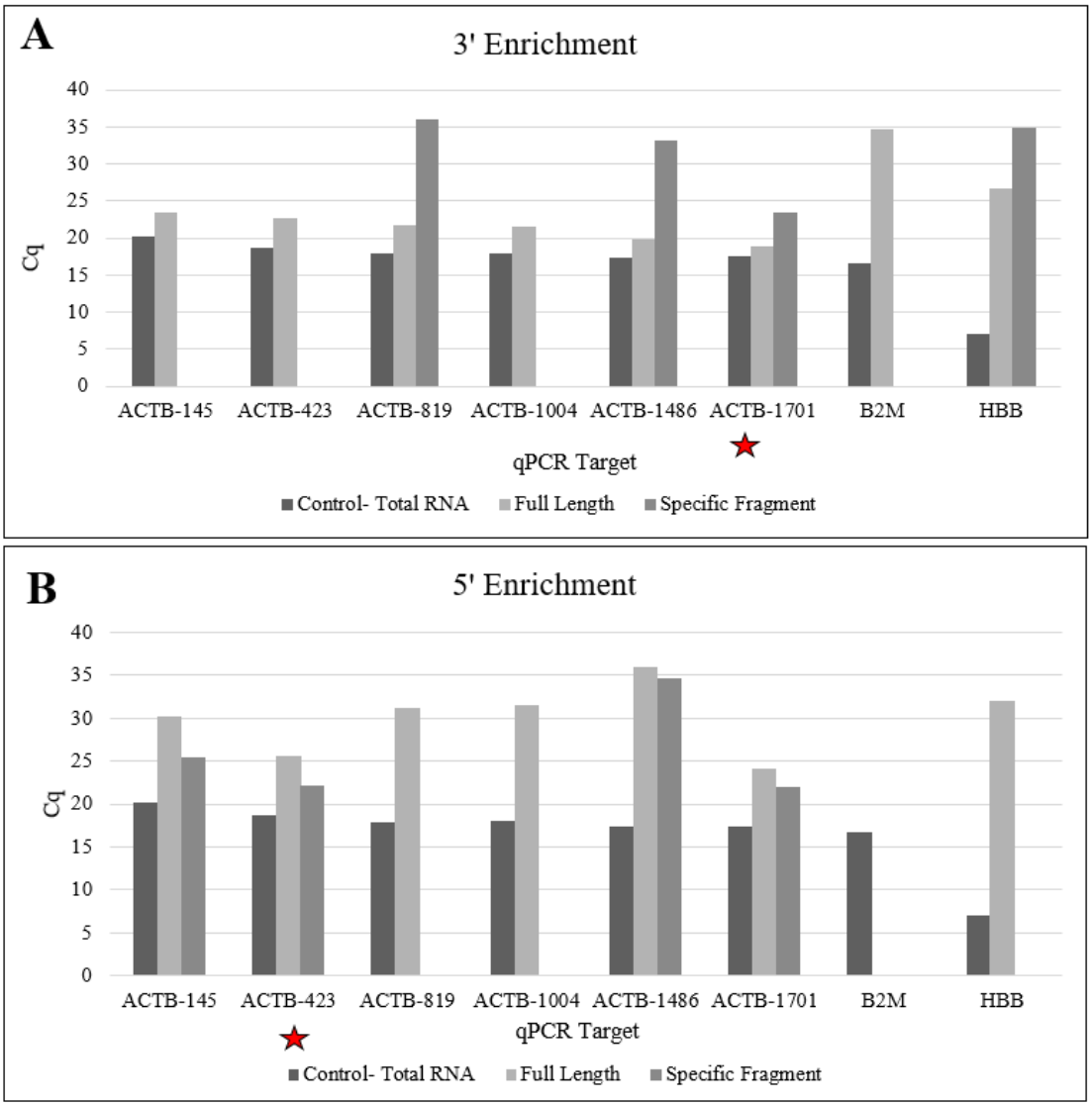


Figure 2A and B shows the comparison of DNase I only treated samples (Full length) and RNase A and DNase I treated samples (Specific fragment) for both 5' and 3' enrichment products. The X axis shows the transcript regions amplified during qPCR. ACTB numbers represent the location on the transcript with 145 and 1701 being the most 5' and 3' targets respectively. Stars represent DNA probe locations.

Results support the conclusion that we were able to successfully enrich *ACTB*, however potential DNA contamination in the final products make it difficult to assess the final amount of enriched target. This is displayed with the amplification data of the probe-only and no-RT products from the final 5' and 3' hybridizations. Figure 4 shows that the Cq values are relatively similar across all products from the 5' (Figure 4A) and 3' (Figure 4B) probe regions. This suggests that leftover, undigested probe may be present in the final hybridization sample.

Figure 4. Experimental and Control Sample Comparison

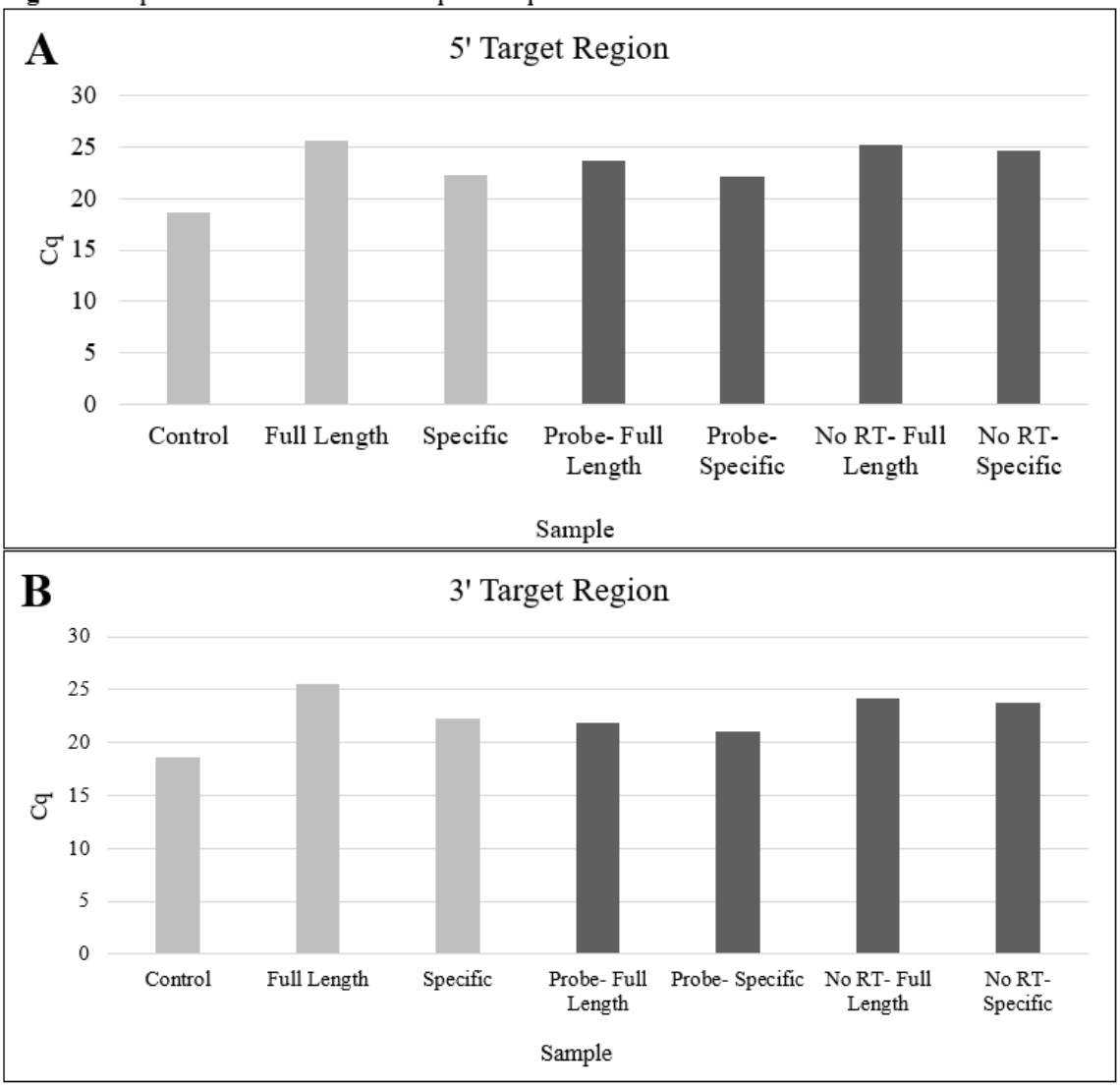


Figure 4A and B shows the comparison of experimental and control hybridization samples. Light grey bars represent products that were reverses transcribed and darker bars are those that were directly analyzed via qPCR and thus amplification shows DNA contamination.

Lastly, quantification using the Qubit H.S. RNA kit did not detect any RNA in the selected transcript suggesting the actual concentration of recovered RNA was very low (data not shown). Therefore, sample concentrations were too low for analysis via LC-MS/MS. For this reason, efforts were shifted to native RNA sequencing for modification detection.

Nanopore Native RNA Sequencing

Sequencing Summaries

Native RNA from three samples was successfully sequenced on the ONT platform, two fresh bloodstains and one 11-week-old aged bloodstain. All libraries were prepped using total RNA as input. The final concentration of library prepped RNA, total amount of data produced, and number of active pores available on the flow cell used in each sequencing run are shown in Table 5.

Sample	Library Input (µg)	Sequencing Input	Reads Generated	Bases Called (Mb)	Failed Base Calls (Mb)	Data Produced (GB)	Active Pores
Fresh- 1	1.8-2.7	150ng*	472.33 k	271.87	9.9	16.57	1478
Fresh-2	2.4-3.3	180ng	218.06 k	127.44	4.94	8.26	1148
Aged	2.3-2.5	64ng	331.69 k	166.1	9.08	11.3	1202

Table 5. Nanopore Sequencing Concentrations and Data Outputs.
*Sample was measured using RNA instead of DNA Qubit kit, concentration may reflect RNA that is not ligated to an adapter.

Looking at the percentage of active pores available throughout each sequencing run, there is a steeper decline for the second fresh sample compared to the other two sequencing runs (Figure 5). As the only difference between the two fresh sequencing runs is the amount of sample input, it is suggested that higher input of RNA in a sequencing run has a negative effect on flow cell health.

Figure 5. Percentage of Active Pores Throughout Sequencing Run

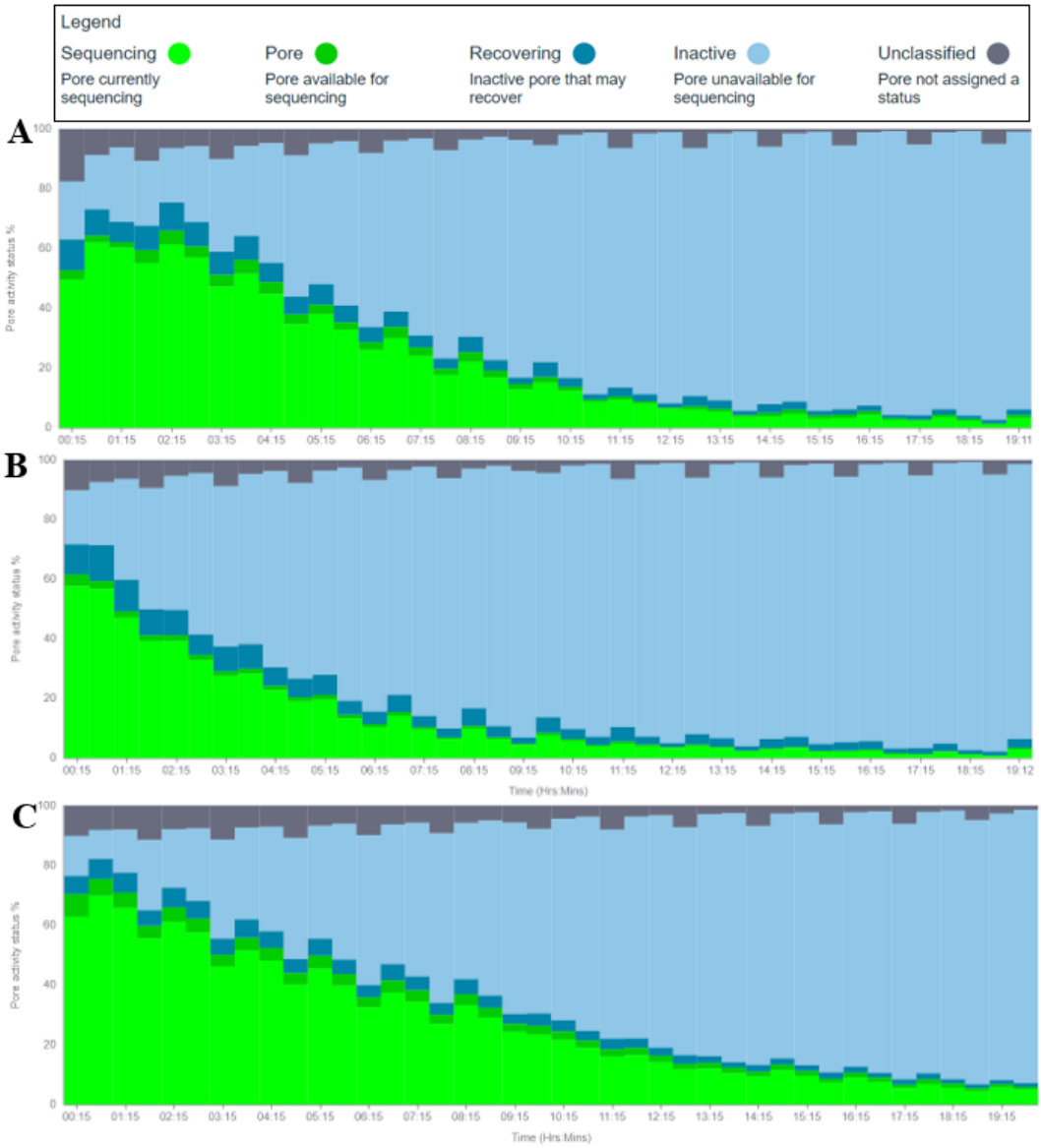


Figure 5A-C shows the percentage of active pores on the flowcell throughout each sequencing run; A- First fresh, B- Second fresh, C-Aged.

The length of transcripts sequenced from native RNA sequencing ONT is shown in Figure 6, displayed in kilobase pairs (kbp). Full-length transcripts were not produced for any of the three samples (Figure 6A-C). Additionally, shorter reads were obtained from the aged bloodstain (Figure 6C) compared to the fresh bloodstains (Figure 6A and B). It appears that sequenced RNA fragments recovered from the aged sample were about one-half the length of

RNA fragments from freshly made stains. This is consistent with our observations concerning RNA degradation in aged stains evaluated using the 5'-3' qPCR assay.

Figure 6. Length of Transcripts Sequenced.

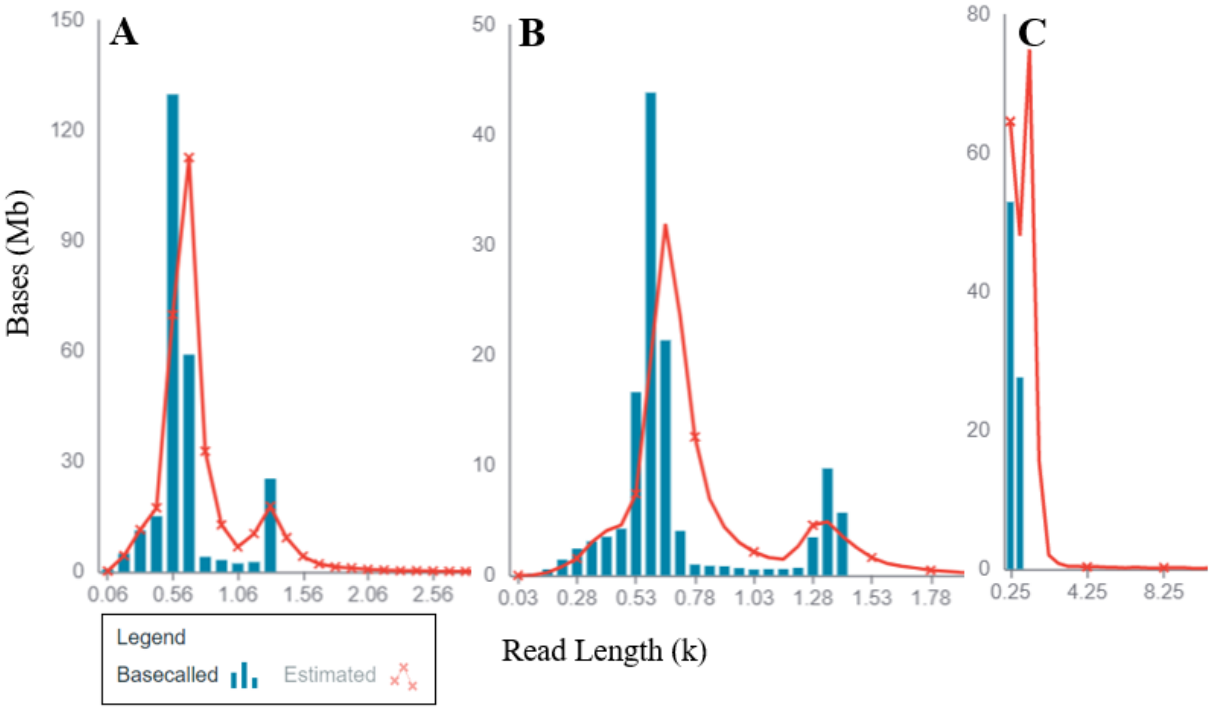


Figure 6 shows the length of transcripts that were sequenced for the first fresh bloodstain (A), second fresh bloodstain (B) and aged bloodstain (C). The peaks at 1.3 k represent the RNA sequencing control.

Sequencing Results

The alignment summary of each sequencing run is shown in Table 6. The data shows average alignment accuracy ranging from 84-86%, which is standard for Nanopore RNA sequencing using an R9 flowcell[46]. Additionally, the number of reads analyzed and successfully aligned from the second sequencing run with a freshly prepared blood stain is about half of that observed for the first fresh sample sequencing run (Table 6). This remains consistent with the overall data produced (Table 5). Additionally, the top transcripts sequenced (in terms of

abundance) remains relatively consistent across each sequencing run for fresh and aged bloodstains (Table 7).

Sample	Reads Analyzed	Reads Successfully Aligned	Avg. Alignment Accuracy	No. of Transcripts Found
Fresh-1	444,326	387,565	84.10%	4,649
Fresh-2	205,174	178,966	82.90%	3,292
Aged	293,513	176,105	86.90%	4,088

Table 6. Reference Alignment Summaries

	Fresh-1	Fresh-2	Aged	Previous RNA Seq
1	HBB	HBB	HBB	HBB
2	HBA2	HBA2	HBA2	MTRN2L8
3	HBA1	HBA1	HBA1	HBA2
4	DHFR	DHFR	DHFR	MTRNR2L1
5	MTRNR2L12	MTRNR2L12	MTRNR2L12	MTRNR2L6
6	MTRNR2L8	MTRNR2L8	MTRNR2L8	MTRNR2L9
7	S100A9	S100A9	FTL	MTRNR2L10
8	S100A8	S100A8	S100A9	HBA1
9	B2M	MT-CO3	UBB	B2M
10	MT-CO3	MTRNR2L1	MTRNR2L1	S100A8

Table 7. Transcript Abundance in Descending Order.
Top 10 transcripts are shown for each Nanopore sequencing run in addition to a previously done RNA-sequencing run. Yellow highlighted columns represent exact matches between transcript and abundance. Grey highlights show transcripts that are found in other samples but in a different order.

The only discrepancies between the two fresh bloodstains are the 9th and 10th most abundant transcripts found, whereas inconsistencies began at the 7th transcript for the aged sample (Table 7). Table 7 also includes the most abundant transcripts found from an RNA-sequencing experiment performed prior in our laboratory using conventional NGS techniques sequencing cDNA reverse transcribed from mRNA extracted from blood stains[9]. Seven of the top transcripts found in the previous RNA sequencing run were also included in the Nanopore top transcripts, however only two of the seven matched in abundance order (Table 7).

Modification Analysis

As we did not have control RNA molecules to sequence for comparison to the data produced from the experimental samples, suggested modification data is preliminary. Additionally, the exact location and specific modification remain unknown. Potential modifications were identified on 12 of the transcripts sequenced (Table 8). Of these, three transcripts were identified in all three sequencing runs (*HBB, S100A9* and *RPS12*), two were only identified in fresh stains (*S100A8* and *TMSB4X*), and one was found in the first fresh and aged sample (*B2M*) (Table 8). More specifically, of the six transcripts that were identified in multiple samples, identical suggested locations for modified bases were found in five of them (Table 8). Furthermore, *S100A9* and *RPS12* had modifications detected in all three samples, with an additional modification identified in only the aged sample (Table 8). Modified bases were suggested from our data in two identical regions in *HBB* in the second fresh and aged samples (Table 8). The *HBB* sequencing data is also significant in that, it is the only transcript for which sequencing (and potential modification data) of the 5' end of the transcript is available.

Transcript	Fresh-1	Fresh-2	Aged RNA
HBB	57	57, 117, 443, 593	57, 117
S100A9	432	432	432, 464
S100A8	319	262, 319	
B2M	640		876
TMSB4X	298, 329	329	
RPS12	278	278	278, 321
RPL41	265		
RPS29	256		
UBB	614		
HBA2		494	557
HLAB			1515
FTL			606

Table 8. Transcripts with Potential Modifications
Numbers represent the estimated nucleotide location identified with a potential modification. If there are multiple numbers in one box, there were multiple modifications identified. Colored numbers in the same row represent the same location.

Figure 7 shows the estimated locations for the modifications identified on *HBB* (Figure 7A), *S100A9* (Figure 7B) and *S100A8* (Figure 7C). Also included on Figure 7 is the approximate number of reads that mapped to each region for all three sequencing runs. The data shows that read lengths were relatively consistent across all samples, however the number of reads that mapped to each were lower in the aged sample compared to the fresh. This was consistent across all transcripts identified with potential modifications. All potential modifications found were located near the 5' or 3' end, no modifications were identified in the middle of the transcripts (Figure7). This was also consistent with the other transcripts identified with potential modifications. A total of four possible modifications were identified on *HBB*, however the potential modifications detected at the 3' end were only found in fresh samples (Figure 7A). Two modifications were identified on *S100A9*, both of which are located within 200bp of the 3' tail, and were found in all three samples (Figure 7B). Two modifications were also identified in S*100A8*, also located within 200bp of the 3' tail, however one was only found in fresh-2 while the other was found in both fresh samples (Figure 7C).

Figure 7. Approximate Modification Locations

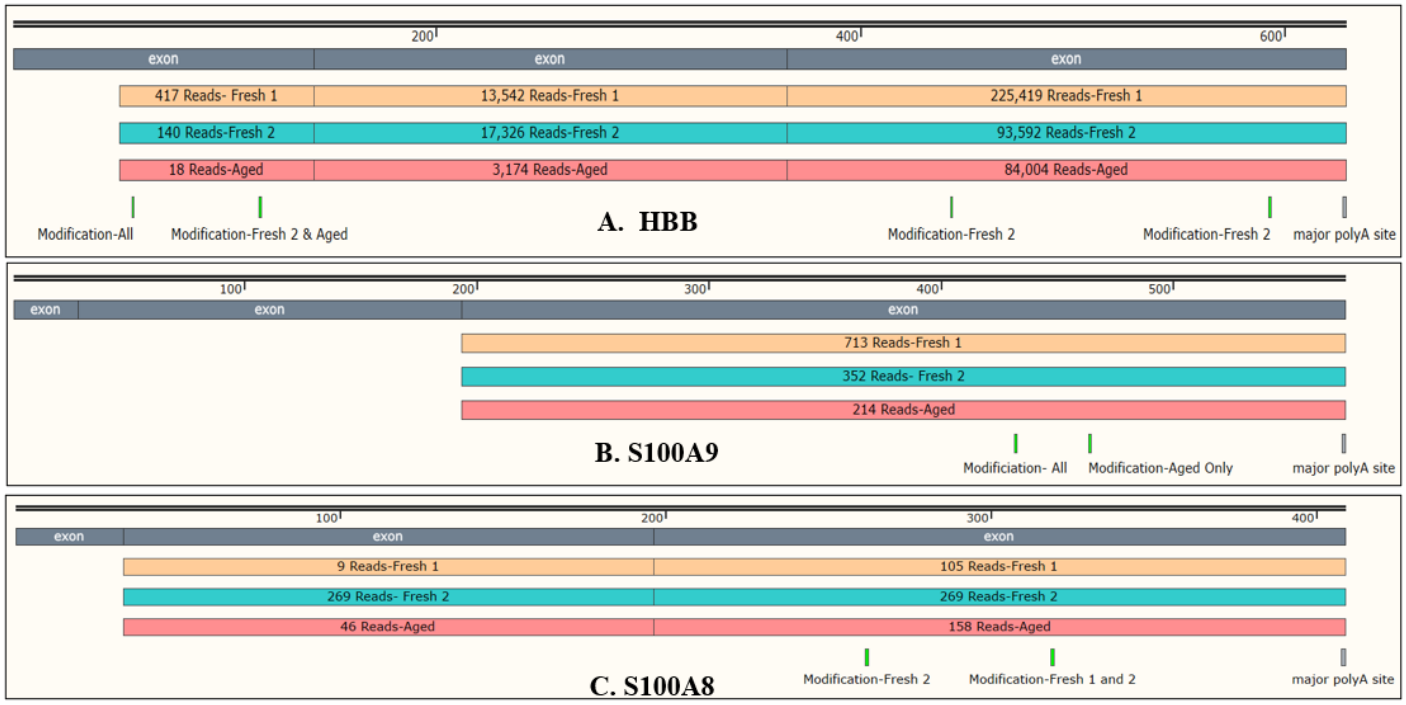


Figure 7 shows the estimated locations of modifications on HBB (A), S100A9 (B) and S100A8 (C). Number of reads mapping to each region are also included for fresh-1 (Tan), fresh-2 (Teal) and aged (pink) samples.

The visual outputs from the modification analyses are electrical signal plots (called squigglegrams) and box-and-whisker diagrams. The electrical signal plots show the raw current data that was recorded for each nucleotide as it passed through the sequencing pore. Comparing the transcripts and regions identified in multiple samples, the data shows consistent electric signal patterns across each sample (Figure 8). When the electrical current graphs are superimposed on top of each other, it is easy to see this consistency (Figure 9). This duplicated pattern is seen for all identical modifications identified on *HBB* and *S100A8* as well, shown in Appendix A and B respectively. Box-and whisker diagrams were also created for each potential modification identified. Each plot shows the statistical output from the model used for comparison next to the sample. When the electrical signal graphs are compared to the box-and-whisker plots similar patterns are observed (Figure 10). Box-and-whisker diagrams for all potential modification sites identified are provided in Appendix C.

Figure 8. S100A9 Electrical Signal Plots

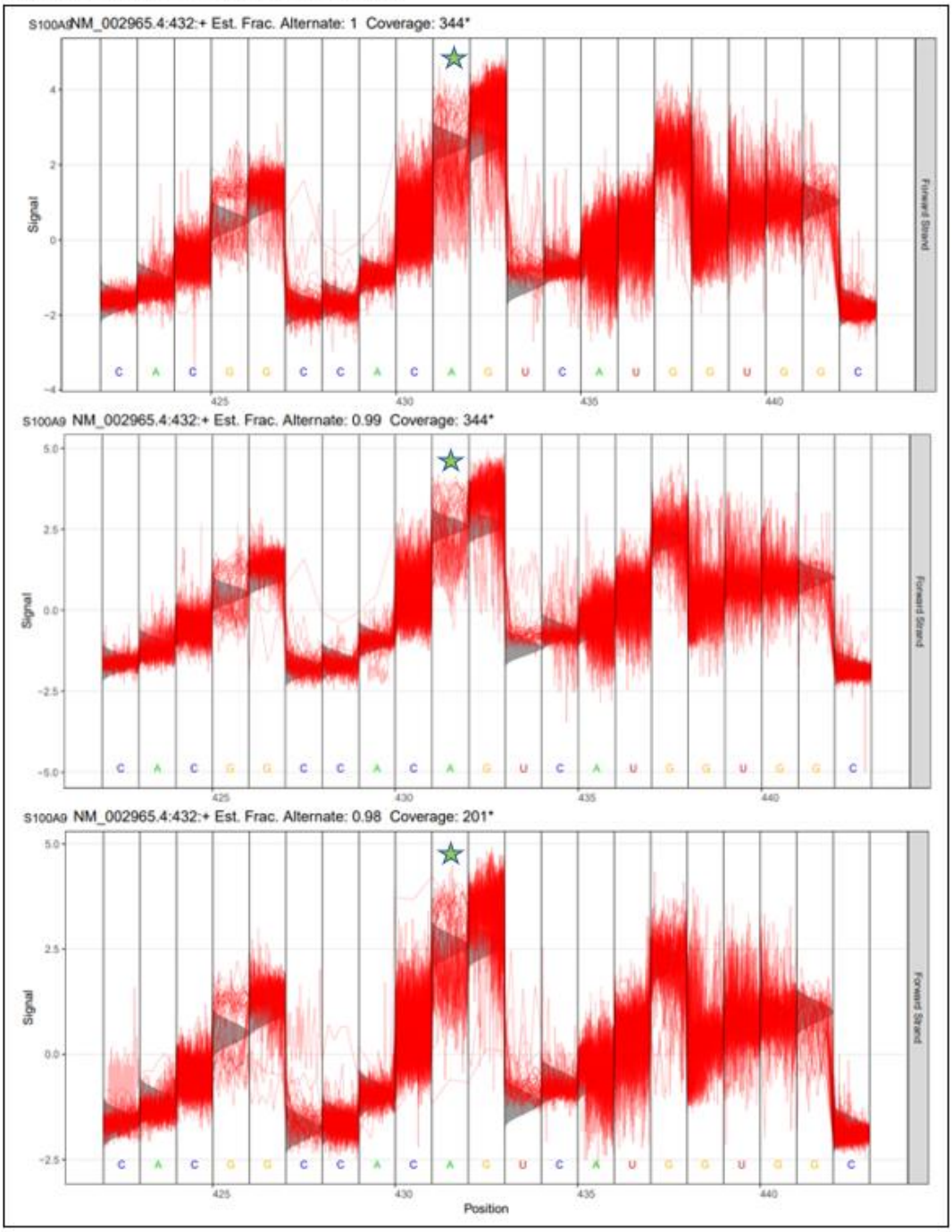


Figure 8 shows the raw electrical signal plots for the potential modification located at the 3' end of S100A9 that was found in all samples; fresh-1 (Top), fresh-2 (Middle) and aged (Bottom). Stars represent the estimated location.

Figure 9. Overlay of Electrical Signals of 3' S100A9 Modification

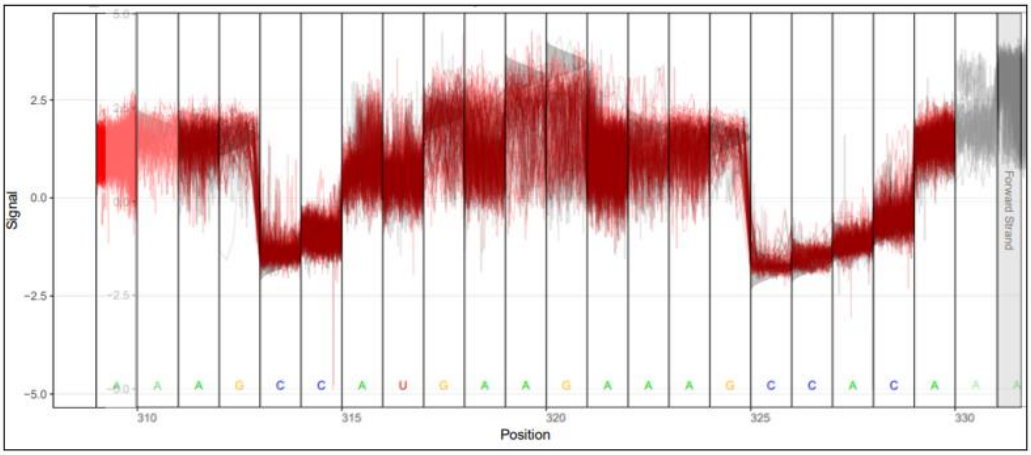


Figure 10. S100A9 Electrical Signal and Box-and-Whisker Plot Comparison

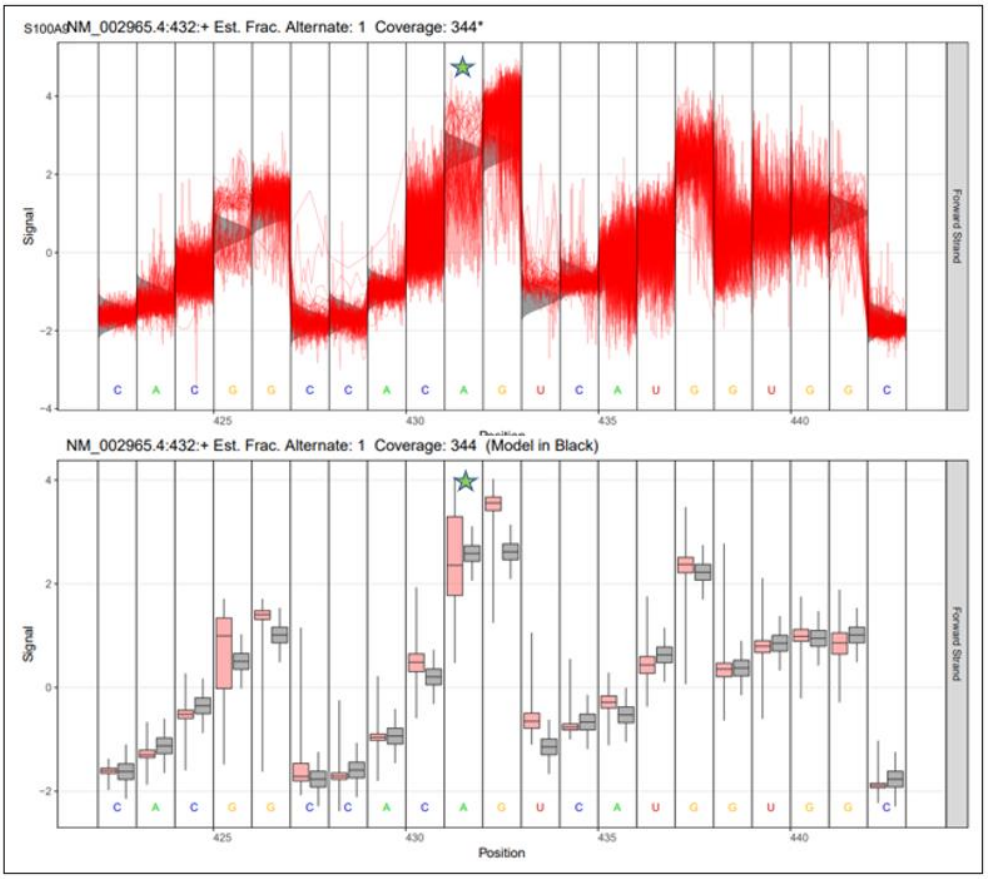


Figure 10 shows the comparison of the raw electrical signal and box-and-whisker plots for the potential modification identified on the 3' end of S100A9.

The star in the box plot denotes a suggested modification site according to Tombo.

Discussion of Findings

mRNA Enrichment Needs Further Optimization

Even with following experimental suggestions from Tan et al (2019)[14], our hybrid-capture protocol needs further optimization. A possible explanation for this is the difference in sample types/input between the two applications. Tan et al (2019)[14] utilized total RNA extracted from cultured cells, allowing for a 30μg sample input[14] compared to our ~1μg inputs (Table 1). Unfortunately, achieving higher sample concentrations was not feasible for us as we were working with dried bloodstains. Additionally, their increased sample input enabled them to double their hybridization time for a total of four hours[14]. This change can help with the specificity and abundance of mRNA enrichment. However, when we tried to increase hybridization time, we did not see a difference in on-target or off-target abundance; shown in Appendix D.

The additional issue observed from our targeted enrichment is the presence of undigested DNA probe in our final purified product (Figure 4). We followed the guidelines provided by Tan et al (2019)[14], however again, incubations had to be altered due to sample sensitivity. Furthermore, supplementary experiments were performed to test the effectiveness of different DNase I procedures, but we were never successful in fully digesting the DNA probes; shown in Appendix E. Finally, since we were not able to recover a sufficient amount of enriched mRNA for LC-MS/MS analysis, we decided that no further optimization was warranted, and switched our efforts to sequencing native RNA for modification detection.

Nanopore Sequencing Summaries

Results show that there was a large decrease, average 2μg, in concentration between initial sample input into library preparation and the final amount of adapter-ligated RNA prepped for sequencing (Table 5). This is expected, however, as only RNA with a poly(A) tail were

ligated to the sequencing adapter, which is about 1-5% of total RNA. Additionally, total data generated from the second fresh and aged sequencing runs is about half of that produced from the first sequencing run. While this is expected for the aged sample, as it has likely started to degrade, the difference between sample input and number of active pores between the first and second fresh bloodstain suggests an optimal active pore to sample ratio. Furthermore, the percentage of active pores for the second fresh sequencing run decreased at a steeper rate compared to the other two samples (Figure 5A-C). This is likely due to the clogging of pores from the RNA forming secondary structures and/or degrading during sequencing. Thus, more sample input is not always beneficial when sequencing native RNA using ONT.

The length of sequences obtained from each run is also of something to note. The average length of reads remains consistent between the two fresh samples (Figure 6A and B), however the average read lengths from the aged sample is shorter (Figure 6C). This is expected and agrees with previously published data on RNA degradation kinetics in dried bloodstains[5,9]. In summary, RNA degrades faster at the 5' end of transcripts in dried bloodstains. Thus, as samples are sequenced beginning at the 3' end, shorter reads will be produced as the 5' end degrades in older samples. Since the aged sample was only 11 weeks old, some longer reads were produced (Figure 7), however the number of longer reads was reduced, causing the average read-length to decrease. Although Nanopore specializes in long-read sequencing, shorter RNA transcripts can be expected, due to the fragile nature of RNA and the way the molecule is pulled through the pore for sequencing.

Preliminary Modification Results

As previously mentioned, as a result of not having controls for statistical comparison, results are preliminary and can only suggest the existence and possible locations for RNA modifications. This is displayed in Figure 10, which shows a squigglegram with its corresponding

box-and-whisker diagram. The estimated location of the modified base pair is nucleotide 432, marked with a star. However, at this location, the box-and-whisker diagram shows overlapping signals between the theoretical control and experimental sample. However, looking at the nucleotide directly adjacent to the estimated location (nucleotide 433), the box-and-whisker diagram no longer shows overlapping signals. This possible inaccuracy by the software is a result of not having control samples for statistical comparison. The model used is only able to estimate approximate location, and although it was close, visual observation suggests that the modification is located on nucleotide 433 and not 432 because of the wide separation of the boxes in the plot. Similar discrepancies were seen in some of the other modifications suggested, but not all.

Use of the model-free version Tombo introduces the possibility for a high false positive rate for detecting modified sites in an RNA sequence[41]; however, the reproducibility of squigglegram data for each transcript strengthens the interpretation of the results provided (Figures 8-10). Of the transcripts identified with potential modifications, half were found in more than one sample (Table 8). Furthermore, duplicate locations were identified six times (Table 8). Additionally, identical electrical current patterns are observed at suggested modified sites detected repeatedly and can be superimposed onto one another. Thus, although the data presented here is preliminary, it provides a solid foundation for future studies.

Potential modifications were identified for 12 different transcripts; however, this does not mean that no additional RNA modifications exist. The bioinformatic package used relies on realigning the raw electrical current data to the reference genome to identify any interruptions in the signal. These interruptions are then compared to either the model data set (de novo detection) or to control samples (methylated and unmethylated) of identical sequence to confirm differences in the electronic signal and thereby confirm a modified site. It was hypothesized that modified transcripts could have been missed in the computational analyses if too few reads were mapped to that region. However, looking at the number of reads for transcripts identified with modifications

versus non-modified transcripts shows this may not be the case. For example, based on previous degradation data, we believed we would identify potential modifications on *ACTB* and *S100A12*. We got read depths of 96, 48 and 188 for *ACTB* for fresh-1, fresh-2 and aged samples respectively. Additionally, we found 59, 22, and 9 reads mapping to *S100A12*. However, the software identified potential modifications on transcripts with as low as 2-7 reads mapping to them. Therefore, our hypothesis does not hold true about insignificant read-depth, and *ACTB* and *S100A12* may not be modified. Nonetheless, these results are preliminary, and experiments should be repeated with the proper controls. Furthermore, *HBB* was the only transcript identified with potential modifications located at the 5' end. However, the 5' end was not sequenced for most of the transcripts and therefore we cannot conclude if 5' modifications do exist on other transcripts.

Degradation of Potentially Modified Transcripts

The transcripts identified in this study with potential modifications were not originally chosen for validation of the 5'-3' assay[5], therefore limited degradation data is available. Nevertheless, early results show that both *S100A8* and *S100A9* follow similar degradation patterns to those previously seen for other transcripts in aged and dried bloodstains (Figure 11A and B). Although data is only available through 6-weeks of aging, *S100A8* and *S100A9* show correlation values of 0.85 (Figure 11A) and 0.84 (Figure 11B) respectively when degradation is followed using the 5'-3' assay. Data is included with permission from Dr. Jun Fu.

Figure 11. Degradation Rates for S100A8 and S100A9

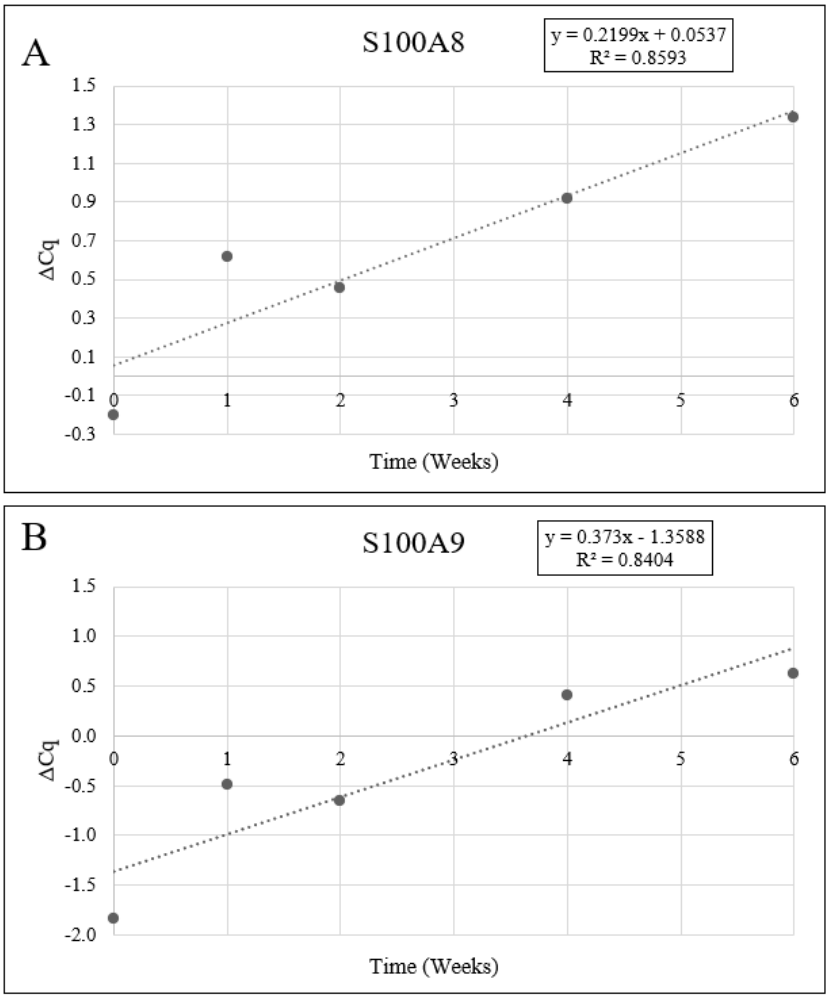


Figure 11 shows the available degradation data for S100A8 and S100A9.

There is also limited degradation data available for *HBB.* Since *HBB* is extremely abundant in blood, it is harder to detect subtle differences in abundance levels of 5' and 3' RNA fragments as they degrade, since the changes are small compared to the transcript's total abundance. However, early results suggest that the two ends of the *HBB* transcript degrade at similar rates up to around 200 days (Figure 12A and B). Whether the observed similarity between the 5' and 3' degradation rates of *HBB* is due to the presence of modifications on both ends of the transcript (Figure 7) or an artifact of the transcripts high overall abundance cannot be determined until further studies are done. Data is included with permission from Dr. Jun Fu.

Figure 12. HBB Degradation Rate

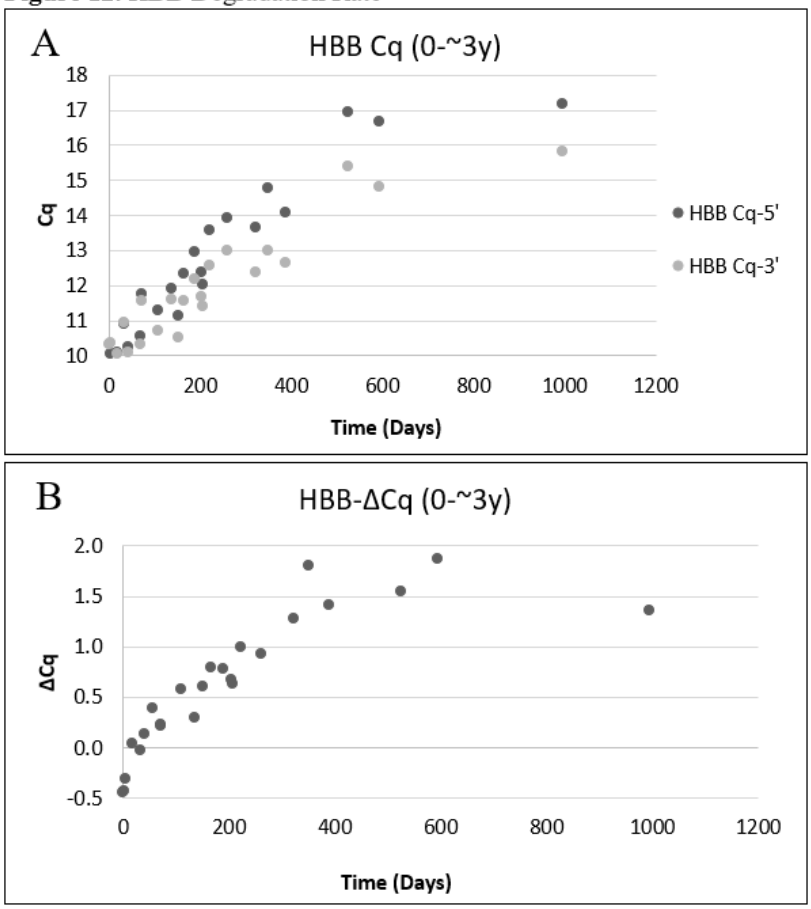


Figure 12 shows the available degradation data for HBB. 12A shows the Cq values for the 5' and 3' ends of HBB, while 12B shows the delta-Cq.

Future Directions

The data produced in this study can serve as the basis for subsequent studies aimed at interrogating the epitranscriptome of forensically relevant transcripts. Additional research is needed in order to confirm the presence of modified transcripts in dried bloodstains shown here. Future studies should include controls to be sequenced and analyzed with the experimental samples for an improved accuracy in identifying modified sites. Including controls will also allow for the exact location and modification type to be identified. Different methods should also be employed to investigate the epitranscriptome of less abundant transcripts in dried bloodstains.

One such option is using Nanopore's alternative native RNA sequencing method, which utilizes sequence-specific adapters to only ligate to your transcript of interest. However, this approach is more labor intensive as your target sequence must be at the very 3' end of your transcript for successful adapter ligation. Therefore, if the target mRNA has a poly(A) tail, as ours do, experimental manipulation is required prior to library preparation. Finally, similar studies should be done in other tissue types that are relevant to a forensic investigation, such as saliva and semen. This can help to expand the knowledge surrounding RNA in various tissues in order to fully understand the molecule's structure and degradation kinetics, helping to improve assays designed to age biological samples.

CHAPTER V

CONCLUSIONS

This study works to expand the knowledge and tools available for studying RNA degradation kinetics in a forensic setting using hybrid capture and native RNA NGS. Although further optimization is needed, the targeted mRNA enrichment approach discussed shows that manipulation of RNA for further analyses is possible, even from an aged bloodstain. If validated, this approach can enable in-depth analyses of targeted RNA transcripts in a relatively low cost and labor efficient manner. However, as validating this approach was beyond the scope of this study, we chose to focus on native RNA sequencing using ONT to interrogate the transcriptome of all poly(A) tailed transcripts within the human exome.

The work provided here shows that sequencing native RNA from forensically relevant samples is possible using the Nanopore platform. Furthermore, we showed that these sequencing reads can then be bioinformatically analyzed to predict potential modification sites. Comparing these results to degradation data from Oklahoma State University, School of Forensic Science, our data suggest that RNA modifications may be influencing differential degradation of the 5' and 3' ends of transcripts in dried bloodstains. Understanding how and why RNA degrades the way it does can be complex, but a thorough understanding of degradation kinetics is crucial for the design of an assay able to estimate the TSD for biological evidence recovered from a crime scene.

REFERENCES

1. Counsil TI, McKillip JL. Forensic blood evidence analysis using RNA targets and novel molecular tools. Biologia. 2010. pp. 175–182. doi:10.2478/s11756-010-0001-2

2. Roeder AD, Haas C. MRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. Int J Legal Med. 2013;127: 707–721. doi:10.1007/s00414-012-0794-3

3. Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, et al. Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. PLoS One. 2014;9. doi:10.1371/journal.pone.0100123

4. Haas C, Klesser B, Kratzer A, Bär W. mRNA profiling for body fluid identification. Forensic Sci Int Genet Suppl Ser. 2008;1: 37–38. doi:10.1016/j.fsigss.2007.10.064

5. Fu J, Allen RW. A method to estimate the age of bloodstains using quantitative PCR. Forensic Sci Int Genet. 2019;39. doi:10.1016/j.fsigen.2018.12.004

6. Alshehhi S, Haddrill PR. Estimating time since deposition using quantification of RNA degradation in body fluid-specific markers. Forensic Sci Int. 2019;298: 58–63. doi:10.1016/j.forsciint.2019.02.046

7. Heneghan N, Fu J, Pritchard J, Payton M, Allen RW. The effect of environmental conditions on the rate of RNA degradation in dried blood stains. Forensic Sci Int Genet. 2021;51. doi:10.1016/j.fsigen.2020.102456

8. Bird TAG, Walton-Williams L, Williams G. Time since deposition of biological fluids using RNA degradation. Forensic Sci Int Genet Suppl Ser. 2019;7: 401–402. doi:10.1016/j.fsigss.2019.10.028

9. Weinbrecht KD, Fu J, Payton M, Allen R. Time-dependent loss of mRNA transcripts from forensic stains. Res Reports Forensic Med Sci. 2017;Volume 7: 1–12. doi:10.2147/rrfms.s125782

10. Gagliardi D, Dziembowski A. 5′ and 3′ modifications controlling RNA degradation: from safeguards to executioners. Philos Trans R Soc B Biol Sci. 2018;373. doi:10.1098/rstb.2018.0160

11. Boo SH, Kim YK. The emerging role of RNA modifications in the regulation of mRNA stability. Experimental and Molecular Medicine. Springer Nature; 2020. pp. 400–408. doi:10.1038/s12276-020-0407-z

12. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014;505. doi:10.1038/nature12730

13. Du H, Zhao Y, He J, Zhang Y, Xi H, Liu M, et al. YTHDF2 destabilizes m6A-containing RNA through direct recruitment of the CCR4–NOT deadenylase complex. Nat Commun. 2016;7. doi:10.1038/ncomms12626

14. Tan CCS, Maurer-Stroh S, Wan Y, Sessions OM, de Sessions PF. A novel method for the capture-based purification of whole viral native RNA genomes. AMB Express. 2019;9: 45. doi:10.1186/s13568-019-0772-y

15. Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Miano V, et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat Commun. 2021;12: 7198. doi:10.1038/s41467-021-27393-3

16. Xu L, Seki M. Recent advances in the detection of base modifications using the Nanopore sequencer. J Hum Genet. 2020;65: 25–33. doi:10.1038/s10038-019-0679-0

17. Takamiya M, Saigusa K, Kumagai R, Nakayashiki N, Aoki Y. Studies on mRNA expression of tissue-type plasminogen activator in bruises for wound age estimation. Int J Legal Med. 2005;119. doi:10.1007/s00414-004-0453-4

18. Bauer M, Gramlich I, Polzin S, Patzelt D. Quantification of mRNA degradation as possible indicator of postmortem interval—a pilot study. Leg Med. 2003;5. doi:10.1016/j.legalmed.2003.08.001

19. Zhao D, Zhu B-L, Ishikawa T, Quan L, Li D-R, Maeda H. Real-time RT-PCR quantitative assays and postmortem degradation profiles of erythropoietin, vascular endothelial growth factor and hypoxia-inducible factor 1 alpha mRNA transcripts in forensic autopsy materials. Leg Med. 2006;8. doi:10.1016/j.legalmed.2005.09.001

20. Becker J, Schmidt P, Musshoff F, Fitzenreiter M, Madea B. MOR1 receptor mRNA expression in human brains of drug-related fatalities—a real-time PCR quantification. Forensic Sci Int. 2004;140. doi:10.1016/j.forsciint.2003.10.012

21. Sampaio-Silva F, Magalhães T, Carvalho F, Dinis-Oliveira RJ, Silvestre R. Profiling of RNA Degradation for Estimation of Post Morterm Interval. PLoS One. 2013;8. doi:10.1371/journal.pone.0056507

22. Qi B, Kong L, Lu Y. Gender-related difference in bloodstain RNA ratio stored under uncontrolled room conditions for 28 days. J Forensic Leg Med. 2013;20. doi:10.1016/j.jflm.2012.09.014

23. Houseley J, Tollervey D. The Many Pathways of RNA Degradation. Cell. Elsevier B.V.; 2009. pp. 763–776. doi:10.1016/j.cell.2009.01.019

24. Sidova M, Tomankova S, Abaffy P, Kubista M, Sindelka R. Effects of post-mortem and physical degradation on RNA integrity and quality. Biomol Detect Quantif. 2015;5: 3–9. doi:10.1016/j.bdq.2015.08.002

25. Song H, Liu D, Dong S, Zeng L, Wu Z, Zhao P, et al. Epitranscriptomics and epiproteomics in cancer drug resistance: therapeutic implications. Signal Transduct Target Ther. 2020;5. doi:10.1038/s41392-020-00300-w

26. Boccaletto P, Bagiński B. MODOMICS: An Operational Guide to the Use of the RNA Modification Pathways Database. 2021. pp. 481–505. doi:10.1007/978-1-0716-1307-8_26

27. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways—2013 update. Nucleic Acids Res. 2012;41. doi:10.1093/nar/gks1007

28. Zhang W, Qian Y, Jia G. The detection and functions of RNA modification m6A based on m6A writers and erasers. J Biol Chem. 2021;297. doi:10.1016/j.jbc.2021.100973

29. Knuckles P, Bühler M. Adenosine methylation as a molecular imprint defining the fate of RNA. FEBS Letters. Wiley Blackwell; 2018. pp. 2845–2859. doi:10.1002/1873-3468.13107

30. Nachtergaele S, He C. Chemical Modifications in the Life of an mRNA Transcript. Annu Rev Genet. 2018;52. doi:10.1146/annurev-genet-120417-031522

31. Jia G, Fu Y, He C. Reversible RNA adenosine methylation in biological regulation. Trends Genet. 2013;29. doi:10.1016/j.tig.2012.11.003

32. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell. 2012;149: 1635–1646. doi:10.1016/j.cell.2012.05.003

33. Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, et al. A majority of m 6 A residues are in the last exons, allowing the potential for 3′ UTR regulation. Genes Dev. 2015;29. doi:10.1101/gad.269415.115

34. Roost C, Lynch SR, Batista PJ, Qu K, Chang HY, Kool ET. Structure and Thermodynamics of N 6 -Methyladenosine in RNA: A Spring-Loaded Base Modification. J Am Chem Soc. 2015;137. doi:10.1021/ja513080v

35. McIntyre ABR, Gokhale NS, Cerchietti L, Jaffrey SR, Horner SM, Mason CE. Limits in the detection of m6A changes using MeRIP/m6A-seq. Sci Rep. 2020;10. doi:10.1038/s41598-020-63355-3

36. Gilbert W V., Bell TA, Schaening C. Messenger RNA modifications: Form, distribution, and function. Science (80-). 2016;352. doi:10.1126/science.aad8711

37. Hu Y, Fang Z, Mu J, Huang Y, Zheng S, Yuan Y, et al. Quantitative Analysis of Methylated Adenosine Modifications Revealed Increased Levels of N6-Methyladenosine (m6A) and N6,2′-O-Dimethyladenosine (m6Am) in Serum From Colorectal Cancer and Gastric Cancer Patients. Front Cell Dev Biol. 2021;9. doi:10.3389/fcell.2021.694673

38. Meng Z. Mass spectrometry of RNA: linking the genome to the proteome. Briefings Funct Genomics Proteomics. 2006;5. doi:10.1093/bfgp/ell012

39. Jora M, Lobue PA, Ross RL, Williams B, Addepalli B. Detection of ribonucleoside modifications by liquid chromatography coupled with mass spectrometry. Biochim Biophys Acta - Gene Regul Mech. 2019;1862. doi:10.1016/j.bbagrm.2018.10.012

40. Guo A, Gu H, Zhou J, Mulhern D, Wang Y, Lee KA, et al. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. Mol Cell Proteomics. 2014;13. doi:10.1074/mcp.O113.027870

41. Furlan M, Delgado-Tejedor A, Mulroney L, Pelizzola M, Novoa EM, Leonardi T. Computational methods for RNA modification detection from nanopore direct RNA sequencing data. RNA Biol. 2021;18: 31–40. doi:10.1080/15476286.2021.1978215

42. Lauman R, Garcia BA. Unraveling the RNA modification code with mass spectrometry. Mol Omi. 2020;16. doi:10.1039/C8MO00247A

43. Wein S, Andrews B, Sachsenberg T, Santos-Rosa H, Kohlbacher O, Kouzarides T, et al. A computational platform for high-throughput analysis of RNA sequences and modifications by mass spectrometry. Nat Commun. 2020;11. doi:10.1038/s41467-020-14665-7

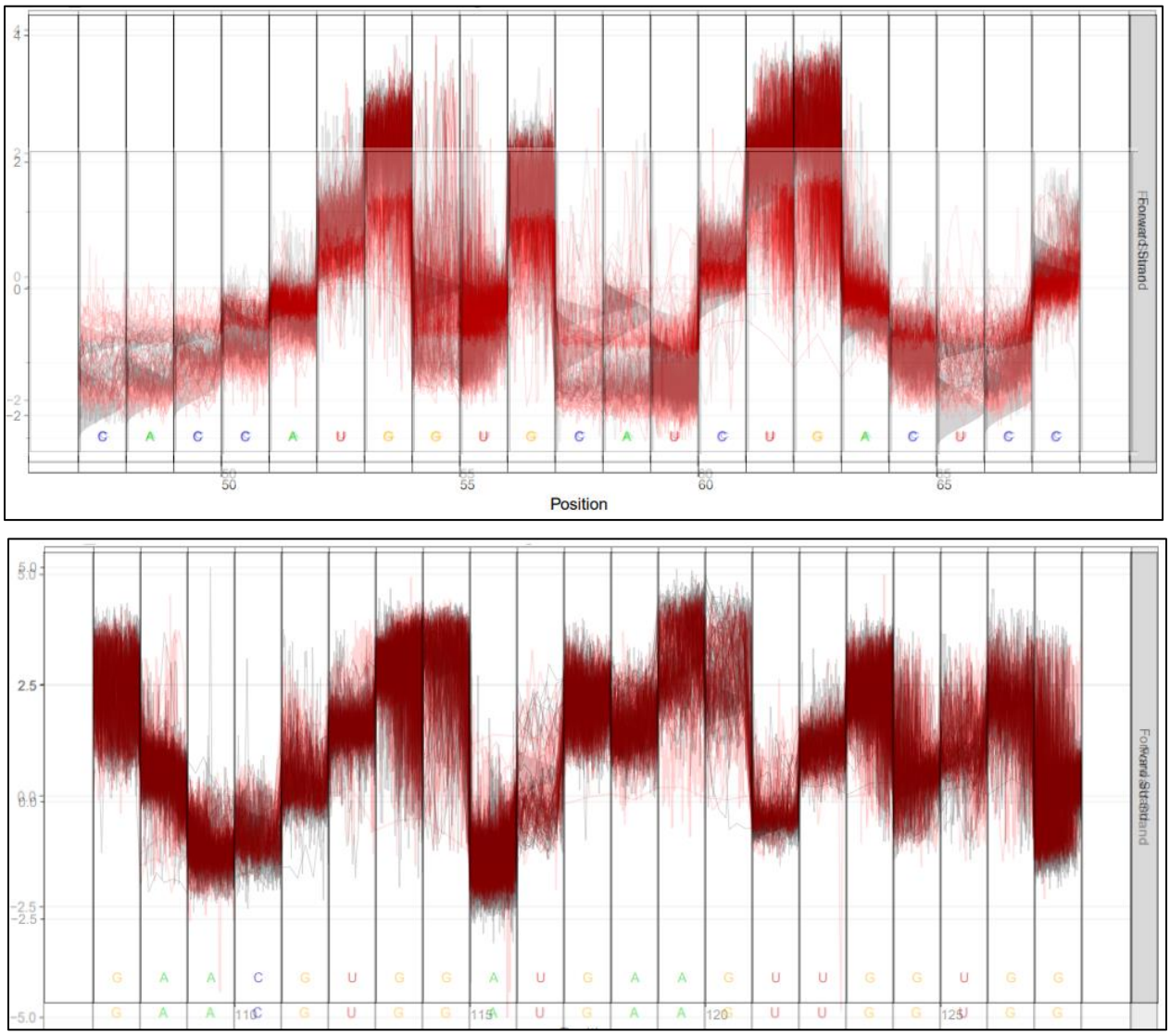
44. Urban PL. Quantitative mass spectrometry: an overview. Philos Trans R Soc A Math Phys Eng Sci. 2016;374: 20150382. doi:10.1098/rsta.2015.0382

45. Soneson C, Yao Y, Bratus-Neuenschwander A, Patrignani A, Robinson MD, Hussain S. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. Nat Commun. 2019;10: 3359. doi:10.1038/s41467-019-11272-z

46. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39: 1348–1365. doi:10.1038/s41587-021-01108-x

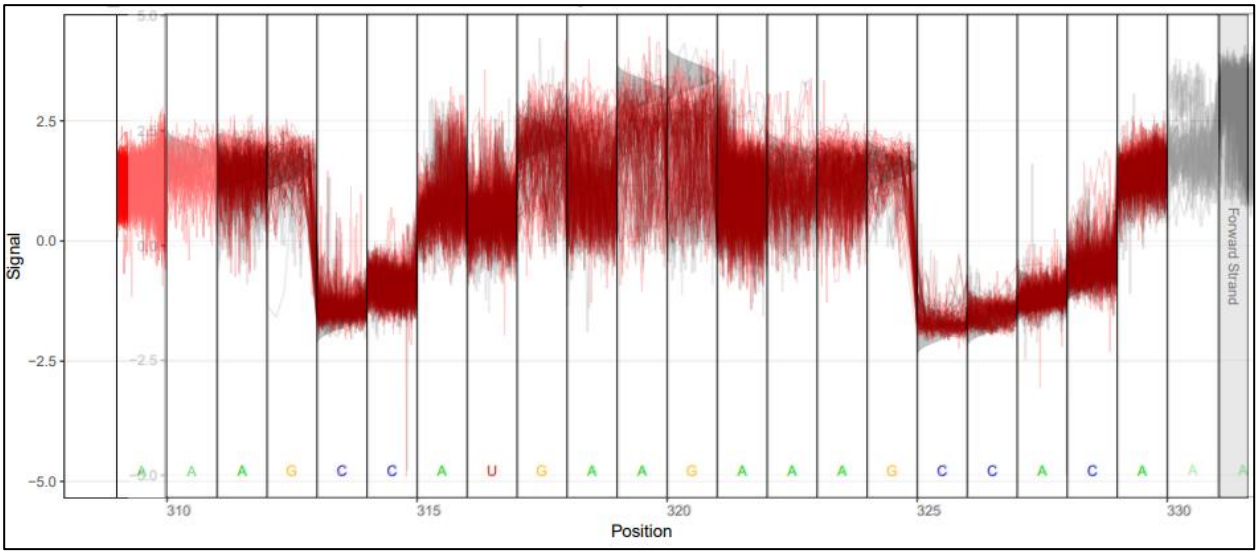
APPENDICES

Appendix A. Superimposed Electrical Current Plots for Duplicate Modifications on HBB



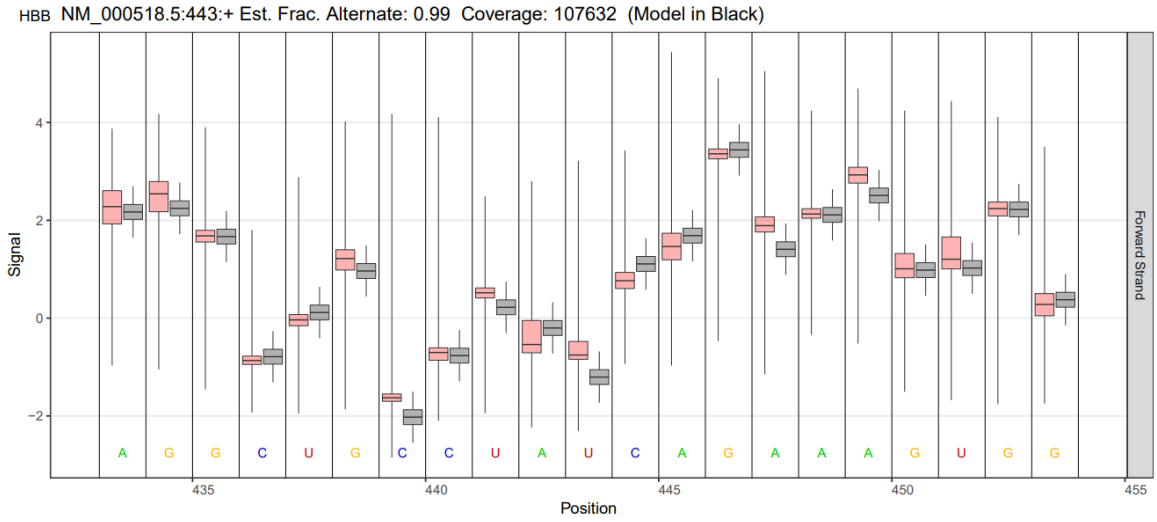
Electrical current plots for modifications identified on HBB found in more than one sample. Top: HBB 5'-found in all three samples, bottom: HBB 3'-Found in fresh 2 and aged.

Appendix B. Superimposed Electrical Current Plots for S100A8

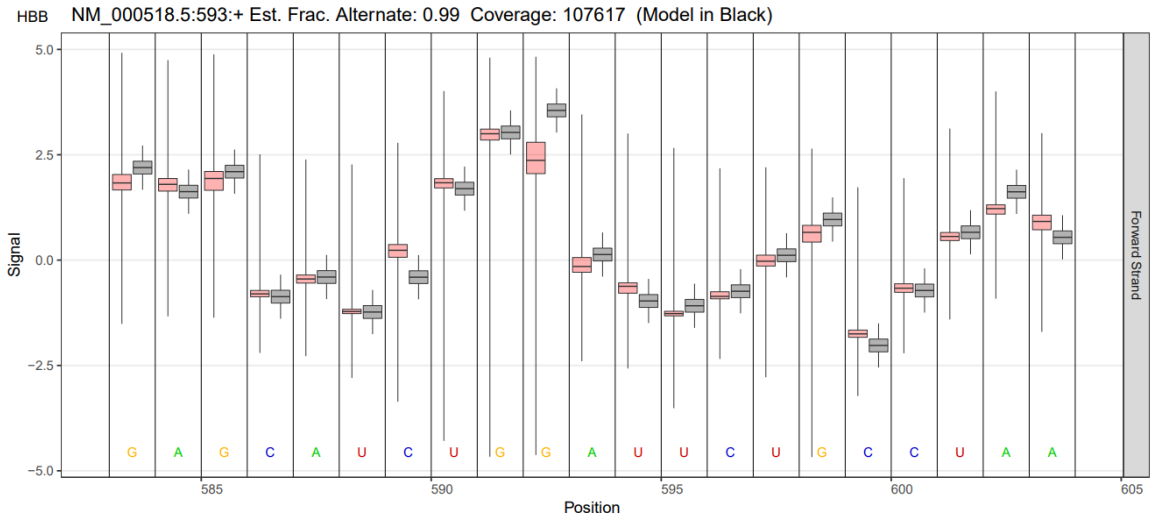


Electrical current plots for modifications identified on S100A8 found in more than one sample. 3'-Found in both fresh samples.

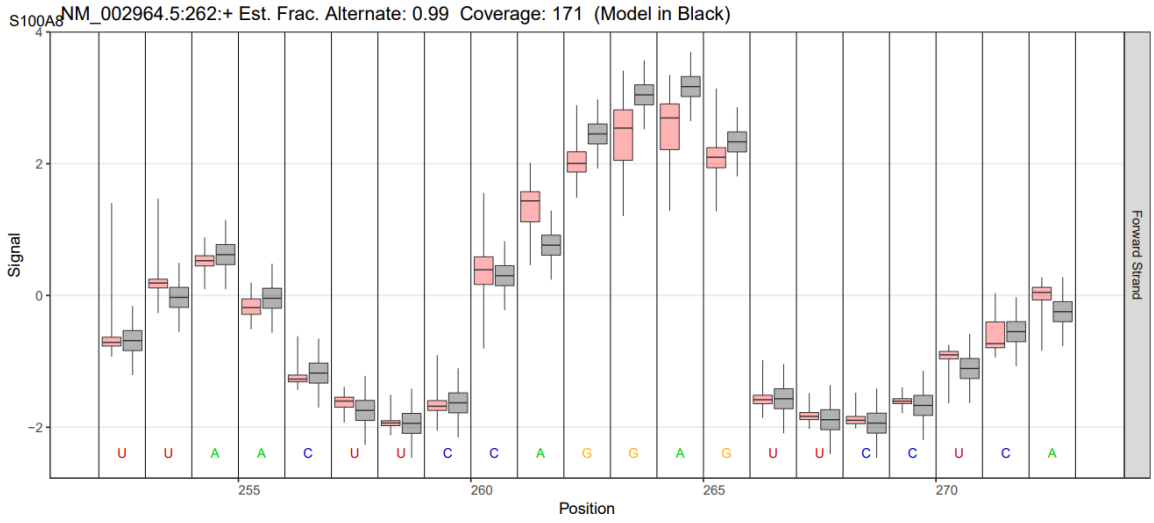
Appendix C. Box-and-Whisker Diagrams for Identified Modifications



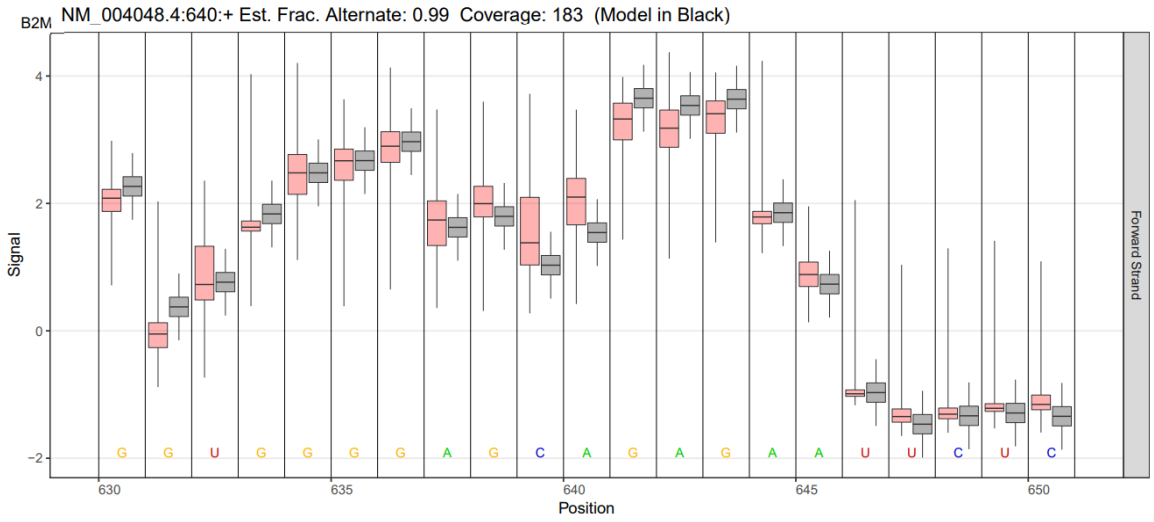
HBB 3'- Only found on fresh 2.



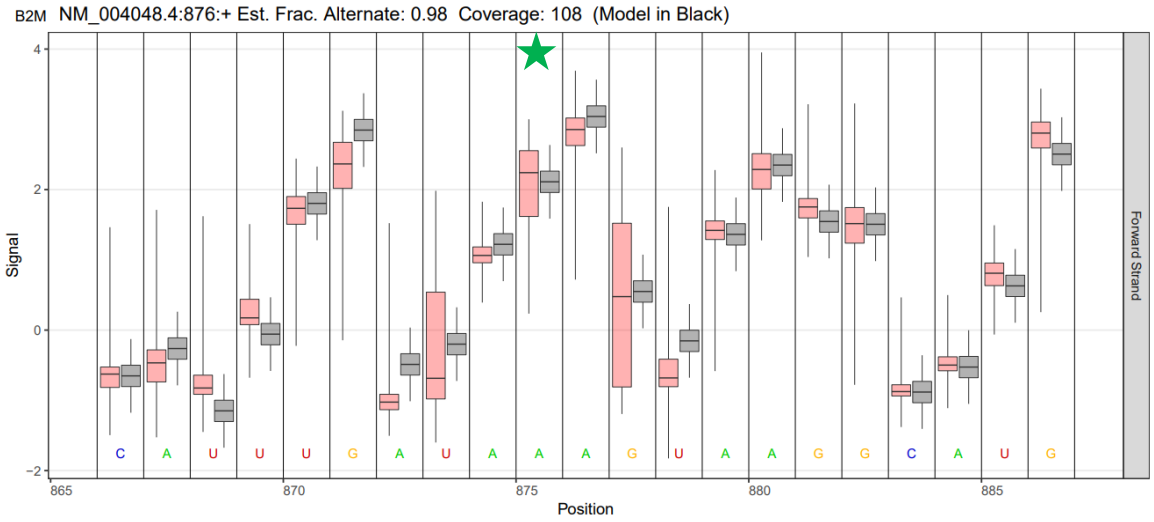
HBB 3'- Only found on fresh 2.



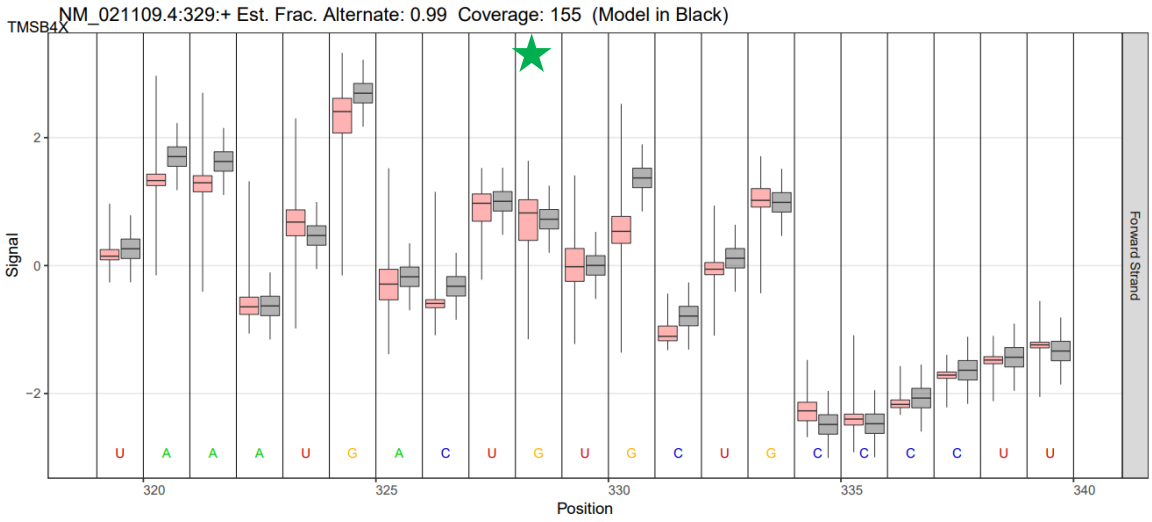
S100A8 3'- Only found on fresh 2.



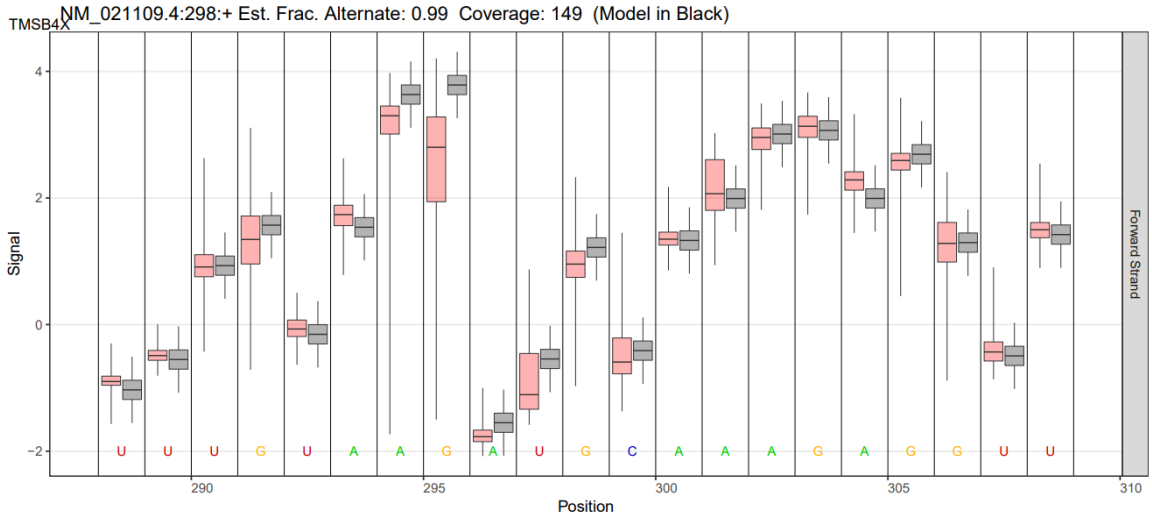
B2M 3'- Only found on fresh 1.



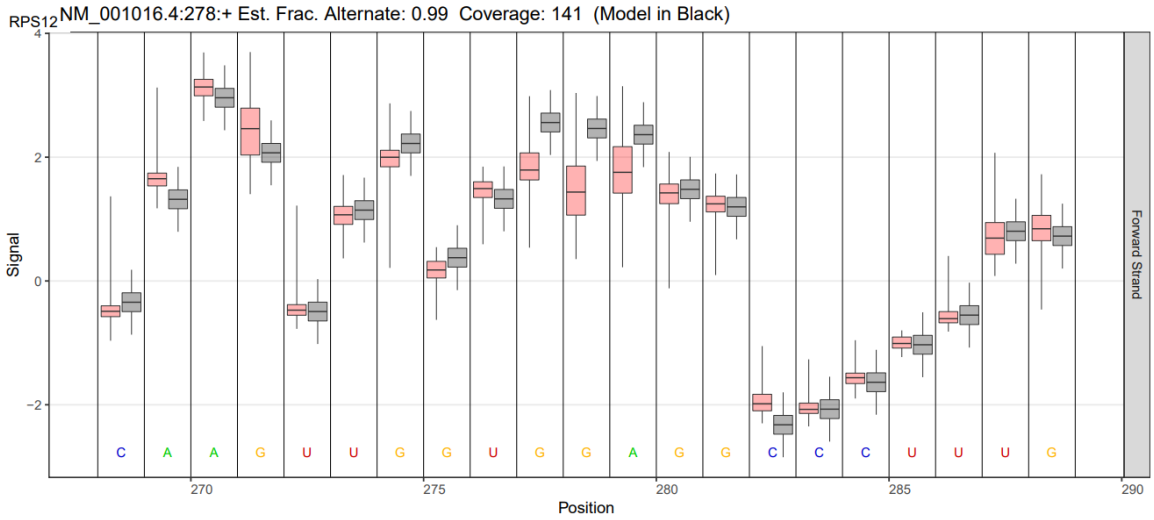
B2M 3'- Only found on aged.



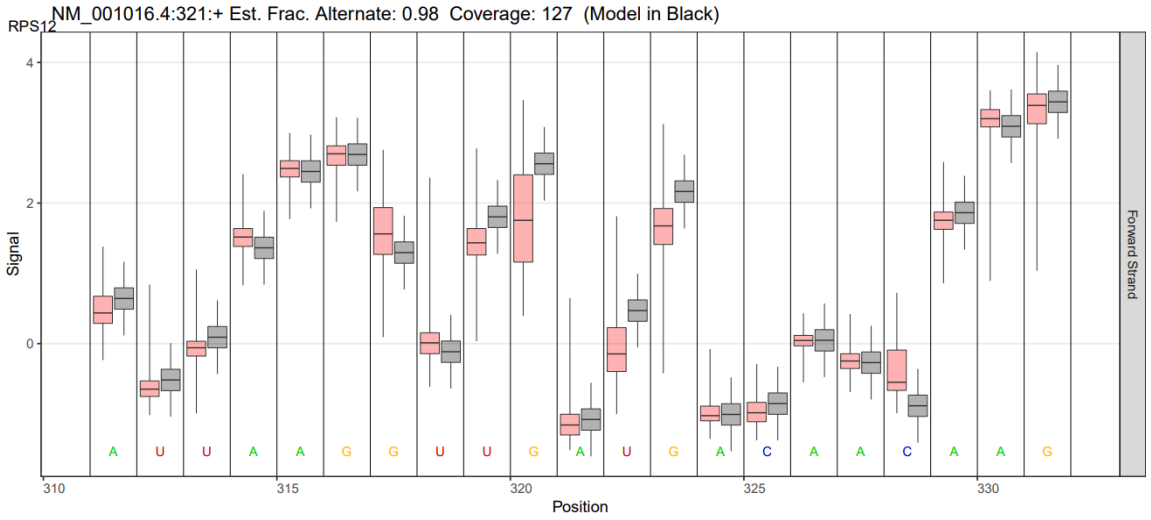
TMSB4X 3'- Found on both fresh samples.



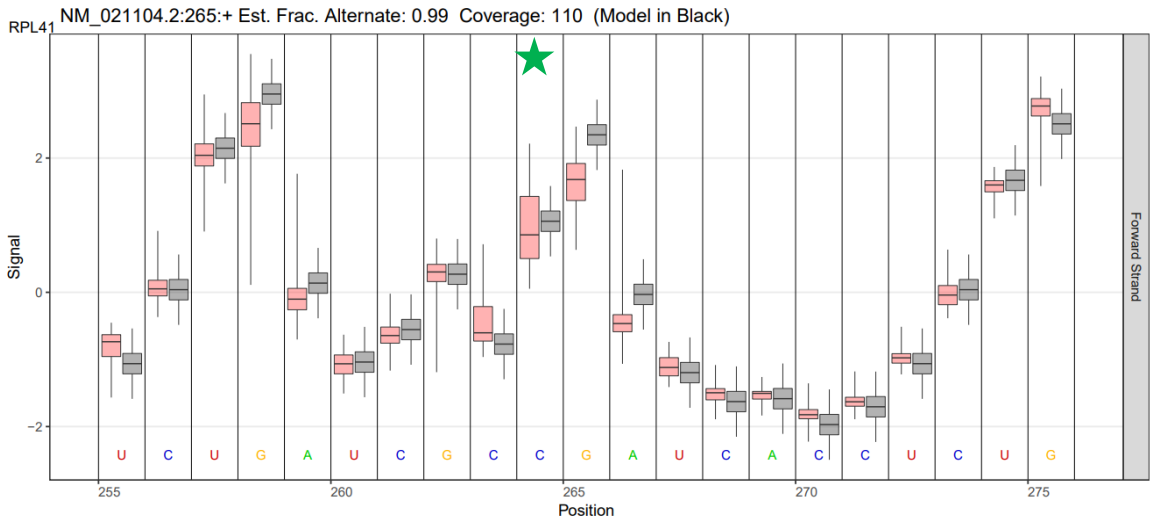
TMSB4X 3'- Only found on fresh 1.



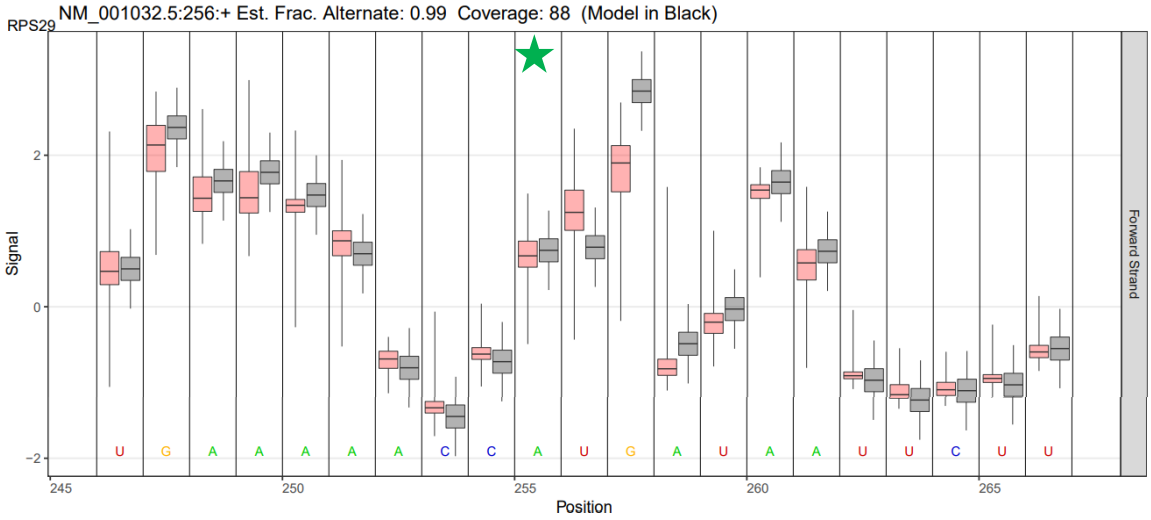
RPS12 3'- Found on all samples.



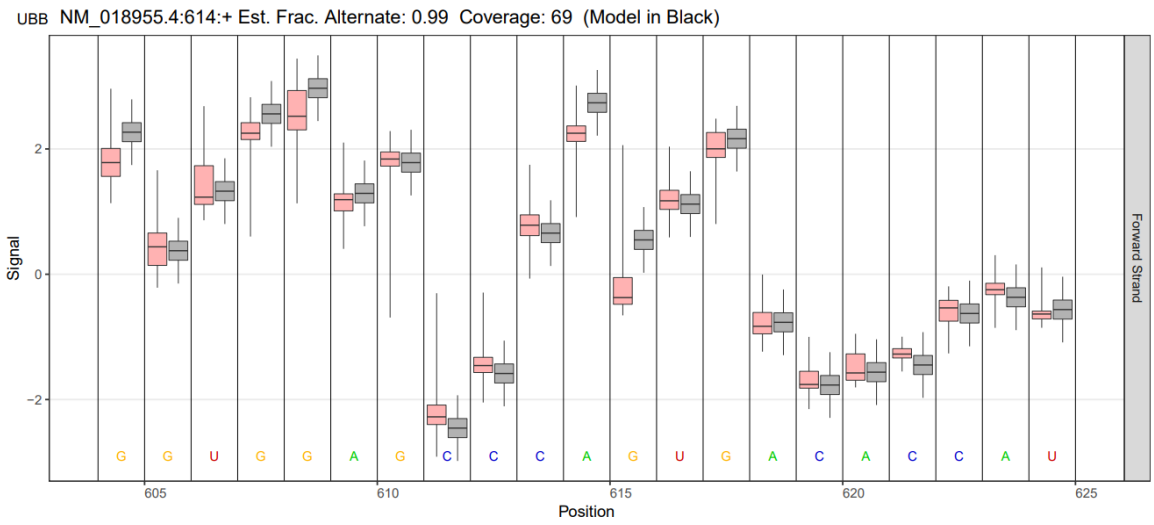
RPS12 3'- Only found on aged.



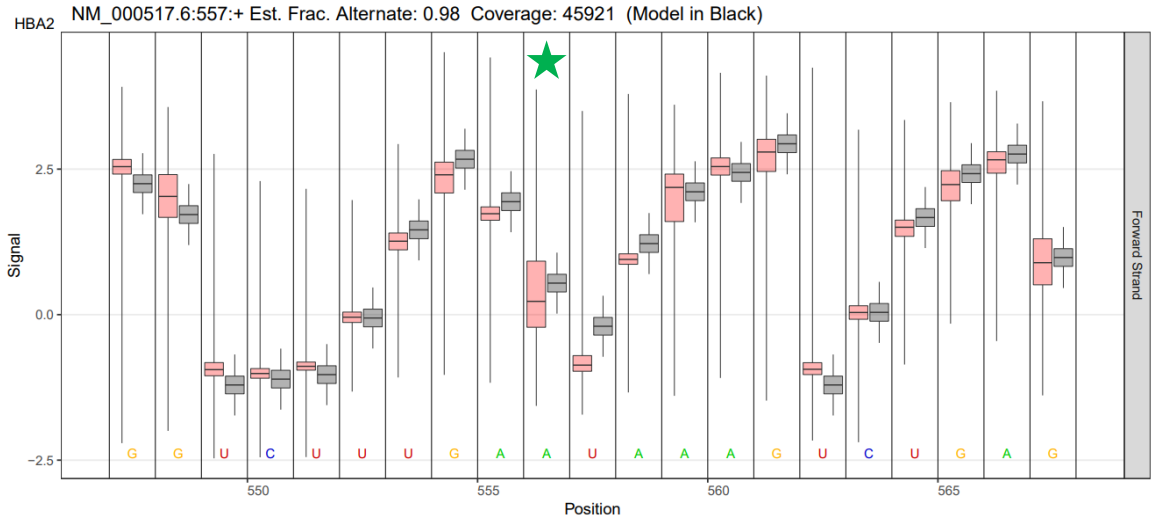
RPL41 3'- Only found on fresh 1.



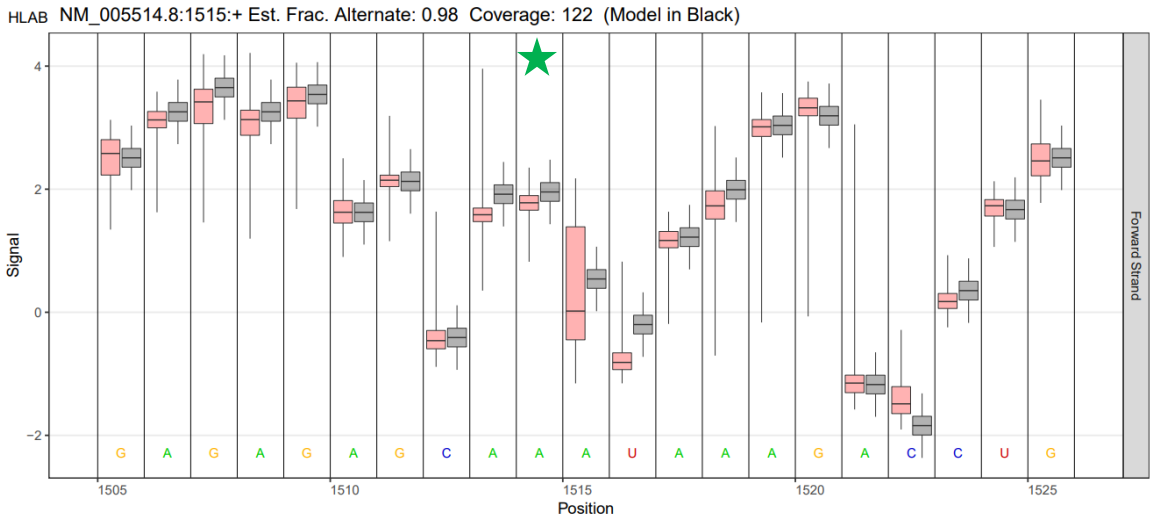
RPS29 3'- Only found on fresh 1.



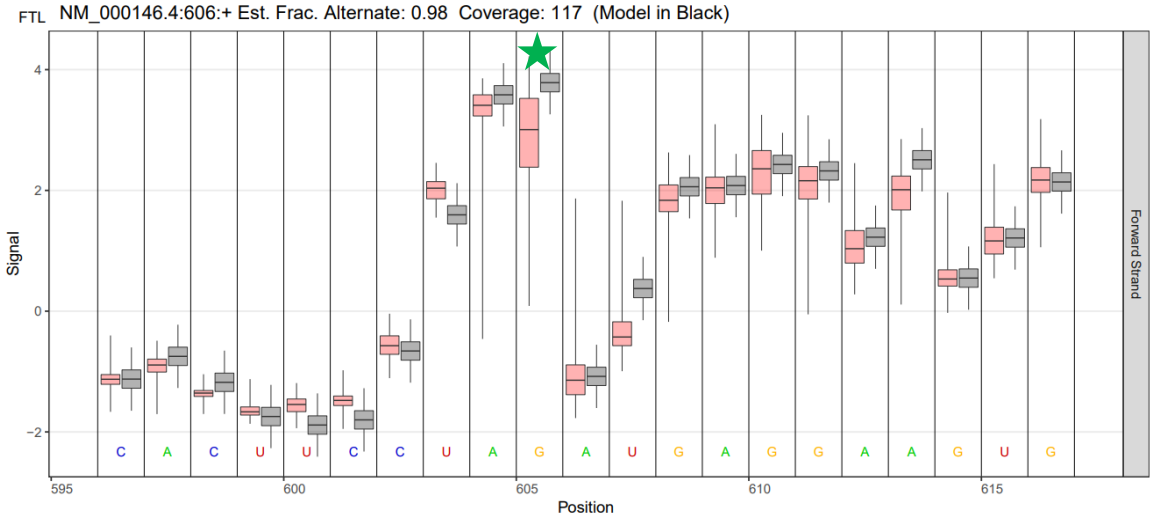
UBB 3'- Only found on fresh 1.



HBA-2 3'- Only found on aged.

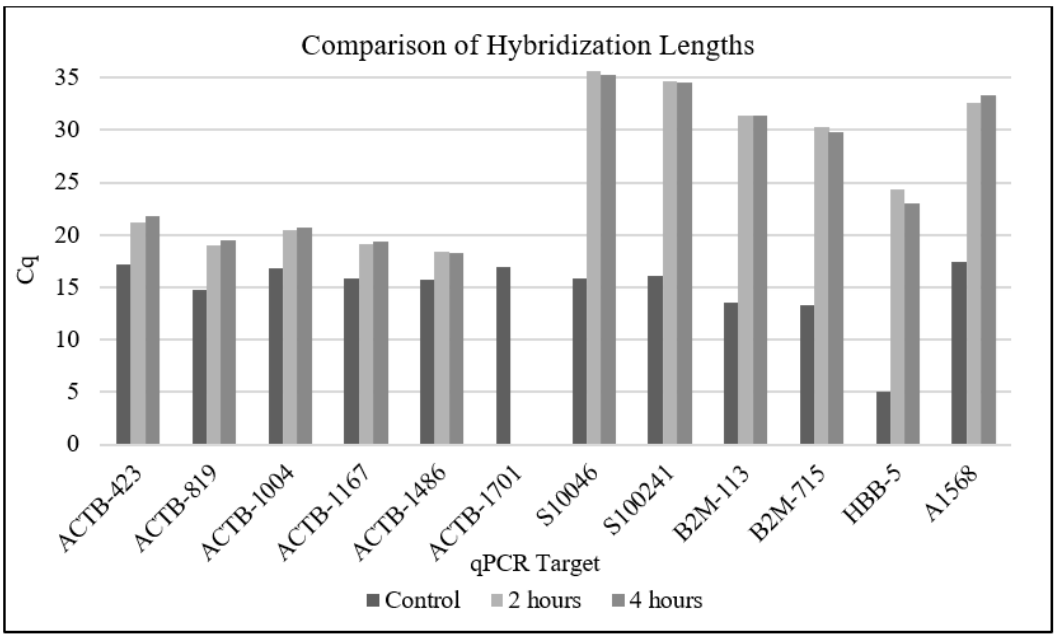


HLAB 3'- Only found on aged.



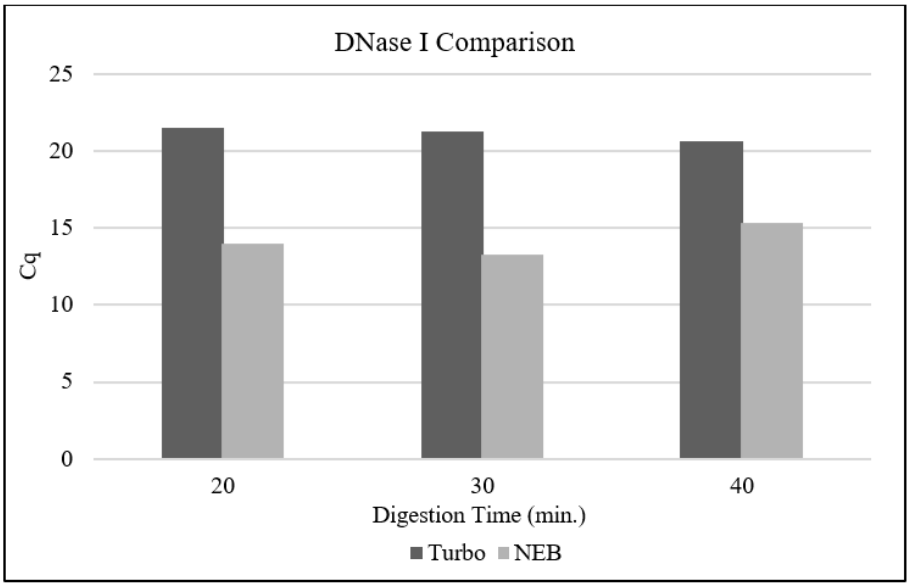
FTL 3'- Only found on aged.

Appendix D. Probe Hybridization Protocol Optimization



Results from hybrid capture optimization. Increased hybridization time does not decrease off-target products or increase on-target abundance. No products are shown for ACTB-1701 because the sample was not digested with DNase I to remove the probe.

Appendix E. DNase I Comparison



Results from DNase I Comparison. TURBO DNase I does a better job of digesting the DNA probe, however DNA contamination is still seen in the final product.

VITA

Chaelynne Lohr

Candidate for the Degree of

Master of Science

Thesis: DETECTION OF RNA METHYLATION PATTERNS IN FORENSICALLY RELEVANT TRANSCRIPTS

Major Field: Forensic Sciences

Biographical:

Education:

Completed the requirements for the Master of Science in Forensic Sciences, Emphasis in Forensic Biology/DNA at Oklahoma State University – Center for Health Sciences, Tulsa, Oklahoma in July, 2022.

Completed the requirements for the Bachelor of Science in Human Biology at Michigan State University, East Lansing, Michigan in May, 2019.