DEPARTMENT OF PURE & APPLIED CHEMISTRY

University of Strathclyde Science

Forensic RNA analysis for estimating the age of body fluid stains

Suaad Alshehhi

Centre for Forensic Science
Department of Pure & Applied Chemistry
University of Strathclyde

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2019

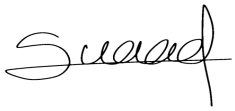
Centre for Forensic Science
Department of Pure & Applied Chemistry
University of Strathclyde

Forensic RNA analysis for estimating the age of body fluid stains

PhD Thesis

Suaad Alshehhi

2019

Signed:  Date: 02/09/2019

ACKNOWLEDGEMENTS

First and foremost, I thank Allah the greatest of all for having made everything possible by giving me strength and courage to do this work.

I want to express gratitude to my family for their continuous support and for believing in me.

I would like to give an enormous thanks to my loving husband Hamdan for his constant support, inspiration and for all his care and concern throughout my PhD journey.

I would like to thank my supervisor Dr Haddrill for her valuable advised and help throughout my PhD. I have been extremely lucky to have a supervisor who have been a tremendous mentor for me and responded to my queries promptly.

I also would like to give special thanks to Dr McCallum who didn't hesitate to help and advised me at earlier stage for my research.

PUBLICATIONS AND PRESENTATIONS

Publications:

Suaad Alshehhi, Nicola A. McCallum, Penelope R. Haddrill (2017), Quantification of RNA degradation of blood-specific markers to indicate the age of bloodstains, *Forensic Science International: Genetics Supplement Series*, 2017;6: e453-e55

Suaad Alshehhi, Penelope R. Haddrill (2019), Estimating time since deposition using quantification of RNA degradation in body fluid-specific markers, *Forensic Science International*, 2019;298: 58-63

Suaad Alshehhi, Penelope R. Haddrill, Evaluating the effect of body fluid mixture on the relative expression ratio of blood-specific RNA markers, Submitted to *Forensic Science International*, Status under review.

Oral presentations:

Quantification of RNA degradation of blood-specific markers to indicate the age of bloodstains. Policing Research and Police Scotland student award 2017, Edinburgh, Scotland (December 2017).

Quantification of RNA degradation of body fluid-specific markers to indicate the age of biological stains. 4th Scottish Student Forensic Research Symposium, Dundee, Scotland (March 2018).

Quantification of RNA degradation of body fluid-specific markers to indicate the age of biological stains. 8th European Academy of Forensic Science Conference, Lyon, France (August 2018).

Poster presentation:

Quantification of RNA degradation of blood-specific markers to indicate the age of bloodstains. Emirates International Forensic Conference & Exhibition 2017, Dubai, United Arab Emirates (March 2017).

Quantification of degradation in blood-specific markers for estimating the age of bloodstains. The 27th Congress of the International Society for Forensic Genetics, Seoul, South Korea (August 2017).

ABSTRACT

The first appearance of ribonucleic acid (RNA) in forensic science research was in 1984 in the study of post-mortem tissues. Since then, many studies have explored the role of gene expression and its potential applications in forensic science. Some RNA types such as messenger RNA (mRNA) and microRNA (miRNA) have been subject to increasing interest in the forensic science community.

It has been shown that analysis of RNA molecules can also be applied to determine the time when a biological stain was deposited at a crime scene, by analysing their stability and degradation rate. Estimating the age of a biological stain can provide essential information to an investigation. This research has analysed the expression and degradation level of different types of body fluid-specific RNA markers, including mRNA and miRNA markers, with the aim of identifying RNA markers that can be used to estimate the age of three types of body fluid (blood, saliva and semen). In this approach, reverse transcription quantitative PCR (RT-qPCR) has been utilised using appropriate primers and TaqMan® probes. The relative expression ratio (RER) between different types of body fluid-specific RNA markers was then examined in aged body fluid samples across one-year storage interval, with the aim of developing a method to estimate the age of body fluid stains recovered at crime scenes.

Each RNA marker exhibited unique degradation behaviour across ageing time points for blood, saliva and semen samples, with miRNAs showing high stability. Statistically significant correlations were found between the RERs of RNA markers and ageing time points, exhibiting a non-linear relationship. Overall, the RERs of body fluid-specific markers can be considered as a potential method for estimating the age of biological stains, and this has a wide range of applications in forensic science.

ABBREVIATION

°C	Degrees Celsius
18S rRNA	18S ribosomal RNA
3' UTR	3' untranslated regions
A_{230}	Absorbance, in optical densities, at 230 nm
A_{260}	Absorbance, in optical densities, at 260 nm
A_{280}	Absorbance, in optical densities, at 280 nm
ACTB	ß-actin mRNA
ANOVA	Analysis of variance
asRNA	Anti-sense RNA
ceRNA	Competing endogenous RNA
C_q	Quantification cycle
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
dRn (ΔRn)	Normalised reporter signal minus the baseline
dsDNA	Double-stranded DNA
dsRNA	Double stranded RNA
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
HBA	Hemoglobin alpha
HBB	Hemoglobin beta
HBD-1	Human beta-defensin 1
HMBS	Hydroxymethylbilane Synthase
hnRNA	Heterogeneous nuclear RNA
HTN3	Histatin 3
KLK3	Kallikrein 3
MIQE	Minimum information for publication of Quantitative Real-Time PCR Experiment
min	Minutes
miRNA	MicroRNA
mL	Millilitre

mRNA	messenger RNA
***MMP**-11*	Matrix metalloproteinase 11
***MMP**-7*	Matrix metalloproteinase 7
mRNA	Messenger RNA
ng	Nanogram
nt	Nucleotides
PBGD	Porphobilinogen-Deaminase.
PCR	Polymerase chain reaction
PRM1	Protamine 1
PRM2	Protamine 2
PPIA	Peptidylprolyl Isomerase A
qPCR	Quantitative real-time polymerase chain reaction
RER	Relative expression ratio
RIN	RNA integrity number
RNA	Ribonucleic acid
RNase	Ribonucleases
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcription–quantitative real time-polymerase chain reaction
s	Seconds
S	Standard error of regression
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
STATH	Statherin
TDMD	Target-directed miRNA degradation
Tm	Melting temperature
tRNA	Transfer RNA
µL	Microlitre

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter one: Introduction

1.1 Thesis overview

The main aim of this project was to develop a method to estimate the age of body fluid stains (blood, saliva, and semen) by analysing the degradation of RNA markers. This would allow estimation of the time of deposition of biological stains at crime scenes. This type of information can be crucial in a criminal investigation as it can provide information regarding when a crime occurred or help to exclude samples that do not correspond to the time of the crime. A small number of existing publications have examined the degradation level of RNA markers over time in an attempt to estimate the age of biological stains. These studies are reviewed in this thesis, discussing their findings and limitations, to provide a rationale for the current project.

Additionally, the identification of different body fluid types can also be very important to crime investigations. Many RNA markers have been identified that can indicate the presence of different types of body fluids. This thesis also reviewed the literature in this field, and discussed the papers that have identified RNA markers for different body fluids such as blood, saliva, semen, menstrual blood and vaginal secretions.

Since the discovery of deoxyribonucleic acid (DNA) structure by Watson and Crick in 1953 [1], it has been well understood how genetic information is encoded and transferred from one cell to another and from one generation to another by the three dimensional structure of DNA. It is also accepted that the conversion of genetic information into proteins is mediated by ribonucleic acid (RNA), however it is not fully understood how gene expression is regulated. There are 22,000-25,000 protein-coding genes that have been identified by the human genome project [2]. RNA molecules play a crucial role in translating these genes into proteins, and this process involves a variety of different

mechanisms. In order to understand the function of RNA and its potential application in forensic science, it is important to initially understand its structure and the different types of RNA molecules that are known; these details are presented in the current Chapter.

The instability and integrity of RNA makes it notoriously difficult to work with, and analysing RNA from challenging samples such as crime scene samples makes it even more difficult. Therefore, various RNA methodologies were compared and analysed in Chapter three to determine which method is more suitable for analysis of RNA from body fluid samples. Additionally, the degradation behaviour of individual RNA markers across ageing time points in three different types of body fluid was explored in detail in Chapters four to six, in order to determine whether there is a relationship between RNA degradation rate and ageing time points. Finally, reference genes play an important role in gene expression analysis and selecting suitable genes for RNA studies is crucial. Chapter seven therefore investigated the degradation rate of two commonly used reference genes, and their stability in aged body fluid stains.

1.2 The molecular biology and the structure of RNA

RNA is similar to DNA in structure. It is made up of individual subunits called nucleotides, each consisting of a ribose sugar (rather than the deoxyribose sugar found in DNA), a phosphate group and a nitrogenous base (Figure 1.1) [3]. The β-glycosidic bond holds the base to the ribose sugar, while a phosphodiester bond holds the phosphate group to the ribose sugar.

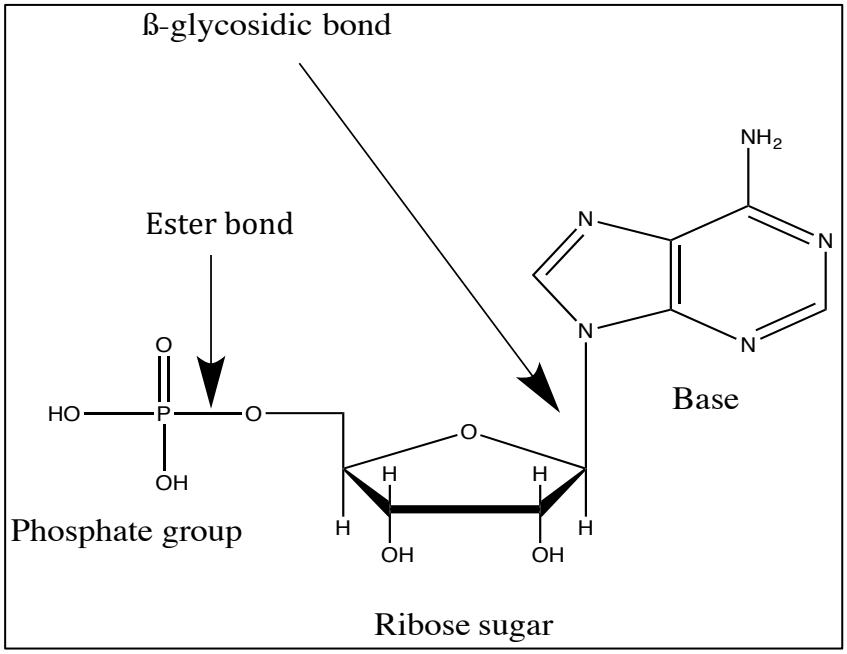


Figure 1.1: **General structure of an RNA nucleotide.** The β-glycosidic bond links the base to the ribose sugar, and the phosphodiester bond links the phosphate group to the ribose sugar. Illustration was drawn in ChemDraw Professional v.15.

There are four different nitrogenous bases that are found in RNA molecules: adenine (A), cytosine (C), guanine (G), and uracil (U) (Figure 1.2), and each ribonucleotide carries one of these four bases.

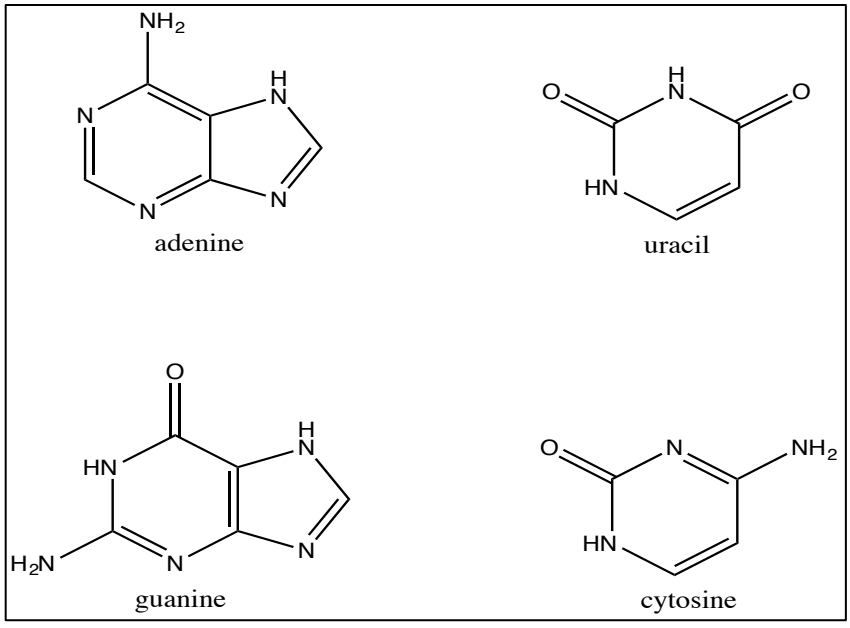


Figure 1.2: **The four nitrogenous bases that that are found in RNA molecules.** Illustration was drawn in ChemDraw Professional v.15.

RNA nucleotides are joined together through a phosphodiester backbone, which plays a critical role in RNA biology [4]. These phosphodiester bonds give RNA molecules a directional polarity denoted 5′-3′, because every RNA molecule starts with a 5′ phosphate group and ends with a 3′-OH group in the polynucleotide chains (i.e. long nucleic acid chains) [4].

Additionally, unlike DNA the RNA molecule is able to produce a secondary and tertiary structure. Due to base pairing (hydrogen bonds) either intramolecular (within single molecules of RNA) or intermolecular (between different RNA molecules), the RNA molecule can form a three dimensional secondary structure [4]. Examples of these secondary forms are the helix, loop, pseudoknot, kissing loop complex and helical junction [4]. This secondary structure is important for the interaction of RNA molecules with other molecules [5]. Furthermore, folding up RNA secondary structures into a very compact shape produces a tertiary structure. The tertiary shape is maintained by hydrogen bonding and metal ions where, the latter are positively charged, interacting with the negatively charged RNA [4, 5].

1.3 RNA and DNA molecules

As stated earlier, RNA is similar to DNA in structure, however there are some fundamental differences between the two molecules, summarised in Table 1.1. One difference is that the presence of a hydroxyl group in RNA nucleotides makes RNA more prone to hydrolysis than DNA, which lacks the hydroxyl group. The stability of RNA is affected by this difference; for example under alkaline conditions DNA molecules would be stable while RNA molecules would be destroyed [6]. Another difference is that the thymine (T) base found in DNA strand is substituted with uracil (U) in RNA.

Nevertheless, unlike DNA, RNA molecules adopt different secondary and tertiary structures, which are mediated by intramolecular hydrogen bonding and RNA binding proteins. These different structures are critical to different

types of RNA molecules and their functions [4], which will be discussed in more detail in this Chapter.

Table 1.1: **Comparing the differences between RNA and DNA molecules.**

	RNA	**DNA**
Strand type	Single-stranded	Double-stranded
Nitrogenous bases	Made up of adenine (A), cytosine (C), guanine (G) and uracil (U)	Made up of adenine (A), cytosine (C), guanine (G) and thymine (T)
Nucleotide sugar	The ribose sugar has a 2′-OH group	The deoxyribose sugar lacks a 2′-OH group
Location	Located in cell nucleus, cytoplasm and ribosomes	Located in cell nucleus and mitochondria

1.4 RNA synthesis

RNA is synthesised through the process of transcription, which uses the activity of RNA polymerase enzymes to assemble a new chain of ribonucleotides complementary to the template DNA of interest. The synthesis of different types of RNA involves different types of RNA polymerase [4, 6, 7]. There are three steps involved in RNA synthesis: initiation, elongation and termination. During the initiation step, the RNA polymerase binds to the promoter region, located upstream of a gene of interest. Transcription factors will already have bound this promoter region in order to assist in the binding of RNA polymerases and initiate RNA transcription. A DNA helicase enzyme then unwinds the double stranded DNA, so the RNA polymerase can form an RNA strand complementary to the DNA strand in the elongation step. Once the RNA polymerase reaches the termination sequence on the DNA, the elongation process stops and the RNA strand separates from the DNA; this step is called termination. There are also some termination factors that can initiate the termination of RNA synthesis [7].

1.5 Different RNA types and their functions

All coding information held within DNA can be translated into specific protein or RNA molecules with the help of different types of RNA. The main three RNA molecules that are involved in the transcription and translation of protein-coding genes are messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). There are many other types of RNA involved in the regulation of gene expression. MicroRNA (miRNA), anti-sense RNA (asRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) all play a role in regulation of gene expression, and their key functions are summarised in Table 1.2.

Table 1.2: **RNA types and their main functions in the cell [3, 8].**

RNA type	Main function
Messenger RNA (mRNA)	Carries the transcribed genetic information to the ribosome
Transfer RNA (tRNA)	Transfers amino acids to the ribosome where translation takes place
Ribosomal RNA (rRNA)	Forms the ribosomes in association with proteins and directs the translation of mRNA
MicroRNA (miRNA)	Translation repression via degradation of mRNA
Anti-sense RNA (asRNA)	Regulation of gene expression by inhibiting translation
Small interfering RNA (siRNA)	Involved in the cleavage of mRNA
Small nuclear RNA (snRNA)	Processing the splicing of pre-mRNA (hnRNA) into mRNA in the nucleus
Small nucleolar RNA (snoRNA)	Modification of rRNA in ribosome biogenesis

1.5.1 Messenger RNA

Messenger RNA (mRNA) is transcribed by RNA polymerase II from DNA template [9]. The coding information that is copied from a DNA template is carried in mRNA in a series of triplet codes called ‘codons’ (three bases), where each codon encodes a specific amino acid to be incorporated into the resulting protein. An mRNA is therefore ‘read’ as a template by the ribosome to guide the synthesis of a polypeptide.

The structure of mRNA is illustrated in Figure 1.3. mRNA is first synthesized as ‘pre-mRNA’, but this primary transcript goes through three steps before it produces a mature mRNA [4, 6]. The first step involves the capping of the 5′ end by adding a specialised nucleotide cap for the ribosomal recognition of mRNA in protein synthesis. The 5′ cap is formed by adding a 7-methylguanosine cap to the first transcribed nucleotide via the action of three capping enzymes [10]. The 5′ cap structure protects the mRNA from the activity of exonuclease enzymes and is believed to be involved in other functions such as promoting transcription and splicing [10].

The second step in the production of mature mRNA is the addition of a poly(A) tail to the new 3′ end after cleaving the precursor mRNA (hnRNA), in a process called polyadenylation. It is thought that both cleavage and polyadenylation processes are coupled [11]. There are two elements with specific sequences in the pre-mRNA molecule that defines the site for cleavage: the AAUAAA and U-rich sequences, which are found upstream and downstream of the cleaved site respectively [11, 12]. Each element is bound by specific cleavage factors to mediate endonucleolytic cleavage [11] followed by the poly(A) polymerase, which generates the poly(A) tail by adding a chain of adenine nucleotides at the 3′ end. The functions of the poly(A) tail are to transport the mRNA to the cytoplasm and maintain its stability [13, 14].

The final step in mRNA maturation is splicing, where introns from the pre-mRNA are removed and exons are linked together. The spliceosome complex,

which is composed of small RNA molecules and proteins, recognises the splice site at the end of each intron to initiate splicing [15]. In the intron sequence, there is a specific adenine nucleotide that interacts with the sugar-phosphate backbone at the 5′ end of the exon to create a loop, leaving a 3′-OH free at the end of the exon, which will join the start of the next exon sequence [15]. The mature mRNA will then be transported to the ribosomes in the cytoplasm by binding to export receptors, which facilitate exportation through the nuclear pores [16].

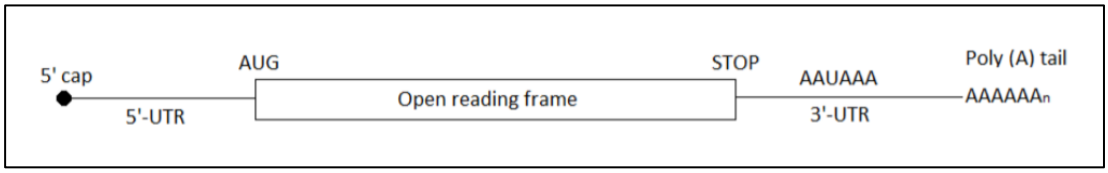


Figure 1.3: **The structure of mRNA showing the 5′ cap, the poly(A) tail, the 5′ and 3′ untranslated transcribed regions (UTRs).** The UTRs flank the coding sequence, indicated by the start codon AUG and one of three stop codons (UAA, UAG or UGA) [4].

1.5.2 Transfer RNA

Transfer RNA (tRNA) is a type of RNA exhibiting primary, secondary and tertiary structure, specifically three hairpin loops that form an ‘L’-shaped structure. It is transcribed by RNA polymerase III and its function is to carry the amino acids to the ribosome [9]. Each tRNA contains a three-base anticodon region (triplet sequence) that codes for a specific amino acid. During translation, tRNA hybridizes to mRNA via this anticodon, which is complementary to a codon on the mRNA at the ribosome site where protein synthesis takes place. Each of the triplet sequences designates the addition of a specific amino acid. The enzyme aminoacyl-tRNA synthetase is responsible for ensuring the correct amino acid is attached to the tRNA [17].

1.5.3 Ribosomal RNA

Ribosomal RNA (rRNA) is a component of the ribosome, the site for protein synthesis in cells. RNA polymerase I transcribes the large species of rRNA

including *28S* and *18S*, while RNA polymerase III transcribes the small species of rRNA including *5S* [9]. The ribosome consists of two subunits (large and small subunits), comprised of both rRNA and proteins (ribosomal proteins). The mRNA docks into the ribosome and moves along it, where it can be 'read' or translated to facilitate the assembly of polypeptide chains. Due to the catalytic properties of rRNA (i.e. not the ribosomal proteins) it forms a peptide bond between peptidyl-tRNA and aminocyl-tRNA [18], and hence ribosomes can produce a polypeptide of amino acids from the coding region in the translation step [8].

1.5.4 MicroRNA

MicroRNAs (miRNA) are small non-coding RNA molecules, 18-24 nucleotides long, derived from hairpin-shaped double stranded RNA (dsRNA) [19]. As illustrated in Figure 1.4, miRNA is transcribed by RNA polymerase II to produce pri-miRNA, which is processed by an enzyme called Drosha into small pre-miRNA (~70 nucleotides), with the ability to form stem loop structures (hairpins) [20]. Pre-miRNA is transported to the cytoplasm by Exportin-5, where it is cleaved again by an enzyme called Dicer to generate mature miRNA (double stranded), and becomes incorporated into a miRNA-induced silencing complex (miRISC) through the interaction with Argonaute proteins [4, 20-22]. The sequence of the specific miRNA in the miRISC complex acts as a guide strand that indicates the target mRNAs to interact with, while the function of the proteins in the miRISC complex is to degrade or silence the target mRNAs [22] .

Despite the fact that only around 2000 miRNAs have been discovered in humans, each miRNA has been found to be responsible for regulating the expression of hundreds of target genes [23]. MicroRNAs therefore play an essential role in regulating many biological processes [4, 24], but the process by which this is achieved is not clear. There are eight types of Argonaute protein (Ago and Piwi subfamilies) found in humans, all of which have been found to associate with miRNA and lead to gene silencing, but only Ago2 leads

to the cleavage of mRNA [25]. miRNAs regulate gene expression by recognising and binding to specific mRNA sequences within the 3′ UTR region, in order to control their degradation or inhibit their translation [5, 26]. The decision whether to degrade or inhibit depends on the complementarity between miRNA and mRNA strands; when perfect or near perfect it will result in degradation, but if partial then it many only lead to translation inhibition [27]. Different mechanisms have been suggested for how miRNA controls gene regulation, including miRNA repression at translation initiation and miRNA translation repression at post-initiation, where both are dependent upon the promoter of the target gene [28].

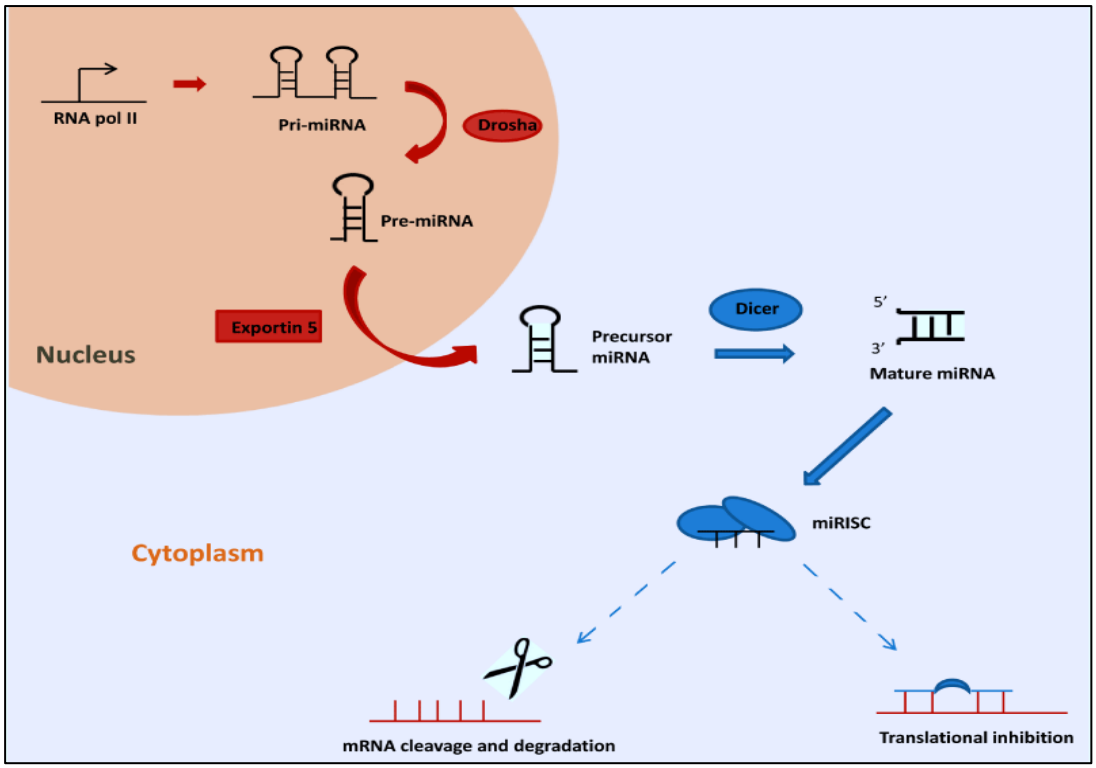


Figure 1.4: **The biogenesis of microRNA.** The RNA polymerase II transcribes the pri-miRNA, which then is cleaved by the Drosha protein to produce pre-miRNA. The pre-miRNA is then exported to the cytoplasm by Exportin-5, where it will be cleaved again by Dicer to produce mature miRNA. The mature miRNA is incorporated into a miRISC complex [22].

1.5.5 Small interfering RNA

Small interfering RNA (siRNA) is another class of small non-coding RNA molecules that is derived from long dsRNA [19]. Similar to miRNA, siRNA inhibits gene expression by silencing the translation of mRNA through binding to specific sequences on the mRNA molecule [29, 30].

1.5.6 Small nuclear RNA

Small nuclear RNA (snRNA) is transcribed by RNA polymerase III [9]. These RNA molecules play an important role in RNA splicing, for example splicing the pre-mRNA into mature mRNA [5].

1.5.7 Small nucleolar RNA

Small nucleolar RNA (snoRNA) is located in the nucleolus, and is 60-300 nucleotides in length [31]. snoRNA molecules are involved in the modification of RNA molecules, for example during the biogenesis of the ribosome in the nucleolus [31].

1.6 RNA stability and degradation

The degradation of RNA is a crucial process and serves as a powerful way to regulate gene expression and protein synthesis [32]. Different RNAs exhibit a wide spectrum of stability levels, varying from several days to a few minutes. This property is very important to cells when they need to be able to express genes quickly by switching them ‘on’ to translate essential proteins in response to different stimuli such as developmental or environmental stress, and switch genes ‘off’ very quickly when a particular protein is not needed anymore. Each type of RNA molecule has its own rate of degradation, which can be estimated by determining the half-life – the period of time that it takes an RNA molecule’s concentration to drop by half from its initial value [4] (Table 1.3). Therefore, the amount of the protein produced from an mRNA molecule often reflects the

stability of that mRNA [33]. In mammalian cells, mRNAs are longer lived with an average half-life of several hours when compared to bacterial mRNAs, which have an average half-life of less than few minutes [32]. A good example is cytoskeletal proteins, the mRNAs for which have a long half-life (up to 10 hours) and are synthesized in large quantities due to their role in the cell cycle [34].

Table 1.3: **The average half-life for different types of RNA.** Each RNA type has a specific half-life depending on its function [35].

RNA type	Half-life
mRNA	Minutes to days
tRNA	Days
rRNA	Days
miRNA	Days to weeks
asRNA	Days to weeks
siRNA	Days to weeks
snRNA	Days

This project is principally focused on the analysis of two types of RNA molecules, mRNA and miRNA, and their application in forensic science. Therefore, the stability and degradation of these two RNA types are discussed in more detail below.

1.6.1 The degradation of mRNA

There are elements within the sequence of each RNA molecule type that determine its half-life [35]. For example, mRNA has unique structures that contribute to the regulation of its stability: the 5′ cap, the poly(A) tail at the 3′ end, the 5′ UTR, the 3′ UTR and specific sequences in the coding and non-coding regions [4, 33, 36], all of which are shown in Figure 1.3 above. There

are also sequence elements known to play a role in regulating mRNA stability. For instance, the AU-rich elements (ARE), which are present in the 3′ UTR [4] are highly abundant in short-lived mRNAs, and often make these molecules targets for rapid degradation [35]. In the closed loop conformation of mRNA, the 5′ cap interacts with the poly(A) tail at the 3′ end. In this conformation, the 5′ cap interacts with the EIF4E protein, the poly(A) tail interacts with poly(A) binding proteins (PABP), and the interaction of these structures stimulates translation and keeps both ends of the mRNA protected against degradation [4, 37].

There have been many different pathways proposed as to how mRNA is degraded. The essential steps involved in the degradation of mRNA are summarized in Figure 1.5, and include deadenylation, decapping and the activity of exo- and endonuclease enzymes [33]. Deadenylation initiates at the 3′ end of mRNA by the attack of polyadenylate ribonuclease (PARN) to shorten the poly(A) tail [4]. Given that the poly(A) tail interacts with both the 3′ end and the 5′ cap, the deadenylation of the poly(A) tail exposes the 5′ cap for removal, hence the RNA will be unable to retain the closed loop conformation [33]. Decapping is performed by the decapping complex (DCP1 and DCP2), which catalyses the removal of the 5′ cap structure [4, 33]. When the 5′ cap structure is removed the mRNA becomes a target for the exonuclease XRN1 to digest and cleave one nucleotide after another in a 5′ $\rightarrow$ 3′ direction, or the exosome (an exoribonuclease complex), which digests mRNA in the 3′ $\rightarrow$ 5′ direction [4, 33].

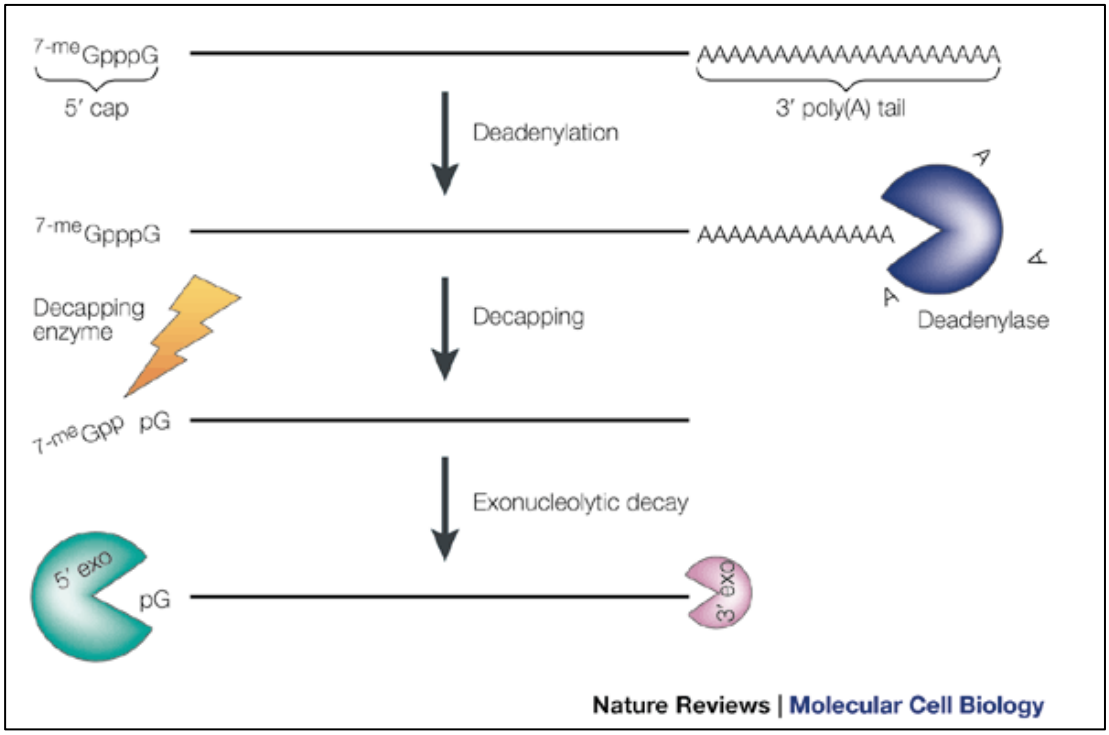


Figure 1.5: **The pathway of mRNA decay.** The poly(A) tail at the 3′ end of the mRNA is shortened by the activity of deadenylase. Once deadenylation is complete the 5′ cap will be removed, and then exonucleases will digest the mRNA for degradation [38].

1.6.2 The degradation of miRNA

As stated above, all RNA types have half-lives – the period of time that it takes a molecule's concentration to drop by half from its initial value, which can be used as an estimate of its rate of degradation. miRNAs follow the same principle; however, the stability and degradation mechanisms of human miRNA remains largely unknown. Due to the important functions of miRNA and their association with different diseases, such as cancer [39] and cardiovascular diseases [40], it is crucial to regulate the expression level of each miRNA at specific times and in response to cellular conditions. It is believed that the miRISC complex and its components (especially Argonaute proteins and several sequences in miRNA molecules themselves) are responsible for regulating the stability of miRNA, protecting it against degradation [41, 42]. However, there are some exoribonucleases that have been identified as miRNA-degrading enzymes [43], which are involved in physiological miRNA turnover.

It was reported that in *Arabidopsis thaliana* and *Caenorhabditis elegans*, the degradation of miRNAs was mediated by the mechanism of 3′ → 5′ and 5′ → 3′ exoribonucleases respectively [43, 44]. The small RNA degrading nucleases (SDNs) mediate miRNA turnover in *Arabidopsis thaliana* and possess 3′ → 5′ exonuclease activity that acts on miRNAs [44]. Additionally, uridylation of the 3′ end of miRNA contributed to the decay of miRNA molecules in *Chlamydomonas reinhardtii* [43]. In contrast, in humans the addition of uridines has been shown not to affect the degradation rate of the miRNA, but did inhibit its activity [41].

Li et al. (2013) attempted to study the stability of a number of human miRNAs and revealed some factors that might affect miRNA stability [45]. They performed genome-wide microarray analysis on 10 cell lines that covered a range of different tissue types (lung, stomach, embryonic kidney, blastocyst, prostate, muscle, breast, embryonic brain, pancreas and brain). Transcription in all cell lines was inhibited for three hours by adding actinomycin D to block the activity of RNA polymerase. The level of RNA was measured at time points 0 h and 3 h and were normalized against *U6*. Their results showed a distribution of miRNA stability, where different cell types had different miRNA stability. It was not possible to identify any clear reason why this diversity was observed but it was suggested that it could be related to the different cell functions. In addition, a correlation analysis was performed to study the relationship between miRNA stability and expression level. This revealed a negative correlation, which they called “rapid production, rapid turnover, slow production, slow turnover”, in other words that more highly expressed miRNAs will tend to turnover faster than the less expressed miRNAs, which remain more stable and degrade slowly. These findings draw a different conclusion from previous research examining the relationship between proteins [46] and mRNA [47], where a positive correlation has been found.

More recently, research have been done to understand the degradation dynamics of miRNAs [48-50], and suggested that miRNA decay could be determined two mechanisms. The first mechanism is the competing endogenous RNA (ceRNA), where endogenous RNA (coding or noncoding) targets compete with the real RNA targets and bind to miRNA, therefore, the activity of miRNA is influenced by the change in the expression of the competing RNA targets [50]. The second mechanism is the target-directed miRNA degradation (TDMD), where the interaction of miRNAs with their mRNA targets promotes miRNA degradation. Post-transcriptional modifications of miRNAs are proposed to be accompanied with TDMD, such as adding or removing nucleotides to modify miRNAs ends, leading to the separation of Argonaute proteins from miRNA and hence cause degradation of the latter [48, 49]. Despite this, the details of miRNA decay mechanisms remain obscure.

1.7 The applications of RNA in forensic science

Recently, there has been increasing interest in the use of RNA in the forensic science community, for various applications. Some RNA types such as mRNA have demonstrated the ability to reveal the activities of genes and their respective cells/tissues, which might help to give an indication of pathological states [51] or any condition that leads to death [52]. In addition, RNA can be used to identify the origin of a body fluid [35, 53-57] and give some indication of the time it was deposited [58-60]. Furthermore, the mechanism of mRNA degradation may also help determining the time of death [61]. RNA first appeared in forensic research in 1984, when its synthesis in post-mortem tissues was described [62]. Since then, many more studies have been carried out to try and understand how patterns of gene expression can be useful in forensic science.

1.7.1 Identification of body fluids

Identification of the types of biological stains recovered from crime scenes can be very important to criminal investigations, i.e. whether they originate from blood, saliva, semen, etc. This kind of information may help reconstruct events that occurred at a crime scene. Current methods used in forensic laboratories for body fluid identification depend on catalytic or enzymatic tests, to identify proteins or compounds present in one specific body fluid. For example, the Kastle-Meyer presumptive test for blood indicates the presence of haemoglobin. Some of these methods lack specificity and sensitivity, and may also destroy precious samples, precluding the ability to perform subsequent DNA profiling [63]. As such, RNA is of increasing interest in forensic science as a novel means to identify body fluids.

1.7.1.1 Application of mRNA to body fluid identification

It has been shown that the identification of cell-specific mRNAs can provide high specificity for body fluid identification, due to different gene expression patterns in different tissue types [53, 64-66]. This follows the principle that cells will switch ‘on’ genes encoding proteins essential for their functional role, with these genes being transcribed into RNA. As a result, each body fluid has its own specific gene expression pattern that can be defined by the presence of mRNAs that encode for proteins with functions specific to that body fluid. A number of recently published papers have shown that RNA can be used in detecting and identifying biological stains and body fluids [55-57]. Table 1.4 shows a selection of some markers that have been evaluated using reverse transcription endpoint polymerase chain reaction technology (RT-PCR) and reverse transcription quantitative real-time polymerase chain reaction technology (RT-qPCR) to identify different types of body fluids.

Table 1.4: **Different molecular markers used for the identification of specific body fluids [53, 64, 65].**

Body fluid type	mRNA markers
Blood	ß-Spectrin (*SPTB*) Hydroxymethylbilane Synthase (*HMBS*), which is also known as PBGD Haemoglobin Subunit Alpha (*HBA*) Haemoglobin Subunit Beta (*HBB*)
Menstrual blood	Matrix metalloproteinase 7 (*MMP*-7) Matrix metalloproteinase 11 (*MMP*-11)
Saliva	Statherin (*STATH*) Histatin (*HTN3*)
Semen	Protamine 1 (*PRM1*) Protamine 2 (*PRM2*) Kallikrein 3/prostate specific antigen (*KLK3*/*PSA*) Semenogelin-1 (SEMG1)
Vaginal secretions	Human beta-defensin 1 (*HBD*-1) Mucin 4 (*MUC4*)

Both the sensitivity and specificity of these mRNA methods varies depending on the body fluid type that is examined, and the RNA markers targeted. When the sensitivity of mRNA methods was compared to some of the current presumptive tests, comparable results were seen, and higher sensitivity was shown to depend on the choice of markers [64]. With regards to specificity, mRNA markers for some body fluid types such as blood, saliva and semen have been demonstrated to have good specificity with few cross-reactive results in non-target body fluids [65, 67], whereas less specificity has been observed for vaginal secretion samples [68].

The potential application of RNA in forensic science was ignored for many years under a false perception that RNA is extremely unstable, and subject to rapid degradation. Despite the effect of ribonucleases (RNases) on mRNA

stability, Zubakov et al. [69] were able to identify nine stable mRNA markers for bloodstain identification and five for saliva identification. They applied whole-genome gene expression analysis to degraded blood and saliva stains aged for up to 180 days and found that mRNA can remain stable for long periods of time in dried blood and saliva stains, due to reduced RNase activity. Some of these markers were even successfully amplified in 16-year-old bloodstains [54]. Setzer et al. (2008) examined the influence of environmental factors such as temperature, UV/visible light and weather conditions such as rain on the recovery rate of mRNA in different types of biological stains [70]. They were able to detect mRNA from saliva, blood and semen after 1, 3 and 7 days respectively, after they were stored outside and exposed to heat, light, humidity and rain. Even though unfavourable environmental conditions had a negative impact on mRNA recovery rate, mRNA markers were still detected after a few days. Overall, research has therefore shed a positive light on the use of RNA to identify aged stains.

One major advantage of using mRNA for body fluid identification in the forensic field is the ability to extract both DNA and mRNA simultaneously from the same stain, which is very useful especially where there is limited material in the sample. Another key advantage for using mRNA profiling for body fluid identification is the ability to create a multiplex RT-qPCR assay to provide information about the presence/absence of RNAs associated with many body fluids in one reaction, saving time and reducing sample consumption [64, 68, 71, 72]. Identifying a panel of different body fluids in one reaction can be extremely beneficial for forensic analysis, as many crime scene stains comprise a mixture of cell types, often from different persons.

However, when it comes to crime scene stains, body fluid identification is challenging, primarily because crime scene stains will degrade over time and may give false negative results [35]. Bauer et al. (2008) have suggested a solution to distinguish true and false negative results in the identification of menstrual blood using quantitative analysis [73]. Using the gene *GAPDH* as a

reference gene, they found that all samples with positive results for menstrual blood markers had a relative expression ratio ranged between 0.125 and 250 to the reference gene. The ratio of 250 was found to be at quantification cycle (C_q) 31 for *GAPDH*. Therefore, under their study conditions, they defined a cut-off value for the *GAPDH* marker and any samples with negative results for menstrual blood marker and a *GAPDH* C_q below 31, would be truly negative results. They were successfully able to identify only 2 out of 6 samples to be truly false negative and the rest were inconclusive. However, forensic casework samples collected from crime scenes are often very small in quantity and may also be degraded, so it would be unrealistic to depend on the threshold of reference genes, which could also be degraded and provide inaccurate data. Another way to overcome false negative/positive results in body fluid identification is by using multiple markers for each body fluid simultaneously. This can be done through developing a multiplex system, as it is less likely that all the selected markers per body fluid type would indicate the presence of specific body fluid falsely [74].

1.7.1.2 Application of miRNA to body fluid identification

In a small number of published gene expression association studies, it has been found that similar to mRNA, a number of miRNAs exhibit expression restricted to one cell type [21], making them useful markers for body fluid identification. Hanson et al. [66] were the first group to explore the role of miRNA in the identification of body fluids for forensic applications, in 2009. They examined five different body fluid stains: blood, saliva, semen, vaginal secretions and menstrual blood. By comparing the expression level of different miRNA markers to the reference gene *U6b* in each body fluid, they were able to create an assay of nine potential miRNA markers to indicate the presence of each type of body fluid: blood (*miR16* and *miR451*), saliva (*miR205* and *miR658*), semen (*miR10b* and *miR135*), vaginal secretions (*miR124a* and *miR372*) and menstrual blood (*miR412*).

The identified markers *miR16*, *miR451*, *miR205* and *miR658* showed higher expression levels than *U6b* in blood and saliva samples respectively, and lower expression levels than *U6b* in other body fluids. The identified markers for semen (*miR10b* and *miR135*) showed relatively higher expression levels in semen samples than other body fluids, but were not higher than *U6b*. For vaginal secretions and menstrual blood, the markers showed a relatively higher expression level when compared to other body fluids, but only the level of *miR124a* was higher than *U6b*.

To evaluate the specificity of their miRNA selected panel, they examined the expression profiles of these miRNAs in 21 different human tissues. All miRNA assays exhibited different expression profiles (i.e. lower expression) in tissues when compared to body fluids, confirming the high degree of specificity of the selected potential markers. One limitation of this study is that they used only one reference gene (*U6b*), which may affect normalisation and lead to inaccurate results, and recent work has shown that it is more reliable to use several reference genes for accurate normalisation [75]. However, the findings of this study shed light on the potential use of miRNA analysis in body fluid identification.

In 2010 and 2013, the findings of Hanson et al. (2009) [66] were replicated by Zubakov et al. (2010) [76] and Bai et al. (2013) [77] respectively. Zubakov et al. (2010) identified the same markers for venous blood (*miR16* and *miR451*) and semen (*miR10b* and *miR135*) but not for other body fluid types. While Bai et al. (2013) were able to identify the same markers for blood (*miR16* and *miR451*), saliva (*miR205* and *miR658*), and vaginal secretions (*miR124a* and *miR372*) as Hanson et al. (2009), they also identified a new miRNA marker for semen (*miR135b*). In 2012, Wang et al. [78] further examined three miRNA markers identified by Hanson et al. (2009): *miR16* for blood, and *miR658* and *miR205* for saliva. Their findings partially supported the results of Hanson et al. (2009), that *miR16* is specific for blood, but not for *miR658* and *miR205*, which had been indicated to be specific for saliva. They suggested that this

discrepancy could be a result of using different statistical analysis methods in the two studies. According to Wang et al. (2012) [78], the application of an efficiency-calibrated model in data analysis to determine the relative expression ratio (RER) of miRNAs, by incorporating the impact of C_q (quantification cycle) values and PCR efficiencies of miRNAs and reference genes, can affect accurate determination of RER of miRNAs in body fluid identification.

Additionally, Zubakov et al. (2010) [76] performed a larger screen of 718 miRNAs using microarray analysis. 14 candidate miRNA markers exhibiting body fluid-specific expression were selected based on the intensity of hybridisation signals across all body fluids and individual samples, as well as fold-changes of gene expression. These selected markers were then validated by RT-qPCR using TaqMan® Assays (Applied Biosystems) to confirm the results of the microarray analysis. The study was only able to identify two miRNA markers associated with blood (*miR144* and *miR185*) and two miRNA markers for semen (*miR135a* and *miR891a*), but could not identify any markers for saliva, vaginal secretions and menstrual blood. Their results showed no overlap with the putative miRNA markers that were identified by Hanson et al. (2009) [66]. It was suggested that these discrepancies could be due to the small sample size of only six volunteers, so expression variation between individuals could not excluded. Another potential reason is the different technology platforms used by the two groups. Due to the specificity and sensitivity of TaqMan® (used in Zubakov et al. project) and SYBR Green (used in Hanson et al. project), some miRNA markers did show statistically significant differences between these two quantification methods [79]. However, Zubakov et al.'s microarray analysis showed high expression of *miR891a* in a vaginal secretion sample, which contradicted with other literature showing that *miR891a* is expressed only in the epididymis tissue where the sperm cell matures [80]. They validated their microarray analysis by RT-qPCR, where a high expression profile for *miR891a* was found in semen samples. Furthermore, they applied Northern blot analysis to semen and vaginal

secretion samples using LNA^{TM}-modified oligo-nucleotide probes for *miR891a* in order to determine the reason behind the error in the microarray analysis. Similar results to the RT-qPCR analysis were found, however they detected cross-hybridisation between the probes and a non-miRNA, which may be the cause of the error. They raised the issue that LNA^{TM}-modified oligo-nucleotide probes are not human specific and may hybridise with bacterial and fungal RNAs or precursor miRNA. This could cause errors in the microarray analysis and lead to the inability to identify any candidate markers for saliva, vaginal secretions and menstrual blood.

Another study conducted by Courts and Madea (2011) [81] used microarray analysis on two types of body fluid, blood and saliva, performing a global screening of 800 miRNAs to identify those with a differential expression pattern between the two. They were able to differentiate between miRNAs based on their expression level and selected three candidates that showed very high expression in each body fluid. These candidate markers were then evaluated using RT-qPCR. They were successfully able to identify three potential miRNA markers for the identification of blood (*miR451*, *miR150* and *miR126*) and three markers for saliva (*miR205*, *miR203* and *miR200c*). One of the miRNA markers that indicates the presence of blood (*miR451*) and one that indicates saliva (*miR205*) were the same as those identified by Hanson (2009).

Furthermore, Wang et al. (2013) [82] utilized microarrays and RT-PCR methodologies to find suitable miRNA markers for body fluid identification. A total of five miRNA markers were identified, two for blood (*miR16* and *miR486*), two for semen (*miR888* and *miR891a*), and one for menstrual blood (*miR214*), which included three new miRNA markers. They were also able to identify miRNA markers that were supported by previous studies; *miR16* for blood and *miR124a* for vaginal secretions, also identified by Hanson et al. (2009), and *miR891a* for semen identified by Zubakov et al. (2010). In their conclusion, they emphasised the need for more studies to find suitable reference genes for forensic body fluid identification, as their data revealed no reference genes

for miRNA that have stable expression across body fluids, and the success of RT-qPCR analysis depends on proper normalisation of data using a reference gene. More recent work by Wang et al. (2015) [83] was conducted in order to identify miRNA markers that are specific for saliva. Using the same methodologies applied in their previous work [82], they analysed eight potential miRNAs selected from the literature: *miR200c-3p*, *miR203a*, *miR2055p*, *miR658*, *miR138*-2, *miR146b*-3p, *miR206*, and *miR639*. Again, they could not identify any miRNA markers that indicate the presence of saliva, but suggested that a combination of two or three miRNA markers could be used to discriminate saliva from other body fluid types.

Finally, two recent publications have applied next generation sequencing (NGS) or massively parallel sequencing (MPS) to identify miRNA markers for body fluid identification. Wang et al. (2016) utilised MPS using Ion Torrent PGM™ technology, and were able to identify blood-specific miRNA markers that have been reported previously (*miR486*, *miR16*, *miR451a*, *miR144* and *miR126*), confirming their specificity to blood [84]. They also identified 19 miRNA markers for saliva, three of which have been reported previously (*miR2031*, *miR205*, and *miR200c*) and 16 of which were novel markers. The Illumina HiSeq 4000 was used in the project reported by El-Mogy et al. (2018), where they successfully identified the 20 most abundant miRNA markers in blood and saliva samples [85]. In their work, they have identified number of miRNA markers that were identified in previous studies, three for blood (*miR486*, *miR451a* and *miR16*), and two for saliva (*miR203a* and *miR205*), with the addition of 17 novel markers for blood and 18 novel markers for saliva.

Table 1.5 summarizes the main findings of the previous studies reviewed above, showing the similarities and differences with regards to which miRNAs have been identified as exhibiting body fluid-specific expression. These findings highlight the need for more studies in body fluid identification using miRNAs, in order to identify potential markers that can be applied in forensic casework.

Table 1.5: **Summary of miRNA markers identified by different studies for body fluid identification.**

Publication	Method	Reference Genes	Blood	Saliva	Semen	Vaginal Secretions
Hanson et al. (2009) [66]	SYBR Green	*U6b* *RNU44*	*miR16* *miR451*	*miR205* *miR658*	*miR10b* *miR135*	*miR124a* *miR372*
Zubakov et al. (2010) [76]	TaqMan®	*RNU24* *RNU44* *RNU48*	*miR144* *miR185*	-	*miR135a* *miR891a*	-
Courts et al. (2011) [81]	SYBR Green	*U6b*	*miR126* *miR150* *miR451*	*miR200c* *miR203* *miR205*	-	-
Wang et al. (2013) [82]	TaqMan®	*U6*	*miR16* *miR486*	-	*miR888* *miR891a*	*miR124a*
Bai et al. (2013) [77]	SYBR Green	-	*miR16* *miR451*	*miR205* *miR658*	*miR10b* *miR135b*	*miR124a* *miR372*
Wang et al. (2016) [84]	NGS (Ion PGM™)	-	*miR486* *miR16* *miR451a* *miR144* *miR126*	*miR203a* *miR205* *miR141* *miR375* *miR200c**	-	*miR1260b* *miR654*
El-Mogy et al. (2018) [85]	NGS (Illumina HiSeq)	-	*miR486* *miR451a* *miR16* *miR185**	*miR143* *miR203a* *miR375* *miR205**	-	

* More novel miRNAs were identified.

1.7.2 Biological stain age determination

The time since deposition of a body fluid at a crime scene can be crucial for criminal investigations, as it can provide information regarding when a crime occurred. Conversely, samples that do not correspond to the time when the crime is known to have occurred may be excluded. To date, a method for ageing biological samples using RNA analysis has only been applied to bloodstains [58, 86], saliva [87] and hair [60]. Prior to this, the age of bloodstains was estimated using different methods such as variation in solubility, morphological differences in the bloodstains, and analysis of protein degradation, although these methods have some limitations, such as sample consuming, the inability to discriminate blood samples from different species and produce a wide error range when estimating the age [88].

Given that RNA is known to be less stable than DNA and gradually degrades in the environment, it has been proposed that quantifying the level of RNA degradation in a biological stain may be useful as a measure of stain age. This is an emerging area of research with, as yet, a small number of publications investigating the relationship between age and RNA decay in blood, hair and saliva. The review below considers some examples of papers that analysed RNA degradation level in an attempt to estimate the age of biological stains.

1.7.2.1 mRNA application in biological stain age estimation

Anderson et al. (2005) [58] developed a method utilising RT-qPCR to study the expression ratio between two types of RNA molecules, *18S* rRNA and *ACTB* mRNA, to estimate the age of dried bloodstains stored under controlled conditions for 150 days [58]. All bloodstains were stored in a chamber at 25 °C and 50% humidity for a period of 150 days. After the desired age was reached (30, 60, 90, 120, and 150 days), total RNA was extracted and the levels of *18S* and *ACTB* quantified using RT-qPCR. It was found that *ACTB* level reduced over time but *18S* remained stable. Taking the relative ratios of *ACTB* to *18S* using the cycle threshold (C_t) values produced, a linear relationship with

bloodstain age was detected, as shown in Figure 1.6. This finding was proposed as being explained by structural differences, where *18S* is part of the ribosome structure and is surrounded by proteins that protect it against degradation, while *ACTB* does not have such protection and so is more susceptible to degradation. One limitation in this work is that examining the samples every four weeks produces only a crude time estimate. Another, key limitation of this study is the wide time intervals, as they examined the samples every 30 days, which may not be of practical value for application to forensic casework. However, an advantage is that they examined RNA markers that are universally expressed and used as reference genes so, in theory, the same principle can be applied to different tissues or body fluids. The findings of their work have since been replicated, indicating the validity of their hypothesis [89].

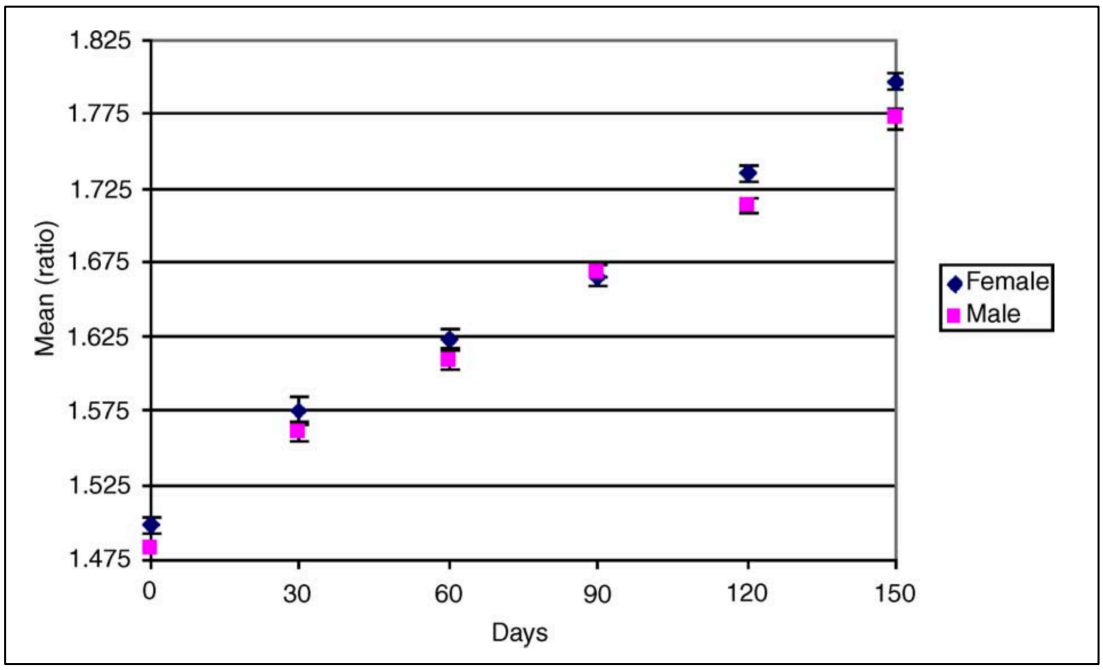


Figure 1.6: **The relationship between the relative expression ratio of *ACTB* to 18S in blood stains aged for up to 150 days [58]**.

Additionally, Anderson et al. (2011) analysed the same RNA markers to examine the relative stability over time of different-sized segments of the two RNAs in bloodstains [59]. Their principle was that a long PCR target amplicon is more susceptible to degradation than a shorter PCR amplicon, and as such,

the ratio of their expression can be used as a measure of fragmentation. Using this approach, they were able to distinguish fresh blood samples from those that were 6 days old, and the 6 days old samples from those aged for 30 days or more.

By applying the same method on hair samples, Hampson et al. (2011) [60] supported the findings of Anderson et al. (2005). They aged hair samples for up to 3 months after plucking, and studied the degradation level of *18S* rRNA and *ACTB* mRNA by extracting RNA from the hair root bulb. Their results demonstrate a linear relationship between the expression ratio of these two RNAs and age (Figure 1.7). It is worth mentioning that even the *18S* started to exhibit signs of degradation after 45 days, which could explain the observance of a plateau effect towards the end of the 3 months.

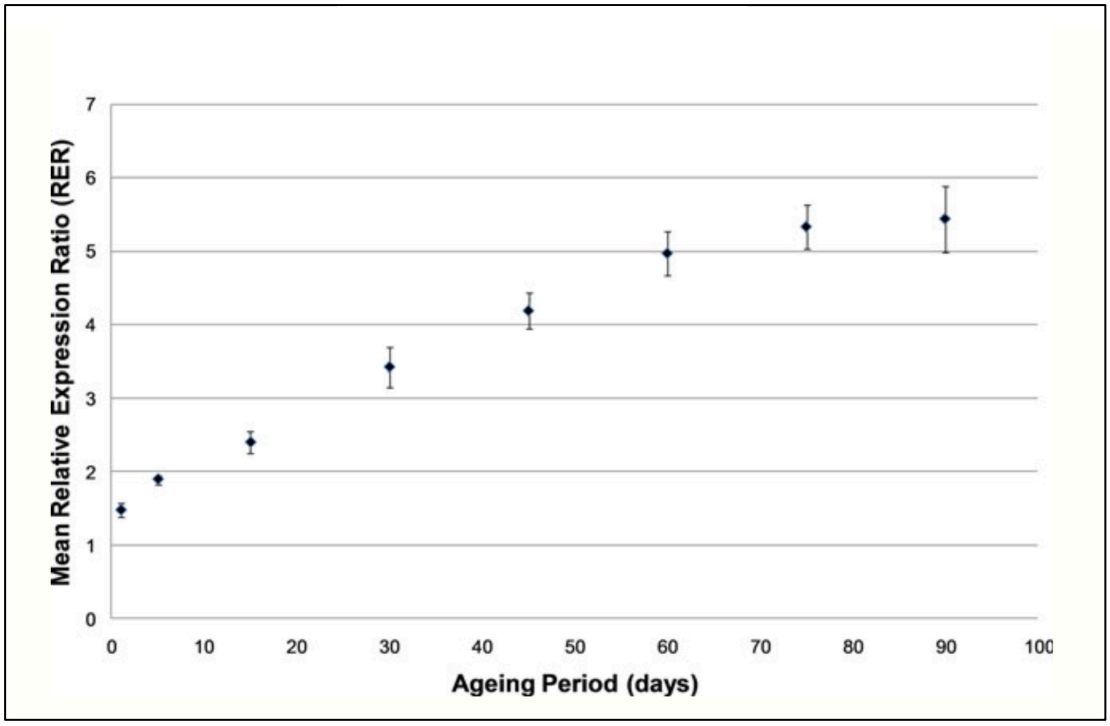


Figure 1.7: **The relationship between the relative expression ratio of *ACTB* to *18S mRNA* in hair samples aged for up to 90 days [60]**.

Furthermore, another study conducted by Simard et al. (2012) [88] showed that fresh bloodstains can be discriminated from stains aged for 14 days or more (Figure 1.8), as well as semen stains aged for 0-29 days versus 56 days, when analysing individual RNA markers. They used RT-qPCR to analyse *GAPDH* and *PPIA* (Peptidylprolyl Isomerase A) mRNAs in addition to the RNAs used in Anderson et al. [58, 59] and Hampson et al. [60], on samples of three different body fluids (blood, saliva and semen) stored at room temperature for up to six months. By calculating the quantification of each RNA marker at each time point relative to the control time point (zero days), their findings allowed them to produce degradation profiles for blood and semen samples but not for saliva, due to the low level of RNA and high RT-qPCR variation in saliva samples. Unlike the findings in Anderson et al. [58, 59] and Hampson et al. [60], Simard et al. [88] found that the degradation rate of rRNA compared to mRNA did not show any significant differences over the time points. Therefore, a correlation between the relative mRNA/rRNA ratios and storage time could not be established in their study. In addition, the same RNA markers showed a lower degradation rate (i.e. higher RNA quantity) in frozen samples over time, when stored at -80 °C for six months compared to those stored at 20 °C, with 91% and 85% of the RNA level being detected in the blood and semen respectively in samples stored at -80 °C for this time interval.

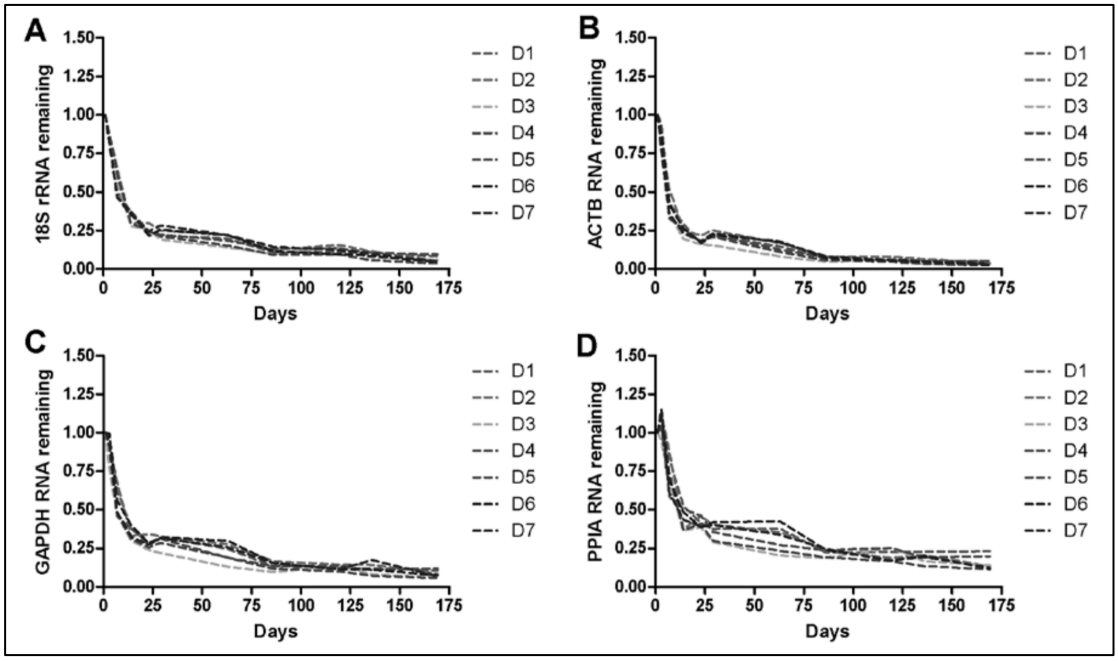


Figure 1.8: **Degradation level of four RNA molecules (A) *18S rRNA*, (B) *ACTB*, (C) *GAPDH* and (D) *PPIA* in blood samples stored at room temperature for increasing time periods up to six months [88]**. The lines represent the relative quantity of RNA markers remaining from 7 bloodstains (D1 to D7).

The findings of the above studies are promising, and future methods using RT-qPCR are likely to be more sensitive for accurate determination of the age of biological stains [53]. The current project focused on the application of body fluid-specific RNA markers to estimate the age of blood, saliva, and semen stains, and increase the accuracy of this age prediction.

1.7.2.2 miRNA application in biological stain age estimation

As discussed in the section above, a number of publications have revealed that the degradation rate of RNA can be useful for estimating the age of biological samples recovered from a crime scene [58-60]. By estimating the age of biological samples, it might be possible to identify either the time at which a crime occurred, or whether a stain is pertinent to a specific criminal investigation. As RNA is prone to degradation, if it were possible to quantify the level of RNA degradation in a biological stain this may provide an indication of the age of a stain.

Despite the fact that miRNA is very stable [90], a study by Nakao et al. (2013) indicated that some miRNA markers could be useful for estimating the age of bloodstains [91]. They examined the degradation rate of two miRNA markers indicative of the presence of blood (*miR16* and *miR451*). The blood samples that they tested were stored in a chamber at 25 °C and 50% humidity from 0 to 28 days. They found that the level of *miR16* significantly decreased 5, 7, and 21 days after incubation ($p < 0.05$), and the level of *miR451* significantly decreased on 7, and 28 days after incubation ($p < 0.05$) relative to the zero time point (Figure 1.9). It was proposed that this significant decrease might be due to the large volume of blood used (1 mL) taking longer to dry, therefore allowing more time for RNase activity. However, these findings showed that both *miR16* and *miR451* could be useful for estimating the age of bloodstains.

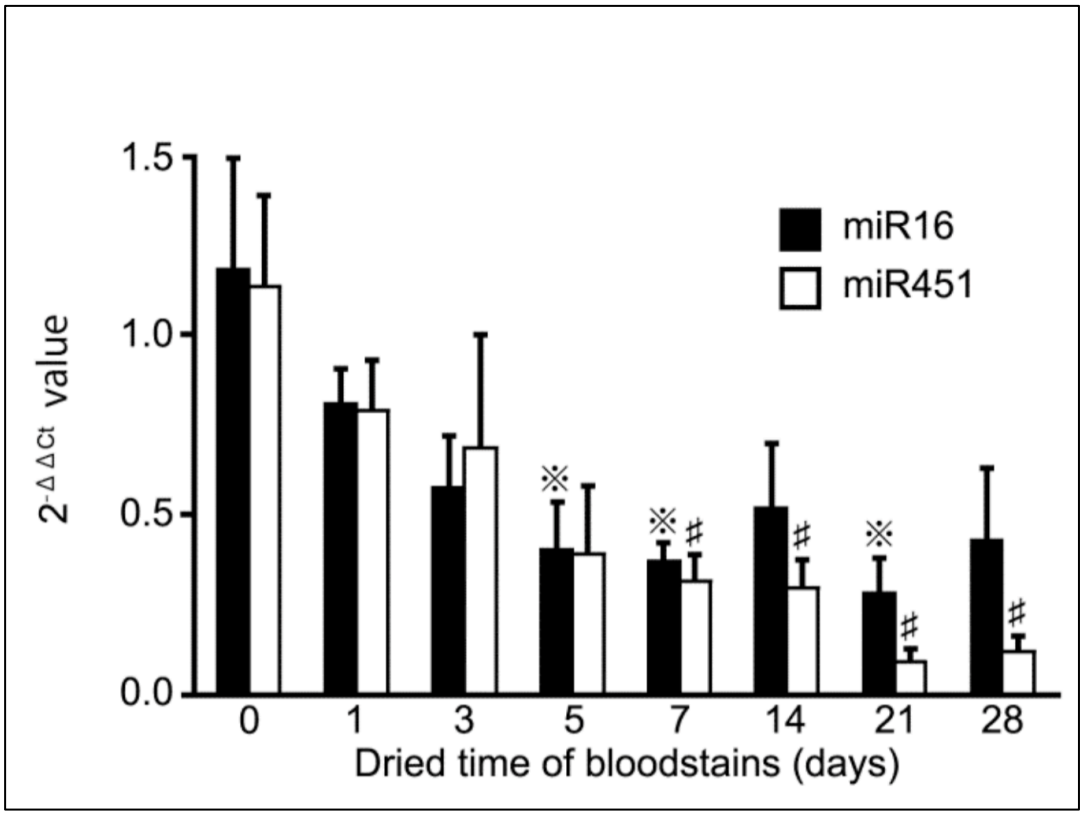


Figure 1.9: **The relative quantification of *miR16* and *miR451* markers to the initial level in bloodstains stored between 0 to 28 days [91].** * $p < 0.05$ compared with the level of *miR16* at day 0 and # $p < 0.05$ compared with the level of *miR451* at day 0.

Another interesting piece of research was conducted by Lech et al. (2014) in an attempt to determine bloodstain deposition time, in terms of whether this occurred during the day or at night [92]. Two blood miRNA markers, *miR1425p*, and *miR541,* were found to exhibit variation in their expression profiles in vitreous humor between individuals who died during the day or at night [93]. Lech et al. investigated the expression levels of these two miRNAs and whether they exhibit diurnal changes in blood. They collected blood samples every 4 h during a 24 h period under controlled wake and sleep lab conditions. Their RT-qPCR results detected no significant differences in the expression level of *miR142-5p* between day and night time. The level of *miR541* was too low in blood and was not used for further analysis because it could not provide reliable results. Therefore, they concluded that neither miRNAs was suitable for estimation of bloodstain deposition time, in terms of whether this was during the day or night.

1.7.2.2.1 MiRNA stability at different storage temperatures

Mraz et al. (2009) analysed the stability of miRNAs in clinical samples of B-lymphocytes [90]. They extracted total RNA from samples that were stored at -80 °C for between 14 days and 10 months; RT-qPCR revealed high stability of miRNAs, with no detected degradation. Ge et al. (2014) examined the stability of four miRNAs (*miR16*, *miR24*, *miR451*, and *miR181a*) by RT-qPCR in exosomes (cell derived vesicles secreted by most cell types) and plasma at three different storage temperatures (4 °C, -20 °C, and -80 °C) for 2 weeks, 2 months, 3 years, and 5 years [94]. At 4 °C the level of *miR24*, *miR451*, and *miR181a* decreased significantly over time compared to other with other storage temperatures, but not *miR16*. Significant degradation was also recorded in long term storage samples at -20 °C for 5 years [94]. Applying a similar approach to human urine samples, Mall et al. (2013) analysed the stability of *miR16* and *miR21* in urine samples at different temperatures (4 °C, and room temperature 25 °C) for up to 5 days [95]. The measured level of both markers decreased on average to 35% of the initial amount after 5 days and

at room temperature storage. At 4 °C, the measured level of *miR16* and *miR21* decreased by 42% and 56% respectively after 5 days of storage at 4 °C. Their findings show that miRNA stability is affected by storage temperature.

The above studies have been conducted from a clinical perspective, looking at how storing samples at different temperatures may affect RNA degradation in clinical samples, but none have been conducted from a forensic perspective. Forensic samples are susceptible to many different environmental conditions, such as high temperature, humidity and rain, which may have a significant impact on RNA degradation. Studies therefore need to be performed in order to examine the effect of these different conditions on miRNA level.

1.7.2.2.2 MiRNA stability in body fluids

With regards to body fluid identification, Zubakov et al. (2010) explored the stability of previously identified miRNA markers for blood (*miR144*, *miR106a* and *miR185*) and semen (*miR10a*, *miR507*, *miR135a* and *miR891a*) [76]. Applying TaqMan® RT-qPCR assays, they determined the level of miRNAs in aged samples that were stored for up to one year at room temperature and under constant humidity. It was found that the absolute expression level of the miRNA markers in aged samples did not decrease, which revealed robust stability in these miRNA markers. Figure 1.10 shows the expression level of one of the tested miRNA marker (*miR891a*) in fresh and aged samples, showing that the C_q of *miR891a* remained stable after one year. The high stability of *miR891a* could be due to being highly over expressed in seminal fluid.

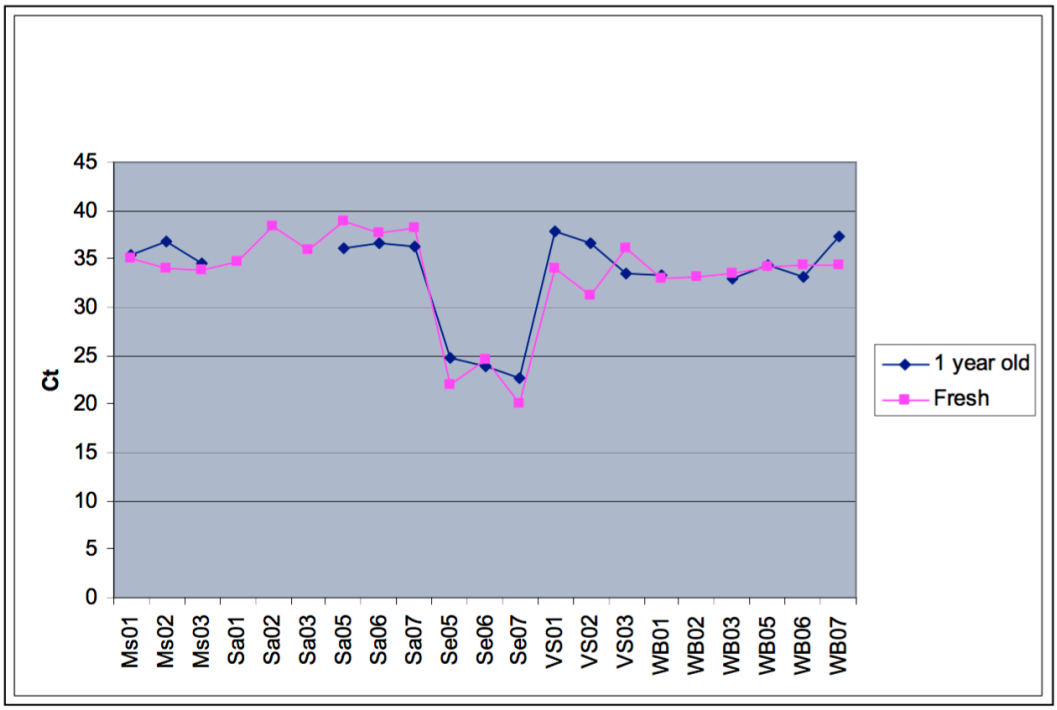


Figure 1.10: **Expression level of *miR891a* in fresh and 1-year old blood (WB), saliva (Sa), semen (Se), and vaginal secretion (VS) samples** [76].

Wang et al. (2013) briefly explored the stability of miRNA markers that they identified for blood (*miR16* and *miR486*), semen (*miR888* and *miR891a*) and menstrual blood (*miR214*) [82]. The samples were stored in the lab at ~15 °C and exposed daily to 10 hours of natural daylight for one month. Using TaqMan® RT-qPCR analysis, a slight decrease was observed in absolute expression levels across the month, reflected in a small increase in C_q values, however this was not statistically significant and the ΔC_q (C_q of the target miRNA – C_q of the reference gene) values remained stable, showing that these markers degraded at similar rates (see Figure 1.11).

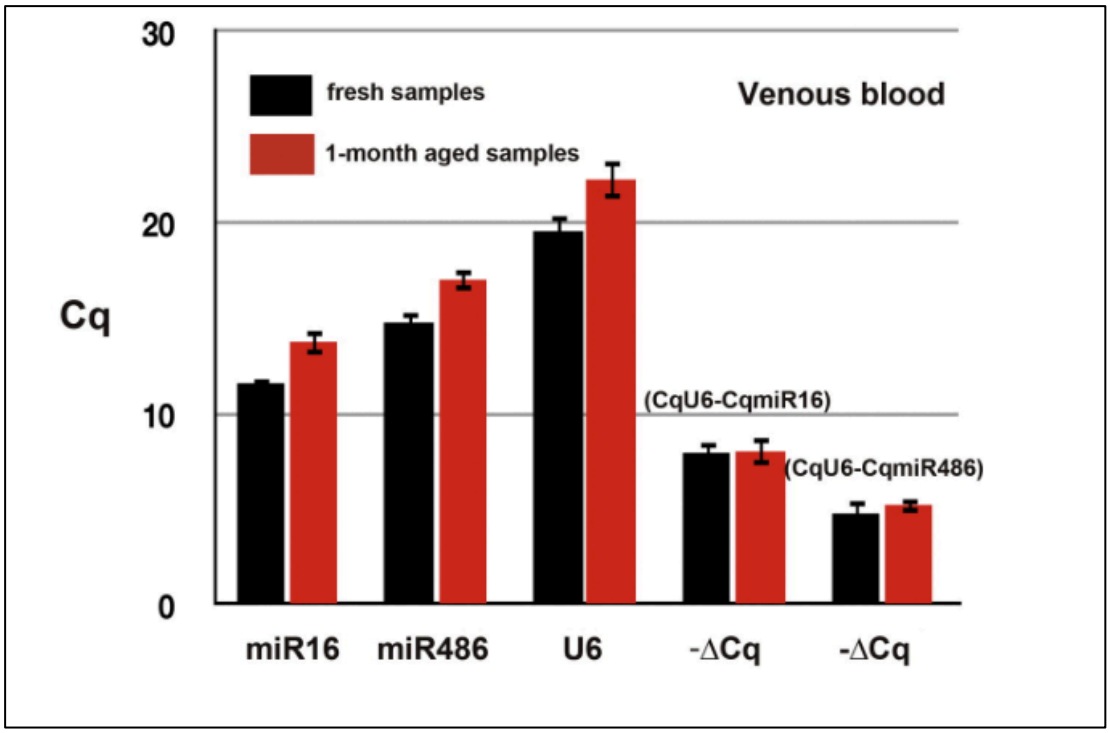


Figure 1.11: **Expression level of miRNAs tested by Wang et al. (2013) (*miR16*, *miR486* and *U6*) in fresh and aged blood samples [82].** Data shows the raw C_q values for fresh bloodstains (in black) and those aged for 1 month (in red), and also as a ΔC_q, normalised to the reference gene (*U6*).

Both of these studies did not examine the effect of any other environmental factors, such as high temperature and high humidity, and did not explore longer storage times. However, a very recent study has investigated the stability of a number of miRNA markers, including *miR451* (a blood marker) and *miR891a* (a semen marker) in samples that were exposed to different environmental conditions (heat, humidity, and sunlight) [96]. They found that miRNA markers were stable and persistent under the different conditions tested, and were still detected after 180 days of storage. Further work is needed on the stability of miRNAs in the conditions encountered in forensic casework

1.8 RNA methodologies

The workflow of RNA analysis starts with the extraction of total RNA from a biological sample. Following this, the concentration of RNA recovered can be quantified, and quality assessment performed to measure whether the RNA is intact or degraded. Finally, total RNA must undergo reverse transcription to synthesise complementary DNA (cDNA), and then qPCR can be used to amplify and quantify the expression of individual RNA targets.

In order to choose which method would be most appropriate for addressing the project aims, different technology platforms that are used in RNA analysis are discussed below, along with their chemistries and the techniques used at different stages in the RNA analysis pipeline.

1.8.1 RNA isolation

RNA extraction has similarities to DNA extraction, but there are a number of additional considerations that should be taken in account. RNA molecules are less stable than DNA and more susceptible to the aggressive activity of RNases, and also the reactivity of the 2′-hydroxyl group on the ribose sugar [97, 98], both of which make the isolation of RNAs more difficult. Since this project analysed the degradation rate of RNA markers, all equipment and reagents were RNase free to avoid any contamination with RNases that exist in the environment. A separate area was designated for RNA work, and the bench surface and all glassware were treated with an RNase inactivating agent (e.g. RNaseZAP). A lab coat, mask, and gloves were worn all the times, and sterile, disposable plasticware were used for liquid handling.

In addition to the above considerations, human body fluids contain a range of components (e.g. proteins, enzymes) that may interfere with the analysis of RNAs, and it is important that the appropriate method for isolation should be selected to remove these. Only methods that isolate total RNA including

miRNA are described in this section, as this project examined mRNA, miRNA and rRNA markers.

The most common RNA extraction method uses TRI Reagent®, which is a cheap and a convenient reagent for use not only for RNA isolation but also for isolating DNA and proteins from different tissue and cell types. The reagent consists of a guanidine thiocyanate and phenol mixture in a monophasic solution, which rapidly inhibits RNase activity [99]. The first step in this procedure is the lysis or homogenization step, where the biological sample is lysed in TRI Reagent®. The homogenate is then separated by the addition of 1-bromo-3-chloropropane (BCP) into three phases, the aqueous phase containing RNA, the interphase containing DNA, and the organic phase containing proteins. Following this are the precipitation and washing steps, with isopropanol and ethanol respectively, to remove any DNA or protein contamination. This method is considered to be effective for isolating different RNA molecules from 0.1-15kbp in length [99].

Additionally, there are widely available commercial kits for total RNA extraction including miRNA, such as miRNeasy (Qiagen) and mirVana™ (Ambion), which are based on chemical extraction using TRIzol and QIAzol reagents, followed by purification on silica columns [26, 100]. These methods all involve a first step in which lysis buffer is used to disrupt the cell components, inactivate RNases and maintain the integrity of RNA, but differ in the techniques used for purifying the total RNA.

The miRNeasy Mini Kit (Qiagen) combines phenol/guanidine-based lysis of samples and silica membrane-based purification of total RNA. Figure 1.12 illustrates the steps involved in extraction of total RNA using the miRNeasy Mini Kit. It uses QIAzol lysis reagent, a monophasic solution of phenol and guanidine thiocyanate similar to TRI Reagent, which is designed to facilitate lysis of cells and tissues, inhibit RNases, and remove most of the cellular DNA and proteins from the lysate by organic extraction [101]. After the addition of

chloroform, the sample undergoes centrifugation to separate three phases: RNA molecules remain in the aqueous phase, while DNA and proteins are found in the interphase and organic phase respectively. The aqueous layer is then transferred into a new tube, and ethanol is added to provide appropriate binding conditions for all RNA molecules that are 18 nucleotides or greater in length. The sample is then applied to an RNeasy Mini spin column, where the total RNA binds to the silica membrane and phenol and other contaminants are efficiently washed away [101].

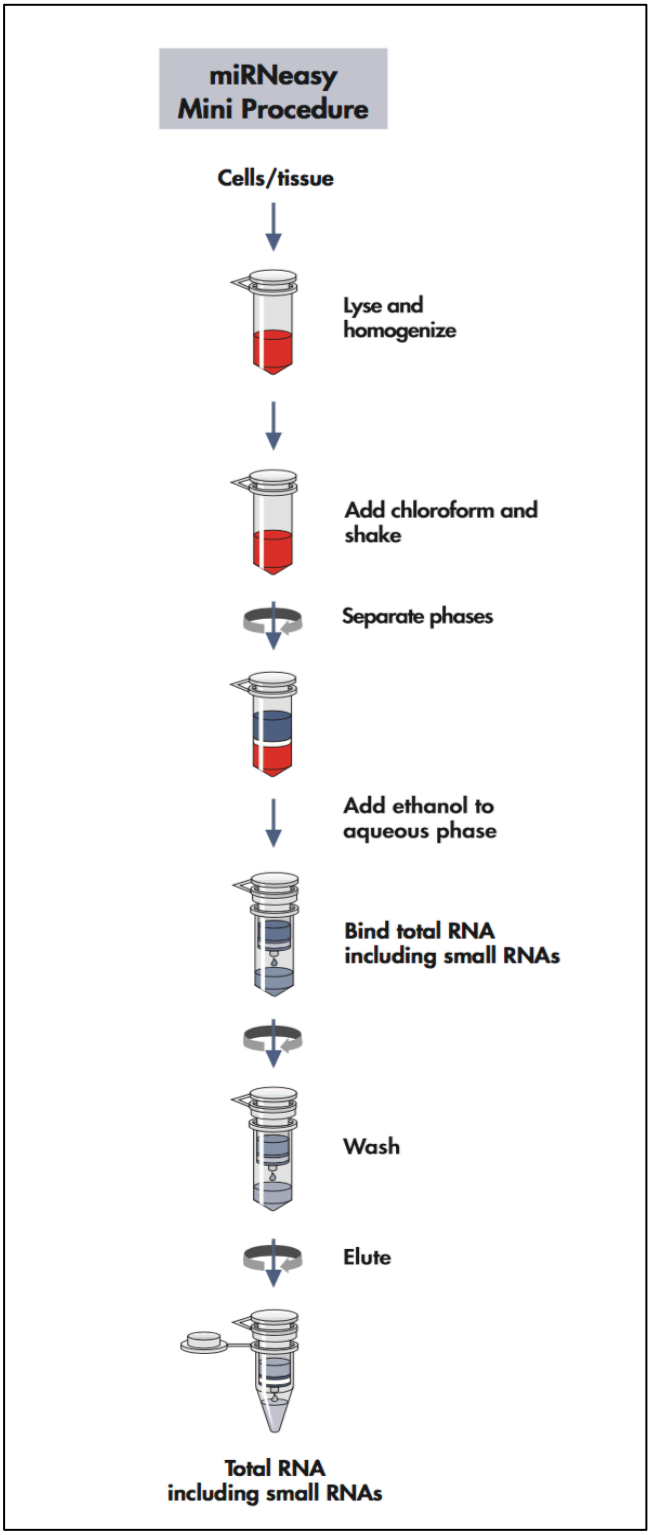


Figure 1.12: **Steps involved in total RNA extraction using the miRNeasy Mini Kit [101].**

1.8.2 Determining the quality and quantity of RNA

In gene expression studies, there has been increasing popularity in the use of real-time PCR, which demands high quality extracted RNA. Endogenous (e.g. proteins) and exogenous (e.g. phenol) contaminants can affect the quality of the total RNA. For example, phenol traces from organic extraction can reduce the efficiency of reverse transcription [102]. Therefore, it is important to determine the quality and quantity of isolated RNA to improve accuracy in RNA profiling studies [26, 103, 104]. There are many methodologies available to measure the yield and degree of RNA integrity, such as spectrophotometry, chip electrophoresis and gel electrophoresis.

1.8.2.1 UV-absorbance

A common technique used to measure the concentration of RNA sample is UV-spectrophotometry, e.g. using a NanoDrop, which measures RNA absorbance of UV at wavelength 260 nm (A_{260}). Both DNA and RNA have a specific absorbance profile in the wavelength range 220 to 350 nm. Pure RNA samples show a characteristic absorbance profile between 230-320 nm (Figure 1.13), and any deviation from this profile indicates the presence of contamination (e.g. proteins or phenol) [3].

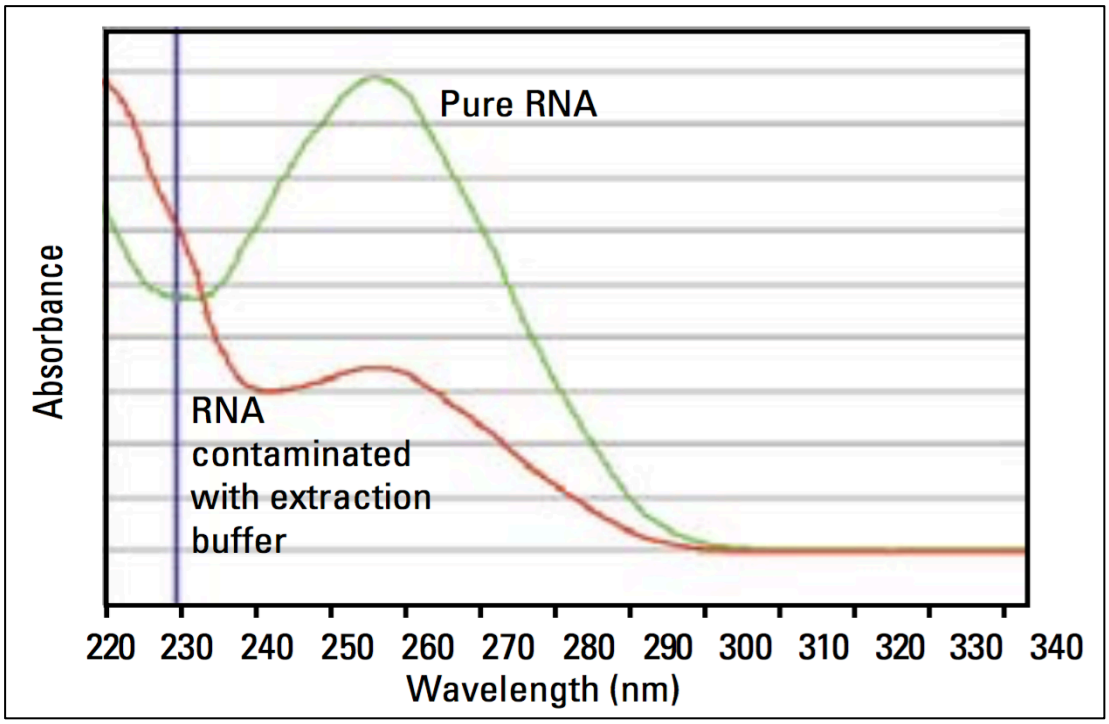


Figure 1.13: **Assessment of the quantity and purity of total RNA by UV-vis spectrophotometry.** The absorbance of the contaminated RNA is higher than the pure RNA at A_{230} [105].

The concentration of RNA in a sample is calculated using the Beer-Lambert law, applying the equation below based on its absorbance at 260 nm [3]:

$$A = e\,c\,L$$

Where: A = UV absorbance at 260 nm

e = average extinction coefficient of RNA (40 ng cm mL^{-1})

c = the concentration of RNA in the sample (ng/µL)

L = light path length

Measuring the ratio between UV absorbance at different wavelengths can also determine the purity of the sample (Table 1.6). Sample preparation for UV-spectrophotometry is very simple and does not require any additional mixing of reagents. No injection or separation is performed, which results in good reproducibility [106]. Although this method is the easiest and most rapid to

perform, it cannot detect RNA specifically as it is not selective and cannot distinguish between DNA, RNA and proteins [107], therefore RNA extracts must be treated with DNase to remove any remaining genomic DNA prior to RNA quantification. Another limitation is that it is influenced by sample contaminants such as genomic DNA or phenol which also absorb at 260 nm [106, 107]. It will also not reflect the integrity of RNA samples since single nucleotides will also contribute to the 260 nm reading [107], leading to difficulties in distinguishing whether a sample is degraded or not. However, a key advantage of using the NanoDrop instrument is that it requires only 0.5-2 µL of sample. For this project, the NanoDrop ND-1000 UV-visible microspectrophotometer (Thermo Fisher Scientific) was used to estimate the concentration of total RNA.

Table 1.6: **The use of UV absorbance ratios to determine purity and quality of RNA samples [3].**

Ratio	Expected ratio for pure RNA	Problem
$\mathbf{A_{260}/A_{280}}$	2.0	Ratio < 1.8 indicates protein contamination
$\mathbf{A_{260}/A_{230}}$	2.0-2.4	Ratio < 1.8 indicates organic compound contamination
$\mathbf{A_{260}/A_{240}}$	1.4	Ratio < 1.4 indicates the presence of large amount of salt protein contamination

1.8.2.2 Agilent 2100 Bioanalyzer

Similar to gel electrophoresis, the Agilent 2100 Bioanalyzer applies laser-induced fluorescence detection of nucleic acids coupled with microfluidic voltage-induced size separation for the quantification of RNA [108]. This technique involves injection and separation steps on sample-specific chips

[106], where an intercalating dye is used to stain RNA molecules and data are produced automatically as electropherograms, as shown in Figure 1.14.

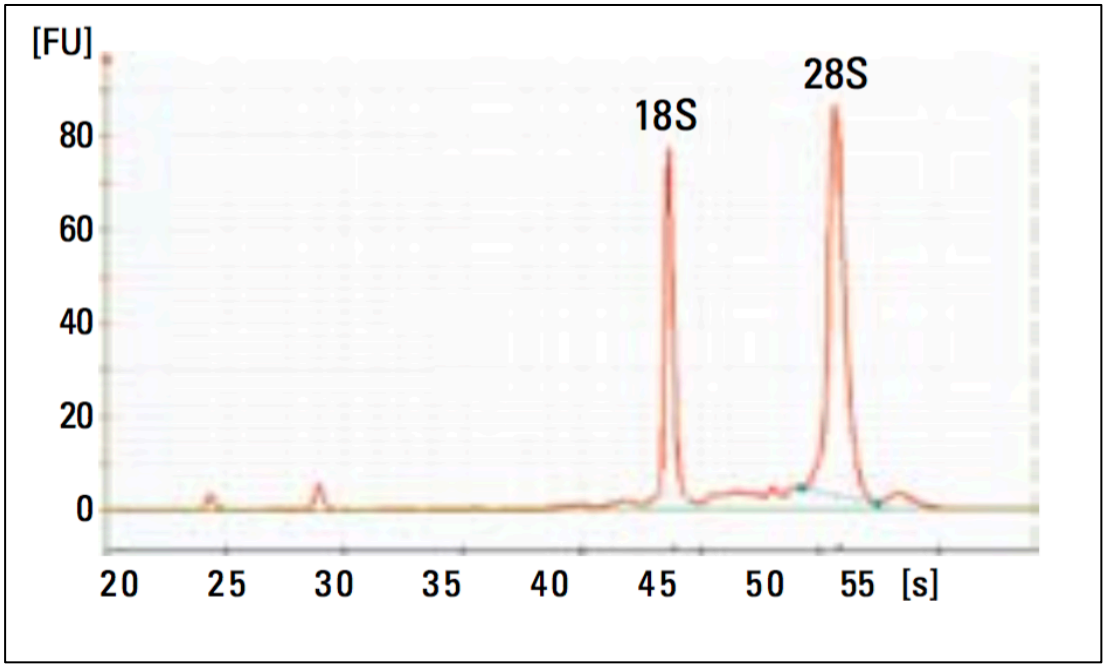


Figure 1.14: **Assessment of the RNA integrity by the Agilent Bioanalyzer 2100.** The *28S* and *18S rRNA* peaks represent intact RNA [105].

For total RNA analysis, the Agilent 2100 Bioanalyzer uses an RNA 6000 Nano Kit or RNA 6000 Pico Kit, depending on the expected concentration of RNA. In addition, the eukaryotic Total RNA Nano Assay can analyse small RNAs (<200 nt) such as miRNA [109].

The software algorithm assigns each RNA sample an RNA integrity number (RIN) from 1 to 10, where 1 is deemed most degraded and 10 most intact [107]. It uses the *18S* and *28S* rRNAs for quality assessment, as these abundant species comprise 90% of a cell's RNA content [4]. Based on the assessment of selected features, such as the total RNA ratio measured by the fraction of *28S* and *18S* peak area to total RNA area, *28S* peak height, and *28S* area ratio, information about RNA integrity can be determined [110].

This method can achieve a high level of sensitivity due to the use of a laser for excitation of intercalating fluorescent dyes, and requires only 1 µL of the RNA sample [106, 107]. The Agilent 2100 Bioanalyzer is not affected by phenol contamination because phenol does not interact with the fluorescent dyes. The ability of this method to quantify RNA degradation is limited when applied to heavily degraded RNA samples, where *28S* and *18S* peaks will hardly be present [107]. In addition, when comparing the reproducibility of the two methods, UV-spectrophotometry has a better level of reproducibility because no additional steps are introduced [106].

1.8.3 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) has become widely used in gene expression studies [111]. It can accurately measure the expression level of RNA molecules (e.g. mRNAs and miRNAs) [112, 113]. Most qPCR technologies measure fluorescence emission at the end of each PCR cycle as a means to 'track' the accumulation of DNA via amplification. TaqMan® and SYBR Green I assay are the two common approaches used to detect qPCR products. Both methodologies are discussed further below.

PCR is a method used to generate DNA 'copies' from a template DNA molecule exponentially, utilising a thermal cycler. Prior to the application of real-time PCR methods in RNA analysis, extracted RNA must be reverse transcribed into complementary DNA (cDNA), in order to perform the amplification process. This step is called reverse transcription (RT), which is the primary step that is performed before qPCR (RT-qPCR). The reverse transcription qPCR reaction can either be performed as a one-step or two-step reaction. In the one-step approach, both the synthesis of cDNA and the qPCR reaction occurs in one tube, while in the two-step approach, the two reactions are separated. Table 1.7 compares the two approaches and highlights the advantages and limitations of each reaction.

Table 1.7: **The advantages and limitations of two approaches for RT-qPCR [102, 114].**

	One-step RT-qPCR	Two-step RT-qPCR
Pros	• Simple and rapid. • Less expensive. • Less handling of samples. • High-throughput screening.	• cDNA stock available. • Highly sensitive. • Flexible priming options. • Reproducible data.
Cons	• Less sensitive. • No stock of cDNA. • Smaller number of targets per sample.	• Time consuming. • More pipetting steps. • Optimisation required.

1.8.3.1 Reverse transcription of RNA into cDNA

RT involves the conversion of an RNA strand into DNA through the enzymatic activity of a reverse transcriptase enzyme. The DNA product of this reaction is called cDNA. There are different types of primers that can be used in the RT reaction: sequence-specific primers, random primers and oligo-dT primers [115], details of which are included in Table 1.8.

Table 1.8: **The specifications of reverse transcription reaction primers [115].**

Primer type for cDNA synthesis	Specifications
Sequence-specific primers	• Only reverse transcribes gene-specific RNA sequences. • Can be used in both one-step and two-step RT-qPCR.
Random primers	• Can reverse transcribe all mRNAs including target and reference genes and *18S* rRNA. • Can transcribe degraded RNA and transcripts with hairpin loop structure.
Oligo-dT primers	• Only reverse transcribes mRNAs with poly(A) tails. • Cannot reverse transcribe *18S* rRNA. • Cannot reverse transcribe degraded RNA or transcripts with hairpin loops.

The specifications of the cDNA produced with each method are significantly different [116], and the choice of primers depends on the purpose of the study. Random primers bind to the target at different sites along the transcript to synthesise more than one cDNA product length per RNA target, therefore this method is non-specific [116]. Random primers can only be used in the two-step RT-qPCR. Oligo-dT primers, in contrast, are more specific than random primers and bind to the poly(A) tail of the RNA. However, high quality RNA is required for oligo-dT primers in order to generate reverse transcripts, and its activity is limited in the presence of secondary structure [116]. Like random primers, oligo-dT primers can only be used in two-step RT-qPCR. It is also possible to use a combination of both random and oligo-dT primers in one approach to produce more versatile cDNA [117] and to detect rare transcripts [118]. Sequence-specific primers allow synthesis of more specific target cDNAs, but limit RT to targeting only one gene of interest. In this work, random primers were used to produce cDNA from mRNA and rRNA markers, as it can produce short fragments of cDNA from all mRNAs and is not biased in terms of whether a transcript contains a poly(A) tail or not, or to specific sequences.

In contrast, for miRNAs, sequence-specific RT primers were used in the current project. The reverse primers have a stem-loop conformation, which provides high specificity as they can pick up as little as one nucleotide difference between related miRNAs [119]. They have an artificial 5´ end, which can fold on itself to form a stem loop, and a 3´ end that binds to 6 bp of the miRNA sequence (Figure 1.15). It has been shown that the specificity of stem-loop primers is not affected by genomic DNA contamination [119]. Additionally, the forward primer enhances the specificity by adding additional length with extra nucleotides to optimize and adjust the melting temperature [120].

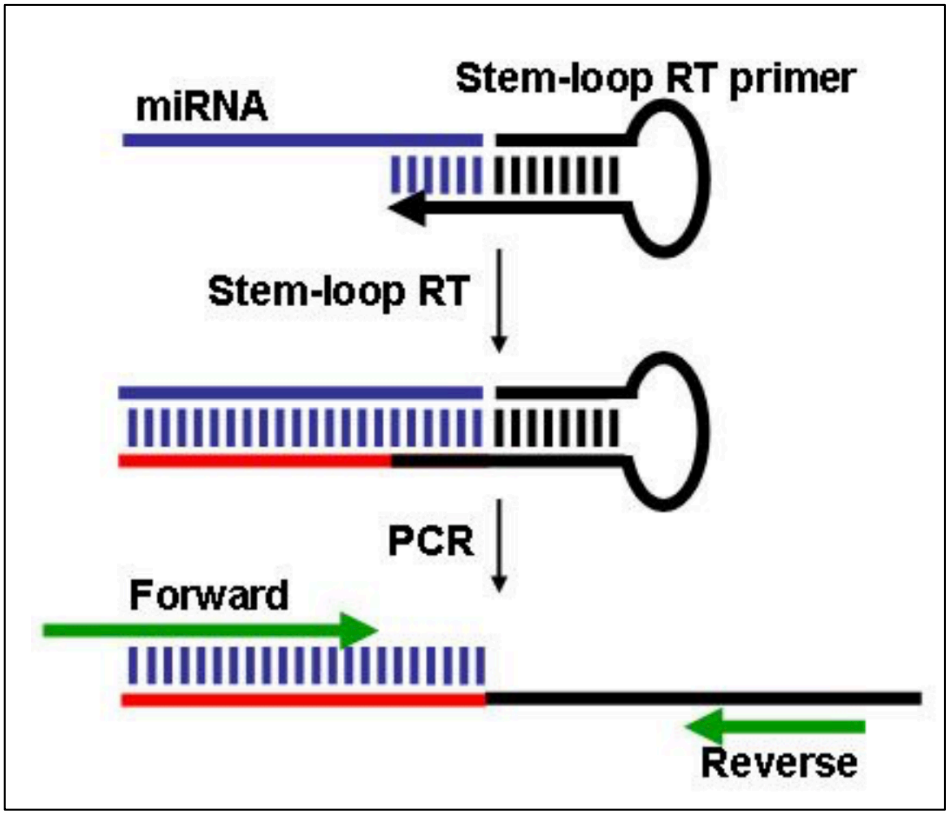


Figure 1.15: **The structure of a stem-loop primer for reverse transcription of miRNAs [121].**

During the RT procedure, the extracted RNA sample is incubated with RT buffer, dNTPs, reverse transcriptase and the selected primers. At 25 °C, the primer annealing process occurs, followed by initiation of RT of RNA into cDNA at ~ 37-50 °C. After a primer extension period during which reverse transcriptase synthesises the complementary cDNA strand using individual dNTPs, the reaction is heated to 85 °C to denature the reverse transcriptase and terminate the reaction. The cDNA product can be recognised by DNA polymerases and can be amplified in real-time PCR.

1.8.4 RT-qPCR detection methods

In an exponential fashion, PCR can generate copies of DNA from a cDNA template in a cyclical process using a thermal cycler for amplification and a laser source for excitation of fluorescent dyes (i.e. attached to PCR probes or bound to double-stranded amplicon) to detect and quantify PCR products. In

each cycle the temperature of the reaction mix is raised to 96 °C to denature the cDNA into single-stranded form, followed by an annealing step when the reaction mix is cooled to 55-70 °C to mediate the binding of the primers to the specific target sequence. The primers are then extended with dNTPs by *Taq* polymerase activity at 72 °C to replicate the target sequence. At the end of each extension step, a light source excites the reporter dye at dye-probe-complex at specific wavelengths to provide the energy needed for an electron transition within the dye molecule, which is then reversed leading to emission of light at a specific wavelength characteristic to that dye [122].

The software incorporated into qPCR instruments generates an amplification plot that presents the data for the fluorescence emission at the end of every amplification cycle. Figure 1.16 shows some important terms that are associated with the qPCR data. The baseline, or background noise, is the period early in the PCR when the fluorescent signal is below the detection level [123]. The y-axis of the plot represents the fluorescence emission, which is measured by a parameter known as the delta normalised reporter (∆Rn). The ∆Rn can be calculated using the below formula [123]:

$$\Delta Rn = Rnf - Rnb$$

Where: *Rnf* = the target's fluorescence emission at each cycle

Rnb = the baseline fluorescence emission

Another important term is the threshold, which represents the limit of detection. Once any fluorescent signal is detected above the threshold it is considered as a real signal and used to determine the threshold cycle (C_t), also known as the cycle of quantification (C_q) or the crossing point (C_p). The C_q value is the number of PCR cycles required to amplify DNA to a level whereby the fluorescence signal is above the baseline and the reaction is still in the exponential phase [102]. The fluorescence intensity correlates to the number of PCR products.

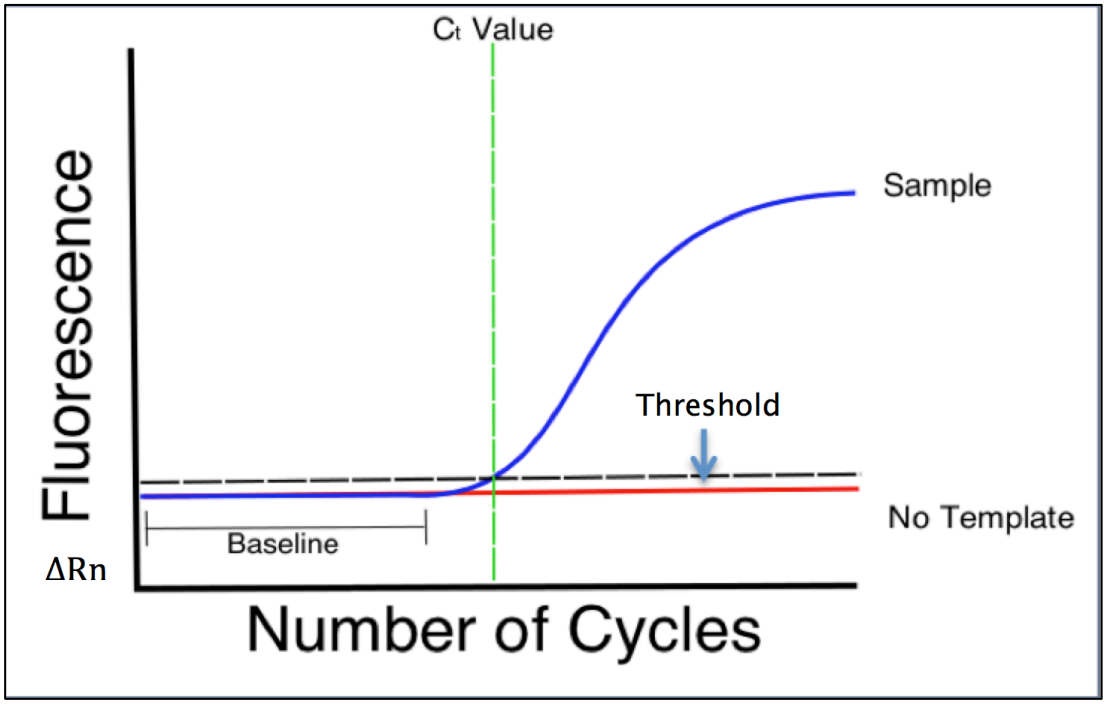


Figure 1.16: **Example amplification plot for qPCR [124].** Shows the C_t value of a sample, which represents the number of PCR cycles required to cross the threshold.

The products of the qPCR can be measured while the reaction is still in the exponential phase. Therefore, qPCR combines both amplification and detection into one step. Figure 1.17 illustrates the four phases that divide the qPCR amplification curve: the linear ground phase, early exponential phase, exponential phase and plateau phase. During the linear ground phase, PCR has just begun, and the fluorescence emission is still below the baseline or background noise. When the fluorescence emission crosses the threshold, the early exponential phase starts and the C_q value can be determined. The optimal amplification of the PCR product is reached during the exponential phase and with each cycle of PCR, theoretically the PCR product is doubled in quantity. The final phase is the plateau, which represents the end of the PCR where no more significant increase in fluorescence emission is detected due to reagent limitation.

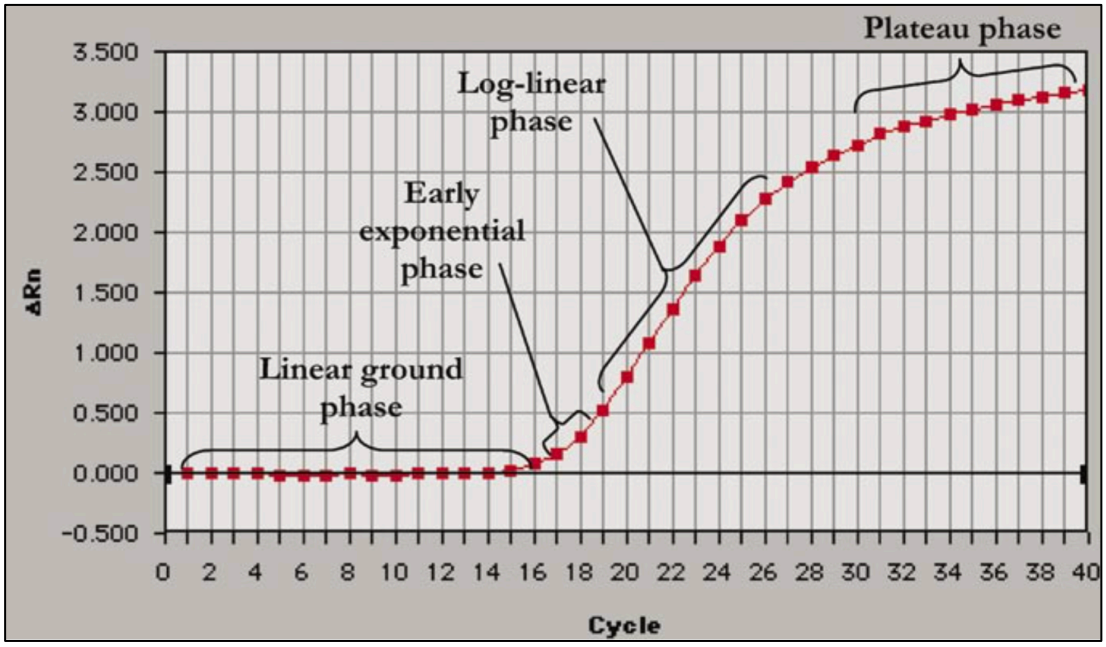


Figure 1.17: **The phases of real-time PCR amplification [125].** Each phase relates to the fluorescence emission during the amplification reaction.

The two general approaches that are used to detect RT-qPCR products are to use the non-specific DNA binding dye SYBR® Green I and TaqMan® Assays. These methods are used to quantify the expression of the studied genes based on qPCR, applying the same principles discussed above, where the fluorescence signal from the sample is measured to determine the C_q value (Figure 1.17). Both technologies are discussed in further detail in this Chapter.

For quantifying gene expression, qPCR is now widely used for several reasons. It is considered to be highly sensitive [126] when compared to other methods such as RNase protection assays [127], and dot-blot hybridisation [128]. Furthermore, it has high-throughput and requires less RNA template than many other methods for the analysis of gene expression [125]. However, the instruments and the reagents needed for real-time PCR are expensive, and for accurate results data normalisation should be fully understood [125].

1.8.4.1 TaqMan® Assays

TaqMan® assays use fluorescently-labelled hydrolysis probes for the detection of PCR products. These hydrolysis probes are sequence-specific (20 to 30 bases in length), containing a fluorescent reporter dye (e.g. FAM, VIC, TET) at the 5′ end and a quencher dye (TAMRA) at the 3′ end, as shown in Figure 1.18. Similar to SYBR® Green I, the reporter dye is excited by a specific wavelength of light, generated by the qPCR instrument. When the probe is intact, the quencher reduces the intensity of the reporter dye fluorescence [125, 129]. During the annealing step of PCR, the probe is bound to the specific target between the forward and reverse primers, and due to the 5′ exonuclease activity of the *Taq* polymerase enzyme, the probe is cleaved. This removes the connection between the fluorophore and quencher, leading to fluorescence emission at a higher wavelength than the light generated by the instrument, based on the principle of fluorescence resonance energy transfer (FRET) [130]. The increase in the fluorescence signal corresponds to the increase in the quantity of PCR products.

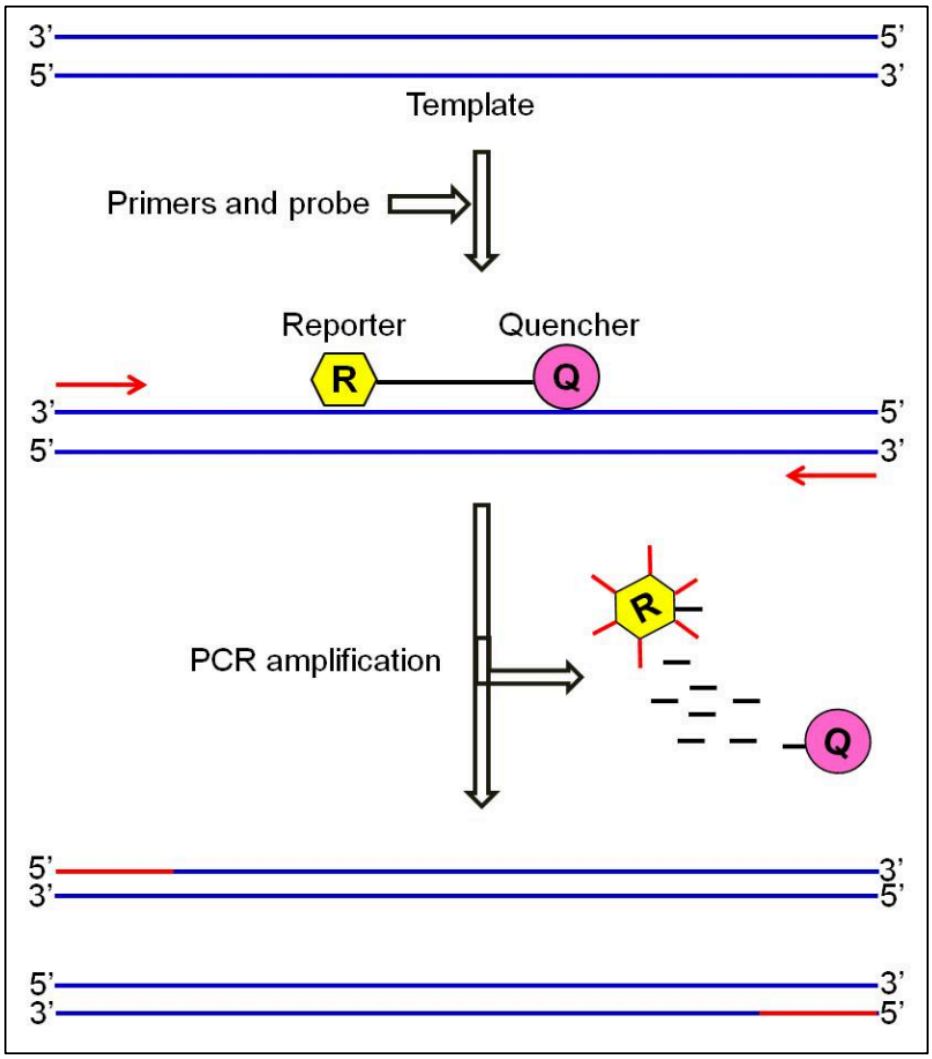


Figure 1.18: **TaqMan® chemistry [131]**. At the 5′ and 3′ ends of the probe, the reporter dye and quencher are present respectively. When amplification occurs, these are separated, and fluorescence emitted from the reporter dye.

The probes that are used in the current project are TaqMan® assays, which are minor groove-binding probes (MGB) labelled with FAM (excited at 494 nm wavelength and emits at 518 nm). These probes contain a minor groove-binder molecule at the 3′ end. This molecule folds into the minor groove of the double-stranded DNA (dsDNA) to stabilise the probe-target complex [123]. Using these kinds of probes also increases the melting temperature (T_m) of the probe and stabilises probe-target hybrids, to allow accurate allelic discrimination using shorter probes [123]. TaqMan® assays offer very specific hybridisation

due to the sequence-specific nature of the target regions. However, the cost of these assays is very high compared to other methods.

1.8.4.2 SYBR® Green I Assays

The SYBR® Green I method detects the amplification products as they accumulate during PCR cycles, by utilising a DNA intercalating dye, as shown in Figure 1.19. The fluorescence of the intercalating dye SYBR® Green I increases when it binds to the minor groove of dsDNA, is excited by a light source at a wavelength of 497 nm, and emits light of wavelength 520 nm [132, 133]. The accumulation of PCR products is thus monitored by the emission of light of wavelength 520 nm by SYBR® Green I, and the measured level of fluorescence is proportional to the number of PCR amplicons produced. This method does not require the addition of fluorescently-labelled oligonucleotides [134, 135], meaning it has lower setup and running costs compared to other technologies (e.g. TaqMan® assays). Additionally, the assays are not complicated and are easy to design. However, there are some limitations when using SYBR® Green I assays, one being that the dye is not specific and can bind to any dsDNA including non-specific products such as primer-dimers [135], which may lead to false positive signals. Therefore, the SYBR® Green I method requires primer optimisation to ensure accurate quantification and robust, highly specific amplification. Another limitation is that when working with multiplex assays SYBR® Green I cannot be used, and also that the length of the amplicon can affect the intensity of the amplification signal [135].

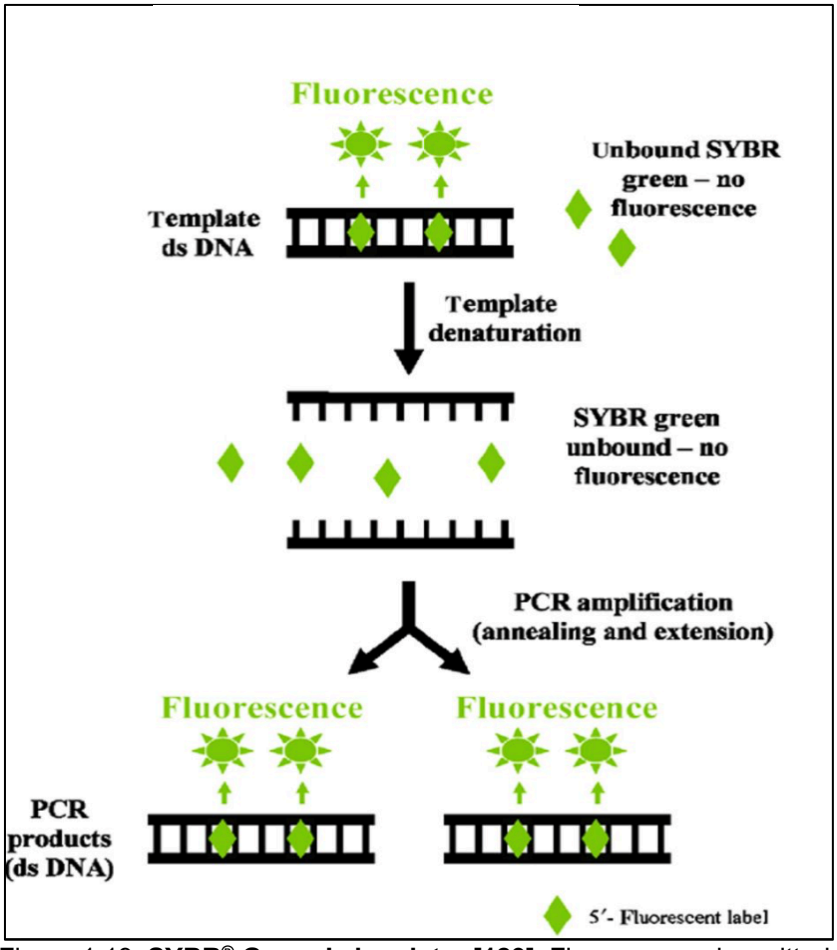


Figure 1.19: **SYBR® Green I chemistry [129].** Fluorescence is emitted when the SYBR Green dye binds to dsDNA.

1.8.5 Amplification efficiency

The overall performance of qPCR can be evaluated using a number of parameters, one of which is the amplification efficiency, which measures the rate at which PCR amplicons are generated. In an ideal reaction the amplification efficiency is assumed to be equal to 100%, which means that the PCR product is doubled with every cycle during the exponential phase. However, this assumption cannot be applied to many PCR reactions because not all reactions can be carried out under ideal conditions, and variation exists from sample to sample due to RNA extraction methods, the presence of PCR inhibitors, and many other factors that can affect amplification efficiency [136]. Therefore the data should be processed with an appropriate correction factor using either a standard curve or raw PCR data to avoid inaccurate

quantification results [125]. A real-time PCR standard curve is represented as a semi-log regression line plot of C_q values vs. the log of total RNA input. The value of a standard curve slope indicates the efficiency percentage; a slope of -3.32 indicates 100% efficiency, whereas a slope > -3.32 indicates an efficiency less than 100%. The amplification efficiency is calculated using data collected from a standard curve made from a dilution series of cDNA by applying the equation of:

$$\text{Efficiency} = 10^{(-1/\text{slope})}$$

The efficiency gives an indication of the proportion of PCR products that are doubled every PCR cycle. The acceptable range of an assay efficiency is between 90-110%, which corresponds to standard curve slopes of -3.6 to -3.1 [137]. The linearity of qPCR is measured by the R^2 value of the regression line, which should be ≥ 0.98 [137].

Furthermore, another factor that affects the reaction efficiency is the primer itself, which should be designed very carefully to avoid non-specific hybridization or primer dimers [138]. There are specific guidelines for designing a primer to avoid primer-dimer formation; the primer should be 18-28 nucleotides in length, have 50% GC content, avoid stretches of repeats, and sequence complementarity between all primers in an assay should be avoided [122]. Applying an incorrect amplification efficiency will lead to unreliable estimates of gene expression changes [136], therefore a reliable correction should always be included in the procedure.

1.8.6 RT-qPCR quantification methods

Analysing the data correctly from RT-qPCR experiments is critical in order to produce accurate and reliable results. There are two strategies that can be applied in gene expression data analysis; absolute and relative quantification.

The strategy that is most appropriate depends on the specific aims of the gene expression study, and how the investigator wishes to present the results.

1.8.6.1 Absolute quantification

This approach determines the absolute copy number of an RNA of interest, or the unknown RNA concentration of a sample, by comparison to a standard curve produced from a serial dilution of known RNA concentrations [136, 139]. The standard curve generates a linear relationship between the C_q value and the initial RNA concentration, and based on the C_q value of unknown samples, their concentration can be determined by comparison to the standard curve [125, 136, 139]. This method assumes that the standards and the samples have equal amplification efficiency and is usually used when knowledge of the exact quantity of sample is required.

1.8.6.2 Relative quantification

Relative quantification allows researchers to quantify differences in the gene expression level of a specific target gene between different samples. The relative quantification strategy does not use a standard curve, but instead measures the change in the expression level of a target gene relative to an internal reference gene i.e. one which is ubiquitously expressed in all cell types, and expected to be stably expressed regardless of the experimental variables under investigation [136, 139]. This approach is considered to be more suitable when studying the physiological changes in gene expression for any particular gene [136]. The results of relative quantification can be presented as a difference (ΔC_q) or ratio (relative expression ratio (RER)).

The ΔC_q method calculates the C_q difference between two different RNA samples after normalising them against a reference gene:

$$\Delta C_q = C_q \text{ (target)} - C_q \text{ (reference)}$$

If the experiment involves measuring gene expression under two different experimental conditions, for example at different time points following the deposition of a biological stain, $\Delta\Delta C_q$ can be calculated as shown in the equations below:

$$\Delta C_q = C_{qT1}\,(\text{target}) - C_{qT1}\,(\text{reference})$$

$$\Delta C_q = C_{qTx}\,(\text{target}) - C_{qTx}\,(\text{reference})$$

$$\Delta\Delta C_q = \Delta C_{qTx}\,(\text{target}) - \Delta C_{qT1}\,(\text{reference})$$

Where: T1 = time point zero
Tx = any other time point

$\Delta\Delta C_q$ can also be presented as a fold change between the two different RNAs using the following formula:

$$\text{Fold change} = 2^{-\Delta\Delta Cq}$$

Using this method, the data can be presented as the fold change upwards/downwards in expression, in response to some experimental variable, which is more relevant for measuring gene expression than the absolute quantity. The qPCR efficiency in this model is assumed to be close to 100% and the amplification efficiencies of the target and the internal control gene are assumed to be similar, or relatively equivalent, during gene expression analysis [139].

Additionally, the variation in gene expression quantification based on the ΔC_q of the treated sample (i.e. the target) and an internal or external control (i.e. reference gene), by calculating the relative expression ratio (RER) shown below. Both Anderson et al. (2005) [58] and Hampson et al. (2011) [60] applied the RER to their aged bloodstains and hair samples, respectively, to identify

the relative quantity of *18S* and *ACTB* (*ß-actin*) [58, 60]. They calculated the RER of *ACTB* to *18S* by dividing the C_q values of *ACTB* by the C_q values of *18S* rRNA. Their C_q values were corrected against the determined efficiencies of both *ACTB* and *18S*:

$$\text{RER} = \frac{Cq\ value\ of\ ACTB}{Cq\ value\ of\ 18S}$$

1.8.7 Normalization and reference genes

In order to correct for or minimise sample variation due to differing amounts of input RNA or specific experimental errors that have been introduced in the RT-qPCR, a normalisation step should be performed [123]. Table 1.9 summarises different normalisation strategies including the benefits and limitations of each strategy. An internal reference gene is amplified along with the target. The main specification for the reference genes is that their expression level should be stable under different experimental or environmental conditions. To generate accurate data, the reference genes must be selected very carefully depending on the conditions of the experiments and the type of samples. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ß-actin (*ACTB*) and *18S* are among the most commonly used reference genes due to the assumption of constant expression level across different experimental conditions, however it has been illustrated that these genes are not consistently expressed across body fluids and tissue types [140].

Table 1.9: **Normalisation strategies for gene expression studies [141].**

Normalisation strategy	Details	Benefits	Limitations
Similar sample size	- Similar sample volume or weight is used	- Easy	- Difficult to estimate accurately
Total RNA	- Similar total RNA input is used	- Similar reverse transcriptase input	- Errors at RT or qPCR are not controlled - The rRNA/mRNA ratio variations are not indicated
Reference genes (rRNA)	- Target C_q values to rRNA C_q values	- Internal control gene is subject to same conditions as the target RNA - Its level is measured by qPCR	- Validation must be performed to ensure rRNA stability
Reference genes (mRNA)	- Target C_q values to reference mRNA C_q values	- Internal control gene is subject to same conditions as the target RNA - Its level is measured by qPCR	- Validation must be performed to ensure reference mRNA stability
'Spiked' in RNA	- Use artificial molecule incorporated into the sample	- Acts as Internal control gene and subject to same conditions of RT and qPCR. - Its level is measured by qPCR	- Must be identified and synthesised - Unlike RNA, not extracted from the same cell

Reference genes are commonly used for normalisation purposes in many gene expression studies whether in the clinical or forensic field [142-145]. There are a number of software programs that allow the assessment of the variability of RNA reference genes, to help the investigators to select the most stable reference gene for normalisation. The two most common software programs that are used for such a purpose are *geNorm* [146], which calculates the minimum numbers of reference genes required for reliable and accurate normalisation, and *NormFinder* [147] that determines a stability value for each gene. Both programs can identify the most stable RNA and suggest whether to use one or more RNAs for normalisation.

1.9 Aims of this project

The overall purpose of this project was to develop a method to estimate the time since deposition of body fluid stains commonly encountered in forensic casework - blood, saliva and semen - using the application of RNA analysis.

The degradation profiles of a variety of selected RNA markers were analysed, including mRNA and miRNA body fluid specific markers, and reference genes. The aim was to look for a correlation between the age of body fluid stains and the degradation rate of these RNA markers, in order to identify those most useful for ageing body fluid stains, and to assess the possibility of using the degradation rate to determine time since deposition or the age of the stain.

The first aspect of the project was method development, where a number of potential protocols for RNA extraction (TRI Reagent® and miRNeasy) were tested on different types of body fluid and directly compared to determine which purification method produces high yields of total RNA using small initial volumes of body fluids (20 µL). The quality of the extracted total RNA was also measured using the Agilent 2100 Bioanalyzer, which is used as a screening indicator to determine RNA integrity. This step was conducted to investigate whether the obtained RNA quality of very small sample volumes should be

taken into account to decide whether the sample should be included for further RT-qPCR analysis or discarded and whether the measured quality of the purified total RNA may represent the state of the target RNA markers. Overall, based on the outcomes, the suitable protocol (i.e. extraction method) was selected and applied into subsequent experiments to assess the degradation rate of RNA markers.

The second aspect of this project was to analyse the expression level and degradation rate of blood-specific markers in blood samples, namely *HBA*, *HBB*, *HMBS*, *miR16,* and *miR451*. These RNA markers were selected from a thorough literature review, and have been shown to indicate the presence of bloodstains. The blood samples were stored in a dark, dry place at room temperature, to simulate natural ageing until they reached a series of desired ageing time points (0, 3, 6, 15, 30, 90, 180, 270 and 360 days) at which stage total RNA was extracted. A study was then performed on the association between RNA marker stability and the relative expression ratio (RER) of these RNA molecules, in order to provide information about which markers are likely to be more accurate for use in estimating the age of bloodstains, both over the short- and long-term.

The third aspect of the project was to apply the same method used in bloodstains to saliva and semen samples, using RNA markers that have been identified in the literature to indicate their presence. A total of seven RNAs were evaluated for the purposes of studying their degradation rate, including two RNA saliva markers (*STATH*, *miR205*) and five semen RNA markers (*PRM1*, *PRM2*, *SEMG1*, *miR10b*, *miR891a*). Both saliva and semen samples were stored in a dark, dry place at room temperature, to simulate natural ageing until they reached a series of desired ageing time points (0, 7, 14, 28, 90, 180, 270 and 360 days) at which stage total RNA was extracted. The relative expression ratios of these markers were also calculated and analysed to determine whether these specific markers can be used to estimate the age of saliva and semen stains.

The fourth aspect was to investigate the impact of body fluid mixtures on the RER of blood-specific markers. The expression level of selected blood-specific RNA markers was measured in both pure bloodstains, and samples of blood mixed with other body fluids (saliva and semen), and the degradation rate analysed to determine the relative expression ratios of these markers. These RERs were then compared among different sample types to determine whether the presence of other body fluid affects the ratios, and hence estimated time since deposition.

The final aspect of this study was to compare the degradation behaviour of two reference genes commonly used in gene expression studies and forensic casework (*18S* and *ACTB*) across three types of body fluids, to determine which RNA markers are most suitable for use in body fluid identification studies, and whether their relative expression ratio can be used to estimate the age of stains across body fluid types. Both reference genes have shown a level of variation in stability and expression levels in different sample types, leading to the conclusion that neither of the examined genes can be used as reference genes for forensic purposes and the relationship between RER and ageing time points could not be established.

Chapter two: Materials and Methods

This research focused on studying the degradation behaviour of different types of RNA markers in multiple body fluids at different ageing time points. The degradation patterns of these RNA markers were then analysed further to see whether they can provide any information on the age of the biological stains or time since deposition. Likewise, this Chapter covers the research methods and laboratory protocols that were followed in the study for generating degradation profiles of RNA markers from body fluid samples (blood, saliva, and semen). The sections below describe the ethical approval, the collection of different sample types, and the key procedures in RNA analysis, starting with RNA extraction, through reverse transcription and quantification of the expression level of each RNA marker that was selected to be examined in this project. The data collection, instrument used, and the methods used for statistical analysis in this research are also described in this Chapter.

2.1 Ethical approval

It is essential to ensure that all experimental procedures that involve human participants are carried out in an ethical manner. The main ethical issues that should be considered relate to the collection, use and retention of participants' biological samples and any data associated with them. These are addressed by provision of an information sheet explaining the project to potential participants, the requirement that donors sign a consent form to indicate they agree to donate a sample, the pseudo-anonymisation of samples and any associated data, the secure storage of biological samples during the project and destruction of any biological samples once the project was completed, and the secure storage of data. These issues were all addressed in an application for ethical approval from the Department of Pure and Applied Chemistry Departmental Ethics Committee, and these were approved by the committee. Informed consent was acquired by obtaining signed consent sheets from each

donor after they had read the Participant Information Sheet (PIS) (see appendix A1).

2.2 Sample collection

2.2.1 Blood samples

Blood samples were collected using sterile disposable Unistik 3 Comfort lancets (Shandong Lianafa Medical Plastic Products, China). Blood was pipetted from the donor's finger using a sterile disposable pipette tip (Elkay Laboratory Products, Hampshire, UK) and deposited onto sterile cotton swabs (Fisher Scientific, Leicestershire, UK) in 20 µL aliquots. Samples were stored for different time periods up to one-year in a dark dry place at room temperature to simulate natural ageing, until they reached the desired ages (see Chapter three, four and seven).

2.2.2 Saliva samples

Saliva samples were collected by participants depositing samples directly into sterile collection pots. Participants were asked not to eat, drink or smoke for at least one hour prior to samples being collected. Samples were then returned to a designated staff member and placed into the laboratory in a suitable refrigerator. A sterile disposable pipette tip was used to transfer 50 µL aliquots of the sample onto sterile cotton swabs. These were then stored for different time periods up to one-year in a dark dry place at room temperature to simulate natural ageing, until they reached the desired ages (see Chapter three, five and seven).

2.2.3 Semen samples

Semen samples were collected by participants depositing samples directly into sterile collection pots, which they were provided with to take home. Samples were then returned to a designated staff member and placed into the laboratory in a suitable refrigerator. A sterile disposable pipette tip was used

to transfer 50 µL aliquots of the sample onto sterile cotton swabs. These were then stored for different time periods in a dark dry place at room temperature to simulate natural ageing, until they reached the desired ages (see Chapter three, five and seven).

2.2.4 Body fluid mixture samples preparation

Mixtures of different body fluids were also examined during this project (see Chapter six). Firstly, pure samples of the three different body fluids (blood, saliva and semen) were deposited onto swabs, as described in sections 2.2.1-2.2.3 above. Mixture samples were then prepared by pipetting 20 µL of blood onto cotton swabs, and then adding 20 µL of saliva or semen to the same swab. These samples were subject to immediate processing after allowing them to dry for 15 min.

Another set of pure body fluid samples and mixture samples were also prepared and aged for up to two months, by pipetting 20 µL of blood onto cotton swabs, adding 20 µL of saliva or semen, and then stored for different time periods up to 60 days in a dark dry place at room temperature to simulate natural ageing, until they reached the desired ages (see Chapter six).

2.3 RNase treatment

Certified RNase-free plasticware including 0.2 and 1.5 mL tubes, microstrips, microcaps and pipette tips (Elkay Laboratory Products, Hampshire, UK) were used throughout the duration of the study. Other laboratory surfaces and instrumentation were sterilised using RNaseZap® (Life Technologies, Paisley, UK) regularly to eliminate RNase. All apparatus was frequently decontaminated using a UV sterilisation cabinet (BIGNEAT, Hampshire, UK).

2.4 RNA extraction and purification

For total RNA extraction, two different procedures were tested and compared in Chapter three: TRI Reagent® (Sigma-Aldrich, Gillingham, UK) and miRNeasy Mini Kit (Qiagen, Manchester, UK).

2.4.1 TRI Reagent® procedure

Extraction and purification of total RNA from body fluid using TRI Reagent® was performed following the manufacturer's guidelines (Sigma-Aldrich, Gillingham, UK) [99]. Cotton swabs containing body fluid samples were cut into small pieces by sterile scalpel blade (Fisher Scientific, Leicestershire, UK) and placed into a fresh 1.5 mL tube. 200 µL of RNase-free water (ThermoFisher Scientific, Invitrogen, UK), 3 µL of a polyacryl carrier (Molecular Research Centre, OH, USA) and 750 µL of TRI Reagent® (Sigma-Aldrich, Gillingham, UK) were added to the tube, and the mixture vortexed for 1 min to disrupt the cells, then incubated at 50 °C for 10 min. 100 µL of 1-bromo-3-chloropropane (Sigma-Aldrich, Gillingham, UK) was added to the tube and centrifuged for 15 min at 4 °C at 12,000 x g. Three phases were formed in the tube and the upper aqueous phase (~ 500 µL) was transferred into a new tube while the interphase and the lower organic phase were discarded. 500 µL of cold (4 °C) isopropanol (Sigma-Aldrich) was added into the new tube containing the aqueous layer, which was incubated at room temperature for 7 min, followed by centrifugation for 8 min at 4 °C at 12,000 x g. The supernatant was discarded and 1 mL of 75% ethanol (Sigma-Aldrich) was added to the RNA pellet, pipetted up and down to mix the solution, followed by centrifugation for 5 min at 4 °C at 12,000 x g. The supernatant was again discarded, and the RNA pellet was allowed to dry at room temperature for 5 min, before being resuspended by adding 50 µL of RNase-free water (ThermoFisher Scientific, Invitrogen, UK) and incubating at 55 °C for 10 min.

2.4.2 miRNeasy Mini Kit procedure

Extraction and purification of total RNA from body fluid using miRNeasy Mini Kit was performed following the manufacturer's standard guidelines [101]. Cotton swabs containing body fluid samples were cut into small pieces by sterile scalpel blade (Fisher Scientific, Leicestershire, UK) and placed into a fresh 1.5 mL tube. 700 µL of QIAzol Lysis Reagent (Qiagen, Manchester, UK) was added to the tube, vortexed for 1 min to disrupt the cells and incubated at 56 °C for 10 min. 140 µL of chloroform (Sigma-Aldrich, Gillingham, UK) was added to the tube and vortexed for 15 s. The tube was placed on the benchtop for 3 min at room temperature and then centrifuged for 15 min at 4 °C at 12,000 x g. Three phases were formed, and the upper aqueous phase was transferred into a new tube while the interphase and the lower organic phase were discarded. 525 µL, or 1.5 volumes, of 100% ethanol (Sigma, Gillingham, UK) was added to the tube and mixed thoroughly by pipetting up and down several times. 700 µL of the sample was transferred into an RNeasy Mini spin column (Qiagen, Manchester, UK) in a 2 mL collection tube and centrifuged at 8,000 x g for 15 s at room temperature. The flow-through was discarded and the remainder of the sample was added to the column and the centrifugation and discard steps were repeated. 700 µL of RWT Buffer was added to the column, centrifuged at 8,000 x g for 15 s and the flow-through discarded. 500 µL of RPE Buffer was added to the column, centrifuged at 8,000 x g for 15 s and the flow-through was discarded. Another 500 µL of the RPE buffer was added to sample and centrifuged at 8,000 x g for 2 min. Total RNA was eluted from the column by the addition of 50 µL RNase-free water, incubation at room temperature for 5 min, and centrifugation at 8,000 x g for 1 min. A second elution step was performed by using the first eluate RNA from previous step, by adding the eluted RNA back onto the column, and then incubated at room temperature for 5 min and centrifuged at 8,000 x g for 1 min, to obtain a higher total RNA concentration.

2.5 DNase treatment of RNA samples

The TURBO DNA-free™ Kit (Ambion®, Life Technologies Paisley, UK) was used for the digestion of genomic DNA (gDNA). The presence of gDNA in the extracted RNA sample can affect the sensitivity and specificity RT-qPCR. Non-specific amplification can occur in RNA samples that are contaminated with gDNA leading to overestimation of RNA transcript levels. DNase treatment was done following the manufacturer's protocol [148].

0.1 volume of 10X TURBO DNase Buffer and 1 µL TURBO DNase were added to the extracted RNA, and vortexed briefly. The sample was then incubated at 37 °C for 20-30 min. DNase Inactivation Reagent was resuspended by flicking the tube before dispensing it. The resuspended DNase Inactivation Reagent (0.1 volume) was added to the sample, mixed well and incubated for 5 min at room temperature, mixing by flicking occasionally. The tube was centrifuged for 3 min at 10,000 x g, and the RNA was transferred into a fresh tube, discarding the pellet containing the DNase and DNase Inactivation Reagent and any gDNA. Samples were then stored at -20 °C.

2.6 Quality and quantity of RNA

2.6.1 UV-visible spectrophotometry

The NanoDrop 1000 Spectrophotometer (ThermoFisher Scientific, Loughborough, UK) was used to check the quality and the quantity of the total RNA present in the samples by following the manufacturer's instructions [149] and using 1.5 µL of the extracted sample. The NanoDrop was first initialised and blanked using 1.5 µL of RNase free water (ThermoFisher Scientific, Invitrogen, UK), then 1.5 µL of the extracted sample was analysed and the RNA extract absorbance was measured over the 220 to 350 nm wavelength range. RNA concentration was calculated using Beer-Lambert law and the purity of the sample was determined by measuring the ratio between UV absorbance at different wavelengths (see Chapter one section 1.8.2.1).

2.6.2 Bioanalyzer 2100

The quality and quantity of the extracted RNA was also measured using a Bioanalyzer 2100 coupled with an Agilent RNA 6000 Pico Kit (Agilent Technologies, Wokingham, UK), following the instructions provided by the manufacturer [150]. The RNA 6000 Pico Kit was selected as it has a very low required input concentration of total RNA (up to 5 ng/μL), making it very sensitive for the determination of RNA integrity, which is quantified by the RNA Integrity Number (RIN) (see Chapter one section 1.8.2.2).

1 μL of Pico dye concentrate was mixed into 65 μL filtered Pico gel matrix and centrifuged for 10 mins at 10,000 x g. 9 μL gel-dye mix and 9 μL conditioning solution were added to the designated wells on the chip to prime the Pico chip. 1 μL of each RNA sample was then added to an individual primed chip well, along with 5 μL Pico marker, which is a 25 nt RNA fragment added to all wells for quality control purposes. The chip was then vortexed for 1 min at 2,400 rpm prior loading it into the instrument.

2.7 Reverse transcription (RT)

During this project two different primers types were used for cDNA synthesis (see Chapter one section 1.8.3.1). The cDNA strands were synthesized from total RNA using two different kits: the High-Capacity cDNA Reverse Transcription Kit and the TaqMan® microRNA Reverse Transcription kit (both Applied Biosystems, Life Technologies, Paisley, UK).

2.7.1 High-Capacity cDNA Reverse Transcription Kit

The High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Life Technologies, Paisley, UK) was used to produce cDNA from mRNA and rRNA within the total RNA, by using the addition of random primers (see Chapter one section 1.8.3.1), which allow the user to generate cDNA from the transcriptome in a single RT reaction, following the manufacturer's guidelines [151]. Each reaction was set up as described in Table 2.1 below.

Table 2.1: **The components of each RT reaction using the High-Capacity cDNA kit.**

Component	Volume (µL)
10X RT Buffer	2.0
25X dNTP Mix (100mM)	0.8
10X RT Random Primers	2.0
MultiScribe™ Reverse Transcriptase	1.0
Nuclease-free Water	4.2
Total RNA Extract + H_2O	10.0
Total reaction	20.0

For blood samples and body fluid mixture samples; RNA concentration was normalised to 30 ng/µL, and the maximum volume was taken if samples had a concentration lower than 30 ng/µL, while for saliva and semen samples the normalised concentration was 50 ng/µL. For the RT-qPCR controls, the RT-control was prepared by including all the components above except the MultiScribe™ Reverse Transcriptase and negative controls (NTC) were prepared by including all the components above but replacing the RNA extract with RNase-free water. Samples and reagents were all handled on ice. Reverse transcription was initiated on a 2720 Thermal Cycler (Applied Biosystems, Life Technologies, Paisley, UK) by incubation for 10 min at 25 °C to anneal primers, extension for 120 min at 37 °C, RT denaturation for 5 min at 85 °C, and the samples were then stored at 4 °C.

2.7.2 TaqMan® MicroRNA Reverse Transcription kit

The TaqMan® MicroRNA Reverse Transcription kit (Applied Biosystems, Life Technologies, Paisley, UK) was used to synthesise cDNA from the miRNA within the total RNA, using the stem-loop reverse transcription method (see Chapter one section 1.8.3.1) for miRNA transcripts, following the manufacturer's guidelines [152]. Each reaction was set up as described in Table 2.2 below.

Table 2.2: **The components of each RT reaction using the TaqMan microRNA RT kit.**

Component	Volume (µL)
100 mM dNTPs (with dTTP)	0.15
TaqMan MicroRNA Assay Primer	3.00
MultiScribe™ Reverse Transcriptase, 50 U/µL	1.00
10X RT Buffer	1.50
RNase Inhibitor, 20 U/µL	0.19
Nuclease-free Water	4.16
Total RNA Extract + H_2O	5.00
Total reaction	15.00

RNA concentration was normalised to 10 ng/µL for all sample types. For the RT-qPCR controls, the RT- control was prepared including all the components above except the MultiScribe™ Reverse Transcriptase and negative controls (NTC) were prepared by including all the components above but replacing the RNA extract with RNase-free water. Reverse transcription was initiated on a 2720 Thermal Cycler (Applied Biosystems, Life Technologies, Paisley, UK) by incubation for 30 min at 16 °C to anneal primers, extension for 30 min at 42 °C, RT denaturation for 5 min at 85 °C, and the samples were then stored at 4 °C.

2.8 Quantitative real-time PCR (qPCR)

Real-time PCR was carried out using the TaqMan® Universal PCR Master Mix II Kit, with no AmpErase® UNG (Applied Biosystems, Life Technologies Paisley, UK) along with either TaqMan® Gene Expression Assay or MicroRNA Assay, following the manufacturer's protocol [153]. A reaction volume of 20 µL was prepared as described in Table 2.3 below.

Table 2.3: **The components of each real-time PCR reaction using the TaqMan® Universal PCR Master Mix II kit.**

Component	Volume (µL)
TaqMan Universal Master Mix II (2X)	10.0
TaqMan Gene Expression Assay (20X)/TaqMan MicroRNA Assay	1.0
cDNA template (1-100 ng)	4.0
Nuclease-free water	5.0
Total reaction	20.0

Amplification was performed using a Stratagene Mx3005P (Agilent Technologies, CA, USA). The parameter values used to program the thermal cycler were: 10 min at 95 °C for enzyme activation followed by 40 cycles of denaturation for 15 s at 95 °C and annealing for 60 s at 60 °C; extension occurs during the temperature ramp of the annealing and denaturation steps.

The efficiency of each TaqMan® assay was assessed by generating a standard curve of a particular assay using a serial dilution of cDNA (see Chapter one section 1.8.5). A 1:3 dilution series of human body fluids (blood, saliva and semen) cDNA in RNase-free water was prepared, with eight dilutions in the series ranging in concentration from 85 to 0.035 ng/µL for blood, 40 to 0.018 ng/µL for saliva and 60 to 0.027 ng/µL for semen, using the same sample type for each body fluid-specific assays.

2.9 Data analysis

Reverse transcription quantitative PCR (RT-qPCR) data was analysed using *MxPro* (Agilent Technologies, CA, USA) to allocate a suitable fluorescence threshold. This threshold is set automatically by the qPCR instrumentation using the baseline threshold method, which is 10 times the standard deviation of fluorescence variation over PCR cycles 5 to 9. Raw data from the spectrophotometry and C_q values from the RT-qPCR reactions were analysed in *Microsoft Excel 2016* to present basic data and line graphs. *Minitab Express* (version 1.5.0) and *Minitab®17* (version 17, both Minitab® Inc., State College,

PA, USA) were used for statistical analyses, including the Anderson-Darling normality test, one-way analysis of variance (ANOVA) and regression analysis. For the nonparametric statistical data, Kruskal-Wallis and Dunn's multiple comparison analysis were performed with the R software package (version 1.3.5) [154]. RT-qPCR data were also analysed using *GenEx* statistical software (version 5.4.4; BioEPS GmbH, Munich, Germany), which normalises the raw C_q values against the determined efficiency of each RNA marker. These RT-qPCR efficiencies were assessed by generating a standard curve of a particular assay using *Microsoft Excel 2016* (see section 2.8 above). Furthermore, the software also determines the stability of RNA markers using the *geNorm* software package that is embedded in *GenEx* statistical software (version 4.3.7; BioEPS GmbH, Munich, Germany).

2.9.1 Relative expression ratio (RER)

The relative expression ratio was obtained by dividing the efficiency-corrected C_q values of the less stable RNA marker by the efficiency-corrected C_q values of the more stable RNA marker, or by dividing the C_q values of the body fluid-specific marker by the C_q values of a reference gene, as shown in the equations (1) and (2) below.

$$RER = \frac{Cq\ of\ less\ stable\ marker}{Cq\ of\ more\ stable\ marker} \qquad \text{(Equation 1)}$$

$$RER = \frac{Cq\ of\ body\ fluid\ marker}{Cq\ of\ reference\ gene} \qquad \text{(Equation 2)}$$

2.9.2 Relative quantity over time ($2^{-\Delta Cq}$/$Log2^{-\Delta Cq}$)

The relative quantity over time was calculated using the corrected C_q values to obtain the $2^{-\Delta Cq}$ of different RNA markers, by subtracting the C_q value of the less stable amplicon (target gene) from the more stable amplicon (reference gene) to obtain ΔC_q [59, 139]. Depending on the distribution of the data, the ΔC_q is converted into $2^{-\Delta Cq}$ or Log $2^{-\Delta Cq}$ values for statistical analysis.

$$\Delta C_q = C_{q\ \text{target gene}} - C_{q\ \text{reference gene}} \qquad \text{(Equation 3)}$$

Chapter Three: Analysis of RNA in body fluid stains: Method development

3.1 Introduction

The main aim of this thesis was to examine the degradation behaviour of different RNA molecules in body fluid stains. To accomplish this aim, some preliminary work was carried out in order to develop methodological strategies that could be used to address the objectives of this project.

One purpose of this method development was to compare the outcomes of two different RNA extraction procedures on different body fluid types to determine which extraction method produced a high yield of RNA using small volumes, to mimic the difficult nature of the biological samples that are commonly found at crime scenes in small quantities and degraded states. For a blood sample, only 20 µL of blood was used in both methods, while for saliva and semen samples, two different volumes of fluids were used (20 µL and 50 µL) and compared. TRI Reagent® (Sigma, Gillingham, UK) and miRNeasy Mini Kit (Qiagen, Manchester, UK) were applied in order to compare the yield of total RNA extracted from fresh body fluids deposited on cotton swabs.

Moreover, the quality of the extracted total RNA was measured using Agilent 2100 Bioanalyzer, which is used as a screening indicator to determine RNA integrity. This step was conducted to investigate whether the obtained RNA quality should be taken into account to decide whether the sample should be included for further RT-qPCR analysis or discarded. Alternatively, the measured quality of the purified total RNA may not represent the state of the target RNA markers, suggesting that Bioanalyzer data is not a suitable method for determining whether forensic samples can be analysed further.

Another purpose was to perform quality control assessment of the TaqMan® assays selected for the study, where all TaqMan® assays targeted towards the

rRNAs, mRNAs and miRNAs shown in Table 3.1 and Table 3.2 were assessed to see if they could be successfully used to amplify cDNA samples. These markers have been selected from relevant literature [54-57, 64-66, 76, 77, 81, 82, 155, 156] for the current project. Each assay has its own unique ID given by Applied Biosystems. There are a number of criteria that should be met to permit the selection of the predesigned TaqMan Gene Expression Assays for mRNA markers, Chapter four and Chapter five describes the criteria for RNA specific markers for blood, saliva and semen respectively.

Table 3.1: **Characteristics of the TaqMan® Gene Expression Assays.**

RNA target	Applied Biosystems TaqMan® assay ID	Amplicon length (nt)	Assay location*
ACTB	Hs99999903_m1	171	53
GAPDH	Hs02758991_g1	93	752
18S	Hs99999901_s1	187	604
HMBS	Hs00609296_g1	69	1070
HBA	Hs00361191_g1	156	158
HBB	Hs00758889_s1	95	511
PRM1	Hs00358158_g1	99	205
PRM2	Hs04187294_g1	73	388
SEMG1	Hs00268141_m1	82	1525
HTN3	Hs00264790_m1	136	302
STATH	Hs00162389_m1	90	165

* Refers to the nucleotide location that is the midpoint of the target region.

Table 3.2**: Characteristics of the TaqMan® MicroRNA Assays.**

miRNA target	Applied Biosystems TaqMan® assay ID	Target sequence
U6	001973	GTGCTCGCTTCGGCAGCACATATACTAA AATTGGAACGATACAGAGAAGATTAGCA TGGCCCCTGCGCAAGGATGACACGCAA ATTCGTGAAGCGTTCCATATTTT
miR16	000391	UAGCAGCACGUAAAUAUUGGCG
miR451	001141	AAACCGUUACCAUUACUGAGUU
miR205	000509	UCCUUCAUUCCACCGGAGUCUG
miR658	001513	GGCGGAGGGAAGUAGGUCCGUUGGU
miR10b	002218	UACCCUGUAGAACCGAAUUUGUG
miR891a	002191	UGCAACGAACCUGAGCCACUGA

3.1.1 Purification of RNA from body fluid stains using the TRI Reagent® and miRNeasy Mini Kit: A comparison study

Following the collection of body fluids on cotton swabs, the next step involved the extraction and purification of RNA molecules. Cotton swabs are the most common technique for body fluid collection, whether from a crime scene or from trace evidence recovered from suspects or victims. In some cases, the item itself may be collected for the identification of body fluids directly, such as cloth and tissues. Cotton swabs were selected as the method to be used in this project for the reason that they are easy to use and process, have guaranteed sterility, and are suitable for long-term storage.

3.1.2 Total RNA quality using Agilent 2100 Bioanalyzer and RT-qPCR performance

The quality of total RNA can be assessed using the Agilent 2100 Bioanalyzer, which was described in Chapter one (section 1.8.2.2). By applying this technology, the size distribution of recovered RNA fragments can be examined, and it can also measure RNA degradation by generating an RNA integrity number (RIN). The higher the RIN number the better the quality of the extracted RNA (1 = poor quality or degraded RNA, 10 = excellent quality or intact RNA) [110]. The 2100 Bioanalyzer experiment is usually conducted to examine the degree of the degradation level of the purified total RNA (i.e. the state of the transcriptome as a whole) to determine whether the sample is intact or too fragmented for further analysis. The work in this Chapter investigated whether the measured RNA quality by Agilent 2100 Bioanalyzer correlates with the obtained data on gene expression analysis measured by RT-qPCR for the selected target RNA markers, and whether it is a suitable method for determining the state of forensic samples.

3.1.3 TaqMan® assays quality control assessment

The selected assays included three blood-specific mRNA markers (*HBA*, *HBB*, and *HMBS*), two blood-specific miRNA markers (*miR16* and *miR451*), two saliva-specific mRNA markers (*STATH* and *HTN3*), two saliva-specific miRNA markers (*miR205* and *miR658*), two sperm-specific mRNA markers (*PRM1* and *PRM2*) and one semen fluid mRNA marker (*SEMG1*), two semen specific miRNA markers (*miR10b* and *miR891a*), along with reference genes (*18S* rRNA, *ACTB* mRNA, *GAPDH* mRNA and *U6* snRNA). TaqMan® chemistry was selected to be explored in this project due to its high specificity, especially the microRNA assays, which can discriminate between related miRNA molecules differing in sequence by as little as one nucleotide [157].

The quantitative range of the selected assays was assessed with the development of a standard curve. Then, the amplification efficiency of each assay was determined to show that efficiency was close to 100%. This step was also performed to confirm that the assays do not amplify any genomic DNA (gDNA) by running controls, and also to assess their precision. Determining all of these can confirm that the quantification values produced by a given assay accurately reflect variation in expression level. This is important to ensure that gene expression comparison could be carried out using relative quantification of different assays.

3.2 Material and Methods

3.2.1 Sample collection

Blood samples were collected onto sterile cotton swabs using disposable Unistik 3 comfort lancets. 20 µL of blood samples were pipetted on to the swabs. Both saliva and semen samples were directly deposited into a small container and then different volumes (20 µL and 50 µL) were pipetted onto swabs. Participants were asked not to eat, drink or smoke for at least one hour prior to saliva samples being collected. All swabs were allowed to dry at room temperature for 15 minutes. Another group of blood sample were aged for 4, 8 and 12 days to be used for Agilent 2100 Bioanalyzer quality assessment. The experimental procedures were approved by the Department of Pure and Applied Chemistry Departmental Ethics Committee and signed consent sheets were obtained from each donor after they had read a Participant Information Sheet (PIS) (see Appendix A1).

3.2.2 Total RNA extraction and reverse transcription

RNA extraction was carried out using TRI Reagent® and miRNeasy Mini Kit as described in section (2.4.1). RNA was eluted into RNase-free water in two different volumes; 50 µL and 20 µL. The TURBO DNA-free™ Kit was used to treat the extracted RNA to remove genomic DNA (section 2.5). The reverse

transcription reactions were carried out using High-Capacity cDNA Reverse Transcription Kit for mRNA and rRNA markers, and TaqMan® microRNA Reverse Transcription Kit for miRNA markers (Applied Biosystems) as described in sections 2.7.1 and 2.7.2.

3.2.3 Total RNA quantification

The NanoDrop 1000 Spectrophotometer was used to check the quantity of the total RNA present in the samples by following the manufacturer's instructions and using 1.5 µL of the extracted sample (see section 2.6.1).

3.2.4 Total RNA quality

An Agilent 2100 Bioanalyzer was used to assess the quality of the recovered RNA from blood samples aged for 4, 8 and 12 days. The RNA 6000 Pico Kit was used on this platform following the manufacturer's instructions described in section 2.6.2.

3.2.5 Quantitative real-time PCR (qPCR)

All TaqMan® assays were run in singleplex assays, following the procedure described in section 2.8. Amplification was performed using a Stratagene *MxPro*.

3.2.6 TaqMan® assays efficiency

The efficiency of each assay was calculated by constructing a standard curve generated by using a series of diluted standards of amplified fragments of known concentration for each assay marker, using a 1:3 dilution series with eight points. These dilutions were run in duplicate and the measurements of the C_q values of these standards were plotted against the logarithm of their concentration to determine the slope of the standard curve and the R^2 values

($R^2 \geq 0.98$ is desirable), which were used to calculate the efficiency as shown in the equation below:

$$\text{Efficiency} = 10^{(-1/\text{slope})}$$

3.2.7 Data analysis

The data generated from RT-qPCR was analysed using *MxPro* and *GenEx* software (version 5.4.4) was used for efficiency correction of the raw data. *Microsoft Excel 2016* was used to manipulate raw data from the spectrophotometry and C_q values from the RT-qPCR reactions and to present basic data and line graphs. *Minitab Express* (version 1.5.0) was used for statistical analysis, including the Anderson-Darling normality test and 2-samples t-test (see section 2.9).

3.3 Results and Discussion:

3.3.1 Purification of RNA from blood samples using TRI Reagent® and miRNeasy Mini Kit: A comparison study

3.3.1.1 Extraction methods comparison: Total RNA yield – blood

The main feature to be compared between the TRI Reagent® and miRNeasy Mini Kit extraction methods was the total quantity of RNA purified from blood, saliva and semen samples on swabs. The total RNA yield extracted from a given volume of body fluid deposited onto cotton swabs was quantified using a NanoDrop 1000 Spectrophotometer after the purified RNA was treated with DNase to remove any contaminating genomic DNA.

When comparing the total RNA purified from blood samples using these two different methods, it was found that the extraction method had a significant effect on RNA yield, as illustrated in Figure 3.1.

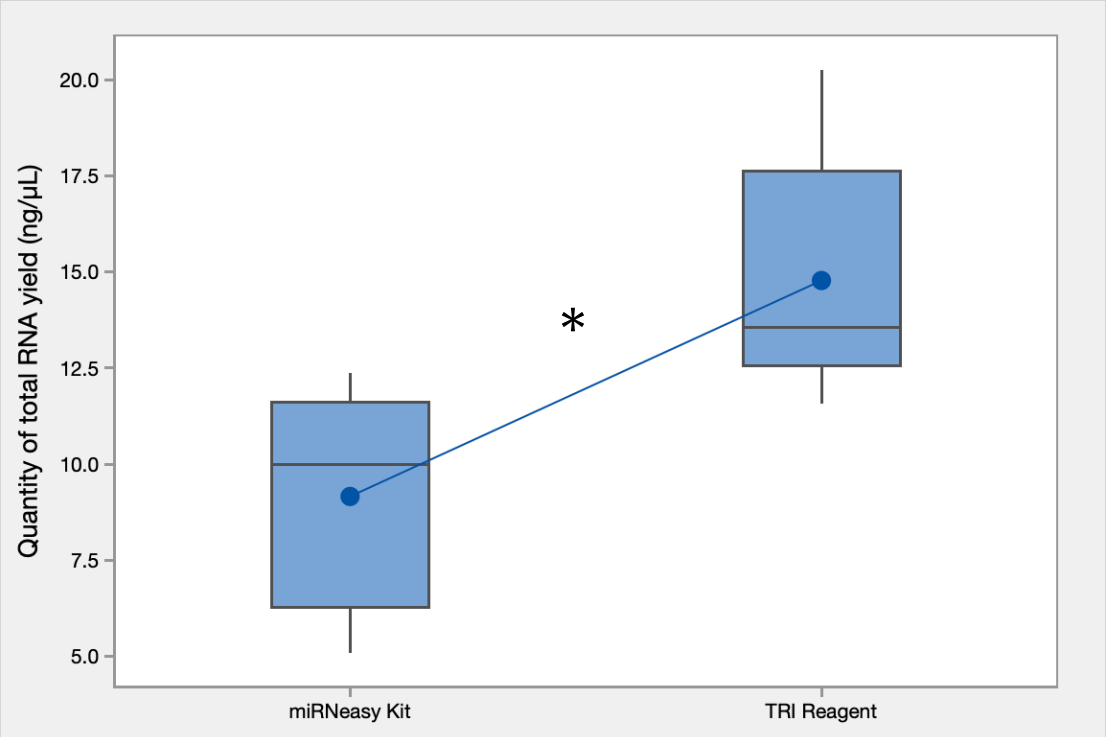


Figure 3.1: **Boxplot showing the quantity of total RNA recovered per 20 µL of blood samples using two RNA extraction methods: TRI Reagent® and miRNeasy Mini Kit** (n = 5). TRI Reagent® showed a significantly higher RNA yield using a 2-sample t-test ($t = -2.87$, $p < 0.05$).

All quantification data were normally distributed when assessed for normality using the Anderson-Darling normality test (TRI Reagent® $p = 0.1652$, miRNeasy Kit $p = 0.7250$). The mean quantity of total RNA extracted from blood samples using TRI reagent® method and the miRNeasy Kit was compared using a 2-sample t-test, which confirmed that TRI Reagent® recovered a significantly higher quantity of RNA from a 20 µL blood sample ($t = -2.87$, $p = 0.02392$).

The manufacturer of the miRNeasy Mini Kit (Qiagen) claims that the reason behind the low quantities of RNA is that this kit is not specific for blood extraction and it works better with other cell and tissue types (personal communication). They suggest processing the blood sample with Red Blood Cell (RBC) Buffer initially, then proceeding with the kit protocol.

Overall, these findings showed that the TRI Reagent® method produced higher yields of total RNA from blood samples when compared to the miRNeasy Mini Kit, an outcome which is supported by other published works. Kim et al. (2014) extracted RNA from frozen blood samples using TRI Reagent® and two different commercial column kits (PAXgene and NucleoSpin) [158]. They quantified total RNA by UV-visible spectrophotometry, similar to that used in this work, and found that TRI reagent® yielded 1.7- to 5.0-fold more RNA than the column-based kits. Similar results were also found when extracting RNA from whole blood using TRI Reagent® by Jakovljevic et al. [159]. The outcomes of these studies are concordant with data shown in this work, which demonstrated that higher RNA recovery is produced when using TRI Reagent®.

In an attempt to increase the concentration of the extracted RNA, the purified RNA was resuspended in 20 µL of RNase-free water instead of 50 µL. Figure 3.2 illustrates the concentration of extracted RNA using TRI Reagent® relative to the miRNeasy kit in two different elution volumes. Both methodologies provided a significantly higher RNA concentration when the extract is suspended in 20 µL RNase-free water. The quantification data from blood samples extracted using the TRI Reagent® and miRNeasy Mini Kit and eluted into different elution volumes (50 µL and 20 µL) were normality distributed when assessed for normality using the Anderson-Darling normality test ($p > 0.05$). The median/mean quantities of purified RNA eluted in two different volumes were compared using the 2-sample t-test ($t = -3.97$, $p = 0.0107$) for TRI Reagent® and ($t = -4.64$, $p = 0.0435$) for miRNeasy Mini Kit, which found a significant increase in RNA concentration when the elution volume was reduced in both the TRI Reagent® and miRNeasy Kit extraction methods.

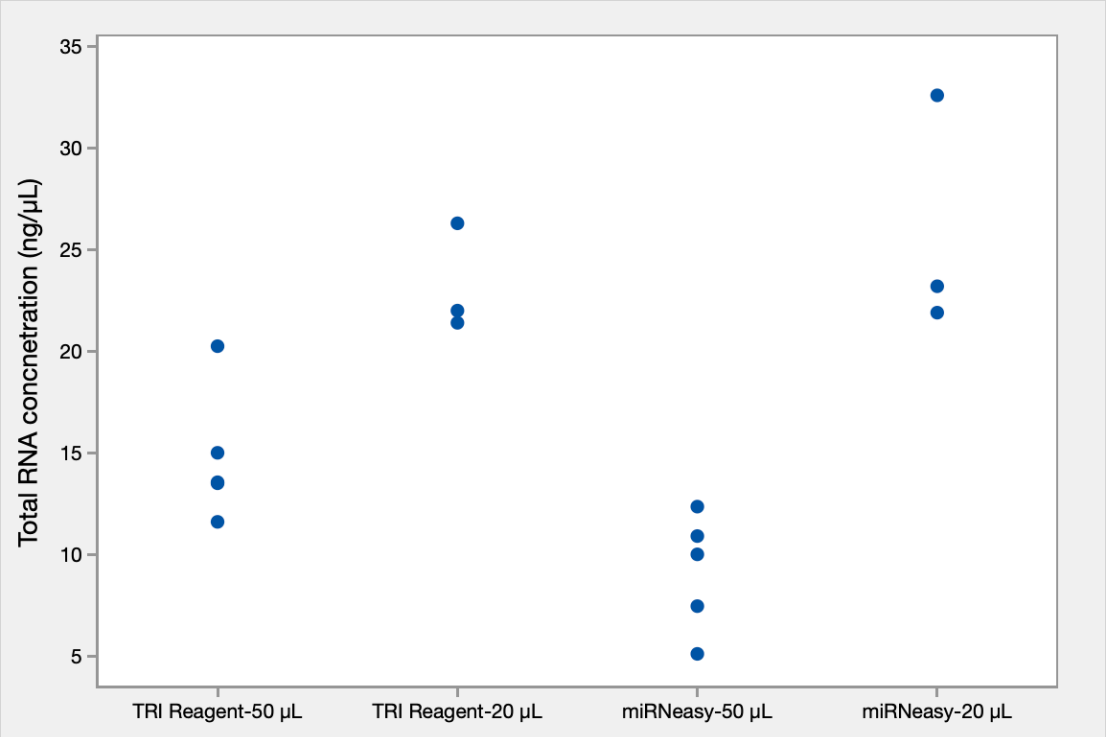


Figure 3.2**: Total RNA concentration recovered from blood samples in two different elution volumes; 50 µl (n = 5) and 20 µL (n = 3) using two RNA extraction methods: TRI Reagent® and miRNeasy Mini Kit.** Both purification methods demonstrated a significant increase in RNA concentration when elution volume was reduced using either a 2-sample t-test; $p < 0.05$.

3.3.1.2 Extraction methods comparison: Total RNA purity - blood

The second feature to be compared between these two extraction methods was the purity of the extracted total RNA. In UV-visible spectrophotometry (NanoDrop-1000) RNA absorbs UV-light at 260 nm, whereas contaminants such as phenol absorb at 230 nm and proteins absorb at 280 nm. Therefore, the ratios of absorbance $A_{260}/_{280}$ and $A_{260}/_{230}$ were used to indicate total RNA purity. Any abnormalities in either ratio usually give an indication of sample contamination either by proteins or other reagents such as phenol.

The ratios of absorbance $A_{260}/_{280}$ and $A_{260}/_{230}$ data were normally distributed when applying Anderson-Darling normality test ($p > 0.05$). The statistical comparison of the ratios has shown that, blood samples extracted using TRI

Reagent® exhibited significantly higher purity than samples obtained using the miRNeasy Mini Kit when performing 2 sample t-test, (A_{260}/A_{280}; t = -3.22, p = 0.0324, A_{260}/A_{230}; t = 4.76, p = 0.0031) (Figure 3.3 and Table 3.3). However, these values are still below the generally accepted level of pure RNA which is ~2.0 for both ratios. The low $A_{260/280}$ in samples extracted using the miRNeasy Mini Kit is indicative of protein contamination, which could be due to unsuccessful disruption of blood samples in the lysis step or incomplete separation/transfer of the upper aqueous layer after layer separation. Another reason that could be behind the low $A_{260/280}$ ratio is the concentration of the total RNA extracted, as a very low concentration (< 10 ng/µL) may produce a low $A_{260/280}$ ratio [160].

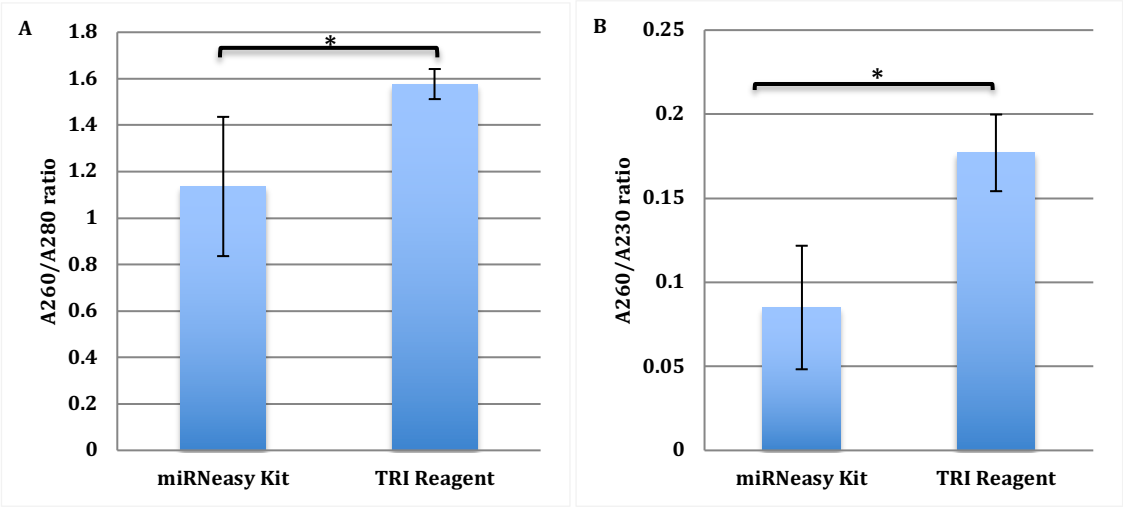


Figure 3.3**: The absorbance ratios of extracted RNA from blood samples using two RNA extraction methods: miRNeasy Mini Kit and TRI Reagent®**. (A) A_{260}/A_{280} and (B) A_{260}/A_{230}, where the graphs show mean ± standard deviation (n = 5). With a 2-sample t-test, the difference in mean ratio was found to be significant ($p < 0.05$).

Table 3.3: **The total RNA yield, A_{260}/A_{280} and A_{260}/A_{230} ratios.** Different blood samples were extracted using two different extraction methods and resuspended in 50 µL of RNase-free water (n = 5).

	$A_{260/280}$		$A_{260/230}$	
Sample	**miRNeasy Kit**	**TRI Reagent**	**miRNeasy Kit**	**TRI Reagent**
1	1.37	1.61	0.14	0.19
2	1.12	1.60	0.08	0.18
3	1.27	1.48	0.06	0.14
4	0.63	1.65	0.05	0.18
5	1.30	1.60	0.10	0.20

Additionally, when comparing the obtained ratios of absorbance $A_{260}/_{280}$ and $A_{260}/_{230}$ in the 20 µL eluted RNA samples, no significant difference was recorded ($p > 0.05$). Among these two extraction methods, the TRI Reagent® yielded the highest quantity of RNA with higher purity compared to RNA extracted using the miRNeasy Mini Kit, despite the fact that one of the potential disadvantages of the TRI Reagent® method is the possibility of residual contamination. Based on these findings, it can be concluded that extraction using TRI Reagent® is a more effective way to recover abundant and high-quality RNA from blood samples.

3.3.1.3 Extraction methods comparison: Reference genes quantity – blood

The expression level of three reference genes, *ACTB*, *GAPDH,* and *18S* were analysed, and their C_q values were recorded to estimate the quantity of these three RNAs recovered from blood samples extracted using TRI Reagent® and the miRNeasy Mini Kit. This step was done to determine whether the expression levels of these genes reflected the total RNA quantities determined in section 3.3.1.1. The UV-visible spectrophotometry quantifies total RNA based on the absorbance of the UV light at 260 nm by the nucleobases without distinguishing whether it is intact (long) or fragmented RNAs (short). Therefore, it would be necessary to prove whether UV-visible spectrophotometry can actually give an indication of the RNAs level in the extracted samples.

As shown in Figure 3.4, the mean C_q values for samples extracted using the TRI Reagent® method and the miRNeasy Mini Kit were: 26.5 and 31.4 for *ACTB*, 30.3 and 38.6 for *GAPDH*, and 20.0 and 29.6 for *18S*, respectively. RNA extracted using the miRNeasy Mini Kit has higher C_q values (~ 4.8- to 9.6-fold) than those extracted using TRI Reagent®, clearly demonstrating that the quantity of RNA extracted utilizing TRI Reagent® is higher than the quantity of RNA extracted utilizing miRNeasy Mini Kit.

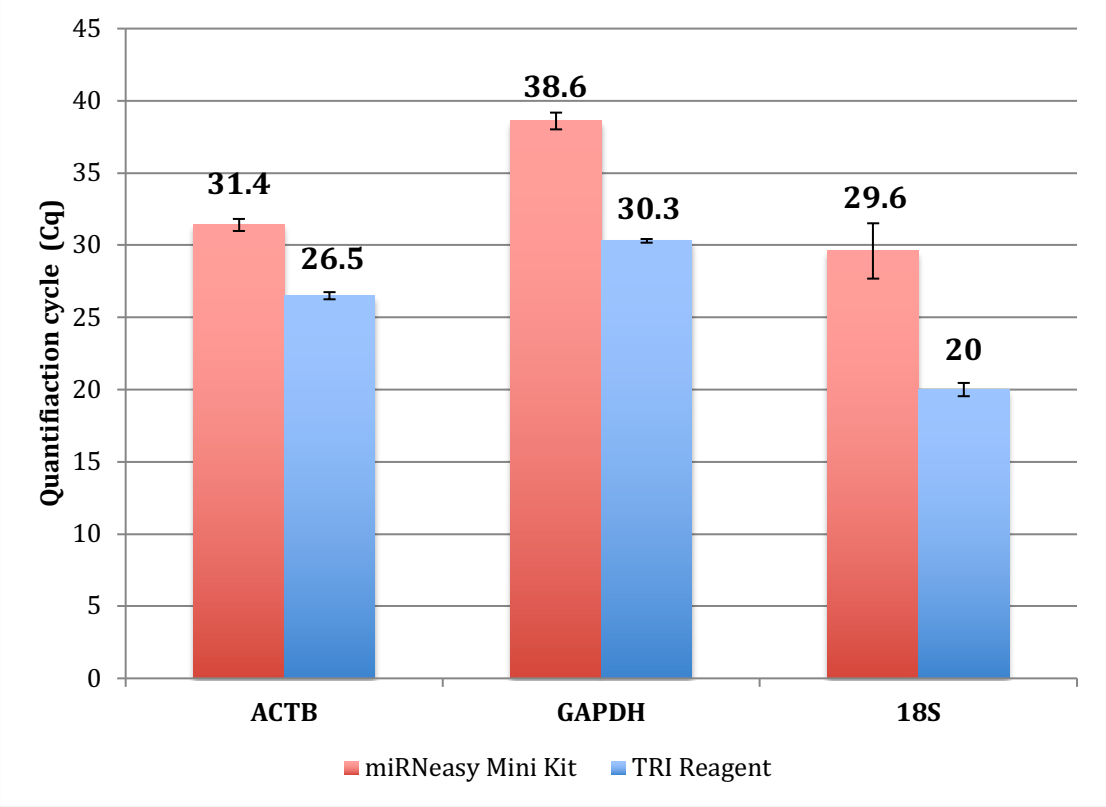


Figure 3.4: **The Quantification cycle (C_q) values of three reference genes assayed from RNA purified using miRNeasy Mini Kit and TRI Reagent® and amplified using RT-qPCR.** Data represent mean ± standard deviation for n = 3.

3.3.2 Purification of RNA from saliva and semen samples using TRI Reagent® and miRNeasy Mini Kit: A comparison study

3.3.2.1 Extraction methods comparison: Total RNA yield – saliva and semen

For saliva and semen samples, different volumes of sample (20 µL and 50 µL) were extracted from swabs using the two methods. When comparing the total RNA purified from saliva and semen samples, it was found that each method produced different RNA concentrations for the different sample volumes, as illustrated in Figure 3.5. The TRI Reagent® method recovered more RNA than the miRNeasy Mini Kit in both sample types and both volumes. The highest RNA yield purified from saliva samples was from the 20 µL samples using TRI Reagent® method, however, the standard deviation for these samples showed

a large range of total RNA recovered, as can be clearly seen in Figure 3.5. On the other hand, the miRNeasy Mini Kit recovered a higher concentration of total RNA in the 50 µL saliva samples when compared to 20 µL samples. A statistical test was not performed due to the small number of samples for each volume.

One reason for the high variation in the saliva results could be due to the heterogeneity of the saliva sample itself, as it composed of proteins, small organic substance, epithelial cells and bacteria [15]. All these compounds vary from one individual to another leading to a high variation in the results. Moreover, a significant level of bacteria is present in saliva, which may increase the degradation rate of the nucleic acids. The level of these bacteria differs from one individual to another, which may lead to such variation in RNA yield.

These results are consistent with a study conducted by Alves et al. (2016) to compare different RNA extraction techniques on buccal mucosa samples, who found that Trizol® (Life Technologies) which is similar to the TRI Reagent® method produced the highest RNA yield when compared to other commercial kits available, such as the RNeasy Mini Kit (Qiagen) [161].

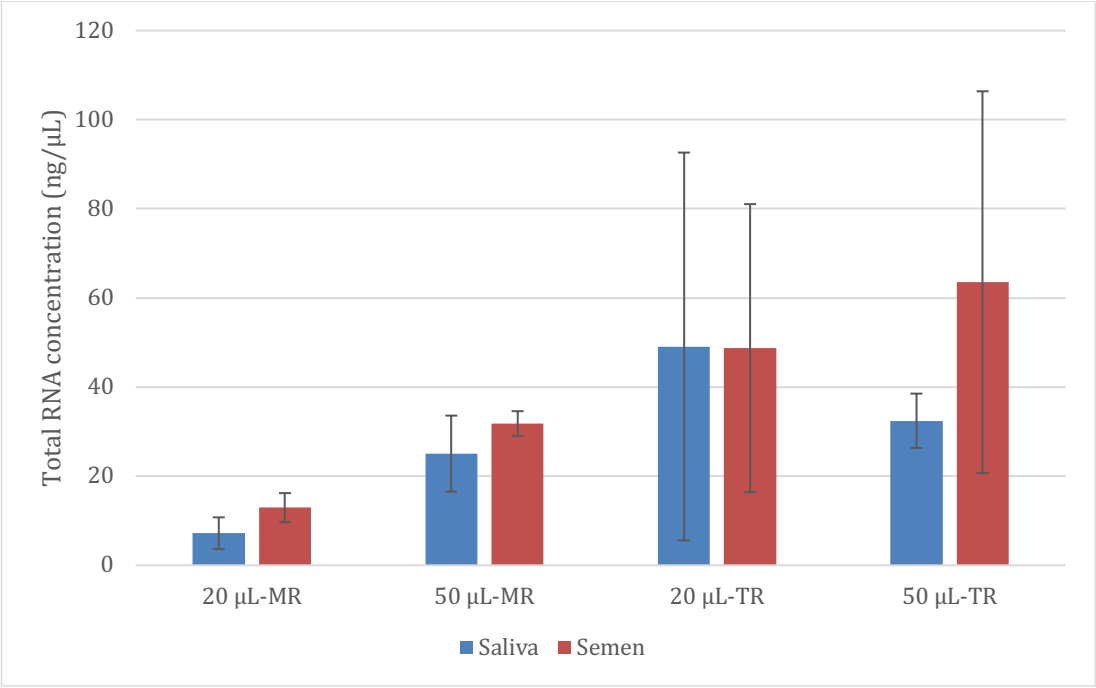


Figure 3.5: **Total RNA yield from saliva and semen samples.** Bars represent the concentration of total RNA recovered from saliva samples using two RNA extraction methods: miRNeasy Mini Kit (MR) and TRI Reagent® (TR) Data represent mean ± standard deviation for n = 3.

When comparing the total RNA yield from semen samples, it was found that the highest quantity of RNA was extracted from 50 µL of semen using TRI Reagent®. Both TRI Reagent® and miRNeasy Mini Kit produced a higher concentration of total RNA from semen samples with a volume of 50 µL than a volume of 20 µL. Unlike the miRNeasy Mini Kit, there was a large range of variation in the total RNA extracted from semen samples using TRI Reagent® as is clearly shown in the standard deviation bars in Figure 3.5. A possible explanation for this variation might be due to the extraction method itself, which may introduce some contaminants that could lead to over- or underestimating the concentration of the purified RNA.

3.3.2.2 Extraction methods comparison: Total RNA purity – saliva and semen

Similar to blood samples, the purity of saliva and semen samples was assessed by measuring the ratios of A_{260}/A_{280} and A_{260}/A_{230}. The purity of total

RNA extracted from both saliva and semen samples using the miRNeasy Mini Kit increased as the volume of the sample increased (Figure 3.6). In the case of TRI Reagent®-extracted saliva samples the purity of RNA extracted from 50 µL was slightly lower than for the 20 µL samples, but not in semen samples where the A_{260}/A_{280} ratio was higher in the 50 µL samples than the 20 µL samples.

Overall, when comparing the purity of RNA extracted from saliva and semen samples, the total RNA extracted with TRI Reagent® had higher A_{260}/A_{280} ratios than samples extracted with the miRNeasy Mini Kit, for both sample volumes (20 µL and 50 µL).

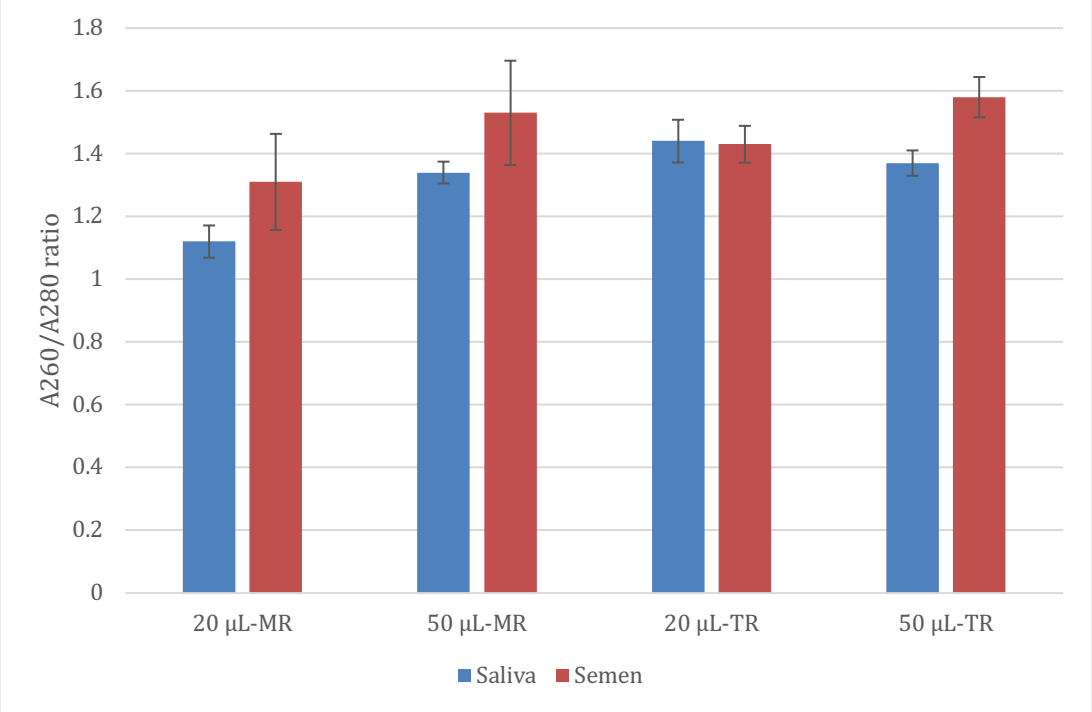


Figure 3.6**: The mean absorbance ratio A_{260}/A_{280} of extracted RNA from saliva and semen samples using two RNA extraction methods**. 20 µL and 50 µL of saliva and semen samples were extracted using miRNeasy Mini Kit and TRI Reagent®. Data represent mean ± standard deviation for n = 3.

Similar results were shown when comparing the A_{260}/A_{230} (Table 3.4), higher ratios were recorded in saliva and semen samples extracted with TRI Reagent® than samples extracted with miRNeasy Mini Kit, for both sample

volumes (20 µL and 50 µL). Both extraction methods had higher A_{260}/A_{230} ratio in 50µL than 20 µL samples.

Table 3.4: **The mean ratio of A_{260}/A_{230} of saliva and semen samples.** 20 µL and 50 µL of saliva and semen samples were extracted using miRNeasy Mini Kit and TRI Reagent®.

	Saliva A260/A230		Semen A260/A230	
Sample volume (µL)	**miRNeasy Kit**	**TRI Reagent**	**miRNeasy Kit**	**TRI Reagent**
20	0.16	0.20	0.09	0.23
50	0.18	0.21	0.21	0.24

These results also seem to indicate that the increased purity of the extracted RNA is due to the increase in the RNA yield. For example, the total RNA extracted from 20 µL saliva samples using TRI Reagent® yielded the highest total RNA quantity among all samples and the A_{260}/A_{280} ratio was the highest for these samples, indicating that the increasing concentration of purified RNA increases the purity RNA. Another example can be seen with the semen samples, where the highest RNA yield was in the 50 µL samples extracted using TRI Reagent® and the highest purity was seen in the same samples, again indicating that the purity of RNA increases with the extracted yield. Similar outcomes were found by Barragan et al. (2015) [162] when they extracted total RNA from semen samples, suggesting that the presence of inhibitors in RNA samples have more pronounced effects in samples with lower RNA yield, leading to a decrease in the purity.

Similar to blood samples, the purity of the extracted RNA from saliva and semen samples using both methods were below the accepted level of pure RNA which is ~2.0, which could be due to protein contamination or low total RNA concentration. The purity and the quality of biological samples that are collected from crime scenes are not ideal as well, as they may originate from

harsh environmental conditions, which may lead to various stages of degradation and contamination. Therefore, the obtained results would be acceptable for the current project.

3.3.2.3 Extraction methods comparison: Reference genes quantity – saliva and semen

Since the total RNA purified from all saliva samples using TRI Reagent® was higher than the samples extracted using miRNeasy Mini Kit, three reference genes (*ACTB*, *18S* and *GAPDH*) were analysed to see if their C_q values reflected these results. The C_q values for the reference genes amplified from RNA extracted from 50 µL saliva samples using TRI Reagent® were the lowest (i.e. they had the highest quantity) (Figure 3.7). These results are not concordant with the total RNA concentration obtained using TRI Reagent® method, as the highest total RNA concentration was extracted from the 20 µL saliva samples rather than the 50 µL samples.

The C_q values of the reference genes were similar in samples from the two different saliva volumes extracted using the miRNeasy Mini Kit. These results also are not concordant with the total RNA concentrations obtained, as the highest total RNA quantity was extracted from the 50 µL saliva samples rather than the 20 µL samples.

These unexpected outcomes might be caused by contamination, which could lead to an overestimation of total RNA yield in these two volumes for the TRI Reagent® method, where the total RNA concentration of 20 µL saliva samples was higher than 50 µL. Moreover, saliva samples are exposed to ribonucleases from bacteria more than any other type of body fluids such as blood and semen, which could lead to an increased rate of nucleic acid degradation, which may lead to inaccurate estimation of total RNA concentration using UV-visible spectrophotometry (NanoDrop-1000).

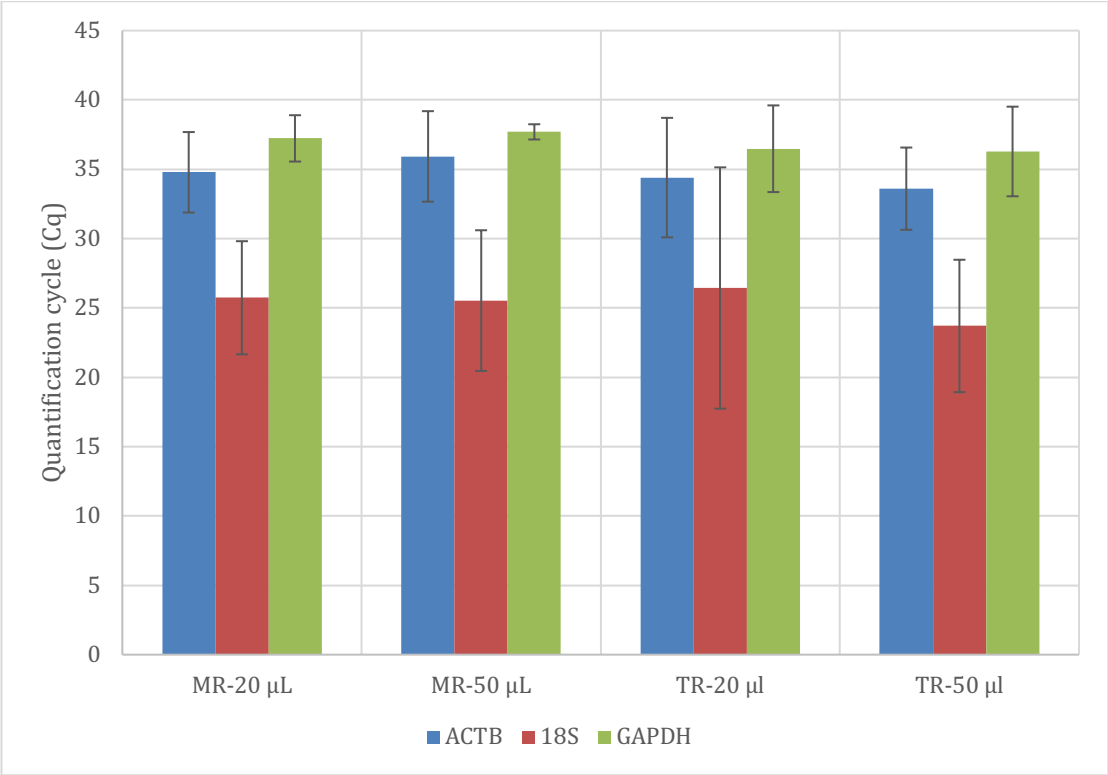


Figure 3.7**: Quantification cycle (C_q) values of three reference genes assayed from RNA purified from saliva samples using TRI Reagent® (TR) and miRNeasy Mini Kit (MR) and amplified using RT-qPCR.** The data represents mean ± standard deviation for n = 3 for each sample volume.

Comparable results were obtained when these three reference genes were amplified from the RNA extracted from semen samples (Figure 3.8). Similar to the saliva samples, the lowest C_q values for all three reference genes were obtained from RNA extracted from 50 µL semen samples using TRI Reagent® followed by the 20 µL samples extracted with the same method. The miRNeasy Mini Kit gave lower C_q values for *ACTB* and *GAPDH* in samples extracted from 50 µL semen compared to 20 µL, but not for *18S* rRNA, which had a lower C_q value in the 20 µL semen samples than the 50 µL samples.

Overall, the C_q values of the reference genes amplified from RNA samples extracted from semen using two different methods were concordant with the data on total RNA quantity.

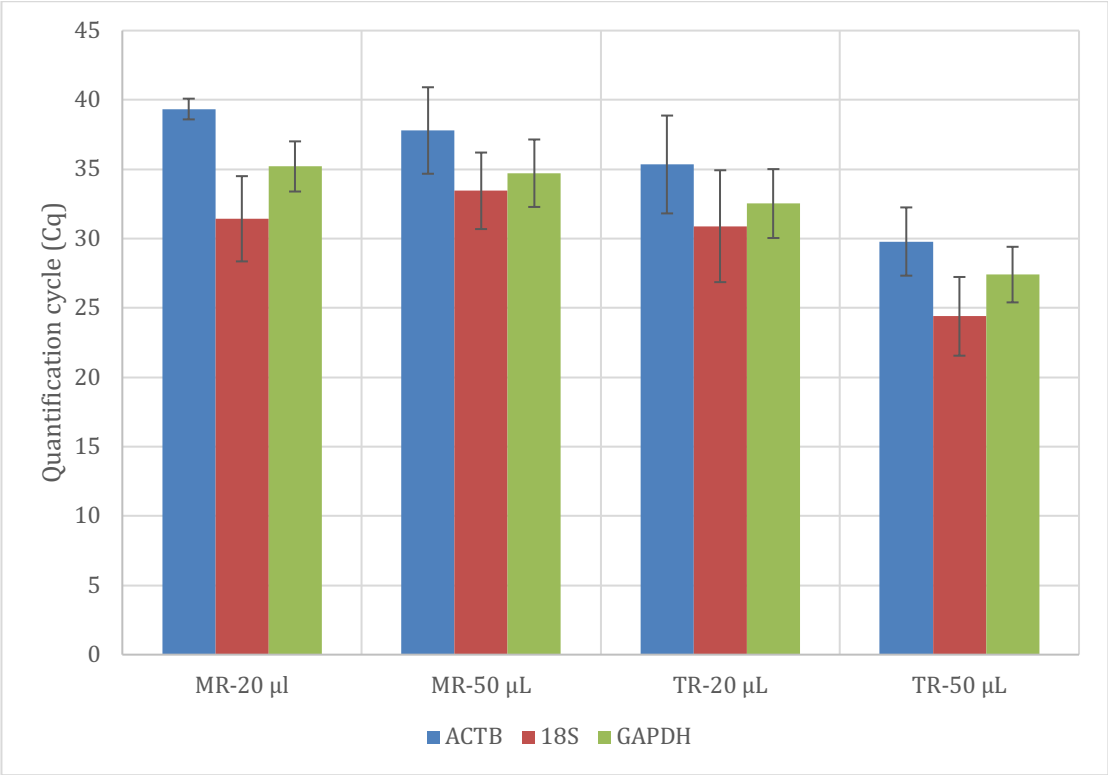


Figure 3.8**: Quantification cycle (C_q) values of three reference genes assayed from RNA purified from semen samples using TRI Reagent® (TR) and miRNeasy Mini Kit (MR) and amplified using RT-qPCR.** The data represents mean ± standard deviation for n = 3 for each sample volume.

3.3.3 Total RNA quality using Agilent 2100 Bioanalyzer

Another aspect investigated as part of this Chapter was RNA integrity. In gene expression studies, determining the RIN using the Agilent 2100 Bioanalyzer can give a good indication about the quality or the degradation level of the recovered RNA. The obtained RIN allows the researcher to decide whether to continue analysing the sample or to exclude it from further analysis, depending on whether it is intact. The Agilent 2100 Bioanalyzer method was applied herein to examine whether it can be applied to forensic samples (i.e. dried bloodstains) and if the RIN actually represents the degradation level of the target RNA markers. Furthermore, the relationship between RIN and C_q values was examined to determine whether there is any correlation between the RT-qPCR analysis and Agilent 2100 Bioanalyzer data.

This experiment was undertaken to determine the relationship between the quality of extracted RNA from fresh and aged bloodstains, measured using RIN, and different ageing time points. Four bloodstains were maintained at room temperature for up to 12 days, and total RNA was isolated at different time points (0, 4, 8 and 12 days). After RNA purification, the quality of the total RNA was measured in triplicate using the Agilent 2100 Bioanalyzer, and the data is illustrated in Table 3.5 and Figure 3.9. It can be seen clearly from the graph that there is no relationship between RIN and ageing time points, and instead that there is fluctuation in RIN values among ageing time points. It was expected that the RIN would decrease as the time increased, due to RNA degradation with time, as a lower RIN represents a more degraded sample.

Starting with the fresh samples, the maximum RIN was 2.40 which is still below the thresholds 5 and 3.95 that were suggested by the literature [104, 163], as samples below these thresholds considered to be too fragmented to be taken forward for further analysis. The RIN value have shown fluctuations as it decreased at day 4 and increased again at day 8 to drop back again at day 12. The coefficient of variation (CV%) values which was used to measure the variation among bloodstains aged for the same time period showed an increase with increasing ageing time points with the highest value (18.69%) at 12 days.

Table 3.5: **The RIN values of fresh and aged bloodstains.** The samples were measured in triplicate n = 3.

		RIN			
Samples	Readings	Day 0	Day 4	Day 8	Day 12
S1	R1	2.20	1.70	2.50	2.50
	R2	2.30	1.50	2.50	2.30
	R3	N/A	N/A	N/A	2.20
S2	R1	2.20	N/A	1.40	1.90
	R2	2.20	1.20	1.10	1.20
	R3	2.20	1.20	1.00	2.00
S3	R1	2.10	1.10	1.30	2.40
	R2	2.20	1.00	1.10	1.30
	R3	2.20	N/A	2.10	1.40
S4	R1	2.20	1.20	2.40	1.00
	R2	2.40	1.80	2.40	1.10
	R3	2.20	1.60	N/A	N/A
Mean		2.25	1.35	1.89	1.70
CV%		2.53	9.62	9.74	18.69

In concordance with this project, Fang et al. (2018) detected very low RIN values, ranging from 1.98 to 3.22, from total RNA isolated from bloodstains using small volume, while total RNA from 1 mL blood samples scored RIN values ranging between 6.08 to 7.39 [164]. Therefore, this may indicate that as the volume of blood samples decrease, the validity and accuracy of the Agilent 2100 Bioanalyzer decreases. In Fujimoto et al. (2018), the average RIN number for their examined forensically body fluid samples was consistently low after the samples were incubated at 31 °C for 30 mins [165]. Therefore, another explanation is bloodstains dry faster leading to faster RNA degradation.

Therefore, based on these findings, the RIN number established by a 2100 Bioanalyzer might be a good tool to assess RNA quality for standard laboratory samples but it should not be used for trace or forensic samples due to their limited in quantity and degraded states. The resolution of this method to quantify RNA degradation is limited when applied to heavily degraded RNA samples, where *28S* and *18S* peaks will hardly be present [107]. This was not the case with RNA samples only but also with DNA samples as the findings in Gorzkiewicz et al. work (2010) shown that the 2100 Bioanalyzer system is not appropriate for accurate determination of degraded DNA concentration from biological traces and their quality was difficult to be assessed [166].

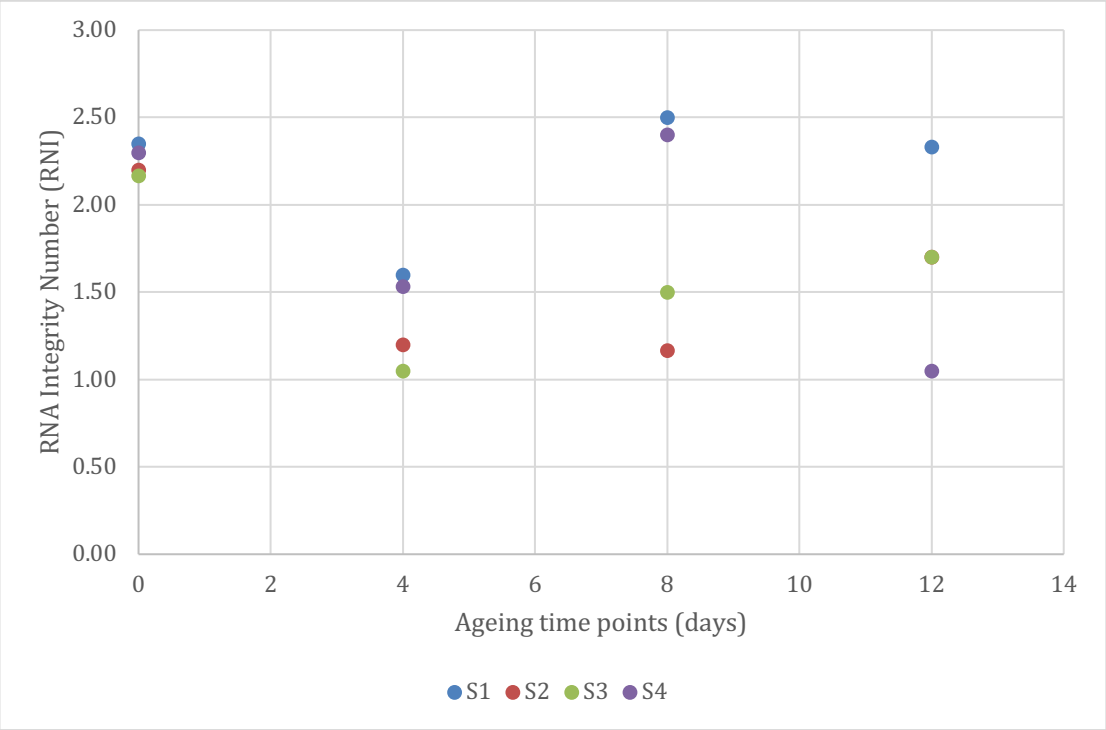


Figure 3.9: **The mean RIN values of total RNA recovered from bloodstains aged for up to 12 days.** Points represent the quality of RNA measured by the Agilent 2100 Bioanalyzer applying the RIN algorithm. Each sample represents the mean of n = 3. Error bars were removed for clarity

In order to examine the validity of the 2100 Bioanalyzer data and to see whether there was a relationship between RIN values and the RNA marker profiles to determine RNA degradation over time, two of the RNA markers

(*HBA* and *18S*) were analysed in the aged blood samples described above, using RT-qPCR. The gene expression profiles of the *HBA* and *18S* markers are shown in Figure 3.10A and Figure 3.10B, respectively. In both profiles, it can be seen that as the time point increases, the mean C_q values (i.e. degradation) increases. Therefore, it can be concluded that the RIN values determined from the bloodstains did not indicate the degradation status of the extracted RNA, as many low RIN samples generated low C_q values (higher marker quantity). For example, sample number 1 (S1) showed an increase in RIN from 2.30 at 0 days to 2.50 at 8 days meaning slightly higher quality of RNA, however, degradation level for both markers increased in the same time period. Therefore, RNA integrity value did not directly influence the obtained data from RT-qPCR.

Since the RIN data does not correspond to the observed RNA degradation and given that a high level of variation was seen between the replicates of a single sample, these results indicate that the Agilent 2100 Bioanalyzer cannot be used for small volume bloodstains. For such samples, Ingold et al. (2018) recommended that the Bioanalyzer method should not be used to quantify extracted RNA as it normally gives low RIN numbers, indicating poor RNA quality [55].

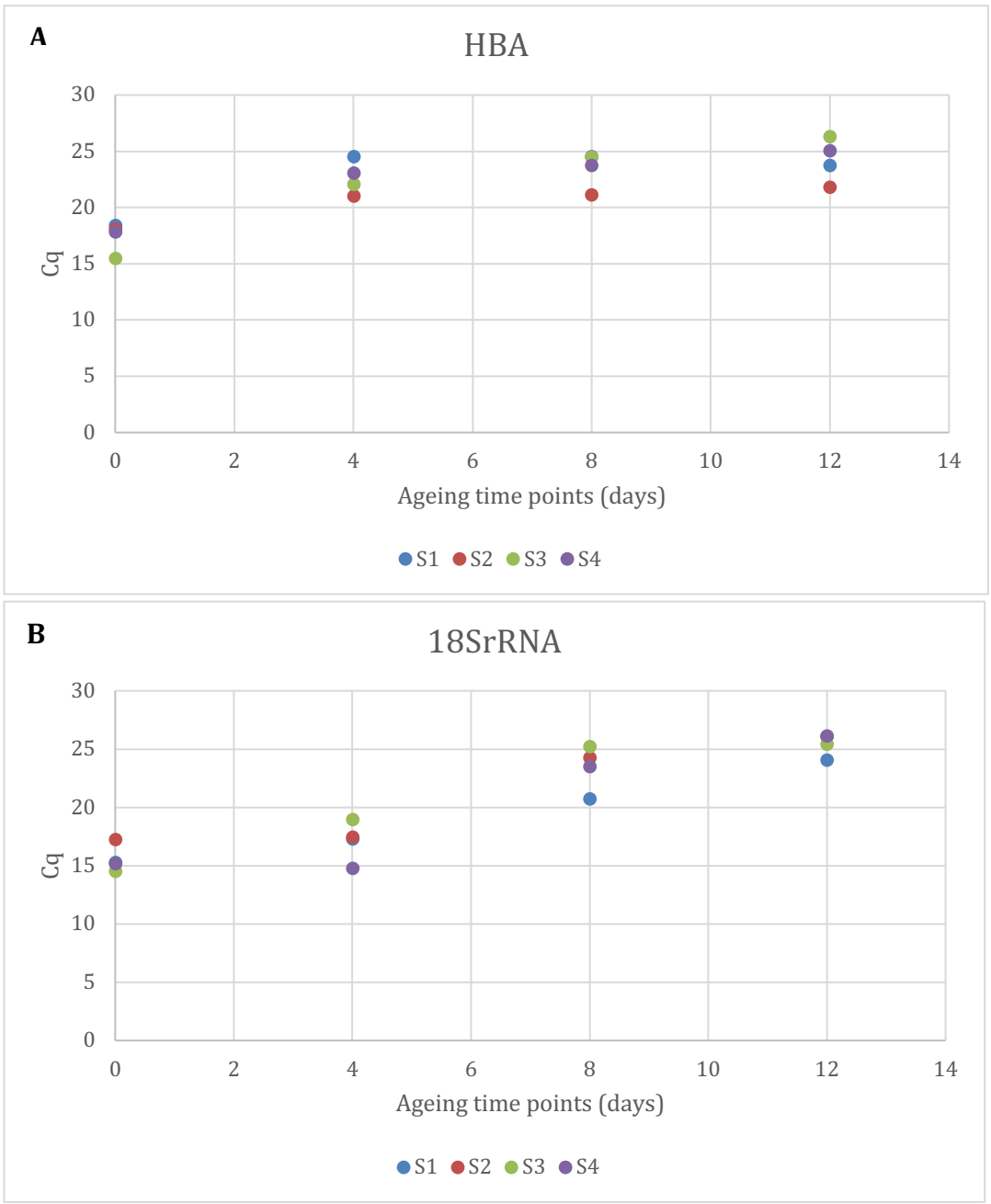


Figure 3.10: **The mean C_q values from gene expression assays of two RNA markers in aged bloodstains.** (A) *HBA* marker and (B) *18SrRNA* gene expression at 0, 4, 8 and 12 days. Each sample represents the mean of n = 2.

3.3.4 Quality control data

In line with MIQE guidelines [167] and for quality control purposes, it is essential that a number of qPCR assay features are characterised. These features include qPCR efficiency, the linearity of qPCR data, reproducibility

and negative control results. The quality control parameters should be obtained for all qPCR assays before utilising them on experimental samples.

3.3.4.1 TaqMan® assays standard curve

Both qPCR efficiency and linearity of qPCR data were assessed by generating a standard curve of each assay using a serial dilution of cDNA. In this study, TaqMan® assay efficiencies were measured using the C_q slope method, which involves generating standard curves (i.e. a dilution series of a template) of each assay (see section 1.8.5, amplification efficiencies) utilising the same sample and determining the C_q for each dilution.

The results of efficiency testing for TaqMan assays amplifying a range of rRNA, mRNA, miRNA and snRNA are shown in Table 3.6. Figure 3.11 shows an example of a standard curve for the *18S* rRNA assay, showing the value of the slope (-3.447) and the R^2 value (0.9921).

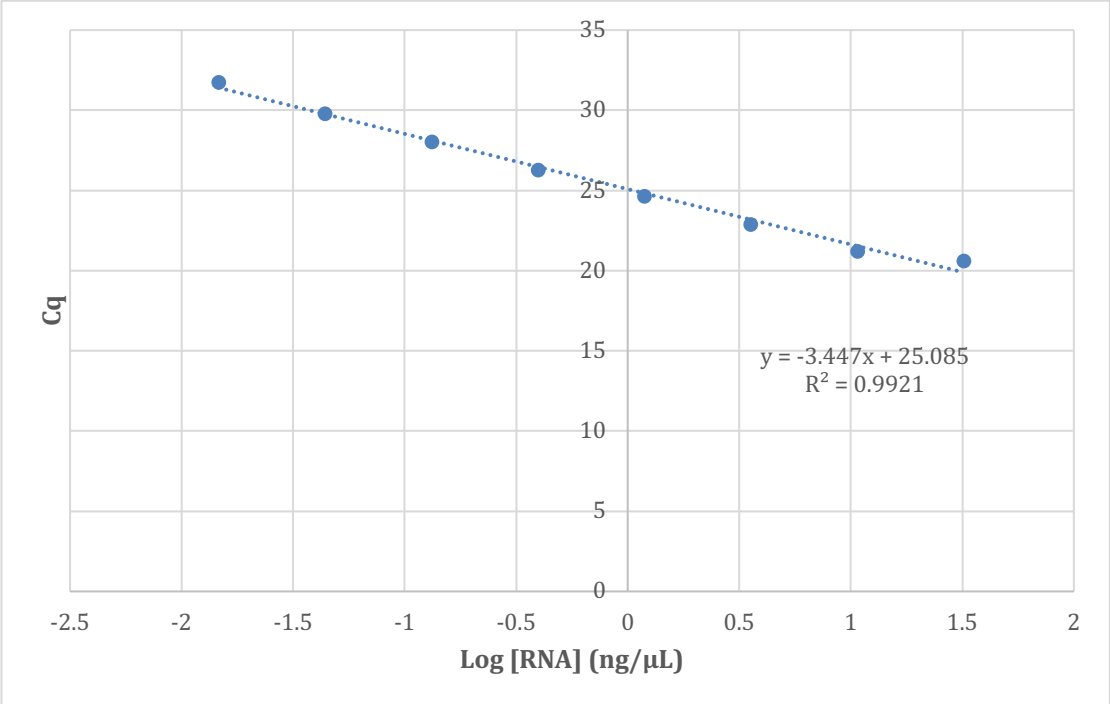


Figure 3.11**: Standard curve for the *18S* rRNA TaqMan® assay.** The curve was produced using a 1:3 dilution series of blood cDNA ranging from 85 ng/µL of starting RNA. Data points illustrate the mean C_q for n = 2.

The relative standard curve for an assay can be used to verify that a template yields results within a linear dynamic range (i.e. the range of RNA input or cDNA the assay can detect) [137]. The standard curve can indicate whether the amount of RNA is too high or too low to run an assay. It can also determine the assay precision by replicating the samples at each input amount, and it can assist in determining an assay's efficiency.

Table 3.6 summarises the quality control data for TaqMan® qPCR assays that were used in this project. For TaqMan® Gene Expression Assays (*ACTB*, *HMBS*, *HBA*, *HBB, PRM1, PRM2, SEMG1, STATH, HTN3* and *18S* rRNA), the maximum concentration of the extracted RNA sample was used, as the kit has the capability to reverse transcribe up to 2 µg. The cDNA was serially diluted 1:3 creating eight different dilutions. For TaqMan® microRNA assays (*miR16*, *miR451*, *miR205*, *miR658*, *miR10b*, *miR891a* and *U6*), The cDNA was serially diluted 1:3 with eight points, where the concentration of the highest standard was 10 ng/µL and the lowest was 0.004 ng/µL, in line with the recommended highest concentration of RNA that should be used with the TaqMan® MicroRNA Reverse Transcription Kit.

The efficiency of the tested TaqMan® assays ranged between 90.0% and 108.10%, which are all within the accepted range for good quality qPCR data (90% to 110%) [137]. Both the *HTN3* and *miR658* markers were excluded from the study, as they exhibited a very high C_q value at the highest concentrations of RNA, and showed no C_q respectively. The efficiencies of these assays could therefore not be determined. These findings showed that the majority of the selected TaqMan® assays all have efficiencies close to 100%.

Table 3.6**: Summary of TaqMan® qPCR assay quality control data.** NTC represents the no template negative control that contains water substituted into the qPCR in place of cDNA to monitor for contamination and/or primer-dimer formation, and the RT- is produced from a reverse transcription without reverse transcriptase, used to monitor for amplification of contaminant genomic DNA.

Target gene	Life Technologies Assay ID	RNA concentration range (ng/µL)	qPCR gradient	% Efficiency	R^2	C_q NTC	C_q RT-
ACTB	Hs99999903_m1	85-0.035	-3.3669	98.16	0.995	No C_q	No C_q
18S rRNA	Hs99999901_s1	85-0.035	-3.447	95.03	0.992	38.53	37.78
HBMS	Hs00609296_g1	85-0.035	-3.1418	108.10	0.990	No C_q	No C_q
HBA	Hs00361191_g1	85-0.035	-3.581	90.22	0.999	No C_q	No C_q
HBB	Hs00758889_s1	85-0.035	-3.5039	92.93	0.997	No C_q	No C_q
PRM1	Hs00358158_g1	60-0.027	-3.4826	93.71	0.999	No C_q	No C_q
PRM2	Hs04187294_g1	60-0.027	-3.5106	92.69	0.999	No C_q	No C_q
SEMG1	Hs00268141_m1	60-0.027	-3.4721	94.09	0.994	No C_q	No C_q
HTN3	Hs00264790_m1	40-0.018	-	-	-	-	-
STATH	Hs00162389_m1	40-0.018	-3.5936	90.00	0.985	No C_q	No C_q
miR16	000391	10-0.004	-3.4536	94.78	0.990	No C_q	No C_q
miR451	001141	10-0.004	-3.474	94.02	0.990	No C_q	No C_q
miR10b	002218	10-0.004	-3.3243	99.90	0.990	No C_q	No C_q
miR891a	002191	10-0.004	-3.3606	98.41	0.997	No C_q	No C_q
miR205	000509	10-0.004	-3.9059	100.67	0.980	No C_q	No C_q
miR658	001513	10-0.004	-	-	-	-	-
U6	001973	10-0.004	-3.5265	92.12	0.989	No C_q	No C_q

3.3.4.2 Negative controls

There are two types of negative controls that are run with each qPCR assay, and both should have a negative value (i.e. no C_q). The first negative control is the no template control (NTC), which is used to assess any contamination of the reagents or primer-dimer formation, by adding water into the reaction instead of the RNA template. The second negative control is the reverse transcription reaction minus the reverse transcriptase (RT-), which is used to assess any amplification of genomic DNA. Table 3.6 shows the results for both NTC and RT- for each assay. As expected only the *18S* rRNA assay gave C_q values for the NTC and RT-, and this is due to *18S* rRNA being so ubiquitous and hyper-abundant in the cell. This means that even a minute amount of biological material could contain enough *18S* to be amplified during qPCR. Therefore, it would be expected that the negative controls of *18S* assay would give C_q values, but these would be well above 35 cycles. The mean C_q for the *18S* rRNA assay carried out on the experimental samples ranged from 20.61-31.75 (see Appendix A5 and A6), which is considerably lower than the NTC and RT- C_q values of 38.53 and 37.78 respectively. The MIQE guidelines [167] suggest recording a threshold cut-off criteria for the amounts of any tolerable contamination, therefore the cut-off C_q values for *18S* assays was set at 37 in this project.

3.3.4.3 Intra-assay and inter-assay variation

The evaluation of the RT-qPCR data reproducibility is also essential and can be obtained by analysing the same cDNA sample repeatedly with the same qPCR assay. The intra-assay variation was measured by running the same sample of cDNA repeatedly within the same RT-qPCR plate, while inter-assay variation was measured by running the same sample of cDNA repeatedly across different RT-qPCR plates and/or days.

The intra-assay and the inter-assay variation measured using relative coefficient of variation (CV%) values, ranged from 0.660 to 2.904 and 1.115 to 2.860 respectively (Table 3.7), indicating a high reproducibility of the assays.

Table 3.7**: TaqMan® qPCR assay quality control data**. Intra-assay and inter-assay variation were calculated based on repeat examination of n = 8 replicates.

TaqMan® assay	Intra-assay variation (CV%)	Inter-assay variation (CV%)
ACTB	0.660	1.237
18S rRNA	1.497	1.708
HMBS	1.094	2.581
HBA	2.904	2.860
HBB	1.762	2.030
PRM1	0.770	1.200
PRM2	1.949	1.634
SEMG1	1.637	1.445
STATH	1.556	1.616
miR16	0.760	1.942
miR451	1.272	1.115
miR10b	1.261	1.888
miR891a	1.632	1.596
miR205	1.249	1.947
U6	1.217	1.345

3.4 Summary and Conclusion

Two RNA extraction methods were evaluated to determine which would perform most favourably with regards to RNA yield and purity, when conducted to extract RNA from a small amount of body fluids. The TRI Reagent® method and the miRNeasy Mini Kit were utilised for total RNA extraction, followed by the use of NanoDrop-1000 to measure the concentration and the purity of the extracted RNA, and the outcomes of both methods were compared. Despite

the disadvantages of using TRI Reagent® for RNA extraction, such as the possibility of residual contamination, it was shown to improve both the RNA yield and purity from blood, saliva and semen samples relative to the miRNeasy Mini Kit.

Furthermore, RT-qPCR was applied to estimate the quantity of three reference genes (*ACTB*, *GAPDH* and *18S*) in RNA extracted from blood, saliva and semen samples. The results again demonstrated that the quantity of RNA extracted from blood and semen samples utilizing TRI Reagent® was higher than the quantity of RNA extracted utilizing the miRNeasy Mini Kit. However, the saliva samples showed some variation that might be caused by the presence of bacteria, which may increase the rate of nucleic acid degradation. Accordingly, the TRI Reagent® extraction method was, therefore, utilised as the extraction system in all further work.

Moreover, the two approaches that were used to measure the quantity and the quality of the extracted RNA suffered from a number of limitations. The UV-visible spectrophotometry (NanoDrop-1000) measured the concentration of the total RNA regardless of whether it was intact or fragmented, so it cannot be used to indicate the degradation state of the samples. Similar conclusion was drawn to the Agilent 2100 Bioanalyzer approach as it did not give a reliable indication about the degradation level in the isolated RNA samples, and showed high levels of variation among bloodstains leading to lack of accuracy and low precision. The high level of variation could be due to the relatively small sample size. Nevertheless, when comparing the obtained RIN values with RT-qPCR data, there was no correlation found. The C_q values of some samples have shown not to be affected by the RIN values, as samples with low RIN (i.e. degraded samples) still gave low C_q values (higher quantity). Therefore, it is not recommended to use the Agilent 2100 Bioanalyzer as a method to quantify total RNA or to check RNA quality for forensic samples types, which will generally be very low in volume and highly degraded.

It is essential to conduct an assessment of the TaqMan® assays to be used in the current project for quality control purposes and to ensure adherence to the MIQE guidelines, in order to enhance the ability to publish any of the work presented here. Using the same RNA sample, replicates of the TaqMan® assays including the Gene Expression Assays and the MicroRNA Assays were carried out. All examined assays generated efficiencies ranging between 90.0% and 108.10%, which is within the accepted range for good quality qPCR data. The assays also gave negative results for the qPCR negative controls (NTC and RT-), with the exception of the *18S* assay, which showed very low levels of amplification in negative control samples; this is not surprising given the high level of *18S* rRNA expected in the sample. Regarding the assessment of the precision of the data by testing intra-assay and inter-assay variation for reproducibility, all tested TaqMan® assays proved to have highly reproducible data. As such, the TaqMan® assays selected demonstrated a reliable and sensitive technology to quantify gene expression in body fluid samples (blood, saliva and semen).

Chapter four: Quantification of degradation in blood-specific RNA markers to estimate bloodstain age using RT-qPCR

4.1 Introduction

The application of RNA analysis in forensic body fluid identification has been well established as discussed previously in Chapter one, as many RNA transcripts have been identified as body fluid/tissue-specific. Identifying the type of the recovered body fluid samples from crime scene provides a relevant information to the investigation questions. Another piece of information that has also become increasingly important to investigators is the time since deposition. Some studies have been carried out to determine the relationship between the degradation rate of RNA transcripts in aged body fluid stains such as blood [58], saliva [87] or other biological materials such as hair [60] and the elapsed time since they were deposited. These studies have applied the approach of relative expression ratio (RER) of two reference genes; *ACTB* mRNA to *18S* rRNA, it has shown to be correlated with the age of bloodstains, saliva and hair. although these studies have suggested a correlation between RNA degradation and time since deposition, to date in a forensic context, there are no practical methods available to estimate the age of biological evidence or the time since deposition of a body fluid at a crime scene.

An evaluation of the feasibility of using blood-specific RNA markers for the estimation of bloodstain age was carried out in this Chapter. The aim was to look for a correlation between the age of the bloodstain and the degradation rate and relative quantity of blood-specific mRNA and miRNA markers. Specifically, the degradation pattern of different RNA molecules in blood samples was characterised via TaqMan® assays in order to assess the possibility of using the degradation rate to determine the time since deposition. TaqMan® Gene Expression Assays were used for blood-specific mRNA

markers, and TaqMan® MicroRNA Assays were used for blood-specific miRNA markers.

In this part of the project, two approaches to relative quantification were applied in order to determine the relative expression of RNA blood-specific markers over time: the $2^{-\Delta Cq}$ method, and the relative expression ratio (RER). Both methods can be used to determine the relative quantification of a less stable marker to a more stable marker across all ageing points.

4.1.1 Blood-specific RNA markers

According to a number of published studies, mRNA molecules can be used to identify the origin of different body fluid types such as blood, semen and saliva [54, 69, 168], and many recent studies have also identified miRNA molecules that are suitable markers for identifying each type of body fluid [66, 76, 77, 81, 82].

In the context of forensic applications, Haemoglobin Subunit Alpha (*HBA*), and Haemoglobin Subunit Beta (*HBB*), which are protein subunits of the haemoglobin molecule, and Hydroxymethylbilane Synthase (*HMBS*), which is an enzyme of the heme biosynthesis pathway [169], are mRNA molecules that have been identified as blood-specific markers [68, 170]. Similarly, a number of miRNAs have been shown to exhibit higher expression in blood, such as *miR16* and *miR451* [66, 77]. Both *miR16* and *miR451* are involved in the regulation of gene expression and affect the stability and translation of mRNAs, and they have also been shown to function as tumour suppressors in cancer cells [171, 172]. These mRNA and miRNA markers were therefore selected for use in degradation assays in aged bloodstains along with *U6* snRNA that is commonly used as a reference gene in gene expression studies.

4.1.2 TaqMan® assays

The TaqMan® Gene Expression Assays and TaqMan® MicroRNA Assays that were used in this study are off-the-shelf, and have been predesigned for each transcript by Applied Biosystems.

Table 4.1 and Table 4.2 show the markers that have been selected from the literature for the current project. Each assay has its own unique ID given by Applied Biosystems. There are a number of specific criteria that were set in order to permit the selection of predesigned TaqMan® Gene Expression Assays for use in quantifying mRNA markers in the current project. These criteria were set to evaluate the predesigned TaqMan® assays and to limit the cause of the variations in the study. Namely, the selected RNA markers should have the following characteristics:

- They should be body fluid-specific (here blood-specific).
- The amplicon should not be longer than 200 nt; the degradation rate will be analysed across a period of one year of storage, and longer amplicons tend to degrade rapidly and disappear at early stages [59].
- The markers should not have pseudogenes or other DNA homologs, which would affect cDNA amplification, as it generates products that are identical to the cDNA target leading to false-positive in the RT-qPCR results.
- The markers should not detect off-target sequences, thereby ensuring the specificity of the reaction.

Table 4.1: **Characteristics of the TaqMan® Gene Expression Assays for blood-specific markers.**

RNA target	Applied Biosystems TaqMan® assay ID	Amplicon length (nt)	Assay location*
HMBS	Hs00609296_g1	69	1070
HBA	Hs00361191_g1	156	158
HBB	Hs00758889_s1	95	511

* Refers to the nucleotide location that is the midpoint of the target region.

Table 4.2: **Characteristics of the TaqMan® MicroRNA Assays for blood-specific markers and for the *U6* marker.**

miRNA target	Applied Biosystems TaqMan® assay ID	Target sequence
miR16	000391	UAGCAGCACGUAAAUAUUGGCG
miR451	001141	AAACCGUUACCAUUACUGAGUU
U6	001973	GTGCTCGCTTCGGCAGCACATATACTA AAATTGGAACGATACAGAGAAGATTAG CATGGCCCCTGCGCAAGGATGACACG CAAATTCGTGAAGCGTTCCATATTTT

All of the selected mRNA markers met the above criteria, except *HBA* and *HBB*, which both have pseudogenes in the genome, and this may affect the amplification of cDNA. It is therefore necessary to reduce this effect and eliminate genomic DNA contamination, which can be done through treatment of the RNA extraction with DNase (section 2.5). Additionally, a portion of the extracted RNA sample was taken through the cDNA amplification step, including everything in the reaction mixture except the reverse transcriptase enzyme, as a control to ensure the removal of contaminating genomic DNA had been successful.

4.2 Aims and objectives

The overall purpose of this research was to develop a method to estimate the deposition time of biological fluids commonly encountered in forensic casework – blood, saliva and semen – using the application of RNA analysis. Initially, the degradation profiles of multiple selected RNA markers were analysed, including a reference gene and body fluid-specific mRNA and miRNA markers. The aim is to determine whether there is a correlation between the age of the bloodstain and the degradation rate of these RNA markers, in order to identify those most useful for body fluid stain ageing and

to assess the possibility of using the degradation rate to determine the time since deposition.

One aspect of this Chapter was therefore to analyse the expression level and study the degradation behaviour of multiple RNA transcripts in dried bloodstains stored at room temperature for up to one year. *HBA*, *HBB*, *HMBS*, *miR16*, *miR451* and *U6* are RNA markers that have been selected from a thorough literature review where these markers have been shown to indicate the presence of bloodstains and on the basis of Chapter three outcomes. Another aspect was to study the association between mRNA and miRNA stability and to calculate the relative expression of these two different RNA molecule types. This was done in order to determine whether this relative expression ratio can provide information about which markers are likely to be more accurate for use in estimating the age of biological stains, both over the short- and long-term.

Finally, studying the behaviour of RNA transcripts over a period of time can also help to determine for how many days after depositing a bloodstain RNA remains of good enough quality for analysis, and after how long it becomes highly degraded so that expression analysis becomes unreliable.

4.3 Materials and methods

4.3.1 Sample collection

A total of 10 volunteers (4 males and 6 females), donated blood samples. The blood samples were collected onto sterile cotton swabs using disposable Unistik 3 comfort lancets. A volume of 20 µL of blood was pipetted on to the swabs and the swabs were then stored in a dark dry place at room temperature to simulate natural ageing until they reached the desired ages (0, 3, 6, 15, 30, 90, 180, 270, and 360 days). The experimental procedures were approved by the Department of Pure and Applied Chemistry Departmental Ethics Committee (see Appendix A1).

4.3.2 Total RNA extraction and reverse transcription

RNA extraction was carried out using TRI Reagent®, as described in section 2.4.1. The TURBO DNA-free™ I Kit was used to treat the extracted RNA to remove any genomic DNA. Reverse transcription was carried out using High-Capacity cDNA kit and MicroRNA cDNA Kit, as described in sections 2.7.1 and 2.7.2.

4.3.3 Quantitative real-time PCR (qPCR)

All TaqMan assays were run in singleplex assays, following the procedure described in section 2.8. Amplification was performed using a Stratagene Mx3005P.

4.3.4 Data analysis

The data generated from RT-qPCR was analysed using *MxPro* and *GenEx* software (version 5.4.4) was used for efficiency correction of the raw data and the calculation of relative quantities (see section 2.9). *Microsoft Excel 2016* was used to manipulate raw data from the spectrophotometry and C_q values from the RT-qPCR reactions and to present basic data and line graphs. *Minitab Express* (version 1.5.0) and *Minitab®17* were used for statistical analyses, including the Anderson-Darling normality test, correlations and regression analysis.

4.4 Results and Discussion:

4.4.1 Analysis of total RNA yield

One way to determine whether the total RNA yield recovered from dried bloodstains decreased as a result of degradation is to use UV-visible spectrophotometry, which is commonly used in RNA research to indicate the quantity of recovered RNA, and was utilised in this experiment. It can provide a general assessment of the yield of RNA recovered. The RNA yield from blood

samples was assessed for bloodstains left under simulated conditions of natural ageing at room temperature for up to one year.

After storing blood samples over an interval of one year, no significant reduction in total RNA yield could be observed, as illustrated in Figure 4.1. The total RNA concentration between samples was highly variable, as shown by the wide error bars (i.e. larger standard deviations), with time point 0 (fresh samples) showing the most notable variation. This is likely due to blood being a heterogeneous mixture containing plasma, platelets, white blood cells and red blood cells that vary in their relative quantities from one individual to another, and such variations cannot be controlled by the researcher. Additionally, the recovery of RNA using TRI® Reagent is very variable [173] and is affected by the users and how effectively they homogenise the sample and separate the phases, which can increase the variability between samples. In this project the same individual performed phases separation and all samples were homogenised for the same period of time in an attempt to minimise any variation.

When applying the Anderson-Darling normality test to determine whether the data was normality distributed or not, it was established that one of the variables (time ageing points) was not normally distributed ($p = 0.0290$).

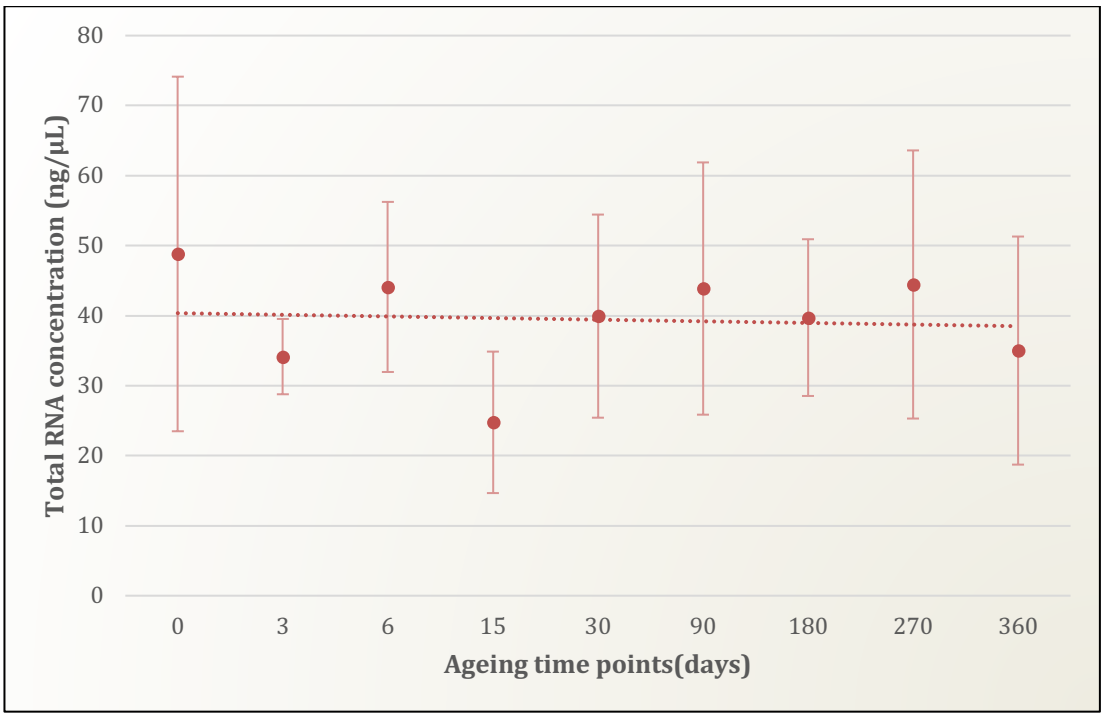


Figure 4.1**: Trend line analysis of total RNA yield from blood samples stored up to one year at room temperature**. Points represent the mean ±SD of total RNA quantity determined by UV-Spectrophotometry (ng/µL) from blood samples (n = 10 at each time point), the dotted line represents the best-fit line).

A Spearman's correlation test showed that no significant correlation was found between storage time and the total RNA yield recovered (r = -0.0167, p = 0.9661).

In general, these results were expected in this investigation. The UV-spectrophotometry system that was used for RNA quantification in this experiment is very simple to perform and available in most forensic/clinical laboratories. However, it depends on the absorbance of ultraviolet light at 260 nm regardless of whether the RNA exists as an intact or fragmented molecule. Therefore, UV-spectrophotometry system is commonly used to indicate the extracted RNA sample quantity for downstream analysis rather than RNA sample quality, as it cannot indicate RNA degradation. However, the key outcome here was that the RNA yield was not affected by ageing of bloodstains, and so if there would be any differences in the quantification level

of individual RNA markers, that would reflect the degradation of these markers and not a reduction in quantity due to reduced yield of the RNA extraction.

4.4.2 Degradation rate of individual RNA transcripts at different time points

In living cells, the degradation of RNA transcripts is regulated by various mechanisms such as deadenylation, decapping and exo- and endonucleases, at different rates [37] which gives each RNA molecule its own half-life depending on the specific elements within RNA sequence (see section 1.6). In contrast, the degradation behaviour of RNA transcripts in dead cells (*ex vivo*) remains poorly understood and it is unclear whether the majority of RNA molecules degrade at a similar rate. Although it is considered that RNA degradation *ex vivo* is a non-regulated process, recent findings by Romero et al. (2014) suggest that RNA degradation in dried peripheral blood mononuclear cells (PBMC) is a non-random process [174].

In dried biological stains, chemical and physical factors determine the rate of degradation of RNA molecules, while normal regulation is disrupted [54, 86]. In this work, the expression level of individual RNA transcripts was quantified by RT-qPCR in blood samples stored at room temperature for up to one year. The gradual increase of C_q value across storage time can characterise the degradation of RNA. The assays used in this work for mRNA transcripts were designed to amplify only sections of the mRNA markers and not the whole transcripts. The size range of the amplified mRNA was 69 to 156 nucleotides (see Table 4.1) and all were the '3' Most Assay' i.e. the assay closest to the 3' end of the relevant transcript. The RT-qPCR data for the blood-specific markers is presented in this section and demonstrated that each RNA transcript showed a unique pattern of degradation behaviour.

Figure 4.2 illustrates the mean C_q data for *HBA*, *HBB*, *HMBS*, *miR16*, *miR451* and *U6* after efficiency correction (C_q values were corrected against the obtained efficiency of each assay, see section 3.3.4.1). At day 0, which

represents the control samples (fresh samples), each of the RNAs examined had a different starting expression level. The microRNA marker *miR451* exhibited the highest expression level (lowest C_q value) followed by *miR16*, and *HMBS* had the lowest expression level (highest C_q value). Both RT- and negative controls that were performed to monitor possible contamination and residual genomic DNA showed no amplification, indicating that the amplification reflects only the specific RNA molecules expressed within a given sample.

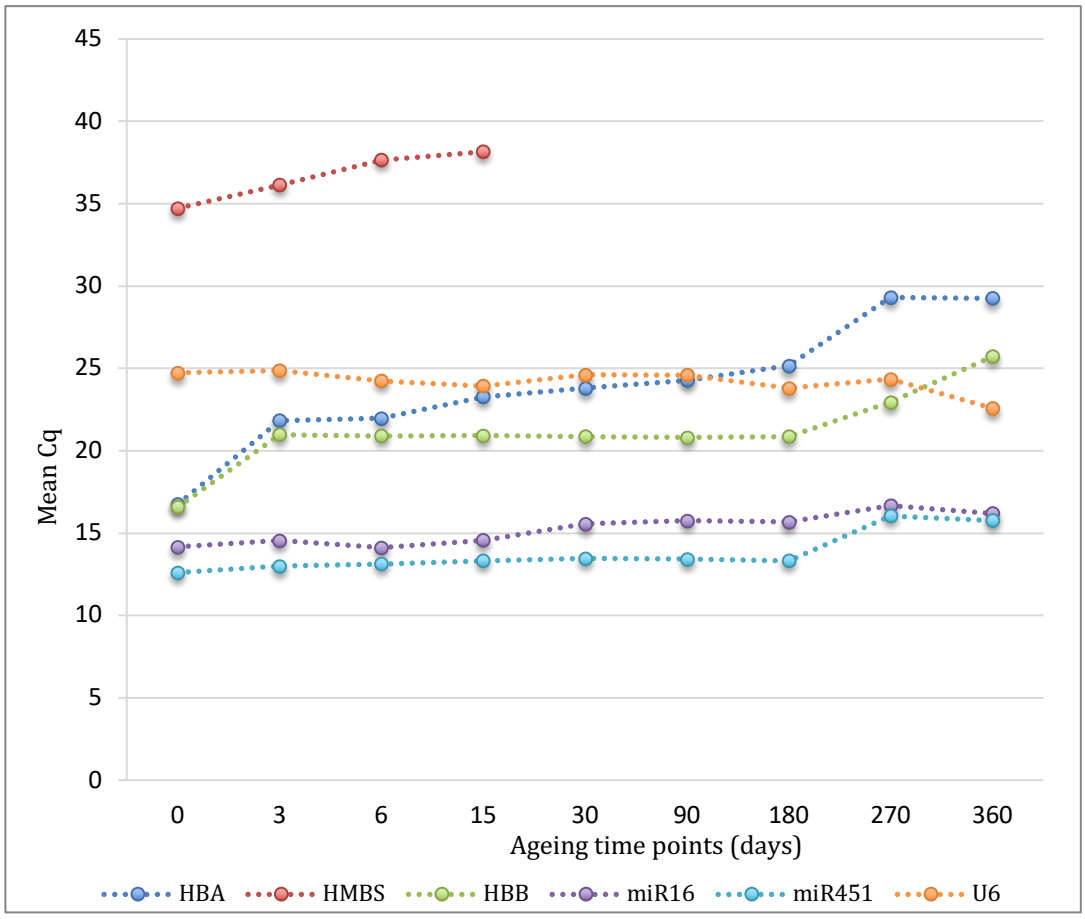


Figure 4.2: **The mean C_q data of *HBA*, *HBB*, *HMBS*, *miR16*, *miR451* and *U6* in bloodstains stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. HMBS markers was not detected after 30 days. Error bars were removed for clarity.

The difference in C_q (ΔC_q), displayed in Figure 4.3, can illustrate more clearly the degradation level of the RNA at each time point relative to the control time point (T = 0 days). Across 360 days of storage, the ΔC_q value for *U6* remained

around zero, which indicates no significant decrease in its quantity. Similarly, *miR451* remained stable across the first 180 days and started to degrade after that. Even though *miR16* showed high stability in the first 15 days, it showed slight degradation after that point, as its ΔC_q values started to increase. When looking at the mRNA markers, the level of *HBA* dropped dramatically across ageing time points, while *HBB* degraded in the first 3 days and then remained stable at the same level until 270 days, when it started to degrade again. *HMBS* was rapidly degraded in the first 15 days and returned no C_q values after 30 days, showing it was degraded to below the sensitivity of the assay used for its detection. Across the two Figures 4.2 and 4.3, standard error bars were removed for clarity, however there was no significant differences among the margin of error between all replicates (data not shown).

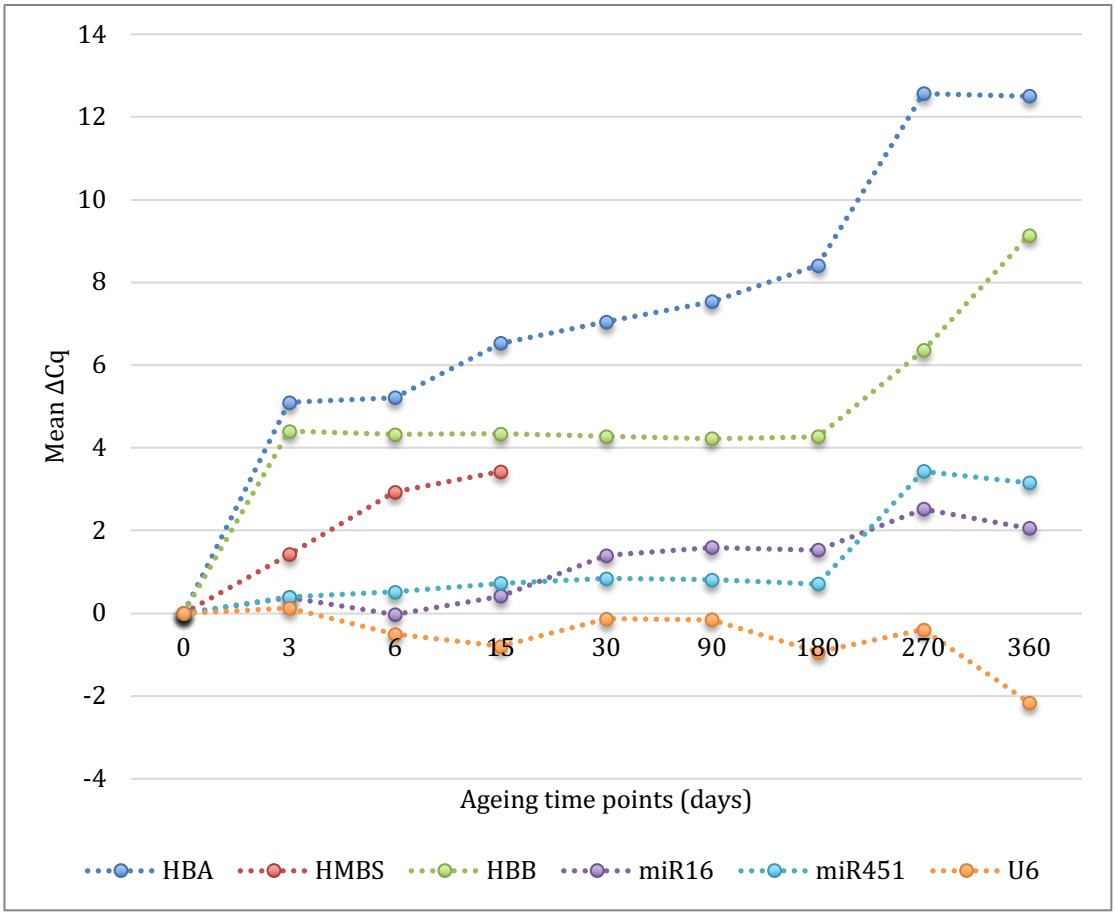


Figure 4.3: **The mean ΔC_q data of *HBA*, *HBB*, *HMBS*, *miR16*, *miR451* and *U6* in bloodstains stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. Error bars were removed for clarity.

A quick visual assessment of the outcomes presented in Figure 4.3 suggests that the degradation rate of the *HBA* marker proceeded in a linear fashion across ageing points and reached a plateau after one year of storage. *HMBS* also showed linear decay behaviour in the first 15 days, before showing no C_q after that. In contrast, the decay behaviour of *HBB* was quite different. It was expected to show a degradation pattern similar to the other mRNA markers; however, it degraded in the first 3 days and then remained at the same level across 180 days and then started to degrade again up to 360 days. Supporting these findings, Haas et al. (2011) have shown that both *HBA* and *HBB* mRNAs were still detected in blood samples that were stored at room temperature and 37 °C for one year and exposed to other environmental conditions for one month [170]. There are some studies that have shown that the *HBB* marker can still be detected in 3 year old samples [175] or 30-50 year old samples [176], confirming that the *HBB* marker can remain stable for longer periods of time than other mRNA markers. This stability might be due to the sequence differences in the 3' untranslated regions (3' UTR) between ∂-globin and ß-globin mRNAs (i.e. produced by *HBA* and *HBB* genes respectively), which give ∂-globin mRNA a half-life less than one-third of that of ß-globin mRNA. It was proven that both *HBA* and *HBB* mRNAs have specific sequences or elements at the 3' UTR region that are responsible for their distinctive stability, where *HBA* is destabilised by single-site mutation with in its 3' UTR, while *HBB* mRNA is destabilised by multiple elements within its 3' UTR, giving it higher stability than *HBA* [177]. Even at the genome level, both genes have shown differences at the promoter sites.

Interestingly, another study has also shown that increases in the length of the protein coding region, the length of the 3' UTR and the %GC content are correlated with higher degradation rates, but the total transcript length is not associated with degradation behaviour [174].

With regards to the miRNA markers, high stability was observed in these markers during the one year storage period. However, even though miRNA markers are considered among the most stable molecules in blood, Nakao et

al. (2013) [91] showed that degradation of the *miR16* and *miR451* markers can be observed in blood samples stored for up to 28 days at 25 °C and 50% humidity. Additionally, Patnaik et al. (2010) [178] determined that miRNA markers tend to degrade under high humidity in bloodstains, if the stains did not dry completely. Wet bloodstains are more vulnerable to ribonuclease activity, therefore degrade faster than dried bloodstains. In this study, only a small volume of blood was used and all samples were dried before storage at room temperature. However, low-level degradation still occurred in the *miR16* marker when the samples aged for 30 days were extracted, indicating that slight degradation can occur in miRNA markers after 15 days.

The stability of miRNA markers observed in this work is in concordance with the outcomes published by Wan et al. (2013) [82], who illustrated that the C_q value of *miR16* increased slightly after storing samples for one month. The current data also confirmed the findings of Courts et al. (2011)'s study, which showed that *miR451* has high stability in aged blood samples [81].

Finally, the very high stability of the *U6* marker across all ageing time points can be explained by its structure, as its first nucleotide is methylated and the 3' end of this molecule is bound by a protein called the La protein, which acts to increase the stability of the *U6* molecule [4].

4.4.3 Inter-donor variation at each time point

At each time point, the inter-donor variation was calculated using the corrected C_q values to obtain the coefficient of variation (CV%). Table 4.3 shows the mean inter-donor variation across all ageing time points for each of the selected RNA markers.

Table 4.3: **The mean inter-donor variation in RNA marker degradation rate across all ageing time points.** The coefficient of variation was calculated using the equation CV = standard deviation/mean x 100.

RNA marker	Mean inter-donor variation (Coefficient of Variation %)
HBA	9.08
HBB	8.71
HMBS*	4.17
miR16	6.14
miR451	8.20
U6	4.02

* Only for up to 15 days.

When considering how the inter-donor variation in each RNA marker changes across age time points, it was observed that this fluctuates widely across the ageing period, as illustrated in Figure 4.4. The highest inter-donor variation was observed at 270 days for *HBA* and *HBB*, at 0 and 15 days for *miR16*, and at 15 days for both *miR451* and *U6*.

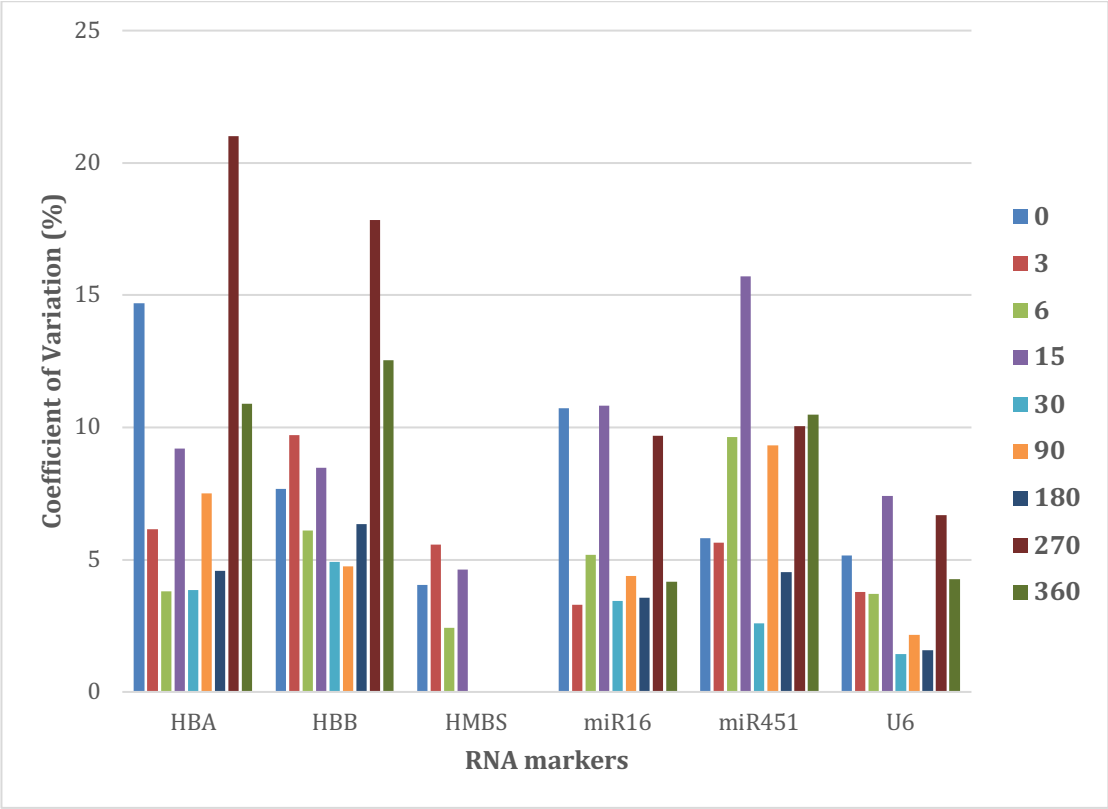


Figure 4.4: **Inter-donor variability in the degradation rate of RNA markers in bloodstains.** The coefficient of variation was calculated using the mean of n = 10 for samples aged up to 360 days (*HMBS* only up to 15 days) at room temperature. Coloured bars represent the different ageing time points.

These fluctuations could be explained by variation at the cellular level, where the composition of ribonucleases differs between individuals, and these molecules may be responsible for RNA degradation. Another explanation could be laboratory error, which may be considered as another source of variation. Many manual steps are involved in the procedure used during this study, which can lead to variation in the volumes/quantities of samples transferred between steps. Variation may occur during reverse transcription or qPCR step. Some of these variations are inevitable, however the researcher should try to minimise them as much as possible, for instance increasing sample number may be one way of decreasing inter-donor variation in future experiments.

4.4.4 Relative quantification analysis

In this project, two approaches to relative quantification were applied to determine the relative expression of RNA markers over time: Log $2^{-\Delta Cq}$ (see section 2.9.2) and relative expression ratio (RER) (see section 2.9.1). Both methods calculate the relative quantification of the less stable RNA marker to the more stable marker across all ageing time points, to determine which method is more suitable to estimate the age of a bloodstain. Therefore, the stability of each RNA marker should be identified, to determine which marker is more stable than the other.

The mean C_q values for blood-specific markers measured by RT-qPCR at 0, 3, 6, 15, 30, 90, 180, 270, and 360 days are shown in Table 4.4 for mRNA and Table 4.5 for miRNA markers. These data were corrected with the determined efficiency of each assay using *GenEx* statistical software (version 5.4.4).

Table 4.4: **The mean C_q values of mRNA blood-specific markers at 0, 3, 6, 15, 30, 90, 180, 270, 360 days for the duplicates of 10 samples after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

Age points	Markers	Sample number/Mean C_q values									
		1	2	3	4	5	6	7	8	9	10
Day 0	*HBA*	17.27	16.80	20.26	16.14	15.44	14.74	14.93	15.96	14.15	21.76
	HBB	15.96	16.07	19.55	16.01	15.76	15.81	16.36	16.31	15.73	18.23
	HMBS	35.27	34.21	37.60	33.60	34.61	33.85	33.67	34.18	33.50	36.71
Day 3	*HBA*	20.96	22.38	24.16	21.59	23.40	21.84	21.84	22.28	20.24	19.72
	HBB	17.92	21.21	24.69	20.65	21.74	22.89	21.95	21.02	18.60	19.17
	HMBS	33.27	36.58	40.15	36.46	36.56	36.21	37.67	36.46	34.22	33.83
Day 6	*HBA*	21.86	21.36	22.71	21.76	21.29	22.32	22.59	22.95	22.48	20.29
	HBB	21.08	18.96	20.46	20.98	19.02	23.16	21.24	21.40	21.86	20.91
	HMBS	36.69	37.29	No Cq	37.29	37.62	38.40	37.24	38.34	38.58	36.08
Day 15	*HBA*	21.47	23.15	23.38	21.59	22.74	19.82	25.12	27.02	25.61	22.90
	HBB	19.93	20.02	20.98	19.79	19.13	19.15	23.29	24.11	22.20	20.63
	HMBS	36.62	37.19	37.53	37.11	38.41	35.92	No Cq	41.74	No Cq	37.50
Day 30	*HBA*	22.74	23.03	24.56	24.55	24.14	23.40	22.89	25.65	23.58	23.42
	HBB	18.88	20.46	21.51	20.65	22.06	20.92	22.00	21.84	20.41	19.86
	HMBS	37.78	38.85	No Cq	No Cq	No Cq	38.67	No Cq	No Cq	No Cq	37.99
Day 90	*HBA*	23.15	23.18	25.72	26.87	22.51	21.23	25.87	26.14	24.35	23.82
	HBB	19.37	20.33	21.26	21.57	20.39	20.27	22.36	22.21	20.13	20.19
	HMBS	38.81	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
Day 180	*HBA*	26.26	25.33	25.55	23.85	24.74	25.95	25.06	27.24	23.79	23.82
	HBB	20.58	20.21	23.09	20.21	20.52	21.95	21.24	22.43	19.14	19.14
	HMBS	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
Day 270	*HBA*	23.13	32.22	25.55	37.04	29.56	24.58	21.21	37.04	25.74	37.04
	HBB	17.57	22.71	22.24	29.16	25.11	20.89	18.51	27.67	18.79	26.78
	HMBS	36.27	No Cq	No Cq	No Cq	No Cq	No Cq	36.68	No Cq	36.78	No Cq
Day 360	*HBA*	27.64	28.18	30.41	27.82	28.41	27.75	27.91	28.12	28.21	38.04
	HBB	25.06	23.96	26.12	24.63	24.80	24.37	24.72	24.51	24.27	34.75
	HMBS	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq

Table 4.5: **The mean C_q values of miRNA blood-specific markers and U6 at 0, 3, 6, 15, 30, 90, 180, 270, 360 days for the duplicates of 10 samples after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

Age points	Markers	Sample number/Mean C_q values									
		1	2	3	4	5	6	7	8	9	10
Day 0	*miR16*	13.72	14.09	14.53	13.48	12.89	17.89	13.29	14.38	12.40	14.86
	miR451	11.94	12.76	13.08	12.77	12.05	11.59	12.37	13.65	12.07	13.77
	U6	23.94	24.28	27.11	25.36	24.93	26.20	22.82	23.61	24.05	25.10
Day 3	*miR16*	14.08	15.03	14.46	13.86	14.17	14.48	15.14	15.30	14.60	14.29
	miR451	12.34	13.51	13.21	11.68	12.07	13.69	13.81	13.49	13.33	12.88
	U6	23.26	24.00	25.16	24.99	25.25	24.53	24.13	25.91	24.93	26.49
Day 6	*miR16*	13.55	13.19	13.23	14.22	13.93	14.86	15.20	14.50	14.93	13.61
	miR451	12.42	11.73	11.49	12.46	12.63	12.68	15.08	14.38	14.81	13.51
	U6	25.10	23.97	23.98	23.77	23.22	23.17	23.94	24.05	25.23	25.93
Day 15	*miR16*	13.39	12.81	13.75	13.16	14.36	13.61	15.72	15.32	17.91	15.66
	miR451	12.09	11.31	12.37	11.21	13.14	12.13	14.93	13.79	18.19	14.13
	U6	22.68	22.20	23.10	22.24	22.92	22.40	25.48	24.87	26.95	26.60
Day 30	*miR16*	14.85	15.54	15.85	14.96	16.33	15.77	15.46	16.28	15.56	14.88
	miR451	12.92	13.45	13.64	13.45	13.64	13.12	13.64	14.07	13.59	13.00
	U6	24.49	24.69	24.83	24.13	24.59	24.98	25.28	24.48	24.31	24.27
Day 90	*miR16*	15.29	14.98	15.52	15.83	15.42	15.46	16.17	17.43	16.01	15.35
	miR451	12.80	12.17	12.59	13.05	12.76	12.93	13.82	16.63	13.72	13.69
	U6	23.97	24.84	25.31	24.65	24.95	25.07	23.99	24.80	23.70	24.56
Day 180	*miR16*	16.10	15.30	15.91	15.12	15.33	16.14	15.83	16.70	15.40	15.00
	miR451	13.26	13.16	13.48	12.84	12.40	13.03	14.24	14.39	13.16	13.19
	U6	24.13	24.40	23.87	23.39	23.62	24.32	23.74	23.40	23.44	23.69
Day 270	*miR16*	19.00	17.23	17.77	17.78	17.88	17.38	15.56	14.90	15.03	14.18
	miR451	17.31	17.31	18.21	16.71	17.11	15.96	15.96	14.25	14.19	13.36
	U6	25.71	25.41	24.69	25.46	25.99	25.97	22.37	21.78	22.77	23.30
Day 360	*miR16*	17.70	15.86	15.92	15.24	16.27	15.92	16.19	16.27	16.91	15.79
	miR451	20.04	15.05	15.09	13.89	15.97	15.15	15.26	16.48	15.73	14.97
	U6	22.04	21.40	23.32	22.40	21.73	22.87	23.78	21.20	23.69	23.40

4.4.4.1 Identification of the most stable RNA marker

The relative quantification of pairs of markers was calculated using the C_q value of a focal degraded marker and the C_q value of the most stable marker. In gene expression studies, a reference gene or endogenous control RNA has to be identified that is suitable for normalisation purposes. To fulfil this function, these control RNAs have to be stable, less variable than the focal marker, and should be expressed across various conditions. In the current study, the different ageing time points are considered the experimental conditions, so an endogenous control RNA must be stable across time points.

There are software packages available to assess RNA expression variabilities, such as *geNorm* and *NormFinder*. Here, *geNorm* software was used to assess the expression of RNA transcripts and determine which RNA markers were most stable and exhibited the lowest variability across all time points.

Figure 4.5 shows the data from *geNorm*, which measures the stability of a gene by calculating an M-value, which is the expression stability value, and pair-wise variation of a certain gene compared to others using the raw C_q values [147]. The marker that shows the highest variation relative to all other markers is eliminated by *geNorm*. The average stability is then determined by the M-value for each pair-wise comparison; the lower the M-Value, the higher the stability. *geNorm* indicated that *miR451* and *miR16* were the most stable markers across all time points, with the lowest M-Value (both 0.989), followed by *U6* with an M-Value of 1.290. *HMBS* was excluded from the calculation, as it did not give C_q values after 15 days.

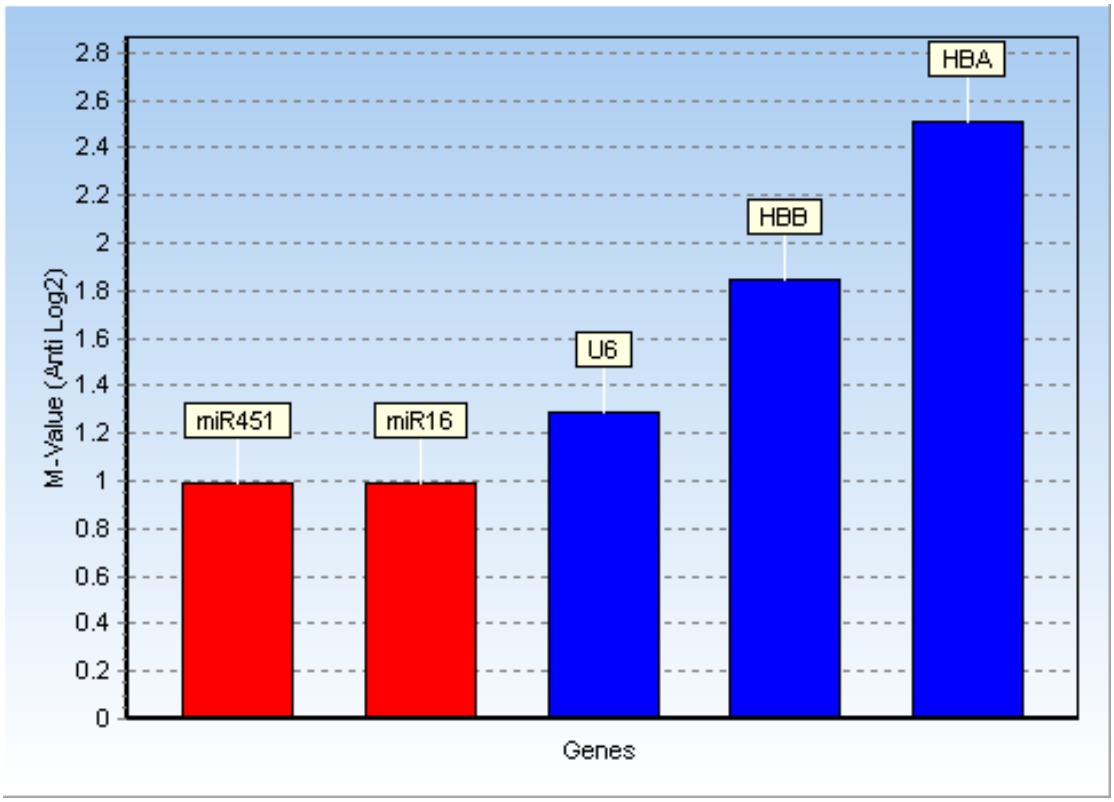


Figure 4.5: **The most stable RNA markers across all ageing time points measured by geNorm.** *miR451* and *miR16* exhibited the lowest M-Value indicating that they were the most stable markers.

4.4.4.2 Relative expression over time (Log $2^{-\Delta Cq}$)

The relative expression over time was calculated using the corrected C_q values to obtain the Log $2^{-\Delta Cq}$ value of different RNA blood-specific markers (see section 2.9.2). The Log $2^{-\Delta Cq}$ was calculated using ΔC_q, where:

$$\Delta C_q = C_{q\ \text{target gene}} - C_{q\ \text{reference gene}}$$

In this study, the less stable RNA marker was treated as the target gene and the more stable marker was treated as the reference gene.

4.4.4.2.1 The Log $2^{-\Delta Cq}$ of RNA blood-specific markers

Initially, only blood-specific markers were analysed. Both miRNA markers (*miR16* and *miR451*) were shown to be the most stable markers across ageing time points, while the *HBA* mRNA marker is the least stable, compared to the miRNA markers and the *HBB* mRNA marker. Table 4.6 shows the mean values of Log $2^{-\Delta Cq}$ ($C_{q\ (HBA)} - C_{q\ (miR16)}$), ($C_{q\ (HBA)} - C_{q\ (miR451)}$) and ($C_{q\ (HBA)} - C_{q\ (HBB)}$) for the 10 samples per each ageing time point.

Table 4.6: **The mean Log $2^{-\Delta Cq}$ values of (*HBA-miR16*), (*HBA-miR451*) and (*HBA-HBB*) across 360 days of ageing.** The corrected C_q values in Table 4.4 and Table 4.5 were used to calculate the $2^{-\Delta Cq}$ values. n = 10.

	Mean Log $2^{-\Delta Cq}$		
Age points	(*HBA - miR16*)	(*HBA - miR451*)	(*HBA - HBB*)
Day 0	-0.78	-1.25	-0.05
Day 3	-1.90	-2.31	-0.26
Day 6	-2.21	-2.51	-0.32
Day 15	-2.62	-3.00	-0.71
Day 30	-2.48	-3.11	-0.88
Day 90	-2.57	-3.29	-1.05
Day 180	-3.00	-3.57	-1.3
Day 270	-3.81	-4.00	-1.92
Day 360	-3.91	-4.06	-1.06

When plotting the results shown in Table 4.6 against time points, as illustrated in Figure 4.6, it can be seen that the mean Log $2^{-\Delta Cq}$ values decrease as the ageing time points increase.

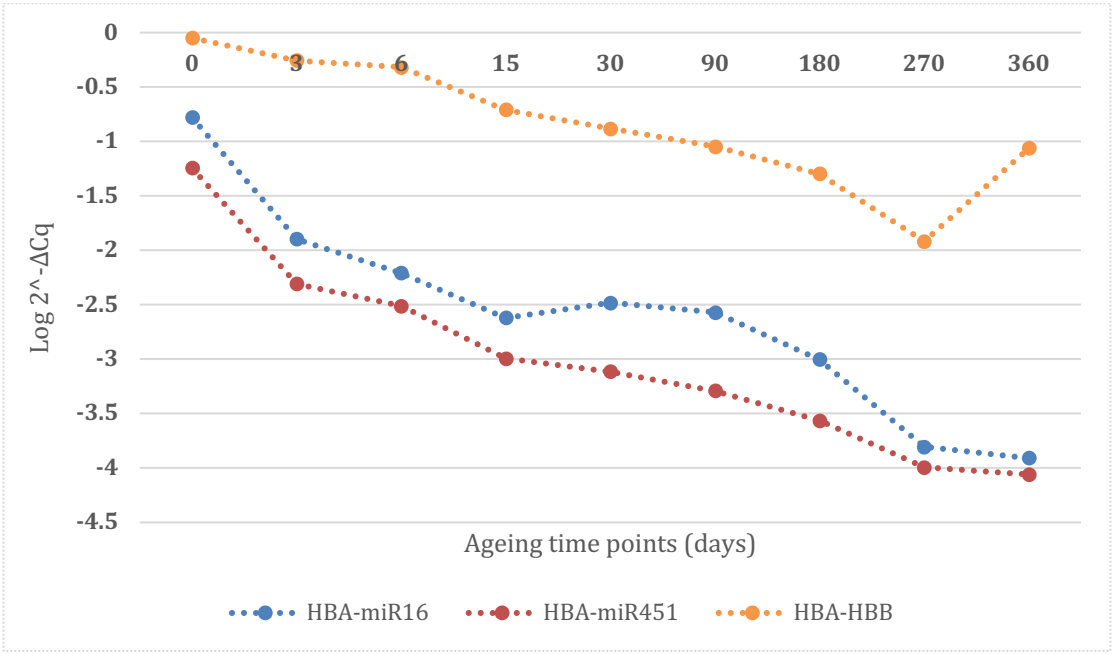


Figure 4.6: **Mean Log $2^{-\Delta Cq}$ values of (*HBA-miR16*), (*HBA-miR451*) and (*HBA-HBB*) over the ageing time points**. In all three cases the Log $2^{-\Delta Cq}$ decreases as the age of the bloodstain increases, with the exception of (*HBA-HBB*) increasing after 270 days. Error bars were removed for clarity.

There was a rapid decrease in the mean Log $2^{-\Delta Cq}$ value in the first 15 days for both (*HBA-miR16*) and (*HBA-miR451*), and the latter continued to decrease gradually, while the mean Log $2^{-\Delta Cq}$ value for (*HBA-miR16*) remained relatively stable between 15 and 90 days and then started to decrease again. The mean Log $2^{-\Delta Cq}$ value for (*HBA-HBB*) decreases in a linear fashion in the first 270 days, and then this value increased as the *HBB* marker started to degrade in the one-year storage samples, as seen in Figure 4.6.

The relationship between the obtained mean values of Log $2^{-\Delta Cq}$ and the ageing time points was statistically evaluated. A normality test was carried out using the Anderson-Darling test; this showed that one of the variables in the data, ageing time points was not normally distributed ($p = 0.029$). Therefore, nonparametric tests were used for statistical analysis.

A Spearman's correlation test indicated that the Log $2^{-\Delta Cq}$ values for (*HBA-miR16*), (*HBA-miR451*) and (*HBA-HBB*) exhibit significant negative

correlations with the age of bloodstains ($r = -0.95$, $p < 0.0001$; $r = -1$, $p < 0.0001$; $r = -0.95$, $p < 0.0001$ respectively). These results suggest that the Log $2^{-\Delta Cq}$ value of RNA blood-specific markers could be used to estimate the age of bloodstains.

4.4.4.2.2 The Log $2^{-\Delta Cq}$ of RNA markers – reference gene (U6)

The Log $2^{-\Delta Cq}$ values of RNA markers (*HBA*, *miR16* and *miR451*) versus the reference gene *U6* were also calculated, to determine whether there is a relationship between the expression rate of each marker to the expression rate of the reference gene. Table 4.7 shows the mean values of Log $2^{-\Delta Cq}$ for ($C_{q\ (HBA)} - C_{q\ (U6)}$), ($C_{q\ (HBB)} - C_{q\ (U6)}$), ($C_{q\ (miR16)} - C_{q\ (U6)}$) and ($C_{q\ (miR451)} - C_{q\ (U6)}$) for the 10 samples per each ageing time point.

Table 4.7: **The mean Log $2^{-\Delta Cq}$ values of (*HBA-U6*), (*HBB-U6*), (*miR16-U6*) and (*miR451-U6*) across 360 days.** The corrected C_q values in Table 4.4 and Table 4.5 were used to calculate the $2^{-\Delta Cq}$ values. n = 10.

	Mean Log $2^{-\Delta Cq}$			
Age points	(*HBA – U6*)	(*HBB – U6*)	(*miR16 – U6*)	(*miR451 – U6*)
Day 0	2.12	2.46	3.19	3.65
Day 3	0.88	1.19	3.13	3.59
Day 6	0.84	1.00	3.04	3.35
Day 15	0.20	0.91	2.82	3.19
Day 30	0.24	1.13	2.73	3.36
Day 90	0.09	1.14	2.66	3.36
Day 180	-0.41	0.89	2.59	3.16
Day 270	-1.49	0.42	2.31	2.50
Day 360	-2.01	-0.94	1.92	2.05

When plotting the results shown in Table 4.7 against ageing time points, as illustrated in Figure 4.7, it can be seen that the mean Log $2^{-\Delta Cq}$ values decrease as the age of the bloodstains increase across all three pairwise comparisons.

The most pronounced reduction was in the mean Log $2^{-\Delta Cq}$ value of (*HBA-U6*) followed by the mean Log $2^{-\Delta Cq}$ value of (*HBB-U6*) then the mean Log $2^{-\Delta Cq}$ of (*miR16-U6*) and (*miR451-U6*).

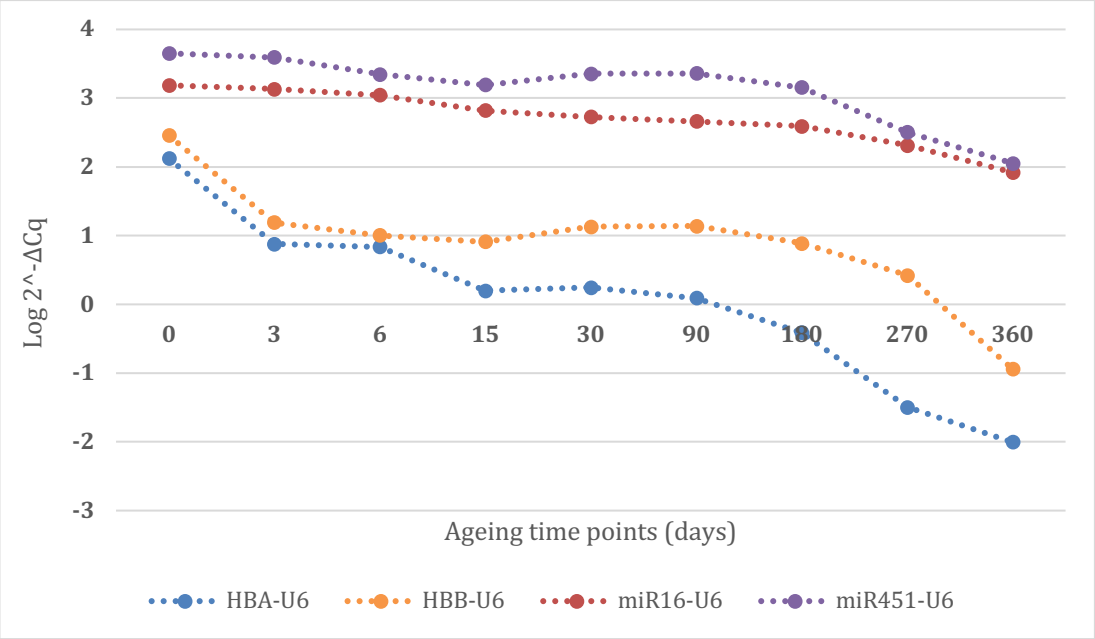


Figure 4.7: **Mean Log $2^{-\Delta Cq}$ values of (*HBA-U6*), (*HBB-U6*), (*miR16-U6*) and (*miR451-U6*) over the ageing time points.** In all three cases the Log $2^{-\Delta Cq}$ decreases as the age of the bloodstains increases. Error bars were removed for clarity.

The relationship between the obtained values of Log $2^{-\Delta Cq}$ and the ageing time points was statistically evaluated. A Spearman's correlation test indicates that the Log $2^{-\Delta Cq}$ values of (*HBA-U6*), (*HBB-U6*), (*miR16-U6*) and (*miR451-U6*) exhibit significant negative correlations with the age of bloodstains ($r = -0.98$, $p < 0.0001$; $r = -0.85$, $p = 0.0037$; $r = -1$, $p < 0.0001$; $r = -0.86$, $p = 0.0028$ respectively). These results suggest that the Log $2^{-\Delta Cq}$ values of RNA blood-specific markers to the reference gene (*U6*) could also be applied to estimate the age of bloodstains.

4.4.4.3 Relative expression ratio (RER)

The second method used to determine the relative quantification of different RNA markers was the relative expression ratio (RER). The ratio was obtained

by dividing the efficiency-corrected C_q values of the less stable RNA marker by the efficiency-corrected C_q values of the more stable RNA marker:

$$RER = \frac{Cq\ of\ less\ stable\ marker}{Cq\ of\ more\ stable\ marker}$$

4.4.4.3.1 RERs of mRNA to miRNA

As above, initially, only blood-specific markers were analysed. Since *miR451* and *miR16* were identified as the most stable marker across all ageing time points, Table 4.8 shows the mean RERs of mRNA marker to miRNA markers, calculated from the corrected C_q values for bloodstains stored up to one year at room temperature. *HMBS* was excluded from the calculation, as it did not produce C_q values after 15 days.

Table 4.8: **Mean RER of the mRNA markers (*HBA* and *HBB*) to the miRNA markers (*miR16* and *miR451*) for 10 samples at ageing time points 0, 3, 6, 15, 30, 90, 180, 270, 360 days.** The RERs were calculated from the mean C_q values in Table 4.4 and Table 4.5.

	RER			
Age points	*HBA*/*miR16*	*HBA*/*miR451*	*HBB*/*miR16*	*HBB*/*miR451*
Day 0	1.19	1.33	1.18	1.32
Day 3	1.50	1.68	1.44	1.62
Day 6	1.56	1.69	1.48	1.60
Day 15	1.60	1.77	1.44	1.59
Day 30	1.53	1.77	1.34	1.55
Day 90	1.54	1.82	1.32	1.56
Day 180	1.60	1.89	1.33	1.57
Day 270	1.81	1.86	1.39	1.45
Day 360	1.81	1.88	1.59	1.65

Figure 4.8 shows the RER values for mRNA markers to miRNA markers. It was found that the RERs of *HBA* to both miRNAs increased with increasing

ageing time points. The relationship between the obtained values of RERs and the ageing time points was statistically evaluated. Since the Anderson-Darling normality test showed that one of the variables in the data, ageing time points was not normally distributed (as shown above), nonparametric analysis was used for statistical analysis.

The RERs of *HBA* to both *miR451* and *miR16* showed a positive correlation between age time points using Spearman's correlation ($r = 0.94$, $p < 0.0001$; and $r = 0.84$, $p = 0.005$, respectively). However, the RERs of *HBB* to both miRNA markers did not show a specific trend, as the *HBB* quantification level remained stable across most ageing time points and did not show any degradation until 270 days. Overall, a wide range of RERs values was most pronounced at 270 days for all different ratios, indicating a high variation at this point. Most of the blood-specific markers have shown the highest degradation rate at this ageing time point.

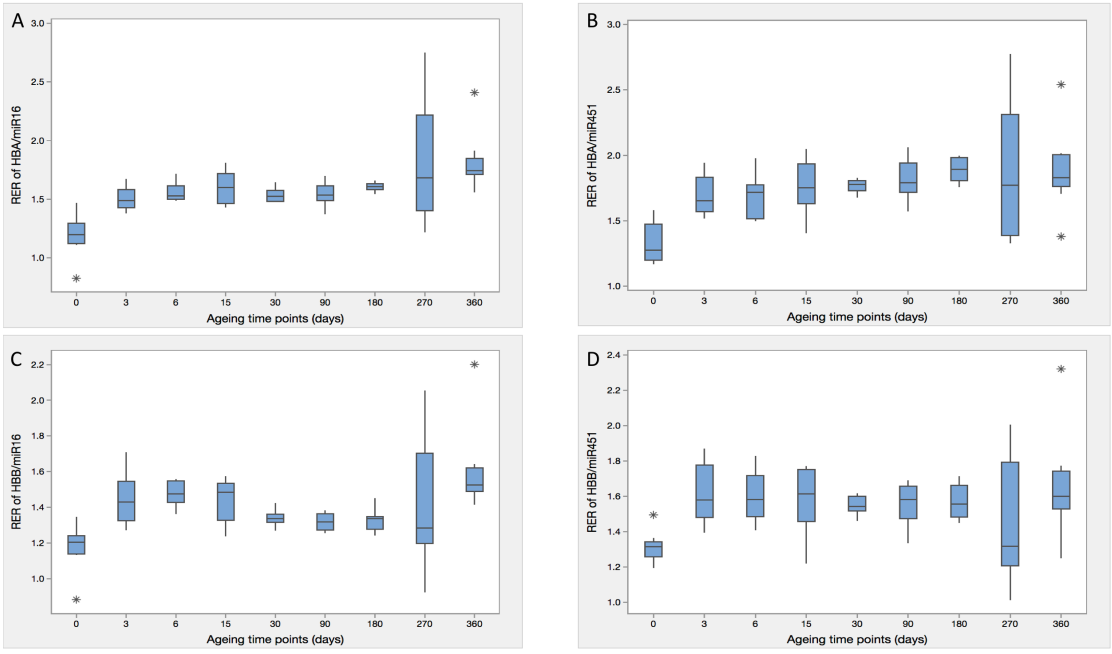


Figure 4.8: **Boxplots showing RER values of dried bloodstains.** RER of (A*) HBA/miR16*, (B) *HBA/miR451*, (C) *HBA/miR16* and (D) *HBB/miR451*. The plots were obtained from data shown in Table 4.8 using Minitab Express, * represents outliers. n = 10.

To investigate further the relationship between the mean RER values of *HBA* to *miR451* and *miR16* and the ageing time points, a regression analysis was performed. The data were fitted using two different models, a linear model (Figure 4.9) and a non-linear model (a quadratic model, also known as a second-order polynomial model; Figure 4.10). These two models were compared to each other, as well as to published studies reporting age prediction equations using linear models [58] and non-linear models [60] to estimate the age of biological samples. The models were compared to determine which is more accurate at predicting the age of bloodstains using RNA blood-specific markers.

When fitting the data using a linear model, the analysis produced a low R^2 value for both the RER of *HBA*/*miR16* (R^2 = 59.9%) and HBA/*miR451* (R^2 = 38.5%).

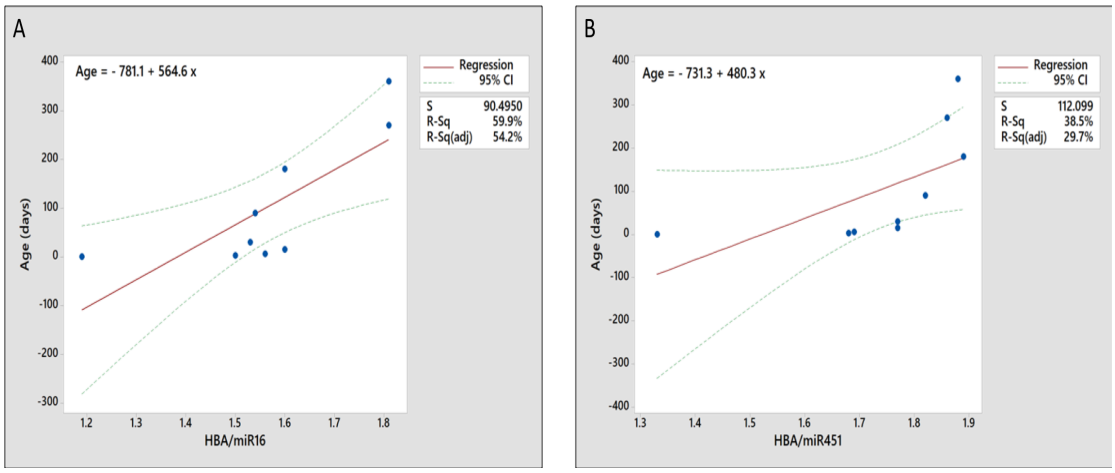


Figure 4.9: **Regression analysis of the RERs of blood samples over ageing time points, using a linear model with 95% confidence intervals.** RER of (A) *HBA/miR16* and (B) *HBA/miR451*. Data represent mean of n = 10.

In contrast, the second-order polynomial model shows a higher R^2 for the RER values for both *HBA*/*miR16* (R^2 = 83.9%) and *HBA*/*miR451* (R^2 = 77.2%). The differences in the 95% confidence intervals were less pronounced in the second model. Narrower confidence intervals and higher values of R^2 in the

second-order polynomial models makes them more reliable in predicting the age of bloodstains.

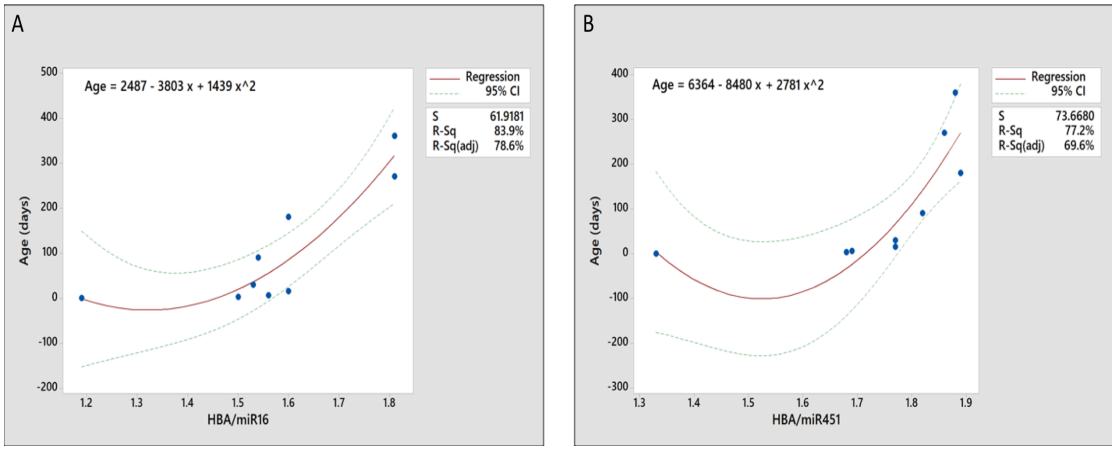


Figure 4.10: **Regression analysis of the RERs of blood samples over ageing time points, using a second-order polynomial model with 95% confidence intervals.** RER of (A) *HBA/miR16* and (B) *HBA/miR451*. Data represent mean of n = 10.

4.4.4.3.2 RERs of mRNA and miRNA markers to *U6*

The *geNorm* analysis described in section (4.4.4.1) above showed that the snRNA molecule *U6* was the next most stable marker after *miR451* and *miR16*, with high stability across all ageing time points during the one-year storage period. Therefore, the RERs of all the blood-specific markers to *U6* were calculated to examine the relationship with the age of bloodstains (Table 4.9), using the equation below:

$$RER = \frac{Cq\ of\ blood-specific\ \ marker}{Cq\ of\ reference\ gene\ U6}$$

Table 4.9: **Mean RER of *miR16*, *miR451*, *HBA* and *HBB* to *U6* for 10 samples at ageing time points 0, 3, 6, 15, 30, 90, 180, 270, 360 days.** The RERs were calculated from the mean C_q values in Table 4.4 and Table 4.5.

	RER			
Age points	*miR16*/*U6*	*miR451*/*U6*	*HBA*/*U6*	*HBB*/*U6*
Day 0	0.57	0.51	0.68	0.67
Day 3	0.59	0.52	0.88	0.84
Day 6	0.58	0.54	0.91	0.86
Day 15	0.61	0.55	0.97	0.88
Day 30	0.63	0.55	0.97	0.85
Day 90	0.64	0.55	0.99	0.85
Day 180	0.66	0.56	1.06	0.88
Day 270	0.68	0.66	1.21	0.95
Day 360	0.72	0.70	1.30	1.14

Figure 4.11 shows the RER values for blood-specific markers to the reference gene (*U6*). It was found that the RERs of *miR16*/*U6*, *miR451*/*U6*, *HBA*/*U6* and *HBB*/*U6* increase with increasing ageing time points. The ratios of *miR16*/*U6*, *miR451*/*U6* and *HBA*/*U6* appeared to increase in a linear fashion. In the early time points, the ratio of *HBB*/*U6* increased initially with ageing time points, then remained at a plateau phase before increasingly sharply again at 360 days. When comparing the variation observed in the RERs values, represented in the interquartile range and the whiskers extension in Figure 4.11, the RERs of the mRNA markers to *U6* had lower levels of variation than the RERs of the miRNA markers to *U6*, with the exception of stains aged for 270 days, which had the highest levels of variation in the *HBA*/*U6* and *HBB*/*U6* ratios.

A Spearman's correlation analysis indicated that there is a significant positive correlation between ageing time points and the RERs of *miR16/U6* and *miR451/U6* (both $r = 0.98$, $p < 0.0001$), and the RER of *HBA/U6* ($r = 0.99$, $p < 0.0001$). The RER of *HBB/U6* also shows a significant positive correlation but with a lower r-value compared to the other markers ($r = 0.84$, $p = 0.005$).

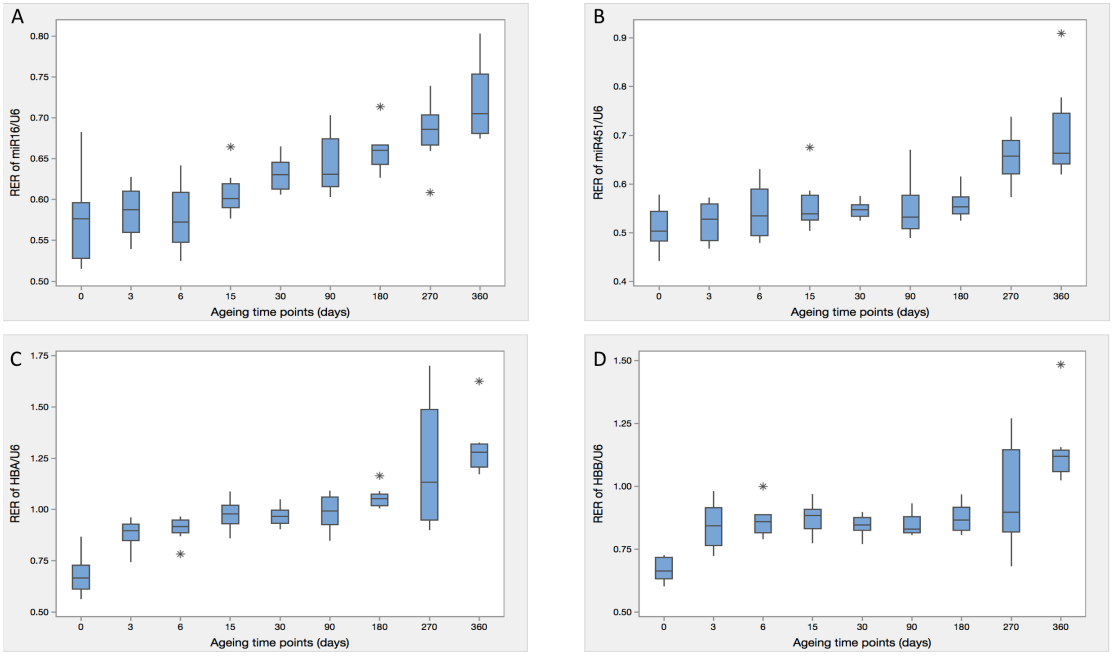


Figure 4.11**: Boxplots showing RER values of dried bloodstains.** RER of (A) *miR16/U6*, (B*) miR451/U6*, (C) *HBA/U6* and (D) *HBB/U6***.** The plots were obtained from data shown in Table 4.9 using Minitab Express, * represents outliers. n = 10.

As above, a regression analysis was performed to examine the relationship between ageing time points and the mean RERs of each miRNA to *U6* and each mRNA to *U6*. Each RER was fitted using both a linear model and a second-order polynomial, to identify the most reliable prediction model. The linear and the second-polynomial models for the RER of *miR16* and *miR451* to *U6* are illustrated in Figure 4.12 and Figure 4.13.

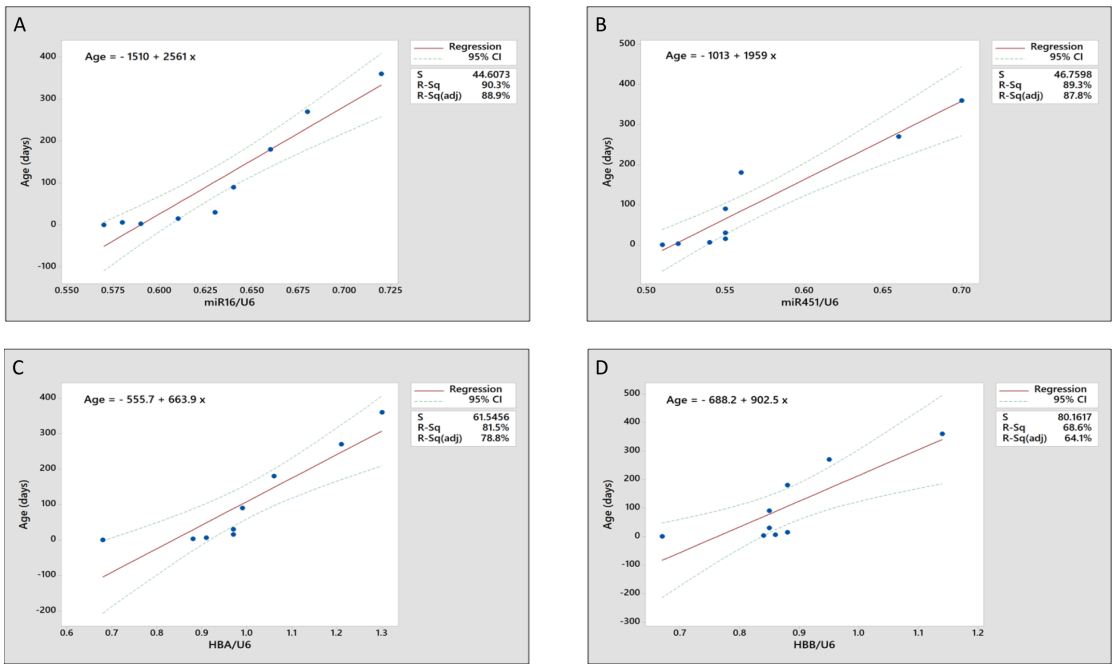


Figure 4.12: **Regression analysis of the RERs of blood samples over ageing time points using a linear model with 95% confidence intervals.** RER of (A) *miR16/U6*, (B) *miR451/U6*, (C) *HBA/U6*, and (D*) HBB/U6*. Data represent mean of n = 10.

Of particular note is that the regression analysis of the RER of *miR16/U6* using a second-order polynomial model gives a high R^2 value (R^2 = 95.5%), which is higher than for the linear model (R^2 = 90.3%). Regression analysis of the RERs of *HBA*/U6 and *HBB/U6* gave similar results, with higher R^2 values for the second-order polynomial models (R^2 = 94.7% and R^2 = 81.5% respectively), and the linear model gave R^2 values of 81.5% and 68.6% for *HBA/U6* and *HBB/U6* respectively. The analysis on the RER of *miR451/U6* produced identical R^2 values for the second-order polynomial and linear models (R^2 = 89.3%).

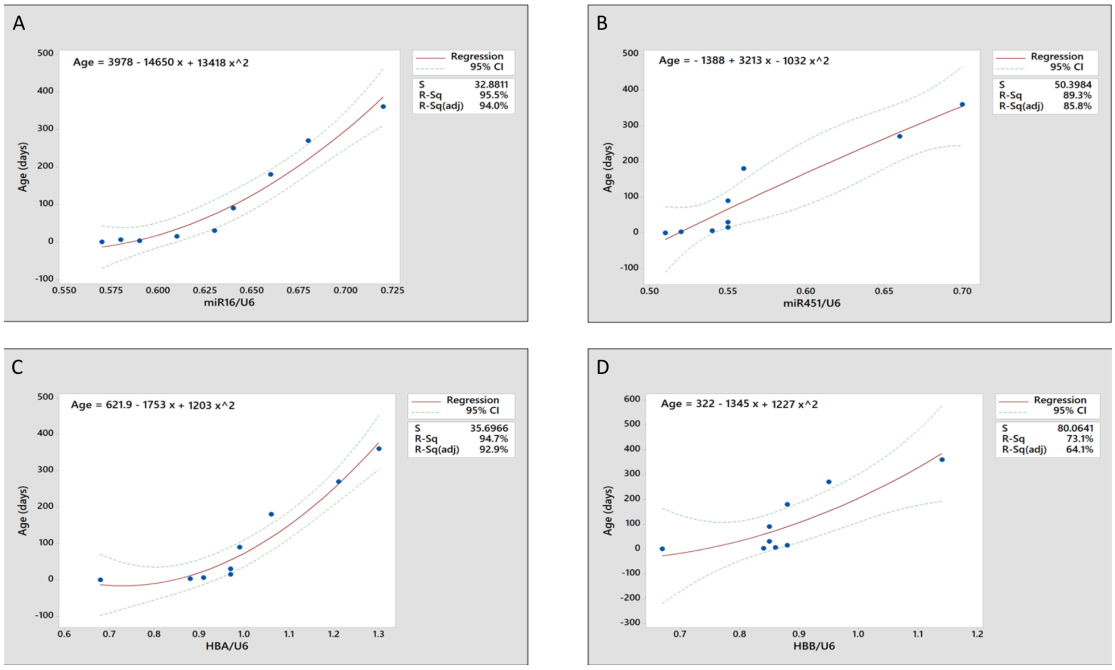


Figure 4.13: **Regression analysis of the RERs of blood samples over ageing time points using a second-order polynomial model with 95% confidence intervals and 95% prediction intervals**. RERs of (A) *miR16/U6*, (B) *miR451/U6*, (C) *HBA/U6*, and *HBB/U6*. Data represent mean of n = 10.

The confidence intervals for the *miR16*/*U6* and *HBA*/*U6* RERs were narrower in the non-linear model, making it more reliable in predicting the age of the bloodstains. This difference was more pronounced in the *miR16*/*U6* ratio, with a relatively small standard error (S = 32.9), suggesting that this ratio should be selected over the others for estimating the age of bloodstains. On the other hand, the confidence intervals were narrower in the linear model or relatively the same in both models in *miR451/U6* and *HBB/U6* respectively.

4.4.4.3.3 RERs of *HBA* to *HBB*

Finally, the ratio of the two blood-specific mRNA markers was also calculated. The stability of *HBB* was higher than both *HBA* and *HMBS*, which degraded rapidly across the 360 day ageing period, or completely degraded after 15 days, respectively. The RER of *HBA* to *HBB* was found to increase with time (see Table 4.10).

Table 4.10: **Mean RER of *HBA* to *HBB* for 10 samples at ageing time points 0, 3, 6, 15, 30, 90, 180, 270, 360 days.** The RERs were calculated from the mean C_q values in Table (4.4) and (4.5).

Age points	RER of *HBA*/*HBB*
Day 0	1.01
Day 3	1.05
Day 6	1.05
Day 15	1.11
Day 30	1.14
Day 90	1.17
Day 180	1.21
Day 270	1.27
Day 360	1.09

Figure 4.14 shows the RER values for *HBA*/*HBB*, which were found to increase with increasing ageing time points. The RER changed in a linear fashion in the first 270 days and then dropped slightly at 360 days due to the degradation of the more stable marker, *HBB*. The variation observed in RER values was relatively small at most ageing points, with the exception being the 270 day time point, which exhibited higher levels of variation..

A Spearman's correlation analysis indicated that there is a significant positive correlation between ageing time points and the RER of *HBA* to *HBB* ($r = 0.75$, $p = 0.02$). When analysing samples only up to 270 days of storage, i.e. before *HBB* started to degrade, the significance was even higher ($r = 0.99$ and $p < 0.0001$). Therefore, the subsequent regression analysis was performed for the ageing time points from 0 to 270 days only.

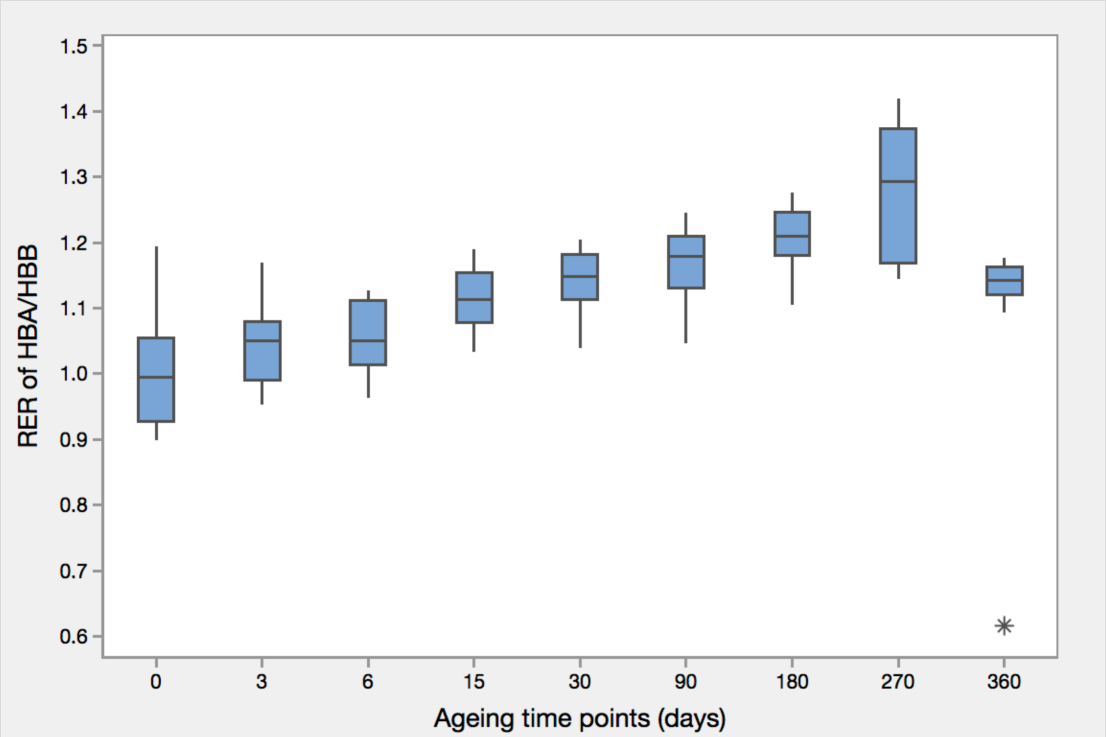


Figure 4.14: **Boxplot showing RER values of *HBA/HBB* in dried bloodstains.** The plot was obtained from data shown in Table 4.10 using Minitab Express. n = 10.

Both linear and second-order polynomial models were fitted to the *HBA*/*HBB* data, to identify the most reliable prediction model. The data shown in Figure 4.15 indicate that the second-order polynomial model gives a higher R^2 (R^2 = 97.9%) than the linear model (R^2 = 84.5%). The 95% confidence intervals in the second-order polynomial model are also smaller, giving narrower confidence limits. The second-order polynomial model also has a substantially smaller standard error (S = 17.33), which indicates that the prediction ability of this model is better for the estimation of the age of bloodstains.

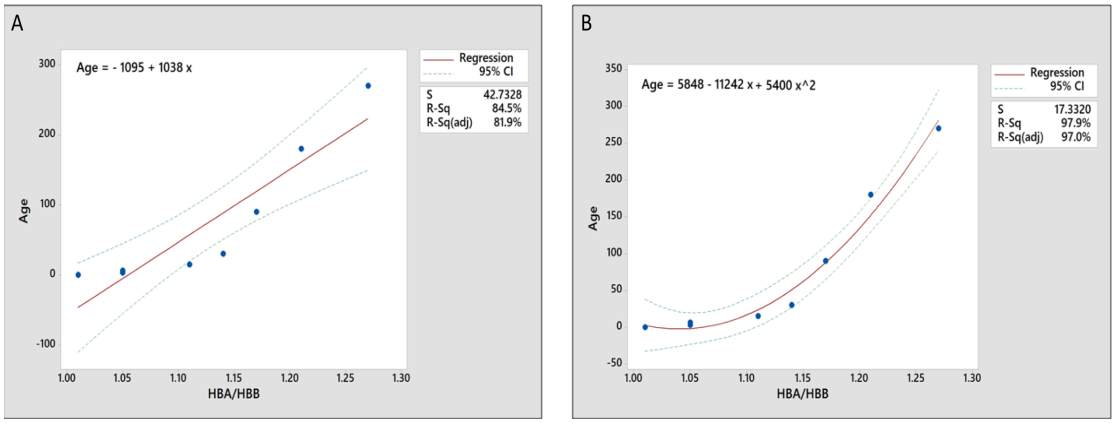


Figure 4.15: **Regression analysis of the RER of *HBA/HBB* of blood samples over ageing time points (up to 270 days) using two models with 95% confidence interval.** (A) linear model and (B) a second-order polynomial model. Data represent mean of n = 10.

4.4.4.3.4 Inter-donor variation at each ageing time point

In order to use the RERs of blood-specific markers in forensic casework for the estimation of the age of bloodstains, the amount of variation in this measure between different individuals should be considered. Therefore, the inter-donor variation for each RER that was discussed above was statistically evaluated at all ageing time points.

Table 4.11 shows the standard deviation and the coefficient of variation for each of the RERs. The RERs that include *HBA* as the numerator showed high variation at day 0 and day 270, which is consistent with the inter-donor variation observed when examining the degradation rate of this marker (section 4.4.3). This variation may be due to the variation in the biological composition of each stain from different individuals, as well as technical variation during setting up experimental reactions.

Table 4.11: **Statistical analysis of RERs for blood-specific RNA markers for the 10 samples at ageing time points 0, 3, 6, 15, 30, 90, 180, 270, 360 days.** The coefficient of variation was calculated using the equation CV% = standard deviation/mean x 100. The data analysis was carried out using Microsoft Excel.

Age points	*HBA/miR16*		*HBA/miR451*		*HBA/HBB*[c]		*miR16/U6*		*miR451/U6*		*HBA/U6*	
	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]
Day 0	0.17	14.53	0.15	11.25	0.09	8.79	0.05	8.91	0.04	7.63	0.09	13.00
Day 3	0.10	6.57	0.14	8.49	0.06	5.89	0.03	4.96	0.04	6.95	0.06	7.20
Day 6	0.07	4.77	0.16	9.33	0.06	5.43	0.04	6.99	0.05	9.26	0.05	5.90
Day 15	0.13	8.03	0.19	10.86	0.05	4.06	0.03	4.16	0.05	8.90	0.07	6.96
Day 30	0.05	3.54	0.05	2.74	0.05	4.35	0.02	3.24	0.02	2.81	0.04	4.52
Day 90	0.09	6.09	0.16	8.63	0.06	4.95	0.03	5.13	0.05	9.76	0.08	8.01
Day 180	0.03	1.98	0.08	4.41	0.05	4.07	0.02	3.63	0.03	5.03	0.05	4.35
Day 270	0.53	29.00	0.52	27.72	0.10	8.24	0.04	5.15	0.05	7.23	0.28	23.45
Day 360	0.07	12.70	0.09	15.64	-	-	0.01	6.14	0.03	12.60	0.04	9.89

[a] Standard deviation [b] Coefficient of variation [c] Up to 270 days only

The relative expression ratios of *miR16/U6* and *HBA/HBB*, which produced the lowest estimation error (i.e. lower standard error, or S values) as shown above, also have low inter-donor variation across ageing time points. This low level of variation can result in a high accuracy in estimating the age of bloodstains, which is a very promising outcome. However, to confirm such findings, more research is needed, to address some of the limitations of this work. For example, the examined number of samples in this project was small (n = 10), therefore further experiments should be carried out, increasing the number of samples as well as applying different environmental conditions (e.g. different temperature, humidity and levels of UV light) to test whether these conditions have an effect on the resulting RER values.

4.5 Summary and Conclusion

The main focus of this project is the application of RNA body fluid-specific markers in determining the age of biological stains, allowing forensic experts to identify the type of body fluid and the time since deposition simultaneously. In the context of forensic science laboratories, these techniques require much more research before consideration could be given to applying them to forensic casework.

To select RNA markers to be used for estimating the age of bloodstains, the degradation rate of these markers needs to be large enough that it is observable across ageing time periods. In this work, the degradation rate of blood-specific RNA markers in aged samples was analysed. A total of ten volunteers (6 females and 4 males) donated blood samples, which were deposited on cotton swabs and then stored in a dark, dry place at room temperature to simulate natural ageing, until they reached a series of desired ages (0, 3, 6, 15, 30, 90, 180, 270, 360 days). The degradation levels of six RNA molecules were analysed using TaqMan® assays and the RERs were calculated to study the degradation behaviour of the markers and their relationship with age.

The outcomes of this research showed that different RNA molecules degrade at different rates in bloodstains, with miRNA markers exhibiting strong stability, likely due to their small size. By applying correlation tests, Log $2^{-\Delta Cq}$ of blood-specific markers have shown a negative correlation with ageing time points, indicating that these markers could be used in determining the age of bloodstains. Similarly, correlation tests and regression analysis have shown that the data from RERs of blood-specific markers have a positive correlation with ageing time points, and may be useful in estimating the age of bloodstains. Regression analysis examining the relationship between the RERs of *miR16*/*U6* and *HBA*/*HBB* with ageing time point produced the highest R^2 values (98.6% and 97.9% respectively) with narrow confidence intervals, using a non-linear model. These results indicate that these two ratios may be the most reliable in bloodstain age estimation.

The RERs of blood-specific markers represent a potential method to estimate the age of bloodstains. The findings of this study therefore emphasise that, in future, methods using RT-qPCR are likely to be more sensitive for the accurate determination of the age of bloodstains than the analysis of protein degradation or variation in solubility and morphological differences in bloodstains [88].

Chapter five: Quantification of degradation in saliva- and semen-specific RNA markers to estimate body fluid stain age using RT-qPCR

5.1 Introduction

In addition to blood, which was considered in Chapter four, saliva and semen are the other biological stains that are commonly found at a crime scene. In a forensic context, there are no practical methods available to estimate the age of saliva or semen evidence, or to determine the time since deposition. However, one study did attempt to use the same method as Anderson et al. (2005) described and apply it to saliva samples [87]. To the author's knowledge, this thesis is therefore the first study to investigate the degradation rate of semen-specific RNA markers and apply it to the estimation of the age of semen stains.

Chapter four evaluated the possibility of using blood-specific RNA markers for estimation of bloodstain age, and in this Chapter, a similar evaluation was carried out using saliva- and semen-specific RNA markers. The aim was to look for a correlation between the age of the saliva and semen stains and the degradation rate and relative quantity of specific mRNA and miRNA markers. The degradation pattern of different RNA molecules in saliva and semen samples was characterised in order to assess the possibility of using the degradation rate to determine the time since deposition of the stains. TaqMan® assays were used for the quantification method: TaqMan® Gene Expression Assays for saliva- and semen-specific mRNA markers, and TaqMan® MicroRNA Assays for saliva- and semen-specific miRNA markers.

The same approach to relative quantification used in the previous Chapter was applied to determine the relative expression of saliva-specific and semen-specific RNA markers over time: the relative expression ratio (RER). The

method can be used to determine the relative quantification of a less sable marker to a more stable marker across ageing points.

5.1.1 Saliva-specific RNA markers

There are a number of mRNA molecules that have been identified as saliva-specific markers, such as Statherin (*STATH*) and Histatin 3 (*HTN3*) [64]. Both genes are protein-coding genes and are associated with the salivary secretion pathway, and their expression is restricted to the salivary glands [169]. Moreover, a number of miRNA markers have been identified as saliva-specific markers, because they exhibit higher expression levels in saliva samples, including *miR205* and *miR658* [66, 81]. As miRNA molecules, these markers are involved in the regulation of gene expression, affecting the stability and translation of mRNAs. *miR205* has been found to act as a crucial tumour suppressor in breast cancer [179], and *miR658* has been found to be overexpressed in gastric cancer [180].

5.1.2 Semen-specific RNA markers

Similarly, many studies have identified mRNA and miRNA markers that are specific to semen and seminal fluid. The common mRNA markers that are associated with sperm are Protamine 1 (*PRM1)* and Protamine 2 (*PRM2)*, and Kallikrein 3 (*KLK3)* and Semenogelin 1 (*SEMG1)* for seminal fluids. The expression of both *PRM1* and *PRM2* is restricted to testes, as they code for protamine peptides that form a highly condensed and stable complex with DNA in spermatids and spermatozoa [181]. *KLK3*, which is also known as Prostate-specific Antigen (*PSA*) is highly expressed in the prostate and, with *SEMG1,* is responsible for encasing ejaculated spermatozoa and allowing them to acquire progressive motility [169]. Additionally, there are a number of miRNA markers that have been identified as semen-specific, including *miR10b* and *miR891a*. These two miRNAs have been reported to have higher expression levels in the epididymis (male reproductive tract) than any other body fluids or tissues [66, 82, 182].

5.1.3 TaqMan® Assays

Similar to the blood-specific markers, the TaqMan® Gene Expression Assays and TaqMan® MicroRNA Assays that were used in this study are off-the-shelf and have been predesigned for each transcript by Applied Biosystems.

Table 5.1 and Table 5.2 show the markers that have been selected from the literature for the current project. Each assay has its own unique ID given by Applied Biosystems. The criteria that should be met in order to permit the selection of the predesigned TaqMan® Gene Expression Assays for use in quantifying mRNA markers are the same as those stated for blood-specific markers (section 4.1.2).

Table 5.1: **Characteristics of the TaqMan® Gene Expression Assays for saliva and semen-specific markers.**

RNA target	Body fkuid-specific	Applied Biosystems TaqMan® assay ID	Amplicon length (nt)	Assay location*
STATH	saliva	Hs00162389_m1	90	165
HTN3	saliva	Hs00264790_m1	136	302
PRM1	semen	Hs00358158_g1	99	205
PRM2	semen	Hs04187294_g1	73	388
SEMG1	semen	Hs00268141_m1	82	1525

* Refers to the nucleotide location that is the midpoint of the target region.

Table 5.2: **Characteristics of the TaqMan® MicroRNA Assays for saliva and semen-specific markers and *U6*.**

miRNA target	Body fluid-specific	Applied Biosystems TaqMan® assay ID	Target sequence
miR205	saliva	000509	UCCUUCAUUCCACCGGAG UCUG
miR10b	semen	002218	UACCCUGUAGAACCGAAU UUGUG
miR891a	semen	002191	UGCAACGAACCUGAGCCA CUGA
U6	Reference gene	001973	GTGCTCGCTTCGGCAGCA CATATACTAAAATTGGAAC GATACAGAGAAGATTAGC ATGGCCCCTGCGCAAGGA TGACACGCAAATTCGTGA AGCGTTCCATATTTT

Most of the selected mRNA markers met the above criteria, except *HTN3,* which can very weakly detect *HTN1* (i.e. the C_q values would be 10-15 cycles greater compared to the C_q values of *HTN3* expression level). It was shown in the validation work reported in Chapter three that this marker had a low detection level or low expression level in the saliva samples examined, therefore, the *HTN3* marker was removed from further analysis. The TaqMan assays of *PRM1* and *PRM2* may detect genomic DNA, which could generate false positive data. It is therefore necessary to reduce this effect and eliminate genomic DNA contamination, which can be done through treatment with DNase (section 2.5). Additionally, a portion of the extracted RNA sample was taken through the cDNA amplification step, including everything in the reaction mixture except the reverse transcriptase enzyme, as a control to ensure the removal of contaminating genomic DNA had been successful.

5.2 Aims and objectives

The overall purpose of this research was to develop a method to estimate the deposition time of biological fluids commonly encountered in forensic casework – blood, saliva and semen – using the application of RNA analysis. In Chapter four, the degradation profiles of multiple selected blood-specific

RNA markers (mRNAs and miRNAs) were analysed, including a reference gene (*U6*). In this Chapter, the degradation profiles of saliva- and semen-specific mRNA and miRNA markers were analysed, along with *U6*, to look for a correlation between the age of the saliva/semen stain and the degradation rate of the selected markers. This was carried out in order to identify those markers most useful for body fluid stain ageing and to assess the possibility of using the degradation rate to determine the time since deposition.

One aspect of this Chapter was therefore to analyse the expression level and study the degradation behaviour of multiple RNA transcripts in dried saliva and semen stains stored at room temperature for up to one year. Saliva-specific markers (*STATH*, *miR205*) and semen-specific markers (*PRM1*, *PRM2*, *SEMG1*, *miR10b*, *miR891a*) have been selected from a thorough literature review and the outcomes of Chapter three, and have been shown to indicate the presence of saliva/semen. Another aspect was to study the association between mRNA and miRNA stability, and to calculate the relative expression of these two different RNA molecules. This was done in order to determine whether this relative expression ratio can provide information about which markers are likely to be more accurate for use in estimating the age of biological stains, both over the short- and long- term.

Finally, studying the behaviour of RNA transcripts over a period of time can help to determine for how many days after depositing a saliva or semen stain the RNA within them remains of good enough quality for analysis, and how quickly it becomes so highly degraded that the expression analysis becomes unreliable.

5.2 Materials and methods

5.2.1 Sample collection

A total of 19 volunteers (5 females and 14 males), 10 donors donated saliva samples and 9 donors donated semen samples. The saliva and semen

samples were deposited into sterile collection pots. A volume of 50 µL of the samples were pipetted on to sterile cotton swabs and then stored in a dark dry place at room temperature to simulate natural ageing until they reached the desired ages (0, 7, 14, 28, 90, 180, 270, and 360 days). The experimental procedures were approved by the Department of Pure and Applied Chemistry Departmental Ethics Committee (see Appendix A1).

5.2.2 Total RNA extraction and reverse transcription

RNA extraction was carried out using TRI Reagent® as described in section 2.4.1. The TURBO DNA-free™ I Kit was used to treat the extracted RNA to remove any genomic DNA (section 2.5). Reverse transcription was carried out using High-Capacity cDNA kit and MicroRNA cDNA Kit as described in sections 2.7.1 and 2.7.2.

5.2.3 Quantitative real-time PCR (qPCR)

All TaqMan assays were run in singleplex assays, following the procedure described in section 2.8. Amplification was performed using a Stratagene Mx3005P.

5.2.4 Data analysis

The data generated from RT-qPCR was analysed using *MxPro* and *GenEx* software (version 5.4.4) was used for efficiency correction of the raw data (see section 2.9). *Microsoft Excel 2016* was used to manipulate raw C_q data from the spectrophotometry and C_q values from the RT-qPCR reactions and to present basic data and line graphs. *Minitab®17* and *Minitab Express* (version 1.5.0) were used for statistical analysis, including the Anderson-Darling normality test, correlations and regression analysis.

5.3 Results and Discussion

5.3.1 Analysis of total RNA yield

Similar to the blood sample analysis described in Chapter four, the total RNA yield recovered was measured to determine whether the concentration of RNA extracted from dried saliva and semen decreased through degradation over time. UV-visible spectrophotometry was used to indicate the quantity of recovered RNA, and it can provide a general assessment of the RNA yield. The RNA yield was assessed for saliva and semen samples left under simulated conditions of natural ageing at room temperature for up to one year.

5.3.1.1 Total RNA yield: Saliva

After storing saliva samples over an interval of one year, no significant reduction in total RNA yield could be observed, as illustrated in Figure 5.1. The total RNA concentration between samples was highly variable, as shown by the wide error bars (i.e. larger standard deviations), with the time point at 14 days showing the largest amount of variation. This variation might be explained by the composition of ribonucleases in saliva, which includes both endogenous and exogenous ribonucleases [183]. These enzymes have variable activity and they also vary in their relative quantities from one individual to another, therefore RNA molecules may degrade at different rates and the quantity of RNA molecules may vary between samples. Additionally, the recovery of RNA using TRI® Reagent is very variable [173] and is affected by the user and how effectively they homogenise the sample and separate the phases, which can increase the variability between samples. In this project the same individual performed phases separation and all samples were homogenised for the same period of time in an attempt to minimise any variation. Moreover, saliva has a very high viscosity that may cause some difficulties in pipetting, which can lead to inaccurate volume transfer onto each swab, meaning that some samples may have lower volume than others. In an attempt to resolve this issue, the tips were cut in order to make them wider prior to pipetting samples.

When applying the Anderson-Darling normality test to determine whether the data was normality distributed or not, it was established that the yield of total RNA obtained at different time points was normally distributed (p = 0.9328).

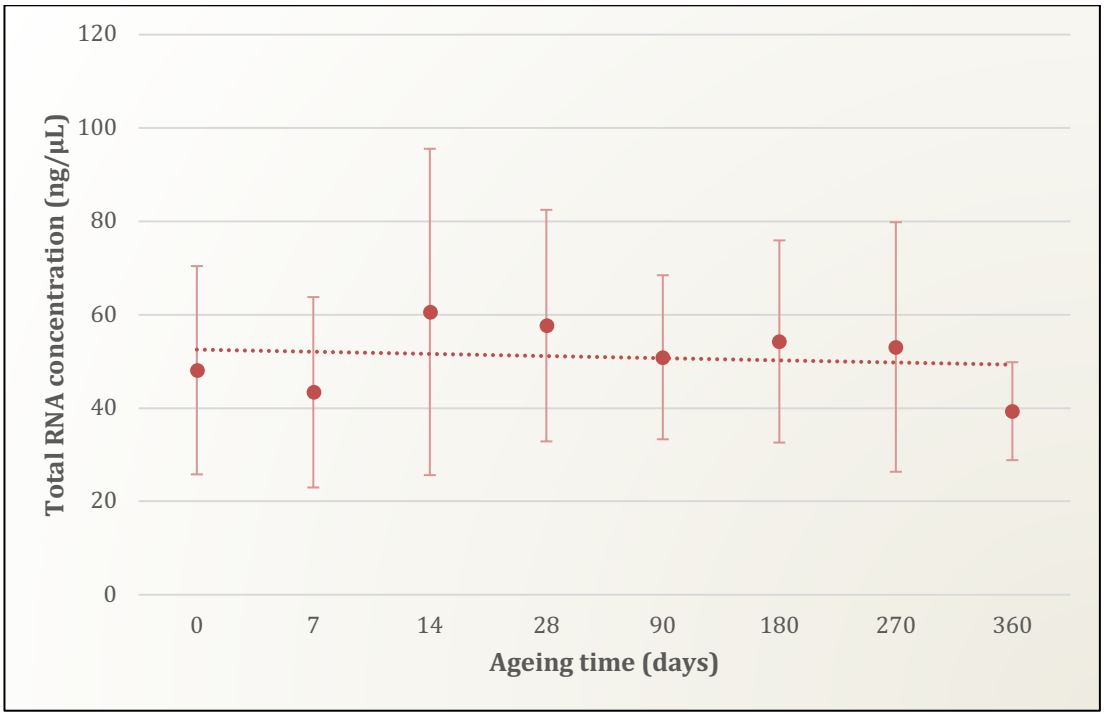


Figure 5.1: **Trend line analysis of total RNA yield from saliva samples stored up to one year at room temperature.** Points represent the mean ±SD of total RNA quantity determined by UV-Spectrophotometry (ng/µL) of saliva samples (n = 10 at each time point), the dotted line represents the best-fit line).

A Pearson's correlation test showed that no significant correlation was found between storage time and the total RNA yield (r = -0.3974, p = 0.3296).

In general, these results were expected in this investigation. The UV-spectrophotometry system that was used for RNA quantification in this experiment is very simple to perform and available in most forensic/clinical laboratories. However, it depends on the absorbance of ultraviolet light at 260 nm regardless of whether the RNA exists as an intact or fragmented molecule. Therefore, UV-spectrophotometry system is commonly used to indicate the extracted RNA sample quantity for downstream analysis, rather than RNA

sample quality, as it cannot indicate RNA degradation. However, the key outcome here is that the RNA yield is not affected by ageing of saliva samples, and so if there would be any differences in the quantification level of individual RNA markers, that would reflect the degradation of these markers and not a reduction in quantity due to reduced yield of the RNA extraction.

5.3.1.2 Total RNA yield: Semen

The total RNA yield was also recovered from semen samples that were stored for up to one year. Again, no significant reduction in total RNA yield could be observed, as illustrated in Figure 5.2. The total RNA concentration between samples was highly variable, as shown by the wide error bars (i.e. larger standard deviations), with the samples at time points 0 and 14 days showing the largest amount of variation. The heterogeneity of RNA within an individual sperm samples could explain these variations, as each spermatozoon contains low abundances of RNAs some of which are localised to the nucleus and does not contain ribosomes [184]. Another explanation for this observed variation is the difference between semen and other sample types in the timing between collection of the samples and their preparation by pipetting them on cotton swabs. Samples were kept in the fridge for a maximum of three hours before pipetting onto swabs, but it is possible that might allow a greater level of ribonucleases to degrade the RNA molecules in the sample. Moreover, TRI® Reagent was the procedure that was utilised for RNA extraction, which can be very variable [173], and due to the highly disulphide bonds, lysis extraction could be difficult [184]. Finally, similarly to saliva, semen is a very viscous fluid making pipetting difficult, which may introduce additional variation between samples. Similarly, in an attempt to resolve this issue, the tips were cut in order to make them wider prior to pipetting samples.

When applying the Anderson-Darling normality test to determine whether the data was normality distributed or not, it was established that the yield of total RNA obtained at different time points was normally distributed (p = 0.5263).

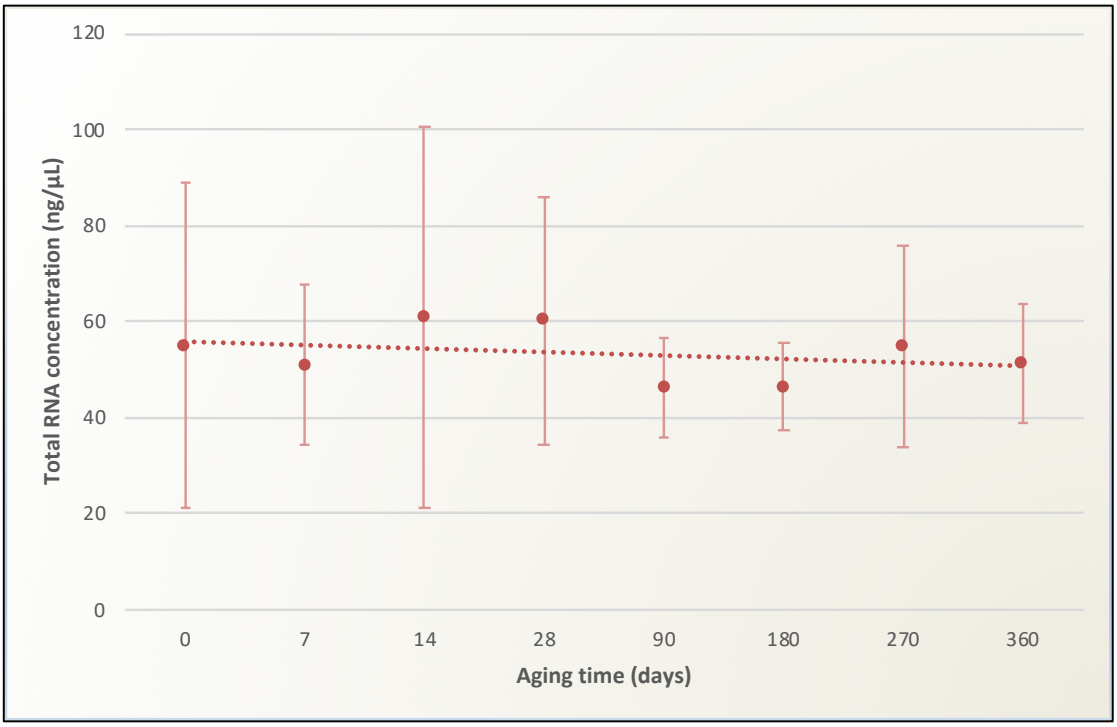


Figure 5.2**: Trend line analysis of total RNA yield from semen samples stored up to one year at room temperature.** Points represent the mean ±SD of total RNA quantity determined by UV-Spectrophotometry (ng/µL) of semen samples (n = 9 at each time point), the dotted line represents the best-fit line).

A Pearson's correlation test showed that no significant correlation was found between storage time and the total RNA yield (r = -0.3435, p = 0.4048).

Similar to the findings for the saliva samples, these results were expected in this investigation, as the UV-spectrophotometry system that was used for RNA quantification in this experiment, which depends on the absorbance of ultraviolet light at 260 nm regardless of whether the RNA exists as an intact or fragmented molecule. Therefore, it is commonly used to indicate the extracted RNA sample quantity for downstream analysis, rather than RNA sample quality, as it cannot indicate RNA degradation.

5.3.2 Degradation rate of individual RNA transcripts at different ageing time points

5.3.2.1 Saliva-specific markers

In this Chapter, the expression level of individual RNA transcripts was quantified by RT-qPCR in saliva and semen samples stored at room temperature for up to one year. The gradual increase of C_q value across storage time can characterise the degradation of RNA. The assays used in this Chapter were designed to amplify only sections of the RNA and not the whole transcripts. The size range of the amplified mRNA was 73 to 136 nucleotides (Table 5.1) and all were the '3' Most Assay' i.e. the assay closest to the 3' end of the relevant transcript. The RT-qPCR data for the saliva-specific markers is presented in this section and demonstrated that each RNA transcript showed a unique pattern of degradation behaviour.

Figure 5.3 illustrates the mean C_q data for *STATH*, *miR205* and *U6* after efficiency correction (C_q values were corrected against the obtained efficiency of each assay, see section 3.3.4.1). At day 0, which represents the control samples (fresh samples), each of the RNAs examined had a different starting expression level. The miRNA marker *miR205* and the reference gene *U6* exhibited the highest expression level (lowest C_q value), and *STATH* had the lowest expression level (highest C_q value). Both RT- and negative controls that were performed to monitor possible contamination and residual genomic DNA showed no amplification, indicating that the amplification observed reflects only the specific RNA molecules expressed within a given sample.

The *STATH* mRNA marker showed very interesting behaviour, where its expression level remained stable in the first 14 days and it then started to degrade after that, showing high stability. The high stability of *STATH* makes it a very good candidate as an mRNA saliva-specific marker.

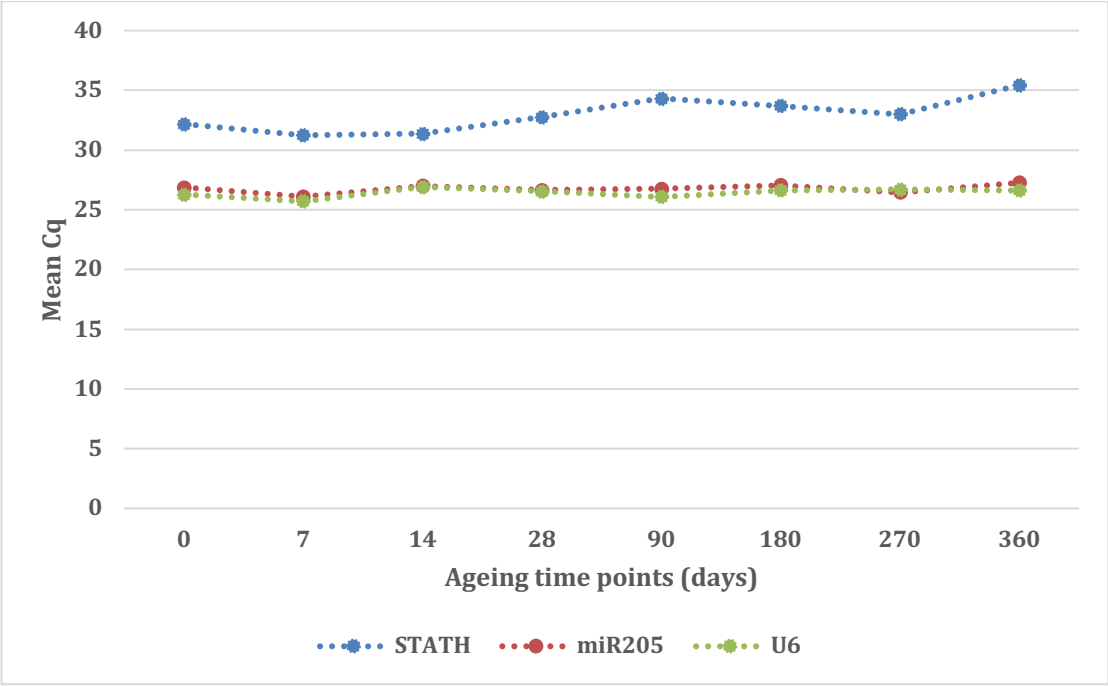


Figure 5.3: **The mean C_q data of *STATH*, *miR205* and *U6* in saliva samples stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. Error bars were removed for clarity.

The difference in C_q (ΔC_q), displayed in Figure 5.4, illustrates more clearly the degradation level of the RNA at each time point relative to the control point (T = 0 days). Across 360 days of storage, the ΔC_q value for *miR205* and *U6* remained around zero, which indicates no significant decrease in their quantity. When looking at the mRNA marker, the level of *STATH* only started to degrade at 28 days up to the 90 days ageing time point, and after that it exhibited small fluctuations in quantity, where the C_q value decreased at 180 and 270 days and increased again at 360 days.

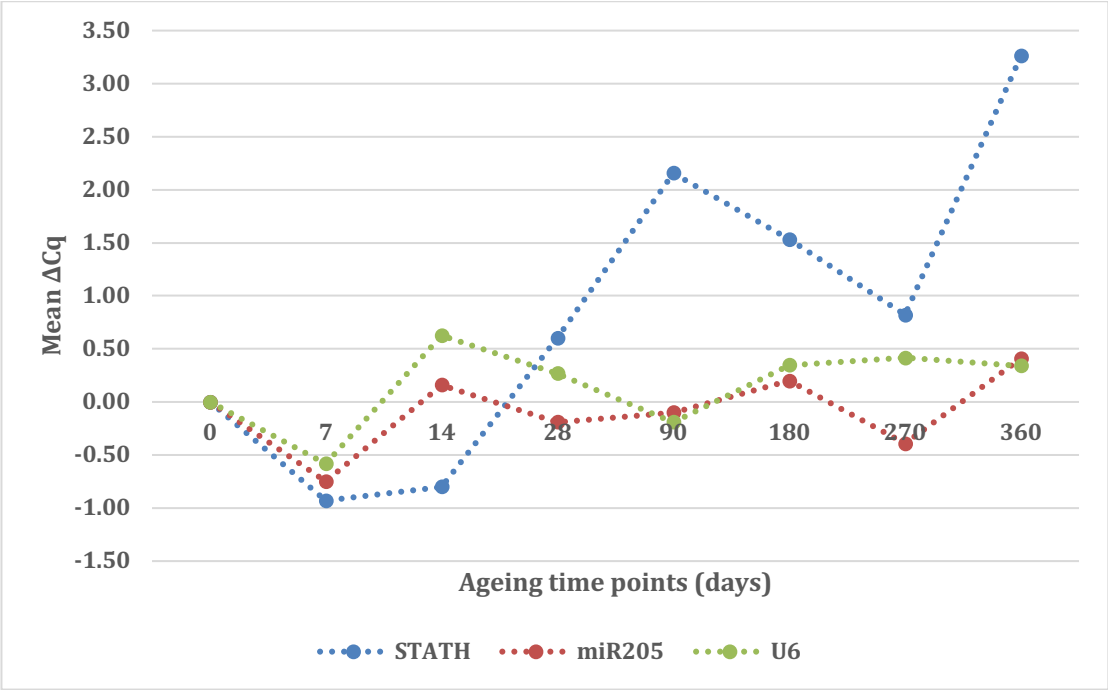


Figure 5.4: **The mean ΔC_q data of *STATH*, *miR205* and *U6* in saliva samples stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. Error bars were removed for clarity.

A quick visual assessment of the outcomes presented in Figure 5.4 suggests that the degradation of *STATH* did not begin until the 14 day ageing time points, indicating that some mRNA markers can stay stable for a few weeks. These findings do not support the results of the study by Sakurada et al. (2009) [185], where the C_q values of *STATH* increased in a linear fashion in saliva samples that were aged for up to one year. However, the results are in concordance with a more recent study by Watanabe et al. (2017) [186], who stored saliva samples for up to one year under dry conditions. Their study showed that the mean C_q of *STATH* remained at the same level for one month, then started to show gradual degradation, and was almost undetectable after one year of ageing.

As expected, the miRNA marker *miR205* showed high stability across all ageing points as its C_q value remained the same. Furthermore, the stability of the *U6* marker across all ageing time points in saliva samples is similar to that seen in blood samples, supporting that its structure makes it highly stable, as

its first nucleotide is methylated and the 3' end of this molecule is bound by a protein called the La protein which increases the stability of *U6* [4].

5.3.2.2 Semen-specific markers

The RT-qPCR data for the semen-specific markers is presented in this section, and demonstrates that each RNA transcript showed a unique pattern of degradation behaviour across ageing time points (Figure 5.5), similar to the results for blood and saliva.

Figure 5.5 illustrates the mean C_q data for *PRM1*, *PRM2*, *SEMG1, miR10*, *miR891a* and *U6* after efficiency correction (C_q values were corrected against the obtained efficiency of each assay, see section 3.3.4.1). At day 0, which represents the control samples (fresh samples), each of the RNAs examined had a different starting expression level. The mRNA marker *PRM2* exhibited the highest expression level (lowest C_q value), and *PRM1 and U6* had the lowest expression level (highest C_q value).

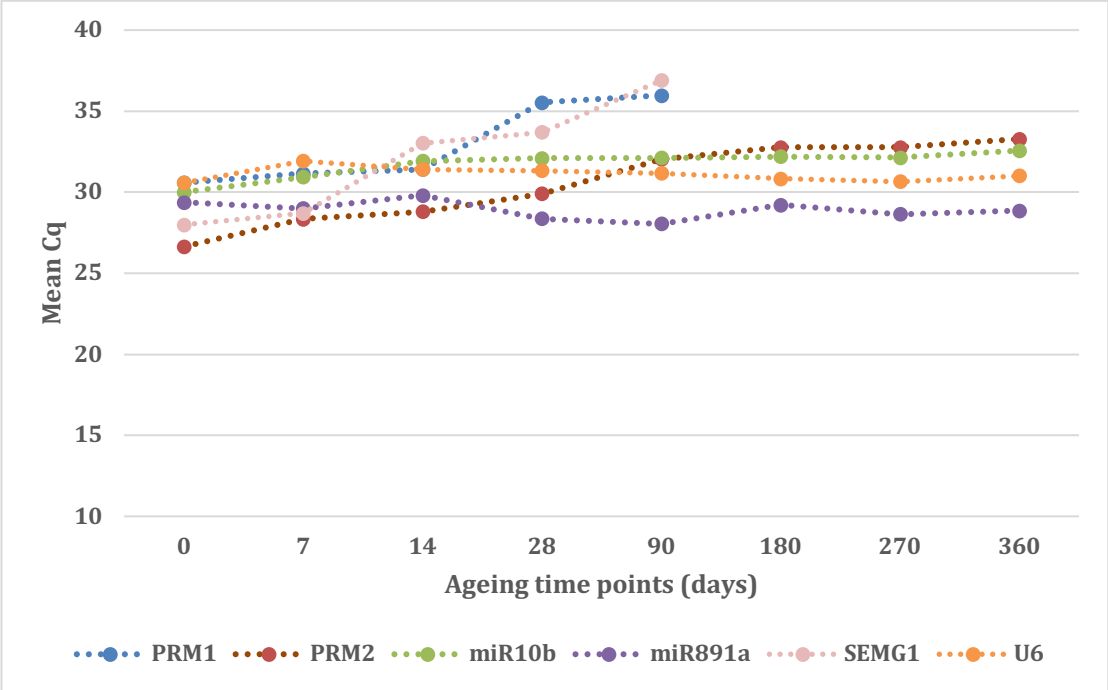


Figure 5.5: **The mean Cq data of *PRM1*, *PRM2*, *SEMG1*, *miR10b*, *miR891a* and *U6* in semen stains stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 9. Error bars were removed for clarity.

The ΔC_q values for the semen RNA markers, displayed in Figure 5.6, can illustrate more clearly the degradation level of the RNA at each time point relative to the control time point (T = 0 days). Across 360 days of storage, the ΔC_q value for *miR891a* and *U6* remained around zero, which indicates no significant decrease in their quantity. The miRNA marker *miR10b* behaved differently from the other miRNA markers examined in saliva and blood samples, as it did show slight degradation early in the ageing time period (T = 7 and 14 days), after which it remained at the same level between 14 to 270 days, before it degraded slightly again at 360 days.

With regards to the mRNA markers, both *PRM1* and *PRM2* degrade in a reasonably linear fashion. The degradation rate of *PRM1* started slowly at the start of the ageing period, but after 14 days its expression level dropped dramatically and fell below sensitivity level of the assay after 90 days. Setzer et al. (2008) found that *PRM1* did not degrade, and it remained stable for up to 180 days, which is not the case in this study [70]. However, Weinbrecht et al. (2014) examined the degradation rate of the semen transcriptome as a whole using next-generation sequencing, and showed that *PRM1* and *PRM2* degraded over 6 months of ageing [187]. The different quantification methods used in these studies could explain the difference in their results, as Setzer et al. (2008) used Ribo-Green® fluorescence assay which detects both human and nonhuman RNAs, while Weinbrecht et al. (2014) used RNA-sequencing. Nonetheless, the data in this work support the findings of the Weinbrecht et al. study.

In contrast, *PRM2* was detected across all ageing time points and it reaches a plateau after 90 days of ageing. The findings of Nakanishi et al. (2014) support the data presented for *PRM2* as they showed that it could still be detected in aged semen samples (33 and 56 years) [188]. The same study also included the *SEMG1* marker, which they showed was not detected in aged semen samples. This is in concordance with the current project, as *SEMG1* degraded rapidly, in linear fashion, and fell below the detection level of the assay after

90 days of ageing. Furthermore, the research of Weinbrecht et al. (2014) also examined the degradation rate of *SEMG1* in semen samples and showed that it did show similar degradation behaviour as *PRM1* and *PRM2* over time and it was still detected after 6 months of ageing [187].

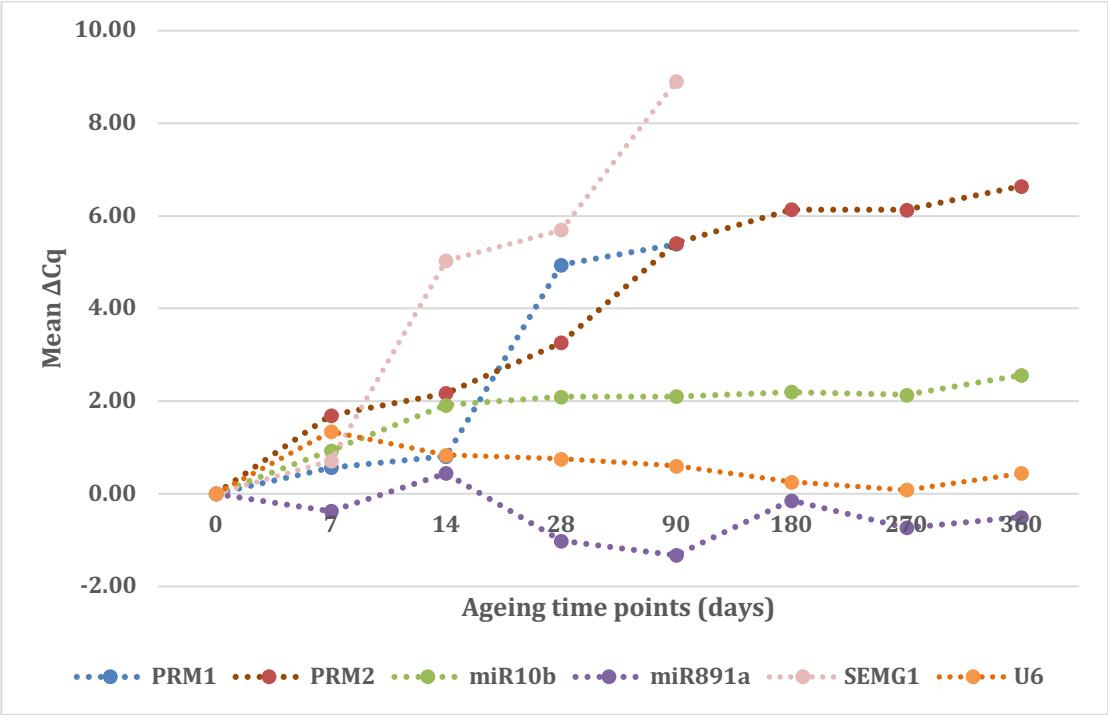


Figure 5.6: **The mean ΔC_q data of *PRM1*, *PRM2*, *SEMG1*, *miR10b*, *miR891a* and *U6* in semen samples stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 9. Error bars were removed for clarity.

Both *PRM1* and *PRM2* are genes specific to spermatozoa [168], whereas *SEMG1* is a gene specific to the prostate and seminal vesicles [189]. The unique degradation behaviour of these mRNA markers could be due to the stability of the cell membrane where they are located, as the cell membrane of the spermatozoon is more rigid than prostate gland [188].

With regards to the miRNA markers, high stability was observed in these markers during the one year storage period. Despite this, there was some variability in the stability of the different miRNAs, which were identified as semen-specific. The *miR891a* marker remained stable across all ageing

points, while *miR10b* started to degrade slightly after 7 days. Li et al. (2013) suggested that the different stability of some miRNAs could be explained by their different biological functions, and unstable miRNAs in most cell types tend to be involved in more functions than less stable miRNA (rapid production, rapid turnover; slow production, slow turnover) [45]. However, once a biological fluid is deposited and starts to dry, gene expression and biological processes stop, and only the activity of exonuclease and endonuclease enzymes continue, which degrade RNA molecules.

The high stability of *miR891a* demonstrated in this work agreed with the findings of Nashwa et al. (2017), where *miR891a* remained at the same level in semen samples that were aged for up to six months [190]. In contrast to this study, the *miR10b* marker showed high stability in semen samples stored for one year in the study by Tong et al. (2015) [191], allowing them to conclude that *miR10b* is a good candidate for a semen-specific marker. However, even though the difference in C_q measurements of *miR10b* at different ageing points in this study is not significant, it did show slight degradation over time. The explanation behind this unique behaviour of *miR10b* is unknown and could be due to the variation during sample handling, as mentioned above in section 5.3.1.2, it could be due to the timing between collection of the samples and their preparation by pipetting them on cotton swabs

Both RT- and negative controls that were performed to monitor possible contamination and residual genomic DNA showed no amplification, indicating that the amplification reflects only the specific RNA molecules expressed within a given sample.

5.3.3 Inter-donor variation at each time point

At each time point for each RNA marker, the inter-donor variation was calculated using the corrected C_q values to obtain the coefficient of variation

(CV%). Table 5.3 shows the mean inter-donor variation across all ageing time points for each of the selected RNA markers.

Table 5.3: **The mean inter-donor variation in saliva- and semen-specific RNA marker degradation rate across all ageing time points.** The coefficient of variation was calculated using the equation CV% = standard deviation/mean x 100.

RNA markers		Mean inter-donor variation (Coefficient of Variation %)
Saliva	*STATH*	7.51
	miR205	11.44
	U6	12.01
Semen	*PRM1*	7.75*
	PRM2	10.76
	SEMG1	6.69*
	miR10b	6.96
	miR891a	4.66
	U6	9.91

*Mean calculated up to 90 days as these markers were not detected after this time point.

The inter-donor variation in the saliva-specific mRNA marker (*STATH*) fluctuates widely across ageing time points, as illustrated in Figure 5.7, whereas the variation in the saliva-specific miRNA markers is more consistent over time. The highest inter-donor variation in *STATH* was observed at 28 days, in *miR205* at 90 and 270 days, and at 270 days for *U6*. There are a number of reasons why the composition of saliva could vary among donors, and it might be expected that inter-donor variation would be highest in saliva compared to other body fluid types, due the presence of large number of bacteria and different cell types. This is more noticeable with *U6* marker, as it

gave higher CV% value in saliva samples (12.01), while CV% value was 9.91 and 4.02 in semen and blood samples (see Chapter four) respectively. Even though all participants were instructed not to eat, smoke or brush their teeth for an hour before depositing a saliva sample, in order to minimise the inter-donor variation, there was still a high level of variation among donors. It is also worth considering that the presence of a high quantity of bacteria in saliva may inhibit PCR and increase variation among donors.

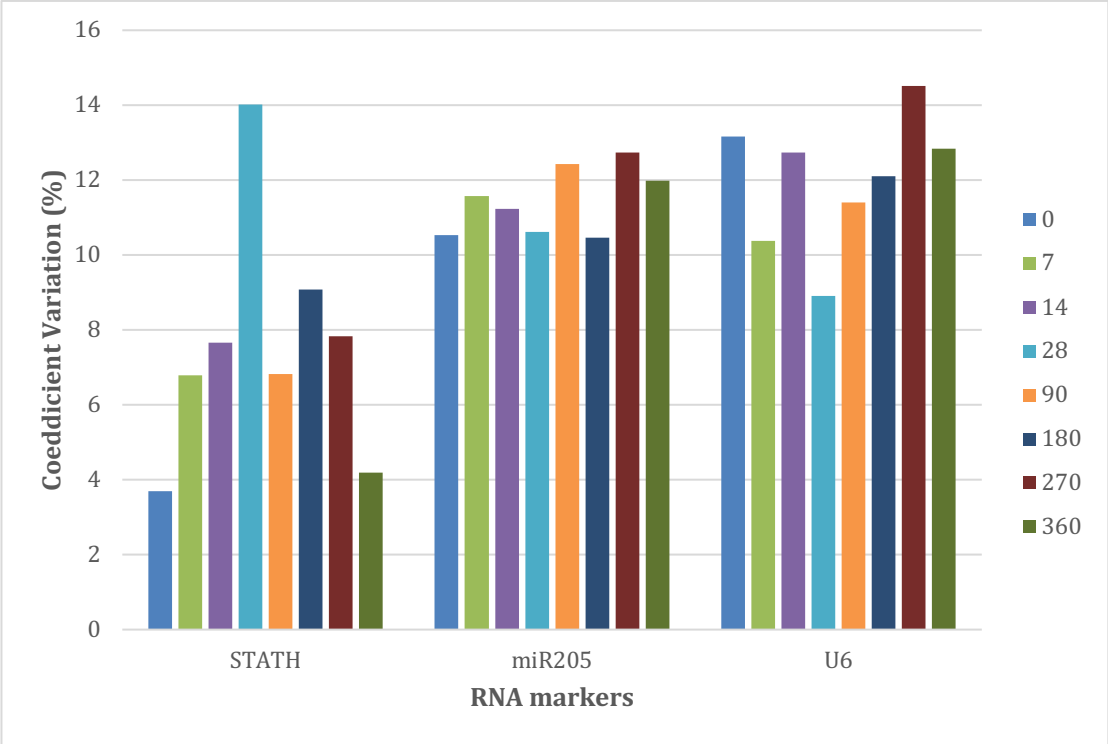


Figure 5.7: **Inter-donor variability in the degradation rate of RNA markers in saliva samples.** The Coefficient of variation was calculated using the mean of n = 10 for samples aged up to 360 days at room temperature. Coloured bars represent the different ageing time points.

The inter-donor variation in the semen-specific markers is shown in Figure 5.8. All markers showed a different level of variation among donors at all ageing tine points, and there were no consistent patterns. The highest inter-donor variation was observed at 180 days for the *PRM2* marker, and the lowest inter-donor variation was at 90 days in the *PRM1* marker.

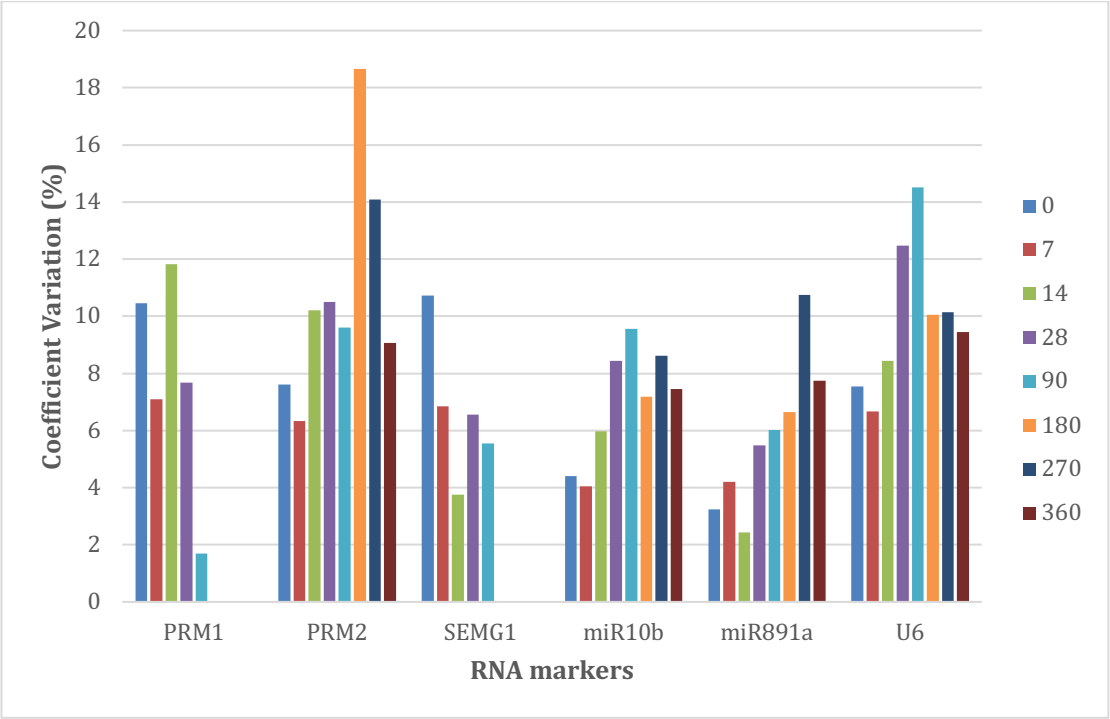


Figure 5.8: **Inter-donor variability in the degradation rate of RNA markers in semen samples.** The Coefficient of variation was calculated using the mean of n = 9 for samples aged up to 360 days at room temperature. Coloured bars represent the different ageing time points.

Generally speaking, factors such as overall health, age, diet and frequency of ejaculation can cause variation among participants in the composition of semen. Another explanation for the high inter-donor variation that is observed in semen samples could be laboratory error, which may be considered as a major source of variation. Many manual steps are involved in the preparation of the samples for ageing, which could lead to such variation. The sample size used in this experiment is relatively small (i.e. 10 saliva samples and 9 semen sample) and increasing sample number may be one way of decreasing inter-donor variation in future experiments.

5.3.4 Relative expression Ratio (RER)

The relative expression ratio (RER) was applied to determine the relative expression of different RNA markers for both saliva and semen over time. This approach was used to determine whether the RER could be used to estimate

the age of saliva or semen samples. The relative expression of a less stable RNA marker to a more stable marker across all ageing time points was calculated, after identifying the stability of each RNA marker. The relative expression of RNA markers over time: $2^{-\Delta Cq}$ was not applied in this experiment as it did not show any specific pattern (data not shown).

5.3.4.1 Saliva-specific markers

The mean C_q values for saliva-specific markers measured by qPCR at 0, 7, 14, 28, 90, 180, 270, and 360 days are shown in Table 5.4 for both mRNA and miRNA markers, along with *U6*. These data were corrected with the determined efficiency of each assay using *GenEx* statistical software (version 5.4.4) (see section 3.3.4.1).

Table 5.4: **The mean C_q values of saliva-specific mRNA and miRNA markers at 0, 7, 14, 28, 90, 180, 270, 360 days for the duplicates of 10 samples after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

Age points	*Markers*	Sample number/Mean C_q values									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Day 0	***STATH***	31.14	32.10	30.87	32.30	30.90	33.67	32.67	31.47	34.47	32.05
	miR205	25.44	25.32	25.63	25.79	26.40	32.15	32.19	24.97	25.15	25.50
	U6	22.57	25.45	22.38	25.66	22.98	31.69	31.93	25.08	26.79	28.27
Day 7	***STATH***	29.29	29.16	29.71	30.01	29.03	33.18	34.27	30.78	32.89	34.03
	miR205	25.88	25.71	24.31	27.08	25.08	31.70	30.37	21.67	23.35	25.87
	U6	22.93	26.19	22.57	26.86	22.85	30.39	29.25	24.53	25.04	26.36
Day 14	***STATH***	29.70	30.26	31.41	31.17	26.90	32.47	35.64	30.32	31.74	34.04
	miR205	26.07	28.06	26.29	27.06	24.68	31.40	32.55	22.86	24.25	26.87
	U6	23.67	23.74	23.74	27.55	22.61	30.74	33.01	26.13	25.82	27.94
Day 28	***STATH***	29.83	31.02	22.62	33.40	29.36	36.28	37.04	35.62	36.69	35.81
	miR205	26.74	26.99	26.33	30.29	26.02	27.70	31.84	22.67	23.51	24.52
	U6	24.45	27.07	23.93	29.10	24.16	29.55	30.16	24.48	26.08	26.49
Day 90	***STATH***	32.32	33.49	33.42	29.09	34.79	35.76	36.39	36.04	35.03	36.87
	miR205	28.03	28.95	27.74	30.62	26.70	29.14	29.73	21.98	21.96	22.67
	U6	25.32	30.20	24.79	30.16	25.75	27.49	28.82	22.01	22.56	23.80
Day 180	***STATH***	36.99	37.31	33.60	37.10	35.90	33.31	33.03	29.87	29.74	30.10
	miR205	27.35	26.89	25.84	27.66	25.42	31.38	31.56	22.32	24.64	27.44
	U6	22.96	26.58	22.60	28.68	23.09	31.72	30.99	26.26	25.67	27.73
Day 270	***STATH***	32.65	32.81	29.40	31.14	28.76	34.46	34.56	34.57	34.46	37.04
	miR205	29.04	29.17	27.26	28.67	27.00	29.62	27.74	19.54	22.18	24.36
	U6	28.87	32.56	27.30	32.03	27.79	26.78	24.97	21.21	21.92	23.53
Day 360	***STATH***	35.93	35.93	32.85	37.04	33.64	37.04	37.04	34.23	35.03	35.54
	miR205	30.60	30.31	29.47	26.23	25.70	29.44	30.96	22.62	22.90	24.39
	U6	22.75	32.47	28.67	26.92	22.85	30.88	27.66	22.61	26.20	25.19

5.3.4.1.1 Identification of the most stable RNA markers in saliva

The relative quantification of pairs of markers was calculated using the C_q value of a focal degraded marker and the C_q value of the most stable marker. In gene expression studies, a reference gene or endogenous control RNA has to be identified that is suitable for normalisation purposes. To fulfil this function, these control RNAs have to be stable, less variable than the focal marker, and should be expressed across various conditions. In the current study, the different ageing time points are considered the experimental conditions, so an endogenous control RNA must be stable across time points.

The software *geNorm* was used to assess the expression of RNA transcripts, and determine which RNA markers were most stable and exhibited the lowest variability across all time points. Figure 5.9 shows the data from *geNorm*, which measures the stability of a gene by calculating an M-value. The marker that shows the highest variation relative to all other markers is eliminated by *geNorm*. The average stability is then determined by the M-value for each pair-wise comparison; the lower the M-Value, the higher the stability. The software indicated that *miR205* and *U6* are the most stable markers across all time points, with the lowest M-Value (both 2.11).

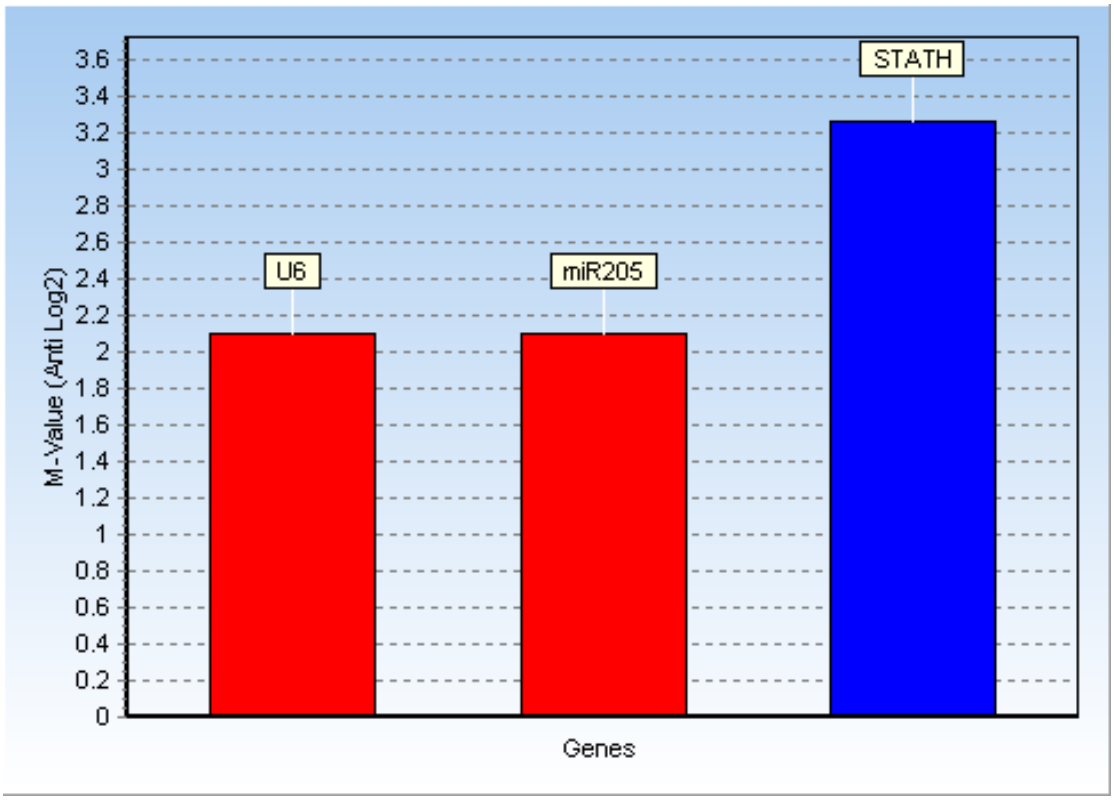


Figure 5.9: **The most stable RNA markers in saliva samples across all ageing time points measured by geNorm.** *miR205* and *U6* exhibited the lowest M-Values, indicating that they were the most stable markers.

5.3.4.1.2 Relative expression ratio (RER)

The ratio was obtained by dividing the efficiency-corrected C_q values of the less stable RNA marker by the efficiency-corrected C_q values of the more stable RNA marker;

$$RER = \frac{Cq\ of\ less\ stable\ marker}{Cq\ of\ more\ stable\ marker}$$

Only values up to 90 days were included in the analysis as the *STATH* marker fluctuated after this ageing time point, which may lead to unreliable data.

5.3.4.1.2.1 RER of mRNA to miRNA

The only saliva-specific miRNA marker that was examined was *miR205,* and this marker was also identified as the most stable across all ageing time points. Table 5.5 shows the mean RERs of *STATH* to *miR205,* calculated from the corrected C_q values for saliva samples stored up to 90 days at room temperature.

Table 5.5: **Mean RER of the mRNA marker *STATH* to *miR205* for 10 saliva samples at ageing time points 0, 7, 14, 28, 90 days.** The RERs were calculated from the mean C_q values in Table 5.4

Age points	***RER of STATH/miR205***
Day 0	1.21
Day 7	1.21
Day 14	1.17
Day 28	1.24
Day 90	1.31

The relationship between the obtained values of the RER and the ageing time points was statistically evaluated. Since an Anderson-Darling normality test showed that the data were normally distributed ($p = 0.3871$), a parametric analysis was used for statistical analysis.

Figure 5.10 presents the RER values for *STATH* to *miR205*. The RER of *STATH* to *miR205* increased slightly with increasing ageing time points, and this was a statistically significant positive correlation using a Pearson's correlation test ($r = 0.896$, $p = 0.039$). Despite the fact that the *STATH* quantification level remained stable across initial ageing time points, and did

not show any degradation until 28 days, the RER of *STATH*/*miR205* did show a specific trend up to 90 days.

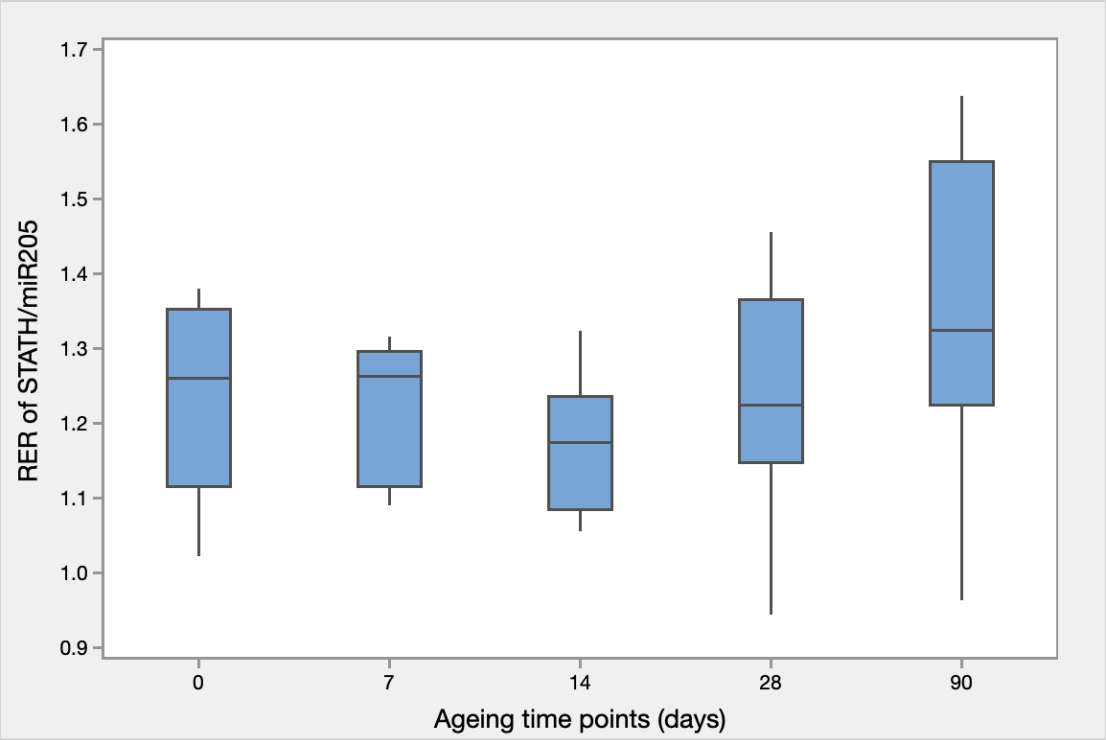


Figure 5.10: **Boxplots showing RER values for *STATH/miR205* in dried saliva samples.** The plots were obtained from data shown in Table 5.4 using Minitab Express. n = 10.

To investigate the relationship between the mean RER values of *STATH* to *miR205* and the ageing time points further, a regression analysis was performed. The data were fitted using two different models, a linear model (Figure 5.11) and a non-linear model (a second-order polynomial model, Figure 5.12). These two models were compared to each other, as well as to published studies reporting age prediction equations using linear models [58] and non-linear models [60] to estimate the age of biological samples. The models were compared to determine which is more accurate at predicting the age of saliva using saliva-specific RNA markers.

When fitting the data using a linear model, the analysis produced a high R^2 value for the RER of *STATH*/*miR205* (R^2 = 80.2%).

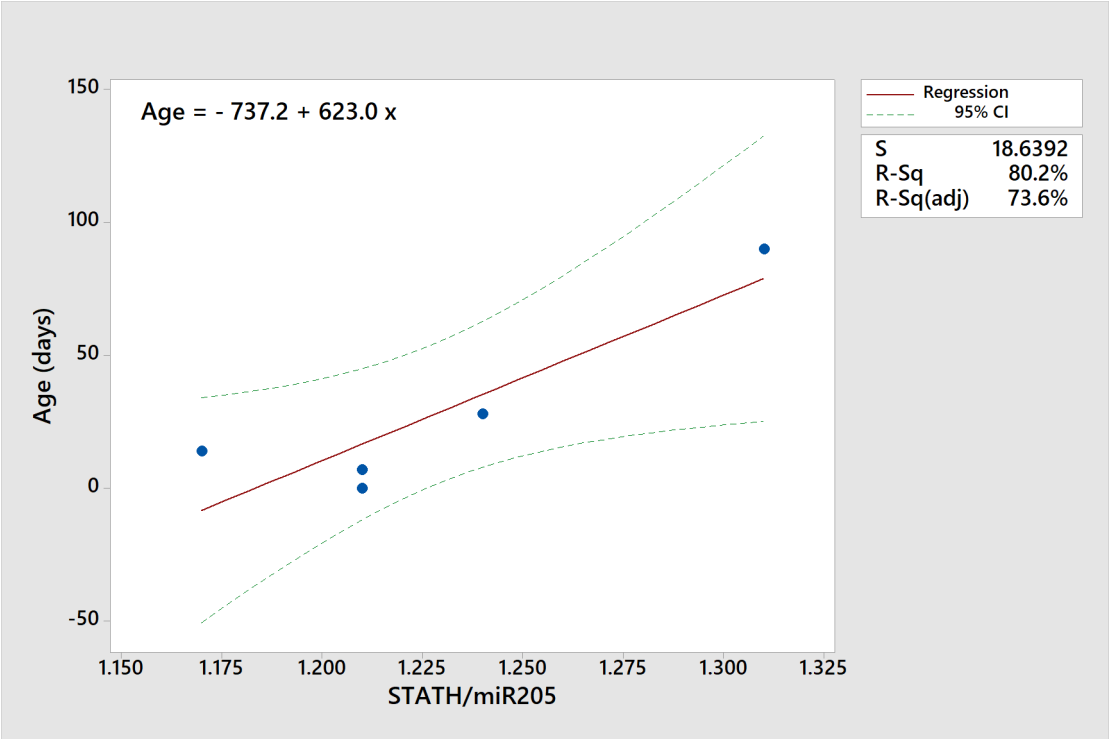


Figure 5.11: **Regression analysis of the RER of *STATH/miR205* in saliva samples over ageing time points, using a linear model with 95% confidence intervals**. n = 10.

In contrast, the second-order polynomial model shows a higher R^2 for the RER value for *STATH/miR205* (R^2 = 97.2%). The differences in the 95% confidence interval were less pronounced in the linear model (95% confidence intervals are narrower in the linear model), but the higher R^2 values in the second-order polynomial model makes it more reliable in predicting the age of saliva stains, with standard error of S = 8.6.

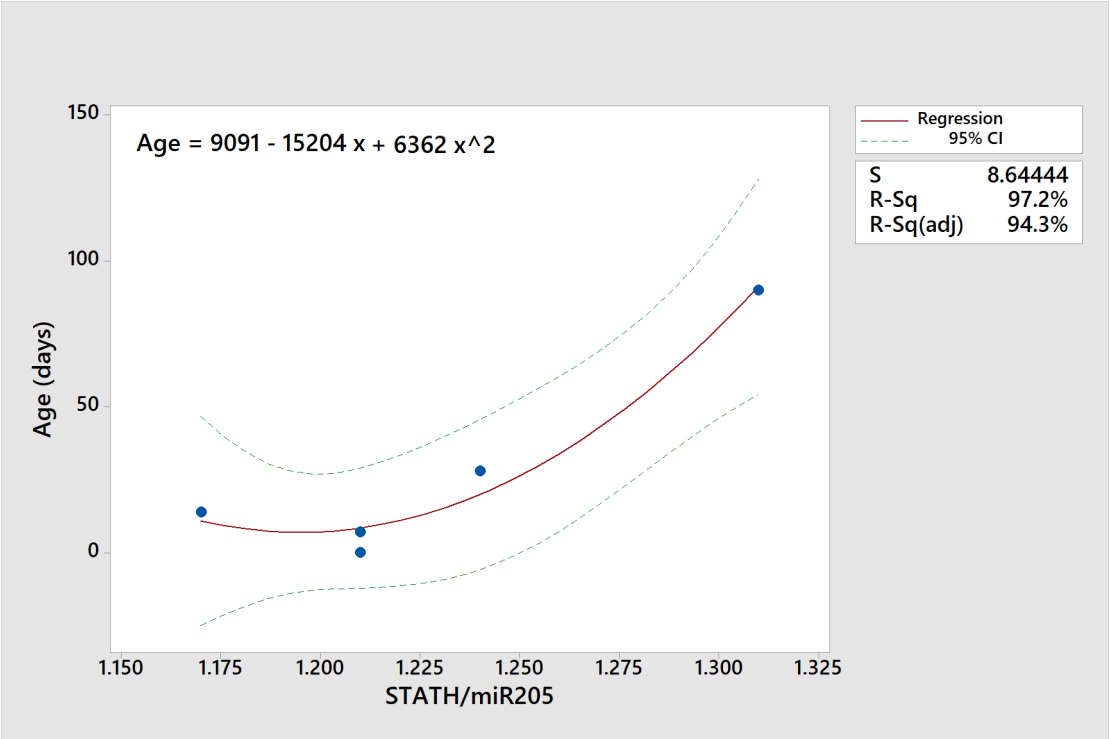


Figure 5.12: **Regression analysis of the RER of STATH/miR205 in saliva samples over ageing time points, using a second-order polynomial model with 95% confidence interval.** n = 10.

5.3.4.1.2.2 RERs of mRNA and miRNA markers to *U6*

The *geNorm* analysis described in section (5.3.4.1.1) above showed that the snRNA molecule *U6* was one of the most stable markers, along with *miR205*, with high stability across all ageing time points during the one year storage period. Therefore, the RERs of all saliva-specific markers to *U6* were calculated to examine the relationship with the age of bloodstains (Table 5.6), using the equation below.

$$RER = \frac{Cq\ of\ saliva - specific\ \ marker}{Cq\ of\ reference\ gene\ U6}$$

Table 5.6: **Mean RER of *STATH* and *miR205* to *U6* for 10 saliva samples at ageing time points 0, 7, 14, 28, 90 days.** The RERs were calculated from the mean C_q values in Table 5.4.

	RER	
Age points	*STATH/U6**	*miR205/U6*
Day 0	1.24	1.03
Day 7	1.22	1.00
Day 14	1.17	1.01
Day 28	1.23	1.01
Day 90	1.34	1.03
Day 180	-	1.01
Day 270	-	1.00
Day 360	-	1.03

*RER was calculated only up to 90 days as this marker fluctuated after this time point.

Figure 5.13 presents the RER values for saliva-specific markers to the reference gene (*U6*). It was found that the RERs of *STATH/U6* and *miR205/U6* did not increase with increasing ageing time points, and the ratios of *STATH/U6* and *miR205*/*U6* did not behave in a linear fashion. In the early stages, there was no increase in the *STATH/U6* ratio with time, as *STATH* remained stable at the same level. The ratio showed an increase only after 28 days when *STATH* started to degrade. When comparing the variation in RER values, the RERs of *STATH* to *U6* had higher levels of variation than *miR205* to *U6*, with the exception of stains aged for 7 days, which had the lowest level of variation in the *STATH/U6* ratio.

A Pearson's correlation analysis indicates that there is no significant correlation between ageing time points and the RERs of both *STATH/U6* ($r = 0.844$, $p = 0.072$), and *miR205/U6* ($r = 0.460$, $p = 0.435$).

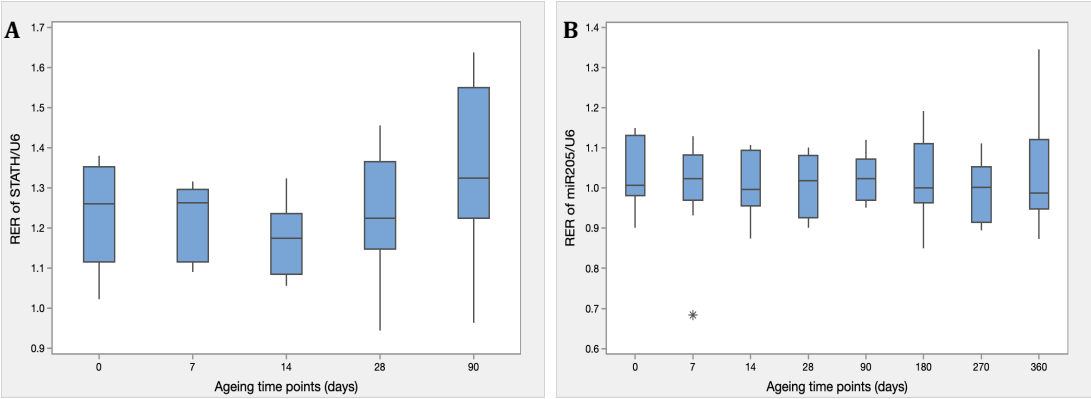


Figure 5.13: **Boxplots showing RER values of dried saliva samples.** RER of (A) *STATH/U6*, (B) *miR205/U6*. The plots were obtained from data shown in Table 5.4 using Minitab Express. *represents outliers. n = 10.

Since a significant correlation between the ratios of saliva-specific markers to *U6* and the ageing time points was not detected, regression analysis was not performed to examine the relationship between the mean RERs of *STATH* and *miR205* to *U6*.

5.3.4.1.3 Inter-donor variation at each ageing time point

In order to apply the RERs of saliva-specific markers in forensic casework for the estimation of the age of saliva stains, the amount of variation in this measure between different individuals should be considered. Therefore, the inter-donor variation for the RER of *STATH/miR205* was statistically evaluated at all ageing time points, as it was the only RER that was significantly correlated with ageing time points.

Table 5.7 shows the standard deviation and the coefficient of variation for the RER of *STATH/miR205*. Since the *STATH* marker was as the numerator (the less stable marker), the highest variation was observed at day 28, which is consistent with the inter-donor variation observed when examining the degradation rate of this marker (section 5.3.3). This variation may be due to variation in the biological composition of each stain from different individuals, as well as technical variation during setting up experimental reactions.

Table 5.7: **Statistical analysis of RERs for saliva-specific RNA markers for the 10 samples at ageing time points 0, 7, 14, 28, 90 days.** The data analysis was carried out using Microsoft Excel.

Age points	*STATH/miR205*	
	SD[a]	CV%[b]
Day 0	0.11	8.81
Day7	0.13	10.81
Day 14	0.10	8.74
Day 28	0.23	18.44
Day 90	0.23	17.90

[a] Standard deviation [b] Coefficient of variation

*Only values up to 90 days were included, as the *STATH* marker fluctuated after 90 days

The relative expression ratio of *STATH*/*miR205*, which produced a high R^2 value (R^2 = 97.2%) as shown above, also had relatively high inter-donor variation across ageing time points. This high level of variation can result in low accuracy in estimating the age of saliva samples, however, to confirm such findings more research is needed, as the examined number of samples in this project was small (n = 10). Further experiments should therefore be carried out, increasing the number of samples as well as applying different environmental conditions (e.g. different temperature, humidity and levels of UV light) to test whether these conditions have an effect on the resulting RER values.

5.3.4.2 Semen-specific markers

The mean C_q values for semen-specific markers measured by RT-qPCR at 0, 7, 14, 28, 90, 180, 270, and 360 days are shown in Table 5.8 for mRNA marker, and Table 5.9 for miRNA markers. These data were corrected with the determined efficiency of each assay using *GenEx* statistical software (version 5.4.4) (see section 3.3.4.1).

Table 5.8: **The mean C_q values of semen-specific mRNA markers at 0, 7, 14, 28, 90, 180, 270, 360 days for the duplicates of 9 samples after efficiency correction**. The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

Age points	Markers	Sample number/Mean C_q values								
		1	2	3	4	5	6	7	8	9
Day 0	*PRM1*	32.01	33.83	31.73	32.90	24.59	27.28	27.80	32.55	32.59
	PRM2	28.05	28.79	27.41	27.89	23.43	23.78	24.97	27.31	28.14
	SEMG1	30.25	30.57	26.27	30.02	22.76	25.70	29.81	31.22	25.34
Day 7	*PRM1*	30.33	35.42	32.73	32.18	28.93	31.76	29.32	31.33	28.35
	PRM2	26.85	32.19	26.10	29.98	28.16	28.16	27.46	28.13	27.96
	SEMG1	27.35	32.86	26.60	30.48	29.16	27.66	28.06	27.18	28.96
Day 14	*PRM1*	29.11	33.50	29.47	38.24	28.32	32.78	34.91	26.57	29.68
	PRM2	28.10	31.27	25.95	34.22	26.19	28.77	31.13	25.25	28.37
	SEMG1	33.52	33.82	32.73	34.32	31.45	33.53	32.57	34.41	30.83
Day 28	*PRM1*	33.45	38.24	29.65	37.46	34.83	35.31	35.69	36.86	38.24
	PRM2	27.68	33.32	24.92	34.54	28.28	28.63	30.71	28.29	32.76
	SEMG1	33.76	35.18	30.83	36.46	30.66	35.77	35.74	32.40	32.42
Day 90	*PRM1*	35.74	36.90	37.01	36.07	35.58	35.47	36.10	35.31	35.58
	PRM2	31.59	34.78	29.15	37.42	29.78	29.20	30.63	30.30	35.60
	SEMG1	37.26	36.60	31.92	36.88	36.20	38.26	38.20	38.04	38.70
Day 180	*PRM1*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	PRM2	39.38	40.00	35.18	40.00	29.84	22.96	28.70	28.50	30.41
	SEMG1	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
Day 270	*PRM1*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	PRM2	32.43	40.00	30.35	40.00	29.27	26.46	31.01	31.17	34.23
	SEMG1	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
Day 360	*PRM1*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	PRM2	29.29	36.47	31.99	37.94	31.22	29.97	34.52	32.48	35.66
	SEMG1	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq

Table 5.9: **The mean C_q values of semen-specific miRNA markers at 0, 7, 14, 28, 90, 180, 270, 360 days for the duplicates of 9 samples after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

Age points	Markers	Sample number/Mean C_q values								
		1	2	3	4	5	6	7	8	9
Day 0	*miR10b*	30.10	32.09	28.42	30.22	28.91	28.25	30.00	30.46	31.63
	miR891a	28.85	28.95	28.50	28.31	30.77	28.55	29.76	29.89	30.74
	U6	30.85	29.37	27.64	28.98	29.04	30.01	30.80	34.42	34.09
Day 7	*miR10b*	31.06	32.75	29.83	30.59	29.43	29.93	30.41	31.49	32.95
	miR891a	26.46	29.20	27.52	28.90	29.74	29.46	29.78	29.85	30.06
	U6	30.23	31.28	29.32	34.57	28.68	30.91	33.87	29.36	32.56
Day 14	*miR10b*	32.93	32.66	30.65	35.11	29.46	29.82	33.75	30.64	32.31
	miR891a	28.99	29.71	29.36	30.24	29.63	29.38	30.87	29.14	30.95
	U6	32.07	29.85	29.85	34.07	28.44	28.98	34.24	27.31	33.56
Day 28	*miR10b*	34.10	35.25	30.43	36.14	29.02	29.63	32.47	29.02	32.84
	miR891a	28.97	30.27	28.50	30.94	26.75	27.04	28.18	26.30	28.24
	U6	33.40	35.98	29.63	35.50	26.59	26.86	32.32	26.84	34.81
Day 90	*miR10b*	32.08	34.34	30.70	38.21	29.57	28.23	31.51	30.01	34.34
	miR891a	28.76	29.42	28.38	31.18	26.58	25.73	26.67	27.18	28.48
	U6	34.51	33.75	30.87	37.64	27.95	23.90	29.89	26.62	35.44
Day 180	*miR10b*	31.58	35.70	30.53	34.21	29.21	29.61	32.18	31.87	34.96
	miR891a	26.99	30.06	26.05	28.42	29.55	29.10	32.03	29.20	31.60
	U6	33.14	34.65	30.79	33.80	27.57	26.83	29.58	27.42	33.72
Day 270	*miR10b*	31.87	35.30	27.61	34.64	29.66	30.99	32.15	30.97	36.11
	miR891a	28.09	29.35	26.17	28.70	28.03	27.55	30.63	29.55	29.62
	U6	33.10	35.18	31.32	29.67	28.40	26.26	30.62	27.17	34.23
Day 360	*miR10b*	29.87	36.77	30.70	34.16	30.83	29.93	32.67	33.29	34.93
	miR891a	26.07	30.74	27.24	30.96	26.87	26.39	31.45	29.00	31.04
	U6	29.89	35.54	33.16	30.73	28.51	26.40	31.34	29.23	34.39

5.3.4.2.1 Identification of the most stable RNA markers in semen

The relative quantification of pairs of markers was calculated using the C_q value of a focal degraded marker and the C_q value of the most stable marker. Again, *geNorm* software was used to assess the expression of RNA transcripts, and determine which RNA markers were most stable and exhibited the lowest variability across all time points. Figure 5.14 shows the data from *geNorm*, with the M-values for each marker. The marker that shows the highest variation relative to all other markers is eliminated by *geNorm*. *PRM1* and *SEMG1* were not included tin the analysis as it disappeared after 90 days of storage. The software indicates that *miR891a* and *miR10b* are the most stable markers across all time points, with the lowest M-Value (both 2.03), followed by *U6* with an M-Value of 2.25.

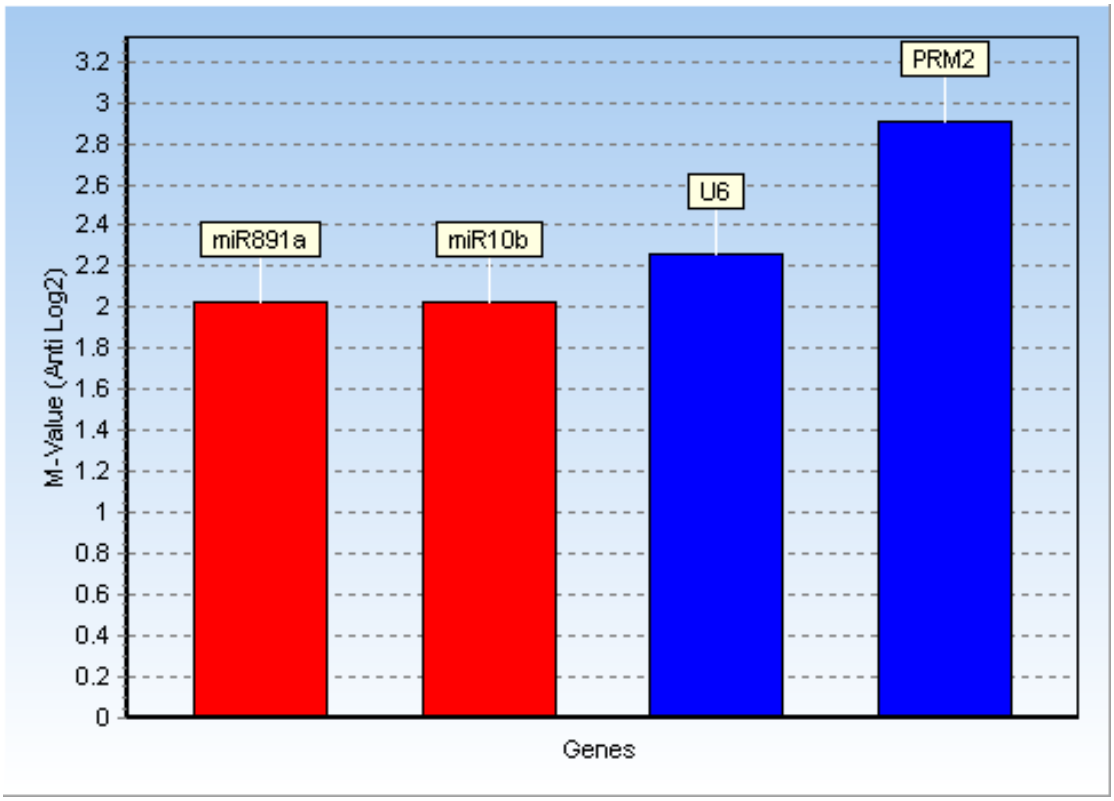


Figure 5.14: **The most stable RNA markers in semen samples across all ageing time points measured by geNorm.** *miR891a* and *miR10b* exhibited the lowest M-Value indicating that they were the most stable markers in semen stains.

5.3.4.2.2 Relative expression ratio (RER)

The ratio was obtained by dividing the efficiency-corrected C_q values of the less stable RNA marker by the efficiency-corrected C_q values of the more stable RNA marker:

$$RER = \frac{Cq\ of\ less\ stable\ marker}{Cq\ of\ more\ stable\ marker}$$

5.3.4.2.2.1 RERs of mRNA to miRNA

Given that the microRNA markers were identified as the most stable semen-specific markers across all ageing time points, Table 5.10 shows the mean RERs of the different mRNA markers to *miR891a* and *miR10b,* calculated from the corrected C_q values for semen samples stored up to one year at room temperature.

Table 5.10: **Mean RER of the mRNA markers (*PRM1*, *PRM2* and *SEMG1*) to the miRNA markers (*miR10b* and *miR891a*) for 9 semen samples at ageing time points 0, 7, 14, 28, 90, 180, 270, 360 days.** The RERs were calculated from the mean C_q values in Table 5.8 and Table 5.9.

	RER					
Age points	*PRM1/miR10b*	*PRM1/miR891a*	*PRM2/miR10b*	*PRM2/miR891a*	*SEMG1/miR10b*	*SEMG1/miR891a*
Day 0	1.02	1.04	0.89	0.91	0.93	0.96
Day 7	1.01	1.08	0.92	0.98	0.89	0.99
Day 14	0.98	1.05	0.90	0.97	1.04	1.11
Day 28	1.11	1.26	0.93	1.05	1.05	1.19
Day 90	1.13	1.29	1.00	1.14	1.16	1.32
Day 180	-	-	1.02	1.13	-	-
Day 270	-	-	1.02	1.14	-	-
Day 360	-	-	1.02	1.15	-	-

*RERs were calculated only up to 90 days as *PRM1* and *SEMG1* markers were not detected after this time point.

Figure 5.15 represents the RER values for mRNA markers to miRNA markers. It was found that the RERs of *PRM2* to both miRNAs increase with increasing ageing time points. The relationship between the obtained values of RERs and the ageing time points was statistically evaluated. Since the Anderson-Darling normality test showed that the data were normally distributed ($p > 0.05$), therefore parametric analysis was used.

The RERs of *PRM2* to both *miR10b* and *miR891a* showed a positive correlation between ageing time points using Pearson's correlation ($r = 0.864$, $p = 0.0057$; and $r = 0.795$, $p = 0.0184$, respectively). In contrast, although the RERs of *PRM1* to miRNA markers showed a slight increase with increasing ageing time points before degrading completely after 90 days, applying a correlation analysis showed no statistically significant relationship.

Furthermore, a similar trend to *PRM1* was shown when analysing the RERs of *SEMG1* to both *miR10b* and *miR891a*, however it showed a marginally significant positive correlation ($r = 0.879$, $p = 0.049$; and $r = 0.920$, $p = 0.026$ respectively) before *SEMG1* fell below the detection level of the assay at 90 days.

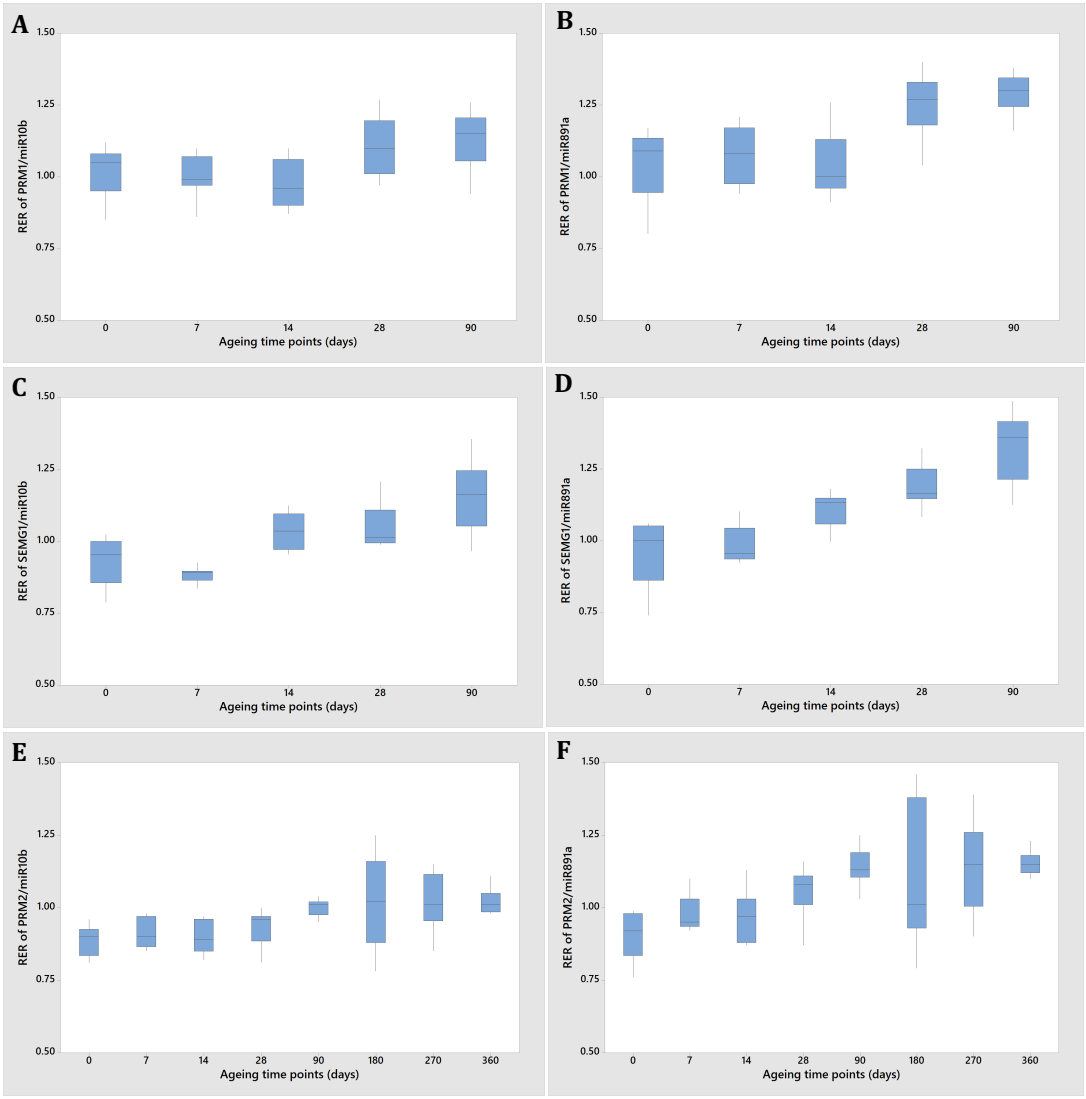


Figure 5.15: **Boxplots showing RER values of dried semen samples.** RER of (A) *PRM1/miR10b*, (B) *PRM1/miR891a*, (C) *SEMG1/miR10b*, (D) *SEMG1/miR891a*, (E) *PRM2/miR10b* and (F) *PRM2/miR891a*. The plots were obtained from data shown in Table 5.8 and Table 5.9 using Minitab®17. n = 9.

To investigate the relationship between the mean RER values of *PRM2* and *SEMG1* to *miR10b* and *miR891a* and the ageing time points further, a regression analysis was performed. The RERs of *PRM1* to miRNA markers were not included as they showed no statistically significant relationship when applying a correlation analysis. The data were fitted using two different models, a linear model (Figure 5.16) and a non-linear model (a second-order polynomial model, Figure 5.17). The models were compared to determine

which is more accurate at predicting the age of semen stains using RNA semen-specific markers.

For the *PRM2* marker, data was only included in the analysis up to 90 days, as the RERs with both *miR10b* and *miR891a* reached a plateau at that point. When fitting the data using a linear model, the analysis produced a relatively low R^2 value for both *PRM2/miR10b* (R^2 = 94.1%) and *PRM2/miR891a* (R^2 = 89.0%). The regression analysis was also applied to the RERs of *SEMG1* to the miRNA markers, as these were also shown to exhibit a positive correlation with ageing time point. The RERs of *SEMG1*/*miR10b* and *SEMG1*/*miR891a* also gave relatively low R^2 values of 77.3% and 84.8% respectively.

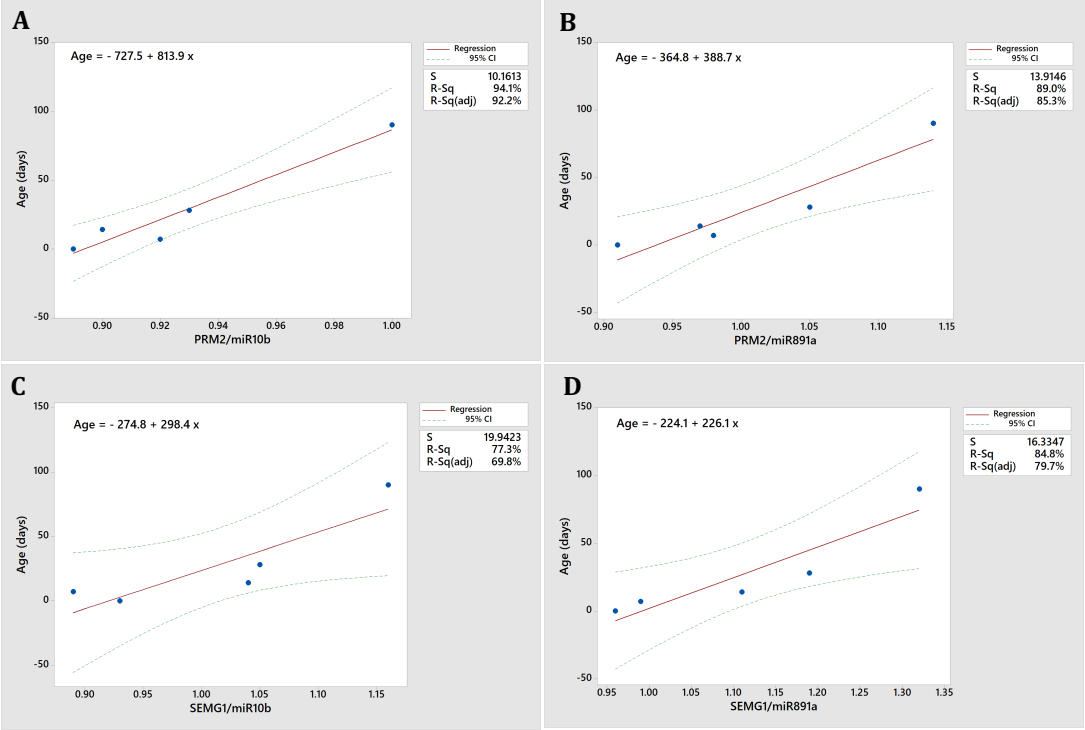


Figure 5.16: **Regression analysis of the RERs of semen samples over ageing time points, using a linear model with a 95% confidence intervals.** RER of (A) *PRM2/miR10b* (B) *PRM2/miR891a*, (C) *SEMG1/miR10b* and (D) *SEMG1/miR891a*. n = 9.

In contrast, the second-order polynomial model shows a higher R^2 for all RERs of the mRNA markers to the miRNA markers; R^2 = 96.6%, 98.4%, 99.0% and 98.9% for the RERs of *PRM2/miR10b*, *PRM2/miR891a*, *SEMG1/miR10b* and *SEMG1/miR891a* respectively, with low standard error S ≤ 9.5. The differences

in the 95% confidence interval were less pronounced in the second-order polynomial models (95% confidence intervals are narrower in the linear model), which in combination with the higher R^2 values make these models more reliable in predicting the age of semen samples.

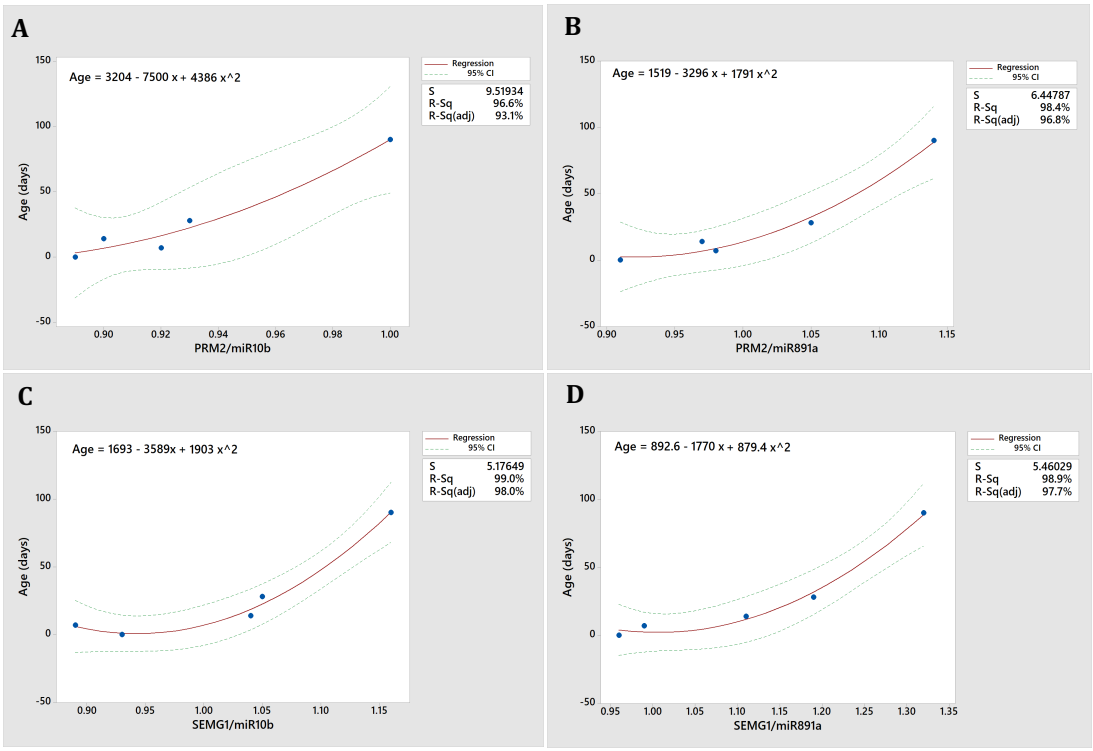


Figure 5.17: ***Regression analysis of the RERs of semen samples over ageing time points, using a second-order polynomial model with 95% confidence intervals.*** *RER of (A)* PRM2/miR10b *(B)* PRM2/miR891a*, (C)* SEMG1/miR10b *and (D)* SEMG1/miR891a*. n = 9.*

5.3.4.2.2.2 RERs of mRNA and miRNA markers to U6

The *geNorm* analysis described in section (5.3.4.2.1) above showed that the snRNA molecule *U6* was allocated a low M-value, placing it right after the most stable markers (*miR891a* and *miR10b)* in terms of stability. Therefore, the RERs of all the semen-specific markers to *U6* were calculated to examine the relationship with the age of semen stains (Table 5.11), using the equation below:

$$RER = \frac{Cq\ of\ semen - specific\ \ marker}{Cq\ of\ reference\ gene\ U6}$$

Table 5.11: **Mean RER of *PRM1*, *PRM2*, *SEMG1*, *miR10b* and *miR891a* to *U6* for 9 samples at ageing time points 0, 7, 14, 28, 90, 180, 270, 360 days.** The RERs were calculated from the mean C_q values in Table 5.8 and Table 5.9.

	RER				
Age points	*PRM1/U6*	*PRM2/U6*	*SEMG1/U6*	*miR10b/U6*	*miR891a/U6*
Day 0	1.00	0.87	0.92	0.98	0.96
Day 7	1.00	0.91	0.90	0.99	0.93
Day 14	1.00	0.92	1.06	1.02	0.95
Day 28	1.15	0.96	1.09	1.03	0.91
Day 90	1.18	1.04	1.21	1.04	0.91
Day 180	-	1.06	-	1.05	0.96
Day 270	-	1.07	-	1.05	0.94
Day 360	-	1.08	-	1.05	0.93

*RERs were calculated only up to 90 days as *PRM1* and *SEMG1* markers were not detected after this time point.

Figure 5.18 presents the RER values for all semen-specific markers to the reference gene (*U6*). It was found that the RERs of *PRM2*/*U6, SEMG1/U6* and *miR10b*/*U6* increase with increasing ageing time points, whereas the RERs of *PRM1/U6* and *miR891a/U6* do not. The increase in the *PRM2*/*U6* ratio with time was steady up to 28 days and then showed a substantial increase at 90 days, while the RER of *miR10b*/*U6* increased up to 180 days and then remained at the same level after this. The RER of *SEMG1*/*U6* did not show an increase until 14 days, and then increased steadily after that point. When comparing the level of variation in the RER values, the RERs of the miRNA markers to *U6* had higher levels of variation than the RERs of the mRNA markers to *U6*.

Since the Anderson-Darling normality test showed that the data were normally distributed ($p > 0.05$), therefore parametric analysis was used. A Pearson's correlation analysis indicated that there was a significant positive correlation between ageing time points and the RER of *PRM2*/*U6* ($r = 0.88$, $p = 0.004$),

SEMG1/U6 (r = 0.890, p = 0.043) and *miR10b*/*U6* (r = 0.76, p = 0.028), but not with the RER of *PRM1/U6* (r = 0.084, p = 0.072) or *miR891a/U6* (r = 0.005, p = 0.99).

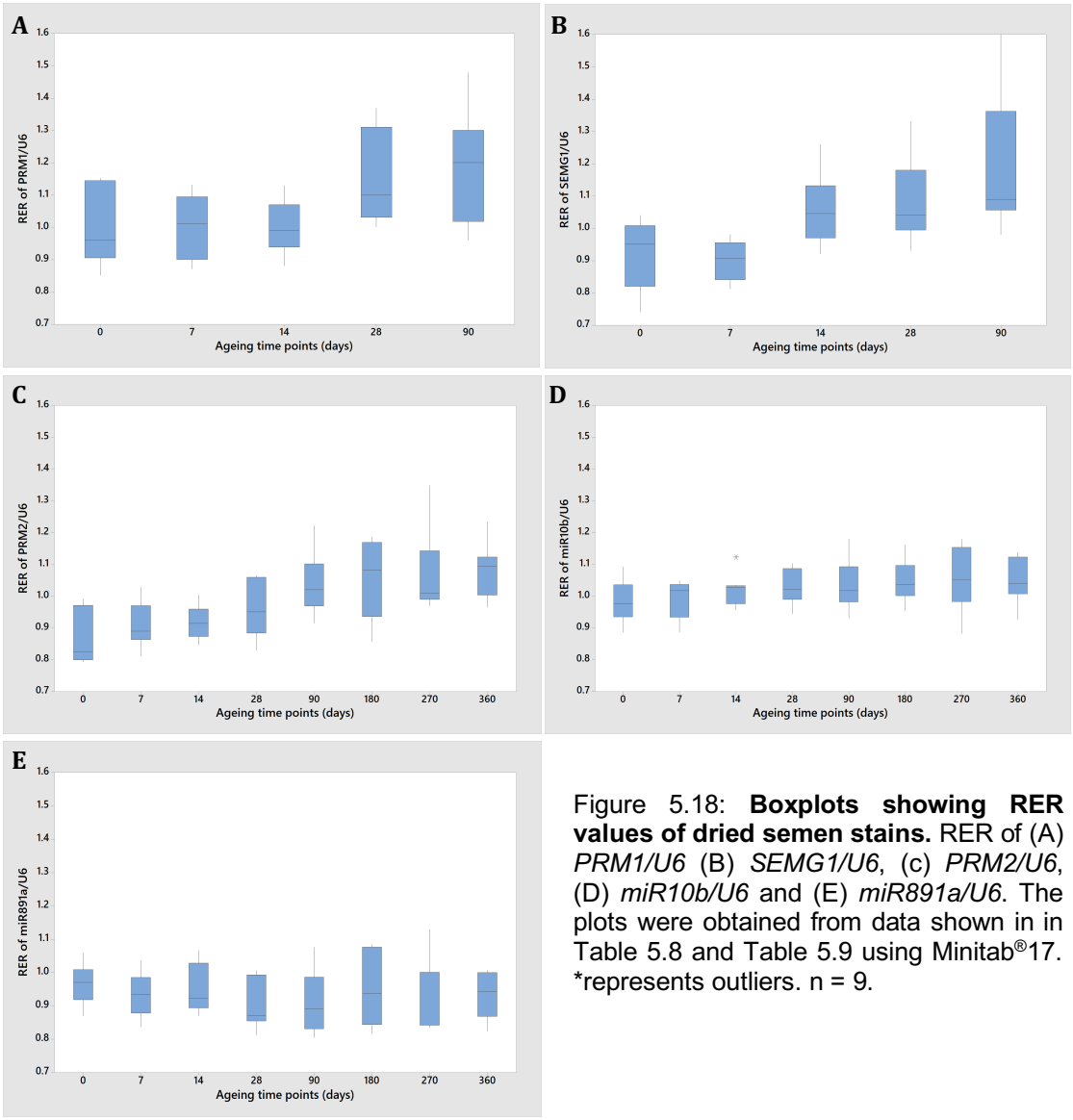


Figure 5.18: **Boxplots showing RER values of dried semen stains.** RER of (A) *PRM1/U6* (B) *SEMG1/U6*, (c) *PRM2/U6*, (D) *miR10b/U6* and (E) *miR891a/U6*. The plots were obtained from data shown in in Table 5.8 and Table 5.9 using Minitab®17. *represents outliers. n = 9.

Since the mean RERs of two of the mRNA markers (*PRM2* and *SEMG1*) and one miRNA marker (*miR10b)* to *U6* have shown a statistically significant relationship with the ageing time points, a regression analysis was performed in order to obtain an age prediction model. Each RER was fitted using both a linear model and a second-order polynomial model, as above, to identify the most reliable prediction model. The linear and second order polynomial models

for the RERs of *PRM2*, *SEMG1* and *miR10b* to *U6* are illustrated in Figure 5.19 and Figure 5.20.

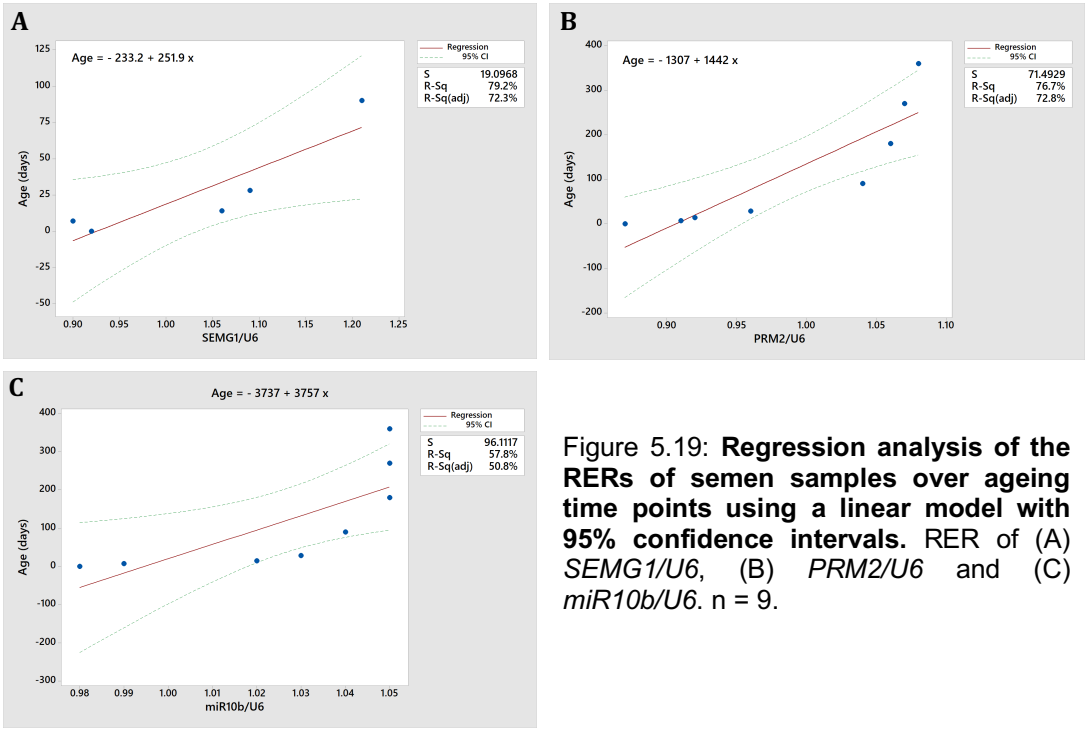


Figure 5.19: **Regression analysis of the RERs of semen samples over ageing time points using a linear model with 95% confidence intervals.** RER of (A) *SEMG1/U6*, (B) *PRM2/U6* and (C) *miR10b/U6*. n = 9.

Of particular note is that the regression analysis of the RER of *SEMG1*/*U6* using a second-order polynomial model gave a high R^2 value (R^2 = 99.6%), which is higher than for the linear model (R^2 = 79.2%), with very low standard error S = 3.0. The same applies to regression analysis of the RERs of *PRM2/U6* and *miR10b/U6*, which gave similar results, showing higher R^2 values for the second-order polynomial model than the linear model; R^2 = 76.7% for *PRM2/U6* and R^2 = 57.8% for *miR10b/U6* in the linear model and R^2 = 90.3% for *PRM2/U6* and R^2 = 83.3% for *miR10b/U6* in the linear model.

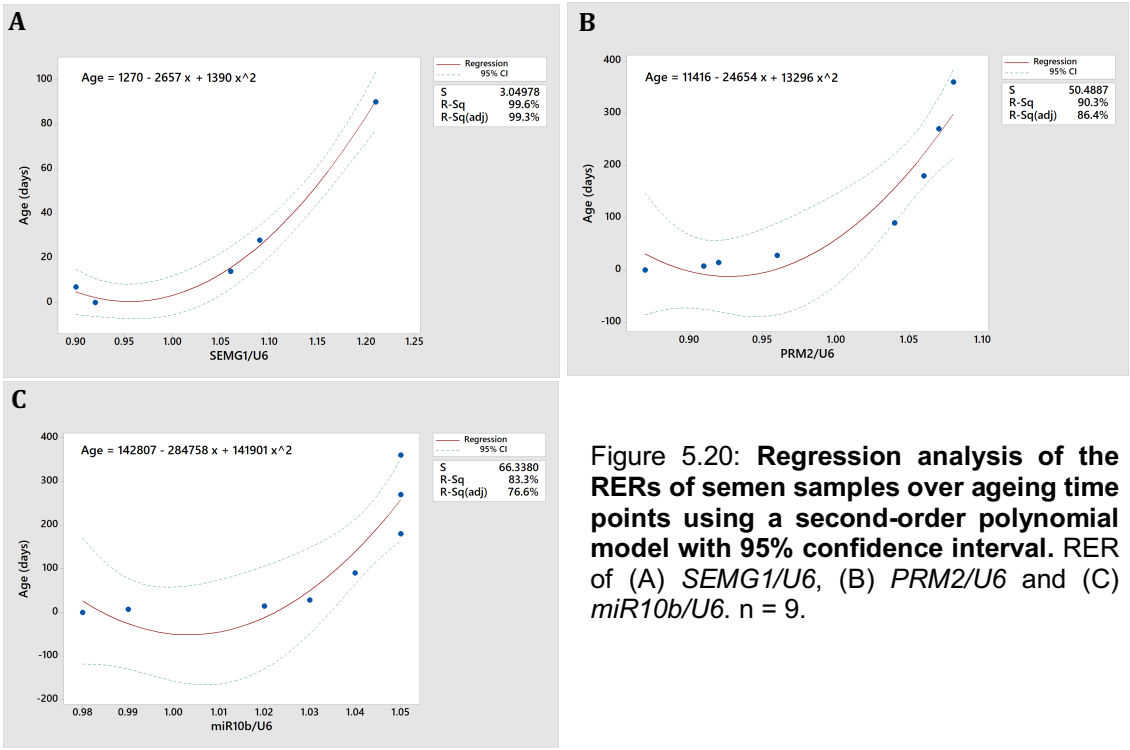


Figure 5.20: **Regression analysis of the RERs of semen samples over ageing time points using a second-order polynomial model with 95% confidence interval.** RER of (A) *SEMG1/U6*, (B) *PRM2/U6* and (C) *miR10b/U6*. n = 9.

5.3.4.2.2.3 RERs of mRNA markers

Finally, the RERs of all three semen-specific mRNA markers were also calculated. None of these ratios was found to have specific behaviour in relation to ageing time points. (Table 5.12).

Table 5.12: **Mean RER of *PRM1/PRM2*, *PRM1/SEMG1* and *PRM2/SEMG1* for 9 samples at ageing time points 0, 7, 14, 28, 90 days.** The RERs were calculated from the mean C_q values in Table 5.8 and Table 5.9.

Age points	*PRM1/PRM2*	*SEMG1/PRM1*	*SEMG1/PRM2*
Day 0	1.15	0.92	1.05
Day 7	1.12	0.92	1.03
Day 14	1.09	1.06	1.16
Day 28	1.19	0.95	1.13
Day 90	1.13	1.03	1.16

The RER values *PRM1*/*PRM2*, *SEMG1*/*PRM1* and *SEMG1*/*PRM2* are presented in Figure 5.21. When performing Pearson's correlation analysis, as the data were normally distributed $p > 0.5$, no significant correlations between ageing time points and the RERs of these mRNA markers were identified (all $p > 0.05$), therefore no further regression analysis was performed on these data.

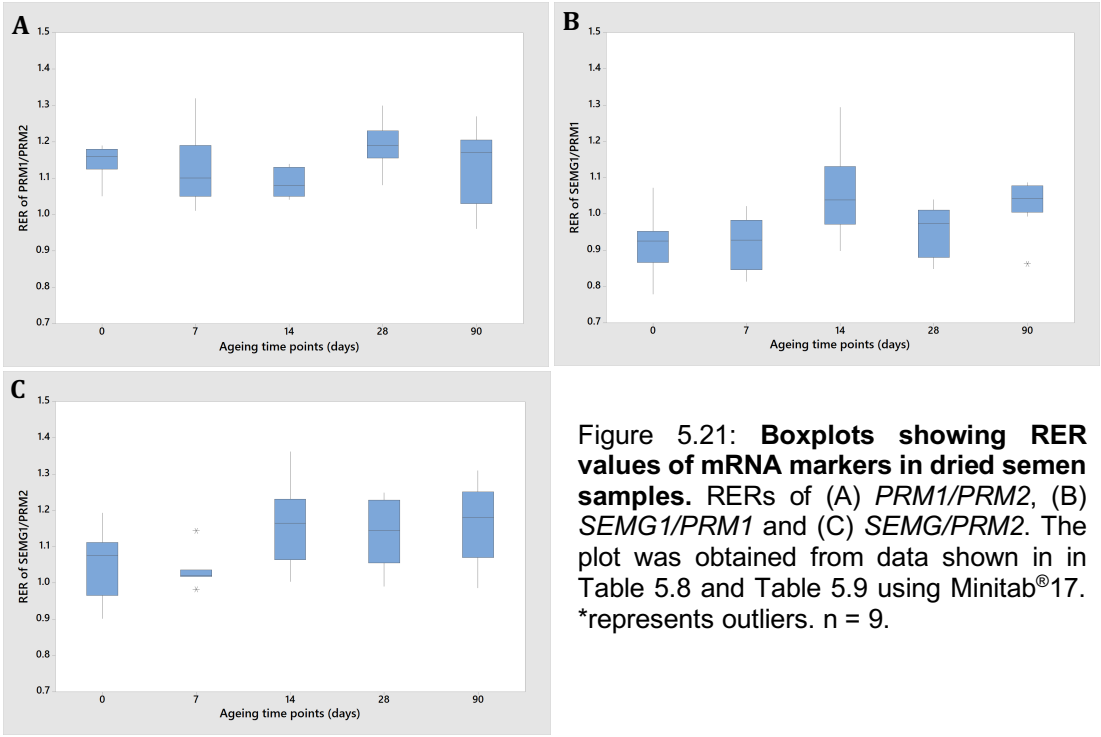


Figure 5.21: **Boxplots showing RER values of mRNA markers in dried semen samples.** RERs of (A) *PRM1/PRM2*, (B) *SEMG1/PRM1* and (C) *SEMG/PRM2*. The plot was obtained from data shown in in Table 5.8 and Table 5.9 using Minitab®17. *represents outliers. n = 9.

5.3.4.2.3 Inter-donor variation at each ageing time point

In order to use the RERs of semen-specific markers in forensic casework for the estimation of the age of semen stains, the amount of variation in this measure between different individuals should be considered. Therefore, the inter-donor variation for each RER that was discussed above was statistically evaluated at all ageing time points.

Table 5.13 shows the standard deviation and the coefficient of variation for each of the RERs. The RERs that include *PRM2* as the numerator showed high variation at day 180, which is consistent with the inter-donor variation observed when examining the degradation rate of this marker (section 5.3.3). This variation may be due to the variation in the biological composition of each stain from different individuals, as well as technical variation during setting up experimental reactions (see section 5.3.3).

The relative expression ratio of mRNA to miRNAs (i.e. semen-specific markers), have produced a high R^2 value as shown above with the second-polynomial model, however they also have shown relatively high inter-donor variation across ageing time points. Similar findings were observed in the RERs of semen-specific markers to *U6*, as they have produced a high R^2 value with the second-polynomial model and have shown relatively high inter-donor variation across ageing time points, with highest CV% values were observed in *SEMG1/U6* ratio. This high level of variation can result in low accuracy in estimating the age of semen samples, however, to confirm such findings more research is needed, as the examined number of samples in this project was small (n = 9). Further experiments should therefore be carried out, increasing the number of samples as well as applying different environmental conditions (e.g. different temperature, humidity and levels of UV light) to test whether these conditions have an effect on the resulting RER values.

Table 5.13: **Statistical analysis of RERs for semen-specific RNA markers for the 9 samples at ageing time points 0, 7, 14, 28, 90, 180, 270, 360 days.** The data analysis was carried out using Microsoft Excel.

Age points	SEMG1/miR10b		SEMG1/miR891a		PRM2/miR10b		PRM2/miR891a		SEMG1/U6		PRM2/U6		miR10b/U6	
	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]	SD[a]	CV%[b]
Day 0	0.09	9.33	0.12	12.15	0.05	5.69	0.08	8.92	0.11	11.48	0.09	9.77	0.06	6.46
Day7	0.03	3.53	0.07	6.86	0.05	5.52	0.06	6.20	0.06	7.00	0.07	7.45	0.06	6.24
Day 14	0.06	6.19	0.06	5.34	0.06	6.20	0.09	9.17	0.11	10.17	0.05	5.80	0.05	4.86
Day 28	0.07	7.11	0.07	6.12	0.07	7.32	0.09	8.29	0.13	11.58	0.09	9.31	0.05	5.17
Day 90	0.12	10.54	0.12	9.02	0.03	2.93	0.06	5.64	0.21	17.15	0.09	9.12	0.08	7.49
Day 180	-	-	-	-	0.16	16.16	0.25	22.35	-	-	0.12	11.72	0.06	6.01
Day 270	-	-	-	-	0.10	9.44	0.15	13.10	-	-	0.12	11.45	0.10	9.30
Day 360	-	-	-	-	0.04	4.21	0.04	3.40	-	-	0.08	7.82	0.07	6.63

[a] Standard deviation [b] Coefficient of variation

5.4 Summary and Conclusion

The main focus of this project is the application of RNA body fluid-specific markers in determining the age of biological stains, allowing forensic experts to identify the type of body fluid and the time since deposition simultaneously. In the context of forensic science laboratories these techniques require much more research before consideration could be given to applying them to forensic casework.

To select RNA markers to be used for estimating the age of saliva and semen samples, the degradation rate of these markers needs to be large enough that it is observable across ageing time periods. In this work, the degradation rate of saliva- and semen-specific RNA markers in aged samples was analysed. A total of 19 volunteers were asked to give saliva and semen samples, which were deposited on cotton swabs and then stored in a dark, dry place at room temperature to simulate natural ageing until they reached a series of desired ages (0, 7, 14, 28, 90, 180, 270, and 360 days). The degradation levels of two saliva-specific and five semen-specific RNA markers were analysed using TaqMan® assays, and the RERs were calculated to study the degradation behaviour of the markers and their relationship with age.

The outcomes of this research showed that different RNA molecules degrade at different rates in saliva and semen samples, with miRNA markers exhibiting strong stability, likely due to their small size. By applying correlation tests and regression analysis, the data indicate that the RERs of saliva and semen-specific markers have a significant relationship with ageing time points. In saliva-specific markers, regression analysis of the relationship between the RER of *STATH*/*miR205* with ageing time period produced a high R^2 value (97.2%) using a non-linear model (i.e. second order polynomial). These results indicate that this ratio could be reliable in the estimation of the age of saliva samples. With regards to semen-specific markers, the RERs of semen-specific markers (*PRM2/miR10b, PRM2/miR891a, SEMG1/miR10b,*

SEMG1/miR891a, PRM2/U6, SEMG1/U6 and miR10b/U6) also showed a positive correlation with ageing time points, with regression analyses giving R^2 values of more than 96% using non-linear models, with low standard error value (i.e S value). These results confirm the reliability of using RERs in estimating the age of body fluid stains.

In both body fluids, the difference between the two types of model examined (linear and non-linear) is very clear, with the non-linear models having substantially higher R^2 values. The higher prediction abilities of these models therefore make them a better choice for the estimation of the age of saliva and semen stains. However, to determine how generally applicable this is across body fluid types, samples collected from crime scene should be studied, including mixture samples such as semen mixed with menstrual blood or vaginal secretions. This study only examined samples that were stored under controlled conditions, therefore, additional environmental factors that might affect the RERs of body fluid-specific markers should be explored as future project, such as UV exposure, humidity and high temperature.

The RERs of saliva and semen-specific markers represent a potential method to estimate the time since deposition of saliva and semen stains. The findings of this study therefore emphasise that, in future, methods using RT-qPCR are likely to be a sensitive technology for the accurate determination of the age of saliva and semen samples.

Chapter six: Evaluating the effect of body fluid mixture on the relative expression ratio (RER) of blood-specific RNA markers

6.1 Introduction

Previous studies have indicated that the RERs of reference genes (*ACTB/18S*) can be used to indicate the age of blood [58, 59], saliva [89] or hair samples [60]. However, it has also been suggested that there are limitations to this method when applied to samples that are mixtures of more than one type [89]. In previous Chapters, the relative expression ratios (RERs) of body fluid-specific RNA markers have also been shown to be a potential method for estimating the age of body fluid stains, or the time since deposition [192, 193]. However, the nature of some forensic samples found at crime scenes could make this challenging, as they frequently occur in a mixture of multiple different body fluid types. For instance, in a physical assault, there could be a mixture of blood and saliva samples, or in sexual assaults there may be a mixture of semen and blood or saliva. Therefore, in order to develop such a method to be considered as a successful approach to estimating the age of biological stains in forensic casework samples, it is important that the impact of body fluid mixtures on RER is evaluated.

In the context of forensic applications, a variety of RNA types have been identified as blood-specific marker, including the mRNA molecules Haemoglobin Subunit Alpha (*HBA*), and Haemoglobin Subunit Beta (*HBB*), which are protein subunits of the haemoglobin molecule, and the microRNA *miR16,* which is involved in gene expression regulation [66, 68, 77, 170]. These RNA molecules have been shown to degrade at different rates in bloodstains, with *miR16* marker exhibiting strong stability [192], likely due to its small size (~22 nucleotides) [23]. Interestingly, the RERs of these blood-

specific markers have bene shown to be positively correlated with the age of bloodstains [192], indicating that they may be reliable in estimating the time since deposition of bloodstains (see Chapter four).

The effect of mixing blood samples with two different body fluids (saliva and semen) and the impact on the RER of blood-specific markers is examined in this Chapter. This will determine whether the RERs are under- or overestimating the age of bloodstains when they are mixed with other body fluid types, and whether it is valid to estimate blood-specific RERs from mixed samples, or if the mixture makes this data unreliable.

6.2 Aim and Objective

The aim of the work in this Chapter was to assess the effect of body fluid mixtures on the RER of blood-specific markers. The expression level of *HBA*, *HBB* and *miR16* along with two reference genes (*18S* and *U6*) was measured in fresh pure body fluid samples and mixed samples, consisting of blood and saliva or blood and semen. Pure blood samples and mixed samples were also aged for up to two months and their expression level was measured. All RNA transcripts expression level was measured using the RT-qPCR approach. The RERs of blood-specific markers were calculated and compared between pure blood samples and mixed samples to determine whether there were any differences, and whether the RERs of the blood-specific markers are under- or overestimated in the mixed samples.

6.3 Materials and Methods

6.3.1 Sample collection

A total of 12 volunteers (6 males and 6 females) donated blood, saliva and semen samples to set up two sample groups, as described in the sections below. The samples and mixture compositions are shown in Table 6.1. The experimental procedures used in this project were approved by the

Departmental Ethics Committee in the Department of Pure and Applied Chemistry at the University of Strathclyde (see Appendix A1).

6.3.1.1 Fresh and pure samples

The first group of samples consisted of fresh body fluid stains (day = 0, i.e. not aged) from single body fluid (i.e. pure stains). Four volunteers donated blood samples in duplicate (total 8 samples). The blood samples were collected onto sterile cotton swabs using disposable Unistik 3 comfort lancets as described in section 2.2.1. Another eight volunteers donated saliva and semen samples in duplicate (four donors for each body fluid). The samples were deposited into sterile collection pots, as described in sections 2.2.2 and 2.2.3 respectively. 20 µL of each body fluid sample was pipetted onto sterile cotton swabs and allowed to dry at room temperature before starting RNA extraction.

6.3.1.2 Pure aged samples

From the same volunteers pure blood, saliva and semen samples (i.e. pure stains) were prepared in duplicate to be stored in a dark dry place at room temperature to simulate natural ageing until they reached a series of desired ageing time points (10, 30 and 60 days), at which stage total RNA was extracted.

6.3.1.3 Mixture samples

The second group of samples consisted of bloodstains mixed with either saliva samples or with semen samples, and each mixture sample consisted of stains from two donors as described below:

A. A mixture of fresh blood and fresh saliva from two volunteers was prepared by adding 20 µL of fresh blood from one volunteer to 20 µL of fresh saliva from another volunteer on cotton swabs. The samples were prepared in duplicate (total 8 samples at each time point).

B. A mixture of fresh blood and fresh semen from two volunteers was prepared by adding 20 µL of fresh blood from one volunteer to 20 µL of fresh semen from another volunteer on cotton swabs. The samples were prepared in duplicate (total 8 samples at each time point).

All mixture samples were stored at room temperature in a dark dry place to simulate natural ageing until they reach a series of desired ageing time points (0, 10, 30 and 60 days), at which stage total RNA was extracted.

Table 6.1: **Sample types and composition.** n = 8 for each samples composition.

Sample age (days)	Sample type	Sample composition
0	Pure blood	20µL of blood
	Pure saliva	20µL of saliva
	Pure semen	20µL of semen
	Blood + saliva	20µL of blood and 20µL of saliva
	Blood + semen	20µL of blood and 20µL of semen
10	Pure blood	20µL of blood
	Pure saliva	20µL of saliva
	Pure semen	20µL of semen
	Blood + saliva	20µL of blood and 20µL of saliva
	Blood + semen	20µL of blood and 20µL of semen
30	Pure blood	20µL of blood
	Pure saliva	20µL of saliva
	Pure semen	20µL of semen
	Blood + saliva	20µL of blood and 20µL of saliva
	Blood + semen	20µL of blood and 20µL of semen
60	Pure blood	20µL of blood
	Pure saliva	20µL of saliva
	Pure semen	20µL of semen
	Blood + saliva	20µL of blood and 20µL of saliva
	Blood + semen	20µL of blood and 20µL of semen

6.3.2 Total RNA extraction and reverse transcription

RNA extraction was carried out using the TRI Reagent® procedure as described in section 2.4.1. TURBO DNA-free™ kit was used to treat the extracted RNA to remove any genomic DNA as described in section 2.5. The reverse transcription was carried out using High-Capacity cDNA Reverse Transcription kit and TaqMan® MicroRNA Reverse Transcription kit as described in sections 2.7.1 and 2.7.2.

6.3.3 Quantitative real-time PCR (qPCR)

All TaqMan® assays were run in singleplex assays, following the procedure described in section 2.8, using a Stratagene Mx3005P.

6.3.3.1 TaqMan Assays®

The TaqMan® Gene Expression Assays used in this Chapter were off-the-shelf, and are predesigned for each transcript. Blood-specific markers and reference genes used in this experiment were from Applied Biosystems (Life Technologies), each with a unique ID. Table 6.2 and Table 6.3 shows the characteristics of the selected RNA markers.

Table 6.2: **Characteristics of the TaqMan® Gene Expression Assays for blood-specific markers.**

Gene	Applied Biosystems TaqMan® assay ID	Amplicon length (nt)	Assay location*
HBA	Hs00361191_g1	156	158
HBB	Hs00758889_s1	95	511
18S	Hs99999901_s1	187	604

* Refers to the nucleotide location that is the midpoint of the target region.

Table 6.3: **Characteristics of the TaqMan® MicroRNA Assays for blood-specific markers and *U6*.**

miRNA target	Applied Biosystems TaqMan® assay ID	Target sequence
miR16	000391	UAGCAGCACGUAAAUAUUGGCG
U6	001973	GTGCTCGCTTCGGCAGCACATATACTAA AATTGGAACGATACAGAGAAGATTAGCA TGGCCCCTGCGCAAGGATGACACGCAA ATTCGTGAAGCGTTCCATATTTT

6.3.4 Data Analysis

The data generated from RT-qPCR was analysed using *MxPro* and *GenEx* software (version 5.4.4) was used for efficiency correction of the raw data (see section 2.9). *Microsoft Excel 2016* was used to manipulate raw C_q values from the RT-qPCR reactions and to present basic data and line graphs. *Minitab Express* (version 1.5.0) was used for statistical analyses, including the Anderson-Darling normality test and one-way analysis of variance (ANOVA) and Dunnett's Multiple Comparisons with a control method.

6.4 Results and Discussion

In this Chapter, the expression level of individual RNA transcripts was quantified by RT-qPCR in pure and mixed body fluid samples stored at room temperature in a dark dry place for up to 60 days. All measured C_q values were corrected against the obtained efficiency of each assay.

6.4.1 The expression of RNA transcripts in pure body fluid samples

The expression levels of the selected blood-specific markers (*HBA*, *HBB* and *miR16*) and two reference genes (*18S* and *U6*) were initially measured in fresh

pure body fluid samples (blood, saliva and semen). The C_q values of each marker in the three body fluid types are recorded in Table 6.4.

As expected, all blood-specific markers exhibited high expression levels in fresh pure bloodstains. In contrast, neither of the blood-specific mRNA markers (*HBA* and *HBB*) exhibited any expression in fresh pure saliva and semen samples. Only *miR16* showed expression in fresh pure saliva and semen samples. However, when comparing the expression level of *miR16* in blood samples to its expression in saliva and semen samples, the expression level of *miR16* in blood was much higher, with average C_q values lower in blood than in saliva and semen by 6.39 and 8.47 cycles, respectively. These findings confirm that the selected markers are truly blood-specific markers, showing low or no expression in other examined body fluids (i.e. saliva and semen). Based on these findings, the expression of the blood-specific markers was not further analysed in aged pure saliva and semen samples, as they showed high or no C_q values in the fresh samples. These findings are in concordance with the EDNAP collaborative exercise [65] and a number of other studies [64, 194-196] that have investigated the expression of blood-specific markers in different body fluids, where no expression of *HBA* and *HBB* has been found in saliva and semen samples.

With regards to the reference genes, *18S* also exhibited higher expression in fresh pure bloodstains (C_q = 14.76) compared to saliva (C_q = 21.70) and semen samples (C_q = 25.88). *U6* on the other hand, showed a similar expression level in blood and saliva samples (C_q = 23.25 and 22.75 respectively), and a lower expression level in semen samples (C_q = 30.12). Both *18S* and *U6* are used as reference genes for mRNA and miRNA studies respectively [66, 191, 197], which means that they should be expressed among all body fluids and tissues at a constant level and should not be affected by experimental conditions. However, the findings of this work suggest otherwise, as the expression of these genes varies between different types of body fluids, with the exception of *U6* in blood and saliva where expression level was relatively similar.

Table 6.4: **The C_q values of blood-specific markers and two reference genes in fresh pure body fluid samples (blood, saliva and semen).** n = 8 for each body fluid type.

		RNA MARKERS				
	Sample	***HBA***	***HBB***	***miR16***	***18S***	***U6***
Blood	B1	18.87	19.31	14.46	17.69	23.07
	B11	18.30	16.47	14.94	18.93	22.50
	B2	17.52	22.00	14.30	17.50	22.58
	B22	21.48	20.46	15.03	16.18	23.87
	B3	21.00	17.13	15.22	15.55	24.00
	B33	18.58	18.49	14.96	16.69	22.63
	B4	16.86	15.63	13.74	18.72	23.38
	B44	19.63	23.58	15.41	18.10	23.99
	Mean	**19.03**	**19.13**	**17.42**	**14.76**	**23.25**
Saliva						
	SV1	No C_q	No C_q	22.57	22.79	22.93
	SV11	No C_q	No C_q	22.06	26.16	23.33
	SV2	No C_q	No C_q	20.93	23.26	22.01
	SV22	No C_q	No C_q	21.53	23.87	22.38
	SV3	No C_q	No C_q	22.62	27.47	23.37
	SV33	No C_q	No C_q	22.52	22.38	22.87
	SV4	No C_q	No C_q	20.38	25.51	22.65
	SV44	No C_q	No C_q	21.00	19.03	22.49
	Mean	**No C_q**	**No C_q**	**23.81**	**21.70**	**22.75**
Semen						
	SE1	No C_q	No C_q	23.91	24.16	30.27
	SE11	No C_q	No C_q	27.64	27.96	31.81
	SE2	No C_q	No C_q	29.02	26.85	31.05
	SE22	No C_q	No C_q	24.74	26.72	30.29
	SE3	No C_q	No C_q	27.55	25.85	30.12
	SE33	No C_q	No C_q	26.56	26.57	30.13
	SE4	No C_q	No C_q	22.90	23.98	27.31
	SE44	No C_q	No C_q	25.84	25.03	30.01
	Mean	**No C_q**	**No C_q**	**25.89**	**25.88**	**30.12**

6.4.2 Degradation rate of individual RNA transcripts at different ageing time points

The main aim of the work in this Chapter was to evaluate the effect of mixtures of body fluids on the RER values of blood-specific markers over time. Therefore, the degradation behaviour of the selected RNA markers was investigated in pure and mixed body fluid samples stored under controlled conditions (i.e. room temperature, in a dark dry place) for up to 60 days. The RT-qPCR data for the blood-specific markers is presented in this section and demonstrates that each RNA transcript showed a unique pattern of degradation behaviour in pure and mixed samples, in concordance with the findings presented in Chapter four.

Figure 6.1 illustrates the C_q data for each examined RNA marker (*HBA*, *HBB*, *miR16*, *18S* and *U6*) after efficiency correction in pure bloodstains and bloodstains mixed with either saliva or semen. At day 0, in the control samples (fresh pure blood samples), each of the RNAs examined had a different starting expression level. The miRNA marker (*miR16)* exhibited the highest expression level (lowest C_q value), and *U6* had the lowest expression level (highest C_q value), confirming the results in Chapter four.

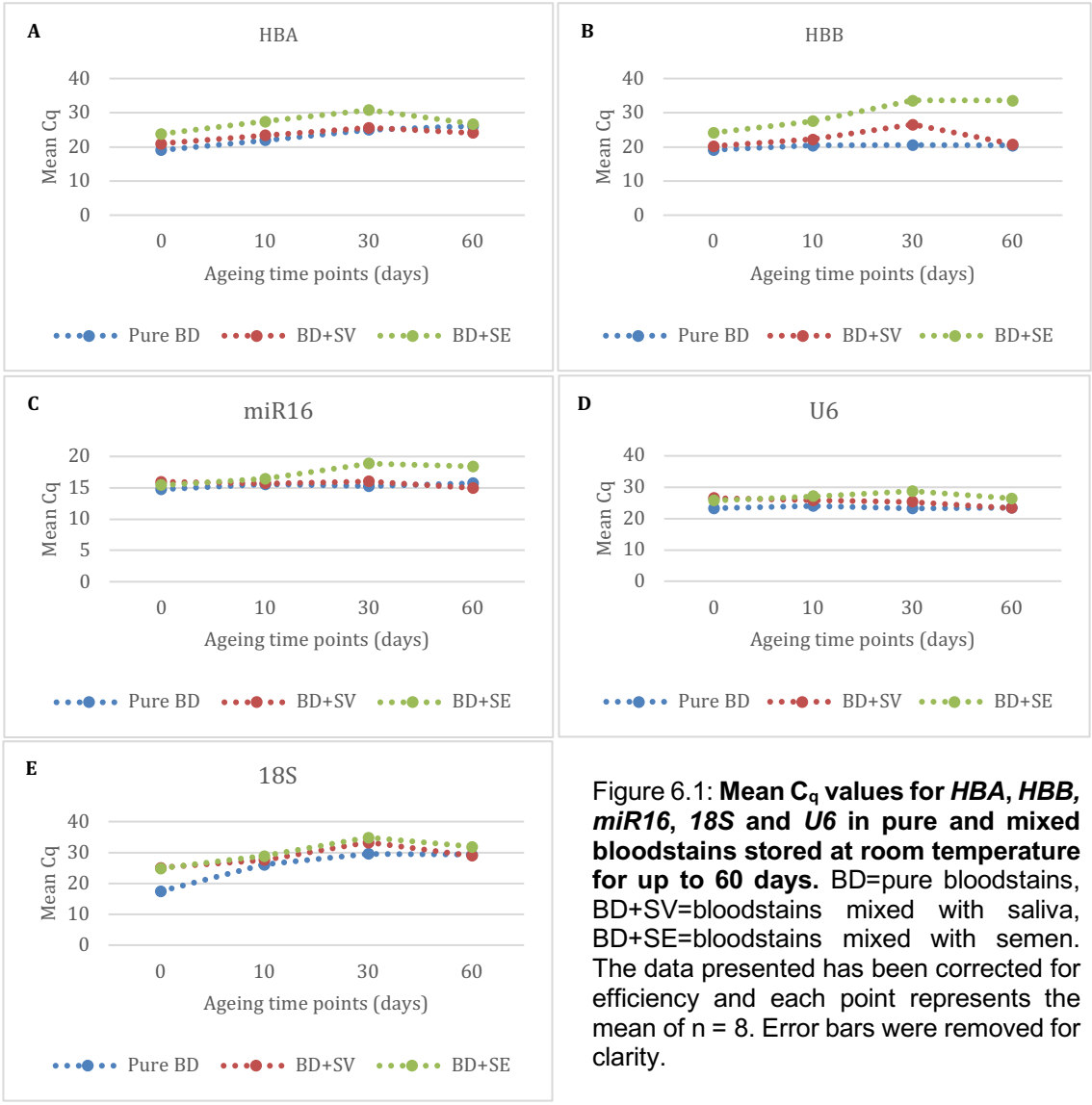


Figure 6.1: **Mean C_q values for *HBA*, *HBB*, *miR16*, *18S* and *U6* in pure and mixed bloodstains stored at room temperature for up to 60 days.** BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. The data presented has been corrected for efficiency and each point represents the mean of n = 8. Error bars were removed for clarity.

The degradation rate of *HBA* was relatively consistent in pure bloodstains (blue dotted line) and bloodstains mixed with saliva (red dotted line), however, this marker exhibited slightly lower expression in bloodstains mixed with semen (Figure 6.1A). The degradation rate of *HBA* in pure bloodstains and mixed samples behaved in a linear manner in the first 30 days of ageing, with the higher C_q values in bloodstains mixed with semen compared to pure bloodstains and bloodstains mixed with saliva being maintained across this time period. At day 60, *HBA* expression level reached a plateau in pure bloodstains and bloodstains mixed with saliva, while in bloodstains mixed with

semen the C_q value decreased at 60 days, i.e. the expression level of *HBA* marker increased.

The second blood-specific mRNA marker (*HBB*) has remained stable across ageing time points in pure bloodstains (Figure 6.1B), confirming the findings in reported in Chapter four. However, it showed degradation in both mixed bloodstain samples across the first 30 days. At 60 days, the quantity level of *HBB* increased (i.e. lower C_q values were obtained) in bloodstains mixed with saliva samples, and stabilised at the same level in bloodstains mixed with semen.

The blood-specific miRNA marker (*miR16*) exhibited very interesting behaviour as it was the only blood-specific marker that remained stable across all ageing time points in all sample types (pure and mixed) with only slight degradation in bloodstains mixed with semen after 30 days of storage (Figure 6.1C).

When exploring the degradation rate of the reference genes in pure and mixed samples, *U6* remained stable across all ageing time points in all sample types (Figure 6.1D), with slight degradation at day 30 in bloodstains mixed with semen. In contrast, *18S*, exhibited a very similar pattern of gradual degradation in pure bloodstains and in both mixed sample types (Figure 6.1E).

6.4.2.1 Statistical analysis

Statistical analysis was performed on each marker in all different sample types. The Anderson-Darling test of normality showed that all data were normally distributed ($p > 0.05$) with the exception of *HBB* data ($p = 0.02$). Therefore, one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test was applied to the *HBA*, *miR16*, *18S* and *U6* data, and a Kruskal-Wallis test on the *HBB* data to determine whether any differences were statistically significant.

When comparing C_q values for *HBA* in pure bloodstains and bloodstains mixed with saliva or semen, no significant differences were found (T = 0.25, p = 0.956 and T = 2.16, p = 0.103 respectively). Similar results were obtained when comparing the C_q values for the *miR16* marker in pure bloodstains and bloodstains mixed with saliva (T = 0.47, p = 0.086) but not when comparing pure bloodstains to bloodstains mixed with semen, where marginally significant differences were identified (T = 2.73, p = 0.042).

Interestingly, the only blood-specific marker that exhibited significant differences in C_q values between pure and mixed bloodstains was *HBB*. The Kruskal-Wallis tests have p-values of 0.022 and 0.021 (with H-value of 7.65 and 7.68) when comparing pure bloodstains to bloodstains mixed with saliva and pure bloodstains to bloodstains mixed with semen, respectively.

Moreover, the statistical analysis of the reference genes showed no significant differences among different sample types (p > 0.05) with the exception of *U6*, which gave a p-value of 0.003 (T-value = 4.47) when comparing its C_q value in pure bloodstains to bloodstains mixed with semen only.

6.4.3 Relative expression Ratio (RER)

The relative expression ratio (RER) was applied to determine the relative expression of RNA markers for pure and mixed bloodstains over time. This approach was used to determine whether mixing bloodstains with other body fluids has an impact on the RER values, and hence on the estimation of bloodstain age. The relative expression of the less stable RNA marker to the more stable marker across all ageing time points was calculated using the same equations used in Chapter four, as shown below.

$$RER = \frac{Cq\ of\ less\ stable\ marker}{C\ of\ more\ stable\ marker}$$

The mean C_q values for blood-specific markers and the reference genes measured by RT-qPCR at 0, 10, 30 and 60 days are shown in Tables 6.5 - 6.7. These data were corrected with the determined efficiency of each assay using *GenEx* statistical software (version 5.4.4) (see section 3.3.4.1).

Table 6.5: **The C_q values of RNA markers at 0, 10, 30, 60 days in pure bloodstains.** Two pure blood samples were analysed from each donor and were tested in duplicate at the RT-qPCR stage (technical repeats). The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

AGEING TIME POINT	RNA MARKERS	SAMPLE NUMBER							
		BD1	BD11	BD2	BD22	BD3	BD33	BD4	BD44
0	***HBA***	18.87	18.30	17.52	21.48	21.00	18.58	16.86	19.63
	HBB	19.31	16.47	22.00	20.46	17.13	18.49	15.63	23.58
	miR16	14.46	14.94	14.30	15.03	15.22	14.96	13.74	15.41
	18S	17.69	18.93	17.50	16.18	15.55	16.69	18.72	18.10
	U6	23.07	22.50	22.58	23.87	24.00	22.63	23.38	23.99
10	***HBA***	21.04	19.90	21.77	24.17	21.32	24.60	17.72	25.12
	HBB	19.13	19.98	22.20	23.25	21.68	21.86	17.82	17.51
	miR16	13.90	15.66	15.44	17.06	15.90	15.62	14.28	16.46
	18S	24.31	24.97	26.22	26.75	25.58	29.92	21.41	29.23
	U6	22.79	22.71	23.84	23.11	23.24	25.34	24.80	26.24
30	***HBA***	15.76	25.95	26.96	27.66	25.97	27.72	25.46	24.87
	HBB	17.11	17.34	20.27	20.96	23.91	23.71	20.86	20.14
	miR16	14.30	14.43	15.80	15.35	15.21	16.72	14.33	15.80
	18S	16.73	26.42	34.59	28.91	28.02	35.72	33.20	32.71
	U6	24.23	21.84	23.12	22.95	22.53	24.06	23.63	23.34
60	***HBA***	27.08	24.95	26.99	26.18	28.12	27.59	24.03	23.82
	HBB	20.38	18.44	20.23	19.66	23.34	22.10	19.73	19.45
	miR16	18.37	15.13	15.24	14.90	15.75	14.40	15.76	16.16
	18S	33.29	24.90	27.40	26.27	35.06	33.32	25.58	28.61
	U6	25.72	22.25	22.43	23.01	22.88	22.32	24.17	24.39

Table 6.6: **The C_q values of RNA markers at 0, 10, 30, 60 days in bloodstains mixed with saliva.** Two pure blood samples were analysed from each donor and were tested in duplicate at the RT-qPCR stage (technical repeats). The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

AGEING TIME POINT	RNA MARKERS	SAMPLE NUMBER							
		BD+SV1	BD+SV11	BD+SV2	BD+SV22	BD+SV3	BD+SV33	BD+SV4	BD+SV44
0	***HBA***	20.34	20.54	21.49	19.79	20.72	23.97	20.76	20.32
	HBB	19.46	19.37	21.17	20.10	19.43	23.32	19.40	19.35
	miR16	16.86	14.33	16.11	16.31	15.95	16.07	15.69	16.36
	18S	24.08	26.31	24.78	25.40	24.82	25.74	23.16	26.67
	U6	27.44	23.98	26.22	25.14	26.55	26.36	28.50	27.96
10	***HBA***	24.67	22.04	24.01	22.40	23.09	24.30	23.17	23.21
	HBB	20.26	20.27	23.68	22.01	22.29	24.98	22.28	22.24
	miR16	16.16	14.60	15.81	16.79	15.28	16.43	14.77	15.53
	18S	29.40	24.02	31.66	26.47	24.10	28.70	28.33	28.03
	U6	26.18	24.33	26.18	26.67	24.77	25.85	25.61	26.66
30	***HBA***	25.87	26.45	27.09	26.08	25.36	25.30	23.08	25.65
	HBB	26.33	26.39	26.19	26.06	27.59	27.70	24.58	27.68
	miR16	16.17	15.37	16.29	15.18	16.26	15.62	16.04	17.07
	18S	33.95	32.98	31.24	31.29	32.41	36.20	31.33	35.06
	U6	25.55	24.15	25.31	24.34	26.01	24.70	25.88	26.21
60	***HBA***	25.57	22.22	23.66	22.88	26.33	22.56	26.45	23.02
	HBB	20.03	20.49	19.92	22.10	21.12	20.55	21.85	20.01
	miR16	15.21	15.08	14.86	14.71	15.22	14.80	15.03	14.80
	18S	27.42	27.97	27.01	33.06	27.83	27.47	31.39	29.52
	U6	24.24	23.60	22.77	23.63	23.90	23.58	24.20	20.57

Table 6.7: **The C_q values of RNA markers at 0, 10, 30, 60 days in bloodstains mixed with semen.** Two pure blood samples were analysed from each donor and were tested in duplicate at the RT-qPCR stage (technical repeats). The data were obtained by correcting the mean raw C_q values measured by RT-qPCR based on the efficiency of each assay.

AGEING TIME POINT	RNA MARKERS	SAMPLE NUMBER							
		BD+SE1	BD+SE11	BD+SE2	BD+SE22	BD+SE3	BD+SE33	BD+SE4	BD+SE44
0	***HBA***	23.50	23.05	22.41	23.97	24.93	23.81	23.20	25.53
	HBB	21.92	25.44	27.09	23.66	24.33	25.08	23.52	22.01
	miR16	16.13	15.77	16.57	16.27	14.40	14.98	15.25	13.72
	18S	27.40	25.16	24.29	24.25	24.06	24.67	24.34	24.88
	U6	27.87	25.11	26.55	26.33	25.24	25.02	25.51	24.35
10	***HBA***	26.28	30.99	31.61	31.65	26.14	28.18	22.43	21.93
	HBB	29.77	29.21	29.16	28.75	29.79	29.57	22.29	22.03
	miR16	16.62	15.51	17.50	17.80	15.65	15.62	18.47	14.22
	18S	29.75	30.54	29.23	28.90	29.81	27.39	30.78	24.77
	U6	26.60	26.56	27.63	28.62	26.56	25.31	29.87	25.53
30	***HBA***	33.60	31.60	32.78	31.34	31.16	28.57	30.89	26.30
	HBB	35.90	35.69	35.34	32.96	34.13	32.29	33.81	28.60
	miR16	18.80	17.75	19.44	17.86	20.61	18.18	21.20	17.03
	18S	36.34	35.01	34.38	35.44	34.83	33.76	35.36	32.69
	U6	27.85	28.62	29.39	28.18	30.00	27.75	31.40	26.76
60	***HBA***	29.48	26.68	29.11	25.03	23.52	27.94	28.23	23.77
	HBB	35.90	34.26	34.39	33.46	33.18	33.24	34.75	29.54
	miR16	20.35	19.03	19.71	18.44	17.07	18.56	17.50	16.51
	18S	35.48	30.31	31.68	29.24	28.48	32.89	36.47	30.54
	U6	29.45	27.56	25.72	24.19	25.39	26.88	25.30	26.26

6.4.3.1 RERs of mRNA to miRNA

The results reported in Chapter four showed that the RER of *HBA/miR16* was positively correlated with ageing time points, but the RER of *HBB/miR16* was not. We therefore examined the effect of mixed samples on the RER of *HBA/miR16* only. Table 6.8 shows the mean RERs of *HBA/miR16* calculated from the corrected C_q values for pure bloodstains and bloodstains mixed with saliva or semen at the different ageing time points.

Table 6.8: **Mean RER of *HBA/miR16* in pure and mixed bloodstain samples at ageing time points 0, 10, 30, 60 days**. The RERs were calculated from the mean C_q values in Tables 6.5 – 6.7. BD=bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. n = 8.

Ageing time point	**RER of *HBA*/*miR16***		
	Pure BD	**BD+SV**	**BD+SE**
0	1.29	1.32	1.56
10	1.41	1.49	1.67
30	1.64	1.60	1.64
60	1.67	1.61	1.45
SD [a]	**0.15**	**0.09**	**0.15**
CV% [b]	**9.66**	**6.25**	**9.59**

[a] Standard deviation [b] Coefficient of variation

When plotting the RER of *HBA/miR16* in the three sample types over time (Figure 6.2), it can be seen that the ratio in pure bloodstains is relatively similar to that in bloodstains mixed with saliva, indicating that the presence of saliva in bloodstains did not affect the *HBA/miR16* ratio. However, this is not the case in bloodstains mixed with semen, as the ratio started with a higher value than in pure bloodstains (1.56) and then over time dropped below this to (1.45) at 60 days.

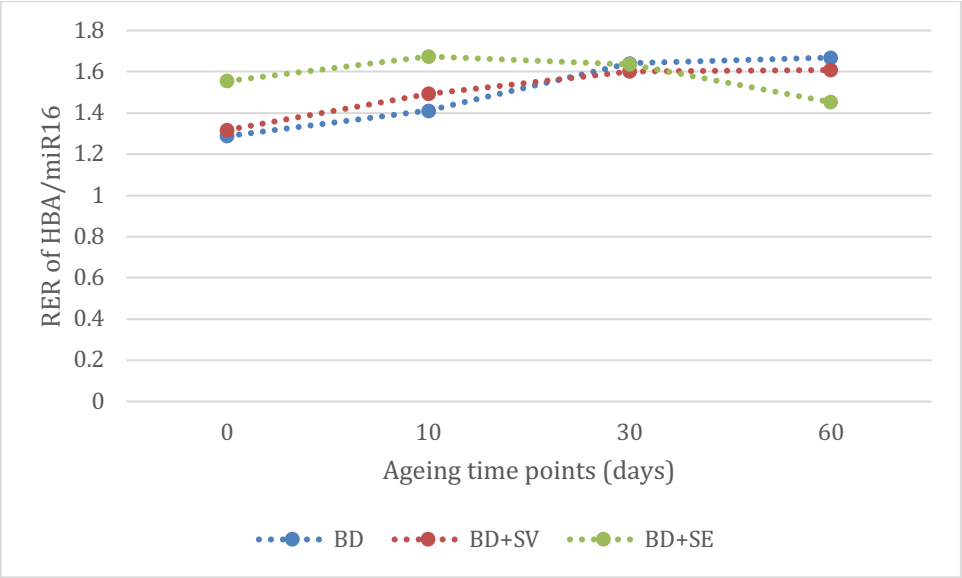


Figure 6.2: **Mean RER of *HBA/miR16* in pure and mixed bloodstains stored at room temperature for up to 60 days.** BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. Each point represents the mean of n = 8. Error bars were omitted for clarity.

The differences between the RERs of *HBA/miR16* in all different sample types were statistically evaluated. When applying the Anderson-Darling normality test, all the data were determined to be normally distributed ($p > 0.05$), therefore parametric analysis was used.

One-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test was used to compare the mean values of the *HBA/miR16* ratio in pure bloodstains to either bloodstains mixed with saliva or bloodstains mixed with semen. There were no significant differences between the RERs of *HBA/miR16* in either mixed sample type when compared with pure bloodstains (Figure 6.3), where $T = 0.02$, $p = 0.999$ comparing pure bloodstains to bloodstains mixed with saliva, and $T = 0.77$, $p = 0.674$ comparing pure bloodstains to bloodstains mixed with semen. These results indicate that the presence of saliva or semen in bloodstains does not affect the RER of *HBA/miR16*.

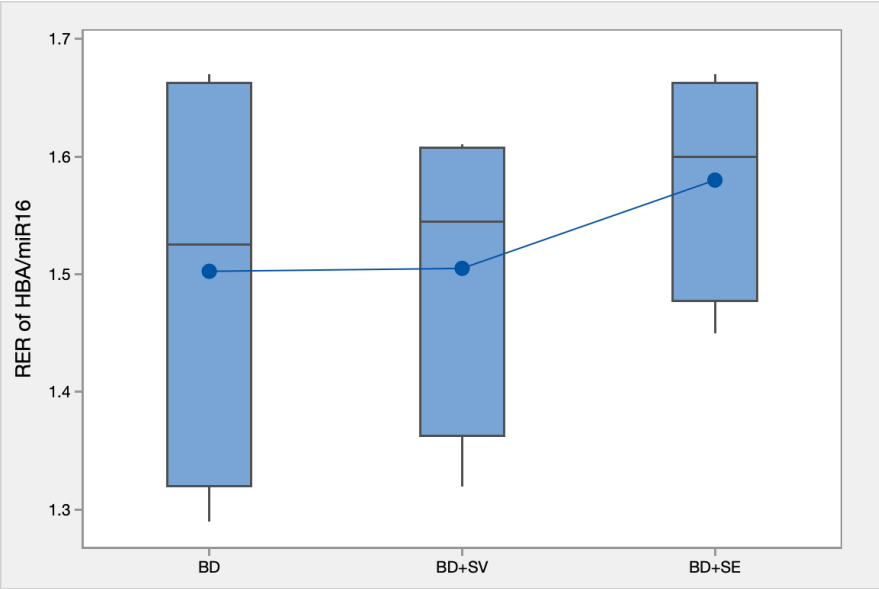


Figure 6.3: **Mean RER of *HBA/miR16* in pure and mixed bloodstains stored at room temperature for up to 60 days.** BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. n = 32.

6.4.3.2 RERs of mRNA and miRNA markers to U6

The RERs of *HBA/U6* and *miR16/U6* have previously been shown to exhibit a significant positive correlation with ageing time points, as presented in Chapter four. Therefore, the effect of mixed body fluids on the RERs of blood-specific markers to the reference gene (*U6*) was also examined. Table 6.9 shows the mean RERs of *HBA* and *miR16* to *U6,* calculated from the corrected C_q values for pure bloodstains and bloodstains mixed with saliva or semen, stored for up to 60 days at room temperature.

Table 6.9: **Mean RER of *HBA* and *miR16* to *U6* in pure and mixed bloodstain samples at ageing time points 0, 10, 30, 60 days.** The RERs were calculated from the mean C_q values in Tables 6.5 – 6.7. BD=bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. n = 8.

Ageing time point	RER of *HBA*/*U6*			RER of *miR16*/*U6*		
	Pure BD	**BD+SV**	**BD+SE**	**Pure BD**	**BD+SV**	**BD+SE**
0	0.78	0.73	0.84	0.63	0.60	0.60
10	0.85	0.90	0.94	0.65	0.61	0.61
30	0.97	1.01	0.99	0.66	0.63	0.66
60	1.00	1.00	0.92	0.67	0.64	0.70
SD [a]	**0.19**	**0.25**	**0.19**	**0.03**	**0.02**	**0.03**
CV% [b]	**20.94**	**26.69**	**20.35**	**5.28**	**3.71**	**4.64**

[a] Standard deviation [b] Coefficient of variation

When plotting the obtained RERs in the three sample types over time (Figure 6.4), there are minimal differences in the RER of *HBA/U6* between pure and mixed bloodstains (Figure 6.4A), indicating that the presence of saliva or semen in bloodstains did not affect this ratio. However, the RER of *miR16/U6* showed some differences when comparing pure bloodstains to mixed bloodstains (Figure 6.4B). In the first 10 days, this ratio was lower in mixed samples compared to pure bloodstains. After 10 days of storage, the RER values increased gradually in bloodstains mixed with saliva, and increased rapidly in bloodstains mixed with semen.

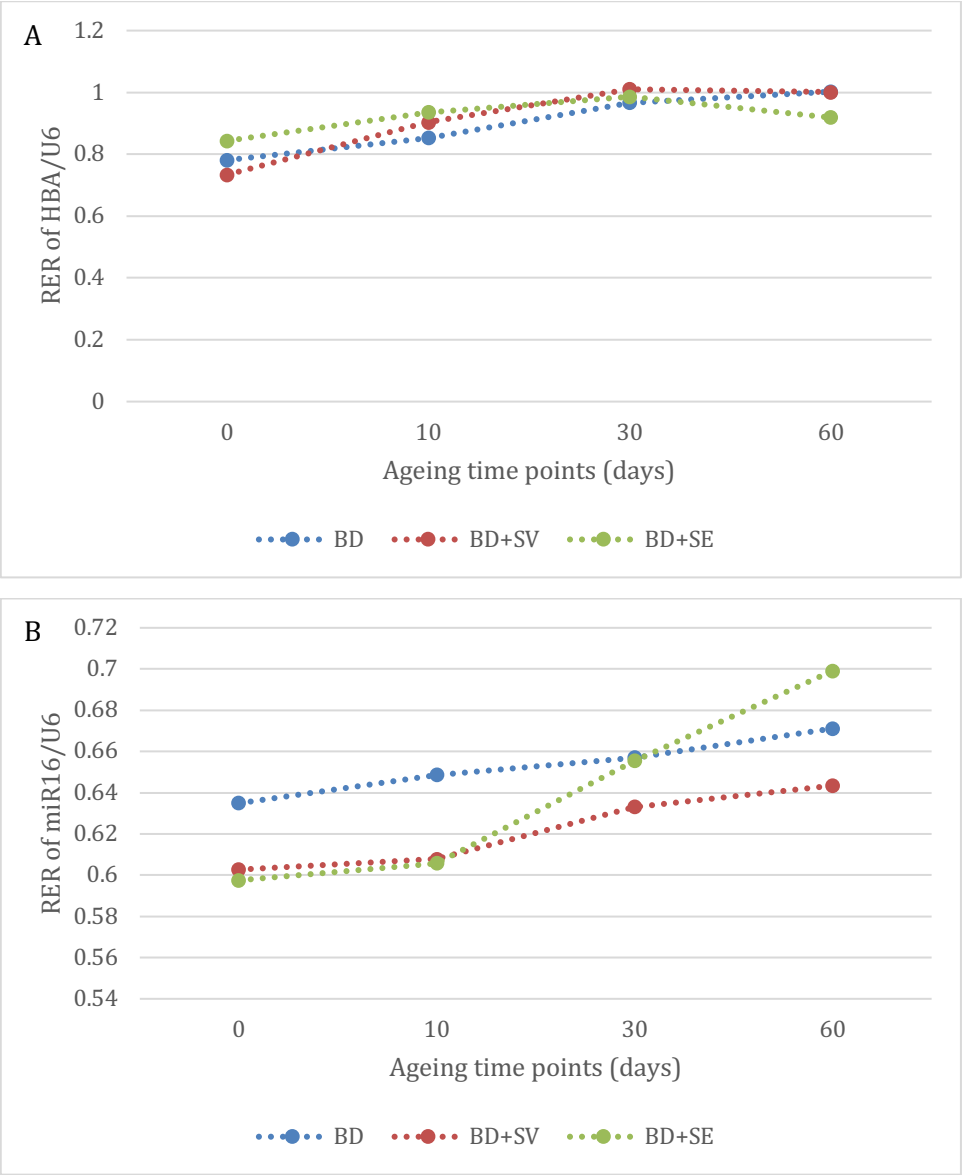


Figure 6.4: **Mean RER of blood-specific markers in pure and mixed bloodstains stored at room temperature for up to 60 days.** RER of (A) *HBA/U6* and (B) *miR16/U6*, BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. Each point represents the mean of n = 8. Error bars were omitted for clarity.

The differences between the RERs of *HBA/U6* and *miR16/U6* in all different sample types were statistically analysed. The data were found to be normally distributed ($p > 0.05$) when applying an Anderson-Darling normality test, so again one-way ANOVA with Dunnett's multiple comparisons test was used to compare the mean values of these ratios in pure bloodstains to either bloodstains mixed with saliva or bloodstains mixed with semen.

There was no significant difference in the RER of *HBA/U6* between pure bloodstains and either of the mixed samples (Figure 6.5A), where T = 0.14, p = 0.986 comparing pure bloodstains to bloodstains mixed with saliva, and T = 0.31, p = 0.932 comparing pure bloodstains to bloodstains mixed with semen. There was a notable difference in the RER of *miR16/U6* between pure bloodstains and mixed bloodstains, however no significant difference was recorded (Figure 6.5B), where T = -0.46, p = 0.858 comparing pure bloodstains to bloodstains mixed with semen, and T = -1.51, p = 0.274 comparing pure bloodstains to bloodstains mixed with saliva.

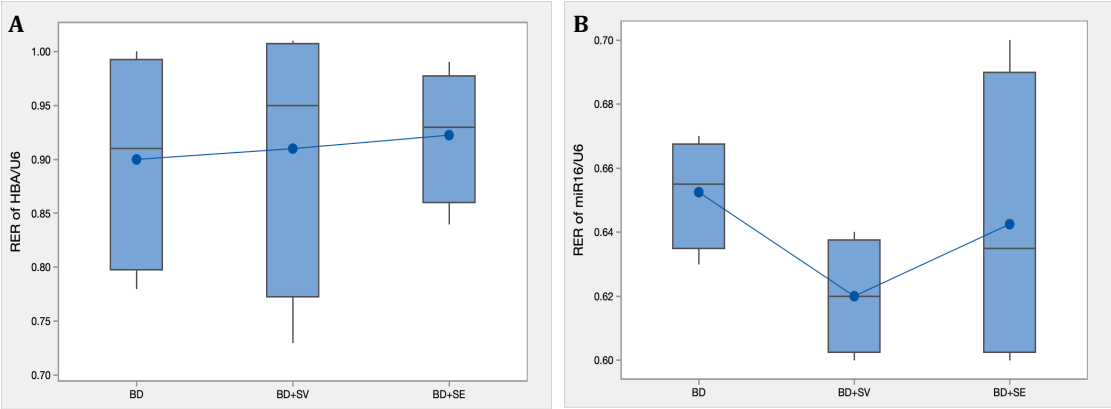


Figure 6.5: **Mean RERs of blood-specific markers to U6 in pure and mixed bloodstains stored at room temperature for up to 60 days.** RER of (A) *HBA/U6* and (B) *miR16/U6*, BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. n = 32.

6.4.3.3 RERs of mRNA markers

Finally, the effect of body fluid mixture on the RER of *HBA/HBB* was examined, as it has also been shown to give a significant positive correlation with ageing time points in blood samples that were aged for up to one year (see Chapter four). Table 6.10 shows the mean RERs of *HBA*/*HBB* calculated from the corrected C_q values for pure bloodstains and bloodstains mixed with saliva or semen stored up to 60 days at room temperature.

Table 6.10: **Mean RER of *HBA* to *HBB* in pure and mixed bloodstain samples at ageing time points 0, 10, 30, 60 days.** The RERs were calculated from the mean C_q values in Tables 6.5 – 6.7. BD=bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. n = 8

Ageing points	RER of *HBA*/*HBB*		
	Pure BD	**BD+SV**	**BD+SE**
0	1.01	1.04	0.99
10	1.08	1.05	1.00
30	1.24	0.96	0.92
60	1.28	1.16	0.80
SD [a]	**0.13**	**0.06**	**0.06**
CV% [b]	**11.86**	**5.45**	**6.84**

[a] Standard deviation [b] Coefficient of variation

The RERs of *HBA/HBB* in the three sample types are shown in Figure 6.6. As expected, in pure bloodstains the ratio increased in a linear fashion with increasing ageing time points. However, this is not the case in mixed samples, as the ratio fluctuated in bloodstains mixed with saliva, and decreased over time in bloodstains mixed with semen. This suggests that the presence of saliva or semen in bloodstains affects the ratio of blood-specific mRNA markers.

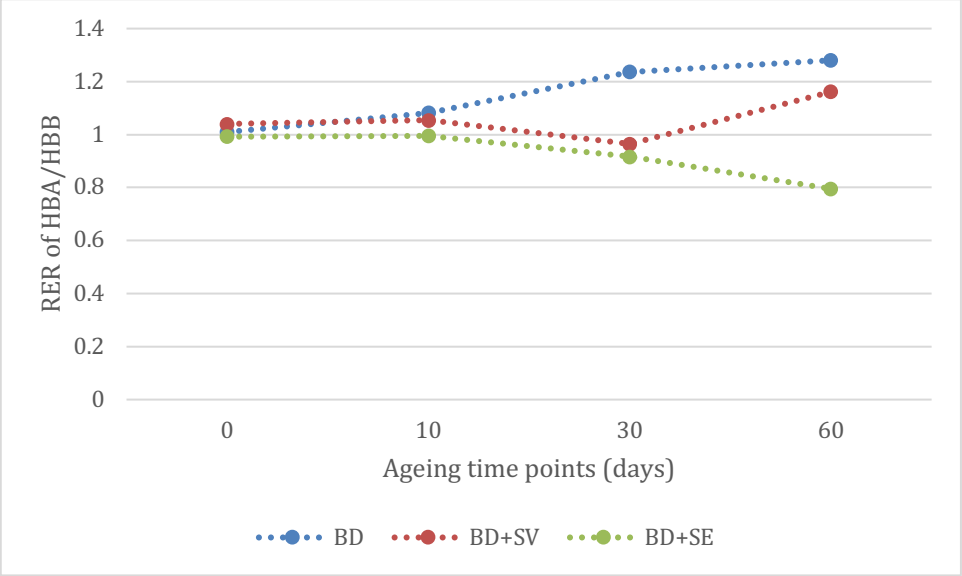


Figure 6.6: **Mean RER of *HBA/HBB* in pure and mixed bloodstains stored at room temperature for up to 60 days.** BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. Each point represents the mean of n=8. Error bars were omitted for clarity.

The differences in the RER of *HBA/HBB* between all different sample types were statistically evaluated. The Anderson-Darling normality test showed that all data were normally distributed ($p > 0.05$). One-way ANOVA with Dunnett's multiple comparisons test was carried out to compare the mean values of *HBA/HBB* ratio in pure bloodstains to both bloodstains mixed with saliva and bloodstains mixed with semen. Despite the different patterns observed in the RER of *HBA/HBB* in pure bloodstains and mixed samples, a significant difference was found only when comparing pure bloodstains to bloodstains mixed with semen (Figure 6.7), where $T = -3.09$, $p = 0.023$. In contrast, no significant difference was found when comparing the same ratio in pure bloodstains to bloodstains mixed with saliva, where $T = -1.37$, $p = 0.330$.

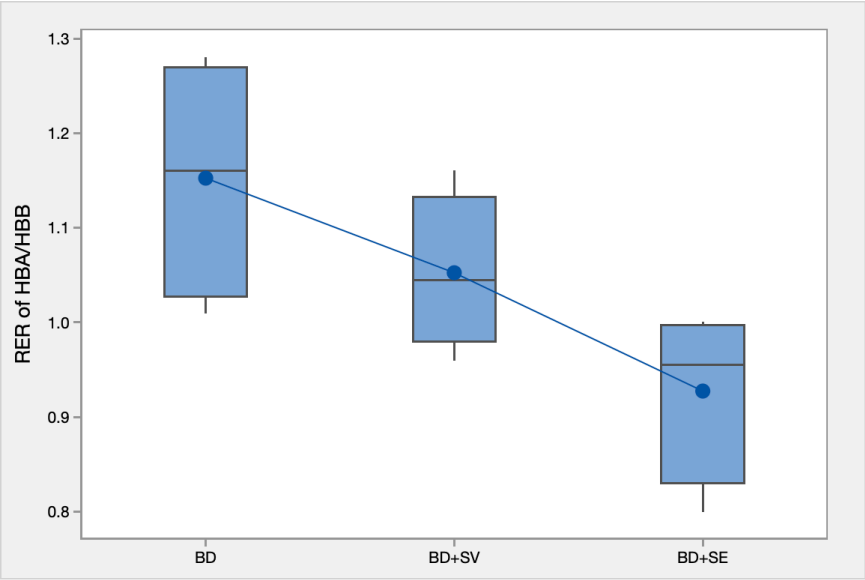


Figure 6.7: **Mean RER of *HBA/HBB* in pure and mixed bloodstains stored at room temperature for up to 60 days.** BD=pure bloodstains, BD+SV=bloodstains mixed with saliva, BD+SE=bloodstains mixed with semen. Each point represents the mean of n = 32.

6.5 Summary and Conclusion

Biological samples that are found at crime scenes are commonly found as mixtures, where two or more body fluids are mixed together, either from the same individual or multiple individuals. This limitation can be resolved by applying a method using body fluid specific markers to both identify body fluid type and estimate the age of the stain simultaneously. The findings reported and published in Chapter four [192] showed that the RERs of blood-specific markers can be considered as a potential method to estimate the age of bloodstains. However, it is important to investigate whether there is any limitation to this proposed method for estimating the time since deposition as a result of body fluids being present in mixed stains. This Chapter therefore evaluated the effect of mixing bloodstains with other body fluid types (saliva and semen) on the RER values of various blood-specific markers, and hence the effect on the estimation of bloodstain age or time since deposition.

Pure bloodstains and bloodstains mixed with saliva or semen were stored for up to 60 days under controlled conditions (i.e. room temperature, in a dark dry

place), and at each desired ageing time point, total RNA was extracted and RNA analysis was performed. The expression level of multiple blood-specific markers (*HBA, HBB* and *miR16*) along with two reference genes (*18S* and *U6*) was quantified by RT-qPCR in all different sample types (i.e. pure and mixed). Blood-specific markers showed low or no expression in the other body fluids (i.e. saliva and semen), confirming that the selected markers are truly blood-specific. These findings are in accordance with the EDNAP collaborative exercise [65] and other studies [64, 194-196] that have investigated the expression of blood-specific markers in different body fluids, where no expression of *HBA* and *HBB* has been found in saliva and semen samples.

The RT-qPCR data for the blood-specific markers demonstrates that each RNA transcript showed a unique pattern of degradation behaviour in pure and mixed bloodstains. *HBA* was the only blood-specific marker which exhibited no significant differences in degradation behaviour between all samples types, while the degradation rate of *miR16* was significantly different between pure bloodstains and bloodstains mixed with semen (but not saliva), and *HBB* degradation was significantly different in all samples types. This pattern indicates that the presence of saliva or semen in bloodstains affects the quantity of blood-specific mRNA and miRNA markers.

The RERs of different combinations of blood-specific markers were calculated for each sample type against the ageing time points: *HBA/miR16*, *HBA/U6*, *miR16/U6,* and *HBA/HBB*. Mixing bloodstains with other types of body fluids such as saliva or semen did not have an effect on the RERs of blood-specific markers, as no significant differences were identified when comparing these RERs in pure bloodstains and bloodstains mixed with either saliva or semen. The only exception was the *HBA/HBB* ratio, which was significantly different between pure bloodstains and bloodstains mixed with semen (but not saliva), due to the different degradation behaviour of *HBB* in the mixed samples.
The findings presented here therefore indicated that the presence of body fluid mixtures did not have an impact on the majority of RERs for blood-specific

RNA markers, suggesting that calculation of these ratios is a promising method for estimating the age of bloodstains. However, to determine how applicable this is across different body fluids, more mixtures should be studied, including sample types such as menstrual blood, vaginal secretions, urine and sweat, as well as mixtures of more than two body fluids. Another potential avenue of research would be to examine different volume ratios of body fluid mixtures, as this study only considered equal volume mixtures of blood with other body fluids. Additionally, there are environmental factors that might affect the RERs of body fluid-specific markers have not been explored as part of this project, such as UV exposure, humidity and temperature. These are all important avenues for future research.

Chapter seven: The stability of reference genes in three types of body fluid.

7.1 Introduction

In gene expression studies, it is crucial to use appropriate reference genes, or control genes, for normalisation in order to be able to reliably interpret the data on expression levels in target genes. This normalisation step corrects and minimises variation among samples that might be introduced by several parameters, such as different amounts of RNA, varying levels of RNA quality or specific experimental errors that may be introduced in the RT-qPCR step [123].

The main specification for the reference genes is that their expression level should be stable between cells in different tissues and under different experimental or environmental conditions. The reference genes must be selected very carefully depending on the conditions of the experiments and the type of samples being examined, in order to generate accurate data. Therefore, the selection of the best reference genes should be based on preliminary experimental validation when comparing different developmental stages or different environmental conditions, and not based on earlier studies or on studies with similar experimental designs.

Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ß-actin (*ACTB*) and *18S* rRNA are among the most frequently selected as single reference genes in many gene expression studies, whether in the clinical or forensic field [142-145], due to the assumption that they show constant expression levels across different experimental conditions. However, some studies have illustrated that these genes are not consistently expressed across body fluids and tissue types, and can vary considerably under different environmental conditions, which may lead to inadequate normalisation [140, 198]. Therefore, this

Chapter has evaluated the stability of two commonly used reference genes (*18S* rRNA and *ACTB* mRNA) in body fluid samples aged for up to one year.

7.1.1 *18S rRNA*

18S is a member of the ribosomal RNA family, which is the most abundant RNA type in the cell (i.e. thousands of copies per cell). It is a component of the ribosome structure, which is the site for protein synthesis in cells. In eukaryotic cells, ribosomes are divided into two subunits, the large (*60S*) and small subunits (*40S*). *18S* rRNA is a component of the small ribosomal subunit (*40S*), while *5S*, *5.8S* and *28S* rRNA form the large ribosomal subunit (*60S*).

18S is often used as a reference gene in many gene expression studies, including many forensic gene expression studies. However, when using this particular gene as a reference gene for mRNA markers, caution should be taken, as there is an imbalance between the rRNA and mRNA fractions within a cell. This is because rRNA is present in much higher abundance than mRNA transcripts, creating difficulties for the normalisation procedure [199]. In addition, *18S* is not present in purified mRNA samples and it has been reported that some biological factors and drugs affect rRNA transcription [147]. The main drawbacks in using *18S* as a single reference gene is therefore that any changes or variation in *18S* expression do not reliably reflect changes in the mRNA, as it makes up almost half of the total RNA in a cell, whereas the mRNA only contributes 1-2% of the total RNA [147], and its transcription is carried out by RNA polymerase I, so its regulation synthesis is independent from mRNA synthesis which carried out by RNA polymerase II [200].

In contrast, some studies have shown that *18S* rRNA is stable and less prone to degradation than mRNA markers when tested in aged biological samples [58, 60, 87]. This could be due to the presence of the ribosomal complex, which protects it from environmental factors. The same studies have also examined the relative expression ratio of *18S* rRNA and the mRNA marker (*ACTB*) in

different biological samples such as blood, saliva and hair, across multiple ageing time points, in an attempt to indicate the age of the biological samples. These studies have successfully found a significant relationship between the *ACTB/18S* ratio and ageing time points.

7.1.2 *ACTB*

The *ACTB,* or *ß-actin,* gene encodes one of the six actin proteins, which are highly conserved and involved in various types of cell motility, structure, integrity and intercellular signalling [169]. *ACTB* is ubiquitously expressed in all eukaryotic cells, and the protein encoded is a major constituent of the contractile apparatus in muscle cells of the heart and one of the two non-muscular cytoskeletal actins.

ACTB has a large number of pseudogenes, which can affect the reliability of *ACTB* as a reference gene. These pseudogenes could be transcribed, and would therefore be amplified from RNA samples during PCR, thereby contributing to the yield of the RT-qPCR. Therefore primers must be selected carefully, so the primer or probe is designed to cross an exon/exon boundary, and the samples must have any genomic DNA contamination removed by incorporating a DNase I digestion step before analysis [201]. Despite the fact that *ACTB* is ubiquitously expressed in all eukaryotic cells, some studies have shown that that the expression level of *ACTB* varies across body fluid and tissue types [202, 203]. However, *ACTB* is still commonly used as a reference gene in body fluid identification studies.

7.2 Aims and Objectives

The aim of this chapter was to assess the stability and the behaviour of two commonly used reference genes, *18S* rRNA and *ACTB,* in aged body fluid samples. The expression level of these reference genes was analysed across

multiple ageing time points using RT-qPCR in three different body fluids (blood, saliva and semen) that were stored for up to one year.

The key objective of this work was to investigate the stability of *18S* rRNA and *ACTB* in dried body fluid samples and to determine for how many days they remained stable, and at what point they became so degraded that their expression level could not be detected. This study therefore evaluated the suitability of *18S* and *ACTB* for normalisation purposes in forensic body fluid studies, as well as whether their relative expression ratio could be used to indicate the age of the biological stains.

7.3 Materials and Methods

7.3.1 Sample collection

A total of 29 volunteers (13 males and 16 females), donated blood (10 donors), saliva (10 donors) and semen samples (9 donors). The blood samples were collected onto sterile cotton swabs using disposable Unistik 3 comfort lancets. 20 µL of blood was pipetted onto the swabs and then stored in a dark dry place at room temperature to simulate natural ageing, until they reached a series of desired ageing time points (0, 3, 6, 15, 30, 90, 180, 270, and 360 days). The saliva and semen samples were deposited into sterile collection pots. 50 µL of each sample was pipetted onto sterile cotton swabs and then stored in a dark, dry place at room temperature to simulate natural ageing until they reached a series of desired ageing time points (0, 7, 14, 28, 90, 180, 270, and 360 days). The experimental procedures used were approved by the Departmental Ethics Committee in the Department of Pure and Applied Chemistry at the University of Strathclyde (see Appendix A1).

7.3.2 Total RNA extraction and reverse transcription

RNA extraction was carried out using TRI Reagent, as described in section 2.4.1. The TURBO DNA-free^{TM} Kit was used to treat the extracted RNA to

remove any genomic DNA, as described in section 2.5. The reverse transcription was carried out using High-Capacity cDNA Reverse Transcription kit and TaqMan® MicroRNA Reverse Transcription cDNA Kit, as described in section 2.7.1 and 2.7.2.

7.3.3 Quantitative real-time PCR (qPCR)

All TaqMan assays were run in singleplex assays, following the procedure described in section 2.8. Amplification was performed in a Stratagene Mx3005P.

7.3.3.1 TaqMan assays®

The TaqMan® Gene Expression Assays used in this Chapter were off-the-shelf, and were predesigned for each transcript. Both *ACTB* and *18S* markers used in this experiment were control reagents from Applied Biosystems (Life Technologies, Paisley, UK), each with a unique ID. Table 7.1 shows the characteristics of the *ACTB* and *18S* assays.

Table 7.1**: Characteristics of the TaqMan® Gene Expression Assays for the reference genes.**

Reference gene	Applied Biosystems TaqMan® assay ID	Amplicon length (nt)	Assay location*
ACTB	Hs99999903_m1	171	53
18S	Hs99999901_s1	187	604

* Refers to the nucleotide location that is the midpoint of the target region.

7.3.4 Data analysis

The data generated from RT-qPCR was analysed using *MxPro,* and *GenEx* software (version 5.4.4) was used for efficiency correction of the raw data (see section 2.9). *Microsoft Excel 2016* was used to manipulate raw C_q values from

the RT-qPCR reactions and to present basic data and line graphs. *Minitab Express* (version 1.5.0) was used for statistical analyses, including the Anderson-Darling normality test and one-way analysis of variance (ANOVA). For the nonparametric statistical data, Kruskal-Wallis and Dunn's multiple comparison analysis were performed with the R software package (version 1.3.5) [154].

7.4 Results and Discussion

7.4.1 Degradation rate of *18S* and *ACTB* RNA transcripts at different ageing time points

In this Chapter, the expression level of the *ACTB* and *18S* RNA transcripts was quantified by RT-qPCR in blood, saliva and semen samples stored in a dark dry place at room temperature for up to one year, in order to determine whether either of them would be a suitable reference gene for body fluid identification studies. The expression levels were measured and compared in fresh and aged samples. All measured C_q values were corrected against the obtained efficiency of each assay.

7.4.1.1 Blood samples

The RT-qPCR data for *ACTB* and *18S* in blood samples is presented in this section and demonstrates that each reference gene showed a unique pattern of degradation behaviour in blood samples. Neither of the selected genes showed stability in their expression level across ageing time points (Figure 7.1).

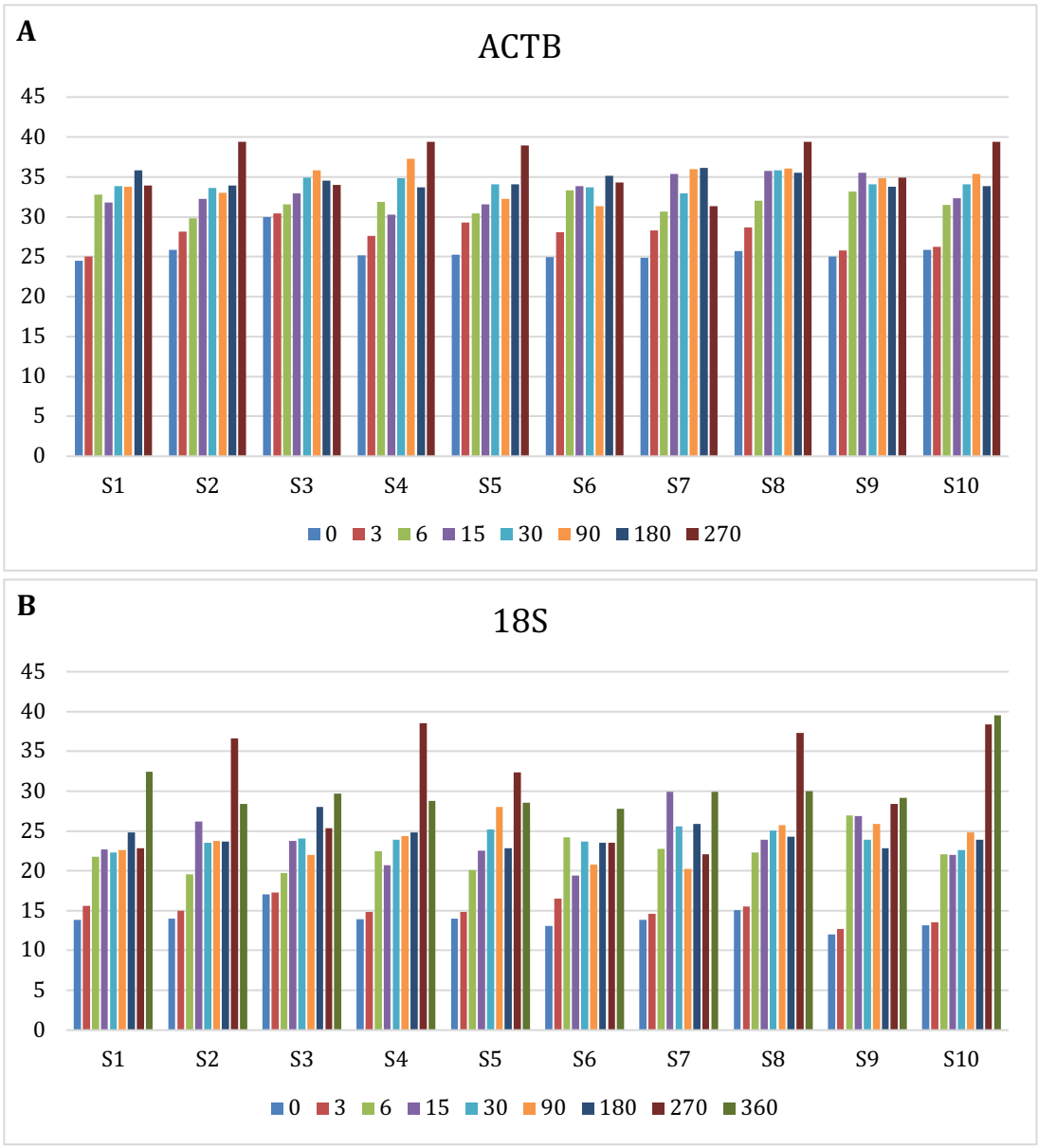


Figure 7.1: **Histogram plots showing C_q values of (A) *ACTB* and (B) *18S* in bloodstains from 10 individuals (S1-S10) stored at room temperature for up to one year.** The data presented have been corrected for efficiency.

The gradual increase in C_q values across storage time can indicate the degradation of RNA. Figure 7.2 illustrates mean C_q data for *ACTB* and *18S* across all donors. At day 0, in the control samples (fresh blood samples), *ACTB* and *18S* had different starting expression levels. The reference gene *18S* exhibited a higher expression level (lower C_q value) than *ACTB*, which had a lower expression level (higher C_q value). These findings are in

accordance with Usarek et al. (2017), who investigated the expression of a number of reference genes including *18S* and *ACTB* in venous blood samples [204].

Neither *18S* nor *ACTB* remained stable across the 360 days in any of the ten samples. The *18S rRNA* showed the largest reduction in quantity over the time period, compared to *ACTB*. As can be seen from Figure 7.2, the mean C_q for *18S* remained at the same level in the first three days and then started to gradually increase after that point, indicating that the *18S* degraded as the age of the bloodstains increased. The highest reduction was observed between 3 to 6 days and 180 to 270 days.

ACTB, in contrast, started to degrade from the first three days and continued to degrade until it reached the detection level of the assay, showing no C_q value after 270 days.

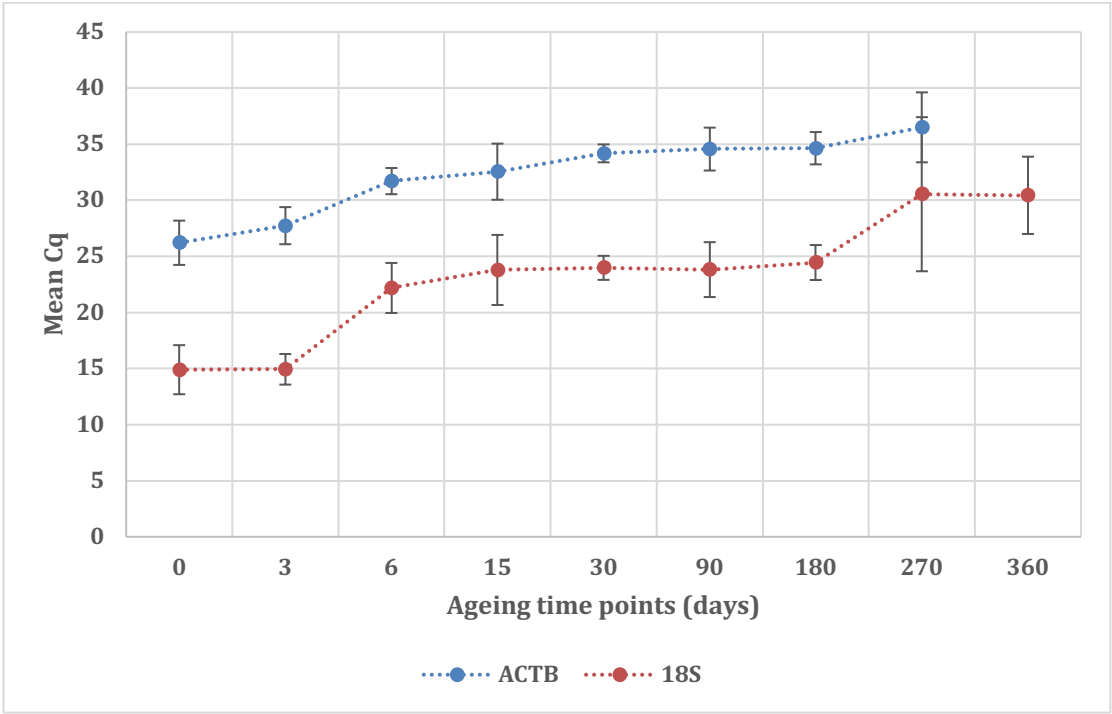


Figure 7.2**: Mean C_q for *ACTB* and *18S* in blood samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 10. Error bars represent standard error.

The difference in C_q values (ΔC_q) is presented in Figure 7.3, and can illustrate more clearly the degradation level of an RNA target at each time point, relative to the control point (T = 0 days). The outcomes presented in Figure 7.3 suggest that the degradation rate of *18S* was most significant between 3 and 6 days and between 180 and 270 days. The quantity of *18S* remained relatively stable between 15 and 180 days. The degradation rate of *ACTB* started in a linear fashion in the first 30 days and then remained stable at the same level between 30 to 180 days, before it started to degrade again.

One-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test was used to examine the difference between the degradation rate at each ageing time point to the control time point (T = 0), for both the *ACTB* and *18S* markers. The test confirmed that there was no significant change in the degradation rate across the first 3 days for either reference gene, which means that they both remained stable for up to 3 days. However, significant differences were found when comparing all other ageing time points to the control time point (all $p < 0.0001$), for both *ACTB* and *18S*.

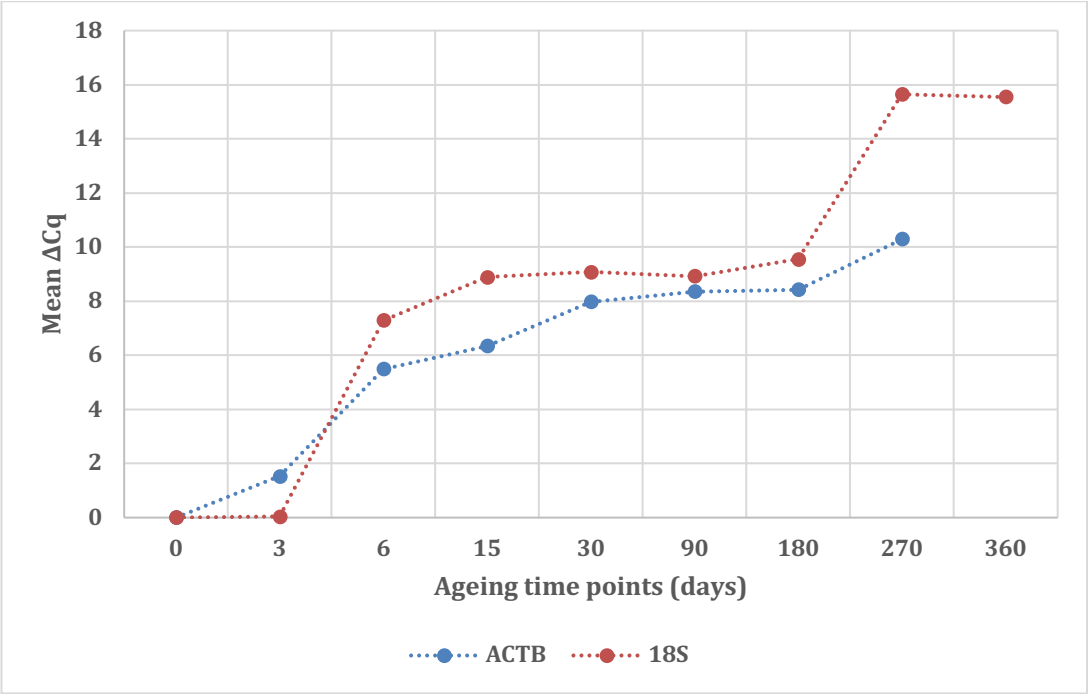


Figure 7.3**: Mean ΔC_q for *ACTB* and *18S* in blood samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 10. Error bars were removed for clarity.

One previous study has investigated the expression level of *18S* in blood samples stored at room temperature for up to 40 hours, and found that the C_q value of *18S* was significantly higher after ageing than in fresh samples, which is not the case in the presented work here. The shortest ageing time point investigated here was 3 days (72 hours), after which the C_q value of *18S* did not show a significant increase compared to the fresh samples.

Another study found that the quantity level of *ACTB* was highly variable in whole blood samples, and did not meet the criteria for use as a reference gene. The inter-donor differences that were observed in their study could have influenced the variability of *ACTB* level in whole blood samples [205]. An individual's genotype at various sites, environmental factors and disease states are all potential causes of variation in expression level among individuals. Moreover, the expression level of *ACTB* was found to significantly decrease over time when tested in newborn blood samples stored at ambient

temperature for time periods ranging from 3 to 16 years [206]. The behaviour of *ACTB* in the bloodstains examined in this work has confirmed these previous findings, as it was shown to degrade across ageing time points and did not remain stable.

7.4.1.2 Saliva samples

The degradation behaviour of the *ACTB* and *18S* markers in saliva samples was different from in blood samples. In all 10 samples, the *ACTB* marker was shown to be reasonably stable in the first 28 days, before giving high C_q values at 28 days, then falling below the detection level of the assay indicating high levels of degradation. *18S* degraded in the first 90 days and then remained at the same level, showing no further degradation. This marker reached a plateau in saliva samples earlier than in blood samples, and was detected across all ageing points (Figure 7.4).

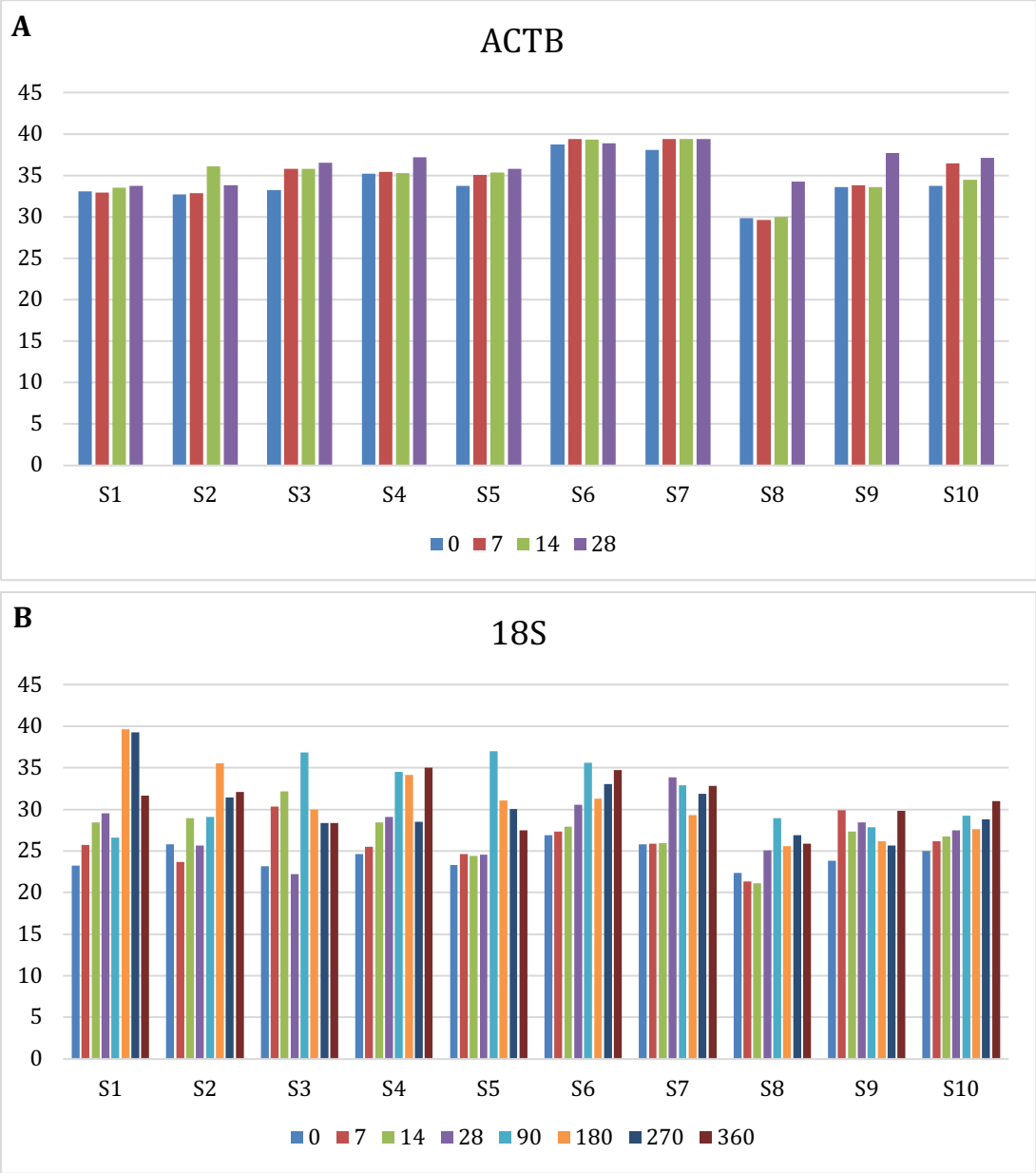


Figure 7.4: **Histogram plots showing C_q values of (A) *ACTB* and (B) *18S* in saliva samples from 10 individuals (S1-S10) stored at room temperature for up to one year.** The data presented have been corrected for efficiency.

Figure 7.5 illustrates the mean C_q for both *ACTB* and *18S* in saliva samples across ageing time points. It can be clearly observed that the quantity levels of both reference genes in fresh saliva samples (0 days) are much lower than in blood samples (higher C_q values). The low expression level of *ACTB* in saliva samples support the data of Lindenbergh et al. (2013), who have shown that the expression of *ACTB* was low in saliva compared to other body fluids

such as blood and menstrual secretion [207]. Whereas *18S* has shown to have a higher expression than *ACTB* in saliva samples, suggesting that this may be a more sensitive marker in challenging body fluid types such as saliva (i.e. due to the presence of a large number of endogenous and exogenous ribonucleases and bacteria). However, despite having low expression levels in saliva, *ACTB* did show high stability in the first 28 days of the current study, and Moreno et al. (2012) were also able to detect *ACTB* in saliva samples after having been exposed to environmental factors for one or two weeks [140]. The quantity of the *18S* marker started to decrease in the first 7 days and remained at the same level before degrading again after 28 days, then stabilising again and reaching a plateau (Figure 7.5).

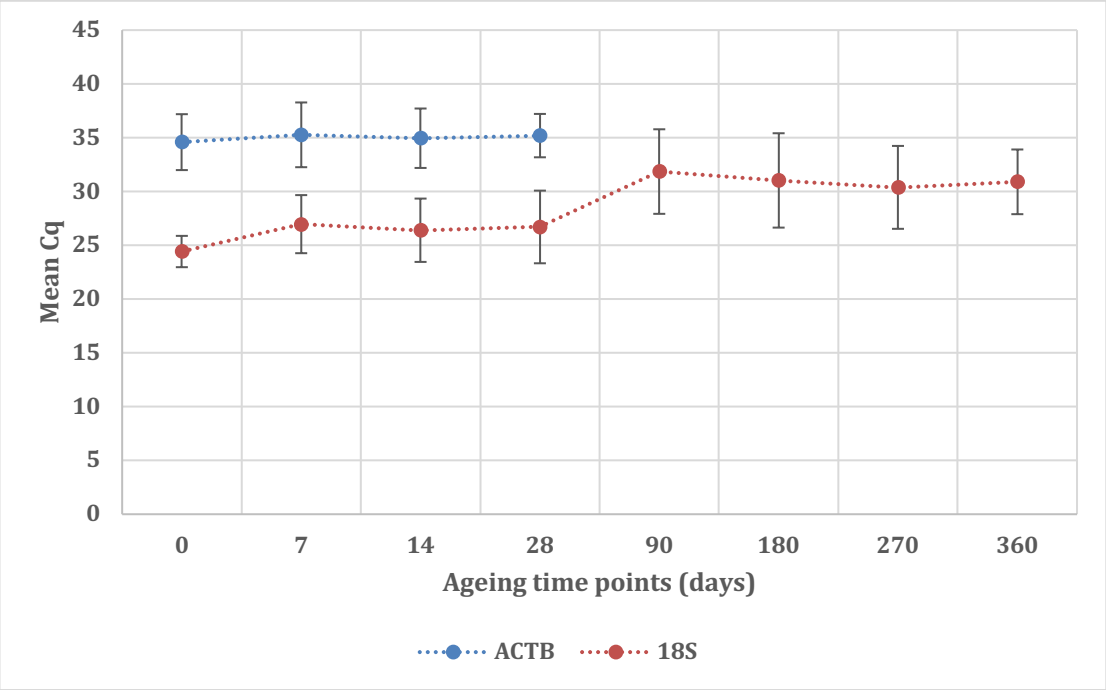


Figure 7.5: **Mean C_q data for *ACTB* and *18S* in saliva samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 10. Error bars represent standard error.

The mean ΔC_q is shown in Figure 7.6, indicating that for *ACTB,* this measure remained around zero up to 28 days, indicating high stability, before the marker became undetectable. For the *18S* gene, a reasonable level of stability

can be seen up to 28 days, before a dramatic increase in ΔC_q indicating that there was a high degradation rate for this gene between 28 and 90 days.

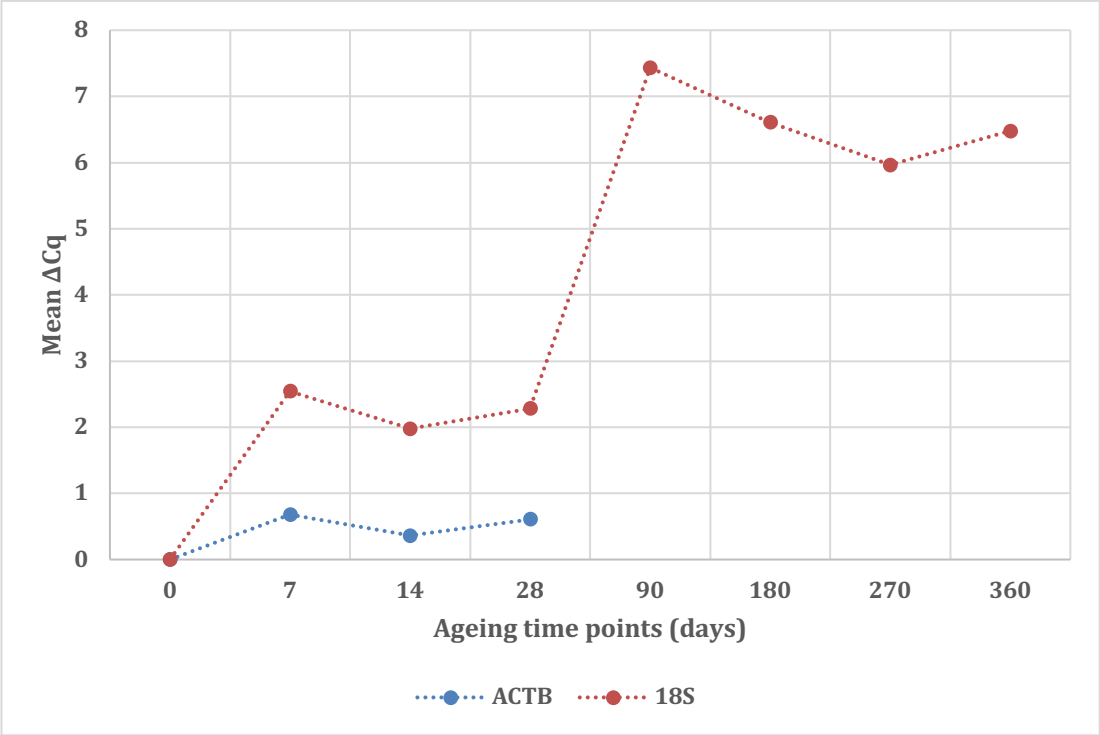


Figure 7.6: **Mean ΔC_q for *ACTB* and *18S* in saliva samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 10. Error bars were removed for clarity.

One-way ANOVA with Dunnett's multiple comparison test was carried out on the saliva sample data to examine the difference between degradation rate at each ageing time point compared to the control time point (T = 0), for both the *ACTB* and *18S* markers. This indicated that there was no significant change in expression level for either *ACTB* or *18S* in the first 28 days. However, after 90 days there was a significant difference in degradation rate compared to the control time point for the *18S* gene (all $p < 0.0001$).

The data presented here is in contrast to the findings of Sakurada et al. (2013), who investigated the stability of *ACTB* and *18S* in body fluids stored at room temperature for up to one year, and found that the C_q values of both reference genes increased significantly with time [208]. In their work, both *ACTB* and

18S were still detected after one year, whereas this is not the case here for *ACTB,* which was completely degraded after 28 days. The difference between these studies could be explained by the different sizes of amplicon length used, as Sakurada et al. used a 77bp amplicon, while a 171bp amplicon was used in the current project, and longer amplicons are likely to degrade faster than shorter amplicons.

7.4.1.3 Semen samples

In semen samples, both *ACTB* and *18S* showed unique degradation patterns when compared to blood and saliva samples. In all nine samples, *ACTB* showed no C_q values after only two weeks of ageing, whereas *18S* was still detected after one year (Figure 7.7). The overall mean expression levels of *ACTB* and *18S* in fresh semen samples (Figure 7.8) were much higher than in saliva samples and lower than in blood samples. When looking at mean C_q values, it can be seen that rapid degradation occurred in the first 7 days for *ACTB* and *18S*, and then *18S* showed another period of rapid degradation between 14 to 90 days, as shown in Figure 7.8.

Moreno et al. (2012) examined the effect of environmental exposure (25.5°C and no precipitation) on gene expression level of a number of reference genes, including *ACTB*, and found a decrease in *ACTB* expression level after only one day, although it was still detected after two weeks, [140]. In this study, the *ACTB* marker also showed an expression profile up to 14 days in samples stored at room temperature, before its level dropped below the detection level of the assay (Figure 7.8).

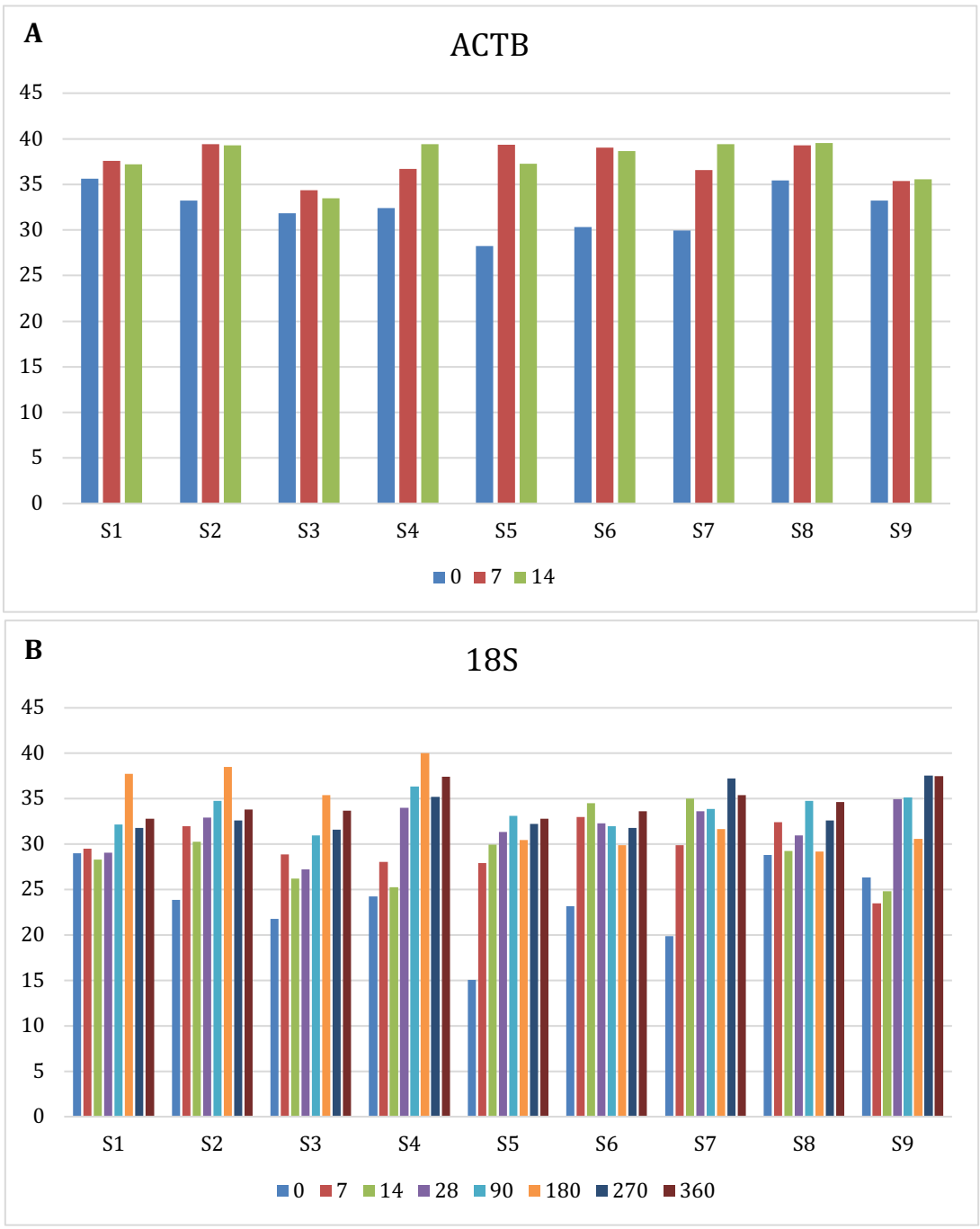


Figure 7.7: **Histogram plots showing C_q values of (A) *ACTB* and (B) *18S* in semen samples from 9 individuals stored at room temperature for up to one year.** The data presented have been corrected for efficiency.

As illustrated in Figure 7.8, *18S* started to show an increase in degradation rate in the first 7 days of ageing, and after 14 days continued to degrade across a period of six months before stabilising at the same level until it reached 360 days of ageing. These findings are in agreement with the study of Sherier

(2016), where *18S* degraded in the early stages of ageing (2 weeks) in semen samples stored at room temperature [209]. Moreover, when investigating very old semen samples, Nakanishi et al. (2014) showed that *18S* could still be detected after ageing semen samples for 33 and 56 years [188].

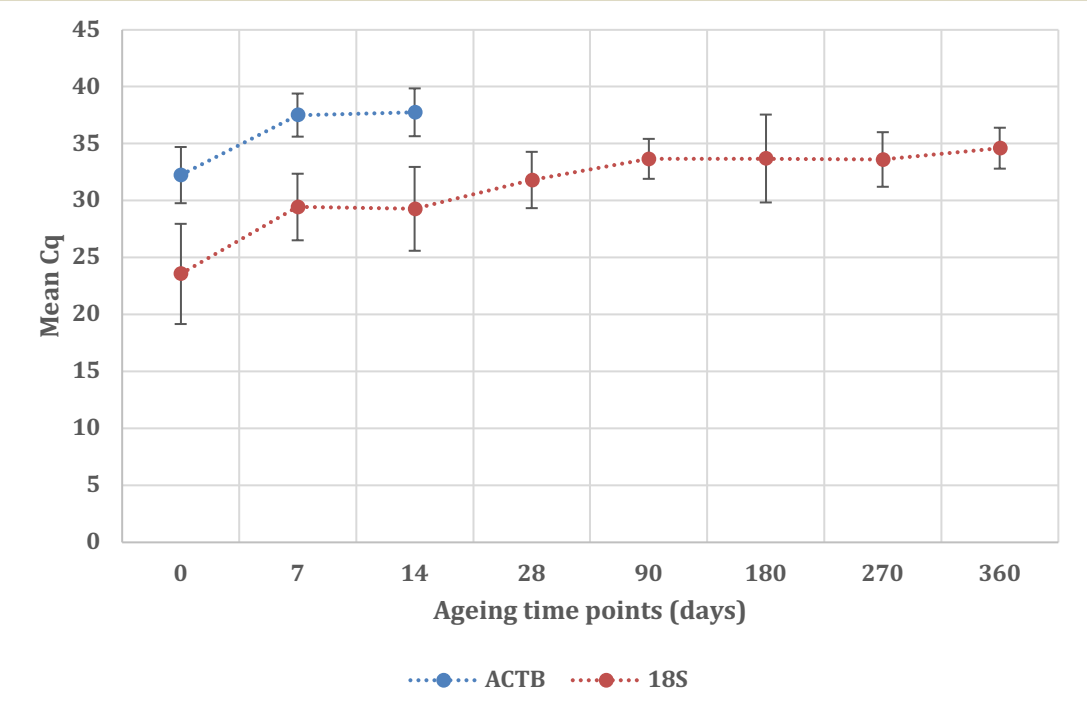


Figure 7.8: **Mean C_q data for *ACTB* and *18S* in semen samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 9. Error bars represent standard error.

The difference in C_q between each time point and the control time point (ΔC_q) is presented in Figure 7.9, and illustrates more clearly that in the first 7 days there was a dramatic increase in degradation for both *ACTB* and *18S*, and for *18S* there was a second period during which degradation increased, between 14 to 90 days.

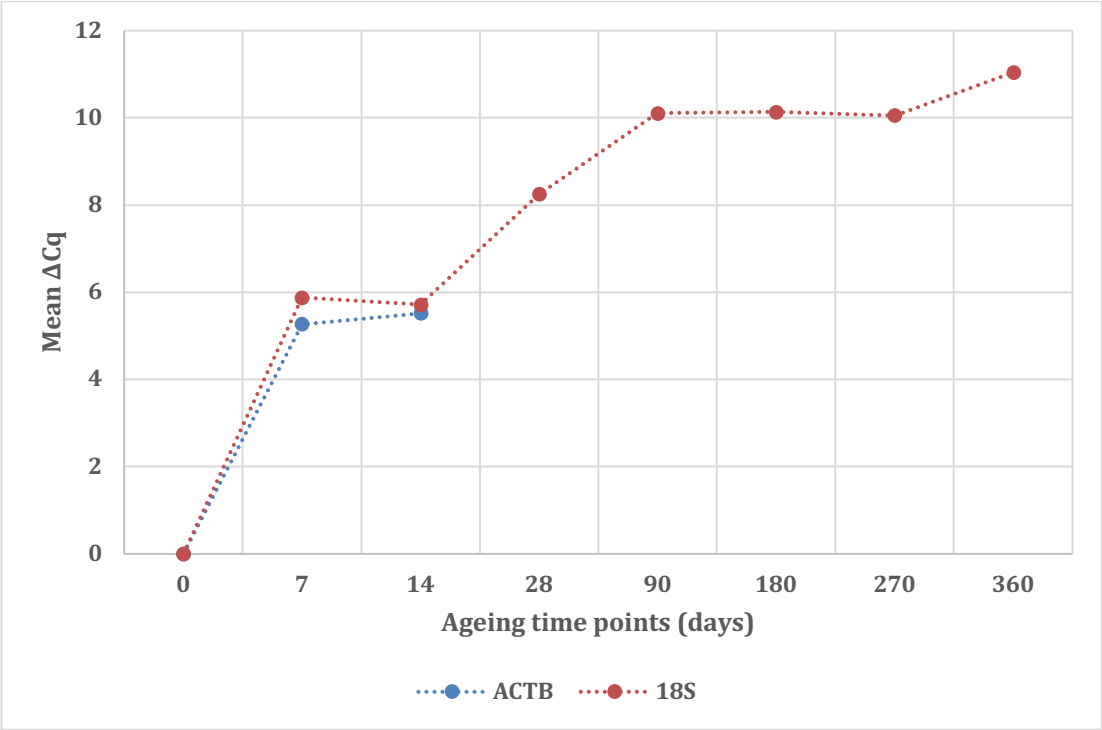


Figure 7.9: **Mean ΔC_q data for *ACTB* and *18S* in semen samples stored at room temperature for up to one year.** The data presented have been corrected for efficiency and each point represents the mean of n = 9. Error bars were removed for clarity.

When one-way ANOVA with Dunnett's multiple comparison test was applied to the semen sample data, significant increases in degradation were detected at all ageing time points when compared to the control time point (T = 0), for both *ACTB* and *18S* (all $p < 0.0001$). These data support the findings of Sakurada et al. (2013), who showed that the mean C_q values of *ACTB* and *18S* increased significantly with time in semen samples stored for up to 1 year [208], although *ACTB* was still detected after one year. These differences could be explained again by the different sizes of the amplicon used, as Sakurada et al. (2013) used a shorter amplicon (77 bp) compared to the current project, which used a longer amplicon (171 bp), and longer amplicons tend to degrade faster than shorter amplicons.

7.4.2 Inter-donor variation

At each time point, the inter-donor variation was calculated using the corrected C_q values to obtain the coefficient of variation (CV%). Figure 7.10 shows CV% across all ageing time points for each of the selected reference genes (*ACTB* and *18S*), in all three body fluids. It can be seen that *ACTB* showed lower inter-donor variation than *18S* in all body fluids. It could be argued that this may be due to the early disappearance of the *ACTB* marker from the samples over time, however, when looking at CV% values only in the ageing time points where *ACTB* was detected, this variation was still lower than *18S*, as shown in Tables 7.2, 7.3, and 7.4.

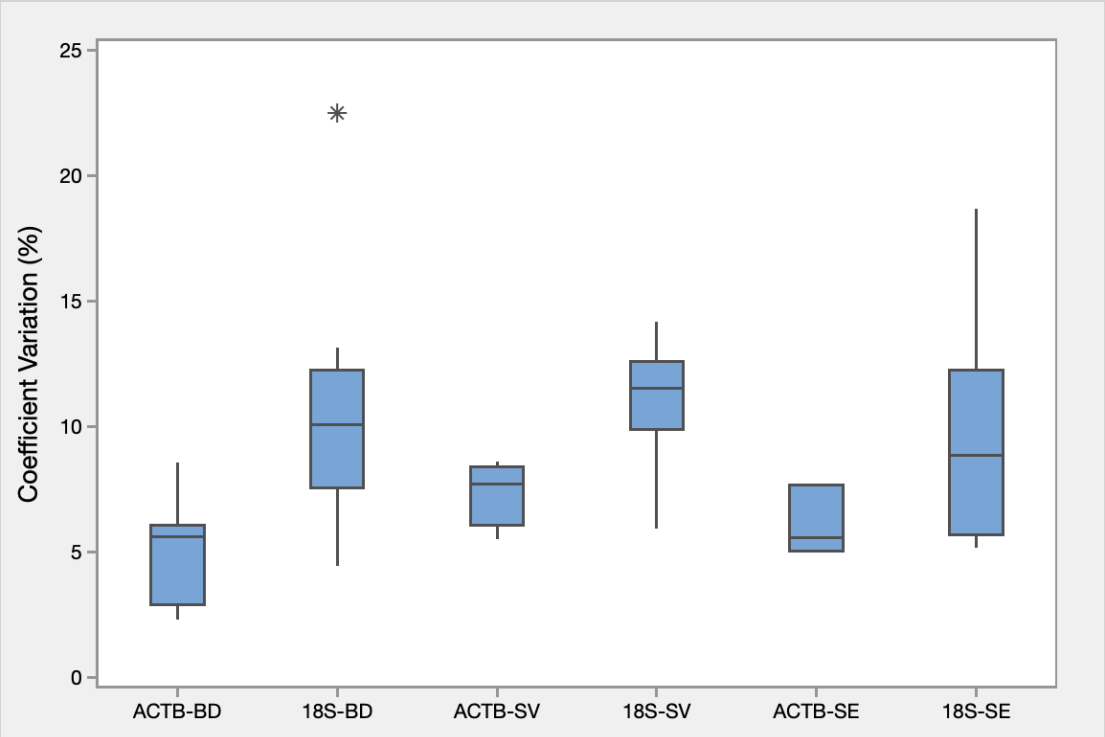


Figure 7.10: **Boxplots showing inter-donor variation for *ACTB* and *18S* in blood, saliva and semen samples.** The plots were obtained using Minitab Express, (BD=blood, SV=saliva, SE=semen). *represents outliers.

Generally speaking, these reference genes have shown variation in expression level in different body fluids across ageing points, suggesting that

neither *ACTB* nor *18S* should be used as reference genes in body fluid identification studies, or in studies that are trying to estimate or predict the age of biological stains. Even in fresh samples (control samples), there was high inter-donor variation among participants for both genes, which could be explained by different factors such as overall health, age and diet, or any other reasons that have been mentioned earlier such as environmental factors and disease states. Another explanation for the observed high variation between samples could be laboratory error. Many manual steps are involved in the procedure, which can lead to variation of this type. This variation was minimised by the same individual performing each step, and all samples being homogenised for the same period of time during the RNA extraction step. Small sample size could be another contributor to the high variability in the expression level of the reference genes and increasing sample number may be one way of reducing inter-donor variation in future experiments.

Table 7.2**: Inter-donor variation for *ACTB* and *18S* in blood.** The coefficient of variation (CV%) was calculated using the mean of n = 10 for samples aged up to 360 days at room temperature.

Age points	***ACTB***	***18S***
Day 0	6.09	9.38
Day 3	5.96	8.71
Day 6	3.69	10.06
Day 15	5.67	13.14
Day 30	2.34	4.47
Day 90	5.53	10.27
Day 180	2.65	6.37
Day 270	8.54	22.49
Day 360	ND	11.31

* ND = Not detected

Table 7.3: **Inter-donor variation for *ACTB* and *18S* in saliva.** The coefficient of variation (CV%) was calculated using the mean of n = 10 for samples aged up to 360 days at room temperature.

Age points	*ACTB*	*18S*
Day 0	7.60	5.96
Day 7	8.58	10.36
Day 14	7.83	10.84
Day 28	5.53	12.21
Day 90	ND	12.34
Day 180	ND	14.13
Day 270	ND	12.68
Day 360	ND	9.73

* ND = Not detected

Table 7.4: **Inter-donor variation for *ACTB* and *18S* in semen.** The coefficient of variation (CV%) was calculated using the mean of n = 9 for samples aged up to 360 days at room temperature.

Age points	*ACTB*	*18S*
Day 0	7.65	18.66
Day 7	5.05	9.93
Day 14	5.56	12.57
Day 28	ND	7.77
Day 90	ND	5.21
Day 180	ND	11.24
Day 270	ND	7.13
Day 360	ND	5.19

* ND = Not detected

7.4.3 Relative expression ratio (RER)

There are a small number of studies that have used the RER of *ACTB* to *18S* in order to estimate the age of biological samples such as blood [58], saliva [87], and hair [60], where *18S* exhibited higher stability than *ACTB*, which showed increasing degradation rate with increasing storage time. Therefore, the same analysis was applied in this work, in order to replicate the findings of the previous work.

7.4.3.1 The RER of *ACTB* to *18S* in bloodstains

Table 7.5 below shows C_q values for *ACTB* and *18S* in fresh and aged bloodstains from 10 donors, measured in duplicate using RT-qPCR at 0, 3, 6, 15, 30, 90, 180, 270, and 360 days. These data were corrected with the determined efficiency of each assay using *GenEx* software (version 5.4.4).

Table 7.5: **Mean of duplicate C_q values for *ACTB* and *18S* in blood samples from 10 donors, aged for 0, 3, 6, 15, 30, 90, 180, 270, and 360 days, after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR.

Age points	Markers	Sample number/Mean C_q values									
		1	2	3	4	5	6	7	8	9	10
Day 0	*ACTB*	24.50	25.83	29.99	25.15	25.22	24.96	24.91	25.74	25.02	25.84
	18S	13.83	13.97	17.01	13.93	13.99	13.11	13.82	15.08	12.06	13.19
Day 3	*ACTB*	25.01	28.12	30.40	27.62	29.30	28.06	28.26	28.65	25.75	26.21
	18S	15.57	15.00	17.26	14.81	14.81	16.48	14.64	15.50	12.71	13.55
Day 6	*ACTB*	32.76	29.82	31.58	31.90	30.46	33.32	30.62	32.01	33.15	31.50
	18S	21.78	19.55	19.68	22.43	20.11	24.19	22.74	22.34	26.95	22.09
Day 15	*ACTB*	31.79	32.23	32.93	30.25	31.55	33.82	35.36	35.72	35.52	32.36
	18S	22.67	26.21	23.74	20.69	22.53	19.38	29.92	23.87	26.87	22.03
Day 30	*ACTB*	33.87	33.62	34.88	34.80	34.09	33.70	32.94	35.81	34.07	34.07
	18S	22.29	23.49	24.06	23.93	25.19	23.69	25.56	25.07	23.94	22.57
Day 90	*ACTB*	33.74	32.99	35.86	37.28	32.24	31.33	35.96	36.03	34.84	35.40
	18S	22.57	23.78	22.00	24.35	28.06	20.80	20.24	25.74	25.87	24.81
Day 180	*ACTB*	35.83	33.95	34.54	33.72	34.03	35.15	36.09	35.50	33.78	33.86
	18S	24.80	23.66	28.01	24.80	22.84	23.50	25.85	24.29	22.87	23.90
Day 270	*ACTB*	33.89	39.42	34.01	39.42	38.96	34.31	31.31	39.42	34.87	39.42
	18S	22.82	36.64	25.34	No Cq	32.37	23.49	22.09	37.33	28.42	No Cq
Day 360	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	32.47	28.40	29.69	28.77	28.57	27.82	29.95	30.02	29.19	No Cq

The mean RERs of *ACTB* to *18S* calculated from the corrected C_q values across all donors for bloodstains stored up to one year at room temperature are shown in Table 7.6. This measure could only be calculated up to 270 days as the expression level of *ACTB* fell below the detection level of the assay after this ageing time point.

Table 7.6**: Mean RER of *ACTB* to *18S* for 10 blood samples at ageing time points 0, 3, 6, 15, 30, 90, 180, 270 days.** The RERs were calculated from the mean C_q values in Table 7.5.

Age points	RER of *ACTB*/*18S*
Day 0	1.84
Day 3	1.85
Day 6	1.44
Day 15	1.41
Day 30	1.43
Day 90	1.46
Day 180	1.42
Day 270	1.23

The pattern of RER values for *ACTB/18S* obtained in this study contrasted with the findings of Anderson et al. (2005), who found that the ratio of *ACTB/18S* in bloodstains increased with increasing age time points. However, their blood samples were stored in a chamber at 25 °C and 50% humidity for a period of 150 days, while the samples in this project were stored at room temperature in a dark dry place for 360 days. The ageing time points considered in their work were 0, 30, 60, 90, 120 and 150 days, whereas shorter ageing time points were considered in the initial stages of the study reported here, giving more details about the degradation rate of both reference genes in the first 30 days of ageing.

Even though *18S* remained stable across the 150 days of ageing examined in Anderson et al.'s work, it only showed high stability in the first three days in the study presented here, and started to degrade after that, reaching a plateau after 270 days. There was a slight reduction in the rate of degradation observed between 15 to 180 days. In contrast, *ACTB* did not show any stability in this study, and degraded in a linear fashion across ageing time points.

When applying the Anderson-Darling normality test to determine whether the data was normally distributed or not, it was established that the obtained RERs of *ACTB/18S* at different ageing time points were not normally distributed ($p = 0.015$). Therefore, nonparametric tests were used for statistical analysis. A Spearman's correlation test indicated that the RER values of *ACTB/18S* exhibited a significant negative correlation with the age of bloodstains ($r = -0.74$, $p = 0.037$). This negative correlation was a result of the degradation pattern of the *18S* marker as it did not remain stable. Therefore, the findings of the current project did not replicate the findings of previous studies, which showed that the RER of *ACTB/18S* was positively correlated with bloodstains age [58, 59].

The findings presented here were not the only study that could not replicate previous work on RER of *ACTB/18S* and bloodstain age, as Simard et al. (2012), who also concluded that the degradation rate of rRNA relative to mRNA did not show any significant differences (i.e. correlation) over the time points when analysing blood samples stored at room temperature for a period of six months. Therefore, a correlation between mRNA/*18S* ratios and storage time could not be established in their study [88]. In this current project, on the other hand, a significant negative correlation was found between the RER of *ACTB/18S* and ageing time points.

7.4.3.2 The RER of *ACTB* to *18S* in saliva

The RERs of *ACTB/18S* were also calculated in saliva samples to determine whether they showed consistent patterns in body fluids other than blood. Table 7.7 below shows the C_q values of both *ACTB* and *18S* in fresh and aged saliva samples from ten donors, measured in duplicate using RT-qPCR at 0, 7, 14, 28, 90, 180, 270, and 360 days. These data were corrected with the determined efficiency of each assay using *GenEx* software (version 5.4.4).

Table 7.7: **Mean of duplicate C_q values for *ACTB* and *18S* in saliva samples from 10 donors, aged for 0, 7, 14, 28, 90, 180, 270, and 360 days, after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR.

Age points	Markers	Sample number/Mean C_q values									
		1	2	3	4	5	6	7	8	9	10
Day 0	*ACTB*	33.10	32.69	33.21	35.20	33.72	38.75	38.05	29.85	33.56	33.77
	18S	23.27	25.79	23.19	24.66	23.29	26.92	25.82	22.39	23.84	25.01
Day 7	*ACTB*	32.91	32.84	35.78	35.40	35.03	39.42	39.42	29.59	33.83	36.44
	18S	25.72	23.69	30.34	25.49	24.63	27.37	25.90	21.31	29.91	26.19
Day 14	*ACTB*	33.50	36.11	35.80	35.31	35.35	39.30	39.42	30.02	33.58	34.51
	18S	28.41	28.96	32.19	28.41	24.38	27.96	25.96	21.14	27.35	26.73
Day 28	*ACTB*	33.72	33.84	36.54	37.19	35.79	38.85	39.42	34.28	37.70	37.11
	18S	29.57	25.64	22.25	29.12	24.54	30.57	33.83	25.09	28.46	27.47
Day 90	*ACTB*	36.11	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	35.15	35.34	35.84
	18S	26.62	29.09	36.82	34.51	36.97	35.59	32.87	28.95	27.87	29.24
Day 180	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	No Cq	35.54	29.95	34.15	31.09	31.29	29.30	25.56	26.17	27.63
Day 270	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	No Cq	31.42	28.34	28.50	30.02	33.07	31.87	26.91	25.68	28.77
Day 360	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	31.65	32.10	28.39	35.03	27.48	34.71	32.84	25.91	29.86	31.00

The mean RERs of *ACTB* to *18S* calculated from the corrected C_q values across all donors for saliva stains stored up to one year at room temperature are shown in Table 7.8. This measure could only be calculated up to 28 days as the expression level of *ACTB* fell below the detection level of the assay after this ageing time point, and no C_q value was recorded.

To determine whether there was a relationship between ageing time points and the RER of *ACTB/18S* in saliva samples, a Pearson's correlation test was applied as an Anderson-Darling normality test showed that the data was normally distributed ($p = 0.08$). The correlation test revealed that there is no relationship between the RER of *ACTB/18S* and the ageing time points ($r = -0.63$, $p = 0.37$). It can be concluded that the findings of the current project did not show any trend (i.e. correlation) between ageing time points and the RER of *ACTB/18S*, therefore, did not replicate the outcomes of a previous study that showed a positive correlation between the RER of *ACTB/18S* and the age of saliva stains [87].

Table 7.8: **Mean RER of *ACTB* to *18S* for 10 saliva samples at ageing time points 0, 7, 14, 28 days.** The RERs were calculated from the mean C_q values in Table 7.7.

Age points	RER of *ACTB*/*18S*
Day 0	1.40
Day 7	1.35
Day 14	1.31
Day 28	1.33

The previous study, Alrowaithi (2013), identified a significant relationship between the RERs of *ACTB/18S* and ageing time points in saliva samples, using the same methodologies as those applied in the current project [87], finding that these RERs increased with increasing ageing time points.

However, Simard et al. (2012) were not able to obtain a relationship between the RER of the selected reference genes (*ACTB* and *18S*) and the age of the saliva samples [88]. The differences in the findings of these studies might be due to factors such as the quantity of total RNA in the saliva samples, limited RNA marker detection due to the low sensitivity of the assays used, and variation between donors. Additionally, the nature of saliva samples means that they are challenging to work with due to the presence of bacterial and viral RNA, which might influence the analysis of RT-qPCR.

7.4.3.3 The RER of *ACTB* to *18S* in semen

Finally, the RERs of *ACTB/18S* were calculated in semen samples to examine how this parameter varied in semen compared to blood and saliva. Table 7.9 below shows the C_q values of both *ACTB* and *18S* in fresh and aged semen samples from 9 donors, measured in duplicate using RT-qPCR at 0, 7,14, 28, 90, 180, 270, and 360 days. These data were corrected with the determined efficiency of each assay using *GenEx* software (version 5.4.4).

Table 7.9: **Mean of duplicate C_q values for *ACTB* and *18S* in semen samples from 9 donors, aged for 0, 7, 14, 28, 90, 180, 270, and 360 days, after efficiency correction.** The data were obtained by correcting the mean raw C_q values measured by RT-qPCR.

Age points	Markers	Sample number/Mean C_q values								
		1	2	3	4	5	6	7	8	9
Day 0	*ACTB*	35.63	33.23	31.85	32.41	28.22	30.32	29.95	35.41	33.20
	18S	29.02	23.83	21.78	24.24	15.07	23.15	19.86	28.83	26.33
Day 7	*ACTB*	37.59	39.42	34.37	36.70	39.34	39.04	36.55	39.29	35.34
	18S	29.49	31.98	28.84	28.05	27.94	33.01	29.85	32.38	23.46
Day 14	*ACTB*	37.21	39.30	33.50	39.42	37.25	38.63	39.42	39.54	35.58
	18S	28.27	30.28	26.23	25.23	29.93	34.50	35.04	29.23	24.82
Day 28	*ACTB*	36.82	No Cq	34.07	No Cq	38.11	No Cq	No Cq	No Cq	No Cq
	18S	29.06	32.94	27.23	34.01	31.32	32.31	33.62	30.95	34.93
Day 90	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	32.14	34.77	30.96	36.33	33.10	31.96	33.90	34.76	35.14
Day 180	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	37.71	No Cq	35.40	No Cq	30.44	29.87	31.63	34.66	30.59
Day 270	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	31.75	32.63	31.57	35.18	32.24	31.79	37.23	32.59	No Cq
Day 360	*ACTB*	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq	No Cq
	18S	32.78	33.84	33.67	No Cq	32.76	33.59	35.36	34.61	No Cq

The mean RERs of *ACTB* to *18S* calculated from the corrected C_q values across all donors for semen stains stored up to one year at room temperature are shown in Table 7.10. As previously, this measure could only be calculated up to the ageing time point at which *ACTB* could still be detected, which for semen was 14 days. Even though the RER data for semen are limited and cannot be used to determine whether a significant relationship exists with ageing time points, it can be seen that the determined values did not show any trend of increasing or decreasing with ageing time points. These findings were expected given that the degradation profile of *18S* was similar to *ACTB* across the first 14 days (see Figure 7.8), therefore, a correlation between the RERs of *ACTB/18S* and ageing time points was unlikely to be obtained.

Since no previous studies have been conducted to determine the relationship between the relative expression ratio of these reference genes and the time of storage for semen samples, it was not possible to compare the findings presented here with any other studies. Overall, the findings indicate that the *ACTB/18S* ratio is not a useful parameter to calculate for estimating the age of the semen samples.

Table 7.10**: Mean RER of *ACTB* to *18S* for 9 semen samples at ageing time points 0, 7, 14 days.** The RERs were calculated from the mean C_q values in Table 7.9.

Age points	RER of *ACTB*/*18S*
Day 0	1.40
Day 7	1.28
Day 14	1.30

7.4.4 Relative quantity over time

As the degradation profile of the rRNA marker (*18S)* was similar to that of the mRNA marker (*ACTB)* in the current project conditions, it was expected that there would be no correlation between the ratio of *ACTB/18S* and the age of the biological stains, in contrast to the work reported previously [58-60, 87]. This was confirmed by the results presented in the previous section, where there was no relationship between the RER of *ACTB* to *18S* and the age of the body fluid stains, with the exception of blood samples where negative correlation was observed. To investigate this relationship further, we calculated an alternative measure; the relative expression over time using $2^{-\Delta Cq}$ values. This was done in order to determine whether there was a significant relationship between the relative quantities of the two reference genes and the age of the body fluid stains (see section 2.9.2), where:

$$\Delta C_q = C_{q\ (T=x)} - C_{q\ (T=0)}$$

This measure examines the remaining quantity of a given reference gene, relative to the quantity of that gene present at the start of the experiment, i.e. ageing time point zero.

7.4.4.1 Blood samples

The relative quantity of *ACTB* and *18S* over time was first calculated in blood samples using the data presented in Table 7.5. The remaining quantity of the *ACTB* and *18S* markers are illustrated in Figure 7.11, where the mean relative quantity was measured at each ageing time point.

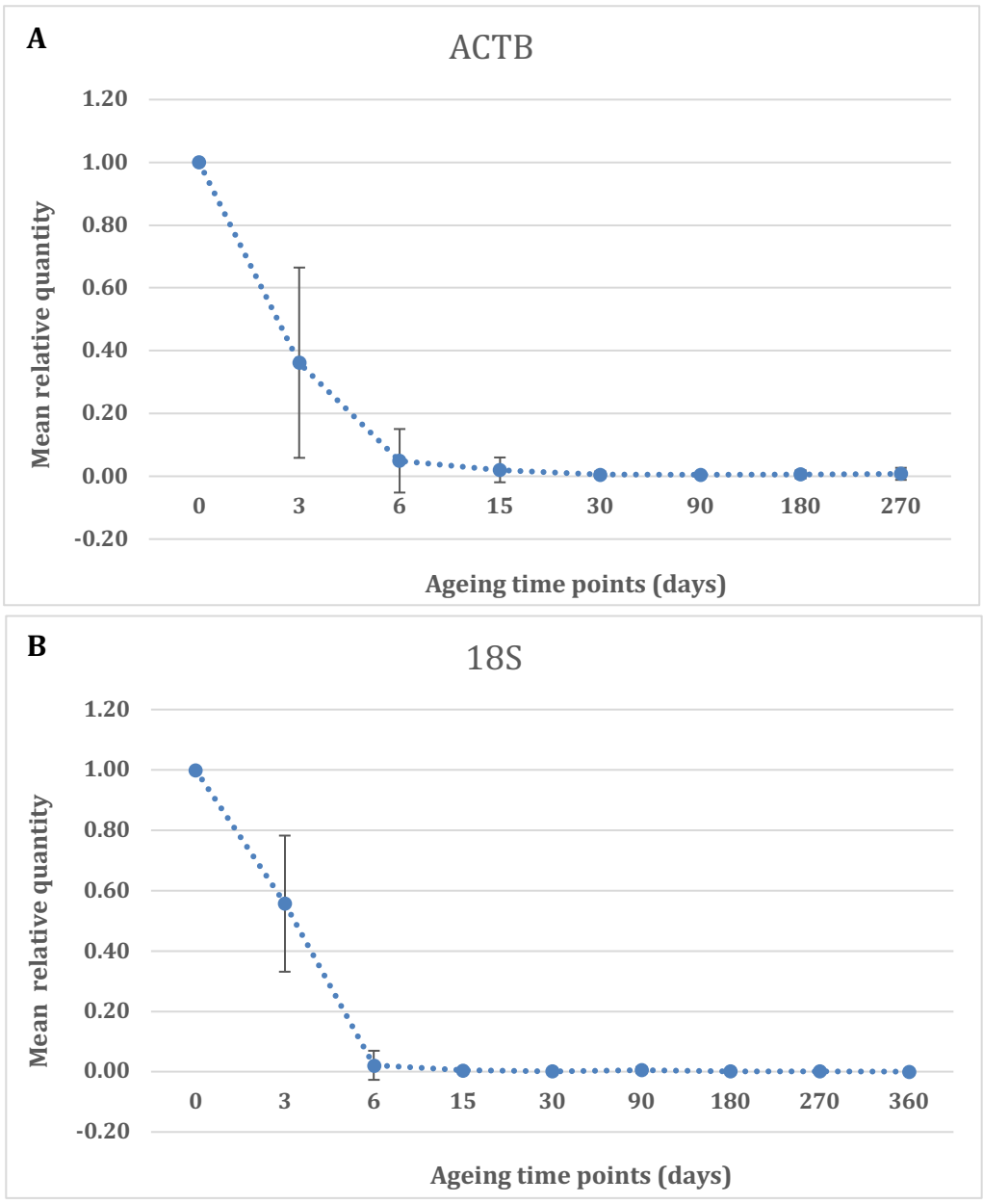


Figure 7.11: **Mean relative quantity of (A) *ACTB* and (B) *18S* in bloodstains stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. Error bars represent standard error.

The plots of both *ACTB* and *18S* show the same trend, and the relative quantity of both markers dropped dramatically in the first 6 days in blood samples. As can be seen in Figure 7.11, the pattern of relative quantity can be divided into two phases; a rapid degradation phase (first phase) followed by a plateau (second phase) where the degradation is slower in both plots.

A Spearman's correlation test was performed as an Anderson-Darling normality test showed that the data was not normally distributed ($p < 0.05$), and indicated that the mean relative quantity of both *ACTB* and *18S* showed a significant decrease with increasing age points ($r = -0.878$, $p = 0.004$, and $r = -0.803$, $p = 0.009$, respectively). A one-way ANOVA with Tukey's multiple comparison test could not be applied as the data was not normally distributed, therefore, Kruskal-Wallis test followed by Dunn's multiple comparisons test was performed instead in order to determine whether the measured relative quantity at each specific ageing point could be significantly discriminated from other ageing time points. For both reference genes, fresh blood samples (day 0) could be significantly distinguished from samples aged for 15 days and above ($p < 0.0001$), but not from younger samples (i.e. 3 and 6 days). Similarly, significant results were found when comparing the relative quantity of *ACTB* in samples aged for 3 days to samples aged for 15 days and above (all $p < 0.05$). In addition, the relative quantity of *18S* in samples aged for 3 days could be significantly discriminated from samples aged for 30 days and above ($p < 0.01$), with the exception of 90 days where $p = 0.08$. Finally, the relative quantity of *18S* in samples aged for 6 days was significantly discriminated from samples aged for 360 days ($p < 0.01$).

7.4.4.2 Saliva samples

In saliva samples (Figure 7.12), the relative quantity of both reference genes behaved differently when compared to each other. The degradation profile of *ACTB* could only be reported for the first 28 days, before the marker fell below the detection level of the assay. This shows only one phase of degradation with no plateau phase. In contrast, the relative quantity of *18S* decreased in the first 90 days before reaching a plateau. It also can be seen from the error bars in Figure 7.12 that there was a high variation between samples in the first 28 days for both markers.

A Pearson's correlation test was performed as an Anderson-Darling normality test showed that the data was normally distributed ($p > 0.05$) and indicated

that the mean relative quantity of both *ACTB* and *18S* showed a significant decrease with increasing age points ($r = 1$, $p < 0.0001$, and $r = -0.922$, $p = 0.001$, respectively). A one-way ANOVA with Tukey's comparison test was also performed, with the relative quantity of *ACTB* being significantly discriminated between fresh samples and samples aged for 14 and 28 days ($p = 0.012$, and $p < 0.0001$, respectively) but not from samples aged for 7 days ($p > 0.05$). However, samples aged for 7 days could be discriminated from those aged for 28 days ($p < 0.05$). In contrast, when comparing the relative quantity of *18S* at different time periods, none of the younger saliva samples (up to 28 days) could be discriminated, but fresh samples and those aged for 7 days could be distinguished from samples aged 90 days and above (all $p < 0.05$).

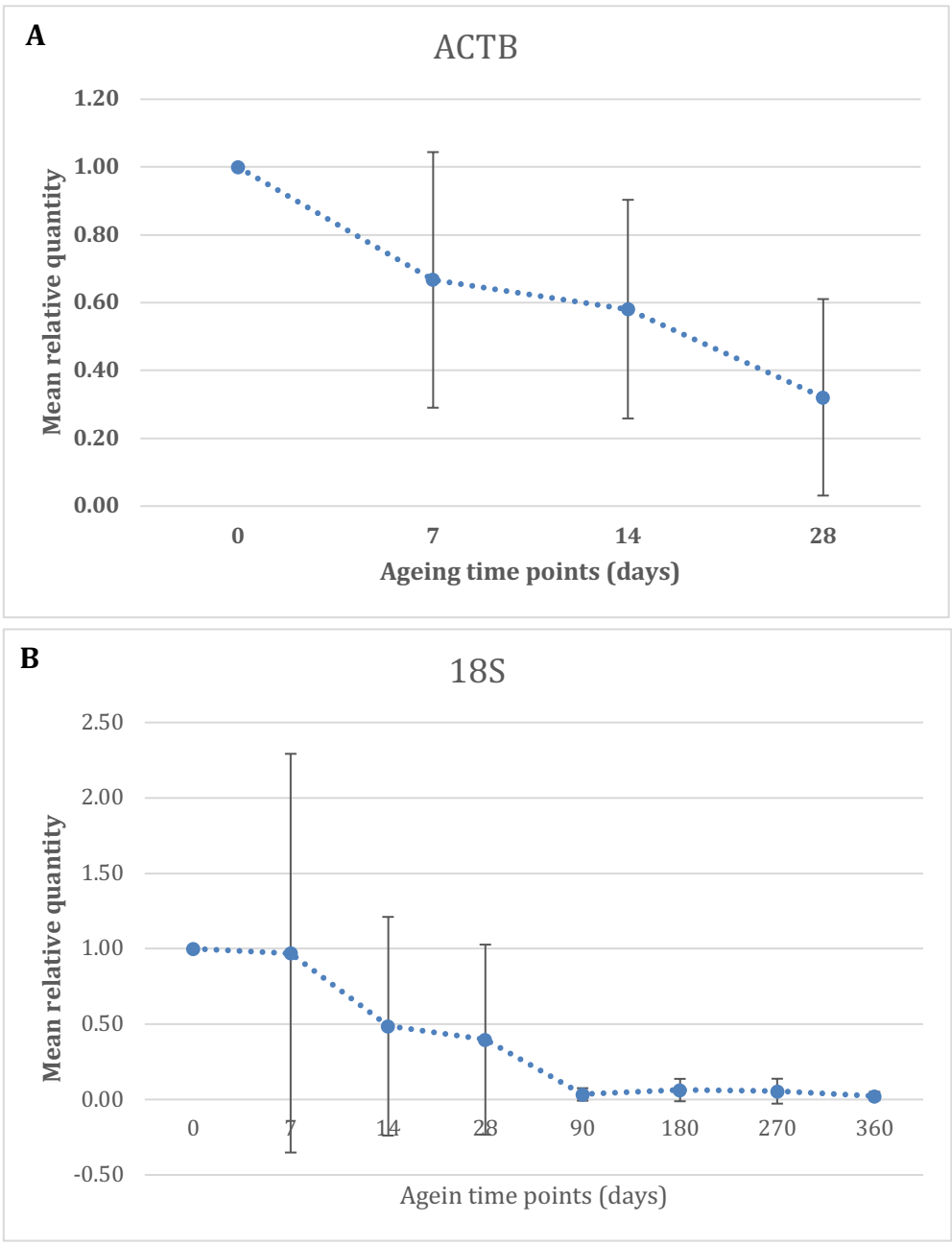


Figure 7.12: **Mean relative quantity of (A) *ACTB* and (B) *18S* in saliva samples stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 10. Error bars represent standard error.

7.4.4.3 Semen samples

The relative quantity of *ACTB* and *18S* over time was calculated in semen samples aged for up to 14 days, before the expression level of *ACTB* dropped below the detection level of the assay. Even though only three ageing points were recorded, it can still be seen (Figure 7.13) that there is a two-phase

decay, reaching a plateau after 7 days. The relative quantity of *18S* also exhibited a two-phase profile, degrading in the first 28 days and reaching a plateau at 90 days.

A Spearman's and Pearson's correlation tests were performed as Anderson-Darling normality test showed that the *18S* data was not normally distributed ($p = 0.01$) but the *ACTB* data was normally distributed ($p = 0.06$). These correlation tests indicated that only the mean relative quantity of *18S* showed a significant decrease with increasing age points ($r = -0.910$, $p = 0.002$), but this was not the case for *ACTB* ($r = -0.861$, $p = 0.339$), which was expected due to the limited ageing time points. A one-way ANOVA with Tukey's comparison test was performed for *ACTB* data, with the relative quantity of *ACTB* in fresh samples being significantly discriminated from samples aged for 7 and 14 days (both $p < 0.0001$). For *18S* data, a Kruskal-Wallis test followed by Dunn's multiple comparisons test was performed as the data was not normally distributed, and only fresh semen samples (day 0) could be significantly discriminated from samples aged for 28 days and above ($p > 0.01$), but could not be distinguished from younger samples (i.e. 7 and 14 days).

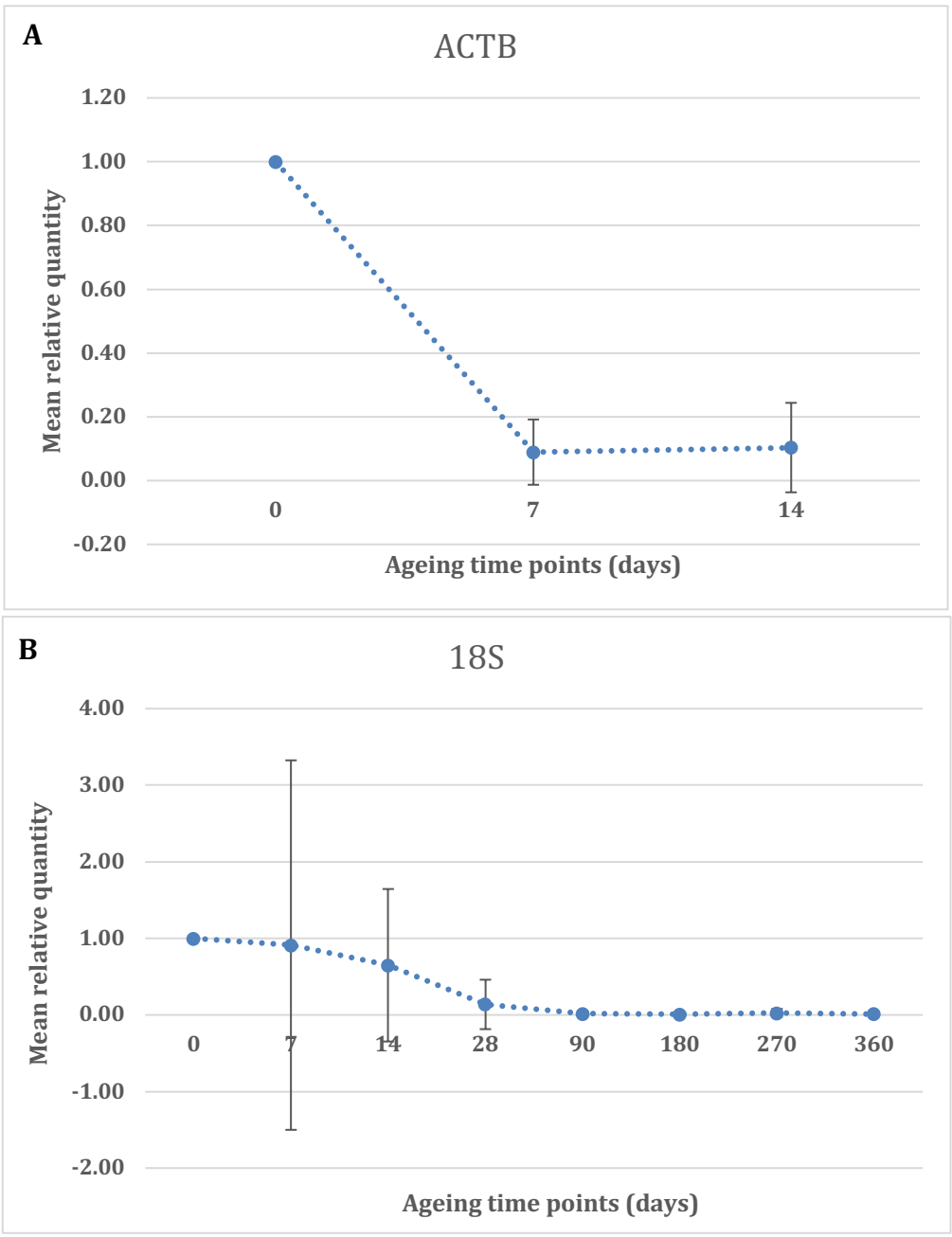


Figure 7.13: **Mean relative quantity of (A) *ACTB* and (B) *18S* in semen samples stored at room temperature for up to one year.** The data presented has been corrected for efficiency and each point represents the mean of n = 9. Error bars represent standard error.

7.5 Summary and Conclusion

Generally, in gene expression studies, reference genes (also known as house-keeping genes) are used to normalise the data for target genes, in order to overcome or minimise the effect of variation among samples and due to

experimental conditions. These types of genes are expected to be ubiquitously expressed and show high stability in the examined samples under different experimental or environmental conditions. This requirement is necessary because it increases the reliability and validity of RT-qPCR data, as it normalises the levels of RNA markers between different samples (i.e. the expression profile of the target gene relative to the reference gene). Therefore, it is crucial to select a suitable reference gene or a group of reference genes for data analysis.

ACTB and *18S* are among the most commonly used reference genes, therefore they were selected here to investigate whether they are suitable for use in forensic sample types. The findings of previous chapters indicate that different RNA markers degrade at different rates and frequently show unique degradation behaviour in different body fluid types. In this Chapter, both *ACTB* and *18S* also showed unique degradation profiles in three different body fluid types (blood, saliva and semen), with abundance being lowest and degradation being highest for both reference genes in semen samples. Similar data were found by Simard et al. (2012) [88] and Sakurada et al. (2013) [208], who studied the stability of a number of reference genes including *ACTB* and *18S* in blood, saliva and semen samples. Sakurada et al. (2013) also showed that *ACTB* had similar stabilities to body fluid-specific genes.

In all three different cell types, *18S* had higher expression than *ACTB*, with the highest being recorded in blood samples. The *ACTB* expression profiles varied in different body fluid types, consistent with the data reported by Lindenbergh et al. (2013) [207], and dropped below the detection threshold of the assay after 14 days and 28 days in semen and saliva samples respectively, and after 270 days in blood samples. On the other hand, *18S* was still detected even after 360 days of storage in all three different samples types. With this level of variation in stability and expression levels in different sample types, it can be concluded that neither of the examined genes should be used as reference genes for forensic purposes.

The RER of *ACTB* to *18S* was also examined to determine whether it is suitable for use in predicting the age of biological stains, as has previously been suggested. Anderson et al. (2005) showed that *18S* was more stable than *ACTB* in blood samples aged for 150 days, proposing that the difference in degradation rate is due to structural differences between the two markers, as *18S* is protected by a protein complex (ribosomal RNA) [58]. In this chapter, *18S* remained stable only in the first 3 days of ageing in blood samples. Since *18S* had low stability in the samples examined here, the positively correlation relationship between RER and ageing time points could not be replicated, showing on the other hand a significant negative correlation between *ACTB/18S* and the age of bloodstains. In saliva and semen samples, the relationship between RER and ageing time points could not be established. However, when calculating the relative quantity over time ($2^{-\Delta Cq}$) for both *ACTB* and *18S*, significant relationships were found with ageing time points, allowing the data to be divided into groups based on ageing time points, such that fresh samples could be discriminated from older samples. Therefore, the relative quantity over time ($2^{-\Delta Cq}$) measure of reference genes could be a potential method for estimating the age of body fluid stains.

In future, the degradation behaviour of a different set of reference genes should be investigated in dried biological samples in order to determine the most suitable reference genes for RT-qPCR analysis. In addition, the effects of sex, age and lifestyle should be further investigated in order to understand inter-donor variation, which was shown to be high in this work, as well as increasing samples size. Furthermore, other techniques such as RNA sequencing could be used for more precise measurement of RNA degradation across multiple ageing points, in order to determine whether there is a relationship between the degradation rate of RNA markers and the age of biological stains.

Chapter eight: General Discussion and Conclusion

The overall purpose of this project was the application of RNA analysis for estimating the age of body fluid stains. This type of information can be crucial in a criminal investigation as it can provide information regarding when a crime occurred or help to exclude samples that do not correspond to the time of the crime. Body fluid-specific RNA markers were selected in this work to determine the age of biological stains, allowing forensic experts to identify the type of body fluid and the time since deposition simultaneously.

RNA specific markers for blood, saliva and semen were selected to be investigated in this project, as these body fluid stains are commonly encountered in forensic casework. The degradation profiles of these selected RNA markers were analysed, including mRNA, miRNA markers and reference genes. RNA degradation profiles might differ in different cell types, therefore, in this thesis, the aim was to provide a more comprehensive study of RNA degradation rate in dried body fluids stains. Multiple RNA markers for each body fluid were examined and their expression profiles were analysed across one year at multiple ageing time points. The data were then analysed to look for a correlation between the age of a body fluid stain and the degradation rate of the RNA markers, in order to identify those most useful for body fluid stain ageing, and to assess the possibility of using the degradation rate to determine time since deposition.

Initially, the quality and quantity of the isolated RNA are very crucial factors in determining the accuracy of RNA profiling studies. Therefore, in Chapter three two RNA extraction methods were evaluated to determine which would perform most favourably with regards to RNA yield and purity, when conducted to extract RNA from a small amount of body fluids. The purpose of using small volumes was to mimic the difficult nature of biological samples that are commonly found at crime scenes, in small quantities and often degraded states. The TRI Reagent® method and the miRNeasy Mini Kit were utilised for

total RNA extraction, followed by the use of NanoDrop-1000 to measure the concentration and the purity of the extracted RNA, and the outcomes of both methods were compared. TRI Reagent® was shown to improve both the RNA yield and purity from blood, saliva and semen samples relative to the miRNeasy Mini Kit. Accordingly, TRI Reagent® was utilised throughout Chapters 4, 5, 6 and 7.

Furthermore, two approaches were used to measure the quantity and the quality of the extracted RNA from blood samples: UV-visible spectrophotometry (NanoDrop-1000) and the Agilent 2100 Bioanalyzer, and both methods suffered from a number of limitations. The UV-visible spectrophotometry (NanoDrop-1000) measured the concentration of the total RNA regardless of whether it was intact or fragmented, so it cannot be used to indicate the degradation state of the extracted samples. A similar conclusion was drawn to the Agilent 2100 Bioanalyzer approach as it did not give a reliable indication about the degradation level in the isolated RNA samples, and showed high levels of variation among bloodstains leading to lack of accuracy and low precision. The high level of variation could be due to the relatively small sample size. Nevertheless, when comparing the obtained RIN values with RT-qPCR data, there was no correlation found. The C_q values of some samples have shown not to be affected by the RIN values, as samples with low RIN (i.e. degraded samples) still gave low C_q values (higher quantity). Therefore, it is not recommended to use the Agilent 2100 Bioanalyzer as a method to quantify total RNA or to check RNA quality for forensic samples types, which will generally be very low in volume and highly degraded.

Additionally, an assessment of the TaqMan® assays to be used in the current project was conducted, as it is essential for quality control purposes and to ensure adherence to the MIQE guidelines [167], in order to enhance the ability to publish any of the work presented here. All examined assays generated efficiencies ranging between 90.0% and 108.10%, which is within the accepted range (90%-110%) for good quality qPCR data. Moreover, when assessing the

precision of the data by testing intra-assay and inter-assay variation for reproducibility, all tested TaqMan® assays proved to have highly reproducible data. As such, the TaqMan® assays selected demonstrated a reliable and sensitive technology to quantify gene expression in body fluid samples (blood, saliva and semen).

When selecting RNA markers to be used for estimating the age of biological stains, the degradation rate of these markers needs to be large enough that it is observable across ageing time periods. In this work, the degradation rate of blood-specific RNA markers in aged samples was analysed (Chapter four). A total of ten volunteers (6 females and 4 males) donated blood samples, which were deposited on cotton swabs and then stored in a dark, dry place at room temperature to simulate natural ageing, until they reached a series of desired ages (0, 3, 6, 15, 30, 90, 180, 270, 360 days). The degradation levels of six RNA molecules (*HBA*, *HBB*, *HMBS*, *miR16*, *miR451* and *U6*) were analysed using TaqMan® assays and the relative expression ratios (RERs) were calculated to study the degradation behaviour of the markers, and their relationship with age. The outcomes of this research showed that different RNA molecules degrade at different rates in bloodstains, with miRNA markers exhibiting strong stability, likely due to their small size. The correlation tests and regression analysis indicated that the RERs of blood-specific markers have a positive correlation with ageing time points, and may be useful in estimating the age of bloodstains. Regression analysis examining the relationship between the RERs of *miR16*/*U6* and *HBA*/*HBB* with ageing time point produced the highest R^2 values (98.6% and 97.9% respectively) with narrow confidence intervals and low standard error (i.e. S value), using a non-linear model (i.e. second order polynomial). These results indicated that these two ratios may be the most reliable in predicting the age of bloodstains.

Therefore, RERs of blood-specific markers represent a potential method to estimate the age of bloodstains. The findings of this study, therefore, emphasise that, in future, methods using RT-qPCR are likely to be more

sensitive for the accurate determination of the age of bloodstains than the analysis of protein degradation or variation in solubility and morphological differences in bloodstains [88].

Following that, RNA markers that are specific for saliva and semen samples were also examined in an attempt to determine the time since deposition of these sample types (Chapter five). A total of 19 volunteers were asked to give saliva or semen samples, which were deposited on cotton swabs and then stored in a dark, dry place at room temperature to simulate natural ageing, until they reached a series of desired ages (0, 7, 14, 28, 90, 180, 270, 360 days). The degradation levels of two saliva-specific (*STATH* and *miR205*) and five semen-specific (*PRM1*, *PRM2*, *SEMG1*, *miR10b* and *miR891a*) RNA markers were analysed using TaqMan® assays, and the RERs were calculated to study the degradation behaviour of the markers, and their relationship with age. The findings of this section showed that different RNA molecules degrade at different rates in saliva and semen samples, with miRNA markers exhibiting strong stability (similar to blood samples). By applying correlation tests and regression analysis, the data indicated that the RERs of saliva and semen-specific markers have a significant relationship with ageing time points. In saliva-specific markers, regression analysis of the relationship between the RER of *STATH/miR205* with ageing time period produced a high R^2 value (97.2%) using a non-linear model. These results indicated that this ratio could be reliable in the estimation of the age of saliva samples. With regards to semen-specific markers, the RERs of semen-specific markers (*PRM2/miR10b, PRM2/miR891a, SEMG1/miR10b, SEMG1/miR891a, PRM2/U6, SEMG1/U and miR10b/U6*) also showed a positive correlation with ageing time points, with regression analyses giving R^2 values of more than 96% using non-linear models. These results confirmed that these RERs could be a potential method to estimate the age of semen stains.

In the examined body fluids (blood, saliva and semen), the difference between the two types of model examined (linear and non-linear) is very clear, with the

non-linear models having substantially higher R^2 values. The higher prediction abilities of these models therefore make them a better choice for the estimation of the age of saliva and semen stains. However, to determine how generally applicable this is across body fluid types, samples collected from a crime scene should be studied, including mixture samples such as semen mixed with menstrual blood or vaginal secretions. This study only examined samples that were stored under controlled conditions, therefore, additional environmental factors that might affect the RERs of body fluid-specific markers should be explored as future projects, such as UV exposure, humidity and high temperature.

The RERs of saliva and semen-specific markers represent a potential method to estimate the time since deposition of saliva and semen stains. The findings of this study therefore emphasise that, in future, methods using RT-qPCR are likely to be a sensitive technology for the accurate determination of the age of saliva and semen samples.

When comparing the findings for blood, saliva and semen-specific markers in Chapters four and five respectively, it was found that blood-specific markers were present in higher quantity (lower C_q values) than saliva- and semen-specific markers at most ageing time points, even though only 20 µL of blood samples was used while 50 µL was used for saliva and semen, with the exception of the *HMBS* marker, which showed C_q values above 34. Similar findings were also found when comparing the quantity of *U6* among different body fluid types, as it was detected in higher quantities in blood samples across all ageing time points when compared to saliva and semen samples. In addition, in all three body fluid samples that were examined, it was observed that miRNA markers were present in greater quantities than mRNA markers. All the selected miRNAs for each body fluid were still detected after one year of ageing, while in contrast, some mRNAs showed no C_q values (i.e. they were not detected) after 90 days of ageing (*PRM1* and *SEMG1* in semen) or as early as 15 days of ageing, which was the case with *HMBS* in blood samples.

In terms of inter-donor variation, high coefficients of variation (CV%) were recorded among individuals in all sample types, which could be due to the examined body fluids being heterogeneous, such as saliva samples which contain different composition of ribonucleases, epithelial cells and large number of bacteria. The overall health of the individual donors, as well as their age, diet and lifestyle could also lead to high inter-donor variation. The sample size used in these experiments is relatively small (i.e. 10 blood samples, 10 saliva samples and 9 semen sample), which also could lead to high variation, along with technical variation, which could be addressed by increasing sample size and the number of technical repeats.

Chapter six in this thesis examined the impact of body fluid mixtures on the obtained RERs. Biological samples that are found at crime scenes are commonly found as mixtures, where two or more body fluids are mixed together, either from the same individual or multiple individuals. Therefore, in order to develop such a method to be considered as a successful approach to estimating the age of biological stains in forensic casework samples, it is important that the impact of body fluid mixtures on RER is evaluated. The published findings reported in Chapter four [192] showed that the RERs of blood-specific markers can be considered as a potential method to estimate the age of bloodstains. However, it is important to investigate whether there is any limitation to this proposed method for estimating the time since deposition as a result of body fluids being present in mixed stains. Chapter six therefore evaluated the effect of mixing bloodstains with other body fluid types (saliva and semen) on the RER values of various blood-specific markers, and hence the effect on the estimation of bloodstain age or time since deposition.

Pure bloodstains and bloodstains mixed with saliva or semen were stored for up to 60 days under controlled conditions (i.e. room temperature, in a dark, dry place), and at each desired ageing time point (0, 10, 30 and 60 days), total RNA was extracted and RNA analysis was performed. The expression level of multiple blood-specific markers (*HBA, HBB* and *miR16*) along with two

reference genes (*18S* and *U6*) was quantified by RT-qPCR in all different sample types (i.e. pure and mixed). Blood-specific markers showed low or no expression in the other body fluids (i.e. saliva and semen), confirming that the selected markers are truly blood-specific, and therefore could be used to identify the presence of blood in a mixed stain. These findings are in accordance with the EDNAP collaborative exercise [65] and other studies [64, 194-196] that have investigated the expression of blood-specific markers in different body fluids, where no expression of *HBA* and *HBB* has been found in saliva and semen samples.

The RT-qPCR data for the blood-specific markers demonstrates that each RNA transcript showed a unique pattern of degradation behaviour in pure and mixed bloodstains. *HBA* was the only blood-specific marker which exhibited no significant differences in degradation behaviour between all samples types, while the degradation rate of *miR16* was significantly different between pure bloodstains and bloodstains mixed with semen (but not saliva), and *HBB* degradation was significantly different in all samples types. The obtained pattern indicated that the presence of saliva or semen in bloodstains affects the quantity of blood-specific mRNA and miRNA markers.

The RERs of different combinations of blood-specific markers were calculated for each sample type against the ageing time points: *HBA/miR16*, *HBA/U6*, *miR16/U6,* and *HBA/HBB*. Mixing bloodstains with other types of body fluids such as saliva or semen did not have an effect on the RERs of blood-specific markers, as no significant differences were identified when comparing these RERs in pure bloodstains and bloodstains mixed with either saliva or semen. The only exception was the *HBA/HBB* ratio, which was significantly different between pure bloodstains and bloodstains mixed with semen (but not saliva), due to the different degradation behaviour of *HBB* in the mixed samples.

The findings presented in Chapter six therefore indicate that the presence of body fluid mixtures did not have an impact on the majority of RERs for blood-

specific RNA markers, suggesting that calculation of these ratios is a promising method for estimating the age of bloodstains. However, to determine how applicable this is across different body fluids, more mixtures should be studied, including sample types such as menstrual blood, vaginal secretions, urine and sweat, as well as mixtures of more than two body fluids. Another potential avenue of research would be to examine different volume ratios of body fluid mixtures, as this study only considered equal volume mixtures of blood with other body fluids. These are all important avenues for future research.

The final Chapter of this thesis (Chapter seven) assessed the expression level and degradation rate of two RNA markers that are commonly used as reference genes (*18S* and *ACTB*). Generally, in gene expression studies, reference genes (also known as house-keeping genes) are used to normalise the data for target genes, in order to overcome or minimise the effect of variation among samples and due to experimental conditions. These types of genes are expected to be ubiquitously expressed and show high stability in the examined samples under different experimental or environmental conditions. This requirement increases the reliability and validity of RT-qPCR data, as it normalises the levels of RNA markers between different samples (i.e. the expression profile of the target gene relative to the reference gene). Therefore, it is crucial to select a suitable reference gene or a group of reference genes for data analysis.

ACTB and *18S* are among the most commonly used reference genes, therefore they were selected here to investigate whether they are suitable for use in forensic sample types. The findings of previous Chapters indicated that different RNA markers degrade at different rates and frequently show unique degradation behaviour in different body fluid types. In Chapter seven, both *ACTB* and *18S* also showed unique degradation profiles in three different body fluid types (blood, saliva and semen), with abundance being lowest and degradation being highest for both reference genes in semen samples. Similar data were found by Simard et al. (2012) [88] and Sakurada et al. (2013) [208],

who studied the stability of a number of reference genes including *ACTB* and *18S* in blood, saliva and semen samples. Sakurada et al. (2013) have also shown that *ACTB* had similar stabilities to body fluid-specific genes.

In all three different body fluids, *18S* had higher expression than *ACTB*, with the highest being recorded in blood samples. The *ACTB* expression profiles varied in different body fluid types, consistent with the data reported by Lindenbergh et al. (2013) [207], and dropped below the detection threshold of the assay after 14 days and 28 days in semen and saliva samples respectively, and after 270 days in blood samples. On the other hand, *18S* was still detected even after 360 days of storage in all three different samples types. With this level of variation in stability and expression levels in different sample types, it can be concluded that neither of the examined genes should be used as reference genes for forensic purposes.

The RER of *ACTB* to *18S* was also examined to determine whether it is suitable for use in predicting the age of biological stains, as has previously been suggested. Anderson et al. (2005) showed that *18S* was more stable than *ACTB* in blood samples aged for 150 days, proposing that the difference in degradation rate is due to structural differences between the two markers, as *18S* is protected by a protein complex (ribosomal RNA) [58]. In Chapter seven, *18S* remained stable only in the first 3 days of ageing in blood samples. Since *18S* did had low stability in the samples examined here, the positive correlation relationship between RER and ageing time points could not be replicated, showing on the other hand a significant negative correlation between *ACTB/18S* and the age of bloodstains. In saliva and semen samples, the relationship between RER and ageing time points could not be established. However, when calculating the relative quantity over time ($2^{-\Delta Cq}$) for both *ACTB* and *18S*, significant relationships were found with ageing time points, allowing the data to be divided into groups based on ageing time points, such that fresh samples could be discriminated from older samples.

For future work, the degradation behaviour of a different set of reference genes should be investigated in dried biological samples in order to determine the most suitable reference genes for RT-qPCR analysis. In addition, the effects of sex, age and lifestyle should be further investigated in order to understand inter-donor variation, which was shown to be high in this work. Furthermore, other techniques such as RNA sequencing could effectively be used to screen large number of RNA markers simultaneously. Such large-scale screening technique can identify RNA markers with degradation profile that is correlated with the age of the stain.

Even so, results from these investigations do suggest that there is a relationship between *ex vivo* sample age and the degradation state of RNA markers demonstrating that there is a global decrease in mRNA abundance in ageing stains. In the context of forensic science laboratories, these techniques require much more research before consideration could be given to applying them to forensic casework. Through further detailed analysis of transcript behaviour in *ex vivo* body fluid stains, it may be possible to better understand the RNA degradation process and apply that understanding to the development of a reliable method for measuring sample age or the postmortem interval. Future studies could examine the effects of different storage conditions (high temperature, exposure to sunlight, humidity, etc.) to mimic the harsh conditions of crime scenes. Additionally, new guidelines are needed to be set to overcome the limitations in applying such techniques in forensic casework samples, which could improve the reproducibility of the results among different researchers.

REFERENCES

[1]Watson J, Crick F. Molecular Structure of Nucliec Acids. Nature. 1953;171(4356):737-38.
[2]International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. Nature. 2004;431:931-45.
[3]Farrell RE. RNA Methdologies: A Laboratory Guide for Isolation and Characterization. London: Elsevier; 2010.
[4]Elliott D, Ladomery M. The Molecular Biology of RNA. 1st ed: New York: Oxford University Press; 2011.
[5]Sharp PA. The Centrality of RNA. Cell. 2009;136(4):577-80.
[6]Dale JW, Schantz MV, Plant N. From Genes to Genomes. 3rd ed: Wiley-Blackwell; 2012.
[7]Miko I, LeJeune L. Essentials of Genetics. Cambridge: MA: NPG Education; 2009.
[8]Lodish H, Berk A, Zipursky S. Molecular Cell Biology. 4th ed. New York: W. H. Freeman; 2000.
[9]Cooper GM. The Cell. 2nd ed. Sunderlan (MA): Sinauer Associates; 2000.
[10]Cowling VH. Regulation of mRNA Cap Methylation. Biochem J. 2010;425(2):295-302.
[11]Wahle E, Keller W. The Biochemistry of Polyadenylation. Trends Biochem. Sci. 1996;21:247-50.
[12]Keller W. No End yet to Messenger RNA 3' Processing. Cell. 1995;81:829-32.
[13]Colgan DF, Manley JL. Mechanism and Regulation of mRNA Polyadenlation. Genes & Development. 1997;11:2755-66.
[14]Lewis JD, Gunderson SI, Mattaj lW. The Influence of 5' and 3' End Structures on Pre-mRNA Metabolism. Journal of Cell Science. 1995;19:13-19.
[15]Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4th ed. New York: Garland Science; 2002.
[16]Kohler A, Hurt E. Exporting RNA from the Nucleus to the Cytoplasm. Nat Rev Mol Cell Biol. 2007;8(10):761-73.
[17]JM B, JL T, L S. Biochemistry. 5th ed. New York: W H Freeman; 2002.
[18]Lewin B, Krebs JE, Goldstein ES, Kilpatrick ST. Lewin's Essential Genes. 2nd ed. Burlington MA: Jones and Bartlett Publishers; 2010.
[19]Kim VN. Small RNAs: Classification, Biogenesis, and Function. Molecules and Cells. 2005;19(1):1-15.
[20]Lee Y, Jeon K, Lee J-T, Kim S, Kim VN. MicroRNA Maturation: Stepwise Processing and Subcelluar Licalization. The EMBO Journal. 2002;21(17):4663-70.

[21]Pillai RS. MicroRNA Function: Multiple Mechanisms for a Tiny RNA? RNA. 2005;11(12):1753-61.
[22]Lynch J, Wang JY. MicroRNA Regulation of Cancer Stem Cell Phenotypes. SOJ Genetic Science. 2014;1(1):1-8.
[23]Hammond SM. An Overview of microRNAs. Adv Drug Deliv Rev. 2015;87:3-14.
[24]Courts C, Madea B. Micro-RNA - a Potential for Forensic Science? Forensic Sci Int. 2010;203(1-3):106-11.
[25]MacFarlane L-A, Murphy PR. MicroRNA: Biogenesis, Function and Role in Cancer. Current Genomics. 2010;11(7):537-61.
[26]Pritchard CC, Cheng HH, Tewari M. MicroRNA Profiling: Approaches and Considerations. Nat Rev Genet. 2012;13(5):358-69.
[27]Hutvagner G, Zamore PD. A microRNA in a Multiple-Turnover Rnai Enzyme Complex. Science. 2002;297(5589):2056-60.
[28]Cannell IG, Kong YW, Bushell M. How Do microRNAs Regulate Gene Expression? Biochem Soc Trans. 2008;36(Pt 6):1224-31.
[29]Agrawal N, Dasaradhi PVN, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK. RNA Interference: Biology, Mechanism, and Applications. Microbiology and Molecular Biology Reviews. 2003;67(4):657-85.
[30]Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21±Nucleotide RNAs Mediate RNA Interference in Cultured Mammalian Cells. Nature. 2001;411:494-98.
[31]Mattick JS, Makunin IV. Non-Coding RNA. Hum Mol Genet. 2006;15(Review Issue 1):R17-29.
[32]Brown TA. Genomes. 2nd ed. Oxford: Wiley-Liss; 2002.
[33]Tourriere H, Chebli K, Tazi J. mRNA Degradation Machines in Eukaryotic Cells. Biochimie. 2002;84:821-37.
[34]Stein GS, Lian JB. Molecular and Cellualr Approaches to the Control of Proliferation and Differentiation. San Diego: Academic Press; 1992.
[35]Vennemann M, Koppelkamm A. mRNA Profiling in Forensic Genetics I: Possibilities and Limitations. Forensic Sci Int. 2010;203(1-3):71-5.
[36]Guhaniyogi J, Brewer G. Regulation of mRNA Stability in Mammalian Cells. Gene. 2001;265:11-23.
[37]Garneau NL, Wilusz J, Wilusz CJ. The Highways and Byways of mRNA Decay. Nat Rev Mol Cell Biol. 2007;8(2):113-26.
[38]Wilusz CJ, Wormington M, Peltz SW. The Cap-to-Tail Guide to mRNA Turnover. Nature Reviews Molecular Cell Biology. 2001;2:237-46.
[39]Visone R, Croce CM. MiRNAs and Cancer. Am J Pathol. 2009;174(4):1131-8.

[40]Pan ZW, Lu YJ, Yang BF. Micrornas: A Novel Class of Potential Therapeutic Targets for Cardiovascular Diseases. Acta Pharmacol Sin. 2010;31(1):1-9.
[41]Zhang Z, Qin YW, Brewer G, Jing Q. MicroRNA Degradation and Turnover: Regulating the Regulators. Wiley Interdiscip Rev RNA. 2012;3(4):593-600.
[42]Guo Y, Liu J, Elfenbein SJ, Ma Y, Zhong M, Qiu C, Ding Y, Lu J. Characterization of the Mammalian miRNA Turnover Landscape. Nucleic Acids Res. 2015;43(4):2326-41.
[43]Ruegger S, Grosshans H. MicroRNA Turnover: When, How, and Why. Trends Biochem Sci. 2012;37(10):436-46.
[44]Ramachandran V, Chen X. Degradation of microRNAs by a Family of Exoribonucleases in Arabidopsis. Science. 2008;321:1490-92.
[45]Li Y, Li Z, Zhou S, Wen J, Geng B, Yang J, Cui Q. Genome-Wide Analysis of Human microRNA Stability. Biomed Res Int. 2013;2013:368975.
[46]Yen H-CS, Xu Q, Chou DM, Zhao Z, Elledge SJ. Global Protein Stability Profiling in Mammalian Cells. Science. 2008;322:918-23.
[47]Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. Genome-Wide Analysis of Long Noncoding RNA Stability. Genome Res. 2012;22(5):885-98.
[48]Marzi MJ, Ghini F, Cerruti B, Pretis Sd, Bonetti P, Giacomelli C, Gorski MM, Kress T, Pelizzola M, Muller H, Amati B, Nicassio1 F. Degradation Dynamics of microRNAs Revealed by a Novel Pulse-Chase Approach. Genome Research. 2016;26:554-65.
[49]de la Mata M, Gaidatzis D, Vitanescu M, Stadler MB, Wentzel C, Scheiffele P, Filipowicz W, Grosshans H. Potent Degradation of Neuronal miRNAs Induced by Highly Complementary Targets. EMBO Rep. 2015;16(4):500-11.
[50]Ghini F, Rubolino C, Climent M, Simeone I, Marzi MJ, Nicassio F. Endogenous Transcripts Control miRNA Levels and Activity in Mammalian Cells by Target-Directed miRNA Degradation. Nat Commun. 2018;9(1):3119.
[51]Liu J, Lewohl JM, Dodd PR, Randall PK, Harris RA, Mayfield RD. Gene Expression Profiling of Individual Cases Reveals Consistent Transcriptional Changes in Alcoholic Human Brain. Journal of neurochemistry. 2004;90(5):1050-58.
[52]Maeda H, Fujita MQ, Zhu B-L, Ishida K, Quan L, Oritani S, Taniguchi M. Pulmonary Surfactant-Associated Protein a as a Marker of Respiratory Distress in Forensic Pathology: Assessment of the Immunohistochemical and Biochemical Findings. Legal Medicine. 2003;5:S318-S21.
[53]Bauer M. RNA in Forensic Science. Forensic Sci Int Genet. 2007;1(1):69-74.

[54]Zubakov D, Kokshoorn M, Kloosterman A, Kayser M. New Markers for Old Stains: Stable mRNA Markers for Blood and Saliva Identification from up to 16-Year-Old Stains. Int J Legal Med. 2009;123(1):71-4.
[55]Ingold S, Dorum G, Hanson E, Berti A, Branicki W, Brito P, Elsmore P, Gettings KB, Giangasparo F, Gross TE, Hansen S, Hanssen EN, Kampmann ML, Kayser M, Laurent FX, Morling N, Mosquera-Miguel A, Parson W, Phillips C, Porto MJ, Pospiech E, Roeder AD, Schneider PM, Schulze Johann K, Steffen CR, Syndercombe-Court D, Trautmann M, van den Berge M, van der Gaag KJ, Vannier J, Verdoliva V, Vidaki A, Xavier C, Ballantyne J, Haas C. Body Fluid Identification Using a Targeted mRNA Massively Parallel Sequencing Approach - Results of a Euroforgen/Ednap Collaborative Exercise. Forensic Sci Int Genet. 2018;34:105-15.
[56]Dorum G, Ingold S, Hanson E, Ballantyne J, Snipen L, Haas C. Predicting the Origin of Stains from Next Generation Sequencing mRNA Data. Forensic Sci Int Genet. 2018;34:37-48.
[57]Hanson E, Ingold S, Haas C, Ballantyne J. Messenger RNA Biomarker Signatures for Forensic Body Fluid Identification Revealed by Targeted RNA Sequencing. Forensic Sci Int Genet. 2018;34:206-21.
[58]Anderson S, Howard B, Hobbs GR, Bishop CP. A Method for Determining the Age of a Bloodstain. Forensic Sci Int. 2005;148(1):37-45.
[59]Anderson SE, Hobbs GR, Bishop CP. Multivariate Analysis for Estimating the Age of a Bloodstain. J Forensic Sci. 2011;56(1):186-93.
[60]Hampson C, Louhelainen J, McColl S. An RNA Expression Method for Aging Forensic Hair Samples. J Forensic Sci. 2011;56(2):359-65.
[61]Bauer M, Gramlich I, Polzin S, Patzelt D. Quantification of mRNA Degradation as Possible Indicator of Postmortem Interval—a Pilot Study. Legal Medicine. 2003;5(4):220-27.
[62]Oehmichen M, Zilles. K. Postmortem DNA and RNA Synthesis. Preliminary Studies in Human Cadavers. Journal of legal medicine. 1983;91.4:287-94.
[63]An J-H, Shin K-J, Yang W-I, Lee H-Y. Body Fluid Identification in Forensics. BMB reports. 2012;45(10):545-53.
[64]Haas C, Klesser B, Maake C, Bar W, Kratzer A. mRNA Profiling for Body Fluid Identification by Reverse Transcription Endpoint PCR and Realtime PCR. Forensic Sci Int Genet. 2009;3(2):80-8.
[65]Haas C, Hanson E, Anjos MJ, Bar W, Banemann R, Berti A, Borges E, Bouakaze C, Carracedo A, Carvalho M, Castella V, Choma A, De Cock G, Dotsch M, Hoff-Olsen P, Johansen P, Kohlmeier F, Lindenbergh PA, Ludes B, Maronas O, Moore D, Morerod ML, Morling N, Niederstatter H, Noel F, Parson W, Patel G, Popielarz C, Salata E, Schneider PM, Sijen T, Sviezena B, Turanska M, Zatkalikova L, Ballantyne

J. RNA/DNA Co-Analysis from Blood Stains–Results of a Second Collaborative Ednap Exercise. Forensic Sci Int Genet. 2012;6(1):70-80.
[66]Hanson EK, Lubenow H, Ballantyne J. Identification of Forensically Relevant Body Fluids Using a Panel of Differentially Expressed microRNAs. Anal Biochem. 2009;387(2):303-14.
[67]Haas C, Hanson E, Anjos MJ, Banemann R, Berti A, Borges E, Carracedo A, Carvalho M, Courts C, De Cock G, Dotsch M, Flynn S, Gomes I, Hollard C, Hjort B, Hoff-Olsen P, Hribikova K, Lindenbergh A, Ludes B, Maronas O, McCallum N, Moore D, Morling N, Niederstatter H, Noel F, Parson W, Popielarz C, Rapone C, Roeder AD, Ruiz Y, Sauer E, Schneider PM, Sijen T, Court DS, Sviezena B, Turanska M, Vidaki A, Zatkalikova L, Ballantyne J. RNA/DNA Co-Analysis from Human Saliva and Semen Stains – Results of a Third Collaborative Ednap Exercise. Forensic Sci Int Genet. 2013;7(2):230-9.
[68]Nussbaumer C, Gharehbaghi-Schnell E, Korschineck I. Messenger RNA Profiling: A Novel Method for Body Fluid Identification by Real-Time PCR. Forensic Sci Int. 2006;157(2-3):181-6.
[69]Zubakov D, Hanekamp E, Kokshoorn M, van Ijcken W, Kayser M. Stable RNA Markers for Identification of Blood and Saliva Stains Revealed from Whole Genome Expression Analysis of Time-Wise Degraded Samples. Int J Legal Med. 2008;122(2):135-42.
[70]Setzer M, Juusola J, Ballantyne J. Recovery and Stability of RNA in Vaginal Swabs and Blood, Semen, and Saliva Stains. J Forensic Sci. 2008;53(2):296-305.
[71]Juusola J, Ballantyne J. Multiplex mRNA Profiling for the Identification of Body Fluids. Forensic Sci Int. 2005;152(1):1-12.
[72]Juusola J, Ballantyne J. mRNA Profiling for Body Fluid Identification by Multiplex Quantitative RT-PCR. J Forensic Sci. 2007;52(6):1252-62.
[73]Bauer M, Patzelt D. Identification of Menstrual Blood by Real Time RT-PCR: Technical Improvements and the Practical Value of Negative Test Results. Forensic Sci Int. 2008;174(1):55-9.
[74]Albani PP, Fleming R. Novel Messenger RNAs for Body Fluid Identification. Sci Justice. 2018;58(2):145-52.
[75]Mahdipour M, van Tol HT, Stout TA, Roelen BA. Validating Reference microRNAs for Normalizing qRT-PCR Data in Bovine Oocytes and Preimplantation Embryos. BMC Dev Biol. 2015;15:25.
[76]Zubakov D, Boersma AW, Choi Y, van Kuijk PF, Wiemer EA, Kayser M. MicroRNA Markers for Forensic Body Fluid Identification Obtained from Microarray Screening and Quantitative RT-PCR Confirmation. Int J Legal Med. 2010;124(3):217-26.

[77]Bai P, Deng W, Wang L, Long B, Liu K, Liang W, Zhang L. Micro RNA Profiling for the Detection and Differentiation of Body Fluids in Forensic Stain Analysis. Forensic Science International: Genetics Supplement Series. 2013;4(1):e216-e17.
[78]Wang Z, Luo H, Pan X, Liao M, Hou Y. A Model for Data Analysis of microRNA Expression in Forensic Body Fluid Identification. Forensic Sci Int Genet. 2012;6(3):419-23.
[79]Li Z, Bai P, Peng D, Long B, Zhang L, Liang W. Influences of Different RT-qPCR Methods on Forensic Body Fluid Identification by microRNA. Forensic Science International: Genetics Supplement Series. 2015;5:e295-e97.
[80]Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. Cell. 2007;129(7):1401-14.
[81]Courts C, Madea B. Specific Micro-RNA Signatures for the Detection of Saliva and Blood in Forensic Body-Fluid Identification. J Forensic Sci. 2011;56(6):1464-70.
[82]Wang Z, Zhang J, Luo H, Ye Y, Yan J, Hou Y. Screening and Confirmation of microRNA Markers for Forensic Body Fluid Identification. Forensic Sci Int Genet. 2013;7(1):116-23.
[83]Wang Z, Zhang J, Wei W, Zhou D, Luo H, Chen X, Hou Y. Identification of Saliva Using MicroRNA Biomarkers for Forensic Purpose. J Forensic Sci. 2015;60(3):702-6.
[84]Wang Z, Zhou D, Cao Y, Hu Z, Zhang S, Bian Y, Hou Y, Li C. Characterization of microRNA Expression Profiles in Blood and Saliva Using the Ion Personal Genome Machine((R)) System (Ion Pgm System). Forensic Sci Int Genet. 2016;20:140-46.
[85]El-Mogy M, Lam B, Haj-Ahmad TA, McGowan S, Yu D, Nosal L, Rghei N, Roberts P, Haj-Ahmad Y. Diversity and Signature of Small RNA in Different Bodily Fluids Using Next Generation Sequencing. BMC Genomics. 2018;19(1):408.
[86]Bauer M, Polzin S, Patzelt D. Quantification of RNA Degradation by Semi-Quantitative Duplex and Competitive RT-PCR: A Possible Indicator of the Age of Bloodstains? Forensic Science International. 2003;138(1-3):94-103.
[87]Alrowaithi MA. The Use of Genetic Materials, RNA, to Determine of the Age of Saliva Stains: University of Strathclyde; 2013.

[88]Simard A-M, DesGroseillers L, Sarafian V. Assessment of RNA Stability for Age Determination of Body Fluid Stains. Canadian Society of Forensic Science Journal. 2012;45(4):179-94.
[89]Alrowaithi MA, McCallum NA, Watson ND. A Method for Determining the Age of a Bloodstain. Forensic Sci Int. 2014;234:e30-1.
[90]Mraz M, Malinova K, Mayer J, Pospisilova S. MicroRNA Isolation and Stability in Stored RNA Samples. Biochem Biophys Res Commun. 2009;390(1):1-4.
[91]Nakao K-I. Experimental Study on Age Estimation of Bloodstains Based on Biological and Toxicological Analysis. The Open Forensic Science Journal. 2013;6(1):6-11.
[92]Lech K, Ackermann K, Wollstein A, Revell VL, Skene DJ, Kayser M. Assessing the Suitability of miRNA-142-5p and miRNA-541 for Bloodstain Deposition Timing. Forensic Sci Int Genet. 2014;12:181-4.
[93]Odriozola An, Riancho JA, Vega Rdl, Agudo G, García-Blanco A, Cos Ed, Fernández F, Sañudo C, Zarrabeitia MaT. miRNA Analysis in Vitreous Humor to Determine the Time of Death: A Proof-of-Concept Pilot Study. Int J Legal Med. 2013;127:573-78.
[94]Ge Q, Zhou Y, Lu J, Bai Y, Xie X, Lu Z. miRNA in Plasma Exosome Is Stable under Different Storage Conditions. Molecules. 2014;19(2):1568-75.
[95]Mall C, Rocke DM, Durbin-Johnson B, Weiss RH. Stability of miRNA in Human Urine Supports Its Biomarker Potential. Biomark Med. 2013;7(4):623-31.
[96]Mayes C, Houston R, Seashols-Williams S, LaRue B, Hughes-Stamm S. The Stability and Persistence of Blood and Semen mRNA and miRNA Targets for Body Fluid Identification in Environmentally Challenged and Laundered Samples. Leg Med (Tokyo). 2019;38:45-50.
[97]Kubista M, Björkman J, Svec D, Sjöback R. RNA Quality Matters. European Pharmaceutical Review. 2012;17(6):63-67.
[98]Bruns DE, Ashwood ER, Burtis CA. Fundamentals of Molecular Diagnostics. 1st ed. USA: Saunders/ Elsevier; 2007.
[99]Sigma-Aldrich. TRI Reagent®. MO, USA.2011.
[100]Grabmuller M, Madea B, Courts C. Comparative Evaluation of Different Extraction and Quantification Methods for Forensic RNA Analysis. Forensic Sci Int Genet. 2015;16:195-202.
[101]Qiagen. miRNeasy Mini Handbook2014 [cited 2016 April 25]; 1090243 12/2014. Available from: https://www.qiagen.com/gb/resources/resourcedetail?id=632801fb-abc5-4e62-b954-ff51f126a34f&lang=en.

[102]Ginzinger DG. Gene Quantification Using Real-Time Quantitative PCR: An Emerging Technology Hits the Mainstream. Experimental Hematology. 2002;30:503-12.
[103]Imbeaud S, Graudens E, Boulanger V, Barlet X, Zaborski P, Eveno E, Mueller O, Schroeder A, Auffray C. Towards Standardization of RNA Quality Assessment Using User-Independent Classifiers of Microcapillary Electrophoresis Traces. Nucleic Acids Res. 2005;33(6):e56.
[104]Fleige S, Pfaffl MW. RNA Integrity and the Effect on the Real-Time qRT-PCR Performance. Mol Aspects Med. 2006;27(2-3):126-39.
[105]SABiosciences. Technical Note: RNA Quality Control[cited 2016 March 22]. Available from: http://www.sabiosciences.com/newsletter/RNAQualityControl.pdf.
[106]Lightfoot S. Quantitation Comparison of Total RNA Using the Agilent 2100 Bioanalyzer, Ribo- Green Analysis and UV Spectrometry. Agilent Application Note. 2002;Publication number 5988-7650EN.
[107]Wieczorek D, Delauriere L, Schagat T. Methods of RNA Quality Assessment,2012 [cited 2016 March 22]. Available from: http://www.promega.com/pubhub.
[108]Agilent Technologies. Agilent RNA 6000 Pico Kit Guide[cited 2016 August 09]. Available from: http://www.chem.agilent.com/library/usermanuals/Public/G2938-90046_RNA600Pico_KG_EN.pdf.
[109]Becker C, Hammerle-Fickinger A, Riedmaier I, Pfaffl MW. mRNA and microRNA Quality Control for RT-qPCR Analysis. Methods. 2010;50(4):237-43.
[110]Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: An RNA Integrity Number for Assigning Integrity Values to RNA Measurements. BMC Mol Biol. 2006;7:3.
[111]Willard M. Freeman SJWaKEV. Quantitative RT-PCR: Pitfalls and Potential. Bio Techniques. 1999;26:122-25.
[112]Schmittgen TD, Lee EJ, Jiang J, Sarkar A, Yang L, Elton TS, Chen C. Real-Time PCR Quantification of Precursor and Mature microRNA. Methods. 2008;44(1):31-8.
[113]Fryer RM, Randall J, Yoshida T, Hsiao L-L, Blumenstock J, Jensen KE, Dimofte T, Jensen RV, Gullans SR. Global Analysis of Gene Expression: Methods, Interpretation, and Pitfalls. Experimental Nephrology. 2002;10:64-74.
[114]Wacker MJ, Godard MP. Analysis of One-Step and Two- Step Real-Time RT-PCR Using Superscript Iii. Journal Of Biomolecular Techniques. 2005;16(3):266-71.
[115]O'Driscoll L. Gene Expression Profiling: Springer, Humana Press; 2011.

[116]Bustina SA, Nolan T. Pitfalls of Quantitative Real- Time Reverse-Transcription Polymerase Chain Reaction. Journal Of Biomolecular Techniques. 2004;15(3):155-66.
[117]Resuehr D, Spiess A-N. A Real-Time Polymerase Chain Reaction-Based Evaluation of cDNA Synthesis Priming Methods. Anal Biochem. 2003;322(2):287-91.
[118]Brink AATP, Oudejans JJ, Jiwa M, Walboomers JMM, Meijer CJLM, Brule AJCvd. Multiprimed cDNA Synthesis Followed by PCR Is the Most Suitable Method for Epstein-Barr Virus Transcript Analysis in Small Lymphoma Biopsies. Molecular and Cellular Probes. 1997;11(1):39-47.
[119]Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ. Real-Time Quantification of microRNAs by Stem-Loop RT-PCR. Nucleic Acids Res. 2005;33(20):e179.
[120]Kramer MF. Stem-Loop RT-qPCR for miRNAs. Curr Protoc Mol Biol. 2011;Chapter 15:Unit 15 10.
[121]Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. Protocol: A Highly Sensitive RT-PCR Method for Detection and Quantification of microRNAs. Plant Methods. 2007;3:12.
[122]Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, Chajut A. Serum Micrornas Are Promising Novel Biomarkers (Serum MicroRNA as Biomarkers). Plos One. 2008;3(9):e3148.
[123]Arya M, Shergill IS, Williamson M, Gommersall L, Arya N, Patel HR. Basic Principles of Real-Time Quantitative PCR. Expert Rev. Mol. Diagn. 2005;5(2):209-19.
[124]Porterfield A. What Is a Ct Value?2015 6th April 2016. Available from: http://bitesizebio.com/24581/what-is-a-ct-value/.
[125]Wong ML, Medrano JF. Real-Time PCR for mRNA Quantitation. Bio Techniques. 2005;39(1):75-85.
[126]Dvořák Zk, Pascussi J-M, Modrianský M. Approaches to Messenger RNA Detection – Comparison of Methods. Biomed. Papers. 2003;147(2):131-35.
[127]Wang T, Brown MJ. mRNA Quantification by Real Time TaqMan Polymerase Chain Reaction: Validation and Comparison with Rnase Protection. Anal Biochem. 1999;269:198-201.
[128]Malinen E. Comparison of Real-Time PCR with SYBR Green I or 5'-Nuclease Assays and Dot-Blot Hybridization with rDNA-Targeted Oligonucleotide Probes in Quantification of Selected Faecal Bacteria. Microbiology. 2003;149(1):269-77.

[129]Smith CJ, Osborn AM. Advantages and Limitations of Quantitative PCR (Q-PCR)-Based Approaches in Microbial Ecology. FEMS Microbiol Ecol. 2009;67(1):6-20.
[130]Cardullo RA, Agrawal S, Flores C, Zamecnik PC, Wolf DE. Detection of Nucleic Acid Hybridization by Nonradiative Fluorescence Resonance Energy Transfer. Biochemistry. 1988;85:8790-94.
[131]Ali S, Ali S. Differential Gene Expression and Its Possible Therapeutic Implications. In: Kang PC, editor. Gene Therapy - Developments and Future Perspectives: InTech; 2011.
[132]Benes V, Castoldi M. Expression Profiling of microRNA Using Real-Time Quantitative PCR, How to Use It and What Is Available. Methods. 2010;50(4):244-9.
[133]Applied Biosystems, Introduction to TaqMan® and SYBR® Green Chemistries for Real-Time PCR. Applied Biosystems; 2010. p. 1-10.
[134]Klein D. Quantification Using Real-Time PCR Technology: Applications and Limitations. TRENDS in Molecular Medicine. 2002;8(6):257-60.
[135]Cao H, Shockey JM. Comparison of TaqMan and SYBR Green qPCR Methods for Quantitative Gene Expression in Tung Tree Tissues. Journal of Agricultural and Food Chemistry. 2012;60(50):12296-303.
[136]Pfaffl MW. Relative Quantification. In: Dorak T, editor. In: Real-Time PCR. USA: International University Line; 2006. p. 63-86.
[137]Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR. Applied Biosystems; 2008.
[138]Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N. Statistical Significance of Quantitative PCR. BMC Bioinformatics. 2007;8:131.
[139]Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.
[140]Moreno LI, Tate CM, Knott EL, McDaniel JE, Rogers SS, Koons BW, Kavlick MF, Craig RL, Robertson JM. Determination of an Effective Housekeeping Gene for the Quantification of mRNA for Forensic Applications. Journal of Forensic Sciences. 2012;57(4):1051-58.
[141]Huggett J, Dheda K, Bustin S, Zumla A. Real-Time RT-PCR Normalisation; Strategies and Considerations. Genes Immun. 2005;6(4):279-84.
[142]Afolabi OA, Roeder AD, Iyengar A, Hadi S. Reference Gene Study for Forensic Body Fluid Identification. Forensic Science International: Genetics Supplement Series. 2015;5:e167-e69.

[143]Egidi MG, Cochetti G, Guelfi G, Zampini D, Diverio S, Poli G, Mearini E. Stability Assessment of Candidate Reference Genes in Urine Sediment of Prostate Cancer Patients for miRNA Applications. Dis Markers. 2015:ID 973597.
[144]Jain M, Nijhawan A, Tyagi AK, Khurana JP. Validation of Housekeeping Genes as Internal Control for Studying Gene Expression in Rice by Quantitative Real-Time PCR. Biochem Biophys Res Commun. 2006;345(2):646-51.
[145]Zheng G, Wang H, Zhang X, Yang Y, Wang L, Du L, Li W, Li J, Qu A, Liu Y, Wang C. Identification and Validation of Reference Genes for qPCR Detection of Serum microRNAs in Colorectal Adenocarcinoma Patients. Plos One. 2013;8(12):e83025.
[146]Andersen CL, Jensen JL, Ørntoft TF. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research. 2004;64:5245-50.
[147]Vandesompele J, Preter KD, Pattyn F, Poppe B, Roy NV, Paepe AD, Speleman F. Accurate Normalization of Real-Time Quantitative Rt- PCR Data by Geometric Averaging of Multiple Internal Control Genes. Genome Biology. 2002;3(7):1-12.
[148]Ambion LT. TURBO DNA-Freetm Kit, TURBOTM DNase Treatment and Removal Reagents[cited 2016 April 25]. Available from: https://tools.thermofisher.com/content/sfs/manuals/cms_055740.pdf.
[149]ThermoFisherScientific. Nucleic Acid-Nanodrop Spectrophtimeters [cited 2019 February]. Available from: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/CAD/manuals/ts-nanodrop-nucleicacid-olv-r2.pdf&title=TmFub0Ryb3AgTnVjbGVpYyBBY2lkIEhhbmRib29r.
[150]Agilent RNA 6000 Pico. Germany: Agilent; 2016.
[151]High Capacity cDNA Reverse Transcription Kits. California, USA: Applied Biosystems; 2010.
[152]TaqMan® MicroRNA Assays. California, USA: Applied Biosystems; 2006.
[153]TaqMan® Gene Expression Assays. California, USA: Applied Biosystems; 2010.
[154]Dinno A. Dunn.Test: Dunn's Test of Multiple Comparisons Using Rank Sums. R Package Version 1.3.5 2017. Available from: https://CRAN.R-project.org/package=dunn.test.
[155]Tian H, Lv M, Li Z, Peng D, Tan Y, Wang H, Li Q, Li F, Liang W. Semen-Specific miRNAs: Suitable for the Distinction of Infertile Semen in the Body Fluid Identification? Forensic Sci Int Genet. 2017;33:161-67.

[156]Sirker M, Fimmers R, Schneider PM, Gomes I. Evaluating the Forensic Application of 19 Target microRNAs as Biomarkers in Body Fluid and Tissue Identification. Forensic Sci Int Genet. 2017;27:41-49.
[157]Wang Z, Yang B. MicroRNA Expression Detection Methods. Germany: Springer; 2010.
[158]Kim J-H, Jin H-O, Park J-A, Chang YH, Hong YJ, Lee JK. Comparison of Three Different Kits for Extraction of High-Quality RNA from Frozen Blood. SpringerPlus. 2014;3(76).
[159]Jakovljevic K, Spasic M, Malisic E, Dobricic J, Krivokuca A, Jankovic R. Comparison of Phenol-Based and Alternative RNA Isolation Methods for Gene Expression Analyses. Journal of the Serbian Chemical Society. 2010;75(8):1053-61.
[160]van der Linden A, Blokker BM, Kap M, Weustink AC, Robertus JL, Riegman PH, Oosterhuis JW. Post-Mortem Tissue Biopsies Obtained at Minimally Invasive Autopsy: An RNA-Quality Analysis. Plos One. 2014;9(12):e115675.
[161]Alves MG, Perez-Sayans M, Padin-Iruegas ME, Reboiras-Lopez MD, Suarez-Penaranda JM, Lopez-Lopez R, Carta CF, Issa JS, Garcia-Garcia A, Almeida JD. Comparison of RNA Extraction Methods for Molecular Analysis of Oral Cytology. Acta Stomatol Croat. 2016;50(2):108-15.
[162]Barragan M, Martinez A, Llonch S, Pujol A, Vernaeve V, Vassena R. Effect of Ribonucleic Acid (RNA) Isolation Methods on Putative Reference Genes Messenger RNA Abundance in Human Spermatozoa. Andrology. 2015;3(4):797-804.
[163]Weis S, Llenos IC, Dulay JR, Elashoff M, Martinez-Murillo F, Miller CL. Quality Control for Microarray Analysis of Human Brain Samples: The Impact of Postmortem Factors, RNA Characteristics, and Histopathology. J Neurosci Methods. 2007;165(2):198-209.
[164]Fang C, Zhao J, Li J, Qian J, Liu X, Sun Q, Liu W, Tian Y, Ji A, Wu H, Yan J. Massively Parallel Sequencing of microRNA in Bloodstains and Evaluation of Environmental Influences on miRNA Candidates Using Realtime Polymerase Chain Reaction. Forensic Sci Int Genet. 2018;38:32-38.
[165]Fujimoto S, Manabe S, Morimoto C, Ozeki M, Hamano Y, Tamaki K. Optimal Small-Molecular Reference RNA for RT-qPCR-Based Body Fluid Identification. Forensic Sci Int Genet. 2018;37:135-42.
[166]Gorzkiewicz M, Duleba A, Rychlicka E, Wozniak M, Grzybowski T, Sliwka K. Evaluation of the Agilent 2100 Bioanalyzer as a Tool for DNA Analysis in Forensic Genetics. Problems of Forensic Sciences. 2010;LXXXI:91-100.
[167]Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The Miqe

Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009;55(4):611-22.
[168]Bauer M, Patzelt D. Protamine mRNA as Molecular Marker for Spermatozoa in Semen Stains. Int J Legal Med. 2003;117(3):175-9.
[169]National Center for Biotechnology Information (NCBI) Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] [Internet]. 2018 [cited 05/02/2018]. Available from: https://www.ncbi.nlm.nih.gov/.
[170]Haas C, Hanson E, Kratzer A, Bar W, Ballantyne J. Selection of Highly Specific and Sensitive mRNA Biomarkers for the Identification of Blood. Forensic Sci Int Genet. 2011;5(5):449-58.
[171]Tian R, Wang J, Yan H, Wu J, Xu Q, Zhan X, Gui Z, Ding M, He J. Differential Expression of miR16 in Glioblastoma and Glioblastoma Stem Cells: Their Correlation with Proliferation, Differentiation, Metastasis and Prognosis. Oncogene. 2017;36(42):5861-73.
[172]Liu X, Zhang A, Xiang J, Lv Y, Zhang X. miR-451 Acts as a Suppressor of Angiogenesis in Hepatocellular Carcinoma by Targeting the Il-6r-Stat3 Pathway. Oncol Rep. 2016;36(3):1385-92.
[173]Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N. Next-Generation Sequencing of Transcriptomes: A Guide to RNA Isolation in Nonmodel Animals. Mol Ecol Resour. 2011;11(4):650-61.
[174]Romero IG, Pai AA, Tung J, Gilad1 Y. RNA-Seq: Impact of RNA Degradation on Transcript Quantification. BMC Biology. 2014;12:42.
[175]Conti T, Buel E. Forensic Stain Identification by RT-PCR Analysis. Washington, DC: U.S. Department of Justice. 2011;Report No.: 2007-DN-BX-K149 (236537).
[176]Zhao H, Wang C, Yao L, Lin Q, Xu X, Hu L, Li W. Identification of Aged Bloodstains through mRNA Profiling: Experiments Results on Selected Markers of 30- and 50-Year-Old Samples. Forensic Sci Int. 2017;272(Supplement C):e1–6.
[177]Russell E, Liebhaber SA. The Stability of Human P-Globin mRNA Is Dependent on Structural Determinants Positioned within Its 3' Untranslated Region. Blood. 1996;87(12):5314-23.
[178]Patnaik SK, Mallick R, Yendamuri S. Detection of microRNAs in Dried Serum Blots. Anal Biochem. 2010;407(1):147-9.
[179]Elgamal OA, Park JK, Gusev Y, Azevedo-Pouly AC, Jiang J, Roopra A, Schmittgen TD. Tumor Suppressive Function of mir-205 in Breast Cancer Is Linked to Hmgb3 Regulation. Plos One. 2013;8(10):e76402.
[180]Rasnic R, Linial N, Linial M. Enhancing Identification of Cancer Types Via Lowly-Expressed microRNAs. Nucleic Acids Res. 2017;45(9):5048-60.

[181]Wouters-Tyroua D, Martinage A, Chevaillier P, Sautibred P. Nuclearbasicproteins Spermiogenes S. biochemie. 1998;80:117-28.
[182]Belleannee C, Calvo E, Thimon V, Cyr DG, Legare C, Garneau L, Sullivan R. Role of microRNAs in Controlling Gene Expression in Different Segments of the Human Epididymis. Plos One. 2012;7(4):e34996.
[183]Park NJ, Li Y, Yu T, Brinkman BM, Wong DT. Characterization of RNA in Saliva. Clin Chem. 2006;52(6):988-94.
[184]Bianchi E, Stermer A, Boekelheide K, Sigman M, Hall SJ, Reyes G, Dere E, Hwang K. High-Quality Human and Rat Spermatozoal RNA Isolation for Functional Genomic Studies. Andrology. 2018;6(2):374-83.
[185]Sakurada K, Ikegaya H, Fukushima H, Akutsu T, Watanabe K, Yoshino M. Evaluation of mRNA-Based Approach for Identification of Saliva and Semen. Leg Med (Tokyo). 2009;11(3):125-8.
[186]Watanabe K, Akutsu T, Takamura A, Sakurada K. Practical Evaluation of an RNA-Based Saliva Identification Method. Science & Justice. 2017;57(6):404-08.
[187]Weinbrecht K. RNA-Seq of Biological Fluids for the Evaluation of mRNA Degradation in Relation to Sample Age: Oklahoma State University; 2014.
[188]Nakanishi H, Hara M, Takahashi S, Takada A, Saito K. Evaluation of Forensic Examination of Extremely Aged Seminal Stains. Leg Med (Tokyo). 2014;16(5):303-7.
[189]Li H, Wu C, Gu X, Xiong C. A Novel Application of Cell-Free Seminal mRNA: Non-Invasive Identification of the Presence of Germ Cells or Complete Obstruction in Men with Azoospermia. Hum Reprod. 2012;27(4):991-7.
[190]El-Khazragy N, Elsayed EA, M.S.Azab S, H.Badway A, FAlakid Y, Halawa HM. Impact of Semen/Vaginal Fluid Mixture on the Expression of Mir891a and Mir124a in Dried Sexual Stains. International Journal of Science and Research. 2017;6(3):1243-49.
[191]Tong D, Jin Y, Xue T, Ma X, Zhang J, Ou X, Cheng J, Sun H. Investigation of the Application of miR10b and miR135b in the Identification of Semen Stains. Plos One. 2015;10(9):e0137067.
[192]Alshehhi S, McCallum NA, Haddrill PR. Quantification of RNA Degradation of Blood-Specific Markers to Indicate the Age of Bloodstains. Forensic Science International: Genetics Supplement Series. 2017;6:e453-e55.
[193]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.
[194]Harteveld J, Lindenbergh A, Sijen T. RNA Cell Typing and DNA Profiling of Mixed Samples: Can Cell Types and Donors Be Associated? Sci Justice. 2013;53(3):261-9.

[195]Matsumura S, Matsusue A, Waters B, Kashiwagi M, Hara K, Kubo S. Application of mRNA Expression Analysis to Human Blood Identification in Degenerated Samples That Were False-Negative by Immunochromatography. J Forensic Sci. 2016;61(4):903-12.
[196]Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, Gu L, Hu W, Bi G, Ge J, Chen X, Zhao Z. Development of Highly Sensitive and Specific mRNA Multiplex System (XCYR1) for Forensic Human Body Fluids and Tissues Identification. Plos One. 2014;9(7):e100123.
[197]Suzuki T, Higgins PJ, Crawford DR. Control Selection for RNA Quantitation. BioTechniques. 2000;29(2):332-37.
[198]Lee PD, Sladek R, Greenwood CMT, Hudson TJ. Control Genes and Variability: Absence of Ubiquitous Reference Transcripts in Diverse Mammalian Expression Studies. Genome Research. 2002;12(1):292-97.
[199]Stamova BS, Apperson M, Walker WL, Tian Y, Xu H, Adamczy P, Zhan X, Liu DZ, Ander BP, Liao IH, Gregg JP, Turner RJ, Jickling G, Lit L, Sharp FR. Identification and Validation of Suitable Endogenous Reference Genes for Gene Expression Studies in Human Peripheral Blood. BMC Med Genomics. 2009;2:49.
[200]Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to Reference Gene Selection for Quantitative Real-Time PCR. Biochem Biophys Res Commun. 2004;313(4):856-62.
[201]Sun Y, Li Y, Luo D, Liao DJ. Pseudogenes as Weaknesses of ACTB (Actb) and GAPDH (Gapdh) Used as Reference Genes in Reverse Transcription and Polymerase Chain Reactions. Plos One. 2012;7(8):e41659.
[202]Eisenberg E, Levanon EY. Human Housekeeping Genes, Revisited. Trends Genet. 2013;29(10):569-74.
[203]Rojas-Hernandez N, Veliz D, Vega-Retter C. Selection of Suitable Reference Genes for Gene Expression Analysis in Gills and Liver of Fish under Field Pollution Conditions. Sci Rep. 2019;9(1):3459.
[204]Usarek E, Baranczyk-Kuzma A, Kazmierczak B, Gajewska B, Kuzma-Kozakiewicz M. Validation of qPCR Reference Genes in Lymphocytes from Patients with Amyotrophic Lateral Sclerosis. Plos One. 2017;12(3):e0174317.
[205]Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of Housekeeping Genes for Normalizing RNA Expression in Real-Time PCR. BioTechniques. 2004;37(1):112-4, 16, 18-9.
[206]Ho NT, Busik JV, Resau JH, Paneth N, Khoo SK. Effect of Storage Time on Gene Expression Data Acquired from Unfrozen Archived Newborn Blood Spots. Mol Genet Metab. 2016;119(3):207-13.
[207]Lindenbergh A, Maaskant P, Sijen T. Implementation of RNA Profiling in Forensic Casework. Forensic Sci Int Genet. 2013;7(1):159-66.

[208]Sakurada K, Akutsu T, Watanabe K, Miyasaka S, Kasai K. Identification of Body Fluid Stains Using Real-Time RT-PCR: Discrimination between Salivary, Nasal, and Vaginal Secretions. Scienc and Technology. 2013;18(1).
[209]Sherier AJ. Patterns of Degradation in PRM1 and SEMG1 Transcripts in Semen During 6 Months of Aging. 2016;Oklahoma State University, ProQuest Dissertations Publishing.

APPENDIX

APPENDIX NUMBER	CONTENT
A1	Ethical forms
A2	MIQE checklist
A3	Quality control in real-time PCR. TaqMan® Gene Expression Assays
A4	Quality control in real-time PCR. TaqMan® MicroRNA Assays
A5	Intra-assay variation
A6	Inter-assay variation

A 1:

Participant Information Sheet

Name of department: Department of Pure and Applied Chemistry

Title of the study: Estimating the age of dried biological stains using RNA markers

Introduction

Chief Investigator:

Name: Dr Penny Haddrill

Status: Teaching Associate

Tel.: 0141 548 4337

E-mail: penny.haddrill@strath.ac.uk

Co-Investigator:

Name: Suaad Alshehhi

Status: PhD Student

Tel.: 0141 958 5992

E-mail: suaad.alshehhi@strath.ac.uk

What is the purpose of this investigation?

At present, no technique exists in forensic science to estimate the time at which a biological stain (such as a blood, saliva or semen stain) was deposited at a potential crime scene. Such a technique would be invaluable, allowing investigators to estimate the time at which a crime was committed or determine whether a biological stain recovered is of relevance to a known crime event. Biological stains encountered in forensic casework contain RNA: a molecule which is known to be unstable in the environment and is subject to continuous degradation. The aim of this study is to develop a method to quantify RNA degradation in biological stains, as a means to estimate the time point at which the stain was deposited – i.e. the 'age' of the stain.

Do you have to take part?

This study involves participants donating biological samples in the form of blood, saliva and semen. Participation in this research is entirely voluntary and you have the right to refuse to participate without giving a reason; refusing to participate will not negatively affect you in any way. You also have the right to withdraw from this research at any time up to the completion of the project without detriment and without giving a reason, and ask for your data to be destroyed.

What will you do in the project?

You will be required to provide one or more biological samples:

- A small blood sample, collected by one of the trained investigators using a lancet from a finger 'prick' and deposited onto swab/fabric.
- A small saliva sample (less than 5 mL), collected by yourself in a sterile tube.
- A small semen sample (less than 5 mL), collected by yourself in a sterile tube.

These samples will be used by the investigators to create blood, saliva and semen stains. The stains will be dried and aged for up to 1 year by leaving them untouched in ambient room conditions, and the RNA within the stains analysed to determine whether it can be used as a predictor of stain age.

This will take place in the Centre for Forensic Science, University of Strathclyde, Royal College Building, 204 George Street, Glasgow, G1 1XW. No payments will be provided for taking part in this research.

Why have you been invited to take part?

In order to study the ageing of blood saliva and semen stains, it is essential that these samples are collected from volunteers; these samples cannot be simulated. All participants over 18 in age are welcomed to participate, in order to provide the necessary biological samples.

What are the potential risks to you in taking part?

The collection of blood using a lancet carries a small risk of infection to the donor, which will be minimised with the use of alcohol wipes and participants will be given a plaster following donation. In addition, the handling of blood, saliva and semen carries a small risk of infection.

What happens to the information in the project?

All participants' information will be kept confidential. In order to preserve anonymity, samples will be labelled with a code that does not contain any information allowing the participant to be identified; except for by the investigators, to allow a participant's data to be identified, removed and destroyed should they subsequently wish to withdraw. For the duration of the study, biological samples will be stored in the DNA suite (R6.24) of the Centre for Forensic Science, University of Strathclyde. This is a restricted laboratory space with keypad access only for authorised laboratory personnel. All data outputs will be stored electronically on a password-protected computer only accessible by the investigators within the DNA suite (R6.24). Electronic data may be retained indefinitely in this form, and no information will be put onto any databases. All biological samples (e.g. biological stains, RNA samples) will be securely disposed of within 1 month of the conclusion of the study, which is estimated to be around the 1st October 2018. Once the project is completed, the codes linking identity with samples will be deleted.

The outcomes of this study will be written into a PhD thesis by co-investigator Suaad Alshehhi. In addition, it is envisaged that the outcomes of this study will be written into journal/conference publication(s). In neither form of publication will any information be included that could allow the participants to be identified.

The University of Strathclyde is registered with the Information Commissioner's Office who implements the Data Protection Act 1998. All personal data on participants will be processed in accordance with the provisions of the Data Protection Act 1998.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

What happens next?

If you are happy to be involved in the project, please sign the consent form provided to confirm this. Please note that participants will not be informed of the specific results of the tests.

If you do not want to be involved in the project, we would like to thank you for your attention.

Researcher contact details:

Suaad Alshehhi, PhD Student

Centre for Forensic Science, Department of Pure and Applied Chemistry

University of Strathclyde

Royal College, 204 George Street, Glasgow, G1 1XW

Telephone: 0141 958 5992

E-mail: suaad.alshehhi@strath.ac.uk

Chief Investigator details:

Dr Penny Haddrill, Teaching Associate

Centre for Forensic Science, Department of Pure and Applied Chemistry

University of Strathclyde

Royal College, 204 George Street, Glasgow, G1 1XW

Telephone: 0141 548 4377

E-mail: penny.haddrill@strath.ac.uk

This investigation was granted ethical approval by the University of Strathclyde Ethics Committee.

If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, please contact:

Secretary to the University Ethics Committee
Research & Knowledge Exchange Services
University of Strathclyde
Graham Hills Building
50 George Street
Glasgow
G1 1QE

Telephone: 0141 548 3707
Email: ethics@strath.ac.uk

Consent Form

Name of department: Department of Pure and Applied Chemistry

Title of the study: Estimating the age of dried biological stains using RNA markers

- I confirm that I have read and understood the information sheet for the above project and the researcher has answered any queries to my satisfaction.
- I understand that my participation is voluntary and that I am free to withdraw from the project at any time, up to the point of completion, without having to give a reason and without any consequences. If I exercise my right to withdraw and I don't want my data to be used, any data which have been collected from me will be destroyed.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I understand that I can withdraw from the study any personal data (i.e. data which identify me personally) at any time.

- I consent to being a participant in the project.
- I understand that I will be asked to donate one or more of the following:
  - A blood sample, collected by the investigators using a sterile lancet.
  - A saliva sample, collected by myself in a sterile tube.
  - A semen sample, collected by myself in a sterile tube.

- I consent to the RNA in my samples being analysed.
- I understand that all of my biological samples will be securely destroyed by 1st October 2018.

I consent to the taking of biological samples from me, and understand that they will be the property of the University of Strathclyde. [Yes / No]

(PRINT NAME)	
Signature of Participant:	Date:

A 2: ***MIQE checklist****. (E) is essential information and must be submitted with the manuscript, (D) is desirable and information can be submitted if available.*

ITEM TO CHECK	IMPORTANCE	INCLUDED ?	COMMENTS
Definition of experimental and control groups	**E**	**Y**	Method: control groups defined as fresh sample, time point 0
Number within each group	**E**	**Y**	Method: 8 samples for each time point
Assay carried out by core lab or investigator's lab?	**D**	**Y**	Investigator lab
Acknowledgement of authors' contributions	**D**	**N**	
Description	**E**	**Y**	Method: body fluids (blood, saliva and semen)
Volume/mass of sample processed	**D**	**Y**	20 µL of blood, 50 µL of saliva and semen
Microdissection or macrodissection	**E**	**N**	
Processing procedure	**E**	**Y**	Method: RNA was extracted immediately after a desired age was reached
If frozen - how and how quickly?	**E**	**N**	
If fixed - with what, how quickly?	**E**	**N**	
Sample storage conditions and duration (especially for FFPE samples)	**E**	**Y**	Method: samples were stored at room temperature in dark dry place for one year/60 days
Procedure and/or instrumentation	**E**	**Y**	TRI Reagent® (Sigma Aldrich)
Name of kit and details of any modifications	**E**	**Y**	miRNeasy Mini Kit (Qiagen)

Source of additional reagents used	**D**	**Y**	Sigma Aldrich, Qiagen and Molecular Research Center
Details of DNase or RNAse treatment	**E**	**Y**	Method: TURBO-DNA-free
Contamination assessment (DNA or RNA)	**E**	**Y**	Method: NanoDrop-1000/Bioanalyzer 2100
Nucleic acid quantification	**E**	**Y**	Method: NanoDrop-1000
Instrument and method	**E**	**Y**	Method: NanoDrop-1000/ Bioanalyzer 2100
Purity (A260/A280)	**D**	**Y**	Results: $A_{260}/_{280}$ data
Yield	**D**	**Y**	Results: total RNA yield
RNA integrity method/instrument	**E**	**Y**	Bioanalyzer 2100
RIN/RQI or Cq of 3' and 5' transcripts	**E**	**N**	
Electrophoresis traces	**D**	**N**	
Inhibition testing (Cq dilutions, spike or other)	**E**	**Y**	Generating standard curve
Complete reaction conditions	**E**	**Y**	Method: following High Capacity cDNA Reverse Transcription Kit and MicroRNA Reverse Transcription Kit (Applied Biosystems)
Amount of RNA and reaction volume	**E**	**Y**	Method: total volume 20 µL
Priming oligonucleotide (if using GSP) and concentration	**E**	**Y**	Method: random primers
Reverse transcriptase and concentration	**E**	**Y**	Method: as per kit
Temperature and time	**E**	**Y**	Method: as per kit
Manufacturer of reagents and catalogue numbers	**D**	**Y**	Method: High Capacity cDNA Reverse Transcription Kit (Applied Biosystems 4368814), and MicroRNA Reverse Transcription Kit (Applied Biosystems 4366596)
Cqs with and without RT	**D***	**Y**	Results: RT- controls performed as part of assay validation
Storage conditions of cDNA	**D**	**Y**	Method: frozen at -20 ˚C

If multiplex, efficiency and LOD of each assay.	**E**	**N/A**	
Sequence accession number	**E**	**Y**	Life technologies assay ID
Location of amplicon	**D**	**Y**	Supplementary data
Amplicon length	**E**	**Y**	Supplementary data
In silico specificity screen (BLAST, etc)	**E**	**N**	
Pseudogenes, retropseudogenes or other homologs?	**D**	**N**	
Sequence alignment	**D**	**N**	
Secondary structure analysis of amplicon	**D**	**N**	
Location of each primer by exon or intron (if applicable)	**E**	**N**	
What splice variants are targeted?	**E**	**N**	
Primer sequences	**E**	**N**	
RTPrimerDB Identification Number	**D**	**Y**	Life technologies assay ID
Probe sequences	**D****	**N**	
Location and identity of any modifications	**E**	**N**	
Manufacturer of oligonucleotides	**D**	**Y**	Life Technologies
Purification method	**D**	**N**	
Complete reaction conditions	**E**	**Y**	Method: TaqMan® Universal qPCR Master Mix used
Reaction volume and amount of cDNA/DNA	**E**	**Y**	20 µL reaction volume, 4 µL cDNA
Primer, (probe), Mg++ and dNTP concentrations	**E**	**Y**	Method: TaqMan® Universal qPCR Master Mix used
Polymerase identity and concentration	**E**	**Y**	Method: TaqMan® Universal qPCR Master Mix used
Buffer/kit identity and manufacturer	**E**	**Y**	Method: TaqMan® Universal qPCR Master Mix used
Exact chemical constitution of the buffer	**D**	**Y**	Method: TaqMan® Universal qPCR Master Mix used
Additives (SYBR Green I, DMSO, etc.)	**E**	**N/A**	
Manufacturer of plates/tubes and catalogue number	**D**	**N**	

Complete thermocycling parameters	**E**	**Y**	Method: as recommended by assay manufacturer
Reaction setup (manual/robotic)	**D**	**Y**	Manual
Manufacturer of qPCR instrument	**E**	**Y**	Method: Stratagene Mx3005P
Evidence of optimisation (from gradients)	**D**	**N**	Inventoried assays were purchased from Life Technologies
Specificity (gel, sequence, melt, or digest)	**E**	**N**	Inventoried assays were purchased from Life Technologies
For SYBR Green I, Cq of the NTC	**E**	**N/A**	
Standard curves with slope and y-intercept	**E**	**Y**	Results: quality assessment
PCR efficiency calculated from slope	**E**	**Y**	Results: quality assessment
Confidence interval for PCR efficiency or standard error	**D**	**Y**	Results: quality assessment
r2 of standard curve	**E**	**Y**	Results: quality assessment
Linear dynamic range	**E**	**Y**	Results: quality assessment
Cq variation at lower limit	**E**	**N**	
Confidence intervals throughout range	**D**	**N**	
Evidence for limit of detection	**E**	**N**	
If multiplex, efficiency and LOD of each assay.	**E**	**N/A**	
qPCR analysis program (source, version)	**E**	**Y**	MxPro (Agilent Technologies)
Cq method determination	**E**	**Y**	MxPro (Agilent Technologies)
Outlier identification and disposition	**E**	**N**	
Results of NTCs	**E**	**Y**	
Justification of number and choice of reference genes	**E**	**N/A**	Only reference genes considered
Description of normalisation method	**E**	**Y**	Normalised against reference genes
Number and concordance of biological replicates	**D**	**N**	

Number and stage (RT or qPCR) of technical replicates	**E**	**Y**	Method: RT performed in singular, qPCR in duplicate
Repeatability (intra-assay variation)	**E**	**Y**	Results: part of assay validation
Reproducibility (inter-assay variation, %CV)	**D**	**Y**	Results: part of assay validation
Power analysis	**D**	**N**	
Statistical methods for result significance	**E**	**Y**	2 sample t-test and Mann Whitney test Correlation tests ANOVA Kruskal-Wallis
Software (source, version)	**E**	**Y**	Minitab Express Miniab17 Microsoft Excel
Cq or raw data submission using RDML	**D**	**N**	

* Assessing the absence of DNA using a no RT assay is essential when first extracting RNA.

** Disclosure of the probe sequence in highly desirable and strongly encourage

A 3: ***Quality control in real-time PCR****. TaqMan® Gene Expression Assays validations and standard curve.*

The following charts illustrate the standard curve of each TaqMan® Gene Expression Assay that has been validated in Chapter 3 of this report. All TaqMan® assays were purchased from Life Technologies.

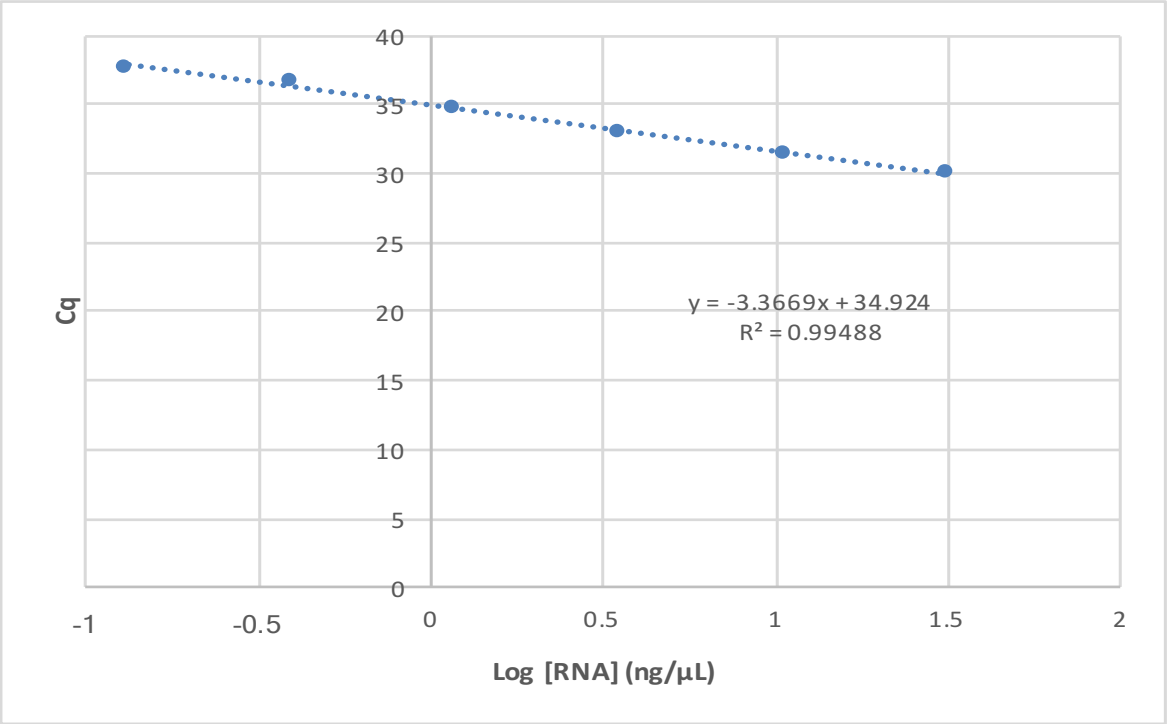


Figure A3. 1: ***Standard curve for ACTB TaqMan® assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 85 ng/µL to 0.035 ng/µL of starting RNA.*

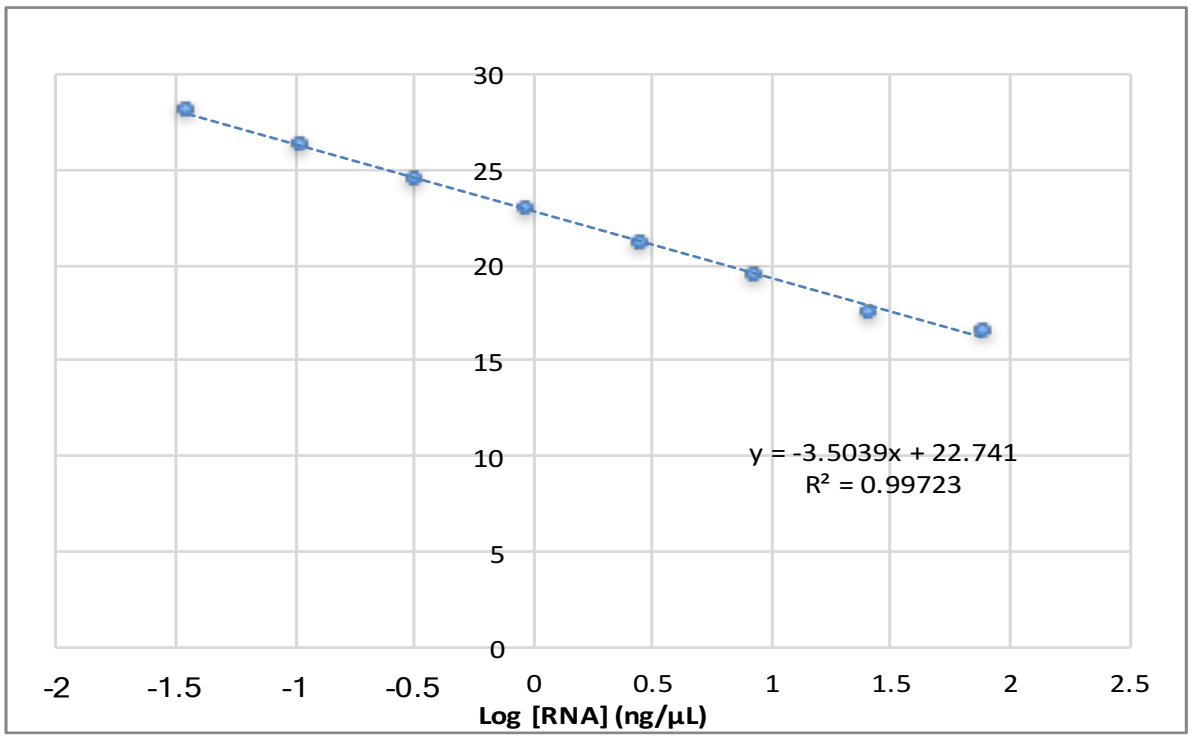


Figure A3. 2: ***Standard curve for HBB TaqMan® assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 85 ng/µL to 0.035 ng/µL of starting RNA.*

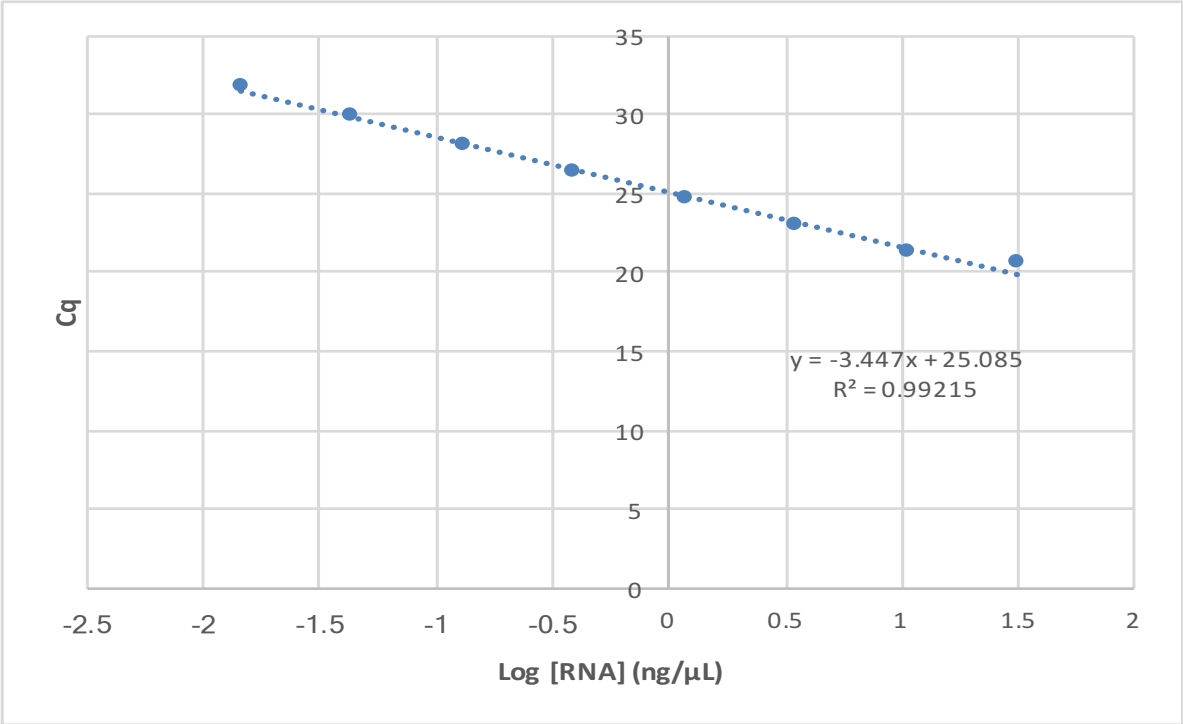


Figure A3. 3: ***Standard curve for 18S TaqMan® assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 85 ng/µL to 0.035 ng/µL of starting RNA.*

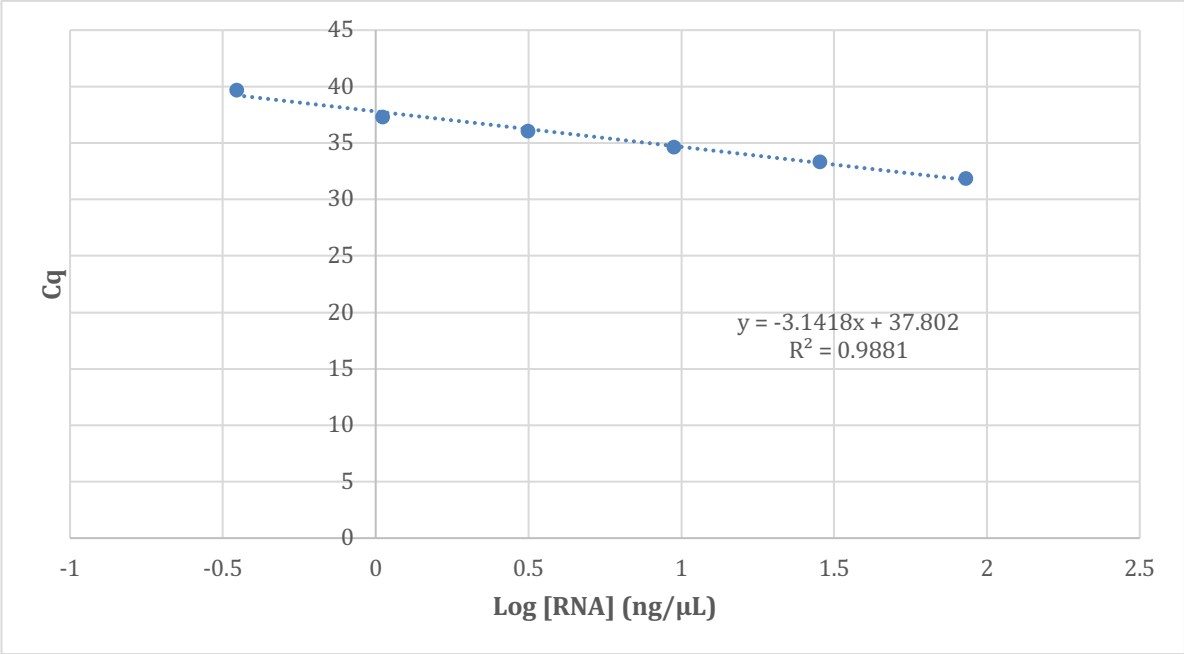


Figure A3. 4: ***Standard curve for HMBS TaqMan® assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 85 ng/µL to 0.035 ng/µL of starting RNA.*

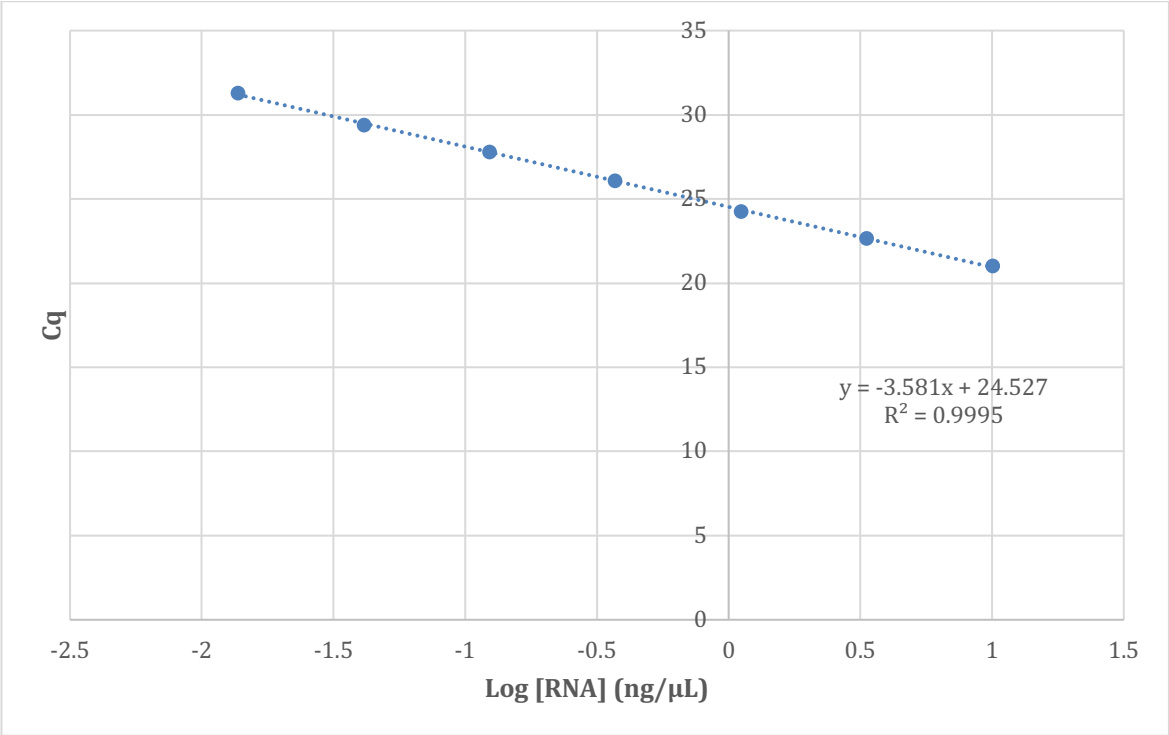


Figure A3. 5: ***Standard curve for HBA TaqMan® assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 85 ng/µL to 0.035 ng/µL of starting RNA.*

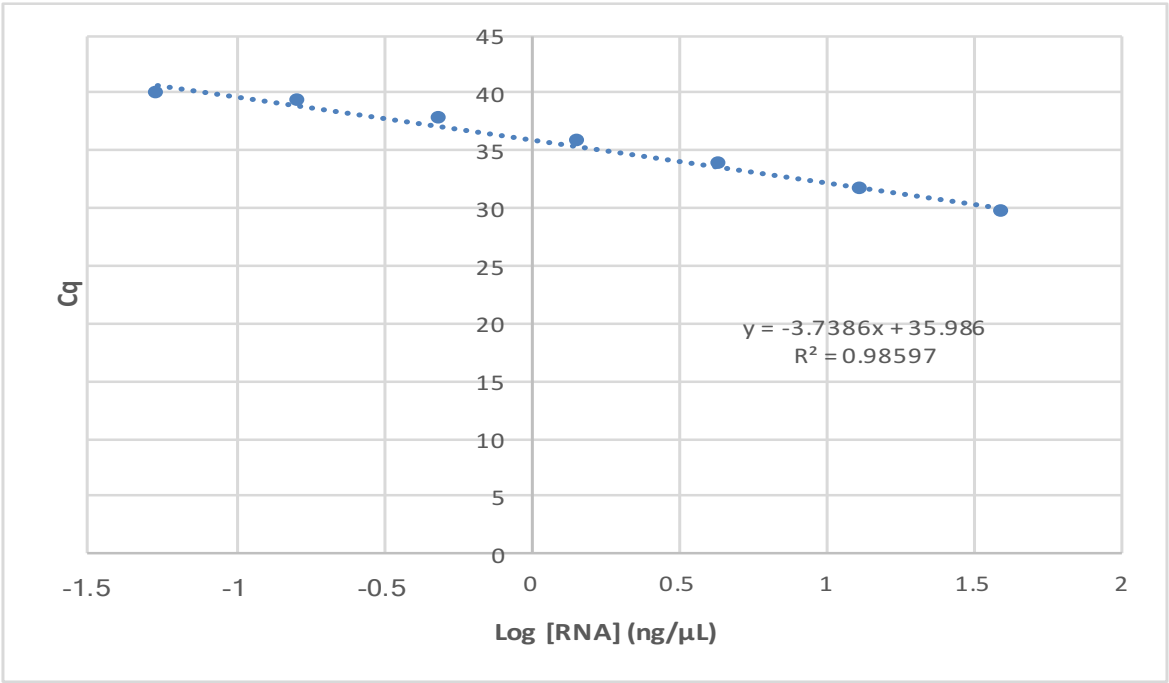


Figure A3. 6: ***Standard curve for STATH TaqMan® assay****. Using 1:3 dilution series of saliva sample cDNA in concentration from 40 ng/µL to 0.018 ng/µL of starting RNA.*

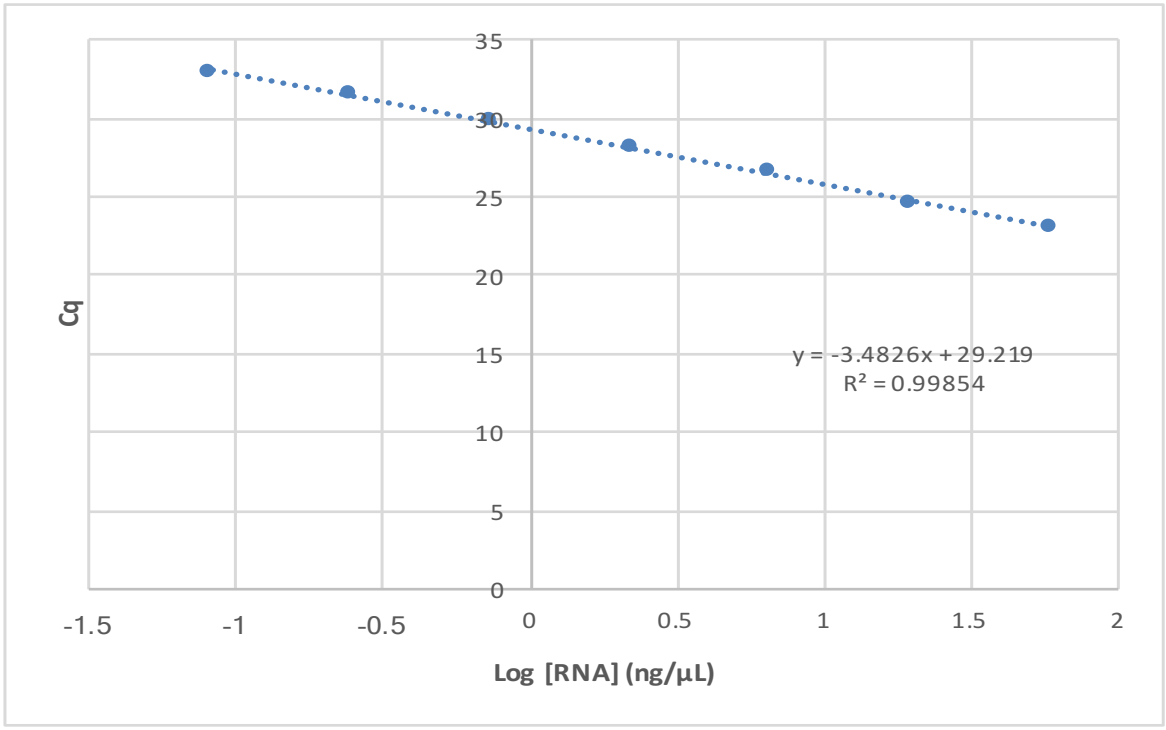


Figure A3. 7: ***Standard curve for PRM1 TaqMan® assay****. Using 1:3 dilution series of semen sample cDNA in concentration from 60 ng/µL to 0.027 ng/µL of starting RNA.*

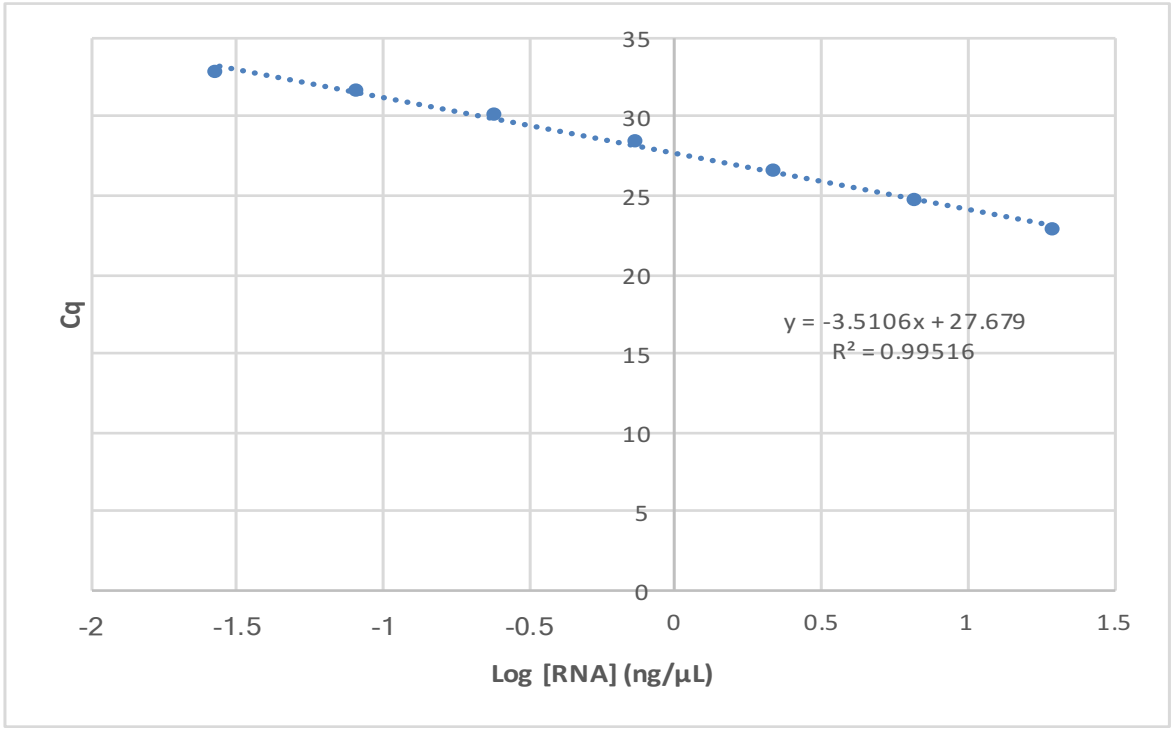


Figure A3. 8: ***Standard curve for PRM2 TaqMan® assay****. Using 1:3 dilution series of semen sample cDNA in concentration from 60 ng/µL to 0.027 ng/µL of starting RNA.*

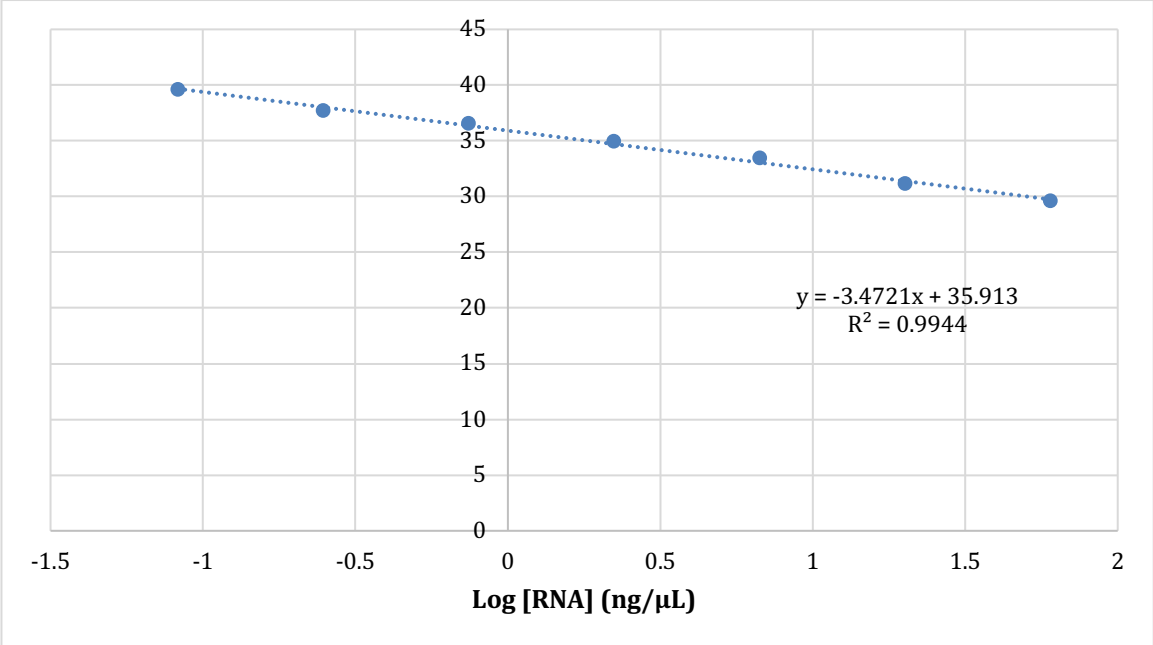


Figure A3. 9: ***Standard curve for SEMG1 TaqMan® assay****. Using 1:3 dilution series of semen sample cDNA in concentration from 60 ng/µL to 0.027 ng/µL of starting RNA.*

A 4: ***Quality control in real-time PCR****. TaqMan® MicroRNA Assays validations and standard curve.*

The following charts illustrate the standard curve of each TaqMan® Gene Expression Assay that has been validated in Chapter 3 of this report. All TaqMan® assays were purchased from Life Technologies.

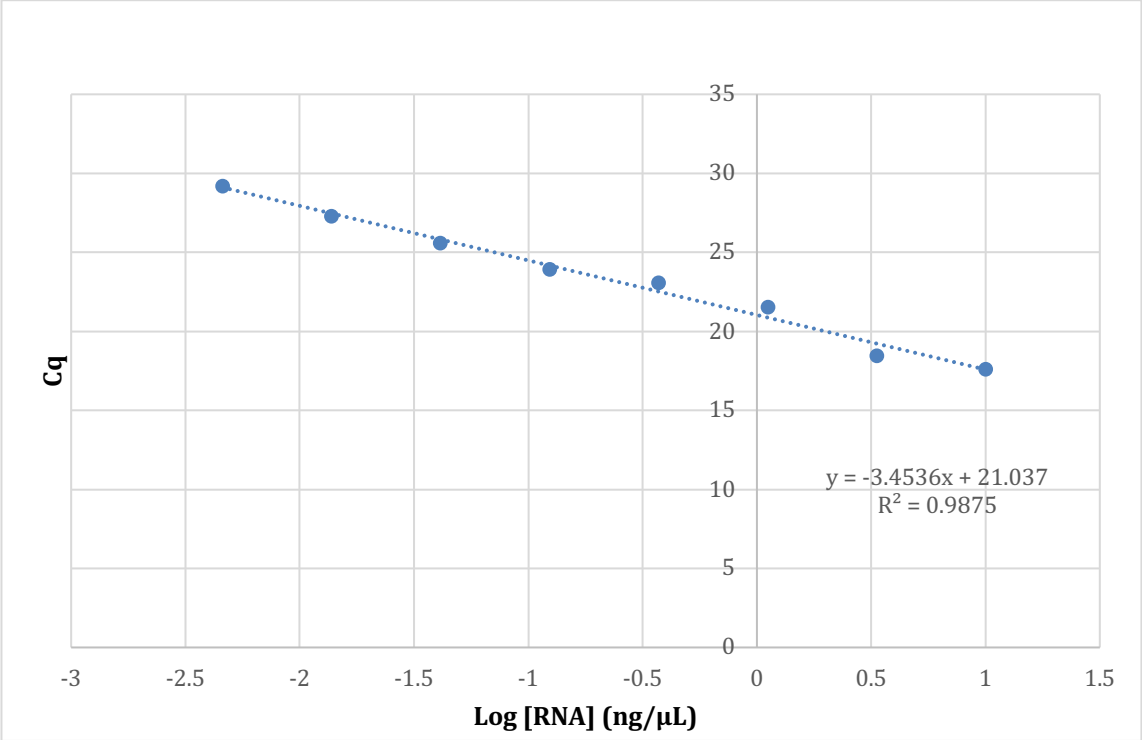


Figure A4. 1: ***Standard curve for miR16 TaqMan® MicroRNA Assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA.*

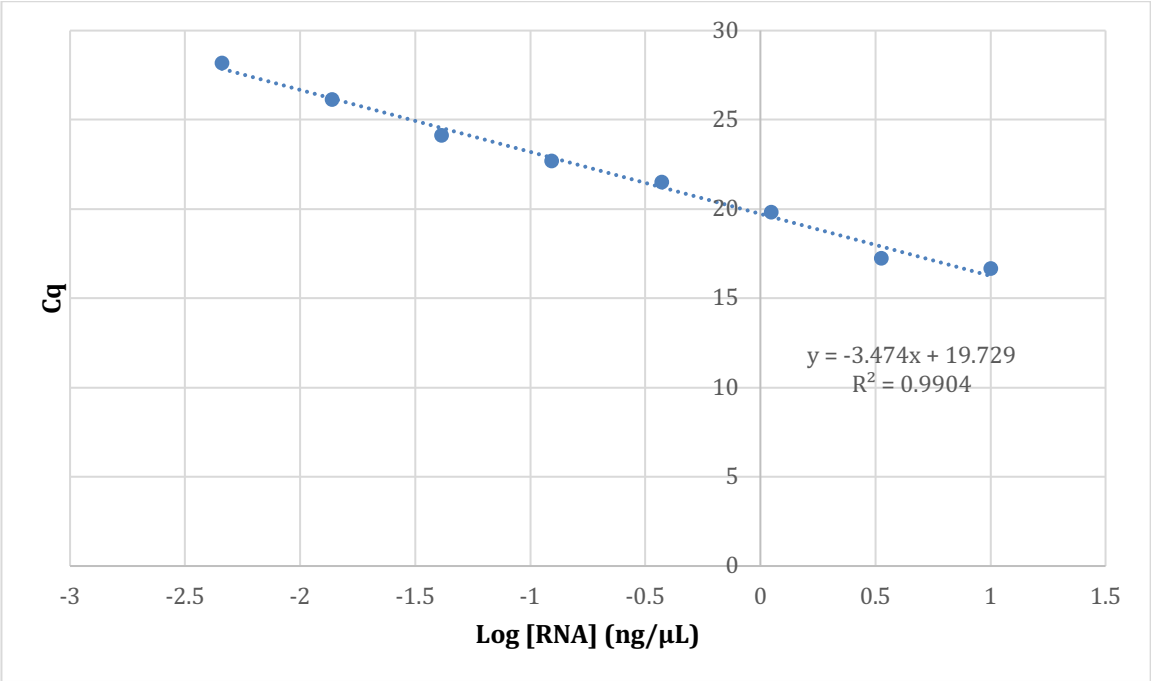


Figure A4. 2: ***Standard curve for miR451 TaqMan® MicroRNA Assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA.*

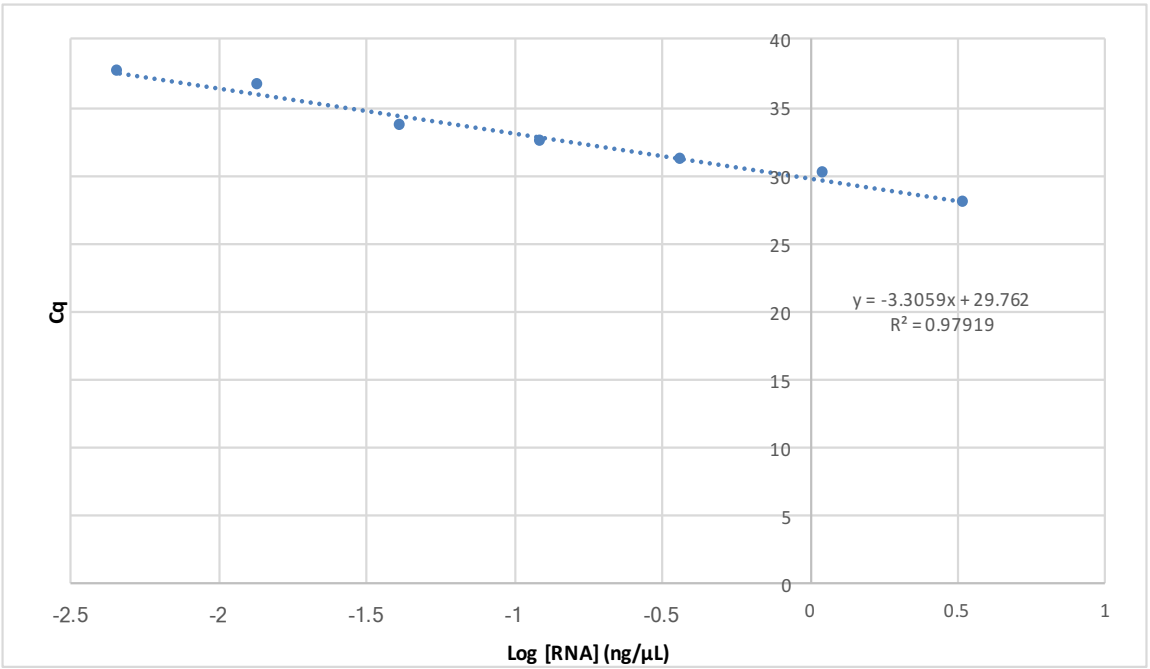


Figure A4. 3: ***Standard curve for miR205 TaqMan® MicroRNA Assay****. Using 1:3 dilution series of saliva sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA*

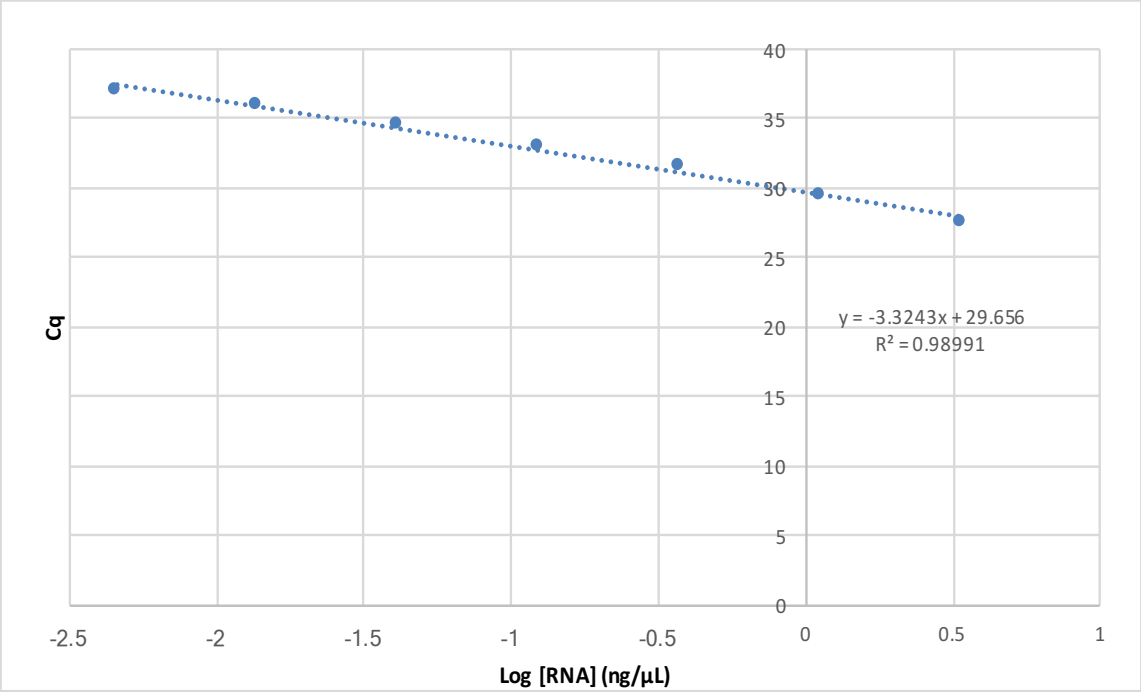


Figure A4. 4: ***Standard curve for miR10b TaqMan® MicroRNA Assay****. Using 1:3 dilution series of semen sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA.*

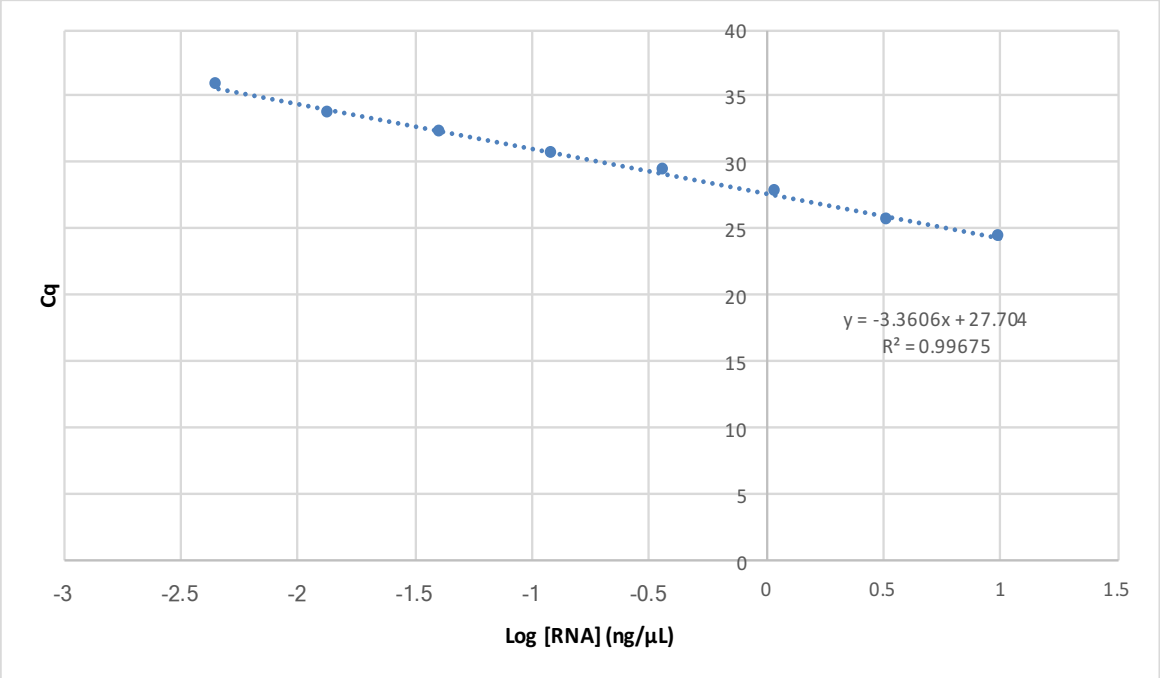


Figure A4. 5: ***Standard curve for miR891a TaqMan® MicroRNA Assay****. Using 1:3 dilution series of semen sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA.*

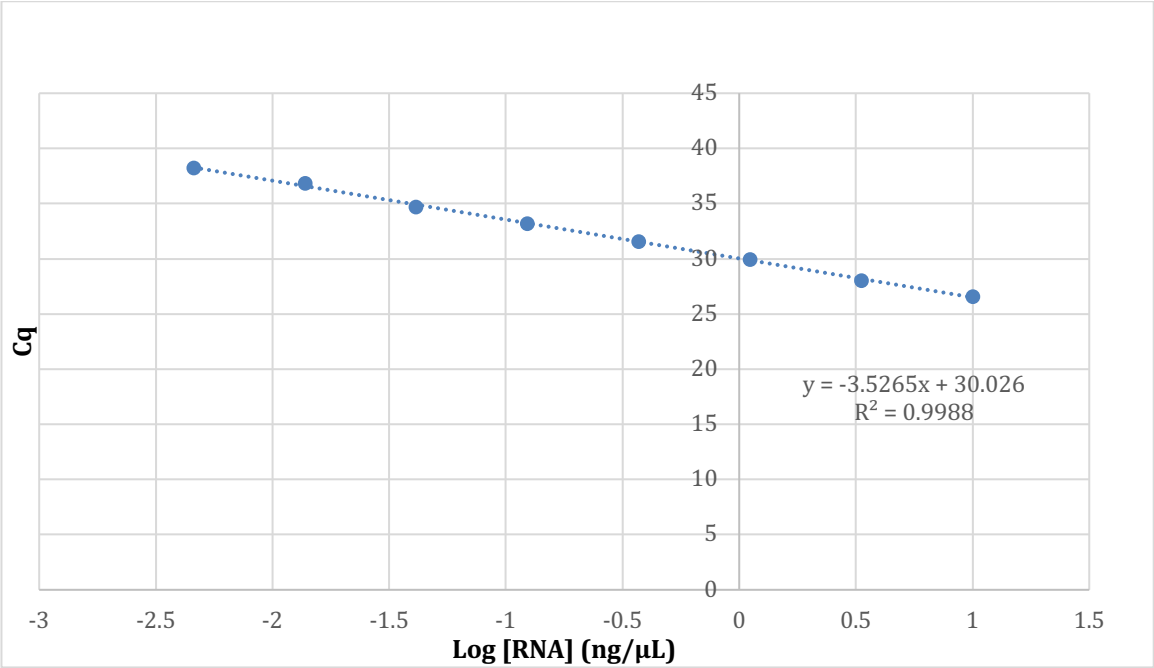


Figure A4. 6: ***Standard curve for U6 TaqMan® MicroRNA Assay****. Using 1:3 dilution series of blood sample cDNA in concentration from 10 ng/µL to 0.004 ng/µL of starting RNA.*

A 5 ***Intra-assay variation.*** *The row C_q results were obtained from repeated quantification of a target from the same cDNA sample within the same qPCR plate and the same master mix.*

Sample	ACTB	HBB	HMBS	HBA	STATH	PRM1	PRM2	SEMG1	18S	miR16	miR451	miR205	miR10b	miR891a	U6
1	23.77	16.22	35.12	21.23	34.64	30.64	24.51	29.54	14.1	15.73	14.39	25.96	25.90	23.23	27.76
2	23.48	16.00	34.38	20.64	33.76	30.50	24.55	29.22	14.15	15.44	14.08	26.63	25.38	23.17	27.18
3	23.32	16.17	34.51	20.43	33.95	30.24	24.59	29.94	14.22	15.41	14.29	25.65	25.21	23.27	26.93
4	23.31	16.20	34.89	20.36	34.91	30.17	25.24	29.76	13.71	15.34	14.38	26.28	25.40	23.13	26.94
5	23.44	15.80	34.55	19.99	34.27	30.29	25.32	30.34	13.65	15.53	14.19	25.76	25.48	23.34	27.04
6	23.31	15.79	33.91	19.72	34.72	30.14	24.61	29.57	13.83	15.43	13.97	25.98	25.79	23.41	27.33
7	23.34	16.04	34.63	19.89	33.89	30.27	24.20	30.14	13.94	15.42	14.16	26.11	25.95	24.00	26.98
8	23.4	16.68	34.94	21.32	33.39	30.78	25.57	30.74	14.04	15.46	14.52	26.38	26.10	24.12	27.68
Mean C_q	**23.42**	**16.11**	**34.62**	**20.45**	**34.19**	**30.38**	**24.82**	**29.91**	**13.96**	**15.47**	**14.25**	**26.09**	**25.65**	**23.46**	**27.23**
SD	**0.16**	**0.28**	**0.38**	**0.59**	**0.53**	**0.23**	**0.48**	**0.49**	**0.21**	**0.12**	**0.18**	**0.33**	**0.32**	**0.38**	**0.33**
CV%	**0.66**	**1.76**	**1.09**	**2.90**	**1.56**	**0.77**	**1.95**	**1.64**	**1.50**	**0.76**	**1.27**	**1.25**	**1.26**	**1.63**	**1.22**

A 6 ***Inter-assay variation.*** *The row C_q results were obtained from repeated quantification of a target from the same cDNA sample across different qPCR plates and different master mix.*

Sample	ACTB	HBB	HMBS	HBA	STATH	PRM1	PRM2	SEMG1	18S	miR16	miR451	miR205	miR10b	miR891a	U6
1	24.54	16.33	35	22.33	33.82	31.92	24.96	29.67	14.63	17.48	16.41	25.12	25.46	24.77	28.40
2	24.64	16.65	35.94	23.94	33.50	31.88	25.55	29.63	14.91	17.5	16.87	25.03	25.46	24.25	28.04
3	24.78	17.16	36.25	24.07	34.96	31.58	24.37	30.42	14.79	17.67	16.77	26.06	25.43	24.25	28.35
4	24.39	17.25	36.54	22.7	33.85	31.66	25.62	29.45	15.07	16.71	16.68	25.33	26.55	25.14	28.99
5	24.24	16.37	34.85	23.05	33.31	31.73	25.13	29.15	15.69	17.14	16.35	25.49	26.55	24.80	28.01
6	23.91	16.50	35.23	22.57	33.81	32.24	25.07	29.32	14.39	17.19	16.75	25.57	25.71	25.30	28.63
7	24.22	16.71	35.86	22.5	33.16	31.45	25.53	30.19	14.86	17.65	16.52	26.50	25.77	25.07	28.70
8	24.76	16.78	35.98	23.24	33.84	32.62	25.30	29.56	15.12	17.65	16.72	25.89	26.31	24.56	29.00
Mean C_q	**24.44**	**16.72**	**35.71**	**23.05**	**33.78**	**31.89**	**25.19**	**29.67**	**14.93**	**17.37**	**16.63**	**25.62**	**25.91**	**24.77**	**28.52**
SD	**0.30**	**0.34**	**0.61**	**0.66**	**0.55**	**0.38**	**0.41**	**0.43**	**0.39**	**0.34**	**0.19**	**0.50**	**0.49**	**0.40**	**0.38**
CV%	**1.24**	**2.03**	**1.71**	**2.86**	**1.62**	**1.20**	**1.63**	**1.45**	**2.58**	**1.94**	**1.12**	**1.95**	**1.89**	**1.60**	**1.35**