



# University of HUDDERSFIELD

## University of Huddersfield Repository

Asaghiar, Fisal E.

Application of forensic RNA analysis as a method for body fluid stain age prediction

### Original Citation

Asaghiar, Fisal E. (2019) Application of forensic RNA analysis as a method for body fluid stain age prediction. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/34926/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: [E.mailbox@hud.ac.uk](mailto:E.mailbox@hud.ac.uk).

<http://eprints.hud.ac.uk/>

# **Application of forensic RNA analysis as a method for body fluid stain age prediction**

---

**A thesis submitted to the University of Huddersfield in Partial Fulfilment of the  
Requirements for the Degree of Doctor of Philosophy**

**Fisal Emhmed Asaghiar**

School of Applied Sciences

**UNIVERSITY OF HUDDERSFIELD**

April – 2018

# Abstract

The basic questions that must be answered during the crime investigation is who left the biological evidence and when. DNA profiling can, in most cases, successfully identify the person who deposited the sample, leaving, therefore, as the main concern question about the length of time between deposition of the stain and its subsequent recovery. The research in the present thesis is concerned with development of the mRNA and miRNA analysis for correct assessment of the age of blood, saliva, and semen samples. It is widely accepted that the level of RNA in the sample decreases over time. Therefore, reverse transcription quantitative polymerase chain reaction (RT-qPCR) method was performed to quantify the selected markers and investigate how they degrade as a function of time. Single and multiple regression analysis were employed in data analysis, which suggested that some of tested markers could be used to predict the stain age. Human specific markers for blood, saliva and semen, as well as oxygen regulated factors such as *vascular endothelial growth factor A* (VEGFA) and *hypoxia inducible factor 1A* (HIF1A) were therefore investigated using TaqMan and SYBR Green chemistries. The predictive equations were derived to determine the age of an unknown sample. Linear regression analysis using relative quantification (RQ) and cycle quantification (C<sub>q</sub>) of the primers revealed the strongest linearity for HIF1A and VEGFA in saliva samples. MicroRNA markers were also explored by targeting miRNA 451, 205, and 891a for blood, saliva and semen samples, respectively, and it was shown that the targets were successfully detected in the samples that were up to 28 days old. Finally, upon development of the predictive models, blind testing was carried out. In blind blood samples, C<sub>q</sub> of selected primers decreased over time and gave accurate prediction of samples' age.

## *Dedication*

*To my Father, mother, beloved wife and children  
who always witness my progress, of which I am  
sure they would be very proud.*

## **Acknowledgments**

I would like to thank everyone that has been involved in this research throughout the last four years without whom I would not be able to get to this stage.

First of all, I would like to thank the Libyan government for funding my studies since 2014, without which this research would have never been possible. A huge thank you to Professor Graham Williams who has been the best and supportive supervisor I could have ever wished for thank you for believing in me and for always keeping your door open. I would also like to thank Dr Stefano Vanin and Dr Douglas Clarke for their guidance throughout the duration of this project.

A huge thank you also goes to Dr Ana Vitlic for her help and feedback. A big high-five to my friends and colleagues in the lab, the volunteers for various body fluids, without whom this work would have not been possible. Finally, a special thanks to my family, my mother and my father for their never wavering support and encouragement.

## List of abbreviations

AAR	Aspartic acid racemisation
ACTB	<i>Actin beta</i>
AFM	Atomic force microscopy
ALAS2	Aminolevulinate delta synthase
ANOVA	Analysis of variance
AQ	Absolute quantification
Ago	Argonaut protein
Bp	Base pair
BFID	Body fluid identification
<i>β-actin</i>	<i>Beta-actin</i>
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
cDNA	Complementary DNA
CHS	Chalcone synthase
CI	Confidence intervals
CP	Crossing point
Cq	quantification cycle
Ct	Cycle threshold
DFR	Dihydroflavonol 4 reductase
dNTPs	Deoxyribonucleotide triphosphates
DNA	Deoxyribonucleic acid
dsDNA	Double strand DNA
DTT	Dithiothreitol
DV	Dependent variable
EPAS1	endothelial PAS domain protein 1
ELISA	Enzyme-linked immunosorbant assay
EPR	Electron paramagnetic resonance
FASN	Fatty acid synthase
FGB	<i>Fibrinogen beta chain</i>
FN1	<i>Fibronectin 1</i>
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEE	Generalised estimating equations
GOI	Gene of interest
G6PD	Glucose-6-phosphate dehydrogenase

GUSB	<i>Glucuronidase Beta</i>
GYPA	<i>Glycophorin A</i>
Hb	Hemoglobin
HBB	<i>Haemoglobin beta chain</i>
HbO <sub>2</sub>	Oxy-haemoglobin
HIF1A1	<i>Hypoxia-inducible factor 1</i>
HLPC	High-performance liquid chromatography
HMGB1	High mobility group box-1
HPRT	Hypoxanthine-guanine phosphoribosyl transferase
HTN3	<i>Histatin 3</i>
IV	Independent variable
MAD	Mean absolute deviation
MetHb	Methaemoglobin
MIQE	Minimum information for publication of quantitative real-time PCR experiments
KLK	<i>Kallikrein</i>
MMLV	Moloney murine leukaemia virus
MMP	Matrix metalloproteinase
MRA	Multiple regression analysis
mRNA	Messenger RNA
miRNA	MicroRNA
MUC7	<i>Mucin 7</i>
NF	Normalisation factor
Nt	Nucleotide
Oligo-dT	Oligo-deoxythymine
PABP	Poly adenylate binding protein
PCR	Polymerase chain reaction
PI	Prediction intervals
PMI	Post mortem interval
Poly A	Polyadenylate
Pre-miRNA	Precursor miRNA
Pri-miRNA	Primary miRNA
PRM 1	<i>Protamine 1</i>
PP2A	Protein phosphatase 2A

PSA	Prostate specific antigen
qPCR	Quantitative polymerase chain reaction
Q-Q	Quantile-Quantile
RISC	RNA inducing silence complex
RBC	Red blood cells
RG	Reference gene
RPLP0	Ribosomal protein large P0
RNA	Ribonucleic acid
RNAi	RNA interference
RQ	Relative quantification
RT	Reverse transcription
RT-PCR	Reverse transcription- polymerase chain reaction
SEMG1	<i>Semenogelin 1</i>
siRNA	short interfering RNA
SDS	Sodium dodecyl sulphate
SI	Stem-loop
ssDNA	Single strand DNA
ssRNA	Single strand RNA
STATH	<i>Statherian</i>
STR	Short tandem repeats
<i>Taq</i>	<i>Thermus aquaticus</i>
TFRC	Transferrin receptor
TGM	Transglutaminase
T <sub>m</sub>	Melting temperature
TSD	Time since deposition
UTR	Untranslated region
UV	Ultra violet light
VEGFA	<i>Vascular endothelial growth factor A</i>
WBC	White blood cell

## Table of Contents

Acknowledgments .....	III
List of abbreviations .....	IV
Table of Contents .....	VII
List of figures.....	XI
List of Tables .....	XIV
Chapter One: Introduction.....	
1.1. Gene Expression Analysis - Messenger RNA.....	2
1.2. mRNA decay pathways in mammalian cells.....	3
1.3. Quantitative PCR (qPCR) .....	7
1.3.1. Types of Real-time Quantification.....	8
1.3.1.1. Absolute quantification.....	9
1.3.1.2. Relative Quantification.....	9
1.3.1.2.1. Normalisation.....	10
1.3.1.2.1.1. Normalisation sample size .....	11
1.3.1.2.1.2. Normalisation reference genes .....	11
1.3.1.2.1.3. Selection of internal control and calibrator for $2^{-\Delta\Delta Cq}$ method.....	12
1.3.1.2.1.4. Geometric mean normalisation .....	13
1.4. Forensic body fluids .....	14
1.4.1. Blood .....	14
1.4.1.1. Previous Techniques Developed to Determine the Blood Age .....	15
1.4.1.2. High-Performance Liquid Chromatography (HPLC).....	16
1.4.1.3. Aspartic acid reaction .....	18
1.4.1.4. Reflectance spectroscopy.....	18
1.4.1.5. Oxygen Electrodes.....	18
1.4.1.6. Atomic Force Microscopy (AFM).....	19
1.4.1.7. Electron paramagnetic resonance (EPR).....	19
1.4.1.8. Enzyme-Linked Immunosorbent Assay (ELISA) .....	19
1.4.1.9. Immunoelectrophoresis .....	20
1.4.1.10. Forensic RNA analysis.....	21
1.4.2. Saliva .....	26
1.4.2.1. Previous studies on saliva using mRNA .....	26
1.4.2.2. Stable RNA markers of saliva stains .....	27

1.4.3. Semen .....	29
1.5. MicroRNA.....	32
1.5.1. Discovery of miRNA.....	34
1.5.2. Micro RNAs as potential biomarker in forensic science.....	36
1.5.3. Micro RNA turnover and degradation.....	41
1.6. Summary .....	42
1.7. Aims of study .....	44
2.1. Ethical approval.....	46
2.2. Sample collection .....	46
2.2.1. Blood samples.....	48
2.2.2. Saliva samples.....	48
2.2.3. Semen samples.....	48
2.3. RNA extraction.....	49
2.4. DNase digestion.....	49
2.5. Quantification and purity assessment of RNA .....	50
2.6. Standard reverse transcription .....	50
2.7. Stem-loop reverse transcription.....	51
2.8. Quantification PCR analysis using TaqMan probe .....	54
2.9. Real-Time PCR analysis using SYBR Green .....	55
2.10. Quantification PCR of stem-loop RT products .....	56
2.11. Data analysis.....	56
Chapter Three: Optimisation and validation of the experiment .....	58
3.1. Introduction .....	59
3.2. Experimental design .....	60
3.3. Influence of different reagents on qRT-PCR experiment .....	60
3.3.1. Reverse transcription-priming strategy .....	60
3.3.2. Assays specificity and sensitivity .....	63
3.3.3. DNA digestion reagents .....	66
3.3.4. Influence of DTT on semen extraction.....	67
3.4. Optimisation of primers .....	70

3.5. Validation of reference genes for normalising RNA in blood samples .....	72
3.6. Validation of $\Delta\Delta Cq$ experiment .....	74
3.7. Discussion .....	78
3.8. Conclusion.....	82
Chapter Four: Degradation of mRNA as an indicator to predict the age of biological stains	
using TaqMan Markers .....	84
4.1. Introduction .....	85
4.2. The problem under investigation.....	86
4.3. Experimental design .....	86
4.4. Results.....	87
4.4.1. Quantification of RNA degradation in one-week-old samples.....	87
4.4.2. Degradation of RNA during one-month period.....	96
4.4.3. Age prediction using TaqMan chemistry .....	96
4.4.3.1. Blood samples .....	97
4.4.3.2. Saliva samples .....	102
4.4.3.3. Semen samples .....	106
4.4.4. Blind samples.....	119
4.4.4.1. Blind blood samples.....	119
4.4.4.2. Blind semen samples.....	121
4.5. Discussion .....	124
4.6. Conclusion.....	127
Chapter Five: Degradation of mRNA as an indicator to predict the age of biological stains by	
using SYBR-Green chemistry. ....	128
5.1. Introduction .....	129
5.2. Experimental design .....	130
5.3. Results.....	130
5.3.1. Blood markers.....	130
5.3.2. Oxygen regulated factors.....	137
5.3.2.1. Hypoxia markers in blood.....	137
5.3.2.2. Hypoxia markers in saliva.....	139
5.3.2.3. Hypoxia markers in semen.....	141

5.3.3. Vascular Endothelial Growth Factors A (VEGFA) .....	143
5.3.4. Hypoxia Inducible Factor 1 Alpha Subunit (HIF1A).....	147
5.3.5. HIF1A with VEGFA .....	151
5.4. Blind samples with SYBR Green .....	159
5.4.1. Blind blood samples .....	159
5.4.2. Blind saliva sample .....	160
5.4.3. Blind semen samples .....	164
5.5. Discussion .....	169
5.6. Conclusion.....	172
Chapter Six: Exploring the degradation pattern of miRNA in biological stains to estimate time since deposition .....	173
6.1. Introduction .....	174
6.2. Experimental design .....	175
6.3. Results.....	175
6.3.1. The patterns of $\Delta$ Ct in microRNA markers as an indicator to predict the age of a stain .....	182
6.3.2. The patterns of RQ of microRNA markers as an indicator to predict the age of a stain .....	185
6.3.3. Geometric mean normalisation strategy.....	191
6.4. Discussion .....	200
6.5. Conclusion.....	204
Chapter Seven: General Discussion.....	205
7.1. Novel work.....	214
7.2. Future work .....	214
Chapter Eight: References.....	216
Appendix 1. Age prediction .....	234
Appendix 2. Blind samples .....	237
Appendix 3. Models generate with geometric mean normalisation .....	240
Appendix 4. Pearson's correlation.....	241
Appendix 5. Normal disruption of RQ of ALAS2/B2M .....	246

## List of figures

Figure 1. Pattern of two mRNA degradation pathways in mammalian cells .....	6
Figure 2. Different methods conducted to determine the age of bloodstains.....	21
Figure 3. The biogenesis of microRNA in eukaryotic cells.....	33
Figure 4. Overview of all miRNA suggested as possible body fluid biomarkers. ....	38
Figure 5. Biochemical pathways for miRNA decay.....	42
Figure 6. Amplifications of ALAS2 and GAPDH in blood samples with different RT primers. .....	61
Figure 7. Comparison between ALAS2 and GAPDH with different RT kits.....	63
Figure 8. Amplifications of three blood markers HBB, GYPA and ALAS2.....	64
Figure 9. Assays specificity in a fresh blood sample using different cDNA synthesis targeting ALAS2, B2M, GAPDH, and ACTB.....	65
Figure 10. Amplification of assays in different templates .....	66
Figure 11. $\Delta Cq$ for primers using different DNA digestion kits.. .....	67
Figure 12. Quantification PCR data using TaqMan primers and probe in fresh and one-week old semen samples.. .....	68
Figure 13. Comparison between two extraction methods of miRNA markers.. .....	69
Figure 14. The melting curve of the amplification of ACTB in blood using SYBR Green.. ..	71
Figure 15. Delta Cq values of TaqMan assays in fresh and one-week old blood samples. ....	73
Figure 16. Box plot analysis of reference genes in fresh blood samples.....	74
Figure 17. Validation plot of $\Delta Cq$ vs.log input amount of RNA.....	76
Figure 18. Validation of the experiment by plotting $\Delta Cq$ vs. log input amount of RNA.....	78
Figure 19. The degradation patterns of ALAS2, GAPDH, and B2M in blood samples using TaqMan chemistry .....	88
Figure 20. Normal Q-Q plot of (A, B, and C) using $\Delta Cq$ of ALAS2, B2M, and GAPDH in blood respectively. ....	91
Figure 21. Simple regression analysis of chosen markers in blood samples over a one-week period.....	93
Figure 22. The age prediction plotted against actual age.....	95
Figure 23. Amplification of ALAS2/HBB and B2M in blood using TaqMan chemistry. ....	98
Figure 24. Regression analysis to obtain models in blood samples using TaqMan chemistry... .....	100
Figure 25. The mean predicted age fitted with actual age in blood models.....	101

Figure 26. $\Delta Cq$ of HTN3 in saliva samples using TaqMan probe.....	103
Figure 27. Regression analysis in saliva .....	105
Figure 28. Age prediction generated by (A) $\Delta Cq$ HTN3 model (B) RQ of HTN3/GAPDH model.....	106
Figure 29. Semen markers with DTT (+DTT). .....	108
Figure 30. Semen markers without DTT (-DTT) over 28 days period.....	109
Figure 31. Simple regression analysis using semen TaqMan markers. ....	110
Figure 32. Age prediction that was obtained in semen (+DTT) with TaqMan chemistry. ....	112
Figure 33. The age prediction in semen without DTT extraction (-DTT). .....	113
Figure 34. Age prediction using multiple regression models.....	116
Figure 35. Mean age prediction calculated using multiple regression models.....	118
Figure 36. Relative quantification (RQ) of blind blood samples using TaqMan chemistry. .	120
Figure 37. Blind semen samples predicted using TaqMan markers .....	123
Figure 38. Delta Cq of blood markers using SYBR Green chemistry over 28 days period..	131
Figure 39. RQ in blood samples over 28 days using markers (A) HBB, (B) FN1 and (4C) EPOS1.....	132
Figure 40. Regression analysis of HBB/ACTB in blood using SYBR Green. ....	133
Figure 41. Age prediction in blood samples using the models obtained .....	136
Figure 42. The degradation patterns of hypoxia markers using SYBR Green in blood samples. ....	138
Figure 43. The degradation patterns of hypoxia markers using SYBR Green in saliva samples. ....	140
Figure 44. Relative quantification in semen samples.. .....	142
Figure 45. Simple regression analysis of VEGFA marker with ACTB as reference gene... ..	144
Figure 46. Age prediction obtained with VEGFA/ACTB model in (A) blood, (B) saliva, and (C) semen samples.. .....	146
Figure 47. Scatterplots for RQ of HIF1A/ACTB against predicted age.....	148
Figure 48. Age prediction generated from HIF1A/ACTB model.....	150
Figure 49. Simple regression analysis of HIF1A with VEGFA as reference gene. ....	153
Figure 50. Age prediction using HIF1A/VEGFA model. ....	154
Figure 51. Multiple regression analysis using hypoxia models.....	156
Figure 52. Age prediction derived from multiple regression analysis.....	158
Figure 53. Relative quantification obtained by using HBB/ACTB model in blind blood sample with SYBR Green chemistry .....	159

Figure 54. The blind blood samples using hypoxia models. ....	162
Figure 55. Saliva blind sample age prediction. ....	164
Figure 56. Blind semen samples using models obtained from hypoxia markers. ....	167
Figure 57. Age prediction using hypoxia models in blind blood, saliva, and semen samples. .....	168
Figure 58. MiRNA in saliva samples.....	177
Figure 59. MiRNA in blood samples.....	179
Figure 60. MiRNA in semen samples with DTT (+DTT).....	180
Figure 61. MiRNA in semen samples without DTT (-DTT), aged up to 28 days.....	181
Figure 62. The mean age prediction using miRNAs markers in saliva. ....	184
Figure 63. The mean RQ of selected miRNA aged to up 28 days.....	186
Figure 64. Age prediction obtained, using SRI. ....	189
Figure 65. The mean age prediction using miRNA models. ....	190
Figure 66. Geometric mean normalisation strategy in saliva samples.....	192
Figure 67. The mean age prediction in saliva samples.. ....	193
Figure 68. Geometric mean normalisation in semen samples.. ....	195
Figure 69. Mean age prediction in semen samples.. ....	196
Figure 70. The new model generated with geometric mean in blood samples. ....	198
Figure 71. New normalisation with GYPA in the blood samples .....	199

## List of Tables

Table 1. Summary of using mRNA to ageing biological stains. ....	25
Table 2. Messenger RNA markers used to identify semen in human. ....	31
Table 3. Information of the samples investigated using TaqMan and SYBR Green chemistry. .....	47
Table 4. MicroRNA primers using in blood, saliva, and semen. ....	53
Table 5. TaqMan probe used for blood, saliva, and semen. ....	54
Table 6. Unlabelled assays for blood and saliva. ....	55
Table 7. Rule of Thumb to interpret the level of a Correlation Coefficient. This table as described by (Mukaka, 2012). ....	57
Table 8. Normality test for Random primers using GAPDH and ALAS2 in blood samples...	62
Table 9. Test of Normality of Delta Cq ( $\Delta Cq$ ) of blood markers in one-week period. ....	90
Table 10. Test of Normality of RQ for blood markers during one-week period. ....	90
Table 11. Simple regression analysis model generated in blood during 7 days. ....	92
Table 12. Summary of models obtained in blood using a simple regression analysis for 28 days period. ....	99
Table 13. Saliva models obtained using simple regression analysis up to 28 days. ....	104
Table 14. Summary of semen models obtained using simple regression analysis over a 28 days period. ....	111
Table 15. Summary of models generated by using multiple regression analysis in blood and semen sample. ....	114
Table 16. Blind blood samples with actual age used for the validation process. ....	119
Table 17. Age prediction of blind blood using the models obtained. ....	121
Table 18. Blind semen samples with TaqMan models. ....	122
Table 19. Summary of the models obtained for blood samples using SYBR Green chemistry. .....	134
Table 20. Models obtained with VEGFA marker in blood, saliva, and semen. ....	143
Table 21. Models obtained with HIFA1/ACTB in blood, saliva, and semen. ....	147
Table 22. The summary of all models obtained by single regression analysis in blood, saliva, and semen samples using SYBR Green chemistry. ....	152
Table 23. Multiple regression for models in blood, semen, and saliva samples using SYBR Green chemistry. ....	157
Table 24. Age prediction of blind blood samples using HBB/ACTB model. ....	159

Table 25. Age prediction for blind saliva samples using hypoxia models.....	163
Table 26. Semen age prediction using SYBR Green models.....	166
Table 27. Models were generated in saliva using Ct forms as indicators.....	183
Table 28. Age prediction model using multiple regression analysis in saliva samples. ....	183
Table 29. The simple regression analysis using RQ of miRNA markers in blood, saliva and semen samples. ....	188
Table 30. Mean age prediction calculated with all mRNA and miRNA models obtained using TaqMan and SYBR-GREEN chemistries. ....	211
Table 31. The mean age prediction in blood using TaqMan chemistry among 1 week. ....	234
Table 32. The mean age prediction in semen (-DTT) using TaqMan. ....	234
Table 33. The mean age prediction using TaqMan chemistry in blood, saliva, and semen among 1 month. ....	235
Table 34. The mean age prediction in blood using SYBRGREEN chemistry among 1 month. .....	235
Table 35. The mean age prediction using SYBRGREEN chemistry in blood, saliva, and semen among 1 month. ....	236
Table 36. The mean age prediction calculated by using miRNA models in blood, saliva, and semen among 1 month. ....	236
Table 37. Age predicted of blind sample using single and multiple regression models with ALAS2/B2M and HBB/B2M. ....	237
Table 38. Blind blood samples using hypoxia models.....	238
Table 39. Blind saliva samples using hypoxia models. ....	238
Table 40. Prediction the blind samples using TaqMan models in semen. ....	238
Table 41. Age prediction using hypoxia models in semen. ....	239
Table 42. The mean age prediction using geometric mean in blood, saliva, and semen. ....	240
Table 43. Relationship between ALAS2, HBB and B2M in blood among one week.....	241
Table 44. Pearson's correlation calculation in miRNA 205 and 44 in saliva samples. ....	242
Table 45. Correlation between miRNA 891a and 44 (+DTT) in semen samples. ....	243
Table 46. Correlation in blood with miRNA markers. ....	244
Table 47. Semen markers without DTT (-DTT).....	245
Table 48. Multiple regression analysis in blood during one week. ....	246
Table 49. Multiple regression analysis in semen samples during 28 days.....	246

# **Chapter One: Introduction**

## **1. Introduction**

According to Locard's principle, "every contact leaves a trace", so the majority of the evidence is deposited through the interactions between the perpetrator and the victim or the perpetrator and the crime scene. However, it is often difficult to determine if the evidence collected from crime scene was left through the same crime, or deposited during a previous interaction with that environment. It can be very important to accurately determine the link between the time of deposition of the sample and suspect's presence at the crime scene, especially when the suspect has close personal relationships with the victim. One of the famous cases in forensic science, Orenthal James (O.J.) Simpson's case is a clear demonstration of the importance of age determination of bloodstains during the criminal investigation. According to the prosecution, the blood of Nicole Brown Simpson, O.J. Simpson's wife, found in the vehicle owned by O.J. Simpson was a clear proof of his guilt, while the argument of the defence was that it could have been there long before the crime was committed. Since no forensic technique existed capable of accurately determining the time since deposition (TSD) of the sample, arguments from either side could not be accepted or rejected.

Two key characteristics of the forensic material found at the crime scene, most commonly human body fluids such as blood and saliva, are of interest in the forensic science. The first is the person identification, which is easily answered by DNA profiling, and helps with the crime scene reconstruction and more in general with the whole criminal investigation and can be relied upon in court (Ackermann *et al.*, 2010, Jobling and Gill, 2004). The second is the age of the biological sample found at the crime scene. The determination of these samples is considered a challenge in the forensic science field. For example, was the particular bloodstain deposited less than 24 hours ago when the alleged incident occurred, or did that happen 2 months ago when the involved party had an accident and cut his finger?

## 1.1. Gene Expression Analysis - Messenger RNA

The central dogma of molecular biology explains the normal flow of biological information by describing three processes. These include deoxyribonucleic acid (DNA) replication when a new DNA is copied; transcription when the information is transcribed into messenger ribonucleic acid (mRNA); and translation when proteins are being synthesised from the mRNA template. Transcription is a complex process that has several stages: pre-initiation, initiation, promoter clearance, elongation, and termination. In pre-initiation phase, the pre-initiation complex is formed of RNA polymerase and transcription factors at the promoter sequence of DNA, followed by promoter melting, which gives access to a single strand DNA molecule. This is followed by binding of RNA polymerase to the promoter sequence with the help of other transcription factors in the process known as initiation. Promoter clearance is the step where RNA polymerase clears the promoter sequence, but only after first bond has been formed, thus preventing premature release of the transcript (Shandilya and Roberts, 2012). In the elongation stage with presence of elongation factors, RNA polymerase with the help of elongation factors uses DNA as a template and starts to copy gene into mRNA. In termination phase, newly synthesised mRNA molecule is released from the complex. Termination is in eukaryotic organisms tightly linked to pre-mRNA processing which includes capping of 5' end, 3' polyadenylation, and splicing reactions to remove the introns. The 5' cap protects pre-mRNA from degradation and allows ribosome to recognise the beginning of the mRNA. Closely linked to termination of transcription, the poly (A) tail addition further increases stability of RNA during its transport from the nucleus into the cytoplasm.

Messenger RNA inside the single cell accounts for only 1% to 5% of total RNA. On average, each cell contains 20,000 to 30,000 different mRNA species with copy numbers varying from 15 to 12,000 (Fleming *et al.*, 1998).

Although all biological cells conform to the rules of central dogma described above, there are few important exceptions. One of them is the process of reverse transcription, in which mature mRNA is used as template to transfer the information into DNA. Here, the transcription of RNA template into DNA is being catalysed by Reverse Transcriptase Enzyme, RNA-dependent DNA polymerase. All retroviruses such as human immunodeficiency virus (HIV) are able to synthesise DNA from RNA template and create a new single strand DNA complementary to the viral RNA. This process of reverse transcription is widely used in research (Su, 2017).

## **1.2. mRNA decay pathways in mammalian cells**

RNA molecule may form many secondary structures, which are important for its function. This could suggest that RNA has better stability. On the contrary, however, several reasons make RNA less stable comparing to DNA. First, ribose sugar in RNA contains one more hydroxyl group on the second carbon (C2'- OH), which makes it more reactive than deoxyribose. This is because hydroxyl bond is prone to hydrolysis, creating C2'- O ion, which reacts with P atom and breaks phosphodiester bond in the sugar-phosphate backbone, cleaving RNA molecule. In addition, unlike DNA which is double-stranded molecule with relatively small grooves, RNA is single stranded with large grooves that provide easier access to the damaging enzymes such as Ribonuclease that can decompose it easily. Therefore, less energy is required to break down RNA than DNA. Finally, half-life of RNA is relatively short and RNA cannot be preserved for a long time due to the abundance of Ribonuclease enzymes (Su, 2017).

Regulation of mRNA decay can be considered the main controlling point in gene expression process. Interactions between the structural elements of RNA and the proteins that bind it determine the stability of the mRNA, which can be specific for that particular mRNA or more

general. Several decay pathways have been characterised for mRNA (Figure 1), which depend on cellular conditions. The regulated decay of mRNA occurs through interactions between structural components of mRNA, namely 5' -cap structure, 5'-untranslated region (UTR), the protein coding region, 3' -UTR and the 3' - poly (A) tail, and the specific trans-acting factors (Guhaniyogi and Brewer, 2001). Sachs (Sachs, 1990) reported that almost all of the mammalian mRNAs are polyadenylated, which plays an important role in nuclear mRNA processing, cytoplasm exporting, translation and cytoplasmic mRNA stability. This is achieved with the help of the poly (A)-binding protein (PABP), shown to be present in high concentration and in excess over its poly (A) binding sites in cytoplasm, and as such binds most of the poly (A) tails (Görlach *et al.*, 1994). Poly (A)-PABP interaction at the 3' end of mRNA molecule has a role in protecting this molecule from ribonucleases and rapid decay *in vitro*, given that the decay of many mammalian mRNA starts with deadenylation (Sachs, 1990, Ross, 1995, Guhaniyogi and Brewer, 2001). Poly (A) plays a crucial role in the stability of mRNA, and together with the cap structure at 5' end facilitates the process of translation. Overall, two main activities of mRNA are serving as template for translation and substrate for cellular degradation (Roy and Jacobson, 2013), and it is not uncommon for polyadenylated mRNA to be targeted by the deadenylation-dependent mechanism of decay while still being engaged in the process of translation. All RNAs can be classified by their stability in the cell. For example, both transfer and ribosomal RNA (tRNA and rRNA, respectively) are considered to be the most stable RNA forms. On the other hand, mRNAs are unstable with half-life, a turnover rate or time required for degrading 50% of the existing mRNA molecule, in eukaryotic cells usually shorter than the generation time. The general pathway for normal mRNA turnover starts with deadenylation of the 3' poly (A) tail and subsequent decapping of the 5' end (Chen and Shyu, 2011).

The ratio of  $\beta$ -actin (ACTB) mRNA to 18S rRNA was one of the newly described methods to determine the relationship between the age of the blood sample and the degradation of RNA molecule. Both of these RNA molecules are commonly used as reference genes, due to their constitutive expression in all type of cells. However, Anderson *et al.* (2005) found that the cycle quantification (Cq ) value of 18S rRNA remained unchanged over the course of 150 days, while the Cq values for  $\beta$ -actin significantly decreased over time (Anderson *et al.*, 2005). This led to the increase in the relative ratio of 18S rRNA to  $\beta$ -actin mRNA over time.

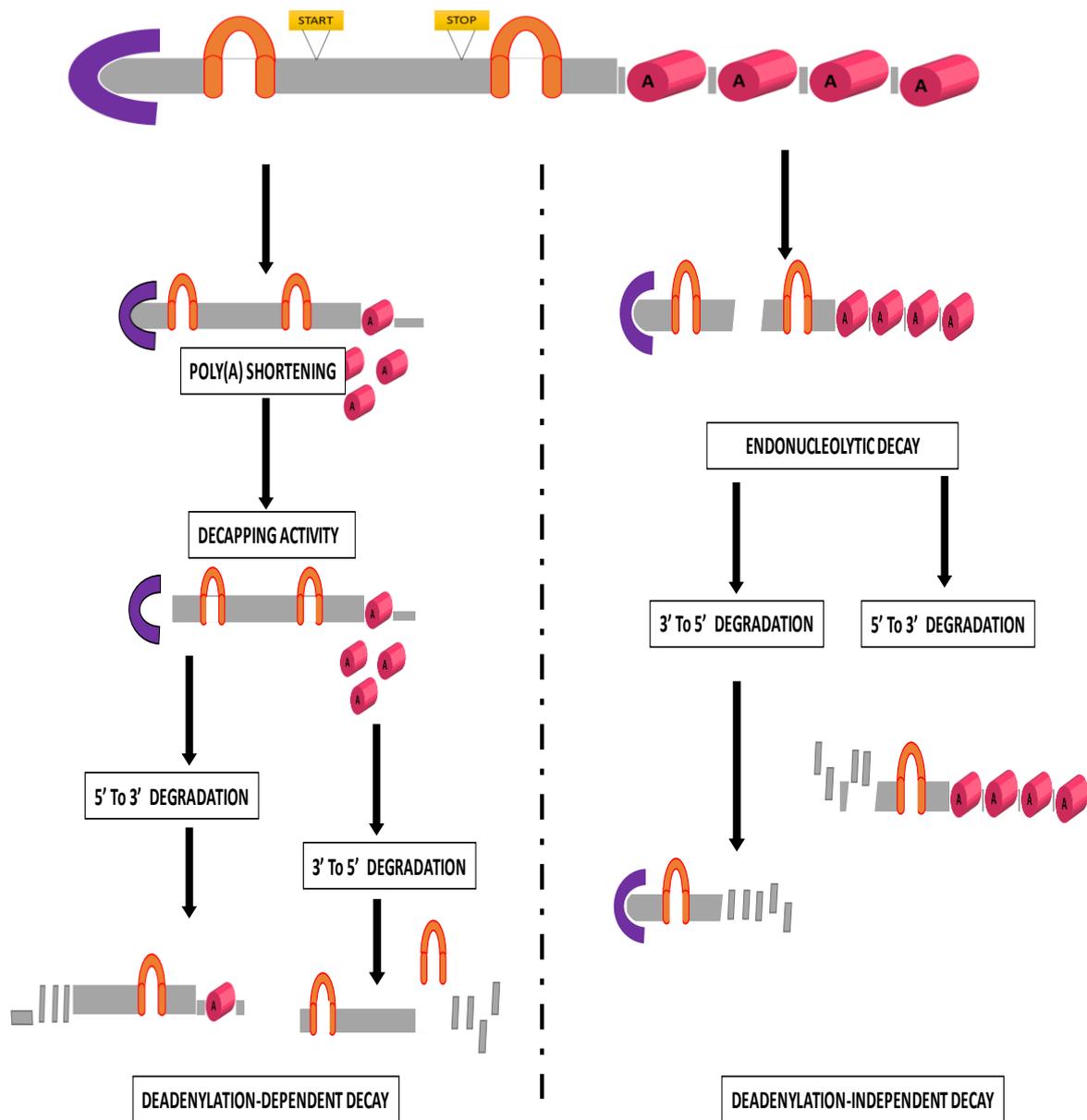


Figure 1. Pattern of two mRNA degradation pathways in mammalian cells. The structure of a mature mRNA molecule consists of a stem-loop and 5-cap structure located at 3 -UTR and 5 -UTR respectively. The starting and stopping codons are both indicated. The left panel describes the major pathway of deadenylation-dependent decay, which shows that decapping occurred after removal of most or all of poly (A) tract. The image on the right side shows the deadenylation-independent decay pathway. Endonucleolytic cleavage of an mRNA generated two products likely to be subjected to exonucleolytic activities. New updating design according to Guhaniyogi and Brewer (2001).

### **1.3. Quantitative PCR (qPCR)**

The process known as polymerase chain reaction (PCR), in which a single or a small number of copies of DNA are amplified generating millions of copies, was first introduced by Kary Mullis in 1985 (Mullis, 1990). For this invention, Mullis won the Nobel Prize in 1993.

Quantitative PCR has become the most important technique to identify and quantify accurately and sensitively RNA and/or DNA. The inherent problems generated by using gels, membrane transfers, radioactive probe hybridization and limitations of film as a detector, are all successfully overcome by using fluorescently labelled primers. In general, there are two problems related to real-time PCR. Firstly, the quantification methods are dependent on the available standards, the integrity of assay and statistical methods used. Second, the correct way needs to be chosen to perform a proper normalisation for different samples to correct the differences between the amounts of RNA for each sample-to-sample input.

The advances in RT-PCR method have obviously influenced the area of quantification of gene expression. Quantitative PCR technique is used to collect the data during the normal PCR run, rather than at the end of PCR, consequently, combining the amplification and detection in a single process (Kubista *et al.*, 2006). This is performed by using a variety of different fluorescent chemistries to detect a specific PCR product as it accumulates during PCR, by relating the amplicon concentration to the fluorescence intensity. The reaction is defined by the cycle value when the amplified target is detected. This point is usually called cycle quantification (Cq) and is defined by the cycle at which fluorescence intensity is greater than background fluorescence. Consequently, the higher target amplification in a sample will exceed the background signal sooner and give lower Cq value (Kubista *et al.*, 2006). The best technique to quantify extracted RNA is qPCR because it produces quantitative data over dynamic range of 7 to 8 log orders of magnitude. Further, in qPCR, post-amplification

analysis is not required and these assays are far more sensitive than some other gene expression analysis, such as RNase protection assays (Morrison *et al.*, 1998).

There are four major steps in qPCR process: linear, early exponential, exponential, and the plateau phase. In linear stage, the first 10-15 cycles, qPCR is starting, and the fluorescence produced during each cycle does not increase over the baseline. During the early exponential phase, fluorescence intensity exceeds the threshold level indicating significant increase in the target molecule. The cycle or point where this change has occurred is known as "crossing point" (Cp), threshold cycle (C<sub>t</sub>), or more adequately now referred to as the quantification point (C<sub>q</sub>) (Bustin *et al.*, 2009). This value is an indicator of an increase in a copy number over original template, and it is used in calculation of experimental results (Heid *et al.*, 1996).

Two common ways to perform RT-qPCR reaction are one-step and two-step methods. One-step method combines both reverse transcription and PCR in a single tube. Two steps method, on the other hand, physically separates process of reverse transcription reaction and PCR amplification. The advantage of one-step qPCR is lower chance of experimental variability, given that the preparation for both reactions occur at the same time. Limitation, however, lies in higher chance of the degradation of RNA molecule used as a template, if not handled properly. Finally, two-step protocol is a method of choice if the samples are expected to be assayed more than once, and if the sensitivity of the assay is imperative (Battaglia *et al.*, 1998).

### **1.3.1. Types of Real-time Quantification**

Reverse transcription followed by PCR is now widely accepted as a successful way of quantification of gene expression. There are two different methods to analyse the data in qPCR experiments: absolute quantification (AQ) and relative quantification (RQ) (Livak and Schmittgen, 2001).

### **1.3.1.1. Absolute quantification**

Absolute quantification (AQ) is normally achieved by preparing serial dilutions of standards molecules of known absolute quantities. These are then used to produce a standard curve. The concept of the standard curve establishes a linear relationship between C<sub>q</sub> and starting amounts of RNA or cDNA (Pfaffl, 2001). The concentration of the unknown samples is then subsequently determined through their C<sub>q</sub> values obtained during the qPCR. This method requires, however, similar amplification efficiencies of standards and samples used. Further, the serial dilutions of the standard must cover all expected ranges for experimental sample they need to be within the range of accuracy quantification.

For standard curve adequate number of dilutions should be prepared to cover the expect range of expression. Therefore, at least a 5-point 10 fold serial standard curve should be prepared, and each dilution tested in duplicate along with positive and negative controls. To control for the pipetting error, the highest and the lowest C<sub>q</sub> value are discarded and average of remaining values is used to calculate final C<sub>q</sub> value. C<sub>q</sub> value is inversely proportional to the logarithm of starting copy numbers, therefore linear regression fit of the standard curve can be used for calculating initial copy numbers and quantification of RNA.

### **1.3.1.2. Relative Quantification**

Relative quantification (RQ), also called comparative quantification is simply defined as changes in gene expression relative to other reference group. The most comprehensive type of relative quantification is so called  $2^{-\Delta\Delta C_t}$  method. Relative quantification does not require the calibration curve, but careful selection of controls, namely calibrator and internal controls instead (Livak and Schmittgen, 2001). Internal controls are reference genes, and when choosing these, care must be taken to maintain high sensitivity of the method, especially in case of studies where subtle physiological changes in gene expression are examined. This is done by adequately assessing and carefully selecting combination of reference genes that are

constantly expressed across all the conditions tested. These can then be co-amplified as endogenous controls in the same reaction tube in a multiplex assay, or as exogenous control when it is amplified in a separate tube or well (Morse *et al.*, 2005). Choice of calibrator for the relative quantification method depends on the type of experiment planned, and it commonly involves untreated control or, as in the case of studies that determine age of biological samples, a fresh sample deposited at zero time point (Livak and Schmittgen, 2001).

#### **1.3.1.2.1. Normalisation**

To obtain an accurate relative quantification result, an appropriate normalisation strategy together with identical cycling conditions during real-time PCR are required to correct for an experimental error (Pfaffl *et al.*, 2004). Additional source of errors are procedures performed prior to the qPCR (i.e. isolation of RNA), as well as RT reaction step and those performed during the PCR setup itself and by cycling process. For precise relative comparison, the normalisation should be performed based on the amount of extracted RNA, on analysed mass of tissue, or a number of cells obtained from a biopsy, cell culture or blood cells used as a source of RNA (Skern *et al.*, 2005). In other words, to ensure that starting material is comparable in concentration and volume, the relative quantification data should be normalised using at least one of the following:

- Amount of the sample, mass/size or volume of the tissue.
- Total amount of isolated RNA
- Reference genes (RG) such as Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), ACTB and  $\beta$ 2 microglobulin (B2M).

#### **1.3.1.2.1.1. Normalisation sample size**

In order to reduce experimental error, similar sample size should be used, such as similar tissue volume or weight. Sometimes it can be difficult to ensure that different samples contain the same amount of cellular material. Fortunately, in case of blood samples it is relatively easy to ensure similar volumes are compared (Huggett *et al.*, 2005).

Further, it is essential to determine accurate quantity and quality of RNA before the reverse transcription step. This is to ensure the similar input amount of extracted RNA during reverse transcription reaction. There are several methods for quantifying RNA, some more accurate than others. For example, ribogreen (Molecular Probes) is considered one of the most sensitive detection dye for determining the amount of RNA in the solution, more accurate than Lab Chip (Agilent 2100). In addition to quantity, RNA integrity should be checked as poor RNA quality can affect downstream measurements. RNA integrity can be determined by running a sample on agarose gel and examining bands for 28S and 18S rRNA. This is the least expensive method although it requires higher quantities of RNA (0.1-1µg) and therefore not suitable if the analysed samples are precious and available in small amounts. Another method to check for integrity that can work with smaller amounts of RNA but is more expensive are bio-analysers, such as Agilent Bioanalyzer.

#### **1.3.1.2.1.2. Normalisation reference genes**

Normalising to RG is a simple and common way of controlling error in RT-qPCR. This procedure is simplified because both the gene of interest and reference genes are measured using RT-qPCR. RG controls for input RNA amounts used in RT step. GAPDH, ACTB, 18S ribosomal RNA, and hypoxanthine-guanine phosphoribosyl transferase (HPRT) are the most commonly used reference genes in Northern blots, RNase protection assays and conventional RT-PCR assays (Riemer *et al.*, 2012). RNA of these genes are expressed at relatively high levels in all cells; this makes them acceptable RG for non/semi-quantitative techniques where

a qualitative change is measured. In other words, they make ideal positive controls if the gene of interest is switched off.

However, in more complex quantitative techniques such as qPCR, care must be taken to ensure that the chosen group of RG is confirmed to be expressed at the similar level in all the conditions used in experiment. For example, Tanaka and co-workers (Tanaka *et al.*, 1975) reported that 18S rRNA increased in expression with cytomegalovirus infection. In 1984, Piechaczyk studied different rat tissues with different amount of mRNA and found that GAPDH transcription occurred at a similar rate (Piechaczyk *et al.*, 1984). On the other hand, Dheda in 2004 validated 13 reference genes including GAPDH and reported that none of them was suitable to be used as a reference gene in blood (Dheda *et al.*, 2004). Another study was stated that GAPDH mRNA was a highly expressed in a skeletal muscle, whereas, it was lower expressed in a breast tissue, with a 15-fold copy numbers is different between both tissues (Barber *et al.*, 2005). On the other hand, profile of nine candidate RG including GAPDH were calculated using four algorithms: geNorm, NormFinder, BestKeeper and the delta C<sub>q</sub> method, and found that GAPDH is the most stable genes (Petriccione *et al.*, 2015).

#### **1.3.1.2.1.3. Selection of internal control and calibrator for $2^{-\Delta\Delta C_q}$ method**

The internal control gene is used to normalise starting amount of input RNA during a reverse transcription PCR reaction. The choice of optimal reference genes used for normalisation is experiment-dependent and each researcher is responsible for determining the best combination by finding those that are consistently expressed across all experimental conditions used. Usually, standard housekeeping genes are appropriate and sufficient as internal control genes. Some examples of internal controls used to successfully normalise real-time quantitative PCR are GAPDH, ACTB,  $\beta$ 2-microglobulin, and 18S rRNA. However, other commonly used reference genes could also be expected to work successfully. The point

is that for the best results, the internal control must be validated for each experiment to ensure that unaffected by experiment treatment are chosen.

The choice of calibrator for the  $2^{-\Delta\Delta C_q}$  method is dependent on the experimental question and the purpose of gene expression analysis, and the most commonly used calibrator is the untreated control. The  $2^{-\Delta\Delta C_q}$  formula presents gene expression as fold change normalised to an endogenous reference gene and relative to an untreated control. Therefore, the result of the formula for untreated control sample which is also used as calibrator, equals one ( $\Delta\Delta C_q = 0$  and  $2^0=1$ ). Similarly, if the aim of the study is to examine the time course of gene expression, then the calibrator should be copy number of transcript expressed at the beginning of experiment, or time zero sample (Livak and Schmittgen, 2001). Finally, this quantification method assumes that the amplification occurs with the same efficiency; therefore, PCR should be performed on dilutions of both target and internal control genes to confirm this. The risk with using reference genes as external standards is that there is no control for internal RT and PCR inhibitors. In other words, some samples may contain substances that affect amplification efficiency in PCR reaction, making it different to that of target gene (Pfaffl *et al.*, 2004).

#### **1.3.1.2.1.4. Geometric mean normalisation**

The normalisation strategy in real time qPCR experiment is an essential step to control for experimental variations. One of the most common methods uses RG, which represents the best possible normaliser. However, the main problem with blindly choosing a single RG as a normaliser is that it does not necessarily have constant expression under all experimental conditions examined, and may depend on the type of samples, amount of starting material, enzymatic efficiencies, and individuals. This problem can be addressed by using multiple RG, confirming, at the same time, their stable expression under experimental conditions of interest

(Vandesompele *et al.*, 2002, Chervoneva *et al.*, 2010). This is not a simple problem but fortunately can be overcome with a help of algorithms and software available for this purpose.

Bustin in the “Minimum Information for publication of Quantitative real-time PCR Experiments” (MIQE) guidelines, recommended the use of RG as internal controls as the most appropriate normalisation strategy (Bustin *et al.*, 2009). A normalisation factor (NFa) that deals with multiple genes represents a geometric mean of relative expression value in the number of target genes in each sample. Several software packages have been established to normalise expression levels using RG, the most common include geNorm (Vandesompele *et al.*, 2002), NormFinder (Andersen *et al.*, 2004), and BestKeeper (Pfaffl *et al.*, 2004). It should be kept in mind, however, that these programs do not always produce the same results, a consequence of different statistical algorithms used by these programs (Xia *et al.*, 2017).

#### **1.4. Forensic body fluids**

Traces of body fluid recovered at the crime scenes may be among the most important type of evidence for forensic investigations. Blood, saliva and semen are the most common body fluids recovered from the crime scene, and they can be used to obtain more information related to the crime. Messenger RNA analysis is the best method to identify body fluids and predict the time since deposition, which is of great importance in forensic science.

##### **1.4.1. Blood**

Blood is most common biological evidence found at the crime scene. It consists of three main components including plasma (54%), erythrocytes (45%), and leukocytes (1%). Leukocytes known as “white blood cells” (WBC) are classified into three major groups: granulocytes, lymphocytes, and monocytes (Adalsteinsson *et al.*, 2012). Unlike erythrocytes, or “red blood cells”, leukocytes contain nucleus and as such are of great importance in the forensic science.

#### **1.4.1.1. Previous Techniques Developed to Determine the Blood Age**

Currently, there are no techniques available that can precisely determine the age of bloodstains or any other biological stains. Attempts have been made to use RNA degradation as an indicator for the age of bloodstain but have so far proven unsuccessful.

The earliest attempt reported was made by Tomellini (1907) who developed a method showing the change in colour from fresh bloodstain to the stain that was one year old by using chart illustration. The main disadvantage of this method is its subjectivity, since it was based on visual observation and dependant on the observer's performance. Consequently, variable results were likely to be produced by different individuals (Tomellini, 1907). A few years later, Leers (Leers, 1910) focused on haemoglobin transformation during ageing of blood stain by comparing the reflectance spectra of dried blood to fresh blood. This is based on different absorption properties of haemoglobin derivatives. In other words, two dips in reflectance spectra caused by oxyhaemoglobin presence in a fresh blood stain reduces over time while in the same time tip caused by presence of methaemoglobin, result of haemoglobin's exposure to the air, becomes more prominent as the bloodstain dries out (Schotsmans *et al.*, 2017). Another pioneer in this area, Schwarzacher, (1930) found that the bloodstain solubility in distilled water decreases as a function of age. If a bloodstain is exposed directly to sunlight, the time period within which the changes in the sample are visible is 20 hours, whereas it would take several weeks for the same process to occur in a sample placed in complete darkness. Thus, Schwarzacher is a first person who emphasised the dependence of the rate of bloodstains ageing on light effects (Schwarzacher, 1930).

In 1937, Schwarz used an enzymatic method called a guaiacum-based assay to determine the catalase and peroxidase activity of haemoglobin in bloodstains (Schwarz, 1937). The colour generated during this reaction varies with the age of the bloodstain. However, this method

was able to differentiate only very recent bloodstains deposited less than 24 hours from those that are at least few weeks old.

Photo-spectrometry was used to record the reflectance spectra of bloodstains and found that temperature, light and humidity play a crucial role in the rate of colour change (Patterson, 1960). A study conducted in 1962 described a way of determining the age of a bloodstain based on the progressive diffusion of chloride ions. The results showed a black border generated by fixing with AgCl upon reduction (Fiori, 1962). The border is observed when the stain is more than two months old, and the size of the border increases gradually up to nine months. The main disadvantage of this method is that it does not take into account the stain size, and it is also not human-specific. The approach developed in 1977 used immunoelectrophoresis method to examine the serum protein profile of a stain which was deposited from 15 days before the analysis up to a period of one year (Rajamannar, 1977). The analysis showed patterns of disappearance for globulin and albumin proteins over time. In contrast, Sensabaugh in 1971 tested albumin through its immunological reaction in dried blood and found that it is detectable in the stains up to eight years old (Sensabaugh *et al.*, 1971).

#### **1.4.1.2. High-Performance Liquid Chromatography (HPLC)**

The whole blood contains three main forms of blood cells, including WBC, platelets and red blood cells (RBC) which are the most abundant cells in human body. They contain no nucleus and therefore no DNA and have a large amount of the oxygen transport protein called haemoglobin. When a person is injured, the bleeding occurs and due to the exposure of RBC to the oxygen in the atmosphere, haemoglobin gets saturated. In addition, the amount of cytochrome b5 is diminished, therefore the conversion from oxy-haemoglobin (HbO<sub>2</sub>) to methaemoglobin (MetHb) to de-oxyhaemoglobin (Hb) stops. Therefore, physical and

chemical changes in RBC can be used as indicators to estimate the age of bloodstains (Morta, 2012).

HPLC is a method in analytical chemistry used to identify and quantify each compound in a mixture depending on a retention time of each compound. Each compound has a specific retention time, defined as a time needed by each particular compound to travel from the separation column into the detector. This time defines from the point at which the display shows a maximum peak height for that compound (Morta, 2012). In addition, the peak area in the measurement is related to the amount of the particular compound present in the sample. This way it is possible to differentiate between the different compounds even if there are many compounds in a mixture analysed, as is the case with the proteins in the blood. Many studies conducted in this area used products of heme degradation in dried bloodstains as markers to estimate the blood age (Kapitulnik, 2004, Andrasko, 1997, Morta, 2012). The results obtained by Inoue and co-workers show that, in bloodstains, the ratio of the peak areas of haemoglobin  $\alpha$ -globin chain and heme protein decreased with age of the bloodstain. In addition, they also concluded that it is possible to differentiate neonatal bloodstains from adult bloodstains by the presence of  $\gamma$ -globin chains in neonates up to 32 weeks old (Inoue *et al.*, 1991). This significance of this protein is that it is only detected in aged bloodstains, and its peak area increases with the age of the bloodstain (Inoue *et al.*, 1992, Andrasko, 1997). This, so called protein 'X' is not affected by the temperature between 0-37°C. The ratio of this protein to heme in fresh blood is 0 and increases to 0.3 when bloodstain becomes 52 weeks old.

The difficulty with HPLC is its reproducibility. This is largely due to the way samples for analysis are being collected. Bloodstain firstly removed from the substrate using a wet swab, and then dissolved in distilled water. This procedure could generate variation in the

concentration of the haemoglobin obtained from samples that leads to increased standard deviation that makes age prediction estimates quite difficult.

#### **1.4.1.3. Aspartic acid reaction**

Arany and Ohtani in 2011 suggested that it could be possible to use decaying amino acids from blood plasma to determine the age of bloodstains through aspartic acid racemisation (AAR). Depending on the D-L-aspartic acid ratio in some tissues such as teeth, it is also possible to predict the age of an individual by using this method. However, of note is that amino acids decay very slowly which is why this method is only suitable for stains over 10 years old (Arany and Ohtani, 2011).

#### **1.4.1.4. Reflectance spectroscopy**

Numerous studies have been conducted focusing on a colour change in bloodstains by using optical spectroscopy methods to visualise their reflectance spectrum (450-700 nm) and relate it to their age (Patterson, 1960, Kind *et al.*, 1972, Hanson and Ballantyne, 2010, Bremmer *et al.*, 2012b). This method is not invasive since it does not require a sample preparation, and therefore no physical contact with the sample. However, the limitation of this method is that for it to work, the bloodstain must be on a white background.

#### **1.4.1.5. Oxygen Electrodes**

This method is based on the ratio of HbO<sub>2</sub> to Hb. Matsuoka *et al.* (1995) demonstrated that the age of blood could be determined by using oxygen electrode immersed in water (MATSUOKA *et al.*, 1995). Cyanomethemoglobin method is a conventional colorimetry method where blood is mixed with a solution that contains potassium ferricyanide and potassium cyanide to generate a stable colour pigment called cyano-met-haemoglobin. The total amount of haemoglobin is then easily determined photometrically at a wavelength 540 nm (Morta, 2012).

#### **1.4.1.6. Atomic Force Microscopy (AFM)**

Atomic Force Microscopy is a microscopic technique that looks at a three-dimensional profile of a surface under investigation. High resolution is achieved by using a sensitive probe that is attached to a cantilever. AFM is used to determine the elasticity of the surface of RBC by determining the frequency resonance of cantilever. Strasser *et al.* (2007) showed that the elasticity changes occur in RBC which increase during 1.5 h after bleeding from 40 KPa to 300 KPa, and then from 600 HPa within 30 h to 2.5 GPa after 30 days. Therefore, they concluded that there is a relationship between the blood elasticity and time since deposition (TSD) (Strasser *et al.*, 2007). The main advantage of this method is that it does not require any sample preparation, but the main disadvantage is that bloodstains must be on the very flat surface, such as glass and tiles.

#### **1.4.1.7. Electron paramagnetic resonance (EPR)**

The haemoglobin molecule contains iron ion whose spin state has changed during the denaturation and generates four significant EPR related to ferric high-spin, ferric low-spin, ferric non-heme and free radical species (Miki *et al.*, 1987). Fujita *et al.* (2005) found that there is a linear relationship between the ratio of ferric low-spin to ferric non-heme when compared with TSD up to 432 days (Fujita *et al.*, 2005). However, environmental factors such as temperature and light exposure play a crucial role in affecting this relationship. Further, this technique is not commonly used in forensic science because bloodstains and other types of samples often cannot withstand conditions during sample preparation such as extremely low temperatures to which the sample must be cooled first (-196°C).

#### **1.4.1.8. Enzyme-Linked Immunosorbent Assay (ELISA)**

The circadian biomarkers such as melatonin and cortisol can be detected by ELISA to determine at which time of a day bloodstain was deposited, up to a period of one month

(Ackermann *et al.*, 2010). Further, high mobility group box-1 (HMGB1) measurement by ELISA can be used to determine post mortem interval (PMI). This is because this DNA-binding protein is released from the cells after they die by necrosis, therefore, the starting concentration of HMGB1 when taken from the blood cells from live humans is very low, and then gradually increase over the period of 7 days when samples are kept at 24°C (Kikuchi *et al.*, 2010).

#### **1.4.1.9. Immunoelectrophoresis**

The blood serum contains proteins which are gradually deteriorating over time. Albumin and globulins are the most common proteins present in bloodstains. Rajamannar in 1977 conducted an experiment to determine the presence of proteins in the bloodstains which are deposited in a period of 15 days up to one year. The study showed that when using immunoelectrophoresis to compare the pattern of various proteins between fresh human serum and bloodstains, the absence of some of the proteins was evident in the latter (Rajamannar, 1977). By determining which particular serum protein is missing, it is possible to determine the age of bloodstain using immunoelectrophoresis.

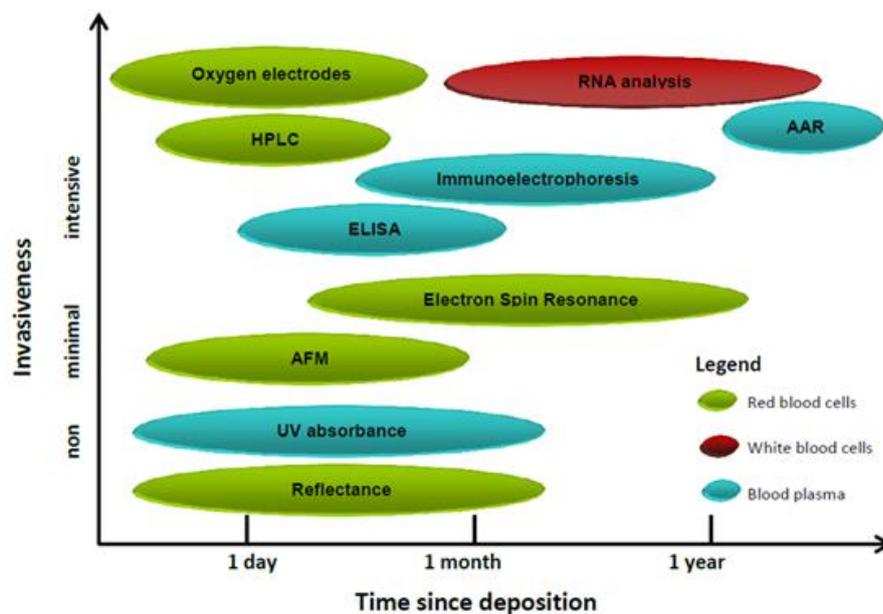


Figure 2. Different methods conducted to determine the age of bloodstains (Morta, 2012).

In summary, all of the methods described above (Figure 2) are limited in its use in the modern forensic science. They lack the specificity and are often unable to differentiate between human blood and that of other species, unable to control for the different amounts of blood therefore leading to erroneous results, while harsh sample preparation methods render some of them inappropriate for use in forensic medicine. This is why further research in this area is warranted and development of the new methods in determining the age of bloodstain is imperative.

#### 1.4.1.10. Forensic RNA analysis

Some of the chemical evidence related to a crime scene such as paint, ink, and drugs tend to age in a chemically predictable way, hence, it is often possible to obtain accurate results for measurement of these samples and to determine their age. On the contrary, determining the age of biological evidence such as semen, saliva and blood is more difficult due to its unpredictability. In the last few decades, attempts have been made to establish the methods for determining the age of bloodstains due to its importance in aiding the crime investigation.

Some methods such as spectroscopic, enzymatic and chromatographic analysis described in the previous sections were developed. Unfortunately, they have a number of limitations which interfere with obtaining accurate results. To date, the most promising work has involved RNA degradation studies (Hampson *et al.*, 2011, Alshehhi *et al.*, 2017).

For many reasons RNA has more advantages over DNA as a forensic material. RNA is found in greater amounts within the cells, it offers high sensitivity with PCR amplification, as well as high specificity with its unique gene expression profile dependent on the cell type. This enables determination of the origin of a sample (e.g. bloodstain from rape or homicide rather than menstrual blood), which can be crucial in determining the link between the crime and the suspect (Bauer, 2007). Another important advantage of RNA in forensic science is its fast rate of degradation. This could be used to estimate the age of biological samples such as bloodstains. In recent years RNA molecule in forensic science research grows substantially, not only for the PMI, determination, identification of body fluids, and the age prediction of a biological sample, but also provide information about the physical characteristics of a subjects such as eye colour, hair colour, skin pigmentation, height, weight and relative age and gender (Morta, 2012). This information can be extremely valuable in cases where there are no known suspects.

DNA is a genetic material present in all nucleated cells. On the other hand, mRNA presence is under both temporal and spatial control. It is, therefore, possible to recognise both type and age of biological stain deposited at a crime scene by analysing the types of mRNA expressed. Quantitative RT-qPCR is one of most effective methods to obtain precise results. In qRT-PCR, labile RNA is firstly converted to more stable form known as cDNA which is obtained through RT process; where a target sequence is amplified and quantified using specific primers and probes.

For highly degraded biological samples, the success of detection by qPCR depends on the size of a target sequence selected. Generally, it is easier to detect smaller sequences than large ones. Modelling the decay as a single molecule stochastic process shows that larger segments are more susceptible to degradation than smaller ones. As a result of this more rapid degradation of larger segments, amplification through PCR will significantly decrease over time. This technique was used to determine the degradation patterns of mRNA segments of different lengths in bloodstains. Environmental factors such as microorganisms, humidity, temperature, UV light, pH, free radical formation, also play an important role in this process of degradation (Anderson *et al.*, 2005).

RT-qPCR was employed to investigate two types of RNA. The result showed that the ratio of 18S rRNA versus  $\beta$ -actin mRNA is changed in a linear fashion over a period of 150 days. The higher degradation rate observed in this study for  $\beta$ -actin mRNA is due to it not forming a complex with the protein, while 18S rRNA is integrated into the small subunit of ribosome and as such better protected from degradation (Anderson *et al.*, 2005). The study has not looked at other body fluids of significance in forensic science, such as saliva and semen, and bloodstains were stored at 50% humidity, while it is known that high humidity can influence the degradation pattern of RNA (Bremmer *et al.*, 2012a, Nakao *et al.*, 2013). Nevertheless, these findings present a significant progress in forensic science offering a novel approach to determining the age of bloodstain that gives higher sensitivity, specificity and applicability of the method.

Simard *et al.* (2012) conducted a similar study to investigate the ratio of mRNA vs. rRNA targets including  $\beta$ -actin mRNA and 18S rRNA in blood, saliva, and semen. The findings from this study showed that the decay rates of selected targets did not exhibit significant differences over a six-month period, and the authors suggested using the individual RNA markers instead of the ratio between different RNA types for estimating the age of the

samples (Simard *et al.*, 2012). It is clear that two previous studies are very similar because they both investigated the bloodstains over the similar time period, and the same markers (18S rRNA and  $\beta$ -actin mRNA) were used. However, two different results were obtained, with one study finding linearly correlated ratio 18S rRNA:  $\beta$ -actin mRNA with the actual age of a sample, whilst other reported no correlation there. Therefore, more investigation is still required in this field by analysing different markers over a different time period. Relevant studies that have been conducted in this area so far are all listed in Table 1.

Table 1. Summary of using mRNA to ageing biological stains.

Research group	Selected marker	Investigation time	Key finding
(Bauer <i>et al.</i> , 2003a)	Fatty acid synthase-messenger RNA 1,2,3, and 4(FASN) multiplex	5 days	Refrigerated post-mortem blood and brain samples were degraded significantly with blood samples taken from living individual in a period of 5 days.
(Bauer <i>et al.</i> , 2003b)	$\beta$ -actin & cyclophilin	5 years, plus 8 and 15-year-old samples	Bloodstains were investigated under different environmental condition. Both targets were detected in aged samples. $\beta$ -actin peak areas in the samples aged up to 8 months were significantly different than sample aged 41 months or older. Cyclophilin marker was significantly lower only in bloodstains older than 59 months.
(Anderson <i>et al.</i> , 2005)	18S rRNA and $\beta$ -actin mRNA	150 days	Bloodstains were tested over a period of time, and the result showed that 18S rRNA is still stable, while $\beta$ -actin mRNA decreased with age of samples increased. Therefore, the ratio 18S/ $\beta$ -actin was linearly correlated with the age increased
(Setzer <i>et al.</i> , 2008)	$\beta$ -spectrin, porphobilinogen (blood), histatin 3, and statherin (saliva). Protamine 1 and 2 (semen), and mucin 4 (vaginal secretions). Ribosomal protein S15 rRNA	547 days	Body fluids including blood were exposed to various environmental conditions targeting specific markers. S15 was the most robust marker detected in bloodstains aged up to 547 days and kept at room temperature condition, 90 days in wet condition outside, and 30 days in dry condition outside. All targets varied between different ages and different environmental factors.
(Anderson <i>et al.</i> , 2011)	18S rRNA and $\beta$ -actin in three varying amplicons	120 days	All three amplicon pairs in bloodstains were generally decreased over time, with the large amplicon showing greater decrease in signal over time.
(Simard <i>et al.</i> , 2012)	3 assays: 18S and cyclophilin A, 18S and GAPDH, 18S and $\beta$ actin	169 days	Blood and semen gave appropriate amount of RNA for amplification, whereas, saliva was limited. No correlation was found for rRNA: mRNA ratio, but individual markers were correlated over the time period.
(Kohlmeier and Schneider, 2012)	$\beta$ -haemoglobin and $\beta$ -spectrin	23 years	B-haemoglobin was detected in all 23 years old bloodstains, but $\beta$ -spectrin could not be detected in any of the samples.
(Morta, 2012)	<i>B-globin</i> (HBB), Glucose-6-phosphatedehydrogenase (G6PD), and 18S rRNA	Days, weeks and months	Selected markers were examined using capillary electrophoresis, and the regression analysis was performed using a peak height to generate an age prediction equation. This equation was the most accurate for prediction the age of bloodstains deposited on cotton cloth and stored at room temperature exposed to light.
(Qi <i>et al.</i> , 2013)	18S rRNA and $\beta$ -actin	28 days	Bloodstains were examined, and it found that 18S is stable over a period of 28 days, and $\beta$ -actin continually increased. 18S: $\beta$ -actin ratio in female was greater than male, however, they generated lower amount of RNA overall than male.
(Alshehhi <i>et al.</i> , 2017)	mRNA marker including <i>a-globin</i> (HBA), porphobilinogen deaminase (PBGD), and HBB. Micro RNA (miRNA) marker (miR16, miRNA 451). Reference gene :18S rRNA, ACTB mRNA, U6 small nuclear RNA (snRNA)	270 days	Blood samples were deposited onto cotton swabs and stored at room temperature. Individual RNA markers showed unique pattern of degradation during the nine-month storage interval, whereas miRNA and U6 markers were more stable over this period.

## 1.4.2. Saliva

### 1.4.2.1. Previous studies on saliva using mRNA

Saliva is produced and secreted from salivary glands which contain water, electrolytes, mucus, and enzymes, all of which flow out of the *acini* into collecting ducts (Tiwari, 2011). Saliva and saliva-stained materials were examined as potential sources to obtain DNA profile in certain forensic settings. However, materials found in crime scene such as envelopes, gags, cans, and cigarette butts could be useful if linked to a time when they deposited. On looking through the literature, a little material is found on the age determination of saliva. A study was conducted by Simard *et al.* (2012) who assessed the possibility of using RNA transcript detection by duplex real-time PCR (RT-qPCR) to determine the age of body fluid stains commonly encountered in forensic biology including saliva. Over six month of storage, the result showed all targets used including the ribosomal 18S RNA and the human  $\beta$ -actin, GAPDH and cyclophilin A mRNAs have a similar rate of RNA decay (Simard *et al.*, 2012).

Each body fluid has a specific gene expression pattern; therefore, mRNA profiling utilises to identify body fluids in a forensically casework. Messenger RNA is more abundant than DNA, because present in many copies while DNA just two copies. Many of the methods already discussed for blood can be applied to saliva. The same RNA and DNA co-isolation method which was described by Alvarez *et al.* (2004) can be also applied for saliva samples (Alvarez *et al.*, 2004).

End-point PCR is an effective method to detect transcription for gene of interest, whereas it gives no information about the level of gene expression. The RT-PCR method proposed by Juusola and Ballantyne in 2005 also applies to saliva, and it is a more accurate and sensitive method of analysis. It allows to measure gene expression levels according to reference genes, which are ubiquitous and responsible for the basic functioning of cells. Saliva-specific genes such *Statherian* (STATH) , HTN3 , and MUC7 were detected and glyceraldehyde 3-

phosphate dehydrogenase (GAPDH) was recommended a common reference gene for mRNA in saliva (Haas *et al.*, 2009). The almost of studies conducted on saliva just for body fluid identification. Bauer and Patzelt (2002) performed the first study in forensic casework using mRNA analysis to identify body fluid. Their study was developed a suitable co-isolation method, but limited for dried bloodstains and epithelia cells (Bauer and Patzelt, 2002).

#### **1.4.2.2. Stable RNA markers of saliva stains**

As described previously one of the main characteristics of mRNA is its instability. In the normal living cell, the activity of the gene is regulated in a number of ways, from transcription step up to the post-translational modification (Porrua *et al.*, 2016). Gene expression regulation, therefore, occurs on the level of synthesis of both mRNA and protein, as wells by orchestrating their degradation (McManus *et al.*, 2015). The lifetime of mRNA is determined by its function in the cell. Hence, some mRNA will be present for minutes such as c-fos mRNA (15 min) while others may be present for hours or even days such as  $\beta$ -globin mRNA (Meyer *et al.*, 2004).

Similarly to mRNA levels in the cell, its decay is regulated by the efficiencies and nucleotide positions at which various stages in gene to protein pathway occur, such as transcription, pre-mRNA splicing, pre-mRNA 3' end formation, post-translational modifications, and export of mRNA from nucleus to cytoplasm (Schoenberg and Maquat, 2012). Messenger RNA decay occurs by two main regulatory mechanisms, one which utilises AU rich elements (ARE) in the 3'UTR (untranslated region) of mRNA, while the other involves removal of the methylguanosine (m7G) cap at the 5' end of mRNA (Jusola and Ballantyne, 2003). Environmental factors including bacteria, ribonucleases, temperature, light and humidity also play an important role for messenger RNA degradation in biological stains. Therefore, stable sample in forensic research is considered the one that survives from time of deposition to time

of collection. The stability of mRNA molecule in body fluids has been subject of many investigations (Visser *et al.*, 2011, Bauer *et al.*, 2003b, Haas *et al.*, 2011a).

The key issue with mRNA molecule in forensic research is its lack of stability and the fact that it is highly prone to degradation. However, over the years it has been shown that it is possible to isolate enough of RNA of an acceptable quality from biological stains even those that are deposited for longer period of time, such as months or years. Zubakov *et al.* (2008) used whole-genome gene expression analyses on RNA isolated from blood and saliva stain samples of different ages in order to identify stable RNA markers. They found five stable mRNA markers showed tissue-specific expression signals in saliva samples aged 180 days (Zubakov *et al.*, 2008).

For determination of their stability, the same five mRNA markers were tested again in much older saliva stain samples, and the result showed successful amplification even in the samples that were 2-6 years old (Zubakov *et al.*, 2009). Another study was performed to determine body fluid identity on dried biological samples involving blood, saliva and semen. The study selected candidate body fluid-specific genes (e.g. STATH and HTN3 for saliva in particular and KRT4, KRT13 and SPRR2A for mucosa in general and similarly chosen markers for other body fluids examined), as well as reference gene (18S rRNA, ACTB and GAPDH) to detect mRNA. Firstly, they successfully developed a 19-plex mRNA multiplex system capable of determining biological origin of body fluids. Their further approach was to combine RNA and DNA profiling in order to give more complete information about both the donor as well as the type of sample that was subjected to analysis, both important in forensic analysis (Lindenbergh *et al.*, 2013, Lindenbergh *et al.*, 2012a).

Setzer *et al.* (2008) examined the influence of different environmental conditions on the stability of mRNA in saliva. Samples were aliquoted and dried onto cotton cloth and then

exposed to the different environmental factors: indoors and room temperature, in light or in the dark, in the humid or non-humid environments, while others were stored outdoors and exposed to humidity, light and heat with or without rain. They were left to decompose under these conditions from 0 to 547 days. When samples were stored indoors, saliva samples were reasonably robust and remained stable for 365 days for all but UV-exposed conditions where the limit was 180 days. On the contrary, when left outdoors, samples degraded quickly, with rain having particular influence on these samples, by decreasing time samples were stable for from 7 days in dry condition to less than a day in the presence of rain. This is most likely due to the variety of factors characteristic for outdoors environment, including intra- and extracellular activities such as presence of bacteria and RNases. These findings have convinced researchers at the time that mRNA analysis cannot be used in forensic research (Setzer *et al.*, 2008).

### **1.4.3. Semen**

Over 1.3 million violent and sexual offences were recorded in 2013/14, as reported by the National Crime Survey (Great Britain, 2014.) Crime scenes related to the most of these offences are likely to contain human body fluids such as blood, saliva, semen and vaginal secretions, that can all play an important role in criminal investigations (Orphanou, 2015).

The average semen volume a healthy male ejaculates is between 2-5 ml (Owen and Katz, 2005). Semen is a complex fluid and presents cellular mixture produced by a variety of different glands within the male reproductive organs (Owen and Katz, 2005). The cellular part of semen comprises predominately of one cell type, spermatozoa. Approximately 15-30% of seminal fluid originates from the prostate and contains highly abundant protein acid phosphatase, prostate-specific antigen and albumin.

RNA has been explored in detail for the identification of biological materials in a forensic context. For example, Bauer and Patzelt in 2003 performed the first study in forensic casework using mRNA analysis to identify body fluid (Bauer and Patzelt, 2002). By selecting specific marker matrix metalloproteinase (MMP) present in endometrium only but not in dried bloodstains and epithelial cells, they demonstrated the possibility of RNA profiling to be used for determining the type of sample found at the scene. Later, they developed a study that expanded the discovery onto semen samples by showing that basic nucleoproteins *protamine* 1 and 2, since exclusively expressed in haploid genome, were suitable markers of spermatozoa (Bauer and Patzelt, 2003).

Bauer *et al.* (2003) were also one of the first to apply the method of RNA analysis for determining bloodstain age. They examined two RNA methods; semi-quantitative duplex RT-PCR with an internal standard and competitive RT-PCR with an external standard. Semi-quantitative duplex RT-PCR takes an advantage of the underrepresented cDNA sequences reverse transcribed from 5' end mRNA, if they are produced by degraded mRNA fragments. Competitive RT-PCR is utilised as a control method to ensure PCR specificity and monitor RT efficiency and ribonuclease contamination (Bauer *et al.*, 2003b).

One of the first documented developments in determining the type of body fluids that used more sensitive, specific and conclusive body fluid identification method that involves RNA analysis was conducted by Juusola and Ballantyne (2003). They focused on mRNA, the molecule that acts as intermediary between genetic information in DNA and proteins specific for that cell type, thereby determining cell-specific gene expression profile (Juusola and Ballantyne, 2003). In other words, as certain genes are expressed in specific tissues and not others, the mRNA profiling will enable differentiation between different body fluids and tissues.

Table 2 shows the common mRNA markers which are used to identify human semen. By using these markers, it is possible to determine the age of semen stains through RNA degradation pattern, one of the key steps in gene regulation process.

Table 2. Messenger RNA markers used to identify semen in human.

mRNA Marker	Gene	Body Fluid Target	Description	Reference
Prostate specific antigen (PSA); also known as kallikrein 3 (KLK)		Semen (seminal fluid)	Produced by prostate epithelia. Hydrolyses <i>Semenogelin</i> during the liquefaction of ejaculated semen.	(Haas <i>et al.</i> , 2013); (Nussbaumer <i>et al.</i> , 2006).
<i>Protamine</i> 1 and 2 (PRM1, PRM 2)		Semen (spermatozoa)	DNA binding protein involved in sperm chromatin condensation.	(Alvarez <i>et al.</i> , 2004); (Fleming and Harbison, 2010); (Haas <i>et al.</i> , 2008, Haas <i>et al.</i> , 2009, Haas <i>et al.</i> , 2013); (Juusola and Ballantyne, 2005); (Lindenbergh <i>et al.</i> , 2013); (Richard <i>et al.</i> , 2012); (Sakurada <i>et al.</i> , 2011); (Vandenberg and Oorschot, 2006).
<i>Semenogelin</i> 1 and 2 (SEMG1, SEMG2)		Semen (seminal fluid)	Coagulation proteins involved in the ejaculation of semen.	(Haas <i>et al.</i> , 2013); (Lindenbergh <i>et al.</i> , 2013); (Sakurada <i>et al.</i> , 2011); (Vandenberg and Oorschot, 2006)
Transglutaminase 4 (TGM4)		Semen (seminal fluid)	Catalyses irreversible cross-links of glutamine residues to peptide-bound lysines or primary amines	(Fang <i>et al.</i> , 2006); (Fleming and Harbison, 2010); (Haas <i>et al.</i> , 2013); (Richard <i>et al.</i> , 2012).

## 1.5. MicroRNA

Micro RNAs (miRNA, miR) are short RNA sequences, 18-25 nucleotides long, present in the non-coding regions of DNA. Together with short interfering RNA (siRNA) they belong to the small non-coding RNA and are mostly responsible for negative regulation of genes in the human body. They are located mostly along the intergenic regions, but can be found within the intronic regions of the human genome (Uchimoto, 2014, Wan *et al.*, 2011). MiRNA plays an essential role in the regulation of gene expression at the post-transcriptional level. They can bind mRNA, usually in its 3' UTR and repress translation through a variety of mechanisms (Han *et al.*, 2018, Moschenross, 2011). MiRNA can bind to its target mRNA resulting in gene silence (Silva *et al.*, 2015, Santos *et al.*, 2014). They are evolutionary conserved, as a way of preserving information from generation to generation (Landgraf *et al.*, 2007, Lagos-Quintana *et al.*, 2001), probably due to their involvement in the important physiological processes such as cell growth, differentiation, glucose homeostasis, fat metabolism and immune regulation (Kannan *et al.*, 2017). On the other hand, they implicated in different pathological conditions such as cancers, infections and autoimmune diseases as well as neurodegenerative disorders (Zubakov *et al.*, 2010, Sun *et al.*, 2010, Silva, 2012, Uchimoto, 2014). Overall, miRNA are major contributors to post-translational repression and degradation of mRNA (Benes and Castoldi, 2010, Courts and Madea, 2010, de Planell-Saguer and Rodicio, 2011).

The biogenesis of miRNA begins in the nucleus (Figure 3) and involves many processing steps. Firstly, primary miRNA (pri-miRNA) which contains 60-80 nucleotide hairpin stem-loop (SL) structure (Sun *et al.*, 2010) is transcribed by RNA polymerase II (Lee *et al.*, 2004). Following this, enzyme Drosha with co-factor DGCR8 are turning it into precursor miRNA (pre-miRNA) ready to be transported into the cytoplasm by cleaving secondary hairpin structure. The transport occurs with the help enzyme Exportin 5 and co-factor Ran-GTP.

There, pre-miR transcript is cleaved with the help of Dicer enzyme and its co-factor TRBP/PACT, and the remaining strand of miRNA binds to Argonaut protein (Ago) to form an RNA inducing silence complex (RISC) (Uchimoto, 2014, Silva *et al.*, 2015). Binding to Ago protein makes miRNA resistant to degradation, much more than is the case with mRNA molecule (Winter and Diederichs, 2011).

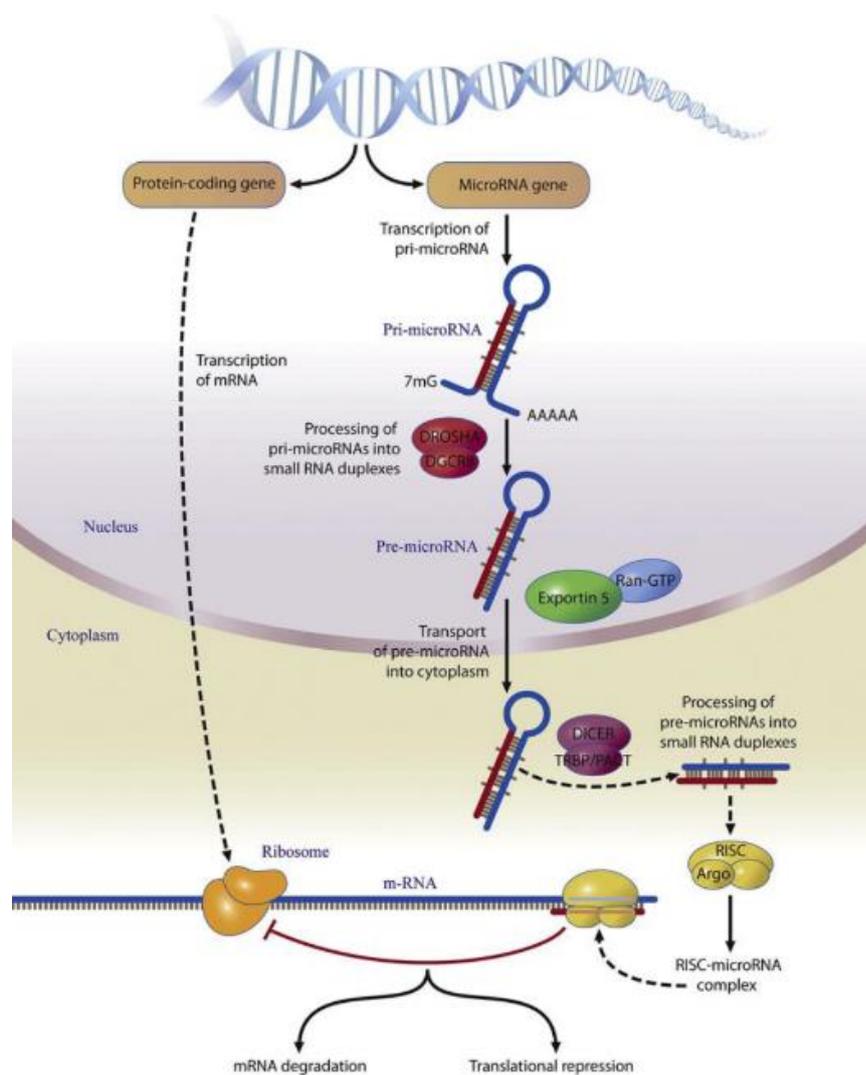


Figure 3. The biogenesis of microRNA in eukaryotic cells. The process starts from the nucleus, where pri-miRNA transcription is made with RNA polymerase II. Following this, Drosha enzyme turns pri-miRNA into Pre-miRNA. Transportation into cytoplasm occurs with Exportin 5 enzyme, after the secondary hairpin structure is cleaved. MiRNA binds into mRNA and if their sequences are completely match, mRNA can be degraded, whereas, if they partial matching this lead to the repression of translation. Image is taken from (Sand *et al.*, 2009).

### 1.5.1. Discovery of miRNA

The gene inhibition was first discovered by Richard Jorgensen and colleagues in petunias, in an attempt to use pigment-producing gene to enhance purple pigment in these flowers. Two flavonoid genes, dihydroflavonol-4-reductase (DFR) and chalcone synthase (CHS) were transferred into the plants. Although in majority transformed samples there was no clear change in flower pigmentation, in about 25% them the reduction in pigmentation was observed and it was accompanied by the reduced expression of both DFR and CHS gene (Van der Krol *et al.*, 1990, Napoli *et al.*, 1990). They called this phenomenon co-suppression, which was later predefined as RNA interference (RNAi). This is how they accidentally discovered a mechanism of gene silencing (Uchimoto, 2014), a process that challenges central dogma of molecular biology by showing that RNA molecule, besides being mediator between genes and proteins, can also interfere with the flow of genetic information.

The first miRNA was identified by Lee *et al.* (1993) during a study that compared the heterochronic patterns of two genes responsible for the development of *Caenorhabditis elegans* larva, lin-4 and lin-14. *Caenorhabditis elegans* is a free-living transparent nematode used as an important model system in biological research. The findings from this study showed that lin-4 transcripts inhibit the translation of lin-14 mRNA. Consequently, lin-4 was the first miRNA to be discovered (Lee *et al.*, 1993). The second miRNA was also discovered in *C. elegans*, when the heterochronic patterns of let-7 with lin-14, lin-28, lin-42 and daf-2 genes were investigated. The results showed that let-7 directly inhibited the 3' UTR of lin-14, lin-28 and lin-41, with the resulting inhibition of lin-41 leading to a complete loss of muscle function (Reinhart *et al.*, 2000).

The gene inhibition mechanism was also studied using *C. elegans*. In this study, *C. elegans* was injected with single stranded RNA (ssRNA) and double stranded RNA (dsRNA). The researchers utilised genes with well-characterised phenotypes such as unc-22 gene, which

codes for myofilament protein found in striated muscle cells. Partial inhibition and full inhibition of this gene will cause twitches and impaired mobility in the muscle, respectively. The result showed reduced activity in *C. elegans* when injected with ssRNA (antisense or sense), whereas, injection with dsRNA resulted in complete loss of muscle activity (Fire *et al.*, 1998). This study was crucial to understanding the gene inhibition mechanism and the researchers involved won the Nobel Prize for Physiology and Medicine in 2006 (Sand *et al.*, 2009).

The same group of researchers further used the same model and injected *C. elegans* with dsRNA derived from the *unc-22* gene. This was to determine at what stage in the gene expression process gene inhibition occurs. After showing no interference with primary DNA sequence, as well as processes of initiation and elongation of transcription, they looked at the levels of accumulation of nascent transcripts in nucleus and cytoplasm using *in situ* hybridization. It was evident that the numbers in the nucleus were affected although not completely, while observing even more dramatic changes in the cytoplasm where the transcripts were virtually eliminated. These findings suggested that gene inhibition occurs at the post-transcriptional stage of mRNA (Montgomery *et al.*, 1998).

The structure features of dsRNA and its effect on gene inhibition in *C. elegans* was also investigated using different factors including RNA bases, length, sequence, homology and helical structure. It was shown that dsRNA from only 25 nt long could cause RNA interference and gene inhibition (Parrish *et al.*, 2000). All miRNA which have been discovered across different species are recorded on a database called miRBase (Kozomara and Griffiths-Jones, 2014, Kozomara and Griffiths-Jones, 2010), which contains with the latest update (v22 released in March 2018) 38589 mature miRNA entries.

### **1.5.2. Micro RNAs as potential biomarker in forensic science**

As we have seen, miRNA plays a major role in many biological processes including degradation of mRNA. The complementarity between miRNA and mRNA occurs mainly in untranslated region (3' UTR) (Benes and Castoldi, 2010). Depending on a type of complementarity, miRNA can perform its function on the targeted mRNA in two ways. If their sequences perfectly match, miRNA cleaves and degrades mRNA, while only partial complementarity of the sequences will lead to the suppressed translation (Hutvagner and Zamore, 2002, Grimson *et al.*, 2007, Silva, 2012, Santos *et al.*, 2014, Silva *et al.*, 2015).

In forensic science, miRNA profiling has a number of advantages even when compared to mRNA analysis. For example, the short sequence of miRNA provides an inherent stability that is greater than mRNA, consequently, makes them less prone to degradation processes and therefore of great interest in forensic science (Courts and Madea, 2010, Zubakov *et al.*, 2010, Uchimoto, 2014, Silva *et al.*, 2015). Further, similarly to mRNA, the gene expression patterns of miRNA are also tissue specific, therefore enabling identification of the type of sample they were isolated from. Finally, the regulatory role of miRNA suggests their relatively high abundance within the cell. All these reasons make miRNA an excellent candidate in forensic casework, especially in cases when low-level samples recovery is the only option, or mixed body fluids are present.

Most of the research groups in forensic science have used miRNA analysis to identify body fluids rather than time since deposition of bloodstains. Only two studies have used miRNA to estimate the age of bloodstains (Nakao *et al.*, 2013, Lech *et al.*, 2014), however, a few that were conducted used them to estimate PMI. In general, the miRNA studies performed have largely used method similar to RT-qPCR described for mRNA, with the exception that they would utilise special primers to accommodate for miRNA shorter size (Uchimoto, 2014). The first study explored miRNA as a potential biomarker to identify body fluids which was

performed by Hanson *et al.* (2009). In this study, five different body fluids were investigated. Nine potential miRNA markers were detected and could be used as a positive indicator of the presence of saliva (miR-658 and miR-205), semen (miR-135 and miR-10b), blood (miR-451 and miR-16), menstrual blood (miR-451 and miR-412) and vaginal secretions (miR-372 and miR-124a) (Hanson *et al.*, 2009). In addition, two potential reference genes were selected including RNU6B for blood and semen; and RNU44 only for blood. RNU44 was also identified as stable reference gene during a comprehensive study conducted with a focus on the miRNA (Zubakov *et al.*, 2010). An important review was published by Silva *et al.* (2015) which summarised the use of miRNA as biomarkers to identify body fluids, when some factors are controlled such as, methodological approaches, physiological conditions, environmental factors, gender, as well as, pathologies and samples storage (Silva *et al.*, 2015). Figure 4 gives an overview of all miRNA suggested as possible body fluid biomarkers to date. It is a clear demonstration of how variable conclusions were obtained from different experimental settings, with only few miRNA markers confirmed by two or more studies. For example, Hanson, Zubakov, Wang and respective co-workers suggested miR-16 as a venous blood-specific marker, and consequently, a potential biomarker for bloodstain (Hanson *et al.*, 2009, Zubakov *et al.*, 2010, Wang *et al.*, 2012). Further, several studies suggested miR-205 as a good biomarker for saliva, with the exclusion of Wang *et al.* (2012) who warned that presence of miR-205 in saliva samples is due to its presence in epithelial cells collected with saliva sample (Wang *et al.*, 2012). No miRNA biomarker was found to be urine-specific, and given the importance of this human fluid in forensics, further research in this area is warranted (Silva *et al.*, 2015). No forensic studies with the aim to predict the age of biological stains as well as to estimate the PMI were mentioned (Silva *et al.*, 2015).

A number of studies was conducted to investigate the stability, sensitivity, and specificity of miRNA with respect to body fluids identification. Dried blood (venous and menstrual) and

semen samples were most commonly used, and samples tested under conditions of controlled temperature and humidity and stored for up to a year showed reasonable stability (Zubakov *et al.*, 2010, Courts and Madea, 2011, Wang *et al.*, 2013b). These results not only confirmed the stability of blood and semen markers used but also suggested that miRNA analysis is more sensitive than mRNA (Zubakov *et al.*, 2010), and confirmed their suitability in detecting blood (miR-126, mi-R150, and mi-R451) and saliva (miR-200c, miR-203 and miR-205) samples (Courts and Madea, 2011).

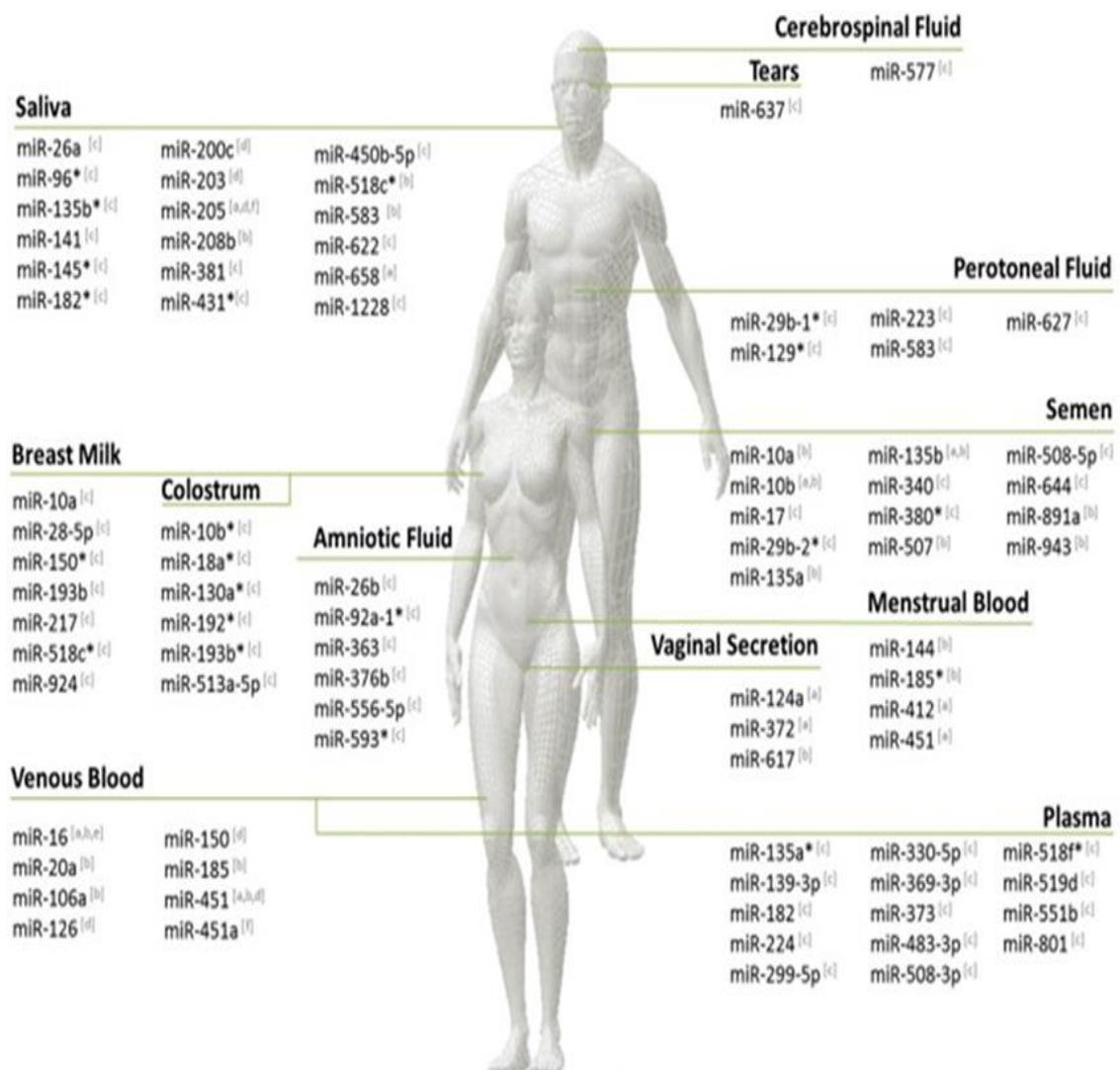


Figure 4. Overview of all miRNA suggested as possible body fluid biomarkers by [a] (Hanson *et al.*, 2009) ; [b] (Zubakov *et al.*, 2010); [c] (Weber *et al.*, 2010); [d] (Courts and Madea, 2010) ; [e] (Wang *et al.*, 2012); and [f](Omelia *et al.*, 2013). Image from Silva *et al.* (2015).

However, it is possible that the stability of miRNA markers is tissue specific, since poorer stability and relatively short half-life was detected in miRNA from brain tissue. This study has analysed specific miRNA abundances and kinetics of miRNA decay in primary cultures of human neural cells and human brain tissues after short post-mortem interval (Sethi and Lukiw, 2009).

Overall, it is accepted that miRNA have higher resistance to degradation in comparison with mRNA but lower than DNA. The determination of PMI in principle requires parameters that change constantly and linearly with time after death (Liu *et al.*, 2007). The degradation of DNA in post mortem samples may change constantly and linearly from the time of death (Liu *et al.*, 2007), but miRNA are also emerging as a new marker with a potential to enable estimation of PMI. Given that both PMI and the estimation of stain age are based on understanding of degradation patterns of markers of interest, further development in one area will undoubtedly lead to advances in the other.

The level of RNA, miRNA, and 18S rRNA were evaluated as an attempt to estimate PMI in heart tissues removed from adult rats at different PMI. The results suggested that there was a pattern of tissue 18S rRNA and miRNA degradation useful in determining the PMI during seven days period (Li *et al.*, 2014). They further described equation of parabola that presented the relationship between Cq values of 18S rRNA and PMI. It is becoming clear that rather than focusing on a single marker, it is possible that developing an appropriate mathematical model to determine PMI that will be more widely applicable would require including both degradation of RNA and miRNA in different matrices. Indeed, the evidence available from the studies conducted so far indicate that combination of stable (miRNA) and more degraded (mRNA) transcripts could offer a successful breakthrough in developing a more robust mathematical model for PMI estimation (Silva, 2012). For example, tissues (rat and human) with known PMI and other parameters were collected. RNA was extracted and tissue specific

reference biomarkers mixture of miRNA and rRNA, were selected using geNorm, while mRNA for ACTB and GAPDH were used as target genes. The mathematical model using target genes (mRNA) normalised to reference biomarkers (miRNA and rRNA) showed significant relationship between the change in their Ct value and PMI (Lv *et al.*, 2016).

Although miRNAs are considered amongst the most stable elements in blood, it is important to have in mind that they too can degrade, especially under conditions of high humidity and when bloodstains are not completely dry (Courts and Madea, 2010, Patnaik *et al.*, 2010). For example, bloodstains collected and incubated in chamber with 25°C and 50% humidity, as an attempt to explore any temporal changes in the relative levels of miR16 and miR-451 during incubation. Relative quantifications of both miR-16 and miR-451 concentrations were performed using the  $2^{-\Delta\Delta Cq}$  method, and expression levels were normalised to that of U6B snRNA. It was shown that the target miRNA markers decreased significantly at day 5 and continued to decrease until the day 28. The correlation coefficients ( $r$ ) of the miR-16 and miR-451 concentrations with dried time of bloodstains were 0.409 and 0.871, respectively. Therefore, it was suggested that the age of bloodstains could be calculated based on these findings (Nakao *et al.*, 2013). Another interesting addition to this study was presence of ethanol, amphetamine sulfate or methamphetamine hydrochloride in these bloodstains in an attempt to examine if the concentration of any of these substances would prove useful in determining bloodstain age. Unfortunately, while ethanol degraded to quickly, within 180 min, other two substances showed constant levels during the time period examined. However, although they could not be used for determining age of forensic sample, they could implicate that person who used them has been present at the scene when the crime was committed.

Another study explored the suitability of miR-142-5p and miR-541 for bloodstain deposition timing. Both markers were normalised against miR-222 and using relative quantification method during the 24 h day/night period. No significant difference was observed with miR-

142-5p, while Cq value was above 35 for miR-541, showing low abundance of this marker in blood. Therefore, both miRNA markers were proved unsuitable for estimating the deposition time of forensic bloodstains (Lech *et al.*, 2014). The studies described above, however, did not explore any other important biological material such as saliva, semen, so further investigation into different miRNA assays as well as others body fluids are still needed.

### **1.5.3. Micro RNA turnover and degradation**

The regulation of miRNA is important for correct cell development and function and for maintaining normal physiology in the cell and govern variety of cell processes such as cell cycle, development, differentiation and apoptosis (Palmero *et al.*, 2011). Due to their function in targeting mRNA and interfering in protein synthesis process in that way, changes in their regulation can lead to diseases (Ha and Kim, 2014). In other words, with their role in dictating fate of mRNA molecules, miRNA are similar to transcription factors that regulate the expression of target genes, only here the regulation is happening on the transcript levels. It is not surprising, therefore, that stability and lifespan of these small molecules is important for smooth functioning of the cells and entire body.

From the point of molecular structure, as described, miRNA are formed from pri- and pre-miRNA precursors in nucleus that, once transported into cytoplasm, are cut out from the stem-loop hairpin structure and packaged with proteins into relatively stable complexes (Krol *et al.*, 2004). Stability of miRNA emerged as an important feature in the early studies. Some examples are miR-208, located within an intron of the gene for myosin heavy chain, that remains in the cells up to 3 weeks of blocking the transcription of the host gene, as well as mature miRNA in human 293T cells, where 8h long blocking of transcription with a chemical inhibitor had no effect on their expression levels (Zhang *et al.*, 2012). Another mechanism involved in miRNA homeostasis, likely due to its higher stability within RISC complex, is



deposition. Answer to when biological stain was deposited may be given through RNA analysis; therefore, many studies have been conducted in this direction. The ability to predict the age of any biological samples would be undoubtedly the most important breakthrough in forensic science, because it could link the sample recovered at the crime scene with the time when it was deposited.

As we have seen, RNA has several advantages over DNA; it is more sensitive, more abundant, and has a fast rate of degradation. In contrast, miRNA with short sequences is more stable than RNA and both have a specific transcription which could be used to identify the type of body fluids, as well as to generate the mathematical model which help in stain age prediction. Up to date, there are no techniques available that can precisely determine the age of any biological stain, with attempts so far still being inaccurate, unspecific, and with low sensitivity. In addition, almost of them require a specific set of conditions not always available in a real crime. Forensic RNA analysis emerges as an appropriate tool for filling this gap. A successful attempt to correlate two types of RNA described with bloodstain age up to six months, that is able to distinguish between recent sample and the sample that is up to six months old, would offer a significant improvement in forensic analysis. Determination of the age of saliva and semen stains is also a very important in forensic investigation, especially in rape and sexual assaults cases. According to the literature, only one study attempted to determine the age of saliva and semen, and no correlation was detected between RNA profile and age of the investigated samples.

Micro RNA analysis is still a relatively new field where not enough is known, especially about the degradation pattern of miRNA. No studies have been conducted using miRNA for saliva and semen age prediction.

## 1.7. Aims of study

Present thesis was conducted with the aim to quantify RNA extracted from blood, saliva, and semen and compare it to the same type of sample deposited at the time zero, in order to predict the age of the biological stain. It focused on identifying and examining specific and more sensitive markers for the samples that have not been thoroughly investigated, such as semen samples.

RNA has been proven to have a fast rate of degradation when compared to DNA and miRNA. Thus, this thesis aimed to further explore the advantages of using mRNA and miRNA in forensic science investigations. An attempt was made to generate models which could be useful to predict the age of blood, saliva, and semen. Relative quantification was employed as a method of choice in order to look at the relationship between the quantity of particular markers and the actual age. This was done using a formula  $2^{-\Delta\Delta Cq}$  following the hypothesis that the overall abundance of RNA in the sample decreases as the age of the sample increases, therefore giving a negative correlation between the age of the sample and the  $\Delta Cq$  value.

This thesis has been divided into three parts.

1. Optimization and validation of the RT-qPCR experiment by investigation of some of the chemical reagents, and their influence on accurate quantification.
2. Generation of the mathematical models for blood, saliva, and semen by monitoring several candidates of a known actual age over a time period. These candidates include mRNA and miRNA markers.
3. A combination between more stable candidates (miRNA) and more degraded candidates (mRNA) by using geometric mean normalisation strategy as an attempt to improve the accuracy of models obtained.

## **Chapter Two: Material and methods**

## **2.1. Ethical approval**

One of the most important factors when embarking on a new research study and experimental test is to consider all potential ethical implications that might arise and make sure that appropriate approval is obtained. This is essential so that development and the nature of the research study are set in the way that will increase the benefit and minimise the risk of harm. The physical and mental safety, as well as the privacy and comfort of the donors must be maintained if the method is to be implemented by the forensic community. For the studies presented in this thesis, the information sheet was provided to participants prior to taking part in the study, and they were encouraged to ask questions related to the research conducted and their involvement. Following their interest in taking part, the consent was taken, and they were further assured that they can withdraw at any time. All studies presented have been approved by the University Of Huddersfield School Of Applied Sciences Ethics Committee.

## **2.2. Sample collection**

Samples were collected from healthy individuals, mostly from the team members within the research group at the University of Huddersfield. This way, it was possible to collect the samples efficiently and effectively, and to minimise chances of contamination from the surrounding areas. This was important since poor sample collection method can be a cause for many unnecessary problems, including low-quality yields during the extraction of sample and can affect downstream processes such as cDNA synthesis and qPCR. Table 3 shows all information about the samples under investigation in both TaqMan and SYBR Green chemistries. The present study was focused only on the environmental conditions such as the temperature, light, and time, whereas other factors involved including age, sex, and ethnicity were ignored.

Table 3. Information of the samples investigated using TaqMan and SYBR Green chemistry.

TaqMan chemistry method (mRNA and miRNA markers)						SYBR Green chemistry method					
Sample type	Frame time	Number of samples	Age	Sex	Ethnicity	Sample type	Frame time	Number of samples	Age	sex	Ethnicity
Blood	One week	56 = 7*8 (7 individuals with 8 time points from fresh until day 7 of storage.	20-45	5 males and 2 females	Arabic , Caucasian , and Japanese	Blood	One week	105 = 21*5 ( 21 individuals with 5 time points from fresh, first, second, third , and fourth week of storage.	20-45	19 males and 2 females	Arabic , and Caucasian
Blood	One month	75 = 15*5 (15 individuals with 5 time points including, fresh, first, second, third, and fourth week.	20-45	10 males and 5 females	Arabic and Caucasian	Saliva	One month	40 = 8*5 ( 8 individuals with 5 time points from fresh, first, second, third , and fourth week of storage.	20-45	5 males and 3 females	Arabic , and Caucasian
Saliva	One month	45 = 9*5 ( 9 individuals with 5 time points from fresh, first, second, third , and fourth week.	20-45	6 males and 3 females	Arabic and Caucasian	Semen	One month	30 = 6*5 ( 68 individuals with 5 time points from fresh, first, second, third , and fourth week of storage.	25-45	6 males	Arabic
Semen	One month	60 = 12*5 ( 12 individuals with 5 time points from fresh, first, second, third , and fourth week.	25-45	12 males	Arabic and Caucasian						

### **2.2.1. Blood samples**

Blood samples were deposited on a sterile filter paper by a finger prick test (Owen Mumford Unistick 3: 1.8 mm depth, 23G gauge). Bloodstain was deposited and covered about 5 cm<sup>2</sup> of the filter paper, and then packaged into a sterile RD polyethylene bag (Fisher Scientific, UK) to minimise the chances of contamination and to decrease the exposure to environmental factors including light and humidity. Sample protected in this way were then stored at room temperature until the extraction. For blood, two time periods were investigated: samples were collected after one week and second one after one month.

### **2.2.2. Saliva samples**

Saliva samples were collected at least one hour after food and drink consumption. This was done in order to minimize any possible contamination that could introduce experimental variations. All saliva samples were collected using sterile buccal swabs (Sarstedt, UK), by swabbing both cheeks for 30s. This was to ensure that a sufficient number of epithelial cells is collected. Samples were then stored in the dark area at room temperature until extraction.

### **2.2.3. Semen samples**

For the semen samples collection, donors were provided with a sterile collection tube. After the sample was collected, sterile cotton swabs were used to prepare five samples for future analyses: one to be analysed fresh, while remaining four cotton swabs were stored for the analyses in the weeks to follow, namely, after one week, two weeks, three weeks, and four weeks, with one cotton swab being analysed each week. All samples (cotton swabs) were stored in a dark at a room temperature until the extraction.

### **2.3. RNA extraction**

Unlike DNA, RNA is rapidly degraded especially in the areas contaminated with RNases, enzymes that catalyse the degradation of RNA into smaller subunits. Hands and dust particles are the most common sources of bacteria and fungi, and as such rich in RNases. For this reason, benches and all equipment used in the experiments were decontaminated with 5% Trigene before each extraction. Care was taken when handling RNA samples before and after the extraction in order to minimise unwanted changes for the further process.

RNeasy Mini Kit from Qiagen was used to isolate total RNA from saliva, blood, and semen samples. The process was conducted according to the manufacturer's protocol, with modification as described by Zubakov *et al.* (2008) (Zubakov *et al.*, 2008). These involved soaking two circles hole-punched from blood stained filter paper, or a whole head cotton swab from saliva and semen samples in 350µl RLT buffer for one hour at 4°C before extraction. For semen extraction, two different methods were applied. First method involved soaking the whole head of cotton swab in 300µl RLT buffer, after which 50 µl Dithiothreitol was added (+DTT), while in the second case no DTT was added (-DTT). This was an attempt to explore the influence of DTT reagent on mRNA and miRNA targets. Blood samples were extracted daily for one week, after which weekly extractions were performed until one month from the start of extraction. Saliva and semen samples underwent only a weekly extraction for one month. RNA was eluted with 40 µl RNase free water and stored -20°C until further analysis.

### **2.4. DNase digestion**

Presence of genomic DNA in the sample after RNA extraction, even traces of DNA molecule can influence the accuracy of sensitive applications such RT-PCR. This is because both RNA and DNA targets may be amplified, resulting in erroneous quantification of the RNA target. Therefore, two strategies were performed to digest DNA. The first method involved using Turbo DNA- Free kit (Life Technologies) as followed: 1µl and 5 µl from 10x Turbo DNase

and Turbo DNase buffer were respectively added into the extracted RNA, followed by vortexing and incubation at 37°C for 25 min. 5µl of re-suspended DNase inactivation reagent was added, and thoroughly mixed by inverting only (DNase is highly sensitive to physical degradation, therefore vortexing was avoided at this stage), and incubated at room temperature for 5 min. Samples were then centrifuged for 1.5 min, and RNA was transferred to a fresh tube and stored at -20°C until needed. The second method was performed using On-Column DNase Digestion with the RNase-Free DNase Set (Qiagen) according to the manufacturer's protocol.

## **2.5. Quantification and purity assessment of RNA**

The integrity and purity of extracted RNA are the most important factors can influence the accuracy of the results. Proteins, carbohydrates, and lipids could be the common factors that affect the purity of RNA. Therefore, the quality and quantity of RNA were measured using NanoDrop ND-1000 software version 3.7.1. The solution used to elute the extracted RNA was used as a blank measurement to calibrate the system. The system measured the absorbance of the sample giving the concentration in ng/µl. The process was repeated three times for each sample, and average value taken as final.

## **2.6. Standard reverse transcription**

Extracted RNA must be converted into cDNA to prevent it from the degradation process as well as to undergo PCR for detection of mRNA. For the reverse transcription reaction, random primers (Hexamer, Decamer, and Oligo dT) were used to prime cDNA individually. Two Reverse Transcriptase Enzymes were also employed, Molony Murine Leukemia Virus (M-MLV) Reverse Transcriptase Enzyme (Life technologies, UK) and Multiscribe Reverse Transcriptase first strand synthesis (Life Technologies, UK). Samples underwent standard reverse transcription following the RETROscript protocol RT (Technologies, 2011). Samples

were initially heat denatured with 50  $\mu$ M of chosen random primer and nuclease-free water on a Veriti thermocycler (Fisher Scientific, UK) for 3 min at 75 °C. This step is commonly used for removing any secondary structure that may inhibit reverse transcription. Other essential reagents including dNTP mix (2.5 mM), RT buffer (10X), RNase inhibitor (10 units/ $\mu$ l), and 100 units/ $\mu$ l of selection RT enzymes were then added to the reaction mix with a total volume of 20  $\mu$ l. Samples then underwent standard RT using the following conditions: 60 min at 42 °C, 10 min at 92 °C. Negative control of reverse transcription was conducted (used water instead of Reverse Transcriptase Enzymes) and negative control (water instead of template) were prepared along with each blood, saliva, and semen samples, and all were stored at -20°C until needed.

## **2.7. Stem-loop reverse transcription**

A mature miRNA is a short oligonucleotide sequence (only 18-25 nucleotide long), which makes it difficult for annealing of primers and accurate synthesis of cDNA strand. A stem-loop reverse transcription is the best choice to overcome this problem. There are two options to perform this method, stem-loop primers (TaqMan® chemistry) or poly-T primers utilised after polyadenylation (SYBR® Green chemistry). In the present thesis, stem-loop primers were selected due to their high specificity and frequent use in forensic RNA studies (Chen *et al.*, 2005, Uchimoto, 2014, Meer *et al.*, 2013, Zubakov *et al.*, 2010).

The stem-loop primer is a universal stem-loop sequence with 6 bases attached that are complementary to the end of the miRNA sequence (sequences marked in bold in Table 4). Forward primer is essentially the DNA version of the miRNA of interest, except for the bases that were used in RT primer binding. During RT, the 3' end of the stem-loop primer binds on the last 6 bp of the mature miRNA sequence while the 5' end, which is artificial (in nature)

will fold on itself. The RT enzymes (e.g. Multiscribe) will then extend this sequence. The stem-loop stays unfolded during PCR running. All miRNA primers used are listed in Table 4.

The TaqMan® microRNA reverse transcription kit and custom microRNA primers from MWG Eurofins were used for miRNA transcripts. The essential reagents were prepared using TaqMan microRNA protocol (Technologies, 2011), and master mix included nuclease-free water, reverse transcription buffer (10x), multiscribe RT (50U/  $\mu$ l), RNase inhibitor (20U/  $\mu$ l), and dNTPs (100 mM). This master mix was combined with a primer (2.5 nM) and extracted RNA in a single tube. Samples then underwent stem-loop RT on a Veriti thermocycler with the same cycling conditions as described in the protocol. Similarly, to the standard reverse transcription assay described above, the negative control was also prepared.

Table 4. MicroRNA primers using in blood, saliva, and semen.

Micro RNA marker	Mature microRNA sequence	Stem-loop reverse transcription	PCR Forward Primer
has-micro-205	UCCUUCAUCCACCGG <b>AGUCUG</b>	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACC <b>AGACT</b>	GCGCGTCCTTCATTCCACC
has-micro-451	AAACCGUUACCAUUACUG <b>AGUU</b>	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACA <b>ACTCA</b>	GCGCGAAACCGTTACCATTAC
has-micro-891a	UGCAACGAACCUGAGCC <b>ACUGA</b>	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACT <b>TCAGTG</b>	GCGCGTGCAACGAACCTGAG
has-RNU44	CTTGTACCAATTAATCG <b>AGGAU</b>	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC <b>ATCCTC</b>	ACTGAACATGAAGGTCTTAATTAGCTC

## 2.8. Quantification PCR analysis using TaqMan probe

The TaqMan universal master mix II protocol (Life Technologies, UK) was used to amplify RT product. All samples and reagents were briefly centrifuged before real-time PCR was performed using RQ method on a 7500 Fast Real-Time PCR system (The Applied Biosystems, Life Technologies, UK). The total volume reaction was 10 µl and contained 2 µl of RT product or 2 µl of RNase free H<sub>2</sub>O (Sigma) for negative controls, 5 µl fast universal master mix (Life Technology, UK), and 1µl assay primer and probe (Life Technology). The quantitative PCR reactions were prepared in triplicate and then transferred to MicroAmp® optical 96-well reaction plate (Life Technologies, UK), and sealed with a MicroAmp® clear adhesive film and applicator (Life Technologies, UK). All reactions were run in the qPCR machine with manufacturer's cycling conditions. All TaqMan probes specific for different body fluids and 6 RG used for normalisation are listed in Table 5. The selection of these markers was based on their success in previous experiments (Anderson *et al.*, 2005, Haas *et al.*, 2008, Haas *et al.*, 2009).

Table 5. TaqMan probe used for blood, saliva, and semen.

Assays	Name	Amplicon length	Specificity
ALAS2	<i>Aminolevulinate, delta, synthase 2</i>	128 bp	Blood marker
GYPA	<i>Glycophorin A</i>	141bp	Blood marker
HBB	<i>Hemoglobin beta</i>		Blood marker
HTN3	<i>Histatin 3</i>	136 bp	Blood marker
MUC7	<i>Mucin 7</i>	77 bp	Saliva
PRM1	<i>Protamine 1</i>	99 bp	Semen (spermatozoa)
SEMG1	<i>Semenogelin 1</i>	82 bp	Semen (seminal fluid)
ACTB	<i>Actin, beta</i>	87 bp	Reference gene
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	93 bp	Reference gene
B2M	<i>Beta 2 microglobulin</i>	95 bp	Reference gene
RPLP0	Ribosomal protein, large, P0	76 bp	Reference gene
FFRC	<i>Transferrin receptor</i>	156 bp	Reference gene
GUSB	<i>Glucuronidase, beta</i>	149 bp	Reference gene

## 2.9. Real-Time PCR analysis using SYBR Green

Real-time PCR was performed using the relative quantitation method with SYBR Green including a melting curve analysis on a 7500 Fast Real-Time PCR system. Reaction master mix was prepared in a total volume of 10 µl containing 2 µl of cDNA (or 2 µl of DNA/RNA free H<sub>2</sub>O (Sigma) for negative controls), 5 µl SYBR Green master mix (Life Technologies, UK), 0.5 µl of forward primer (500 nM), 0.5 µl reverse primer (250 nM) and 2 µl DNA/RNA free H<sub>2</sub>O. All PCR primers used were unlabelled from MWG Eurofins. Table 6 shows a list of these primers with their sequences. Some of the genes play an important role in oxygen homeostasis such as *hypoxia-inducible factor 1*(HIF1A), *vascular endothelial growth factor A* (VEGFA) and *endothelial PAS domain protein 1*(EPAS1).

Table 6. Unlabelled assays for blood and saliva.

Assays	Name	Forward primer sequence	Reveres primer sequence
HIF1A	<i>Hypoxia inducible factor 1 , alpha subunit</i>	GCTCCCTATATCCCAATGGA	GCTTGCGGAACTGCTTTC
EPAS1	<i>Endothelial PAS domain protein 1</i>	GCGCTAGACTCCGAGAACAT	TGGCCACTTACTACCTGACCCTT
FGB	<i>Fibrinogen beta chain</i>	CGGTGGTGGATGGTGGTATAA	AGGTGTACTGTCCACCCAGTAGT
FN1	<i>Fibronectin 1</i>	CCTTCATGGCAGCGGTTT	AGCGTCCTAAAGACTCCATGATCT
VEGFA	<i>Vascular endothelial growth factor A</i>	GCCTTGCCTTGCTGCTCTA	ACTTCGTGATGATTCTGCCCT
F12	<i>Coagulation factor X</i>	AGAACCCTTCGACCTGCT	CCACGATCCTGGTGAGGTTG
HBB	<i>Haemoglobin beta chain</i>	GGCAACCCTAACCCCTAAGGTGAAGGC	GGTGAGCCAGGCCATCACTA
ACTB	<i>Actin beta</i>	ATAGCACAGCCTGGATAGCAACGTAC	CACCTTCTACAATGAGCTGCGTGTG

## 2.10. Quantification PCR of stem-loop RT products

The stem-loop RT products underwent qPCR using TaqMan fast universal master mix (Life Technologies, UK) and a 10X primer solution containing a universal probe (250 nM), a specific forward primer (1500 nM) and universal reverse primer (700 nM). The universal reverse primer was described by Chen *et al.* (2005) (Chen *et al.*, 2005), and the universal probe was described by Jung *et al.* (2013) (Jung *et al.*, 2013). The total reaction volume was 10  $\mu$ l; a negative control containing all the components of the PCR reaction except RT product was included along with each marker. Quantitative PCR was performed on a 7500 fast real-time PCR system using relative quantification method and following the manufacturer's cycling conditions.

## 2.11. Data analysis

Relative quantification (RQ) is generally accepted as the best method of choice to quantify RNA degradation. To evaluate if RNA degradation is a good parameter for the prediction of the age of blood, saliva, and semen samples, Livak or  $\Delta\Delta Cq$  method was used (Livak and Schmittgen, 2001). A fresh sample of blood, saliva, and semen was selected as calibrator to investigate how the transcripts of chosen markers degraded over time. This method involves three steps as follows:

- i. Normalise  $Cq$  (Target) to  $Cq$  (reference gene) for both GOI and calibrator, to generate  $\Delta Cq$  of both.
- ii. Normalise  $\Delta Cq$  of the test sample to  $\Delta Cq$  of the calibrator, to obtain  $\Delta\Delta Cq$ .
- iii. Calculate expression ratio, or fold difference via the equation of  $2^{-\Delta\Delta Cq}$ .

The starting point for this analysis is the cycle threshold ( $Ct$ ) or cycle quantification ( $Cq$ ) of both target and reference genes. This point is termed as a number of cycles required for fluorescence signal to cross the threshold (exceeds the background signal). There are two

ways that  $\Delta Cq$  was calculated, either using the formula  $\Delta Cq = (Cq_{\text{Target}} - Cq_{\text{Reference}})$  or following  $\Delta Cq = Cq_{\text{Max}} - Cq_{\text{GOI}}$ .

Linear regression analysis was used to predict the age of blood, saliva and semen samples. This was achieved by using single and multiple regression analysis with confidence (CI) and prediction interval (PI) at 95%. The Analysis of Variance (ANOVA) and Pearson's correlation were also performed using the SPSS statistical software v 22. The level of significance was set at  $p < 0.05$ , and the interpretation of the Pearson's correlation coefficient (r) was as shown in Table 7 (Mukaka, 2012).

Table 7. Rule of Thumb to interpret the level of a Correlation Coefficient. This table as described by (Mukaka, 2012).

<b>Size of correction</b>	<b>Interpretation</b>
0.90 to 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 to 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 to 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 to 0.30 (-0.00 to -0.30)	Negligible correlation

## **Chapter Three: Optimisation and validation of the experiment**

### **3.1. Introduction**

The method that has revolutionised molecular biology and became a routine laboratory tool to study low abundance gene expression is amplification of cDNA products reverse transcribed from mRNA by real time polymerase chain reaction (RT-PCR). The main advantage of this technique is that it is easy to perform and yet it obtains the necessary accuracy, reproducibility, and rapid quantification of the results. To achieve this, it is necessary to establish not only method that is reproducible but also to identify appropriate mathematical model for data analysis (Pfaffl, 2001).

The quantification of gene expression starts with RNA extraction from biological samples, followed by DNA digestion, cDNA synthesis and PCR. To perform these steps smoothly and to obtain accurate results, optimisation of the experiment is required. A number of protocols and reagents are available from a variety of suppliers and it is important to choose the best combination of reagents and procedures for every experiment. The choice of quantification method is strongly dependent on the gene of interest (GOI), its abundance, the degree of accuracy and whether relative or absolute quantification will be applied. Finally, it is important to properly investigate and compare chemical reagents such as TaqMan probe and SYBR Green in order to select the best option for a particular experimental set up. This chapter aims to investigate the commonly used reagents in the process of quantification RNA. These include DNA digestion kits, cDNA priming reagents and PCR chemistries including TaqMan probe and SYBR Green.

### **3.2. Experimental design**

Blood samples were used to optimise the experimental conditions and investigate the best choice of the reagents. This is because the blood is more sensitive and as such more difficult to optimise than either saliva or semen samples. No extraction kits or methods were investigated, since RNase Mini Kit (Qiagen, UK) was already validated by our group, and the results indicated that RNase Mini Kit is an appropriate kit for extraction of RNA, miRNA as well as DNA (Uchimoto, 2014). Blood samples were collected and extracted as described in chapter two. Two DNA digestion kits, three random primers, and two Reverse Transcriptase enzymes were investigated. In addition, semen extraction with (+DTT) or without (-DTT) reagent was also explored. The stability and sensitivity of some markers were also tested.

### **3.3. Influence of different reagents on qRT-PCR experiment**

#### **3.3.1. Reverse transcription-priming strategy**

In this study, three different primer types were compared on cDNA from the fresh blood samples and those processed one-week post-collection. These included random decamer (10 bp), random hexamer (6 bp) and oligo-dT (12-18 bp) primers. The cDNA synthesised underwent the same protocol which was described earlier (chapter two) using M-MLV RT. cDNA obtained was amplified using TaqMan primers and probe targeting 5'-aminolevulinate synthase 2 (ALAS2) and GAPDH because both markers have poly-A tails. The  $\Delta Cq$  values were obtained for each marker for clearer representation of the data ( $\Delta Cq = 40 - Cq$  of GOI). Value of 40 in the equation is commonly chosen arbitrary value for transforming Cq data and originate in the fact that most qPCR reactions run for 40 cycles. The results showed the highest  $\Delta Cq$  value in both genes when hexamer primers were used, indicating a highest quality of reverse transcription (Figure 6).

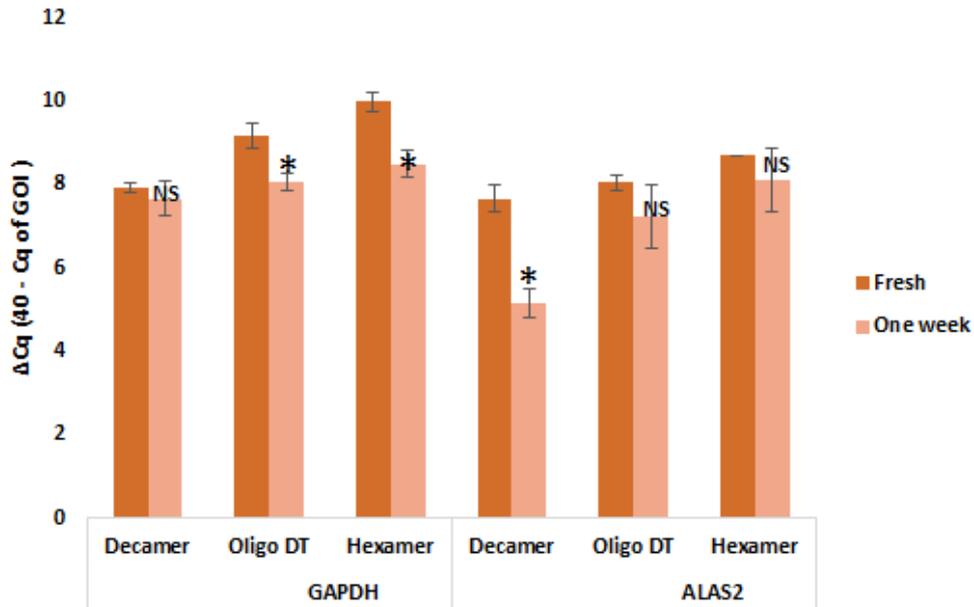


Figure 6. Amplifications of ALAS2 and GAPDH in blood samples with different RT primers. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and that with NS indicates No significance. Error bars represent one standard deviation (N=3).

Cycle threshold value (Cq) is logarithmic in nature, so it is normally distributed (Svingen *et al.*, 2014). To support this statement, the normality of  $\Delta Cq$  was tested using Shapiro-Wilk test (Shapiro and Wilk, 1965), and the results showed that data are normally distributed, ( $p > 0.05$ , Table 8). This finding allowed the usage of the T-test for each marker between fresh samples and those that were one-week old, and the results showed that for Hexamer and oligo dT primers the significance was achieved only when GAPDH marker was used, whereas, with decamers significant difference was achieved with ALAS2 only (Figure 6). All random primers gave higher amplification in fresh blood samples when compared to those processed one week after collection. This is particularly obvious for decamer primers in ALAS2 marker and hexamer primers in GAPDH. The oligo-dT, although to a lesser extent, also showed similar patterns in both markers. The main disadvantage of oligo-dT primer is that the target gene must have poly-A tail, which is not the case for all the genes. In conclusion, hexamer primers emerged as the best choice to prime cDNA, particularly in the aged samples. This is

because these primers have short nucleotides, and therefore the option to bind degraded fragments.

Table 8. Normality test for Random primers using GAPDH and ALAS2 in blood samples.

	Age	Shapiro-Wilk test for normality		
		test value	df	p value
Decamer in GAPDH	.00	.807	3	.132
	7.00	.909	3	.414
Oligo-dT in GAPDH	.00	.933	3	.501
	7.00	.952	3	.579
Hexamer in GAPDH	.00	.908	3	.411
	7.00	.909	3	.414
Decamer in ALAS2	.00	.938	3	.520
	7.00	1.000	3	.977
Oligo-dT in ALAS2	.00	.810	3	.139
	7.00	.945	3	.549
Hexamer in ALAS2	.00	.970	3	.666
	7.00	1.000	3	.994

The second experiment was conducted to investigate two different Reverse Transcriptase enzymes, namely Molony Murine Leukemia Virus (M-MLV) (Life Technologies, UK) and first strand Multiscribe synthesis (Life Technologies, UK). The cDNA was synthesised using hexamer primers targeting ALAS2 and GAPDH markers in fresh and one-week old blood samples. The finding from this experiment demonstrated that  $\Delta Cq$  of both primers with Multiscribe first strand synthesis enzyme was higher than M-MLV enzyme (Figure 7). The T-test also showed there is a significant difference between the primers with both RT enzymes in fresh and one week-old blood samples ( $p < 0.05$ ).

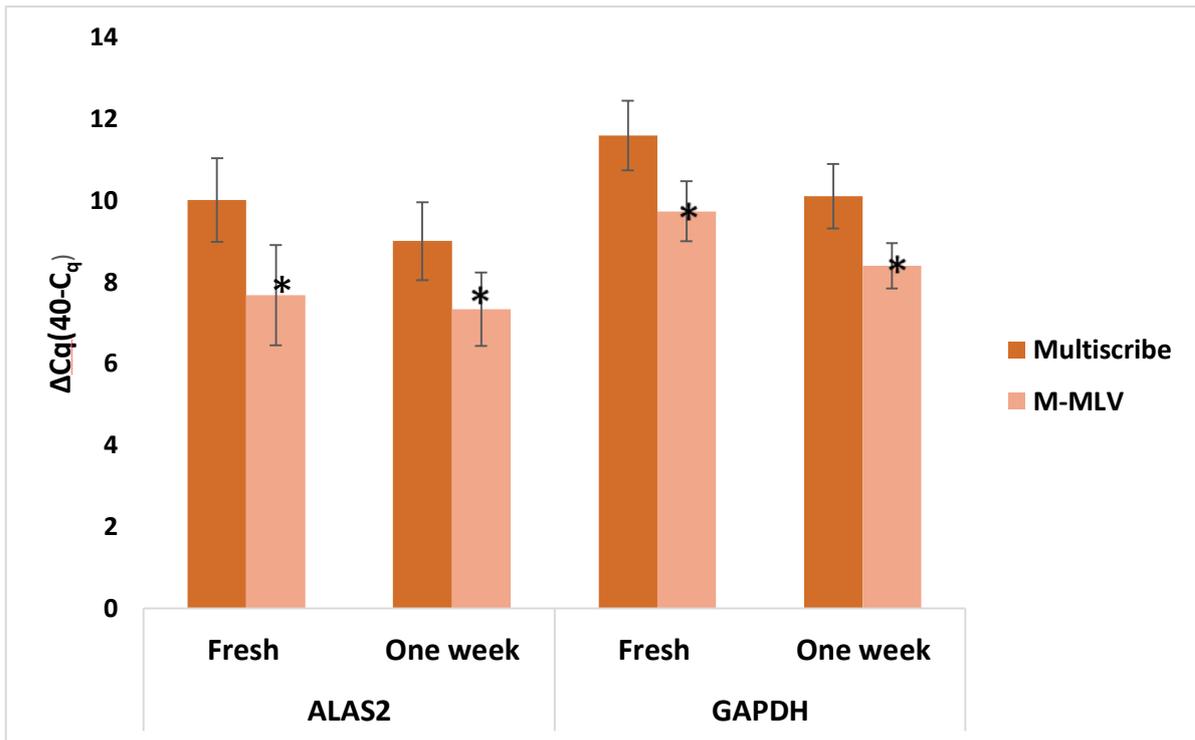


Figure 7. Comparison between ALAS2 and GAPDH with different RT kits. Column denoted with one asterisk (\*) indicates  $p < 0.05$ . Error bars represent one standard deviation (N=3).

### 3.3.2. Assays specificity and sensitivity

New method of interest in forensic research for the body fluids identification is the analysis of cell-specific mRNA expression. To date, few mRNA markers have been proposed for use in blood, saliva, semen, menstrual blood and vaginal secretion (Haas *et al.*, 2011a, Haas *et al.*, 2008, Roeder and Haas, 2016, Wang *et al.*, 2013b). In the present work, all assays gave specific results for detection of blood, saliva, and semen samples. In blood, ALAS2, HBB, and GYPA were investigated. The  $\Delta Cq$  of three markers were calculated, and the results indicated that HBB was abundantly amplified, ALAS2 moderately and GYPA significantly less (Figure 8). This finding was further supported by the study where HBB was the most highly expressed marker in the blood sample when compared with all other markers used (Haas *et al.*, 2011b).

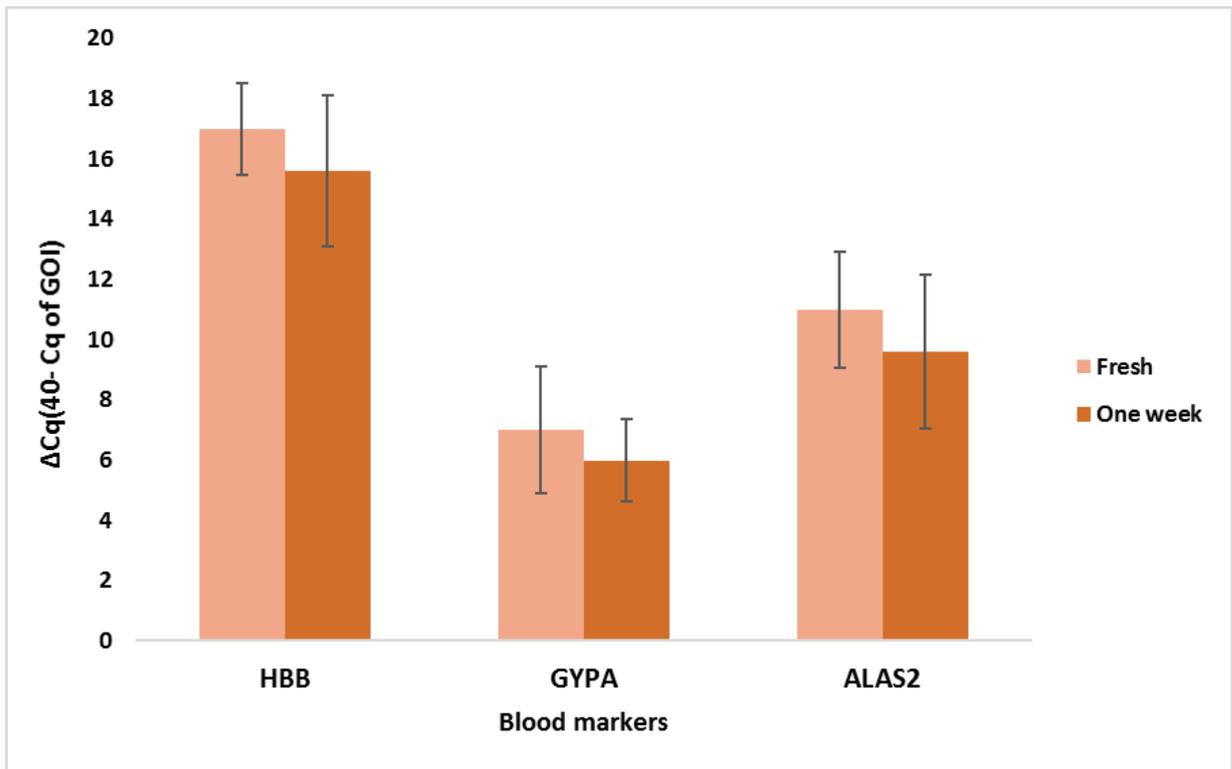


Figure 8. Amplifications of three blood markers HBB, GYPA and ALAS2. Data presented as mean  $\pm$  SD (N=3).

Following this, the specificity of the assay was tested in an experiment designed to examine two different strategies for obtaining cDNA. These were subsequently used in qPCR with primers targeting ALAS2, B2M, GAPDH, and ACTB markers. In the first experiment, following RNA extraction, Turbo DNA free kit (Thermo Fisher Scientific, UK) was used in order to achieve the complete digestion of DNA. The second experiment did not include DNA digestion step. In both cases, reverse transcription was performed in the presence of hexamer primers and Multiscribe RT (Figure 9). The results showed that all markers were successfully amplified when both cDNA were used as templates, but the amplifications occurred at a different rate. It is easy to conclude then that RNase mini kit, despite the claims from a supplier, is a co-isolation kit, meaning it will isolate both RNA and DNA simultaneously. Therefore, the DNA digestion is a necessary step in RNA quantification experiments. Only

ALAS2 and B2M primers gave no amplification with a negative sample when second cDNA template, created without DNA digestion (gDNA only) (Figure 10). This was expected since these primers span exon-exon junction and therefore will not detect genomic DNA (Ginzinger, 2002, Biosystemic, 2008). The  $\Delta C_q$  of all the assays from the second experiment (without DNA digestion step and therefore with both cDNA and gDNA being made in RT step) were higher when compared to those obtained for the first experiment (where DNase was used and only cDNA was present at the end of RT), suggesting a high amplification was detected. These differences could be caused by the presence of Turbo DNase in the first experiment which could inhibited RT reaction (Haas *et al.*, 2011b).

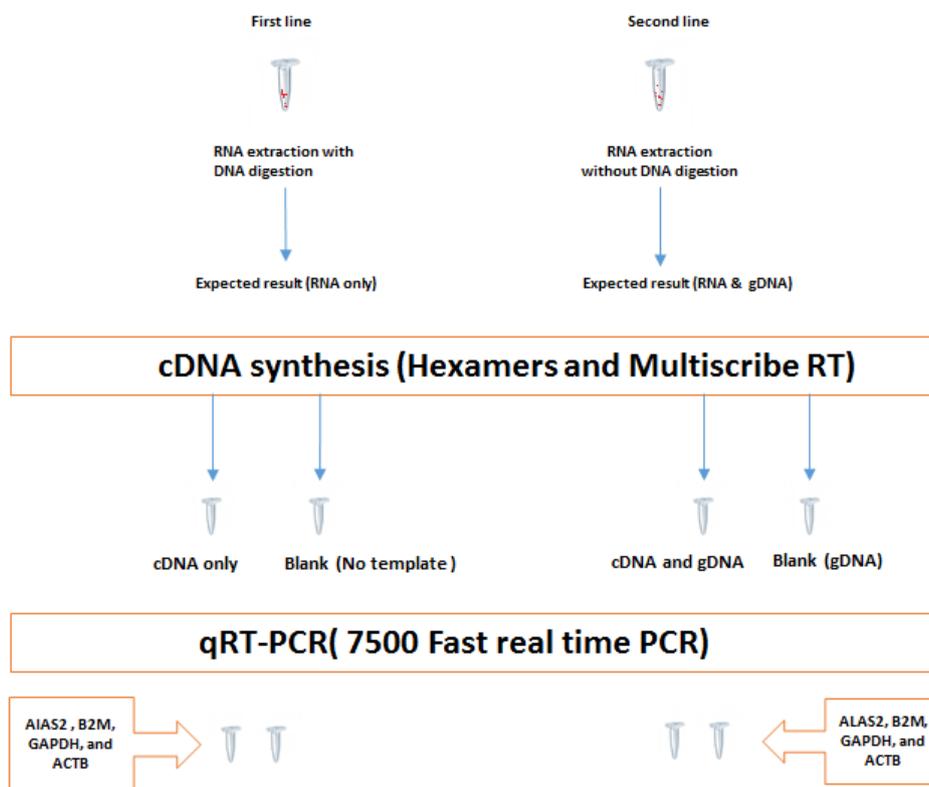


Figure 9. Assays specificity in a fresh blood sample using different cDNA synthesis targeting ALAS2, B2M, GAPDH, and ACTB. Blank control indicates those for which Reverse Transcriptase was not used.

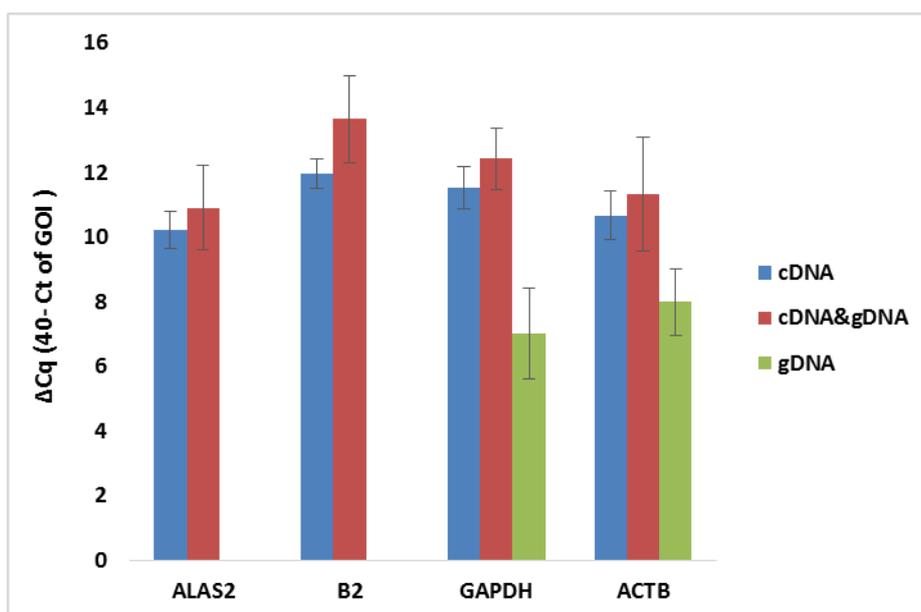


Figure 10. Amplification of assays in different templates (cDNA, cDNA and gDNA; and gDNA). Data are presented as mean  $\pm$  SD (N=one sample with three replicates).

Figure 10 shows the specificity of four markers used. As described, it is highly likely for gDNA to be co-isolated during RNA extraction step. That is why extracted RNA should be treated with DNA digestion kit. If no DNA digestion is performed during RNA extraction step, as was the case with the second experiment, the negative control (blank control) from RT step will contain gDNA. For this reason, the markers whose primers do not span an exon-exon junction will detect this gDNA. This was the case for GAPDH and ACTB in the present experiment (Figure 10).

### 3.3.3. DNA digestion reagents

To examine previously suspected influence of Turbo DNase on the amplification of the markers, two different DNA digestion kits, Turbo DNA free and RNase-Free DNase set, were compared. In this experiment, blood samples underwent RNA extraction process, followed by DNA digestion using each of the kit described as per suppliers' instructions (chapter two for details). Further, cDNA was synthesised using hexamer primers and Multiscribe RT enzyme

as described previously. The quantification RT-PCR was conducted and ALAS2, B2M, GAPDH, and ACTB markers were targeted. The result indicated that  $\Delta Cq$  value ( $40 - Cq$  of GOI) of these markers was higher when using RNase-free DNase kit (Figure 11). Lower  $\Delta Cq$  obtained for the samples treated with Turbo DNA Free suggested that the reagents in this kit act as inhibitors of PCR reaction, the finding already described by Haas *et al.* (2011b) (Haas *et al.*, 2011b).

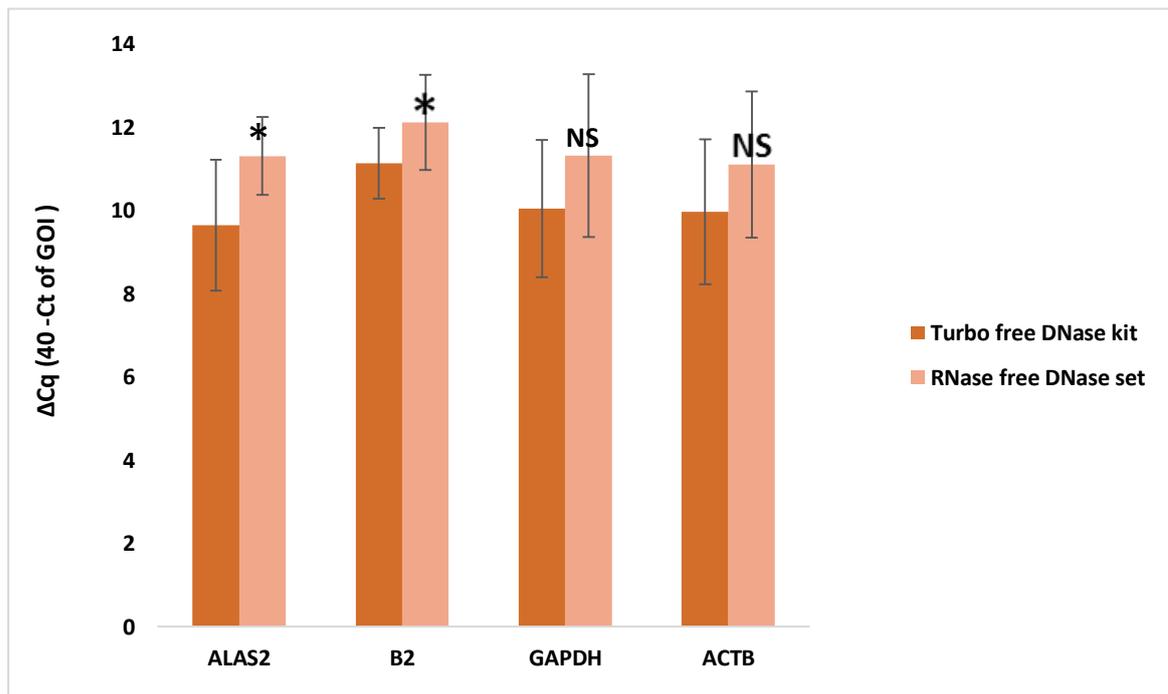


Figure 11.  $\Delta Cq$  for primers using different DNA digestion kits. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and column with NS indicate no statistically significant difference. Data presented as mean  $\pm$  SD (N=3).

### 3.3.4. Influence of DTT on semen extraction

To extract the right amount of RNA/DNA during PCR analysis that targets semen markers, reagents such as carrier RNA, Proteinase K with/without sodium dodecyl sulphate (SDS) and DTT should be taken into considerations. Carrier RNA prevents permanent binding of small amounts of target nucleic acid typically present in semen samples to the silica membrane of the RNA extraction spin column. Therefore, an addition of carrier RNA during sperm RNA

extraction is recommended to increase the recovery of total RNA (Kildemo, 2012) . Up to date, no publication was found on the influence of DTT on RNA and miRNA extraction. Therefore, in the present study, the influence of DTT on RNA/miRNA recovery and their degradation patterns were investigated. Fresh semen samples underwent total RNA extraction with and without DTT followed by cDNA synthesis using hexamer primers and Multiscribe RT. qPCR was conducted and mRNA markers including sperm *protamine 1* (PRM1), *semenogelin 1* (SEMG1), and B2M, as well as miRNA markers (891a and RNU44) were targeted (Figures 12 and 13).

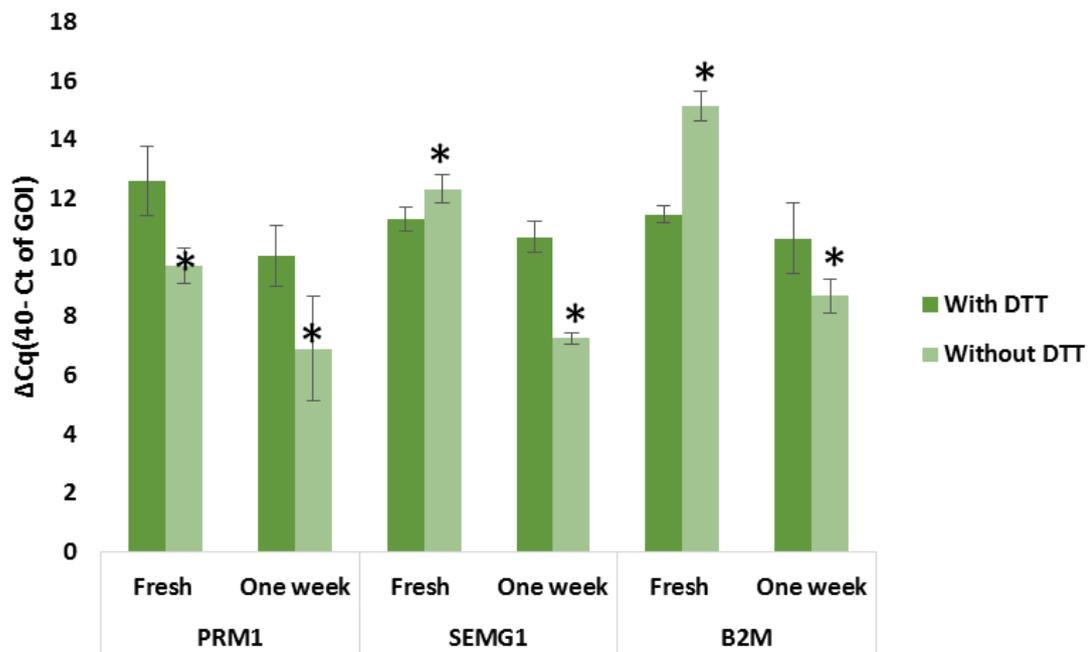


Figure 12. Quantification PCR data using TaqMan primers and probe in fresh and one-week old semen samples.  $\Delta Cq$  of PRM1, SEMG1, and B2M markers in semen with and without DTT. Column denoted with one asterisk (\*) indicates  $p < 0.05$ . Error bars represent one standard deviation (N=3).

As seen in the Figure 12,  $\Delta Cq$  of B2M was the highest, indicated the most abundant B2M in the absence of DTT, whereas the highest yield of PRM1 was achieved with the extraction in the presence of DTT.

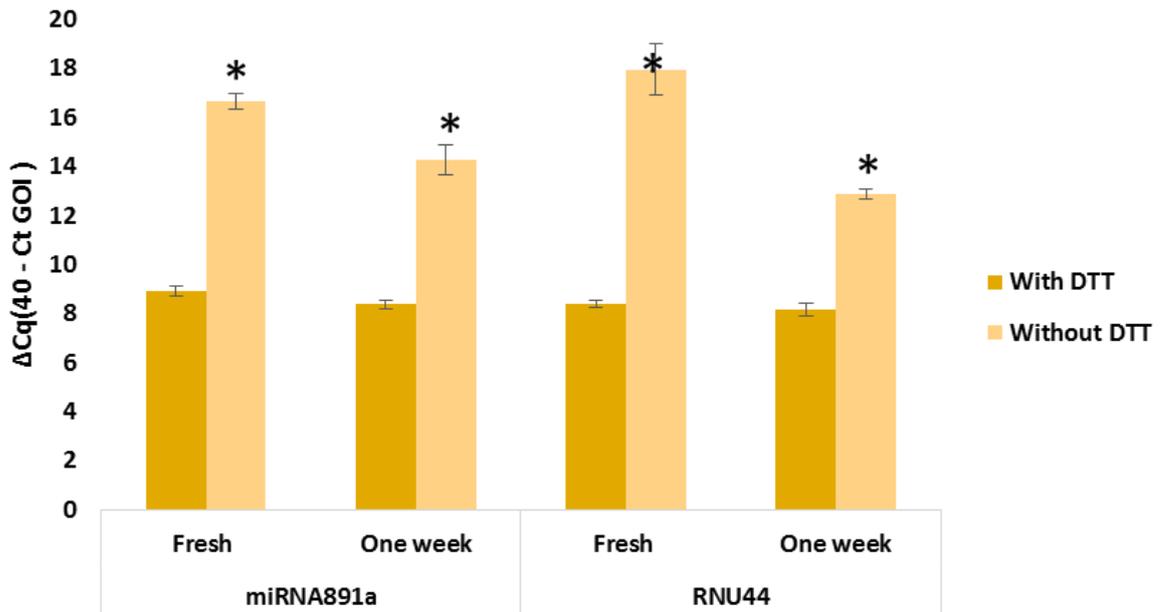


Figure 13. Comparison between two extraction methods of miRNA markers.  $\Delta Cq$  of miRNA 891a and RNU 44 markers were extracted with and without DTT. Column marked with one asterisk (\*) indicates  $p < 0.05$ . Error bars represent one standard deviation (N=3).

Unlike mRNA markers, higher yield for both miRNA was achieved in the absence of DTT. In addition, a significantly higher amount of miRNA 891a was extracted from fresh samples when compared to those that were one week old, and this was true for both extraction methods (Figure 13). These findings indicated that miRNA could be more susceptible to DTT reagent because both markers were the most abundant when DTT was avoided. It is possible that the location, nature and the function of miRNA markers are, at least partially, responsible for this behaviour. Further, the overall abundance of markers was not the same in fresh and one-week old samples, which could be due to the degradation of chosen markers in the older samples.

### 3.4. Optimisation of primers

Optimisation of qPCR conditions is very important for the development of a robust assay. On the other hand, poor optimization leads to the lack of reproducibility between replicates as well as inefficient and insensitive assays. Two main approaches available to overcome this issue are optimisation of primer concentrations and/or annealing temperatures.

All TaqMan assays were performed successfully according to manufacturer's instruction, and no amplification was detected in negative control. For SYBR Green assays optimisation of primer concentration is essential. Typically, using less primers is better and optimal results may require different ratios of forward to reverse primers. Therefore, primers with SYBR Green were optimised into 10 pmol/ $\mu$ l (0.5  $\mu$ l of each primer per reaction) and showed good results in almost of the samples. However, there were few exceptions to this rule where, unexpectedly, amplification was detected in negative controls. A positive signal in the no template controls is the most commonly reported problem with SYBR Green assays. This could be due to contamination in the samples, as well as primer-dimer effect. Alternatively, it could results from the product of the SYBR Green fluorescent dye itself. Indeed, the amplification detected in the negative control samples when using SYBR Green dye in the previous study was attributed to the technique itself rather than contamination of the samples (Connolly and Williams, 2011). To understand this anomaly better, it is the standard practice in experiments with SYBR Green for the melting curve analysis to also be employed. Melting curve is used to confirm that the amplification detected is related to the specific target sequence. The results from the melting curve analysis demonstrated that all amplifications were indeed detected as complementary to the primer used, and that all have similar  $T_m$ . Example of the melting curve in Figure 14 shows the amplification peaks of ACTB marker in blood with approximately the same  $T_m$  confirming that all amplification occurred in the same target sequence, whereas the negative control with the same marker showed peak at different,

and much lower  $T_m$ . It is generally accepted that  $T_m$  in non-template controls that is 5 cycles or higher away from  $T_m$  of target sequence will not affect the interpretation of the data (Bustin and Nolan, 2004).

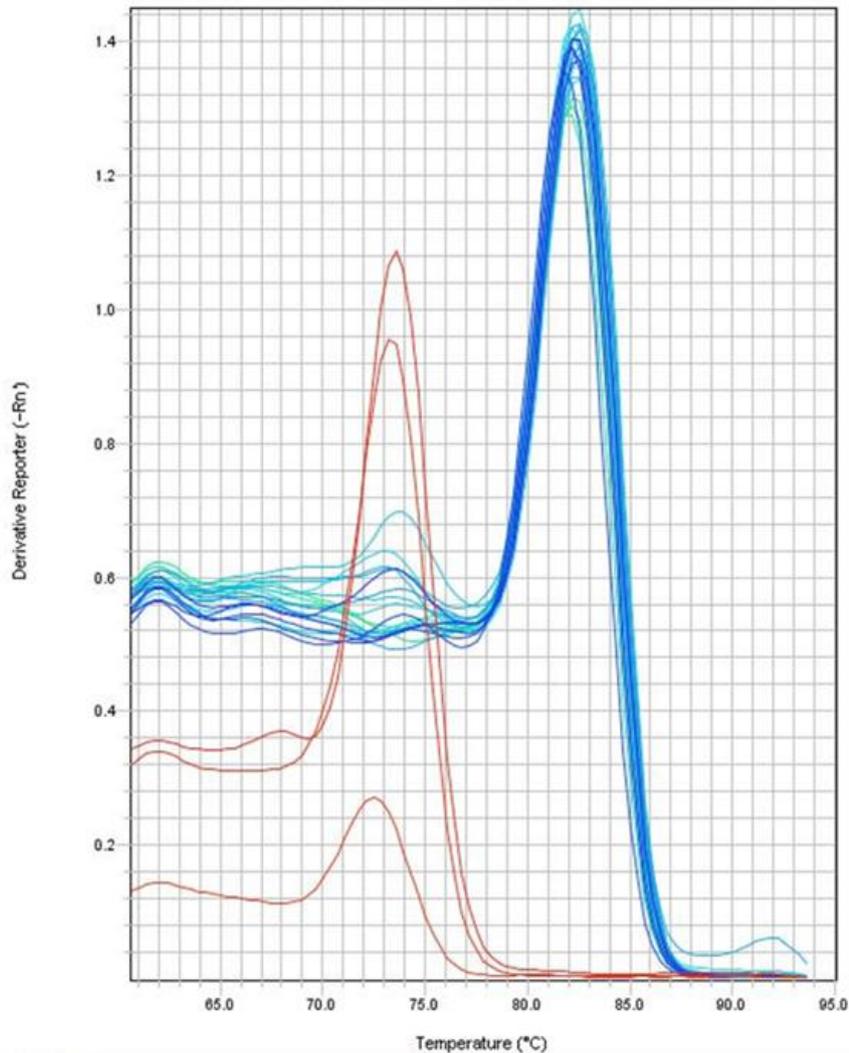


Figure 14. The melting curve of the amplification of ACTB in blood using SYBR Green. Blue peaks represented ACTB and show that all amplification occurred in the sample target sequence. Red peaks represent negative control. For red peaks, the amplification occurred at the lower temperature compared to the blue peaks. This figure was generated from in house data.

In this study, the oxygen regulated factors were also investigated and optimized with SYBR Green chemistry. The HIF1A and VEGFA could be used as useful markers to predict the age of blood, saliva, and semen. These primers were successfully detected in all of the samples in the present study, which will be discussed in more details in chapter five.

### **3.5. Validation of reference genes for normalising RNA in blood samples**

Control genes, previously known as housekeeping genes and now commonly referred to as reference genes, are frequently used to normalise mRNA levels between different samples. However, the expression of some of the genes commonly used as reference is shown to vary among different tissues or cells and can be influenced by the treatment used in the study (Dheda *et al.*, 2004). Thus, the selection of reference genes for quantification of RNA studies is critical for accurate normalisation and correct interpretation of the data. To address this issue, six genes which are the most commonly used in this type of research (Moreno *et al.*, 2012, Roeder and Haas, 2013, Haas *et al.*, 2011a) were investigated in blood. For this, a simple  $\Delta Cq$  approach was applied by comparing relative quantification of all assays in fresh and one-week old samples. All markers were amplified in both types of samples, however,  $\Delta Cq$  values showed the highest gene amplification with B2M, whereas GUSB gave the lowest (Figure 15). The present results were consistent with those previously reported where B2M was shown to exhibit the highest amplification in all body fluids (Moreno *et al.*, 2012). T-test was performed to check for differences between fresh and one-week old blood samples, and the difference was observed only in case of GAPDH.

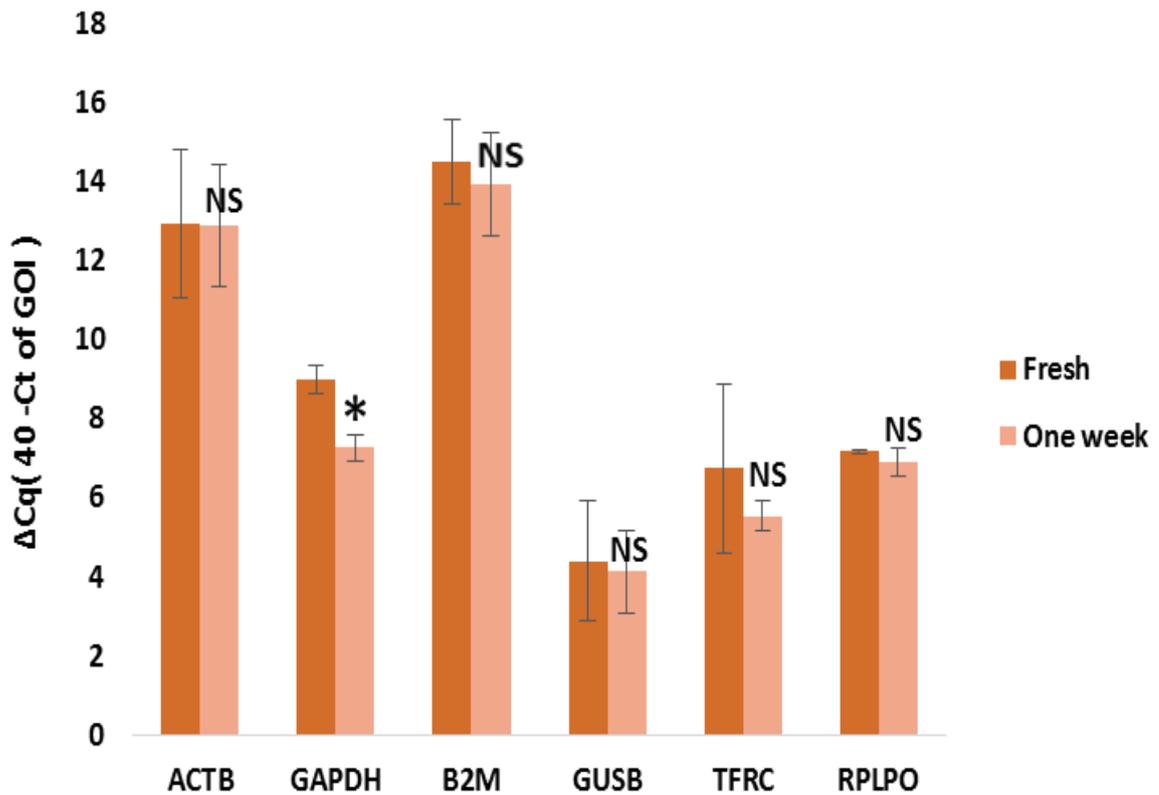


Figure 15. Delta Cq values of TaqMan assays in fresh and one-week old blood samples. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and column with NS indicate no significant difference. Error bars represent one standard deviation (N=3 for each target in each time).

The amplification levels of reference genes should remain stable between the samples obtained from the different tissues and under different experimental conditions (Thellin *et al.*, 1999, Dheda *et al.*, 2004). At a 100% reaction efficiency, 1 cycle threshold indicates a 2-fold change, and the variability of the individual reference genes is reflected as standard deviation and range, expressed as an average fold change or maximum fold change, respectively. To check the stability of assays, the same data underwent the box plot analysis (Figure 16).

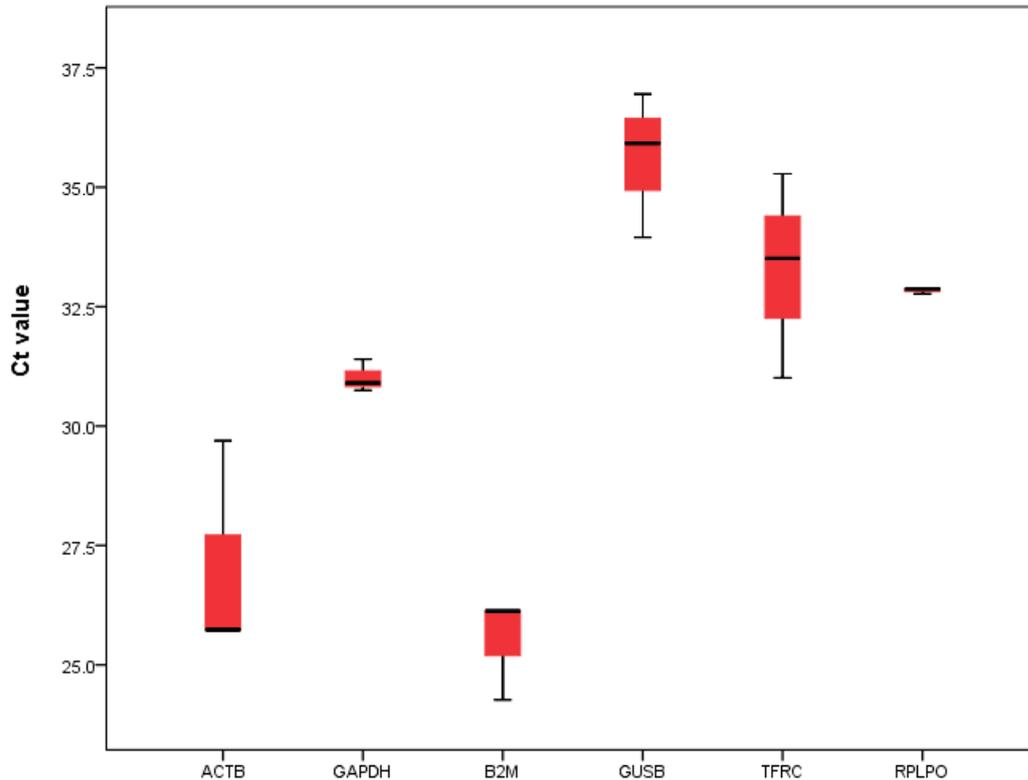


Figure 16. Box plot analysis of reference genes in fresh blood samples. A line across the box is depicted as the median. The box indicates the 25th and 75th percentiles and the whiskers caps represent the maximum and minimum values. (N=3) for each.

### 3.6. Validation of $\Delta\Delta Cq$ experiment

Accurate real-time PCR results are highly dependent on reagents used, experimental set-up, and sample quality. Investigation of most of these factors was already described, and the results showed the importance of optimisation of any experiments. In addition, for the  $\Delta\Delta Cq$  calculations to be valid, the PCR efficiency of target gene must be approximately equal to the PCR efficiency of the reference gene. The adequate method to apply and check if two amplicons have the same PCR efficiency is to look at how  $\Delta Cq$  changes with template dilution. PCR amplification efficiency is determined as a percentage of the rate at which a PCR amplicon is generated. If a particular PCR amplicon is doubled in quantity during the geometric phase of its PCR amplification, then the PCR assay has 100% efficiency. If the amplification efficiency of the two PCR reactions is the same, the scatter plot of log input

amount versus  $\Delta Cq$  has a slope nearly zero. Therefore, in order for a validation experiment to pass, the absolute value of the slope of  $\Delta Cq$  vs. log input should be less than 0.1.

A validation of an experiment is necessary to determine if the  $\Delta\Delta Cq$  calculation is applicable. However, TaqMan® Gene Expression Assays have amplification efficiencies of 100% in high-quality samples. Applied Biosystems (AB) reports 100% efficiency ( $\pm 10\%$ ) in the resulting assays when measured over a 6-log dilution range, provided that samples are free of contaminants, such as carry over phenol, proteins, and inhibitors. In this study, experiments were performed to investigate some primers used, to make sure they have the approximately the same PCR efficiency of reference genes, using the same procedure described by Life Technologies (Biosystems, 2004).

For this experiment, blood samples were selected, RNA extracted, and DNA was digested using on-column DNase digestion with the RNase-Free DNase kit according to the manufacturer's protocol. Reverse transcription reaction was performed using random hexamers and Multiscribe First Strand Synthesis Transcriptase Enzyme (Life Technologies, UK) as described by manufacturer's protocol. Quantification of the cDNA yield was performed using a standard curve method. For this purpose, 1  $\mu$ l of Burkitts Lymphoma (Raji) AM7856 (Life Technologies) was reverse transcribed into cDNA in a 20  $\mu$ l reaction. This cDNA yield was used as a standard to quantify cDNA obtained from a fresh blood sample. The standard curves were generated by amplifying the Raji cDNA with chosen primers. The Raji cDNA concentration was estimated nominally, as 50 ng/ $\mu$ l by assuming that total RNA was reverse transcribed into cDNA. The Raji cDNA was serially diluted by one-third such that the starting concentration was 50 ng/ $\mu$ l and the lowest was 0.023 ng/ $\mu$ l creating 8 different concentrations (50, 16.7, 5.56, 1.85, 0.62, 0.21, 0.069, 0.023 ng/ $\mu$ L). 2  $\mu$ l of the template was used for each qPCR reaction.

As mentioned earlier, for validation of  $\Delta\Delta C_q$  calculations both efficiencies of the target and the reference gene must be approximately equal. In this study, primers for both the target and reference genes were calibrated using the same source (fresh samples). In addition, TaqMan assays which chosen were custom designed (on demand) and already shown to have amplification efficiency of 100% (Biosystems, 2004).

Calculation of the relative efficiencies for amplification of the target gene and the reference (endogenous control) gene was achieved by running standard curves for blood primers (ALAS2 and HBB) with B2M as a reference gene. The  $C_q$  values generated from comparable standard curve mass points (ALAS2 or HBB vs. B2M) are used in the  $\Delta C_q$  calculation. Finally, the  $\Delta C_q$  values are plotted vs. log input amount to create a semi-log regression line. The experiment passed the validation, because the slopes of the resulting semi-log regression lines are  $< 0.1$  (Figure 17).

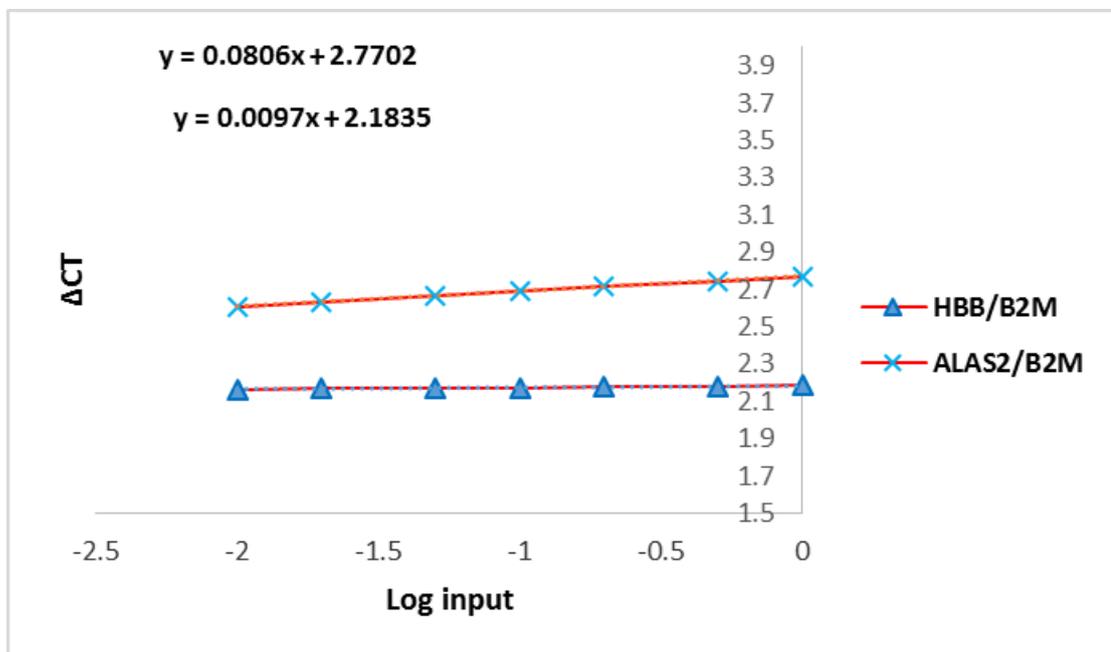


Figure 17. Validation plot of  $\Delta C_q$  vs. log input amount of RNA (HBB with B2M:0.0806; ALAS2 with B2M:0.0097).

The assays with SYBR Green were also validated. Here, two oxygen –regulated factors (VEGFA and HIF1A) were selected and normalised to ACTB as a reference gene. The standard curves for VEGFA, HIF1A and ACTB mRNA transcripts were linear over four orders of magnitude (Figure 18). The amplification efficiency of each target was approximately equal to that of the reference gene (ACTB), which provides the validity of using the  $\Delta\Delta Cq$  method for relative quantification.

Many studies have been conducted to validate RQ method. Validation of GAPDH as a reference gene with some chosen markers including B2M, RPLPO, and ACTB was investigated. Similar efficiencies were obtained for reference and target genes regardless of initial RNA concentration before performing a comparison study, and slopes were consistently close to zero (Moreno *et al.*, 2012). RQ validity was also tested when the standard curves for EPO, VEGF, HIF1A and GAPDH mRNA transcripts were created. These were linear over four orders of magnitude for most of the markers, except for HIF1A. The amplification efficiency of each target was equal to that of the reference gene (GAPDH), which confirmed the validity of using the  $\Delta\Delta Cq$  method for relative quantification (Zhao *et al.*, 2006).

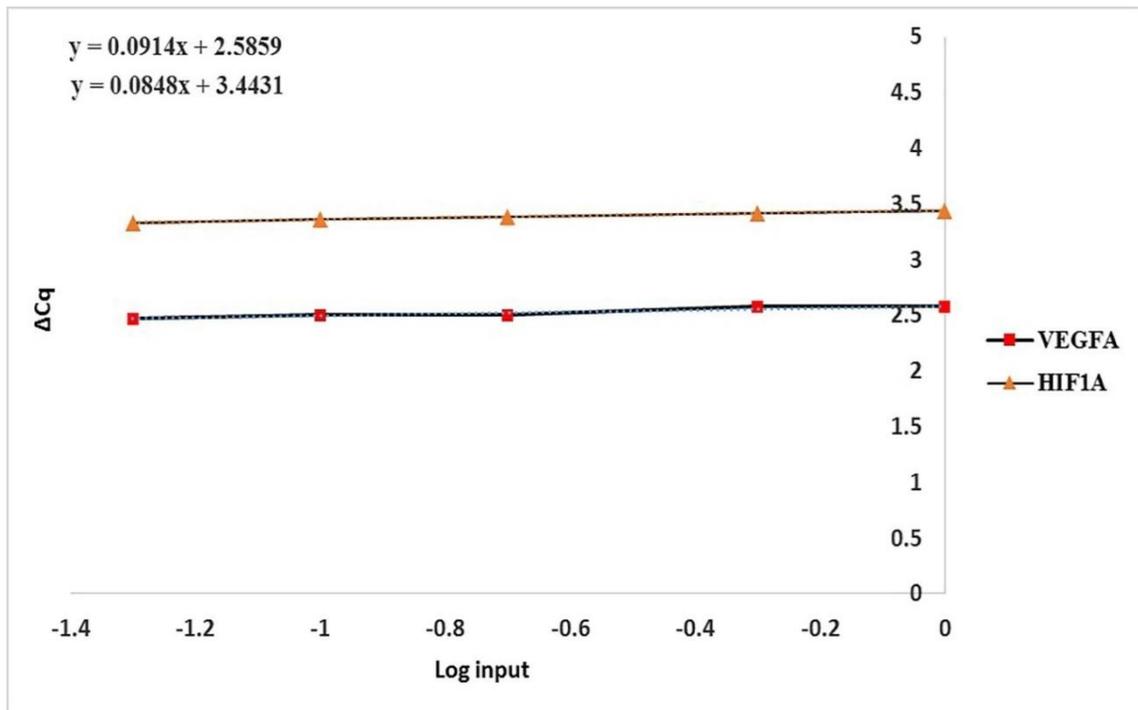


Figure 18. Validation of the experiment by plotting  $\Delta Cq$  vs. log input amount of RNA. All slopes were consistently close to zero (VEGFA with ACTB: 0.091; HIF1A with ACTB: 0.0848).

### 3.7. Discussion

RT-qPCR is a widely used method of quantification of RNA in a target sample. Many reagents and protocols are available from different suppliers that guarantee high quality performance of the assay. The selection of adequate reagents, target and reference genes to be used is one of the most challenging steps. Understanding the patterns of gene expression for any biological samples is very important because it provides information about genes relevant to the biological processes. However, for an adequate interpretation of the results, quantification of RNA studies requires an appropriate normalisation strategy.

The aim of this chapter was divided into two parts; first, to perform the necessary investigation of the reagents needed to conduct RNA quantification experiment and how they influence accuracy and reproducibility of the results. The second part focused on the stability of the reference genes and their validation for RQ analysis. RT-qPCR consists of a number of

stages, all of which need to be completed successfully in order for the accurate results to be obtained. Each of these stages have different reagent and protocols which all need to be carefully chosen and justified. RNeasy mini kit for total RNA extraction (Qiagen) achieved the second highest ranking when five different kits were compared (Grabmüller *et al.*, 2015), and it was selected as the main extraction kit in the present study. Reverse transcription is a critical step that can be performed by using variety of different random primers and RT enzymes available. With this in mind, Stahlberg *et al.* (2004) emphasised the importance of controlling for the variations in RT step of RT-qPCR. For example, oligo-dT primers are used to synthesise long, not fragmented cDNA, since these primers only anneal to the poly-A tail junction of the mRNA template (Stahlberg *et al.*, 2004). All cDNA made in this way will contain the 3' end of the gene. However, if a gene is long, the primer can sometimes fall off, leading to the missing 5' end. Random hexamer primers, on the other hand, randomly bind along the RNA, so that the resulting cDNA represent fragment of the gene, with all its regions possibly not equally represented. Therefore, if GOI is near the 3' end of the gene, then using oligo-dT primers to synthesise cDNA may indicate higher amplification than if random hexamer primers were used and vice versa. In other words, different priming strategies could lead to the differences in the way different regions of the gene are represented in the cDNA that is being synthesised. This could then lead to the differences in PCR-generated amplification levels of these regions.

In this chapter, the comparison of three different random primers was conducted, and the result showed GAPDH and ALAS2 were more abundant when random hexamers were used (Figure 6), confirming them as the best choice to generate RT-qPCR experiment in the present study. These findings were in agreement with the study that reported approximately 19-fold increase in the calculated mRNA copy number from cDNA synthesis reactions primed with random hexamers (Zhang *et al.*, 1999). Another study found that Cq value was not stable

when five genes were compared including  $\beta$ -tubulin, GAPDH, Glut2, CaV1D, and insulin II genes, using random hexamers, oligo-dT, and gene-specific reverse transcription primers. It was suggested that the efficiency of the reverse transcription reaction depends on the priming strategy and also varies among different genes (Ståhlberg *et al.*, 2004).

Okello *et al.* (2010) assessed the sensitivity and reproducibility of 11 commercially available Reverse Transcriptase enzymes in cDNA synthesis from low copy number RNA levels, including M-MLV. Their findings supported use of Accuscript or Superscript III when dealing with low copy number RNA levels rather than using M-MLV (Okello *et al.*, 2010). This was consistent with the data presented in this chapter where for both ALAS2 and GAPDH the highest  $\Delta Cq$  was obtained with Multiscribe RT enzyme (Figure 7).

DNA digestion is an essential step for quantification of RNA experiment, because RNeasy mini kit is a co-isolation kit. Consequently, gDNA would amplify if DNA digestion step was missed. In addition, in this study, not all primers spanned an exon-exon junction expecting to amplify DNA (Figure 10). On-column DNase digestion with the RNase-free DNase kit was proven to be the preferred method of DNA digestion because it did not influence or inhibit reverse transcription and qPCR and it generated acceptable  $\Delta Cq$  when compared with Turbo DNA free kit (Figure 11). This was consistent with the previous reports that showed that an addition of an extra volume of Turbo DNase inhibited the PCR reaction (Haas *et al.*, 2011b).

DTT is a reducing agent and its influence on RT-qPCR was tested in semen markers. The results showed that  $\Delta Cq$  of PRM1 was significantly increased in the presence of DTT, whereas  $\Delta Cq$  of SEMG1 remained relatively stable (Figure 12). This was, perhaps, not surprising, since SEMG1 is a seminal vesicle and prostate specific gene and PRM1 is spermatozoa specific gene. Hence, disulfide cross-links in the *protamine* surrounding the sperm are broken when treated with reducing agent such as DTT (Chapman *et al.*, 1989,

Nakanishi *et al.*, 2014). In miRNA markers of the semen samples, on the other hand,  $\Delta Cq$  of 891a and RNU44 was increased in the absence of DTT (Figure 13). The main reason could be the high sensitivity of miRNA markers to DTT or higher overall volume which may influence the amplification rate of these markers. Treatment of sperm cells with 100 mM DTT resulted in gradual decrease in the percentage of sperm cells/mL cell suspension observed over a period of 20 minutes (De Gannes, 2014).

The selection of GOI and reference gene for normalisation is one of the most important steps during RT-qPCR. Selecting the markers with adequate specificity, sensitivity, and abundance is an imperative. Here, ALAS2 and HBB emerged as specific blood markers and showed higher abundance than GYPA (Figure 8). The specificity of ALAS2 as a blood marker was investigated by evaluating highly specific mRNA markers to identify body fluids; eight markers were tested in a single and multiplex PCR reaction when ALAS2 showed a high specificity to detect only blood samples (Haas *et al.*, 2011b). Another study performed by Roeder and Haas (2013) showed that ALAS2 was detected at a low frequency in a non-target body fluid such as menstrual blood and cervicovaginal fluid (CVF) (Roeder and Haas, 2013).

Successful qPCR was conducted using both fluorescent chemistries, and initial screening for all assays was conducted. Boxplot analysis was carried out to check for outlier data and to provide more information about the stability of sex reference genes. The results demonstrated that GAPDH was the most stable reference gene (Figure 16). This was in contrast with the study conducted to assess the level of 13 reference gene including GAPDH, B2M, and ACTB. The results showed that GAPDH, B2M, and ACTB had an average fold of  $>2$  and a maximum variability of 20-to 35-fold which made them unsuitable as internal references in that particular experiment (Dheda *et al.*, 2004).

DNA melting curve analysis that uses SYBR Green I is widely used in the identification of body fluids (Stewart *et al.*, 2015, Li *et al.*, 2013a, Antunes *et al.*, 2016). Its advantage is that it enables diagnostic even in case of a failed reaction. In the study for determining the applicability of SYBR Green chemistry for Body Fluid Identification test (BFI), the melting curve analysis was performed to ensure that fluorescence detected did indeed correspond to the targeted amplicons (Connolly and Williams, 2011). In the present study, melting curve analysis was performed and the results showed that all of the amplifications were from the targeted markers: a series of peaks with the same  $T_m$  value was detected (Figure 14). If, on the contrary, all of the amplification was the result of contamination then a series of peaks with a variety of  $T_m$  values would be obtained (Connolly and Williams, 2011). For a current study, the melting curve was applied, and the results showed all amplifications were detected in real target sequences.

Relative quantification is the best method to quantify RNA in a biological sample. However, to obtain an accurate result and avoid any unspecific amplification, the experiment should be validated. As mentioned previously, relative quantification is valid only when amplification efficiency of target genes and reference gene is approximately equal (Biosystemic, 2008, Livak and Schmittgen, 2001). Therefore, the variation in efficiencies between markers and the chosen reference genes ALAS2 with B2M, HBB with B2M, HIF1A with ACTB, and VEGFA with ACTB were examined at various input concentrations of cDNA. In all cases, the verification of the experiments passed because the slopes were  $< 0.1$ .

### **3.8. Conclusion**

That experimental variation in reverse transcription-qPCR is essentially attributable to the reverse transcription step. Oligo-dT primers are not recommended as they require presence of poly-A tail for cDNA synthesis, and as such, they are sensitive to RNA degradation. Random hexamer primers, on the other hand, generate a short cDNA, and as such are considered not to

be affected by RNA degradation. Multiscribe Reverse Transcriptase enzyme is also recommended to prime cDNA, since it showed higher  $\Delta C_q$  than M-MLV, and it is suitable for the degraded sample (those with the low copy number).

Reverse transcription followed by PCR is confirmed to be a successful way for quantification any target gene. Relative quantification can be performed using a TaqMan probe or SYBR Green chemistries. In this chapter, both methods gave excellent preliminary results. SYBR Green has many advantages over TaqMan because it is cheaper, simpler, and its specificity can be easily improved by performing post-run melting curve analysis. Finally, it is important to demonstrate that starting RNA concentration has no significant effect on the amplification efficiency of any given marker. Indeed, similar efficiencies were detected between reference and target genes regardless of initial RNA concentration before performing a comparison study utilizing a comparative  $C_q$  method.

**Chapter Four: Degradation of mRNA as an  
indicator to predict the age of biological stains using  
TaqMan Markers**

#### 4.1. Introduction

Establishing age prediction of any biological sample would be one of the most significant findings in forensic science. The basic question that should be answered during the investigation of any crime is, who left the biological evidence and when? DNA analysis using PCR is able to identify the person that biological sample belongs to; however, it provides no information about the time when the sample was deposited.

TaqMan probe is one of the most common methods used to perform gene expression analysis. It relies on more specific detection of signal by the introduction of fluorogenic-labelled probes that hybridise to the target sequence during amplification in real-time PCR systems. The assay uses the 5'-3' exonuclease activity of *Taq* DNA polymerase, to cleave the probe and release the fluorescence signal. The intensity of the signal measured during an exponential phase of PCR is proportional to the amount of target DNA at the start of the assay (Navarro *et al.*, 2015).

When compared to DNA, cellular content of RNA is eight times more abundant. However, this molecule is also more labile and has a fast rate of degradation than DNA, all of which could allow for the estimation of the *ex-vivo* age of biological specimens (Smith, 2010). Many studies have been conducted in an attempt to solve this dilemma. Simard *et al.* (2011-2012) tested the stability of four different RNA targets up to six months old, namely 18S rRNA,  $\beta$ -actin mRNA, GAPDH mRNA and PPIA mRNA obtained from blood, semen, and saliva stains to determine TSD of these body fluids (Simard *et al.*, 2011, Simard *et al.*, 2012). No significant difference between rRNA and mRNA decay rates was observed, whereas Anderson *et al.* (2005) found a good correlation between relative mRNA/18S rRNA in the approximately the same time frame (Anderson *et al.*, 2005). In addition, the ratio of GAPDH

and *β-actin* was also investigated, and the result showed a linear downward expression of both genes over time (Connolly *et al.*, 2011).

#### **4.2. The problem under investigation**

It is usual to find the suspect's DNA in the area where the victim was, but the problem that forensic science faces is how to estimate the age of any biological samples. One example that proves the importance of determining the age of the sample and its impact on the resolution of the crime is Eltham district case in London in which young woman was seriously injured. What followed was the arrest of her friend and alleged lover because his clothes were found to contain number of blood spots that belonged to the victim. Despite this, no conviction occurred since the examiner was unable to claim the moment in which the blood was deposited on his shirt. Therefore, the perpetrator of this crime remains unknown (Guo *et al.*, 2013). In an attempt to overcome this limitation of forensic science, the aim of this chapter was to generate mathematical models based on the degradation patterns of selected TaqMan markers over a period of time at the room temperature.

#### **4.3. Experimental design**

The current study was conducted to quantify degraded RNA in old stains relative to the fresh sample (sample processed at the time zero) as a calibrator to predict the age of any biological stains via a formula of  $2^{-\Delta\Delta Cq}$ .

RT-qPCR technique was conducted using TaqMan probe markers for blood, saliva, and semen in two different time- frames. One-week experiment was only conducted for blood, while one-month long study was performed for blood, saliva, and semen, to investigate the degradation pattern of chosen primers. Simple and multiple regression analysis (SRA & MRA) were employed in an attempt to generate models that could be used to predict the age of the samples under investigation.

#### **4.4. Results**

The photometrical assessment of quantity and purity, as well as the measurement of RNA integrity of each sample was required. Absolute quantification test for RNA is currently not reliable (Uchimoto, 2014), thus, quantity and quality of extracted RNA was verified by an average A260/A280 and A260/A230 for proteins and carbohydrates, respectively. The acceptable values were obtained, because A260/A280 and A260/A230 values greater than 1.8 and 2.0 confirmed lack of proteins and carbohydrates contamination, respectively (Teare *et al.*, 1997, Silva *et al.*, 2015).

##### **4.4.1. Quantification of RNA degradation in one-week-old samples**

In this experiment, relative quantification was performed on blood samples using TaqMan probe only. The ALAS2 marker was selected as a target gene to look at its pattern of degradation over a given period of time (one week), while GAPDH and B2M were chosen as reference genes for normalisation. Therefore, target (ALAS2) level was normalised to endogenous control (GAPDH and B2M) level and then for any given day (e.g. day 7 in the present case) amplified at different level (increase or decrease) of the target gene at day 0. This was achieved by obtaining  $\Delta Cq$  of chosen markers, and subsequent calculation of RQ (Figure 19).

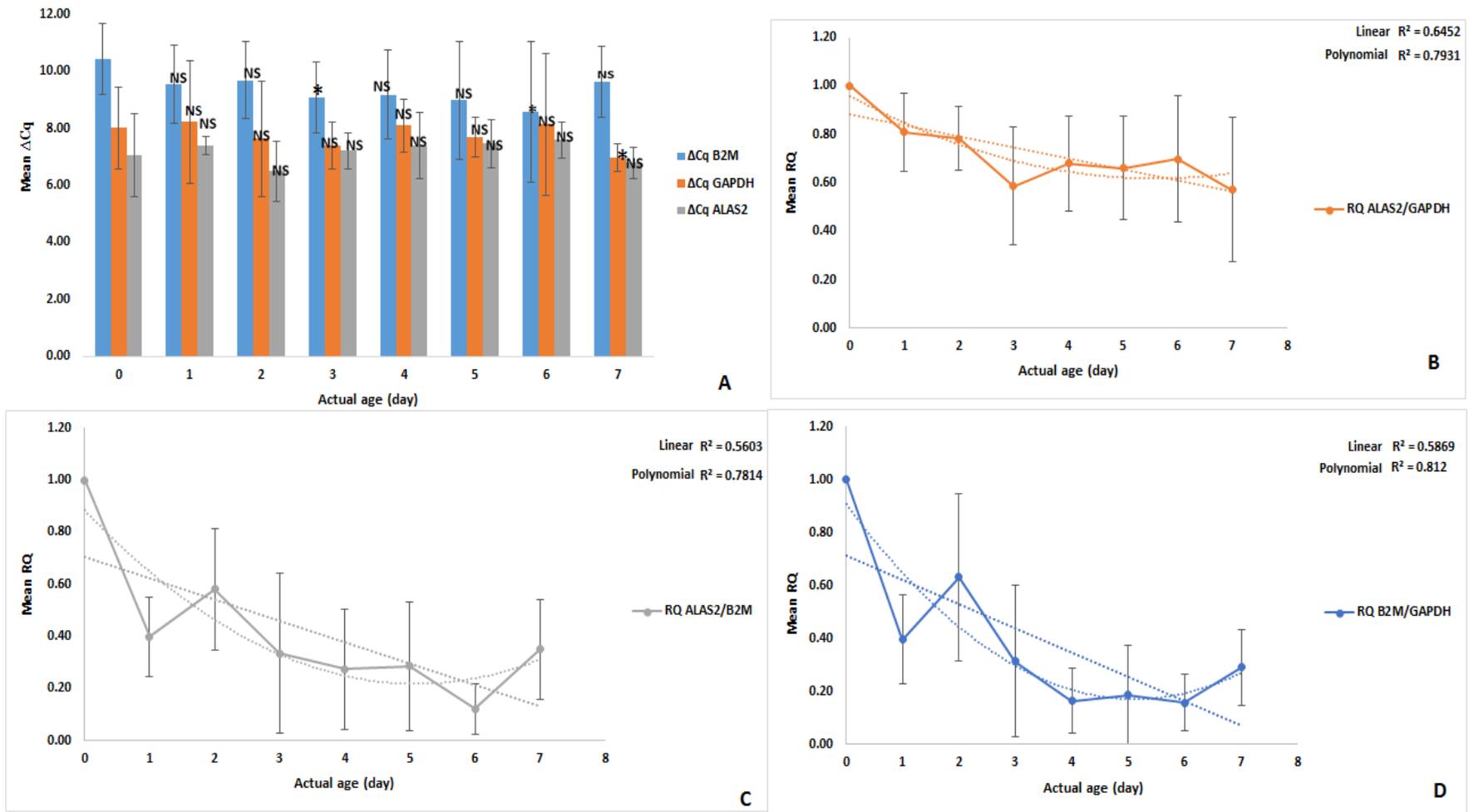


Figure 19. The degradation patterns of ALAS2, GAPDH, and B2M in blood samples using TaqMan chemistry. (A) Delta cycle quantification ( $\Delta Cq$ ) of assays over one week. (B) RQ patterns of ALAS2/GAPDH and (C) B2M (ALAS2/B2M) as reference genes. (D) The RQ patterns of B2M with GAPDH as the reference gene (B2M/GAPDH). Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and column with NS indicates no significant difference. Error bars represent one standard deviation ( $N=56$ ).

As shown in Figure 19 A, the B2M was the most abundant marker detected, and this finding was also supported by a previous study conducted to determine the effective housekeeping genes for the quantification of mRNA for forensic applications (Moreno *et al.*, 2012). The delta-Cq ( $\Delta Cq$ ) in qPCR data had symmetrical and normal distribution, whereas, the distribution of relative gene expression was skewed (Guo *et al.*, 2010). The normality test was employed for  $\Delta Cq$  and RQ of blood markers. The sample size was 56 so the Shapiro-Wilk test (Shapiro and Wilk, 1965) was used to confirm a normal distribution of the  $\Delta Cq$  data ( $p > 0.05$ , Table 9). In addition, the Q-Q plots were produced. When there is a normal distribution, the scatter should be close to the line, which was the case in the present experiment (Figure 20. A, B, and C). In contrast, RQ of blood markers showed skewed distribution, with  $p < 0.5$  (Table 10. Figure 20 D, E, and F).

Statistical differences using the paired samples T-test between fresh samples (calibrator) and other time points over a period of week was employed with  $\Delta Cq$  for each marker. The result showed comparable levels of target gene between different time points, except for GAPDH at day seven and B2M at day three and six. RQ was calculated for ALAS2 marker using GAPDH and B2M for normalisation, ALAS2/GAPDH and ALAS2/B2M, respectively. In addition, both reference genes were also normalised with each other (B2M/GAPDH). The degradation patterns of selected markers were investigated, and two types of curves were explored. The results indicated that ALAS2/GAPDH had the strongest linearity with an  $R^2 = 0.65$ , whereas, B2M/GAPDH had highly polynomial curve with an  $R^2 = 0.81$  (Figure 19. B, and D, respectively). Therefore, ALAS2/GAPDH and B2M/GAPDH emerged as the most appropriate combinations for blood stain age prediction in the period of up to 7 days.

Table 9. Test of Normality of Delta Cq ( $\Delta Cq$ ) of blood markers in one-week period.

	<b>Shapiro-Wilk test</b>		
	<b>test value</b>	<b>df</b>	<b>p value</b>
Delta Cq ( $\Delta Cq$ ) of ALAS2	.969	56	.153
Delta Cq ( $\Delta Cq$ ) of B2M	.971	56	.192
Delta Cq ( $\Delta Cq$ ) of GAPDH	.973	56	.235

Table 10. Test of Normality of RQ for blood markers during one-week period.

	<b>Shapiro-Wilk test</b>		
	<b>test vale</b>	<b>df</b>	<b>p value</b>
RQ of ALAS2/B2M	.911	56	.001
RQ of ALAS2/GAPDH	.914	56	.001
RQ of B2/GAPDH	.853	56	.000

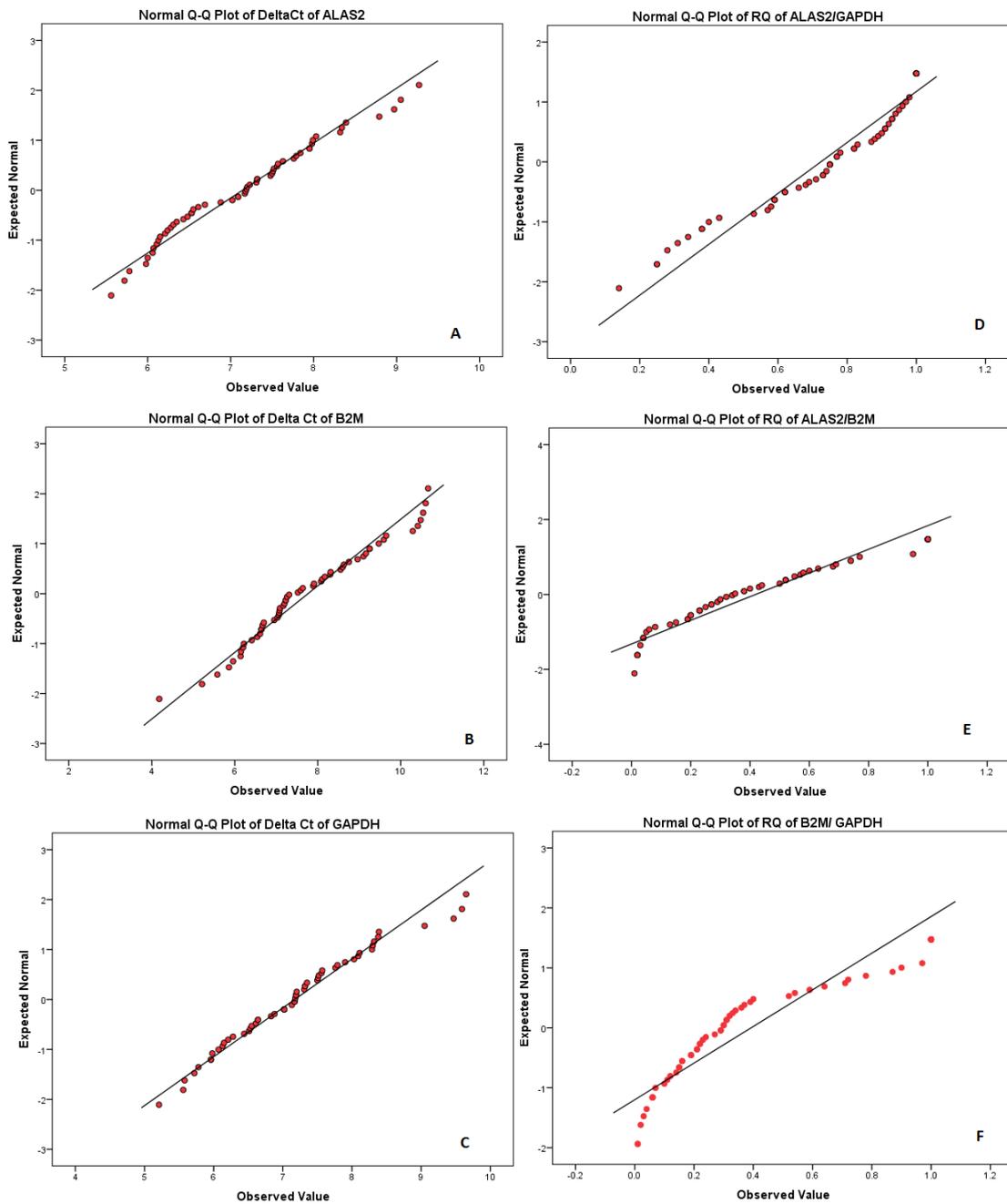


Figure 20. Normal Q-Q plot of (A, B, and C) using  $\Delta$ Cq of ALAS2, B2M, and GAPDH in blood respectively. (D, E, and F) for RQ of ALAS2/GAPDH, ALAS2/B2M, and B2M/GAPDH in blood respectively. The x-axis represents data points. The y-axis is the expected data point if the population distribution of the variable is normal with the population mean and standard deviation. The line is the sample mean  $\pm 2$  x standard deviation.

To generate any model, the correlation between dependent variable (DV) Y and independent variable (IV) X should be detected. Therefore, the relationship between DV (Age) and IV (RQ) was investigated. Pearson's correlation was conducted using RQ of ALAS2/GAPDH, ALAS2/B2M, and B2M/GAPDH against actual age of the sample (Appendix 4. Table 44). Moderate negative correlations  $r = -0.60$ , and  $r = -0.64$ , and low negative correlation  $r = -0.45$  were detected for ALAS2/B2M, B2M/GAPDH, and ALAS2/GAPDH, respectively. A simple regression analysis with CI and PI of 95% was performed to assess the ability of RQ of chosen markers to predict the age of unknown samples. Linear, quadratic and cubic curves were tested to find an optimal curve fitting the model. To achieve that, RQ of ALAS2/GAPDH, ALAS2/B2M, and B2M/GAPDH was plotted individually against the actual age of sample. The graphs show different patterns of degradation for ALAS2 marker, with  $R^2 = 0.39$  and  $R^2 = 0.22$  when normalised to B2M and GAPDH, respectively (Figure 21. A and B), while  $R^2 = 0.43$  for B2M/GAPDH (Figure 21 C). Three equations were generated for regression line and are listed in Table 11. Once regression equation was obtained, the prediction age was calculated for each model and plotted again with the actual age (Figure 21. D, E, and F).

Table 11. Simple regression analysis model generated in blood during 7 days.

Type of sample	Frame time	Number of samples	Pearson's correlation	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Week	56	-0.45**	ALAS2	GAPDH	Quadratic	$y=4.51+3.61x-6.26x^2$	0.22	±1.7
Blood	Week	56	0.60**	ALAS2	B2M	Cubic	$y=5.02-5.08x+9.13x^2-8.84x^3$	0.39	±1.4
Blood	Week	56	-0.64**	B2M	GAPDH	Cubic	$y=5.04-4.84x+7.34x^2-7.3x^3$	0.43	±1.4

Notes: (\*) represented of  $p < 0.05$  and (\*\*)  $p < 0.01$ .

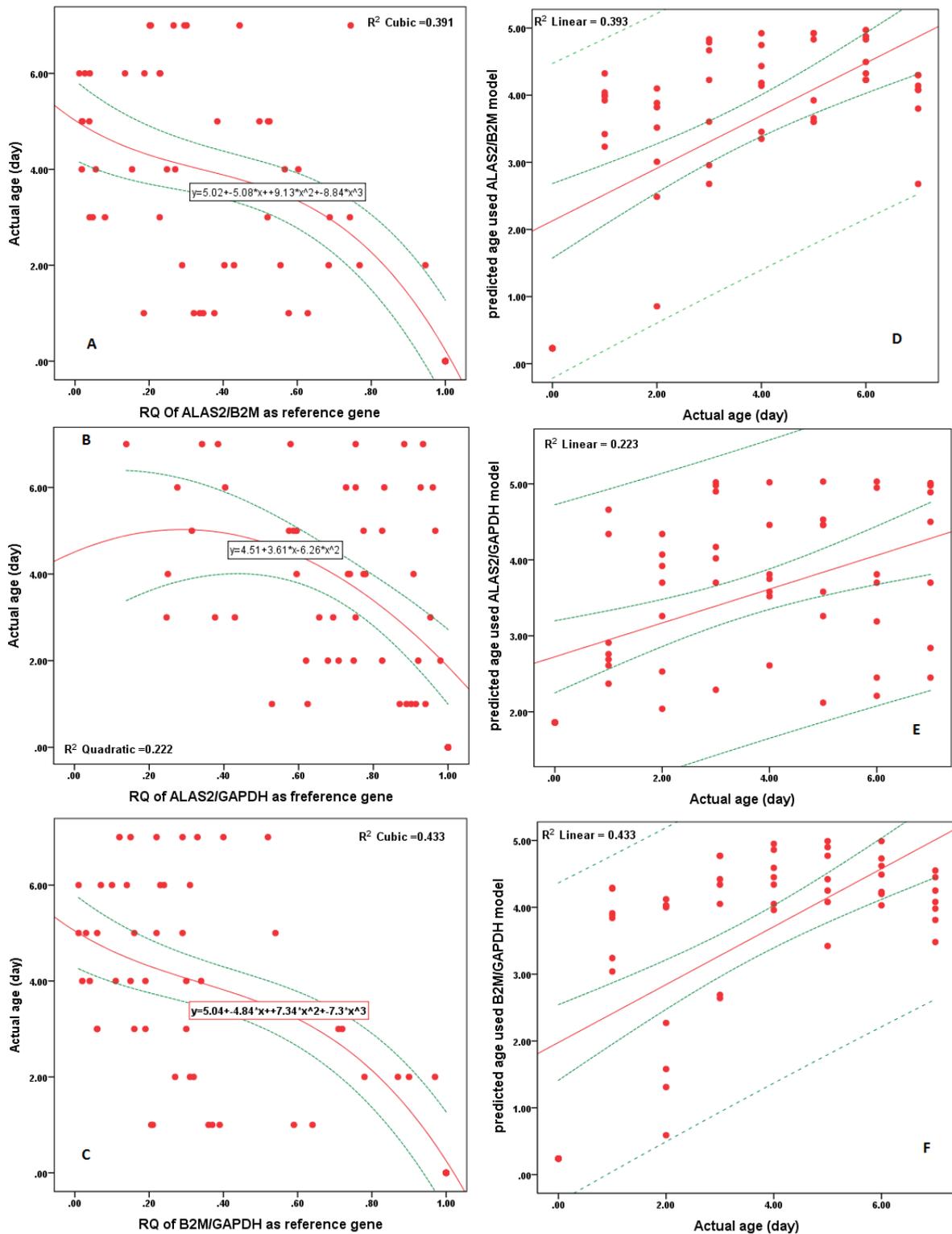


Figure 21. Simple regression analysis of chosen markers in blood samples over a one-week period. RQ of (A) ALAS2/B2, (B) ALAS2/GAPDH, and (C) B2M/GAPDH. (D-F) Age prediction obtained with the same models. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits (N=56).

Wide CI depicted on the graphs above are a consequence of high SD (Motulsky and Christopoulos, 2003). Further, difference in SD values could also explain high  $R^2$  value of B2M/GAPDH when compared to either ALAS2/GAPDH or ALAS2/B2M. Overall, the models presented above were able to predict to a certain extent the actual age of the sample especially in case of samples up to 3 days old.

Once the equation was obtained, the mean age prediction was also calculated (Appendix 1. Table 32). Figure 22 shows the mean age prediction plotted against the actual age. ALAS2/B2M and B2M/ GAPDH were able to predict a fresh sample while ALAS2/GAPDH was not. In addition, models ALAS2/GAPDH, ALAS2/B2M, and B2M/ GAPDH were also able to predict the age at day 2, 3, 4, and 5 reasonably close to observed values. Two parameters were established to evaluate which model was the best, an  $R^2$  value of model linearity and the standard deviation of predicted age. Overall, the models showed good prediction ability with the highest  $R^2 = 0.54$  generated with ALAS2/GAPDH model. The stability of reference gene during any relative quantification study plays a crucial role in obtaining a responsible data analysis. In this experiment, the same marker (ALAS2) obtained two different  $R^2$  values when using two different reference genes.

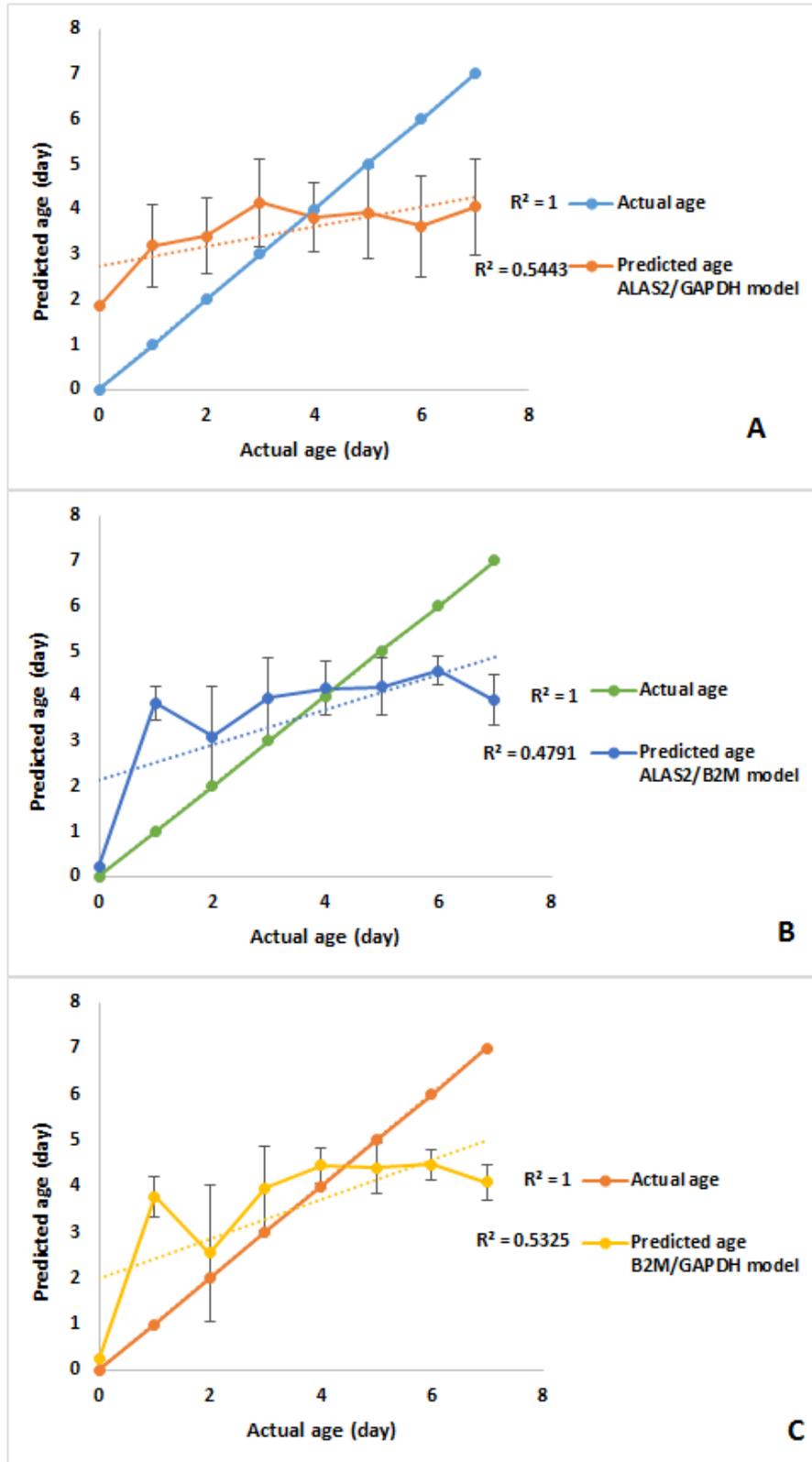


Figure 22. The age prediction plotted against actual age. (A) ALAS2/GAPDH model (B) ALAS2/B2M model(C) B2M/GAPDH model. Data presented as Mean  $\pm$  SD.

#### **4.4.2. Degradation of RNA during one-month period**

The quantification cycle (Cq) is a cycle number at which the exponential amplification is detected above the background. The differences in the Cq value correspond to differences in the amount of template, therefore, the amount of template is the likely cause of the difference in Cq values for short decomposition times. On the other hand, the stain with longer deposition time shows clearer degradation pattern than the recent stain (Simard *et al.*, 2011, Williams *et al.*, 2013). For this reason, a decomposition of time frame was increased to 28 days.

All current qPCR detection systems use fluorescent technologies and there are varieties of different template detection methods. In the present study, relative quantification was performed using TaqMan probe chemistry in blood, saliva and semen stains that were allowed to decompose for up to 28 days. This was similar to the study that used the comparative delta-delta Cq ( $2^{-\Delta\Delta Cq}$ ) method to estimate PMI over 13 days, using the transcript abundances of mRNA, miRNA, 18S rRNA, and U6 snRNA in rat's spleen (Lv *et al.*, 2014).

#### **4.4.3. Age prediction using TaqMan chemistry**

Reverse transcription PCR (RT-qPCR) is a common method to quantify mRNA degradation. The link between stain ageing and the degradation pattern of RNA was detected by Anderson and co-workers in 2005 when the relative amount of 18S rRNA compared to  $\beta$ -actin mRNA was explored (Anderson *et al.*, 2005). Therefore, in the present study, the sets of mRNA TaqMan primers and probes were investigated in blood, saliva, and semen samples.

#### 4.4.3.1. Blood samples

In blood samples, ALAS2, HBB, and GYPA were investigated in an attempt to obtain models which may be useful to predict the unknown age of samples. All transcripts were normalised with B2M as a reference gene. In this experiment, most of the markers examined were successfully detected and showed tendency towards linear correlation with time. This was not the case with GYPA which was less abundant than other transcripts and showed a low correlation with age (Figure 23). Consistent with previous reports, HBB emerged as the most abundant marker in the present study (Haas *et al.*, 2011a). The same statistical analysis which was performed in one-week experiment was also employed, therefore, T-test analysis was performed between fresh samples and later time points for each marker, and the results showed comparable  $\Delta Cq$  values for the samples up to one week old, except in case of GYPA. For samples older than one week, decrease in  $\Delta Cq$  when compared to the fresh samples were observed for most of the markers examined, except HBB and GYPA.

RQ of GYPA with B2M showed weak linear correlation and large standard deviation suggesting that GYPA was not sensitive enough and had the lowest  $\Delta Cq$  value. This finding was consistent with previous reports of GYPA demonstrating only a medium sensitivity for detection limit of RNA (Xu *et al.*, 2014). The RQ for ALAS2 and HBB were also obtained and the trend was checked with two types of curves. The results indicated that HBB marker had a strong linear and polynomial relationship with the actual age with corresponding  $R^2$  values of 0.92 and 0.99 respectively. Further, The ALAS2 also generated strong linear and polynomial correlations with  $R^2$  value of 0.87 and 0.98, respectively (Figure 23. B and C).

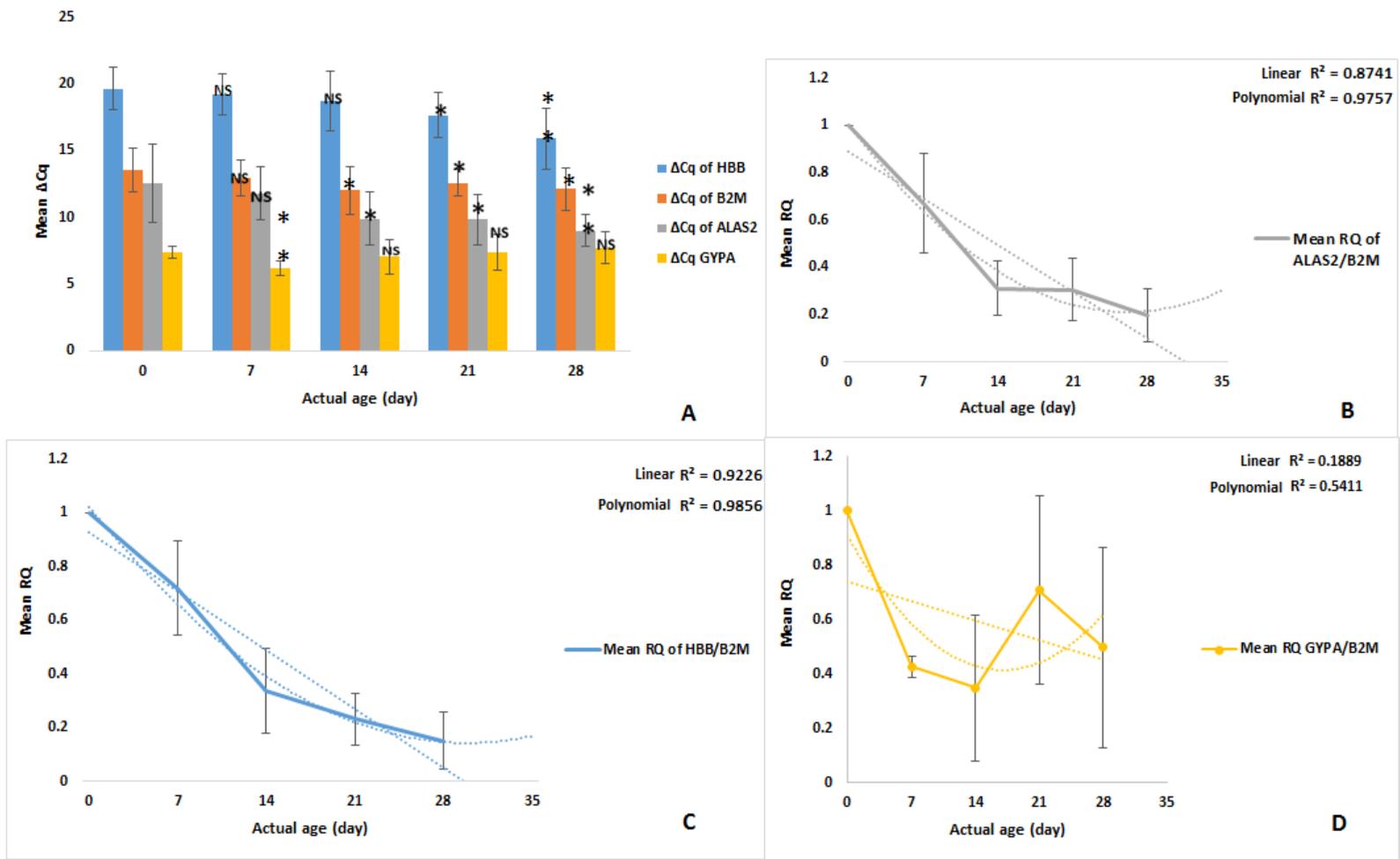


Figure 23. Amplification of ALAS2, HBB, and B2M in blood using TaqMan chemistry. (A) The  $\Delta Cq$  of markers ALAS2, HBB, GYPA and B2M over 28-day period. RQ of (B) ALAS2/B2M (C) HBB/B2M and (D) GYPA/B2M in blood using TaqMan chemistry. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=40 for ALAS2 and HBB; GYPA N=20).

As mentioned earlier, the correlation between dependent (real age) and independent variables (RQ) should be determined. Therefore, Pearson's correlation was also performed, and a strong negative correlation that was highly statistically significant ( $p < 0.01$ ) for RQ of both ALAS2/B2M and HBB/B2M ( $R^2 = -0.86$  and  $-0.90$ , respectively). On the other hand, GYPA was weakly correlated with an  $R^2 = -0.32$  (Table 12). A strong negative correlation means that RQ of selected markers decreases as actual time increases, and this could allow the use of SR analysis to obtain models to predict the age of known bloodstains during one-month period. Therefore, a single regression analysis with 95% confidence and prediction interval was also performed. The predictive equations were obtained to determine the age of an unknown sample resulted from a regression line of RQ parameter. The RQ of ALAS2/B2M and HBB/B2M was plotted against actual age and the result showed a good negative relationship ( $R^2 = 0.75$  and  $0.82$ , respectively) (Figure 24. A and B). Finally, two formulas based on the optimal curve were generated (Table 12), then the predicted age (Appendix 1. Table 34) was plotted against the actual age. Both models generated good predictions, which is confirmed by a number of points located within the confidence and prediction intervals at 95% (Figure 24. C and D).

Table 12. Summary of models obtained in blood using a simple regression analysis for 28 days period.

Type of sample	Frame time	Number of samples	Pearson's correlation	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Month	40	-0.86**	ALAS2	B2M	Linear	$y=27.19-26.5x$	0.75	±3.8
Blood	Month	40	-0.90**	HBB	B2M	Linear	$y=26.76-26.15x$	0.82	±3.2
Blood	Month	20	-0.32	GYPA	B2M	Quadratic	$y=17.23+2.48x-10.23x^2$	0.11	±8.1

Not: \* represented of  $p < 0.05$  and \*\* represented that  $p < 0.01$

Table 12 shows the relationship between Pearson's correlations,  $R^2$  value, and mean absolute deviation (MAD). When Pearson's correlation increases,  $R^2$  becomes a higher, while MAD is decreases, and the vice versa. The strongest correlation was detected in blood with HBB/B2M model ( $r = -0.90$ ) and this gave highest  $R^2$  value (0.82). Therefore, could be considered that

HBB/B2M is the best model of those examined to predict the age of bloodstains that are up to 28 days old, with MAD  $\pm 3.2$ .

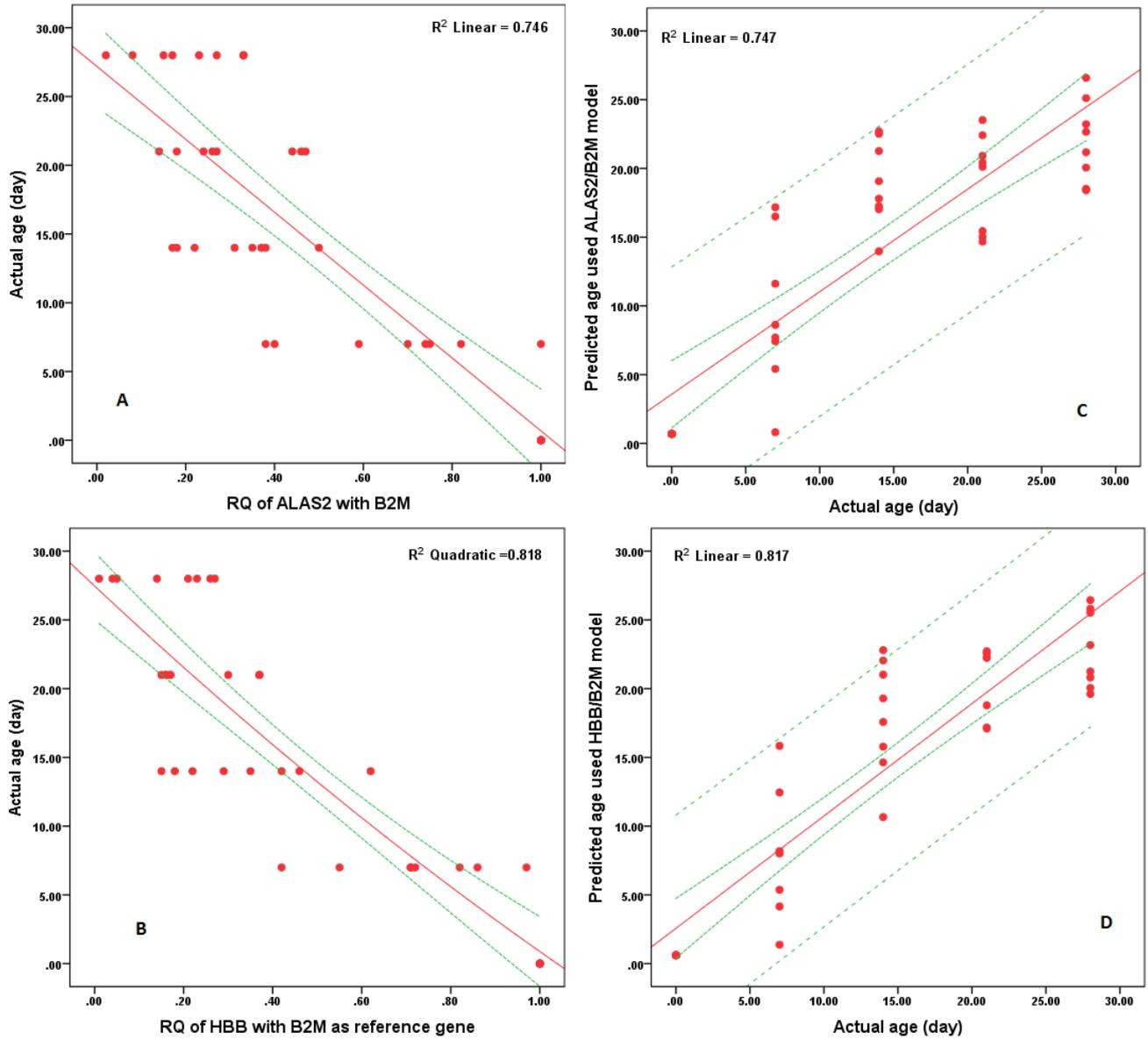


Figure 24. Regression analysis to obtain models in blood samples using TaqMan chemistry. (A and B) Relative quantification of ALAS2 and HBB with B2M as a reference gene, respectively. (C and D) Prediction age generated with ALAS2/B2M and HBB/B2M models, respectively. The solid red line in the graph represents the modelled space where predicted age and real age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

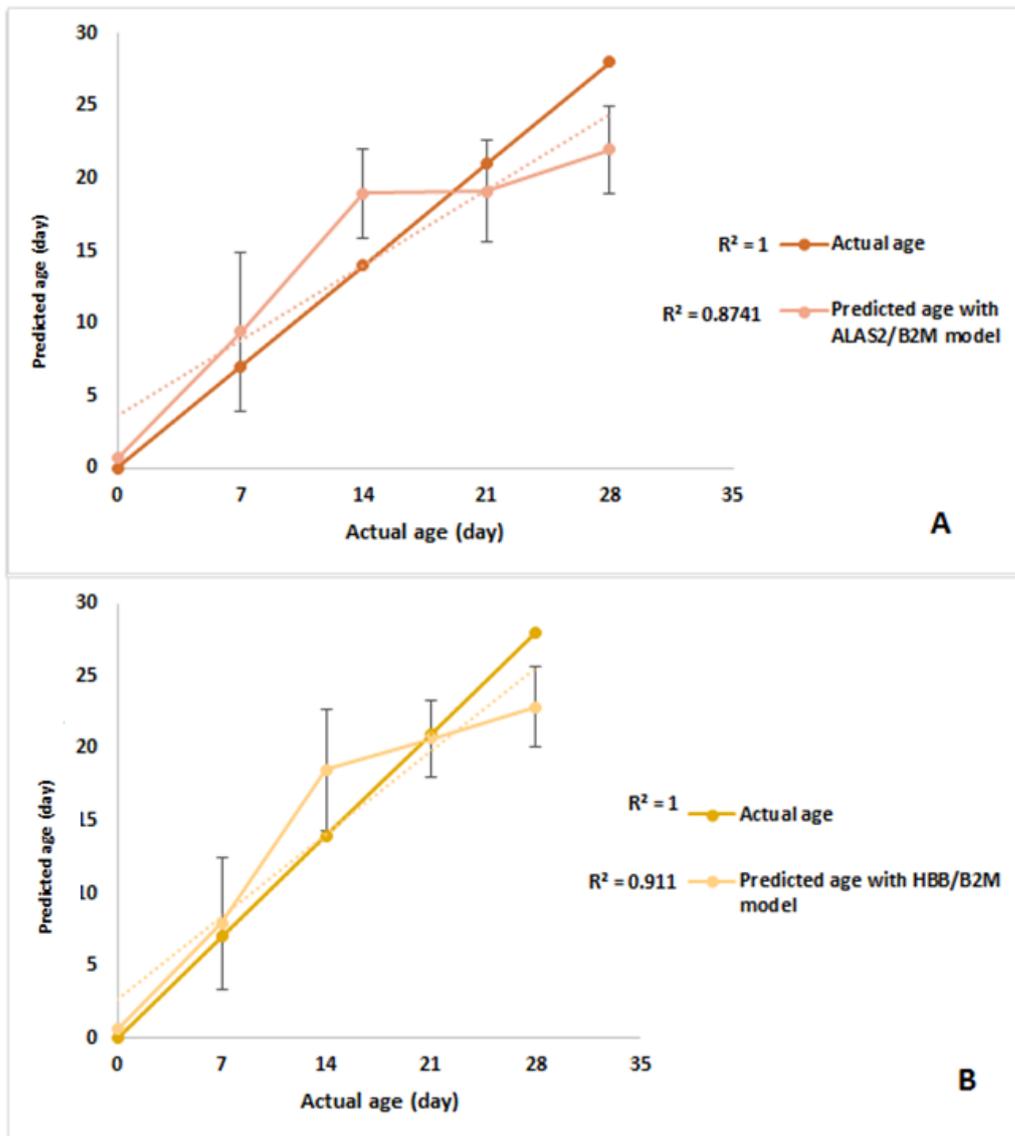


Figure 25. The mean predicted age fitted with actual age in blood models. (A) ALAS2/B2M model (B) HBB/B2M model. Error bars represented one standard deviation.

Figure 25 shows the mean age prediction plotted against the actual age to compare the fit of all the models used. Overall, two models gave good fit with high  $R^2$  values, with some data points, such as point zero (fresh sample), 7 and 21 days in HBB/B2M model falling exactly on the regression line. The comparison between two different times (7 and 28 days) showed an improved  $R^2$  value, which was especially obvious with ALAS2 marker used in both

experiments. It is possible that the deposition time strongly influenced clearer detection of the degradation patterns, as previously reported (Williams *et al.*, 2013).

#### **4.4.3.2. Saliva samples**

Saliva age prediction is critical in the situations when the saliva is only the biological evidence recovered. With saliva samples perhaps more often than with other types of samples, determination of the TSD is important in order for sample to be used as a valid evidence. For example, some samples such as cigarette butts, cans of drinking, and envelopes are likely to be found before the crime was committed, therefore prediction of TSD would allow the forensic investigation to focus only on the relevant evidence that fits into a frame time when crime was committed. In order to achieve this, the same procedures and statistical methods used for blood samples were also performed on saliva in an attempt to predict the age of these stains. Two TaqMan primers and probe were investigated, namely HTN3 and MUC7, and both were normalised to GAPDH as a reference gene. The HTN3 and GAPDH markers successfully detected, whereas MUC7 marker was not sensitive enough and therefore excluded. The HTN3 primer was also previously reported as the most abundant and sensitive primer in saliva samples (Setzer *et al.*, 2008).

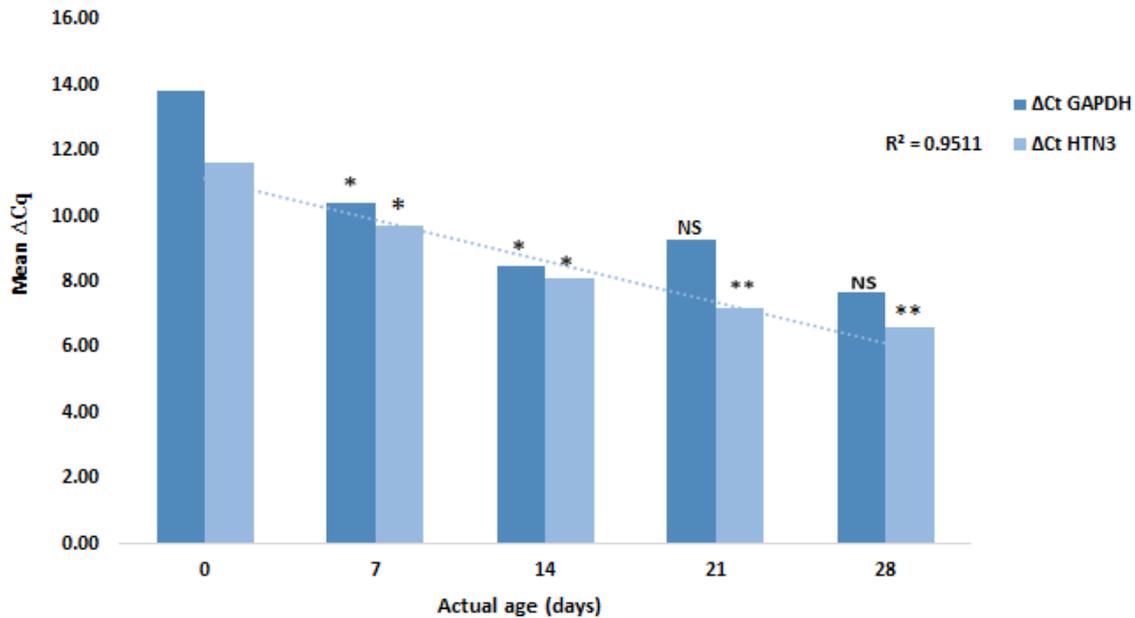


Figure 26.  $\Delta Cq$  of HTN3 in saliva samples using TaqMan probe. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation.  $N = 30$  (six samples with five time points).

Figure 26 shows  $\Delta Cq$  of HTN3 and GAPDH over 28 days period. The result demonstrated that GAPDH was more abundant than HTN3 at all-time points. The T-test analysis showed significant differences of HTN3 marker at all-time points when compared to the fresh sample, whereas, GAPDH was significantly different at 7 and 14 days only. The RQ of HTN3 was plotted and the result showed a weak linear relationship with the actual age, whereas a strong linear relationship with  $R^2 = 0.95$  was observed for the mean  $\Delta Cq$  of HTN3 ( $Cq_{max} - Cq_{target}$ ). This is likely due to decrease in  $\Delta Cq$  of GAPDH over the first three-time points.  $Cq$  of the target gene should be strongly correlated with actual age, but at the same time  $Cq$  of reference gene should be stable in order to obtain the accurate normalisation. In this experiment, the mean  $\Delta Cq$  of HTN3 has a strong  $R^2$  value, which was decreased in  $2^{-\Delta \Delta Cq}$  calculation.

Simple regression was employed using RQ of HTN3/GAPDH and  $\Delta Cq$  of HTN3 (Figure 27. A and B), and the results showed higher  $R^2$  value of  $\Delta Cq$  HTN3 when compared to RQ HTN3/GAPDH. The two equations were obtained and listed in Table 13. The age prediction was calculated (Appendix 1. Table 34), and the data were plotted against the actual age. The results demonstrated that the prediction limits of RQ of HTN3/GAPDH are wider than for  $\Delta Cq$  of HTN3 (Figure 27. C and D). In addition, almost of the data generated from  $\Delta Cq$  of HTN3 model fitted within the set confidence and prediction intervals, especially at day 21.

Table 13. Saliva models obtained using simple regression analysis up to 28 days.

Type of sample	Frame time	Number of samples	Pearson's correlation	Target gene	Reference gene	Optimal curve	Estimated curve equation	$R^2$	MAD (day)
Saliva	Month	30	-0.49**	HTN3	GAPDH	Quadratic	$y=11.5+43.28x-52.29x^2$	0.45	$\pm 5.8$
Saliva	Month	30	0.87**	HTN3	No reference*	Quadratic	$y=85.64-12.2x+0.34x^2$	0.79	$\pm 3.3$

Note: \* represented of  $p < 0.05$  and \*\* represented that  $p < 0.01$ , no reference means HTN3 marker without normalisation.

Figure 28 shows the mean age prediction obtained with  $\Delta Cq$  model was very close to an actual age ( $R^2 = 0.98$ ). These data suggest the possibility of using  $\Delta Cq$  of HTN3 instead of RQ to predict the age of saliva. This finding was also supported by Simard *et al.* (2012) who suggested of using individual markers for age prediction (Simard *et al.*, 2012).

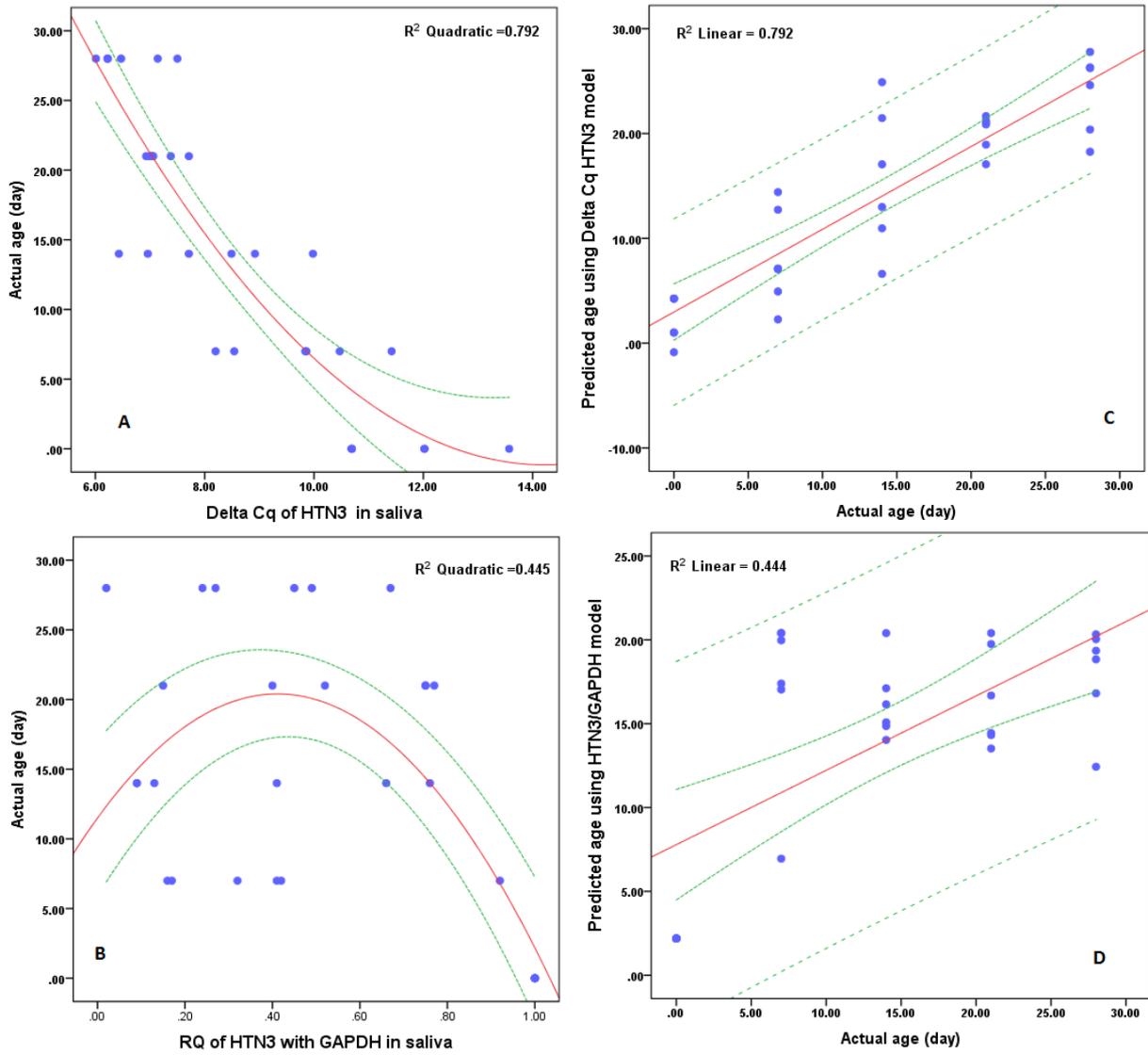


Figure 27. Regression analysis in saliva (A)  $\Delta$ Cq of HTN3 and (B) RQ of HTN3/GAPDH. (C and D) Age prediction using previous models, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

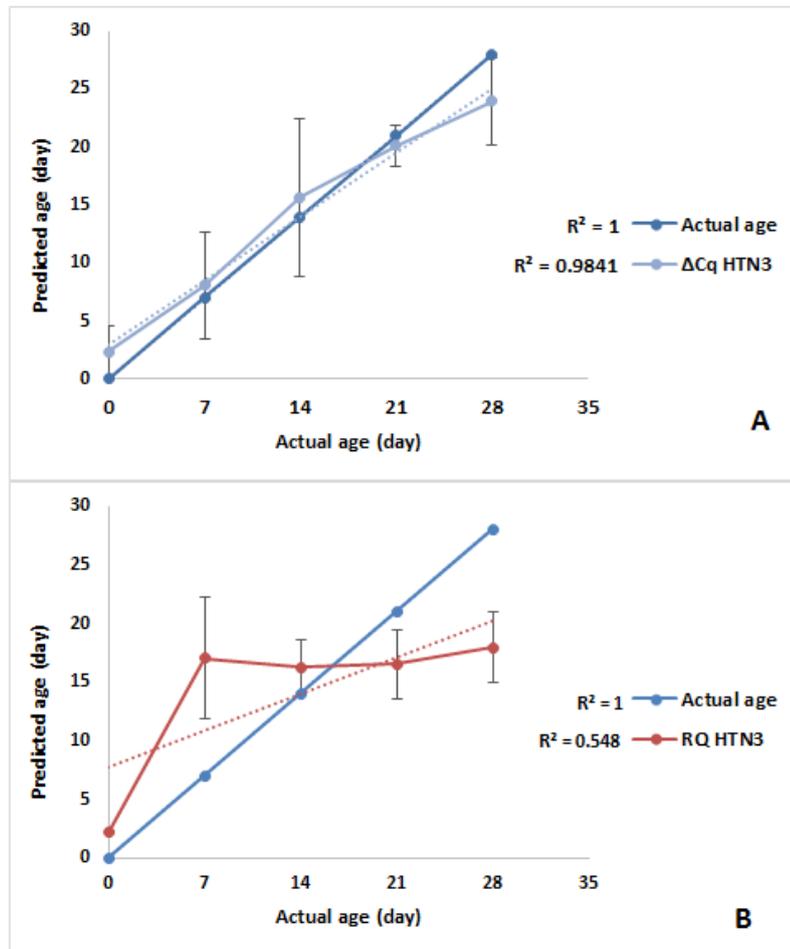


Figure 28. Age prediction generated by (A)  $\Delta Cq$  HTN3 model (B) RQ of HTN3/GAPDH model. Data presented as Mean  $\pm$  SD.

#### 4.4.3.3. Semen samples

Semen age prediction using TaqMan chemistry was also employed, and for this experiment, two semen specific primers and probes, namely PRM1 and SEMG1 were investigated and normalised using B2M as a reference gene. In this experiment, the influence of DTT reagent was also investigated over a 28 days period. In presence of DTT, all markers were detected, and the result showed that PRM1 was the most abundant (Figure 29 A), whereas in the absence of DTT the B2M marker was the most abundant (Figure 30 A).

The same procedure and the same analysis conducted on blood and saliva were also performed in semen samples. In the presence of DTT. The RQ was calculated for PRM1/B2M, SEMG1/B2M and PRM1/SEMG1. Moreover, for all markers selected, a linear downward trend was observed with  $R^2$  value of 0.92, 0.86, and 0.91 respectively (Figure 29. B, C and D). In contrast, in the absence of DTT the correlation of the same markers with the actual age was decreased (Figure 30. B, C, and D).

Pearson's correlation (Table 14) also showed a strong negative relationship between semen markers (+DTT) and actual age, and this finding confirmed the possibility of using regression analysis to obtain model for semen age prediction. Therefore, the RQ was plotted against actual age, and the results showed that all of the markers used were able to generate an optimal cubic curve with  $R^2$  value of 0.64 for both PRM1/B2M and SEMG1/B2M models, and 0.71 for PRM1/SEMG1 (Figure 31. A, B, and C). All three equations obtained are presented in Table 14.

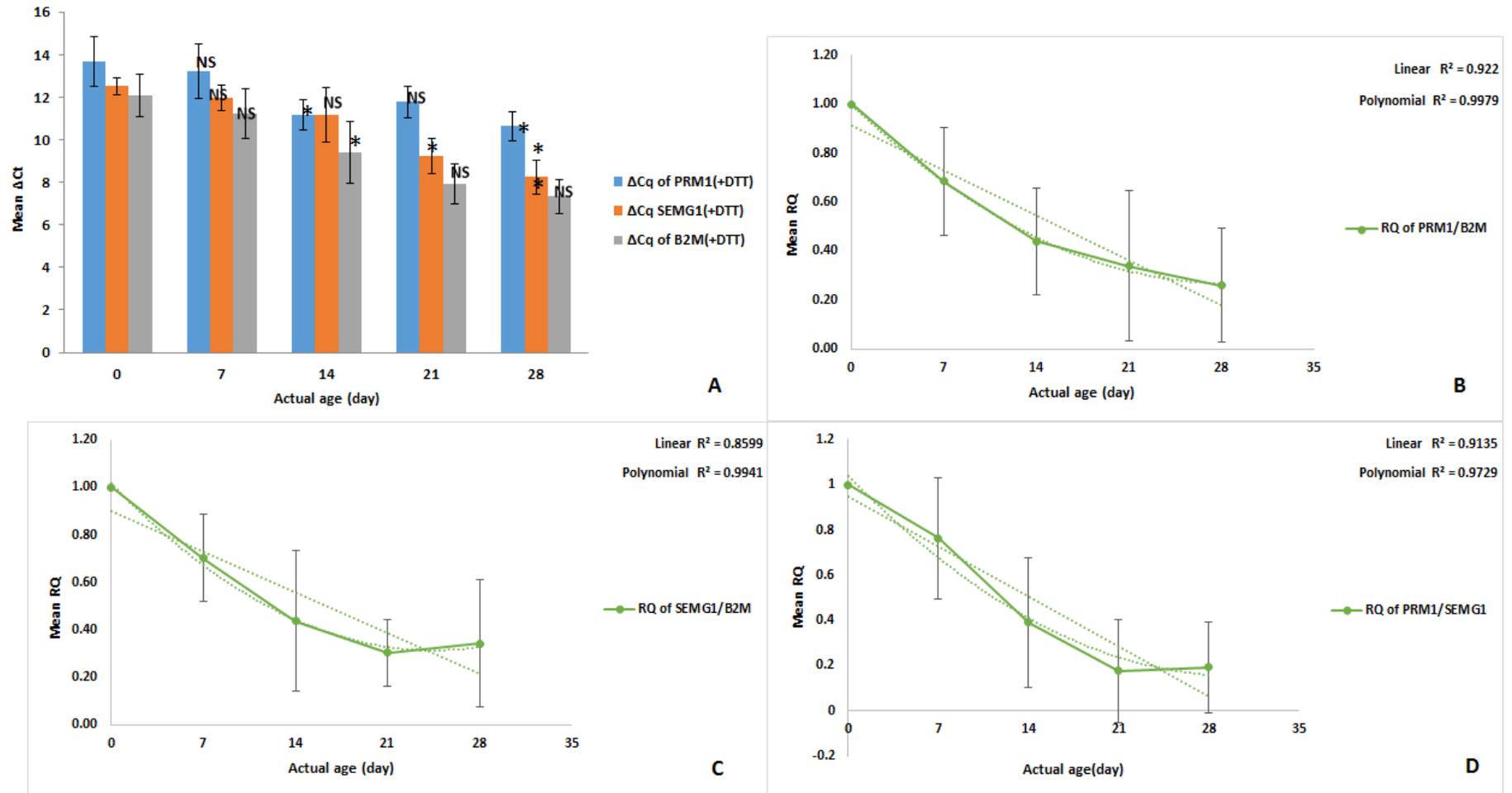
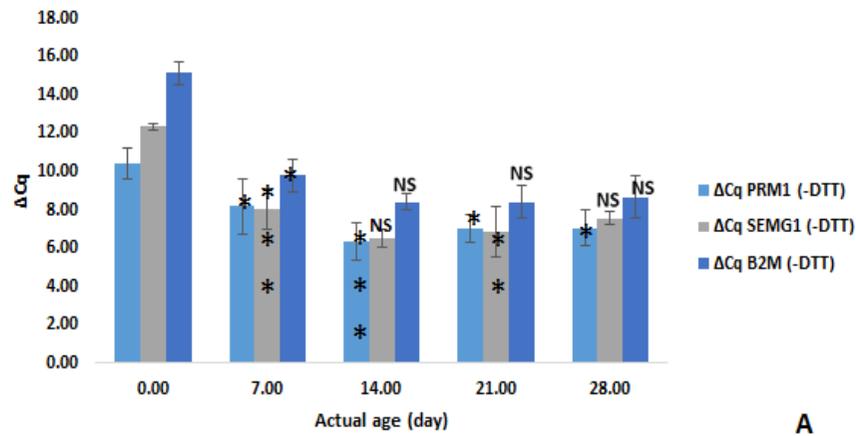
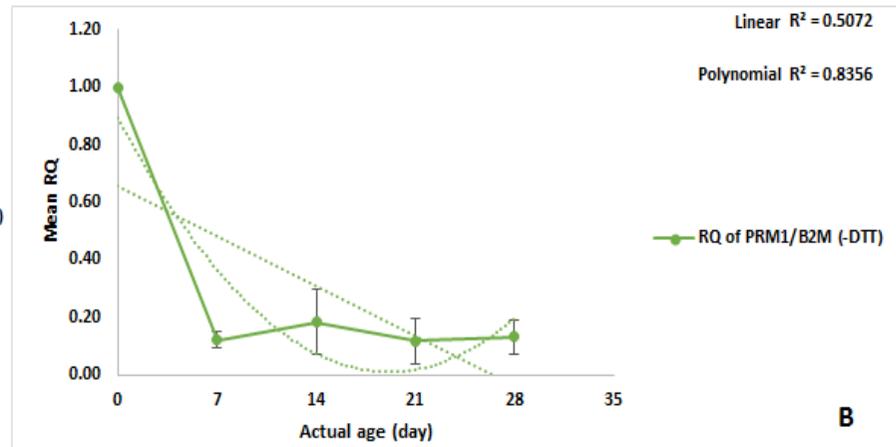


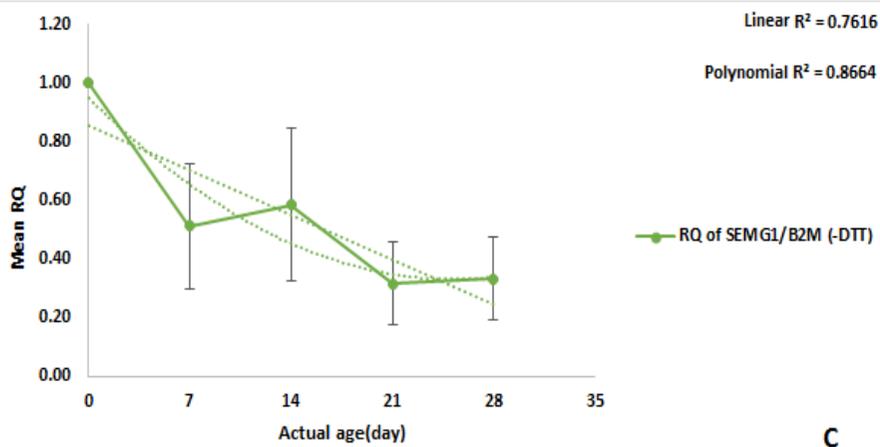
Figure 29. Semen markers with DTT (+DTT). (A)  $\Delta Ct$  of markers PRM1, SEMG1, and B2M in semen samples using TaqMan chemistry. RQ of (B) PRM1/B2M (C) SEMG1/B2M and (D) PRM1/SEMG1. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N= 30).



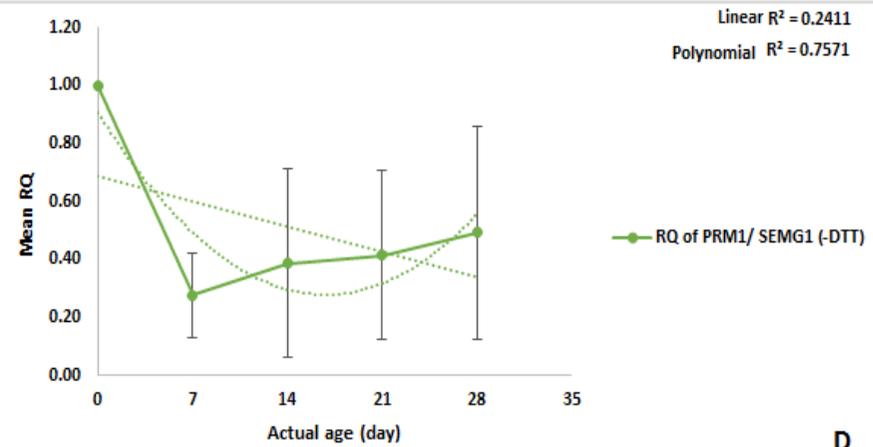
A



B



C



D

Figure 30. Semen markers without DTT (-DTT) over 28 days period. (A)  $\Delta Cq$  of PRM1, SEMG1, and B2M in semen samples. (B and C) RQ of PRM1 and SEMG1 with B2M as reference gene, respectively. (D) RQ of PRM1 with SEMG1 as reference gene. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=30).

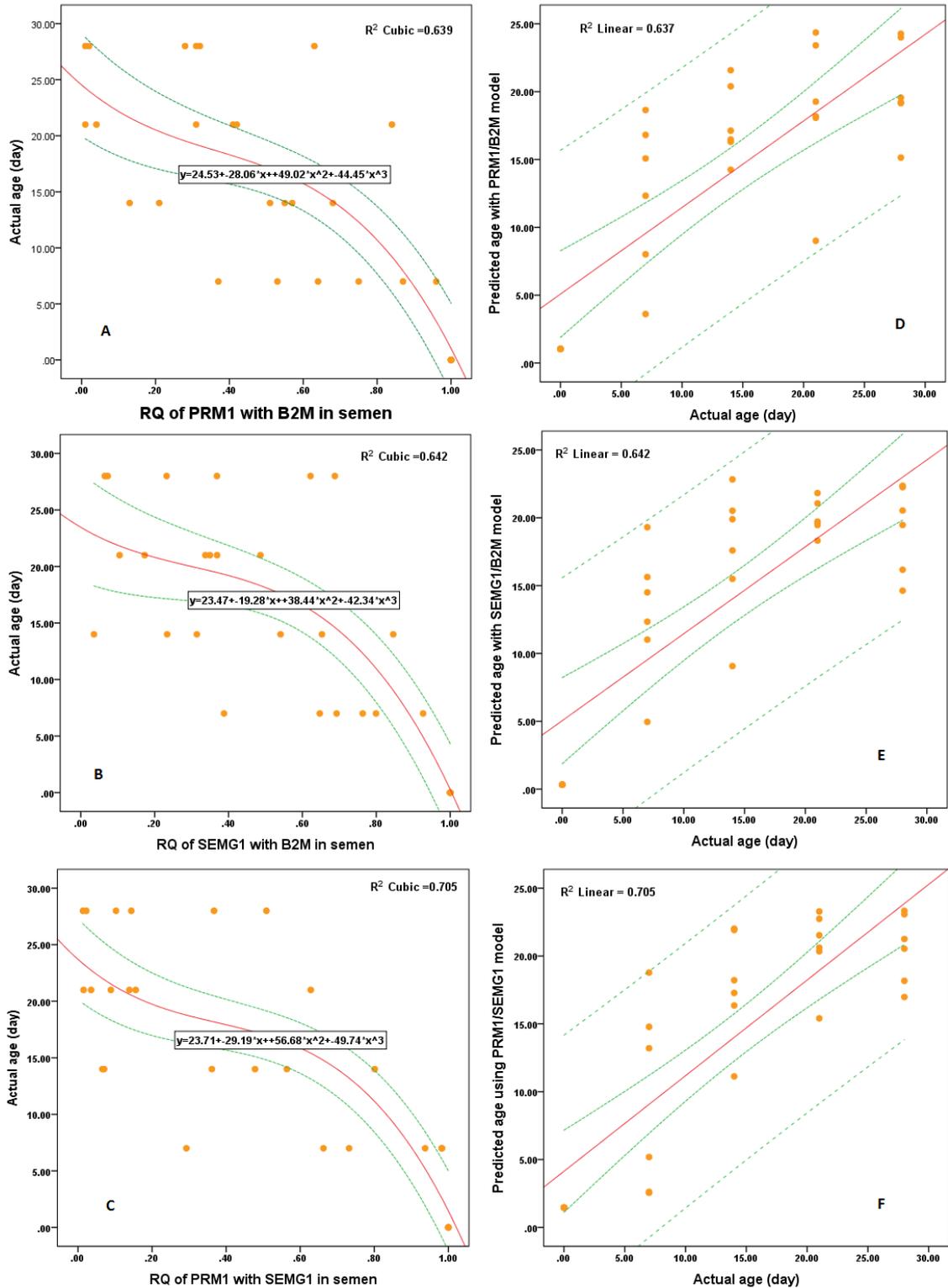


Figure 31. Simple regression analysis using semen TaqMan markers. (A - C) RQ of PRM1/B2M, SEMG1/B2M, and PRM1/SEMG1, respectively. (D - F) Actual age was plotted with predicted age, which was generated using previous models. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits (N=30).

Table 14. Summary of semen models obtained using simple regression analysis over a 28 days period.

Type of sample	Frame time	Number of samples	Pearson's correlation	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Semen	Month	30	-0.77**	PRM1	B2M	Cubic	$y=24.53-28.06x+49.02x^2-44.45x^3$	0.64	±4.7
Semen	Month	30	0.75**	SEMG1	B2M	Cubic	$y=23.47-19.28x+38.44x^2-42.34x^3$	0.64	±4.4
Semen	Month	30	-0.81**	PRM1	SEMG1	Cubic	$y=23.71-29.19x+56.68x^2-49.74x^3$	0.71	±4.3

Note: \* represented of  $p < 0.05$  and \*\* represented that  $p < 0.01$

Figure 30 shows decreased R<sup>2</sup> value when compared to the same value in Figure 29, which is likely due to the large standard deviation of RQ detected within each week. Using the formula generated (Figure 31. A, B, and C), the age prediction of semen was obtained (Appendix 1. Table 34), and a number of data points were located within the confidence intervals region, especially in case of PRM1/B2M and SEMG1/B2M models at day 7 until day 28 (Figure 31. D, E, and F).

To support that, the predicted age using semen models (+DTT) was calculated and plotted against the actual age. As can be seen in Figure 32, all models demonstrated good fit, with some of the data points overlapping actual time points, especially at 0, 14 and 21 days. The best fit was obtained using PRM1/SEMG1 model (R<sup>2</sup> = 0.87).

The comparison between the same marker with and without DTT was performed. The same procedure and the same statistical analysis were conducted on the semen markers in the absence of DTT. As expected, the results showed that RQ of the markers were weakly decreased and generated low R<sup>2</sup> value. The mean age prediction was calculated (Appendix 1. Table 33) and plotted against the actual age, and the result showed that DTT had a clear influence on these models, because no strong R<sup>2</sup> was obtained except with SEMG1/B2M, and large MAD was detected (Figure 33). This finding could be related that

SEMG1 was not more influence, because as mentioned early this gene is located in the surface of semen cell.

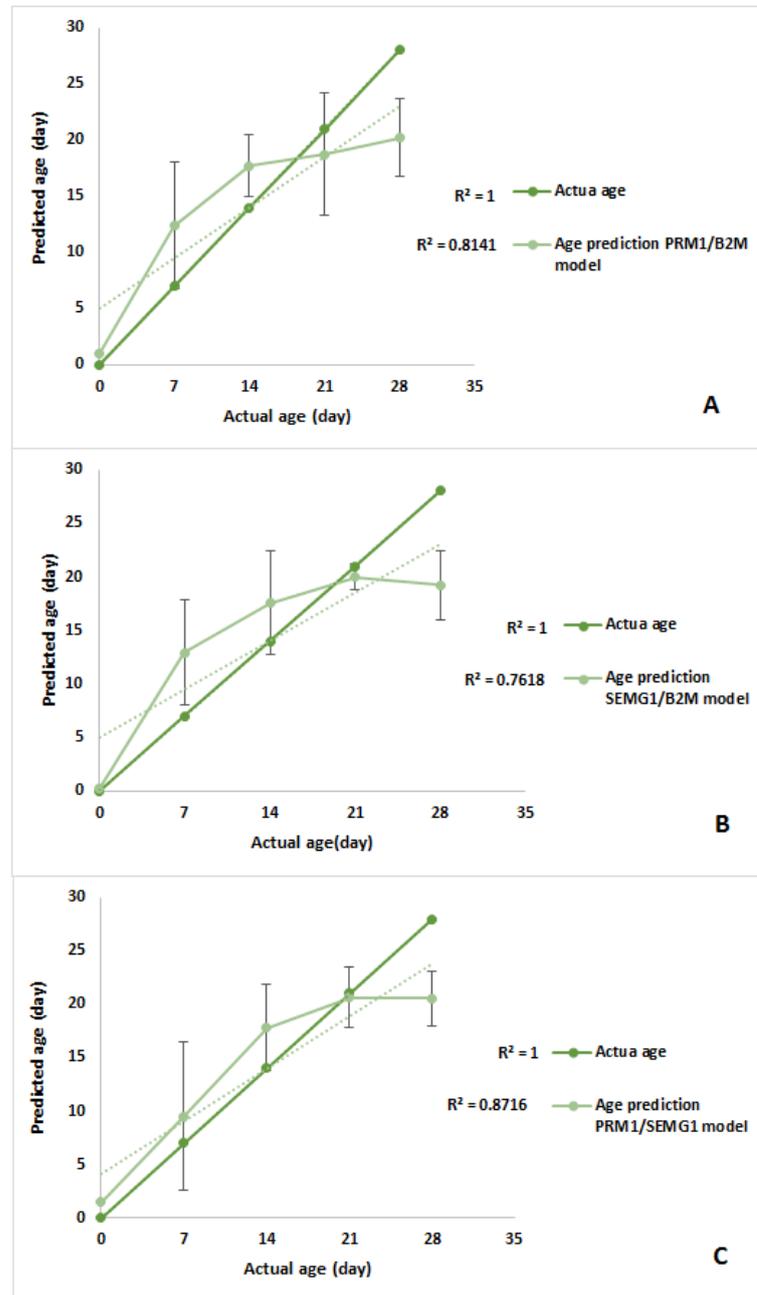


Figure 32. Age prediction that was obtained in semen (+DTT) with TaqMan chemistry. (A) Scatterplot of the mean age prediction using (A) PRM1/B2M, (B) SEMG1/B2M and (C) PRM1/SEMG1 model. Data presented as Mean  $\pm$  SD.

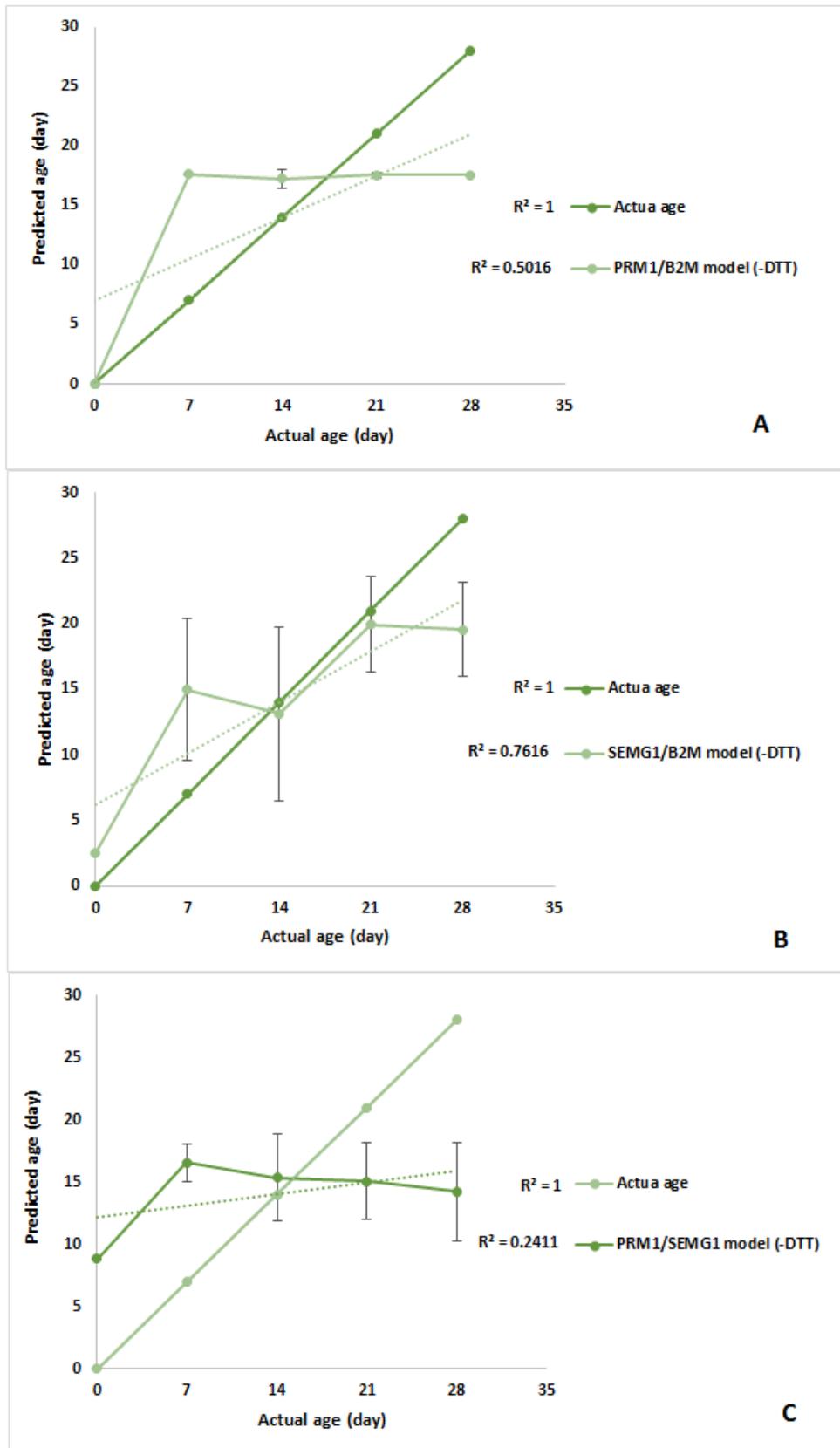


Figure 33. The age prediction in semen without DTT extraction (-DTT). (A, B, and C) Age prediction was calculated using models PRM1/B2M, SEMG1/B2M and PRM1/SEMG1, respectively. Data presented as Mean  $\pm$  SD (N=30).

#### 4.4.3.4. Multiple regression analysis

As an attempt to improve models obtained above, multiple regression analysis (MRA) was also employed. MRA is used to predict value of the variable using more than one independent variable (IV). In the one-week time frame, ALAS2/GAPDH, ALAS2/B2M and B2M/GAPDH were all tested. The data generated from this analysis (Appendix 4. Table 49) showed that two variables, ALAS2/GAPDH and ALAS2/B2M show significant changes over time, suggesting that both contribute in a new model for prediction of the bloodstains age for up to one week. The multiple regression analysis was also employed for blood and semen samples up to one month old, to obtain a combined model of prediction. In semen samples, two models (PRM1/B2M and SEMG1/B2M) demonstrated differences with the actual age, and therefore contributed to a new model for semen age prediction (Appendix 4. Table 50). Three new equations emerged from multiple regression analysis are presented in Table 15.

Table 15. Summary of models generated by using multiple regression analysis in blood and semen sample.

Type of sample	Frame time	Number of samples	Models using	X1	X2	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Week	56	ALAS2/GAPDH*	ALAS2/GAPDH	ALAS2/B2M	y=7.155-2.906(X1) - 3.707(X2)	0.44	±1.4
			ALAS2/B2M**					
Blood	Month	40	HBB/B2M**	ALAS2/B2M	HBB/B2M	y=27.859 -10.465(X1) -17.74 (X2)	0.84	±3.4
			PRM1/B2M**					
Semen	Month	30	SEMG1/B2M*	PRM1/B2M	SEMG1/B2M	y=29.045 -14.207(X1) -13.092(X2)	0.69	±4.4

Note : ( \*) Represented of p< 0.05 and (\*\*) mean p< 0.01

When compared to the single regression analysis, multiple regression had an improved R<sup>2</sup> value, followed by decreased MAD. For example, ALAS2/B2 and HBB/B2M run as a single model in blood gave R<sup>2</sup> value of 0.75 and 0.82 with MAD of ±3.8 and ±3.2, respectively,

whereas multiple regressions analysis improved  $R^2$  to 0.84 with a MAD of  $\pm 3.4$ . Similarly, multiple regression analysis improved  $R^2$  to 0.69 in semen samples, and this was observed in blood samples over one-week period. Two models including B2M/GAPDH and PRM1/SEMG1 for blood and semen, respectively, showed no significant difference with actual age, and they were excluded from the analysis. Once the equations were generated, the prediction age was also calculated in blood and semen samples (Appendix 1. Table 34).

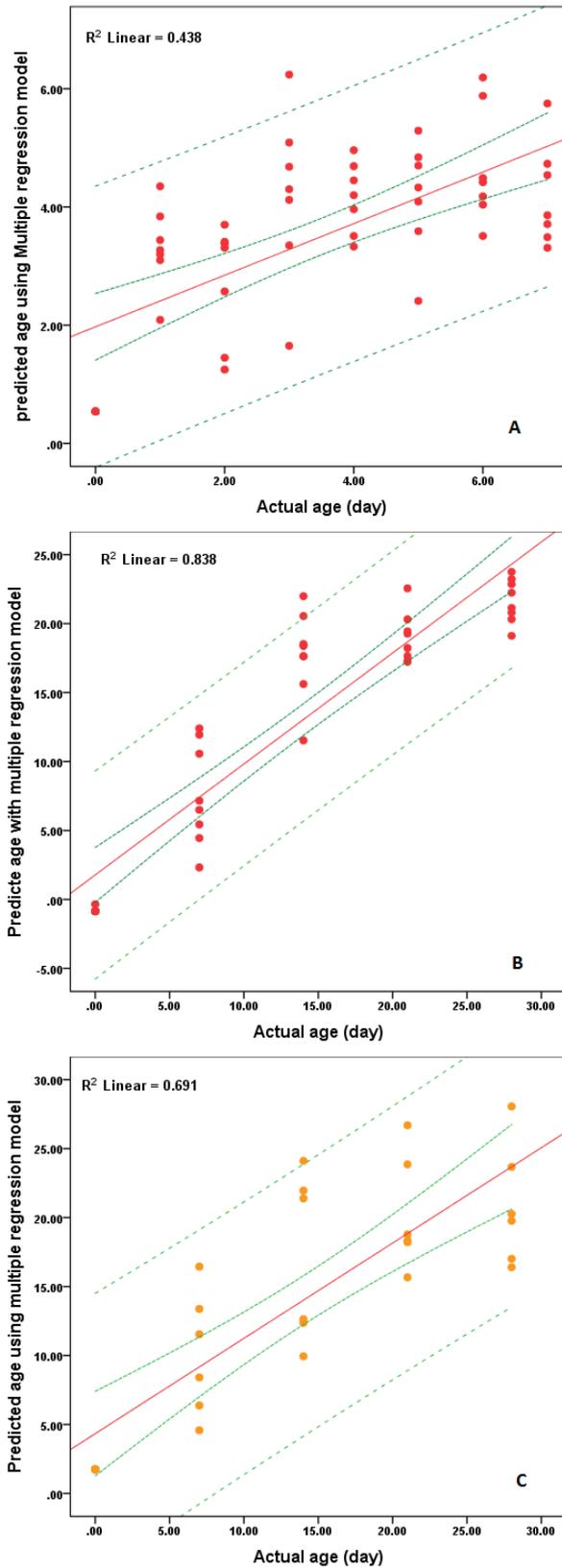


Figure 34. Age prediction using multiple regression models. Age prediction in (A) blood samples during one-week (B and C) blood and semen samples up to one month, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

For one-week age prediction, the graph shows an improved  $R^2 = 0.44$  with  $MAD \pm 1.4$  and fits the data more closely than when each model individually is used (Figure 34 A). Similarly, in blood model over one-month period, the data gave an improved  $R^2 = 0.84$  with  $MAD$  of  $\pm 3.4$  (Figure 34 B). Finally, predicted age in semen became more closely fitted meaning the improved standard deviation for age prediction (Figure 34 C).

The mean age prediction generated by multiple regression models was also calculated (Appendix 1. Tables 32 and 34), as showed in Figure 35. This model gave the highest  $R^2$  value in blood samples over one-month period, with at least three data points closely fitted (Figure 35 B), as well as in semen model where time points were also reasonably closely fitted, especially at 21 day (Figure 35 C). In summary, the multiple regression models overall gave higher  $R^2$  values when compared to a simple regression analysis.

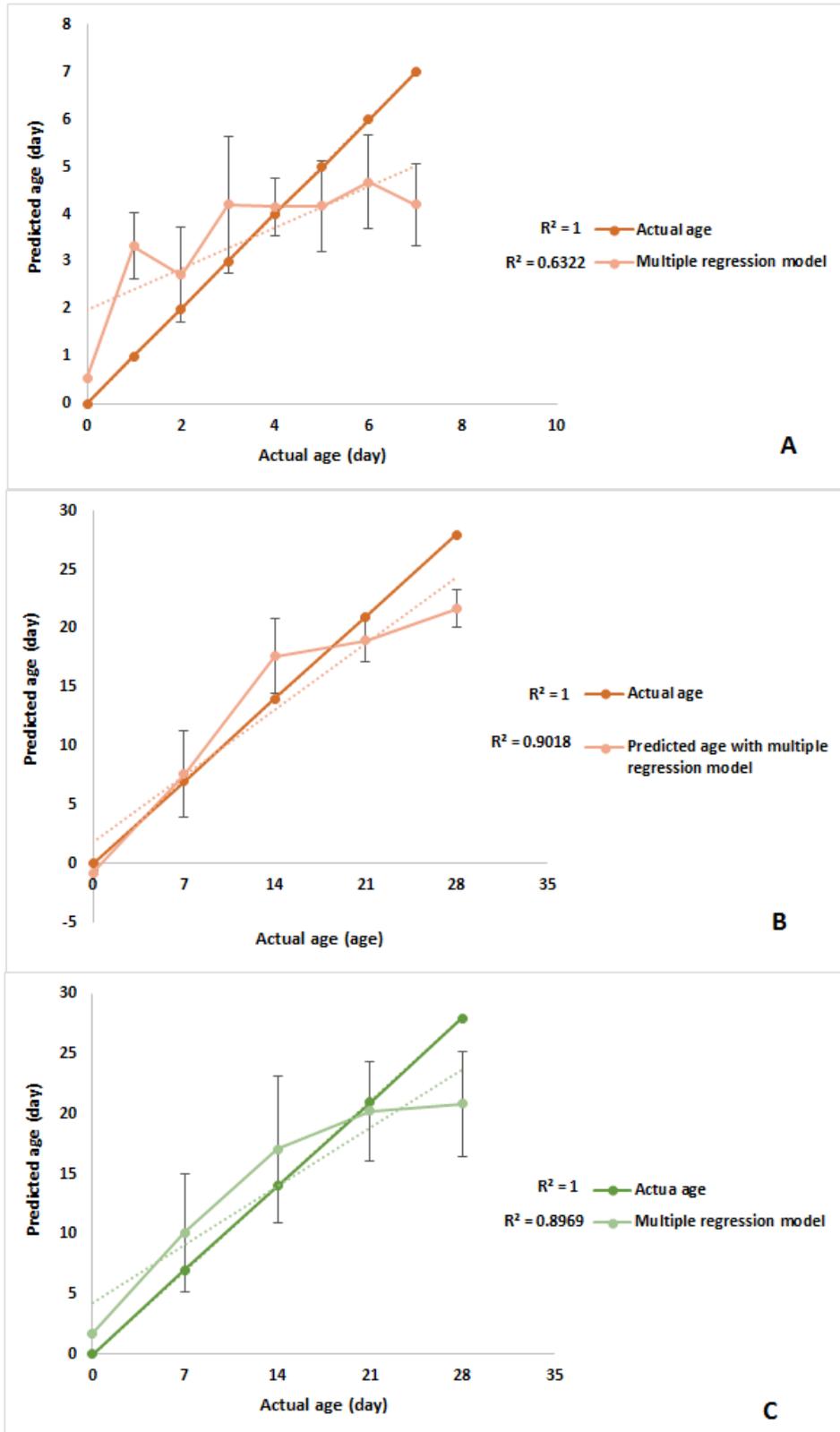


Figure 35. Mean age prediction calculated using multiple regression models (A) Mean prediction in blood samples up to 7 days old. (B and C) Mean age prediction in blood and semen samples up to one-month old, respectively. Error bars represent one standard deviation.

#### 4.4.4. Blind samples

All of the samples examined thus far were of a known age. In order to test the models obtained, blind samples of blood, saliva, and semen were investigated as an attempt to predict their ages using the models obtained previously. The same procedures and statistical analysis were employed, and some of models in blood and semen were validated.

##### 4.4.4.1. Blind blood samples

Five blind blood samples were tested using ALAS2/B2M and HBB/B2M models with TaqMan chemistry. The age of samples used varied from fresh to one-year old (Table 16).

Table 16. Blind blood samples with actual age used for the validation process.

<b>Sample</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
Source	Blood	Blood	Blood	Blood	Blood	Water
Age	Two weeks	One month	One year	Six months	Fresh sample	

All samples underwent RNA extraction, DNA digestion, cDNA synthesis, and qPCR. RQ of selected markers was obtained using B2M as reference gene. Both models with TaqMan chemistry generated similar findings, and both gave no amplification with F sample (Negative control). The samples were ordered according to the RQ of selection markers (Figure 36). The RQ of ALAS2/B2M and HBB/B2M decreased as a function of time and were able to put in the right order most of blind samples when compared to the actual age.

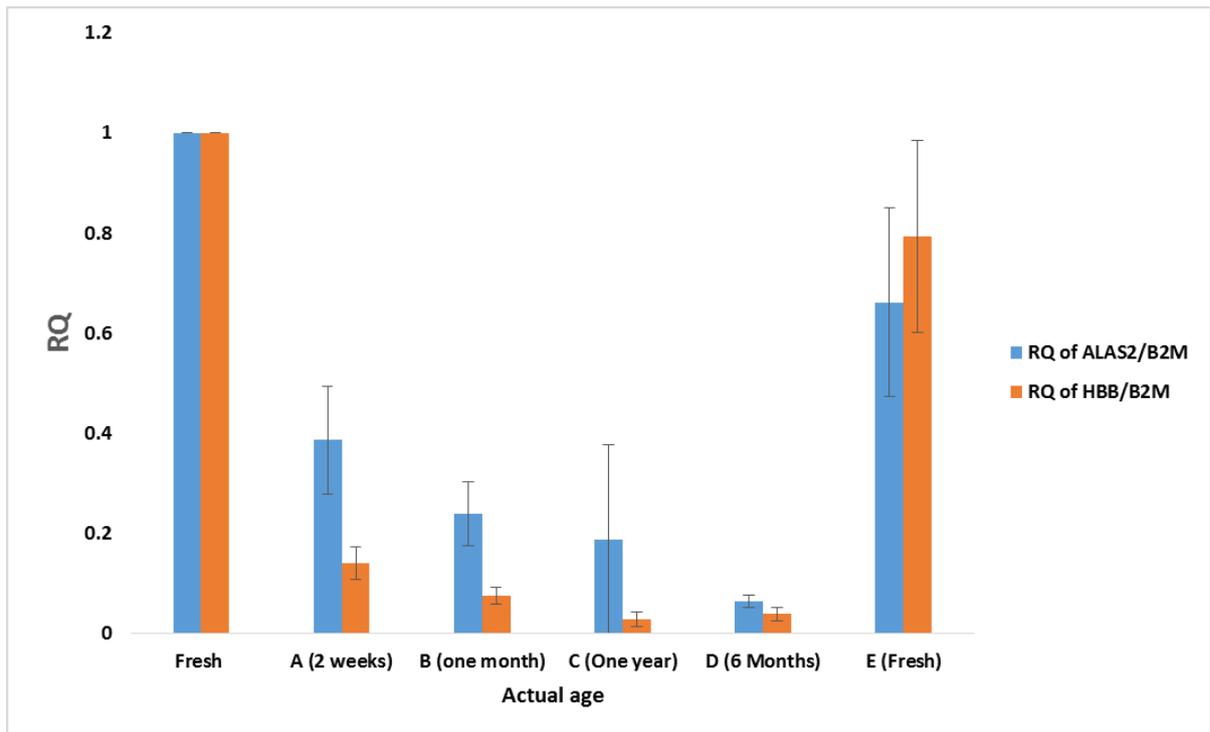


Figure 36. Relative quantification (RQ) of blind blood samples using TaqMan chemistry with ALAS2/B2M and HBB/B2M. Data presented as Mean  $\pm$  SD.

By using the HBB/B2M model, blind blood samples were put into the correct order while this was almost true with ALAS2/B2M model, with only C and D samples being swapped. Therefore, it is reasonable to assume that the RQ could be a method of choice for samples age prediction. With this in mind, to predict the age of the blind samples, RQ value was used as parameter and simple and multiple regression models were used to calculate the prediction age (Appendix 2. Table 38), as presented in Table 17.

Table 17. Age prediction of blind blood using the models obtained.

Sample	Actual age	ALAS2/B2M model	HBB/B2M model	Multiple regression model
A	Two weeks	16.93 ±3.8	23.06 ±3.2	21.30 ±3.4
B	One month	20.85 ±3.8	24.78 ±3.2	24.02 ±3.4
C	One year	22.18 ±3.8	26.01 ±3.2	25.37 ±3.4
D	Six months	25.47 ±3.8	25.73 ±3.2	26.48 ±3.4
E	Fresh	09.66 ±3.8	06.02 ±3.2	06.87 ±3.4

Overall, models were able to reasonably accurately predict the age of the samples that were up to one month old, while having more difficulty with assessing the aged samples. This is perhaps not surprising, since the models used for prediction were generated using samples up to 28 days old. Indeed, the prediction age obtained for these samples was within CI and PI regions previously generated using simple and multiple regression analysis as described.

#### 4.4.4.2. Blind semen samples

As the same as conducted in blood was also performed in semen blind samples, the RQ of PRM1/B2M, and SEMG1/B2M, and PRM1/SEMG1 was obtained as described earlier. The RQ was used as a parameter to predict the blind semen samples with selection models. Prediction age was calculated using all TaqMan semen models (Appendix 2. Table 41) and the data are presented in Table 18.

The best age prediction for semen samples at the time zero and 28 days was generated with PRM1/SEMG1 and multiple regression models (3.17 and 23.85 day, with MAD of ±4.3 and ±4.4, respectively). In addition, an appropriate age prediction for day 14 was generated with PRM1/B2M and SEMG1/B2M models giving values of 15.39 and 15.84 days, respectively. For easier representation of the data, age prediction for all models plotted against actual age is presented in figure 37.

Table 18. Blind semen samples with TaqMan models.

<b>Actual Age</b>	<b>PRM1/B2M Model</b>	<b>Difference between ages</b>	<b>SEMG1/B2M Model</b>	<b>Difference between ages</b>	<b>PRM1/SEMG1M odel</b>	<b>Difference between ages</b>	<b>Multiple regression model</b>	<b>Difference between ages</b>
Fresh (0day)	05.58 ±4.7	05.58	03.78 ±4.4	03.74	03.17 ±4.3	03.17	03.59 ±4.4	03.59
One week (7days)	14.49 ±4.7	07.49	06.65 ±4.4	-00.35	15.94 ±4.3	08.94	07.89 ±4.4	00.89
Two weeks (14days)	15.39 ±4.7	01.39	15.84 ±4.4	01.84	17.96 ±4.3	03.96	11.89 ±4.4	-02.11
Three weeks (21days)	17.07 ±4.7	-03.93	20.02 ±4.4	-00.98	20.41 ±4.3	-00.59	17.90 ±4.4	-03.10
Four weeks(28days)	19.20 ±4.7	-08.80	22.47 ±4.4	-05.53	23.20 ±4.3	-04.80	23.85 ±4.4	-04.15

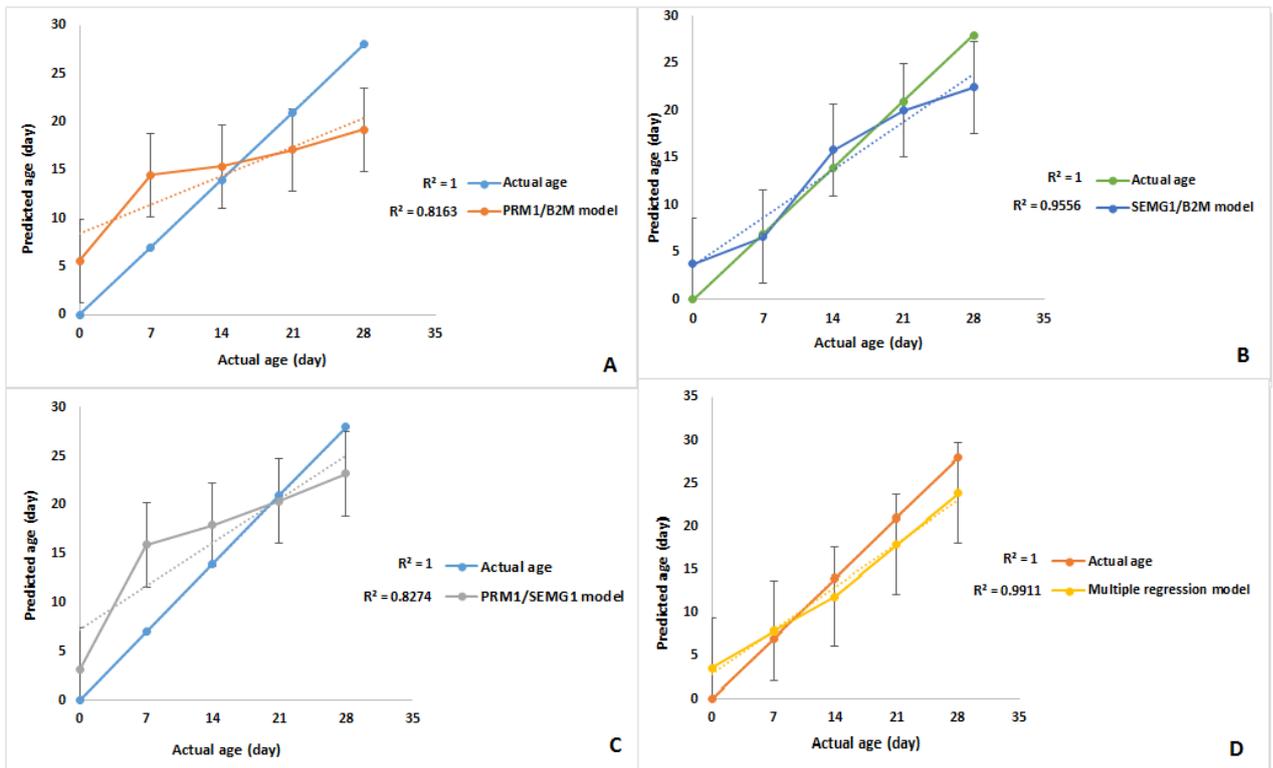


Figure 37. Blind semen samples predicted using TaqMan markers. Age prediction using (A) PRM1/B2M (B) SEMG1/B2M (C) PRM1/SEMG1 (D) multiple regression model. Error bars presented mean absolute deviation (MAD).

The strongest  $R^2$  value for age prediction was obtained with multiple regression model, further confirming the advantage of using the multiple regression over a single regression analysis. The SEMG1/B2M generated a good score, and closely predicted the actual age especially at 7, 14, and 21 days (Figure 37 B). In general, all TaqMan models for semen showed reasonably accurate prediction of the unknown age.

#### 4.5. Discussion

The current study was conducted to quantify degraded RNA in old stains and to compare it to the fresh samples (samples processed at the time zero) in order to create a time calibration curve. The fresh sample was used as a calibrator to predict the age of any biological stains using the formula  $2^{-\Delta\Delta Cq}$ .

In one-week experiment, bloodstains were allowed to decompose at room temperature for up to a week. RNA was then extracted from the samples, and they then underwent cDNA synthesis and qPCR targeting of ALAS2, GAPDH and B2M markers. Lack of strong correlation detected here between RQ and age of the samples (Figure 19. B, C, and D) could be influenced by the reagents used, such as DNA digestion kit (Turbo DNA free) and RT enzyme (M-MLV). As described in chapter Three, M-MLV is not suitable for unbounded gene or low copy gene reverse transcription, while Turbo DNase reagents can inhibit PCR reaction (Okello *et al.*, 2010, Haas *et al.*, 2011b). Indeed, the mean  $\Delta Cq$  of the same markers (ALAS2 and B2M) were different when different reagents were used (Figures 23 A). The influence of reagents was also observed in semen with and without DTT extraction, because the mean age prediction has different  $R^2$  values (Figures 32 and 33).

Another important factor for the selection of reference gene is its stability. Up until recently, it was a common practice to use GAPDH as endogenous control for normalisation of the target genes. This was because, due to playing an important role in cell function and homeostasis, it was considered that GAPDH is constitutively and at similar levels expressed across different cell types. However, studies that would dispute this started to emerge. For example, Dheda *et al.* (2004) (Dheda *et al.*, 2004) stated that GAPDH was unsuitable as a reference gene in the whole blood study, while another group found that GAPDH was not a successful reference gene in one step experiment with saliva and blood samples (Moreno *et al.*, 2012). This could explain the poor correlation detected in saliva samples in the present study when using

HTN3/GAPDH (Figure 27 B). It is now widely accepted that the optimal reference gene should not vary as a function of treatment or condition (time in the present example). It is, however, often difficult to identify a gene or a combination of genes, that would meet this criterion (Thellin *et al.*, 1999).

Further, an optimal primer should be abundantly expressed and sensitive enough even in low template concentration. In addition, its ability to decay should be easy to observe. Half-life and its function are both very important. An appropriate reference gene should be stable in all the samples under investigation. For example, strong correlation was reported for the ratio generated between  $\beta 2$  mRNA, which degraded over time, and 18S rRNA, which remained stable over the course of experiment, up to 150 days (Anderson *et al.*, 2005). The stability of reference genes is also dependent on the main source of samples. For examples, some genes are proven to be good reference genes in blood but were not stable in saliva and vice versa. In this chapter, HTN3 was detected in all samples, but RQ of HTN3/GAPDH was not strongly correlated with actual age. In contrast,  $\Delta Cq$  of NHT3 showed a strong correlation (Figure 26), and as such was used to predict the age of saliva instead of RQ of the same marker (Figure 27 A and C). Surprisingly, perhaps, the best fit was obtained, however, when both genes initially selected as targets were subsequently normalised with each other instead of using common reference genes. This is likely to be due to the similarity of amplification rate between both targets, because when ALAS2 and HBB were normalised with each other (ALAS2/HBB), no correlation was detected.

It is important to note that the gene expression no longer occurs once the stain is completely dried. The environmental factors such as humidity and temperature also influence the drying time of any deposited stains. It is widely accepted that the level of RNA transcription generally decreases with a longer PMI (Bauer *et al.*, 2003a, Sampaio-Silva *et al.*, 2013, Partemi *et al.*, 2010). Therefore, the transcript mRNA could be detected and quantified by

using a fresh sample as calibrator with either TaqMan or SYBR Green chemistries over four weeks period. Therefore, in the follow up experiment, the time frame was increased from one week to one month, and RQ was employed. This was achieved by allowing blood, saliva, and semen stains to decompose at room temperature for up to 28 days.

The stability of mRNA was tested in a number of studies and showed the possibility to detect mRNA expression in aged biological stains. The longest time reported was for HBB marker that remained the most stable when compared to seven different markers including ALAS2, and GYPA, in bloodstains aged from 30-50 years old (Zhao *et al.*, 2017).

In current study, the amplification of selected markers was successfully detected in the samples up to 28 days old, except for MUC7 in saliva. The sensitivity, specificity and stability of this marker could be the main reason for these findings, and it is further confirmed by the previous studies that were able to detect MUC7 was in one negative 1-year old sample, but not in the fresh sample at time 0. This marker also showed low Q value level and was amplified in non-saliva samples (Parker, 2011, Roeder and Haas, 2013, Orphanou, 2015).

Regression analysis is dependent on four principle assumptions: the assumption of normality, assumption of linearity, assumption of constant variance (Age) and assumptions of independence (RQ). However, the violation of the normality assumption may be attributed to the skewed nature of dependent variable. Consequently, the validity of normality can be ignored in the application of linear regression model (Li *et al.*, 2012, Nangia *et al.*, 2011, Sherwin *et al.*, 2011, Zheng *et al.*, 2011). Therefore, when all of the assumptions, except the assumption of normality, were met, the regression analysis was applied. The selection primers showed to be suitable markers for estimating the age of blood, saliva, and semen samples. Wider confidence and prediction intervals obtained throughout the study are likely related to the small amount of starting material, and non-normal distribution of the data. Increasing the

sample size would decrease the sampling error and narrow both confidence and prediction intervals (Wei *et al.*, 2016), as well as improved the normality (Li *et al.*, 2012).

RQ was calculated and used as a parameter to generate models for blood, saliva, and semen samples. However, MAD in blood samples with ALAS2/B2M (7 days) was lower than with the same model at 28 days, although  $R^2$  of ALAS2/B2M (7 days) was the lowest. In general, HBB/B2M emerged as the best model generated for blood ( $R^2 = 0.82$ ), whereas, PRM1/SEMG1 in semen showed the strongest fit ( $R^2 = 0.71$ ) (Tables 12 and 14).

Once the predictive models for blood, saliva, and semen were developed, blind testing was carried out. In the blind blood samples, two-time frames were investigated. The RQ of selected markers clearly decreased over time and positioned most of the blind samples in the right way, despite most of them being over 28 days old (Figure 36). Therefore, the predicted age in the aged stains (over 30 days old) provided good results. To test the models with shorter degradation time, semen models including PRM1/M2M, SEMG1/B2M, PRM1/SEMG1, and multiple regression models were employed to investigate blind sample ranged from fresh up to one-month old. Appropriate results were obtained, and the best age prediction was generated with multiple regression model ( $R^2 = 0.99$ ) (Figure 37 D).

#### **4.6. Conclusion**

Accurate quantification of degraded RNA depends on many factors including reagents, normalisation strategy, and time of deposition. In this chapter, the linearity value of the same markers was improved when different reagents for cDNA synthesis (Random primers, DNA digestion, and RT enzymes) and different deposition times (One week and one month) were investigated.

**Chapter Five: Degradation of mRNA as an indicator to predict the age of biological stains by using SYBR-Green chemistry.**

## 5.1. Introduction

SYBR Green chemistry is the simplest and cheapest method to perform qPCR analysis. In this technique, a fluorescent dye binds with the minor groove of double stranded DNA (dsDNA). When SYBR Green is free in solution where only single-stranded DNA (ssDNA) is present, it emits a low intensity signal. As the PCR progression starts, the quantity of dsDNA increases, so more dye binds with the amplicons generated and hence, the signal intensity increases as a consequence. This dye is the most common dsDNA-binding dye with a long history of using in molecular biology (Dragan *et al.*, 2012). Although the reagents and instruments required for real-time PCR are more expensive than those used in traditional PCR, it is becoming increasingly more affordable, especially when employing SYBR Green as the fluorescent reporter dye.

SYBR Green chemistry has been successfully used in a number of studies to examine gene expression analysis. Yin and co-workers (2001) demonstrated that the reproducible and acceptable quantification could be obtained with SYBR Green I method in cases where the target gene had moderate to high expression levels. Further, they found that the specificity of SYBR Green I detection improved once they optimised the melting temperature (Yin *et al.*, 2001).

A pioneering mathematical model for RQ in RT-qPCR was made using SYBR Green, where the relative expression of target gene transcript was presented in comparison to a reference gene transcript. Here, relative expression ratio was calculated using only the qPCR efficiencies and the crossing point deviation of an unknown sample when compared to the control (Pfaffl, 2001). Therefore, the aim of this chapter is to investigate a number of markers including the hypoxia marker using SYBR Green chemistry for samples that were up to one-month old.

## **5.2. Experimental design**

Sample collection, RNA extraction, cDNA synthesis, and qPCR were performed in blood, saliva and semen. All samples were allowed to dry and left to decompose in a one-month period. Relative quantification was obtained using SYBR Green chemistry followed by a melting curve analysis to increase confidence in the fact that the amplification was detected in a real target sequence as discussed in chapter 3. The same analysis which was performed with the TaqMan probe was conducted in the present study using unlabelled markers listed in Table 6. Some of these markers are related to the hypoxia condition, suggesting the possibility of using the hypoxia as an indicator for age prediction in blood, saliva, and semen.

## **5.3. Results**

### **5.3.1. Blood markers**

Three experiments were conducted on blood samples using SYBR Green. The first experiment was performed using HBB, the second used FN1 and EPOS1 markers, while the third used FGB and CO-A. ACTB was used as a reference gene throughout. All markers were successfully detected and delta Cq was calculated in all cases. HBB emerged as the most abundant marker, and ACTB generally appears to have high amplification rate (Figure 38. A, B, and C).

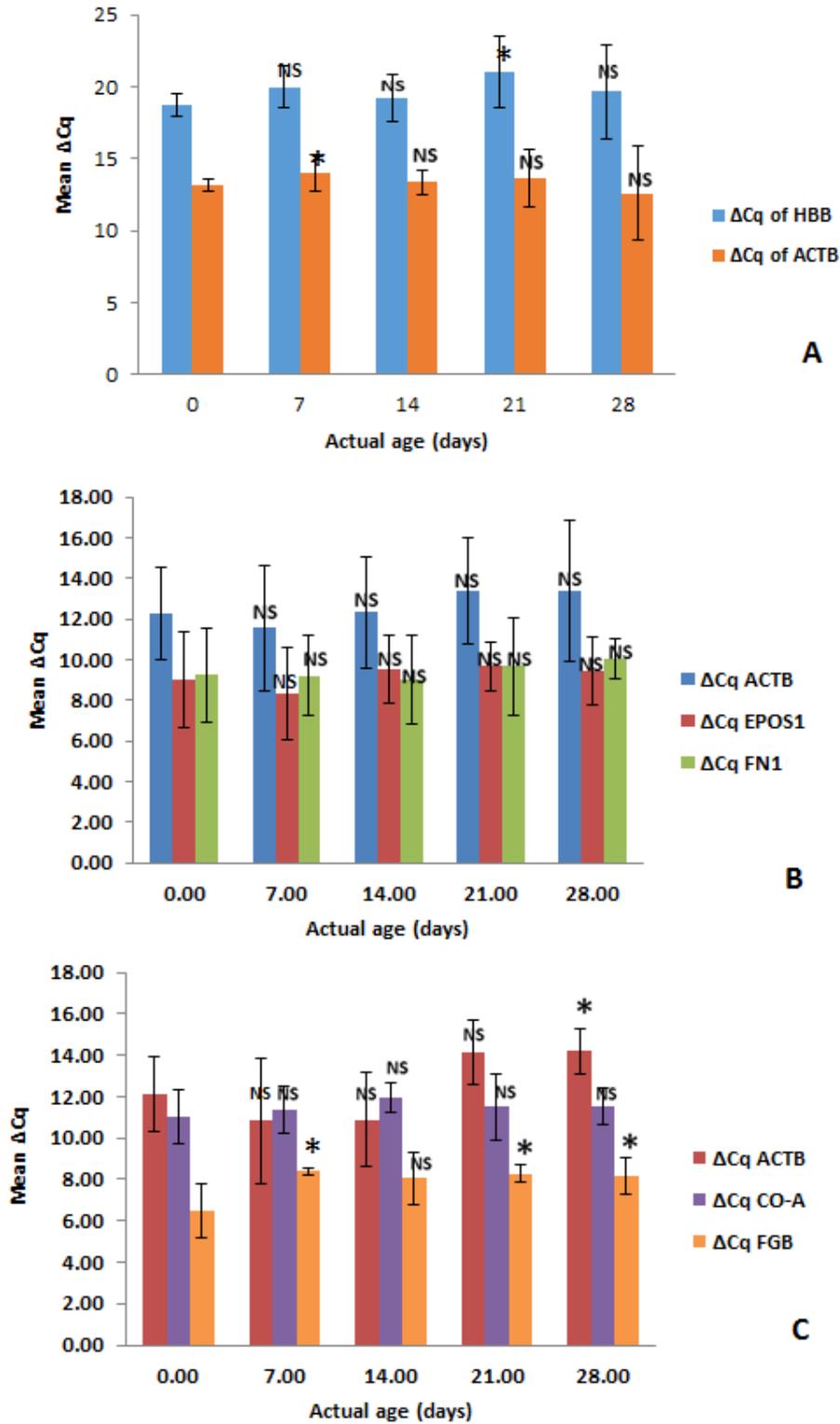


Figure 38. Delta Cq of blood markers using SYBR Green chemistry over a period of 28 days. ΔCq of markers (A) HBB (B) FN1 and EOPS2 and (C) FGB and CO-A markers. ACTB was used as reference gene in all cases. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and column with NS indicates no significant difference. Error bars represent one standard deviation (N= 25 for HBB marker, N = 20 for all other markers).

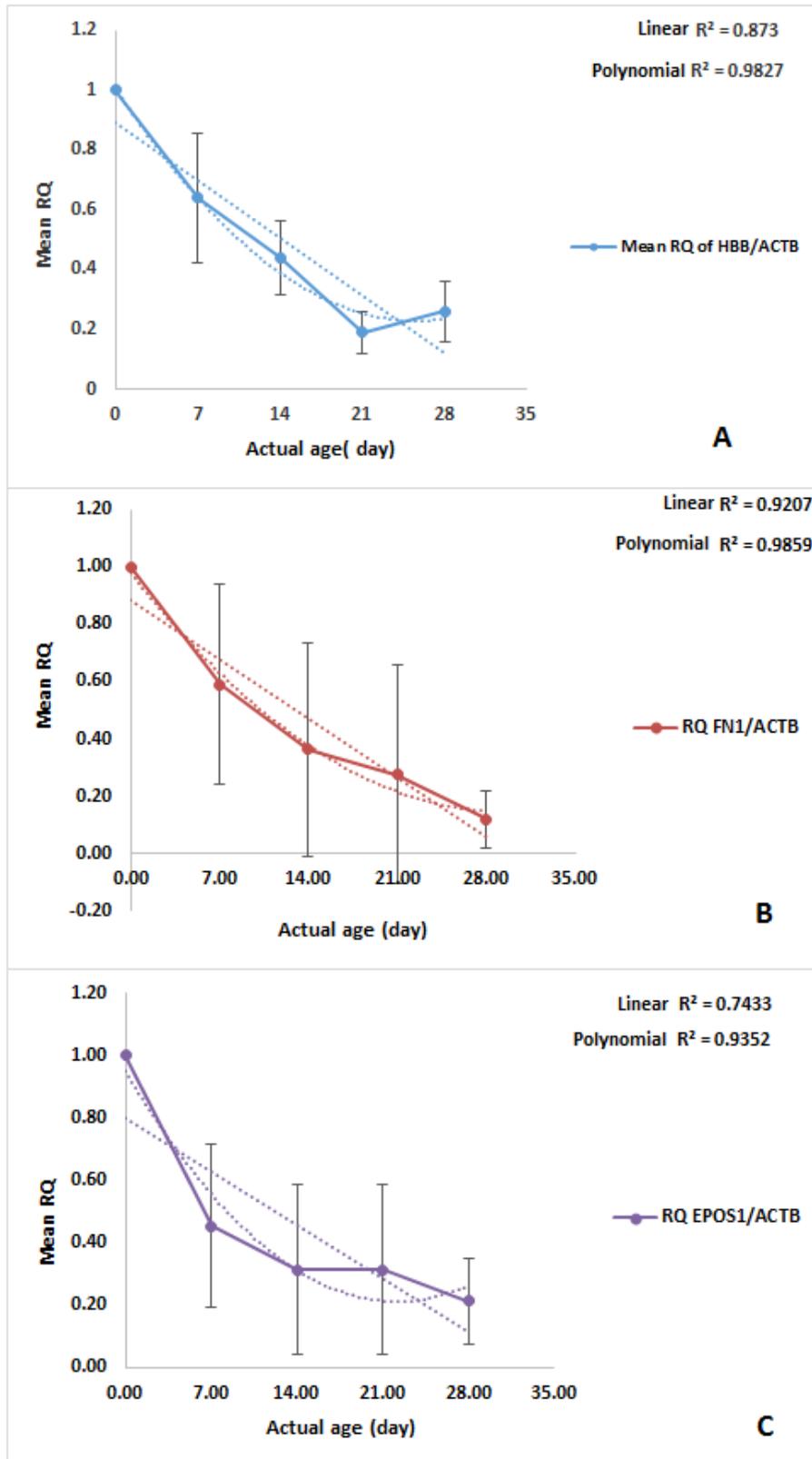


Figure 39. RQ in blood samples over 28 days using markers (A) HBB, (B) FN1 and (C) EPOS1. ACTB was used as reference gene in all cases. Data presented as Mean  $\pm$  SD.

As before, RQ was calculated for selected markers, and the strongest correlation was detected when using FN1 /ACTB ( $R^2 = 0.92$  and  $0.99$  for linear and polynomial relationships, respectively) (Figure 39. A, B, and C).

Pearson's correlation was conducted for all markers and the strongest negative correlation was detected using HBB/ACTB ( $r = -0.87$ ) (Table 19). Therefore, single regression analysis with confidence and prediction intervals at 95% was employed in an attempt to generate models. As expected, the strongest linear curve with an  $R^2 = 0.76$  was obtained for HBB/ACTB (Figure 40 A).

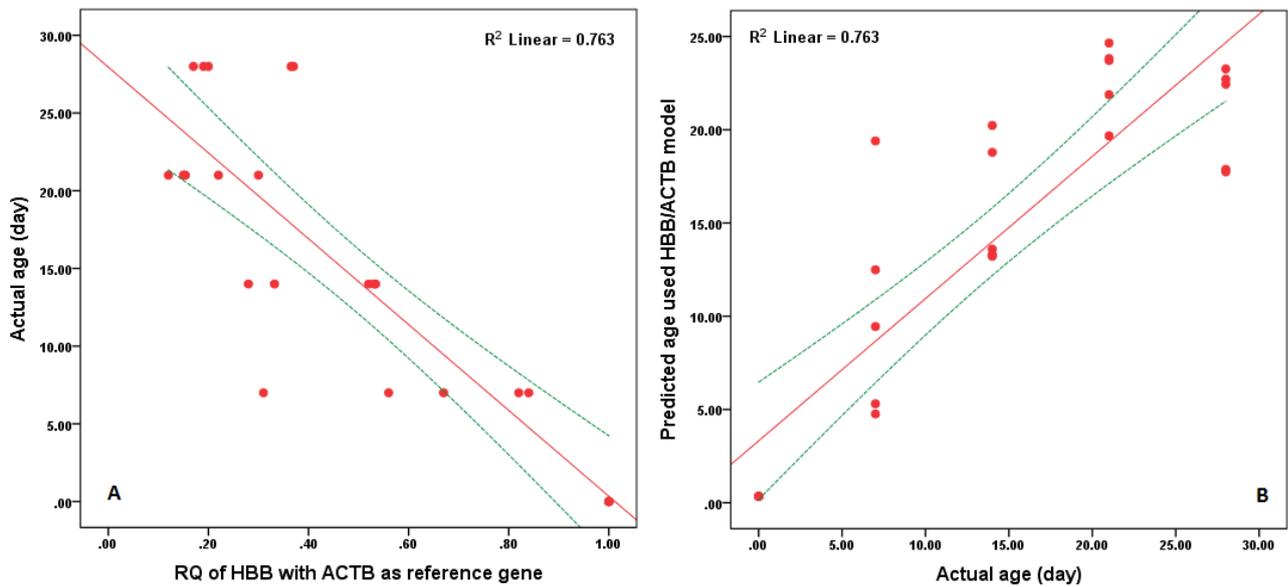


Figure 40. Regression analysis of HBB/ACTB in blood using SYBR Green. Solid red line on the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

RQ of HBB/ACTB was plotted against actual age, resulting in a linear curve with  $R^2 = 0.76$  (Figure 40 A). Age prediction was calculated and plotted against the actual age. It is clear that some of the prediction data points were reasonably close to actual age, especially at 14 and 21 days (Figure 40 B). For all other markers, the same statistical approach was used, and details are presented in Table 19.

Table 19. Summary of the models obtained for blood samples using SYBR Green chemistry.

Type of sample	Time frame	Number of samples	Pearson's correlation(r)	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Month	25	-0.87**	HBB	ACTB	Linear	$y=27.96 - 27.62x$	0.76	±3.4
Blood	Month	20	0.72**	EPOS1	ACTB	Cubic	$y=21.6-20.27x+39.97x^2- 41.46x^3$	0.57	±5
Blood	Month	20	-0.73**	FN1	ACTB	Cubic	$y=20.7-1.7x+0.25x^2-17.9x^3$	0.60	±4.9
Blood	Month	20	-0.65**	FN1	EPOS1	Cubic	$y=17.21-8.61x+61.01x^2-70.1x^3$	0.58	±4.9
Blood	Month	20	-0.43*	FGB	ACTB	Cubic	$y=8.9+33.86x+33.35x^2-75.9x^3$	0.73	±3.9
Blood	Month	20	-0.53*	COA	ACTB	Quadratic	$y=10.15 + 51.56x - 61.51x^2$	0.54	±5.1
Blood	Month	20	-0.43*	FGB	COA	Quadratic	$Y=2.68 -78.18x -76.69x^2$	0.54	±5.8

Note: (\*) represented  $p<0.05$  and (\*\*)  $p<0.01$

It is clear that all models were significantly correlated with the actual age. The HBB/ACTB model has the strongest  $R^2$  value accompanied by the lowest MAD. Therefore, HBB/ACTB could be considered as the best model to predict the age of blood samples. The HBB marker in both chemical chemistries (TaqMan and SYBR Green) was correlated reasonably with the actual age, giving a similar result between the different approaches. This supports the use of SYBR Green chemistry with melting curve analysis, since the increased specificity of the marker used is not compromised, while the cost of the experiment is reduced.

Multiple regression analysis, the method that uses more than one variable, was also conducted for SYBR Green chemistry. Here, however, because the combined models showed no differences with actual age, they could not be used in mixed models to generate prediction.

The mean age prediction was calculated (Appendix 1. Table 35) for an observation of the accuracy of each model individually (Figure 41). Unsurprisingly, HBB/ACTB was the best model obtained. All models had good predictive capabilities overall, with some data points particularly closely fitted to the prediction curve, such as time point of a fresh sample (zero point). The comparison between the HBB/ACTB model obtained with SYBR Green and the HBB/B2M performed with TaqMan, showed that HBB/B2M has stronger correlation and a higher  $R^2$  value, even though MAD prediction values were comparable ( $\pm 3.2$  and  $\pm 3.4$  for HBB/B2M and HBB/ACTB, respectively) (Tables 12 and 19).

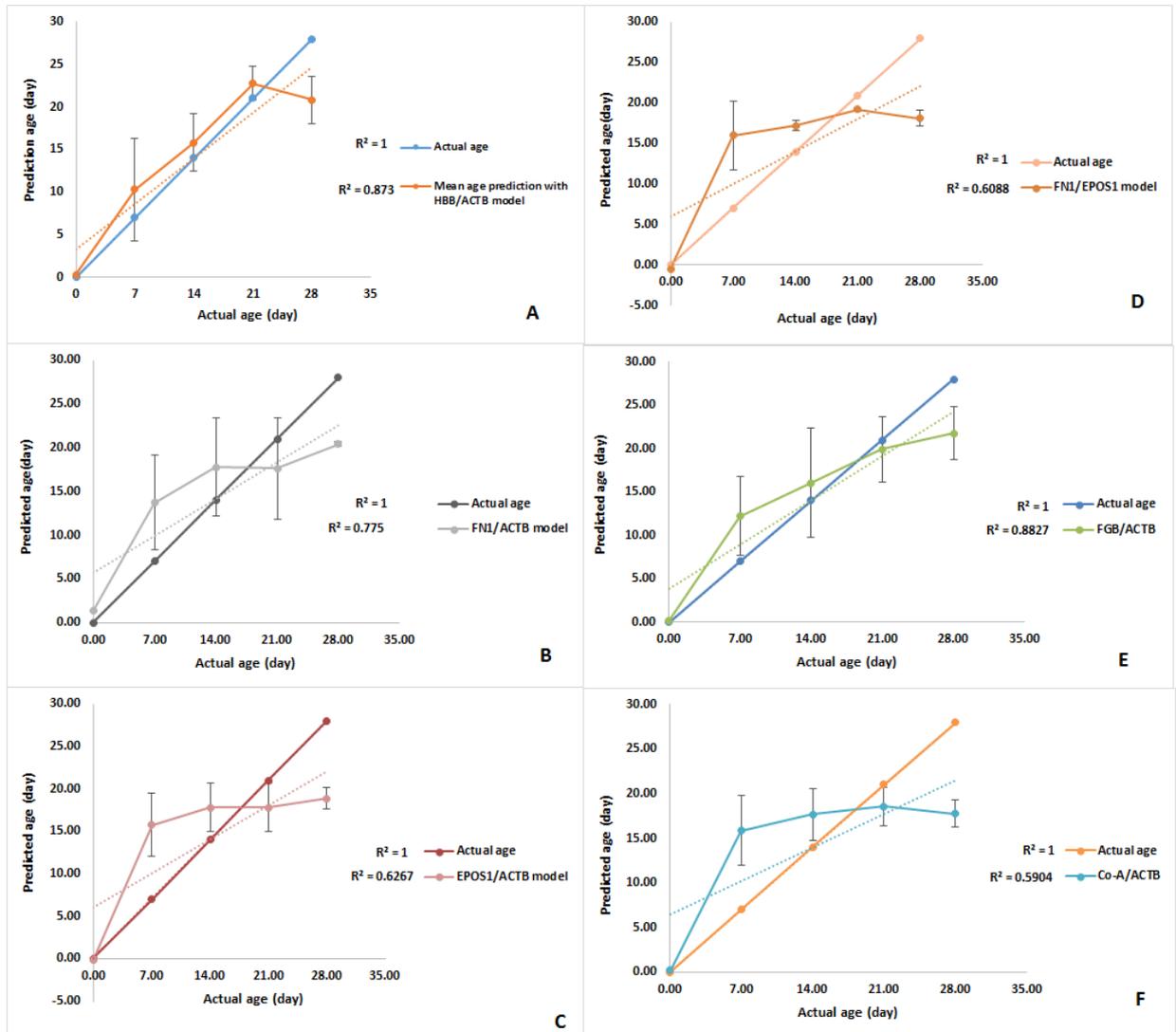


Figure 41. Age prediction in blood samples using the models obtained. (A, B, and C) HBB, FN1, and EPOS1 models generated with ACTB as reference gene. (D) FN1/EPOS1 model. (E and D) FGB and COA models generated with ACTB as reference gene. Data presented as Mean  $\pm$  SD.

### **5.3.2. Oxygen regulated factors**

Many studies have been conducted using mRNA from oxygen-regulated factors such as VEGF, EPO, and HIF1A, to estimate PMI (Thaik-Oo *et al.*, 2002b, Zhao *et al.*, 2006). These factors play a crucial role in regulating oxygen level under the condition of disturbed oxygen homeostasis; i.e, hypoxia. This is why they can be used as useful markers to predict the age of any biological stains. This experiment aimed to investigate VEGFA and HIF1A genes in blood, saliva, and semen using ACTB as a reference gene. The selection of ACTB was based on the study that confirmed ACTB as a stable RG for data normalisation of hypoxia –related genes (Tan *et al.*, 2012). On the contrary, previous reports of the study conducted under hypoxia condition revealed GAPDH as the RG with the highest stability value M, and thus GAPDH becomes the least suitable gene for normalisation of quantitative target gene data (Huth *et al.*, 2013).

#### **5.3.2.1. Hypoxia markers in blood**

In these experiments, all markers were successfully detected in all samples under investigation. In the blood, the mean  $\Delta Cq$  and RQ of hypoxia markers were calculated, with results showing a similar amplification for all of them. Overall, the amplification of ACTB was different for every subsequent time point when compared to the fresh sample, HIF1A was significantly lower at 14 days only, while VEGFA amplification was consistent across a time points (Figure 42 A).

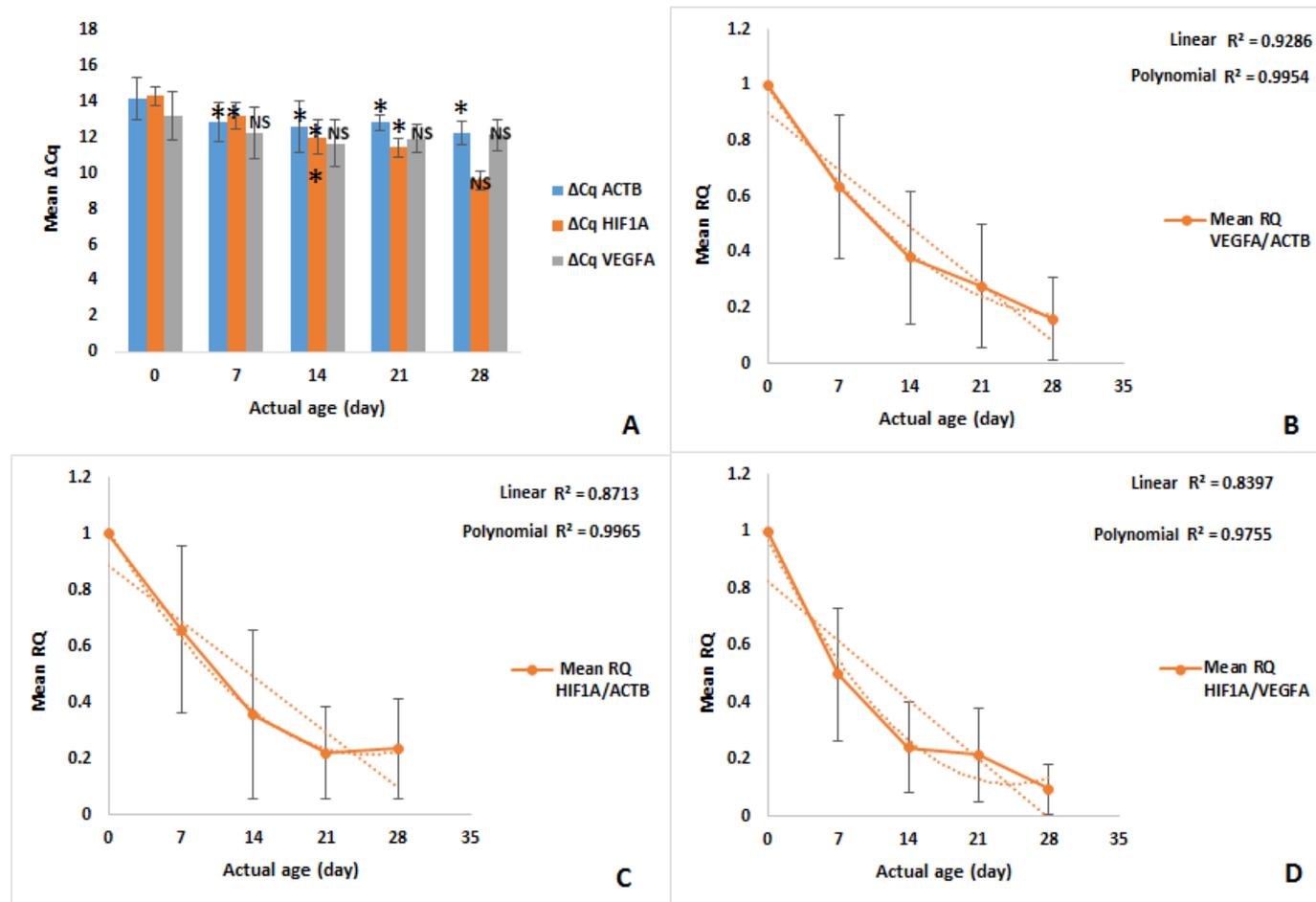


Figure 42. The degradation patterns of hypoxia markers using SYBR Green in blood samples. (A) The mean delta cycle quantification ( $\Delta Cq$ ) of assays over the period of four weeks. (B, C) The RQ patterns of VEGFA and HIF1A with ACTB as reference gene, respectively. (D) The RQ patterns of HIF1A with VEGFA as the reference gene (HIF1A/VEGFA). Column denoted with one asterisk (\*) indicates  $p < 0.05$ , and column with NS indicates no significant difference. Error bars represent one standard deviation (N=40).

Figures 42. B, C and D show the degradation pattern of selected markers in bloodstains over a 28-day period. Markers tested in the study were all tested as the reference gene. In all cases, there is a downward trend in the RQ value, suggesting degradation over time, as expected. The trend appears to be reasonably linear with  $R^2 = 0.928, 0.871$  and  $0.839$  for VEGFA/ACTB, HIF1A/ACTB and HIF1A/VEGFA, respectively. Further, these data indicated that targeting VEGFA with ACTB as the reference gene is the most appropriate combination for bloodstain age prediction. However, the polynomial method gave a higher correlation overall, with  $R^2 = 0.995, 0.997$  and  $0.976$ , for VEGFA/ACTB, HIF1A/ACTB and HIF1A/VEGFA, respectively, which suggests that HIF1A as a target with ACTB as a reference gene is better choice for bloodstain age prediction.

### **5.3.2.2. Hypoxia markers in saliva**

This same procedure described for blood was also performed on saliva samples. The same hypoxia markers were used and both  $\Delta Cq$  and RQ were calculated. It is clear from Figure 43A that the markers were overall more abundant in fresh saliva samples when compared to the fresh blood described above. ACTB as the reference gene was the most abundant marker detected. Unlike in blood samples,  $\Delta Cq$  of hypoxia markers as well as ACTB showed a clear linear function of time in saliva. Hypoxia markers in saliva have the highest correlation with actual age. T-test was used to compare  $\Delta Cq$  values at different time points, showing significant differences when compared to the fresh sample, with the exclusion of VEGFA and ACTB at 7 and 28 days, respectively. Among all selected markers, HIF1A with VEGFA as a reference gene (HIF1A/VEGFA) gave the highest linear and polynomial correlation with  $R^2$  values of  $0.986$  and  $0.996$ , respectively.

Figure 43. B, C and D show the degradation pattern of selected markers in saliva swabs over a 28-day period. As in the blood samples, the RQ value followed a downward trend suggesting

a decrease in the amount of marker present in the sample over time. This trend also appears to be reasonably linear with  $R^2$  values of 0.943, 0.924 and 0.986, respectively. This suggests that targeting HIF1A with VEGFA as the reference gene should be the most appropriate combination for saliva stain age prediction. The polynomial correlation gave 0.977, 0.953 and 0.996, respectively, further confirming HIF1A/VEGFA as the best combination for saliva stain age assessment.

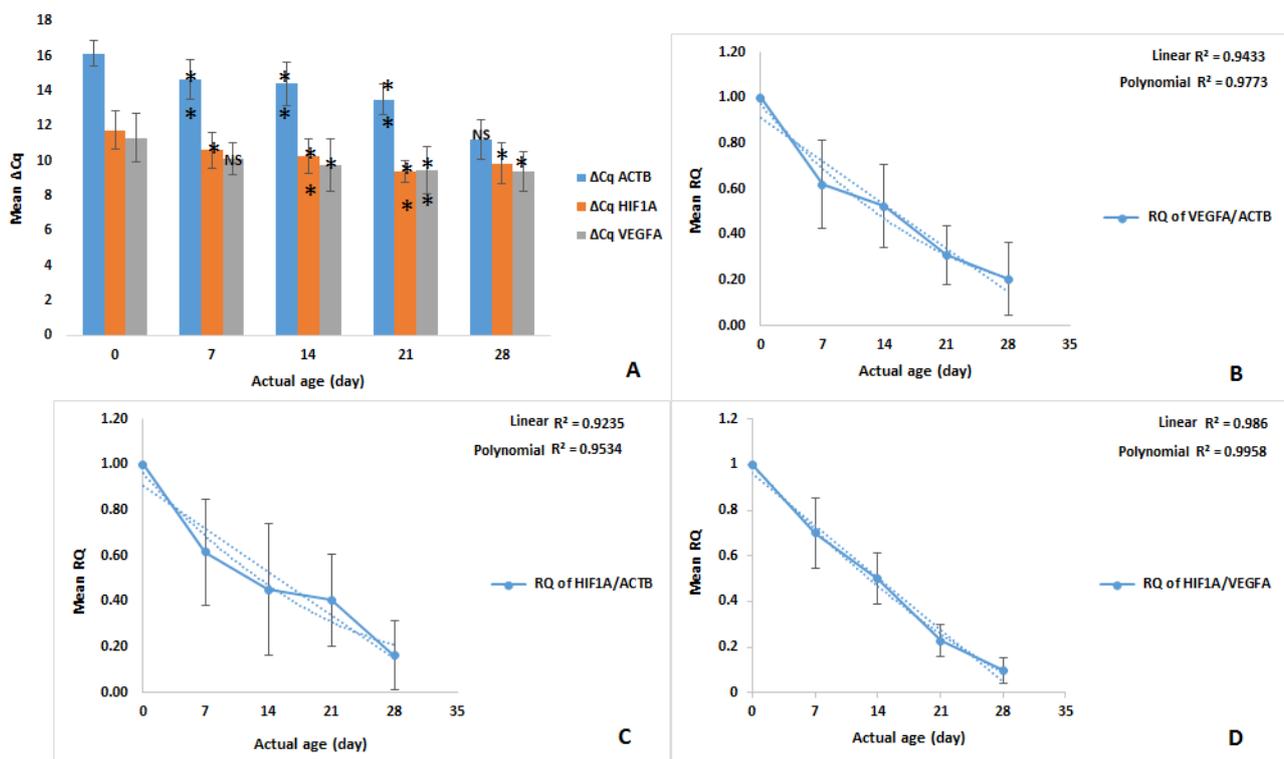


Figure 43. The degradation patterns of hypoxia markers using SYBR Green in saliva samples. (A) Delta cycle quantification ( $\Delta Cq$ ) of assays over four weeks. (B, C) The RQ patterns of VEGFA and HIF1A with ACTB as reference gene, respectively. (D) The RQ patterns of HIF1A with VEGFA as the reference gene (HIF1A/VEGFA). HIF1A normalised with VEGFA had the highest linear and quadratic curves with  $R^2$  values of 0.99. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=40).

### 5.3.2.3. Hypoxia markers in semen

In the semen samples, ACTB gave the highest  $\Delta Cq$  values, while the amplification of VEGFA was the most stable and comparable to the fresh sample, except at 28 days. On the other hand, both HIF1A and ACTB markers showed a decrease in  $\Delta Cq$  values over time (Figure 44 A). Finally, all markers exhibited a downward trend over a 28 day period; however, the correlation values were lower than they were in the blood and saliva samples, with  $R^2 = 0.664, 0.772, 0.732$  for VEGFA/ACTB, HIF1A/ACTB, and HIF1A/VEGFA, respectively. Overall, the results seem to once again point to HIF1A with ACTB as the reference gene for the most appropriate combination for age prediction in semen samples as well. Polynomial correlations gave values of  $R^2 = 0.897, 0.99, \text{ and } 0.97$  for VEGFA/ACTB, HIF1A/ACTB, and HIF1A/VEGFA, respectively, confirming the above conclusion obtained from the linear fit (Figure 44. B, C, and D).

In general, the hypoxia markers showed a strong correlation for all samples under investigation. Further, VEGFA and HIF1A in blood, saliva, and semen samples had a similar amplification pattern, which suggests that normalising these markers with each other should also be considered.

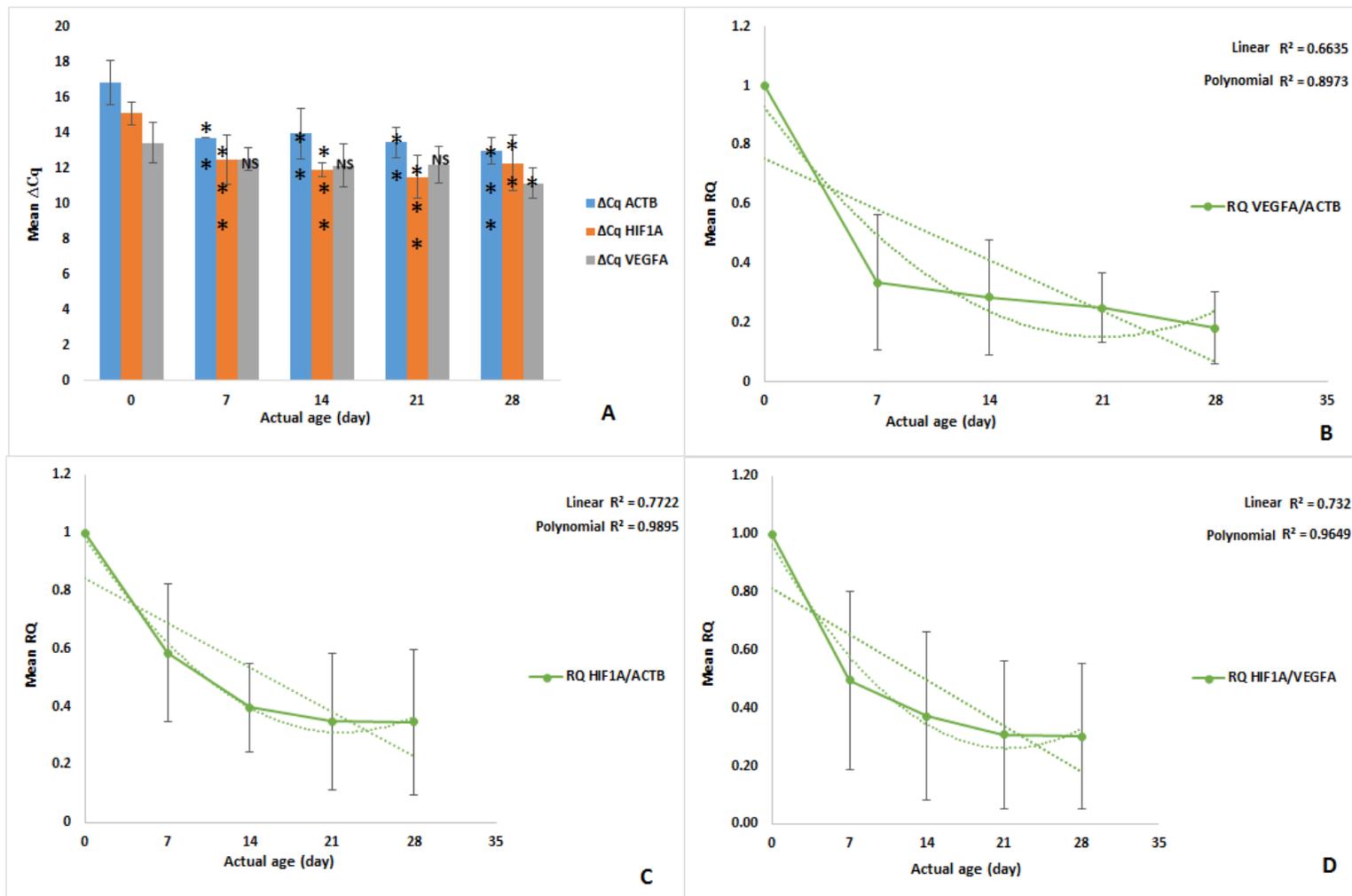


Figure 44. Relative quantification in semen samples. (A)  $\Delta Cq$  of assays over one-month period. (B-C) The pattern of degradation of RQ of VEGFA and HIF1A with ACTB as RG, respectively. (D) The pattern of degradation of HIF1A with VEGFA as RG. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=30).

Pearson's correlation was also performed to test the relationship between hypoxia markers and the actual age of the samples. All markers had a strong negative correlation with time that was statistically significant. Therefore, it is possible to use the regression analysis to obtain models for predicting the age of blood, saliva, and semen samples under hypoxia condition, using oxygen haemostats as indicators. The previous experiments were conducted individually for blood, saliva, and semen, using the same markers. Therefore, it might be clearer to present each marker in blood, saliva and semen at the same time.

### 5.3.3. Vascular Endothelial Growth Factors A (VEGFA)

VEGFA is the member of PDGF/VEGF growth factor family, which is upregulated in many known tumours and its expression is associated with tumour stage and progression. This marker was tested in blood, saliva, and semen samples. Simple regression analysis with confidence and prediction intervals at 95% was performed. The best fit was obtained for VEGFA/ACTB with a cubic curve for blood and saliva ( $R^2 = 0.70$  and  $0.75$ , respectively), and a linear curve for semen samples ( $R^2 = 0.55$ ) (Figure 45. A, B, and C). The equations resulting from the regression analysis are listed in Table 20 below. Using these equations, the prediction age was calculated based on VEGFA/ACTB model and plotted against the actual age (Appendix 1. Table 36).

Table 20. Models obtained with VEGFA marker in blood, saliva, and semen.

Type of sample	Time frame	Number of samples	Pearson's correlation(r)	Target gene	Reference gene	Optimal curve	Estimated curve equation	$R^2$	MAD (day)
Blood	Month	40	-0.82**	VEGFA	ACTB	Cubic	$y=24.39-26.24x+32.14x^2-29.73x^3$	0.70	±4.4
Saliva	Month	40	0.87**	VEGFA	ACTB	Cubic	$y=27.69-22.4x-6.39x^2+1.79x^3$	0.75	±3.9
Semen	Month	30	-0.74**	VEGFA	ACTB	Linear	$y=23.09 - 22.15x$	0.55	±5.4

Note: (\*\*) represented  $p < 0.01$

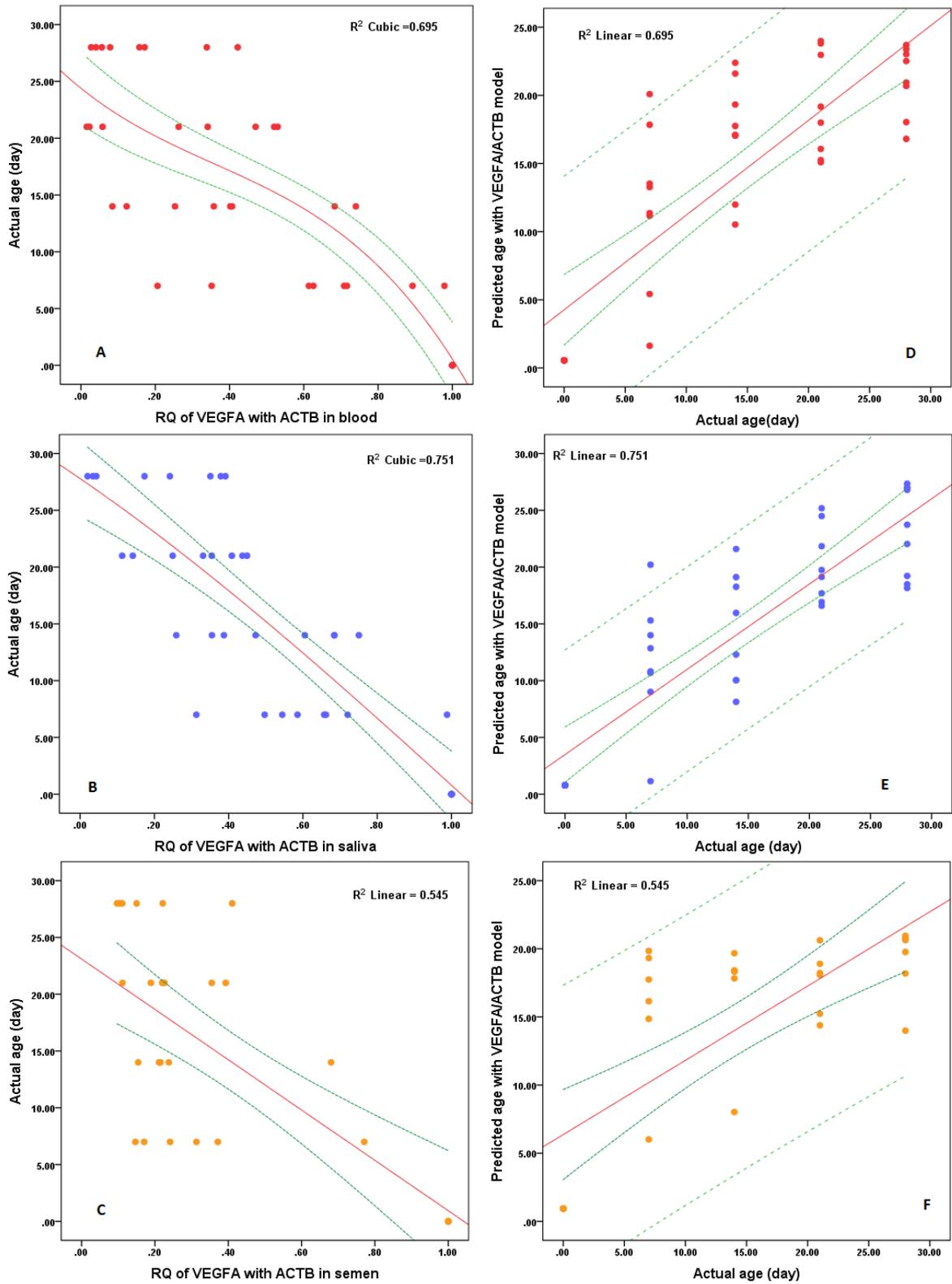


Figure 45. Simple regression analysis of VEGFA marker with ACTB as reference gene. (A) Blood, (B) saliva, and (C) semen. (D, E, and F) Age prediction using VEGFA/ACTB model in blood, saliva, and semen, respectively. Solid red line represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

The higher the  $R^2$  value, the more accurate age prediction becomes. Higher  $R^2$  value in saliva samples overall suggests that these samples can provide better age prediction when compared to blood and semen. A higher number of age prediction points located in the CI region (Figure 45. D, E, and F) also readily confirms this. Further, the mean predicted age was calculated and plotted against the actual age, in order to confirm which samples that have appropriate predictions, and for the evaluation of the model three parameters were used:  $R^2$  value, standard deviation, and MAD.

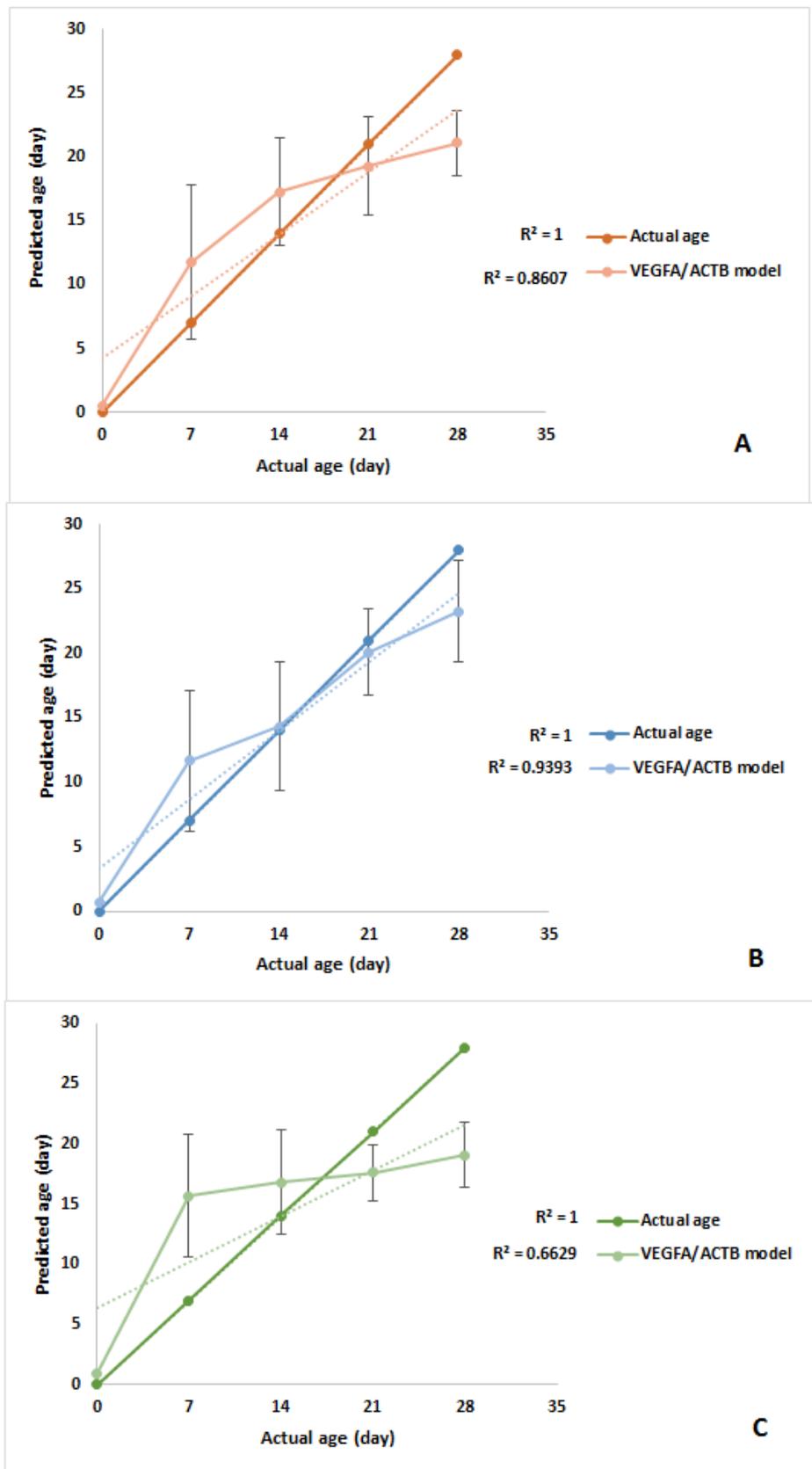


Figure 46. Age prediction obtained with VEGFA/ACTB model in (A) blood, (B) saliva, and (C) semen samples. Data presented as Mean  $\pm$  SD.

In conclusion, Figure 46 supports VEGFA/ACTB as an acceptable model for age prediction in saliva samples with the highest  $R^2$  value (0.94), lowest standard deviation, and lowest MAD ( $\pm 3.9$ ) (Table 20). In saliva samples, it was able to accurately predict three time points at 0, 14, and day 21.

### 5.3.4. Hypoxia Inducible Factor 1 Alpha Subunit (HIF1A)

There are two forms of this gene, an alpha and a beta subunit that form a heterodimer. This gene plays a major role in the hypoxia condition by regulating cellular and systemic responses to the changes in available oxygen. In the present study, HIF1A was also examined in blood, saliva, and semen samples. Pearson's correlation of HIF1A/ACTB was conducted and the strongest negative correlation was observed in saliva samples ( $R^2 = -0.80$ ), while semen samples demonstrated the lowest correlation ( $R^2 = -0.71$ ) (Table 21). A simple regression analysis was performed using RQ of HIF1A/ACTB. The best fit for this factor was obtained with cubic curve in blood ( $R^2 = 0.65$ ), and quadratic curves in saliva and semen ( $R^2 = 0.65$  and  $0.52$ , respectively (Figure 47. A, B, C). The low  $R^2$  values could be due to the high standard deviation at each time point. This is probably why the mean RQ of HIF1A/ACTB for blood, saliva, and semen reported previously (Figures 42C, 43C and 44C) showed a strong correlation with the actual age.

Table 21. Models obtained with HIFA1/ACTB in blood, saliva, and semen.

Type of sample	Time frame	Number of samples	Pearson's correlation(r)	Target gene	Reference gene	Optimal curve	Estimated curve equation	$R^2$	MAD (day)
Blood	Month	40	-0.77**	HIFA1	ACTB	Cubic	$y=23.36-32.89x+74.64x^2-64.25x^3$	0.65	$\pm 4.6$
Saliva	Month	40	0.80**	HIFA1	ACTB	Quadratic	$y=24.43 - 11.48x - 11.23x^2$	0.65	$\pm 4.9$
Semen	Month	30	-0.71**	HIFA1	ACTB	Quadratic	$y=21.39 + 0.44x - 19.91x^2$	0.52	$\pm 5.2$

Note: (\*\*) represented  $p < 0.01$

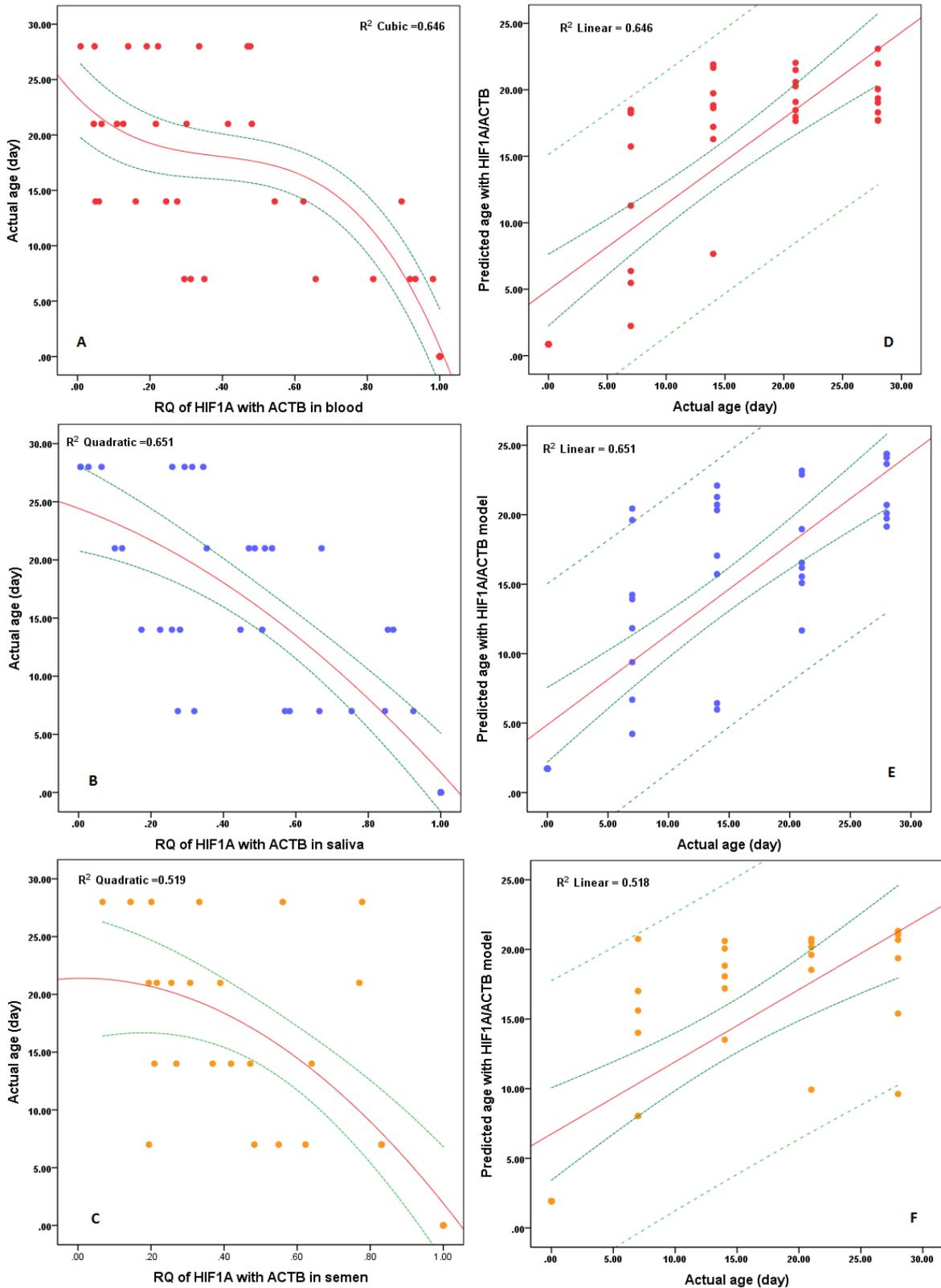


Figure 47. Scatterplots for RQ of HIF1A/ACTB against predicted age. (A-C) RQ of HIF1A in blood, saliva, and semen, respectively. (D-F) Age prediction using HIF1A/ACTB model for blood, saliva and semen samples, respectively. Solid red line represents modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

Equations obtained from regression analysis for blood, saliva, and semen samples are presented in Table 21. Predicted age was calculated (Appendix 1. Table 36) using the HIF1A/ACTB model for blood, saliva, and semen and plotted against actual age (Figure 47. D, E, and F).

In summary, this model was able to fit a number of data points within the confidence interval region, with most of them located close to each other. Therefore, it is reasonable to expect that similar results will be obtained using new samples, especially for saliva and semen as well as samples that are over 7 days old. Mean age predictions for blood, saliva, and semen samples were also calculated, and the  $R^2$  values obtained were 0.80, 0.89 and 0.67 for blood, saliva, and semen, respectively (Figure 48. A, B, and C). The highest  $R^2$  was obtained for saliva; however, the best fit for HIF1A/ACTB model was generated with the blood samples. This is likely to be due to the low SD and MAD in blood (Table 21).

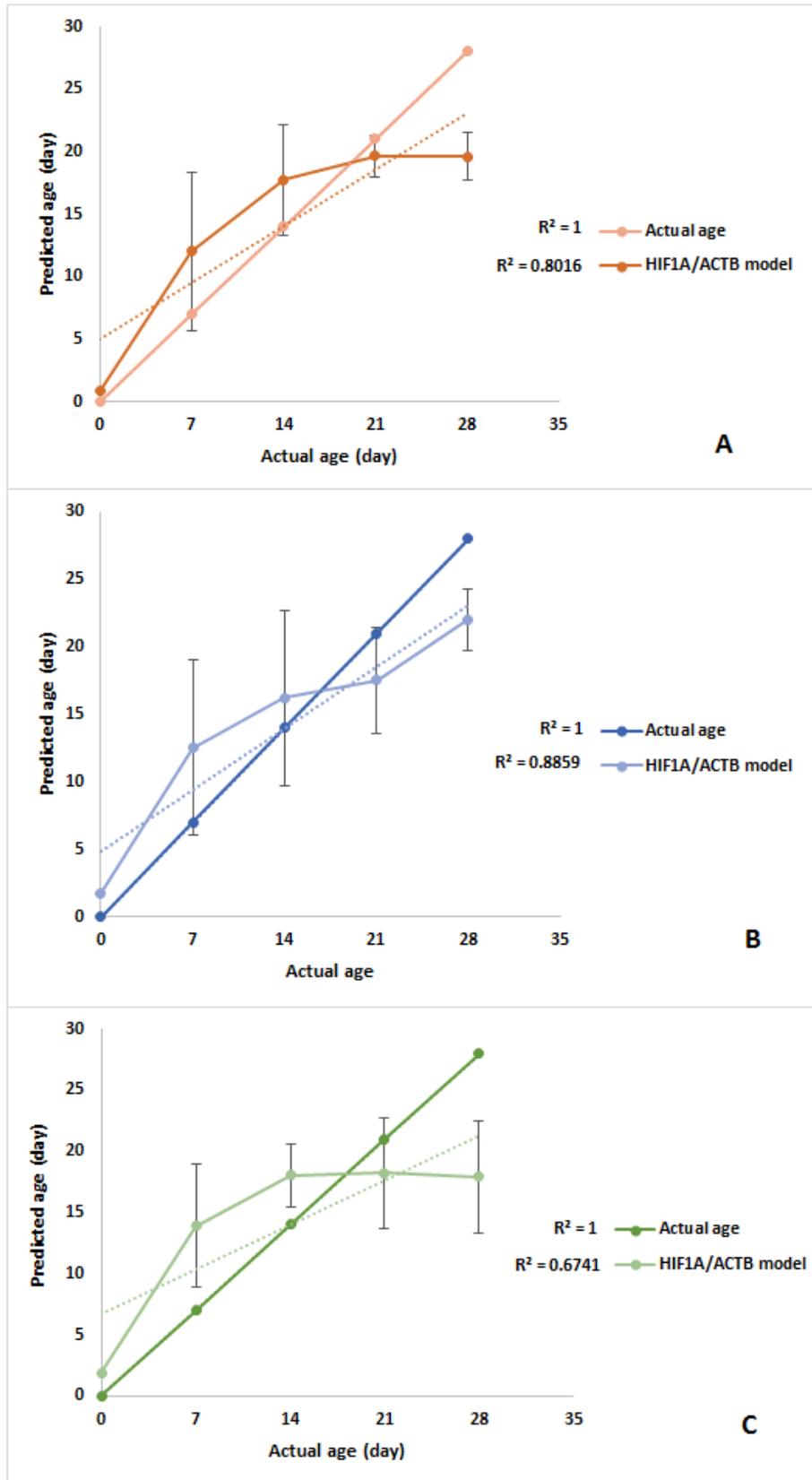


Figure 48. Age prediction generated from HIF1A/ACTB model (A) in blood, (B and C) for saliva, and semen, respectively. Data presented as Mean  $\pm$  SD.

### 5.3.5. HIF1A with VEGFA

As mentioned earlier both HIF1A and VEGFA show similar amplification in blood, saliva, and semen samples. This could mean that better normalisation would be achieved when using VEGFA, instead of ACTB, as a reference gene. However, it is now widely accepted that there is no optimal reference gene for normalisation, and that a unique reference gene, or a combination of several reference genes, should be determined for each experiment separately (Huang *et al.*, 2013, Kozera and Rapacz, 2013, Gilsbach *et al.*, 2006). Therefore, the HIF1A marker in blood, saliva, and semen, was normalised using VEGFA as a reference gene. The same procedures and analysis were performed. The results showed a strong linear relationship with  $R^2 = 0.84, 0.99,$  and  $0.73$  for blood, saliva, and semen respectively (Figures 42D, 43D, and 44D). Pearson's correlation was also performed and showed the strongest correlation in saliva with  $R^2 = -0.96$  (Table 22). Therefore, a simple regression analysis was performed using RQ of HIF1A/VEGFA against the actual age for blood, saliva, and semen. Optimal curves for HIF1A/VEGFA were linear for blood and saliva ( $R^2 = 0.70$  and  $0.92$ , respectively), and quadratic in semen ( $R^2 = 0.54$ ) (Figure 49. A, B, and C).

As a result of the best fitting regression lines, three equations were derived as predictors of the age of blood, saliva, and semen samples, as presented in Table 22. The prediction age was calculated for all samples using the HIF1A/VEGFA model (Appendix 1. Table 36) with data plotted against the actual age to show the fit of the model (Figure 49. D, E, and F). The best scatterplot was obtained for saliva samples, with data points tightly clustered around the line. For blood and semen samples, 21 and 28-day time points had most of the data points located within the confidence interval region. For a clearer representation of the data, mean predicted age was calculated and plotted against the actual age (Figure 50. A, B, and C).

Table 22. The summary of all models obtained by single regression analysis in blood, saliva, and semen samples using SYBR Green chemistry.

Type of sample	Time frame	Number of samples	Pearson's correlation(r)	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Month	40	-0.84**	HIFA1	VEGFA	Linear	$y=23.66 - 23.45x$	0.70	±4.2
Saliva	Month	40	0.96**	HIFA1	VEGFA	Linear	$y=28.23 - 28.21x$	0.92	±2.1
Semen	Month	30	-0.65**	HIFA1	VEGFA	Quadratic	$y=17.07+22.15x-38.54x^2$	0.54	±5.0

Note: (\*\*) represented  $p < 0.01$

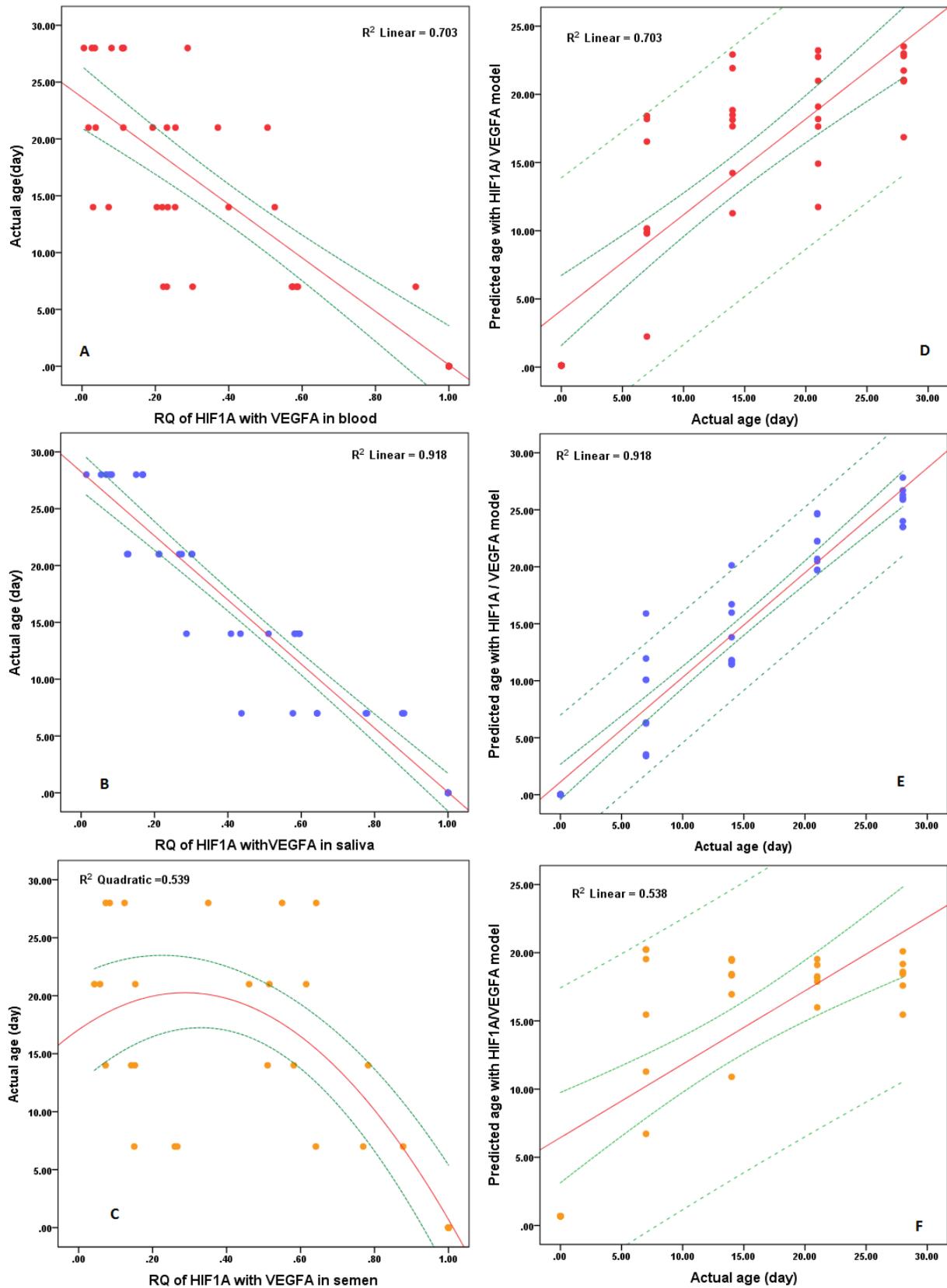


Figure 49. Simple regression analysis of HIF1A with VEGFA as reference gene in (A) blood, (B) saliva, (C) semen. (D-F) age prediction using previous models in blood, saliva, and semen, respectively. Solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

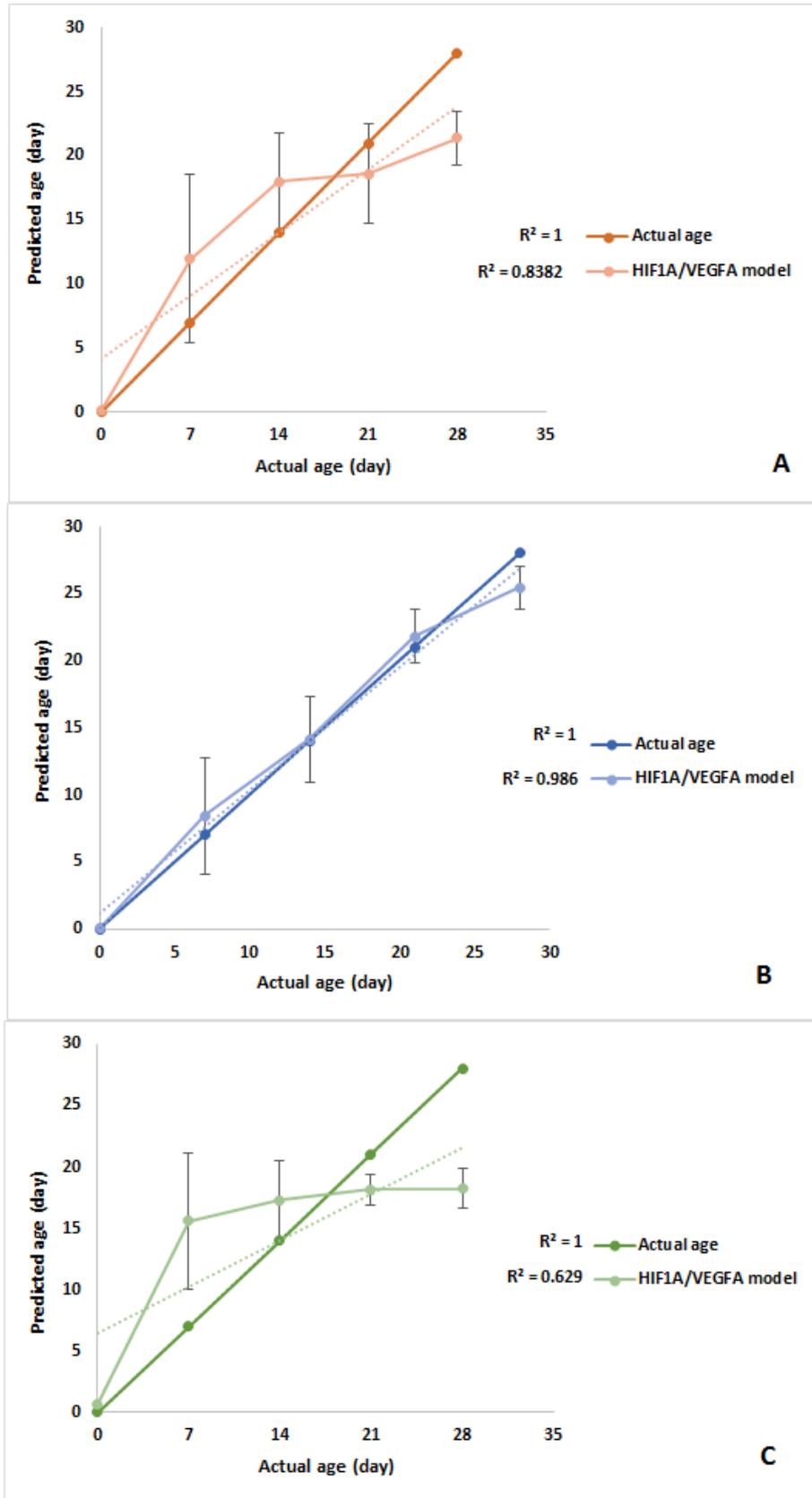


Figure 50. Age prediction using HIF1A/VEGFA model in (A) blood, (B) saliva, and (C) semen. Data presented as Mean  $\pm$  SD.

The HIF1A/VEGFA model was the most stable in saliva, with the highest  $R^2 = 0.99$ , and lowest MAD =  $\pm 2.1$  (Figure 50 B). Overall, HIF1A/VEGFA showed a good fit in blood, saliva, and semen samples, and this was particularly true for time zero.

Tables 20, 21, and 22 showed all models generated with hypoxia markers using SYBR Green chemistry in blood, saliva, and semen samples. The highest Pearson's correlation was detected with HIF1A/VEGFA model in saliva, whereas the lowest was detected with the same model in semen. According to  $R^2$  value and MAD in the present experiment, it seems that HIF1A/VEGFA is an appropriate model to predict the age of blood saliva, and semen, with the best fit emerging in saliva samples.

Further, multiple regression analysis was employed in an attempt to increase the fit of the model using markers in combination as opposed to each individually. Here, the HIF1A/VEGFA model gave comparable values over time, suggesting that it has no contribution to age prediction (blood and saliva), and was subsequently excluded. This left VEGFA/ACTB and HIF1A/ACTB remaining for multiple regression analysis in an attempt to improve the previously obtained model.

Figure 51 shows the strongest  $R^2$  value for saliva (0.82), and the lowest achieved in semen (0.63). Overall, multiple regression analysis improved prediction fit, and lowered MAD value, and this was particularly true for both blood and semen samples. Data points were positioned close to each other, with tighter CI than in single regression analysis. The equations that described the models are presented in Table 23.

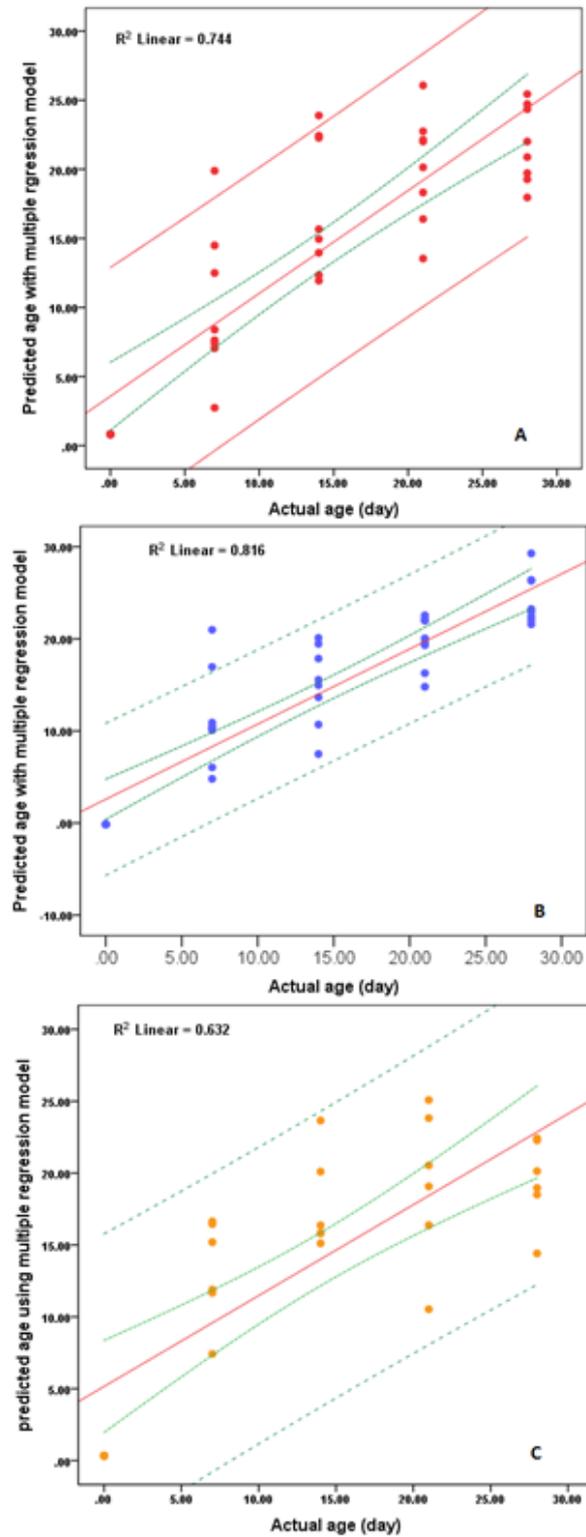


Figure 51. Multiple regression analysis using hypoxia models. (A-B) Age prediction by using VEGFA/ACTB and HIF1A/ACTB models in blood and saliva, respectively. (C) Age prediction using HIF1A/VEGFA and HIF1A/ACTB in semen. Solid red line in the graph represents the modelled space where predicted age and actual age are equal. Small dashed green lines indicate confidence limits. Large dashed green lines indicate prediction limits.

Table 23. Multiple regression for models in blood, semen, and saliva samples using SYBR Green chemistry.

Type of sample	Frame time	Number of samples	Models using	X1	X2	Estimated curve equation	R <sup>2</sup>	MAD (day)
Blood	Month	40	VEGFA/ACTB**	VEGFA/ACTB	HIF1A/ACTB	y=26.794-15.606(X1) – 10.373(X2)	0.74	±3.7
			HIF1A/ACTB**					
Saliva	Month	40	VEGFA/ACTB**	VEGFA/ACTB	HIF1A/ACTB	y=29.992-19.171(X1) -10.973 (X2)	0.82	±3.0
			HIF1A/ACTB**					
Semen	Month	30	HIF1A/VEGFA**	HIF1A/VEGFA	HIF1A/ACTB	y=28.790-11.777 (X1) -16.671 (X2)	0.63	±4.6
			HIF1A/ACTB**					

Note: \* represented of p<0.05 and \*\* represented that p<0.01

It is clear that the best multiple regression model was detected in saliva. Prediction age was calculated for blood, saliva, and semen using the equations derived (Appendix 1. Table 36), and the data were plotted against the actual age. Improved  $R^2$  values were obtained when compared to each model individually (Figure 51. A, B, and C). Mean age prediction was also calculated and plotted against the actual age, and there was an improvement in prediction, particularly for 14 and 21-day data points in blood and semen samples. Finally, the correlation between the predicted age and the actual age was higher than in single regression analysis, particularly for blood and semen where it increased from 0.84 and 0.63 to 0.91 and 0.66, respectively (Figures 50 and 52).

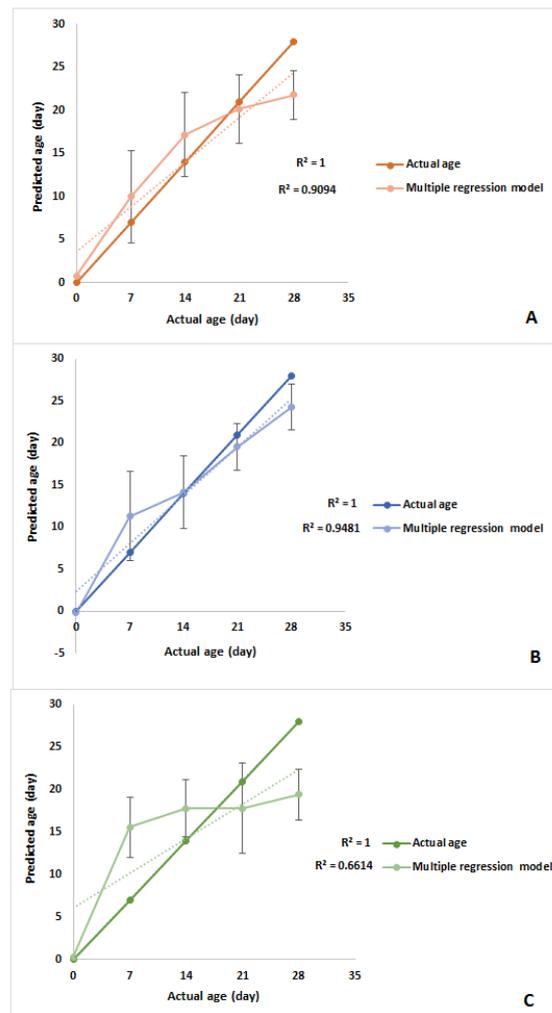


Figure 52. Age prediction derived from multiple regression analysis. (A-C) Age prediction in blood, saliva, and semen, respectively. Data presented as Mean  $\pm$  SD.

## 5.4. Blind samples with SYBR Green

### 5.4.1. Blind blood samples

Similar to the experiment performed with TaqMan markers in blood samples, SYBR Green was used to investigate the age of the blind blood target using the HBB/ACTB model. Based on the obtained RQ it was possible to put the samples in the correct order based on their age (Figure 53). Age prediction was calculated, and the results listed in Table 24.

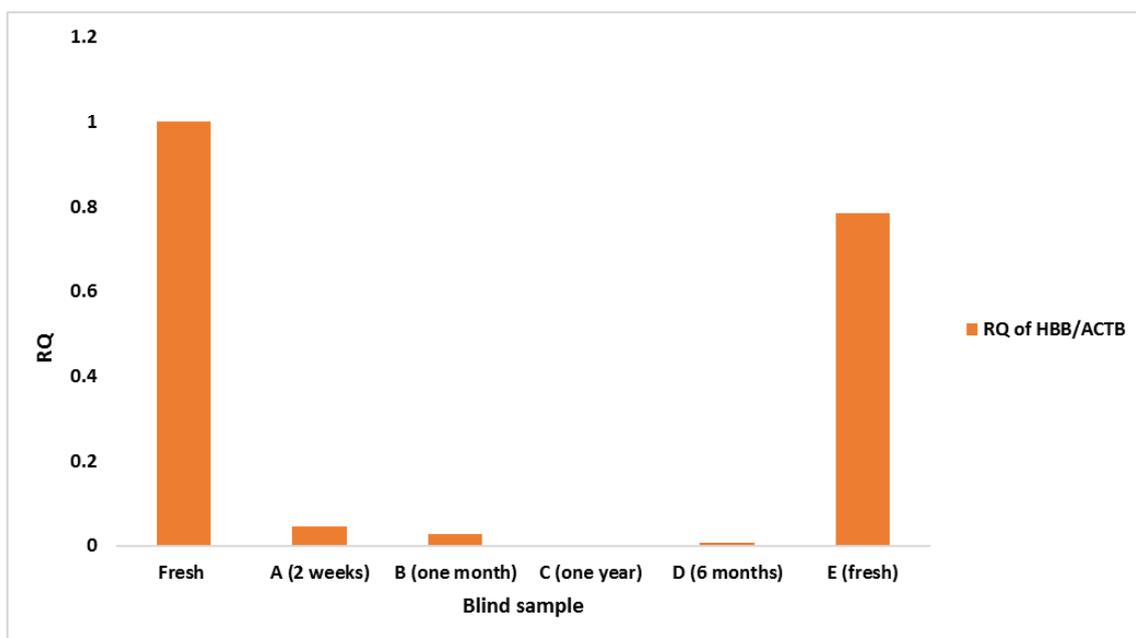


Figure 53. Relative quantification obtained by using HBB/ACTB model in blind blood sample with SYBR Green chemistry.

Table 24. Age prediction of blind blood samples using HBB/ACTB model.

Sample	Actual age	HBB/ACTB model
A	Two weeks	26.12 $\pm$ 3.5
B	One month	26.54 $\pm$ 3.5
C	One year	27.84 $\pm$ 3.5
D	Six months	27.87 $\pm$ 3.5
E	Fresh	07.27 $\pm$ 3.5

The next step was to compare the hypoxia models obtained with blind blood samples from fresh to 28 days old. After obtaining RQ of markers under investigation, the age of the unknown samples was calculated using previously generated models. Cq values were obtained in an experiment and used to calculate RQ of hypoxia markers. RQ value was used as a parameter for age prediction calculation (Appendix 2. Table 39) and the results presented in Table 25. Similarly, as before, the prediction of recent age (less than one month) was more accurate with these models than aged samples (one year old). All models gave acceptable predictions (Table 25). For fresh age prediction the best prediction was generated with a multiple regression model with only 1.8 day over the actual age. Figure 54 shows prediction age generated with hypoxia marker plotted against the actual age. It is clearly demonstrated that the multiple regression model gave the best fit ( $R^2 = 0.88$ ), with accurate age prediction, especially at 14 and 21 days. Other models also produced good  $R^2$  values, with HIF1A/ACTB closely predicting samples age at 21 day, while this was true for fresh and one -week old samples using the HIF1A/VEGFA model (Figure 54. A, B, and C). This is consistent with previously reported strongest  $R^2$  value with VEGFA/ACTB and multiple regression models.

#### **5.4.2. Blind saliva sample**

To test the hypoxia models, blind saliva samples were also investigated for a period of time ranging between time zero and 28 days. Using the same approach, RQ was calculated and used for age prediction with each model (Appendix 2. Table 40), with the data presented in Table 26. The HIF1A/VEGFA model was able to accurately predict the age of the fresh sample and gave the best age prediction at 28 days. In contrast, for one-week old samples, this model was the least accurate with over 5 days in difference. The mean age prediction was also plotted with actual age, and the results demonstrated that the HIF1A/VEGFA and multiple regression analysis models gave the best fit ( $R^2 = 0.90$  and  $0.96$ , respectively) (Figure 55. C and D).

Table 25. Prediction of blood samples using oxygen regulated factors models.

	<b>VEGFA/ACTB</b>	<b>Difference between ages</b>	<b>HIF1A/ACTB</b>	<b>Difference between ages</b>	<b>HIF1A/VEGFA</b>	<b>Difference between ages</b>	<b>Multiple regression model</b>	<b>Difference between ages</b>
Fresh (0day)	2.04 ±4.4	2.04	4.39 ±4.6	4.39	2 ±4.2	2	1.8 ±3.7	1.8
One week (7days)	5.93 ±4.4	-1.07	14.39 ±4.6	7.39	4.54 ±4.2	-2.46	5.59 ±3.7	-1.41
Two weeks (14days)	14.91 ±4.4	0.91	18.24±4.6	4.24	8.24 ±4.2	-5.76	14.74 ±3.7	0.74
Three weeks (21days)	20.03 ±4.4	-0.98	19.89 ±4.6	-1.11	7.58 ±4.2	-13.42	21.96 ±3.7	0.96
Four weeks(28days)	17.13 ±4.4	-10.87	20.75 ±4.6	-7.25	17.61 ±4.2	-10.39	19.51 ±3.7	-8.49

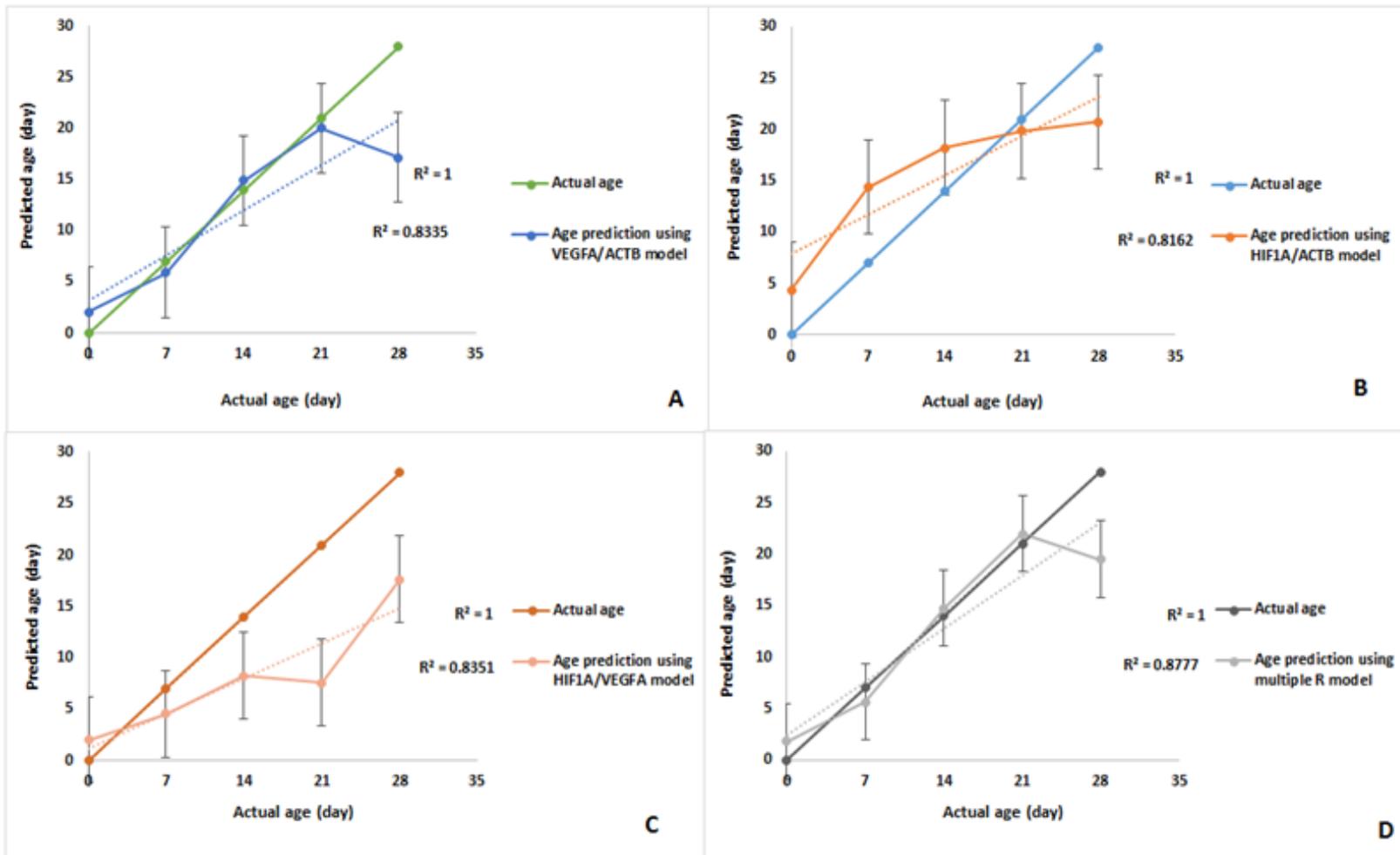


Figure 54. The blind blood samples using hypoxia models. (A) VEGFA/ACTB model (B) HIF1A/ACTB model. (C) HIF1A/VEGFA model. (D) Multiple regression model. Error bars represent MAD.

Table 25. Age prediction for blind saliva samples using hypoxia models.

		<b>Difference</b>		<b>Difference</b>		<b>Difference</b>	<b>Multiple</b>	<b>Difference</b>
	<b>VEGFA/ACTB</b>	<b>between ages</b>	<b>HIF1A/ACTB</b>	<b>between ages</b>	<b>HIF1A/VEGFA</b>	<b>between ages</b>	<b>regression model</b>	<b>between ages</b>
Fresh (0day)	03.62 ±3.9	03.62	05.55 ±4.9	05.55	0.6 ±2.1	00.6	03.03 ±3.0	03.03
One week (7days)	09.98 ±3.9	02.98	10.56 ±4.9	03.56	1.17 ±2.1	-05.83	09.08 ±3.0	02.08
Two weeks (14days)	12.14 ±3.9	-01.86	07.86 ±4.9	-06.14	14.42 ±2.1	00.42	09.50 ±3.0	-04.50
Three weeks (21days)	12.83 ±3.9	-12.73	22.33 ±4.9	01.33	24.92 ±2.1	03.92	17.09 ±3.0	-03.91
Four weeks(28days)	23.00 ±3.9	-05.00	22.11 ±4.9	-05.89	24.34 ±2.1	-03.66	24.28 ±3.0	-03.72

The other two models used were also reasonably good at predicting the age of the samples, with VEGFA/ACTB being the closest at 14 days, and HIF1A/ACTB best at predicting 21-day old saliva samples (Figure 55. A and B). Overall, HIF1A/VEGFA and multiple regression models were the best models to predict the age of the saliva stains. The same model emerged as the best fit for previously reported samples of the known age. The multiple regression model in blind blood and saliva samples obtained the best result, further confirming the advantage of using multiple variables instead of a single variable.

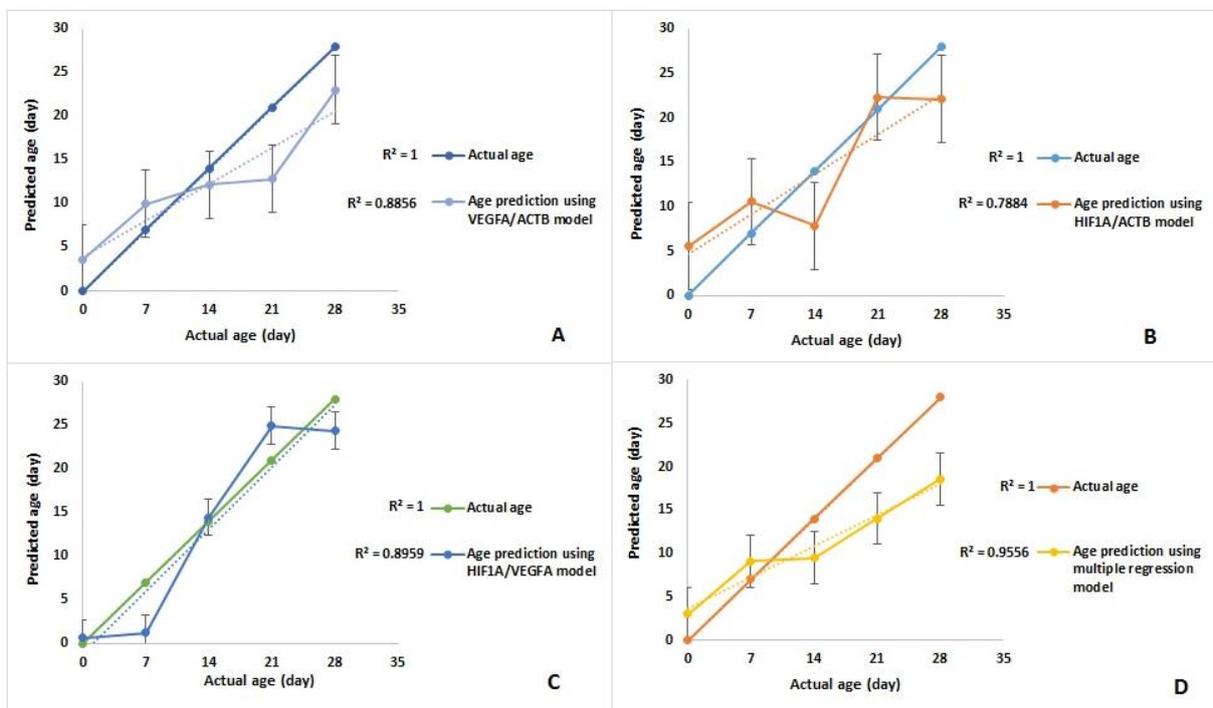


Figure 55. Saliva blind sample age prediction. (A) VEGFA/ACTB model (B) HIF1A/ACTB model (C) HIF1A/VEGFA model (D) Multiple regression model. Error bars present mean absolute deviation (MAD).

### 5.4.3. Blind semen samples

Hypoxia models for blind semen samples were also investigated, and RQ of VEGFA/ACTB, HIF1A/ACTB, and HIF1A/VEGFA were generated and used to predict the unknown age for semen. Age prediction was calculated (Appendix 2. Table 42) with the help of these models and the data presented in Table 27. Good prediction for a fresh semen sample was observed with the VEGFA/ACTB model (1.69 days), whereas the multiple regression model was the

most accurate at 28 days (with a value of 26.8 days). Further, VEGFA/ACTB gave accurate prediction at 14 and 21 days, while this was true for HIF1A/VEGFA at day 21. Figure 56 shows predicted age plotted against the actual age, demonstrating that the VEGFA/ACTB model has the strongest  $R^2$  value (0.89), and the HIF1A/ACTB model the weakest fit ( $R^2 = 0.50$ ).

For easier comparison of all data points described above, age prediction for all blind samples was plotted against actual age (Figure 57). In saliva samples, age prediction becomes more accurate with an increased in the real age, which was especially obvious for the oldest sample used (28-day sample). The best age prediction at time zero was observed in saliva samples with the HIF1A/VEGFA model, while the closest prediction at day 7 was generated with the VEGFA/ACTB model in blood. At day 14, an appropriate age prediction obtained with HIF1A/VEGFA in saliva, and the optimal age prediction at day 21 was seen in semen with multiple regression analysis. Overall, across 5 different time points, the best prediction was obtained with the multiple regression model which accurately predicted age at 2 of the time points (out of total of 5).

Table 26. Semen age prediction using SYBR Green models.

	<b>VEGFA/ACTB</b>	<b>Difference between ages</b>	<b>HIF1A/ACTB</b>	<b>Difference between ages</b>	<b>HIF1A/VEGFA</b>	<b>Difference between ages</b>	<b>Multiple regression model</b>	<b>Difference between ages</b>
Fresh (0day)	1.69 ±5.4	01.69	03.74 ±5.2	03.74	04.82 ±5	04.82	02.07 ±4.6	02.07
One week (7days)	13.04 ±5.4	06.04	18.81 ±5.2	11.81	09.40 ±5	02.41	12.97 ±4.6	05.97
Two weeks (14days)	14.52 ±5.4	00.52	20.67±5.2	06.67	18.14 ±5	04.15	19.29 ±4.6	19.29
Three weeks (21days)	20.59 ±5.4	-00.40	16.14 ±5.2	-04.86	20.05 ±5	-00.95	17.51 ±4.6	-03.49
Four weeks(28days)	21.67 ±5.4	-06.33	21.18 ±5.2	-06.82	17.23 ±5	-10.77	26.80 ±4.6	-01.20

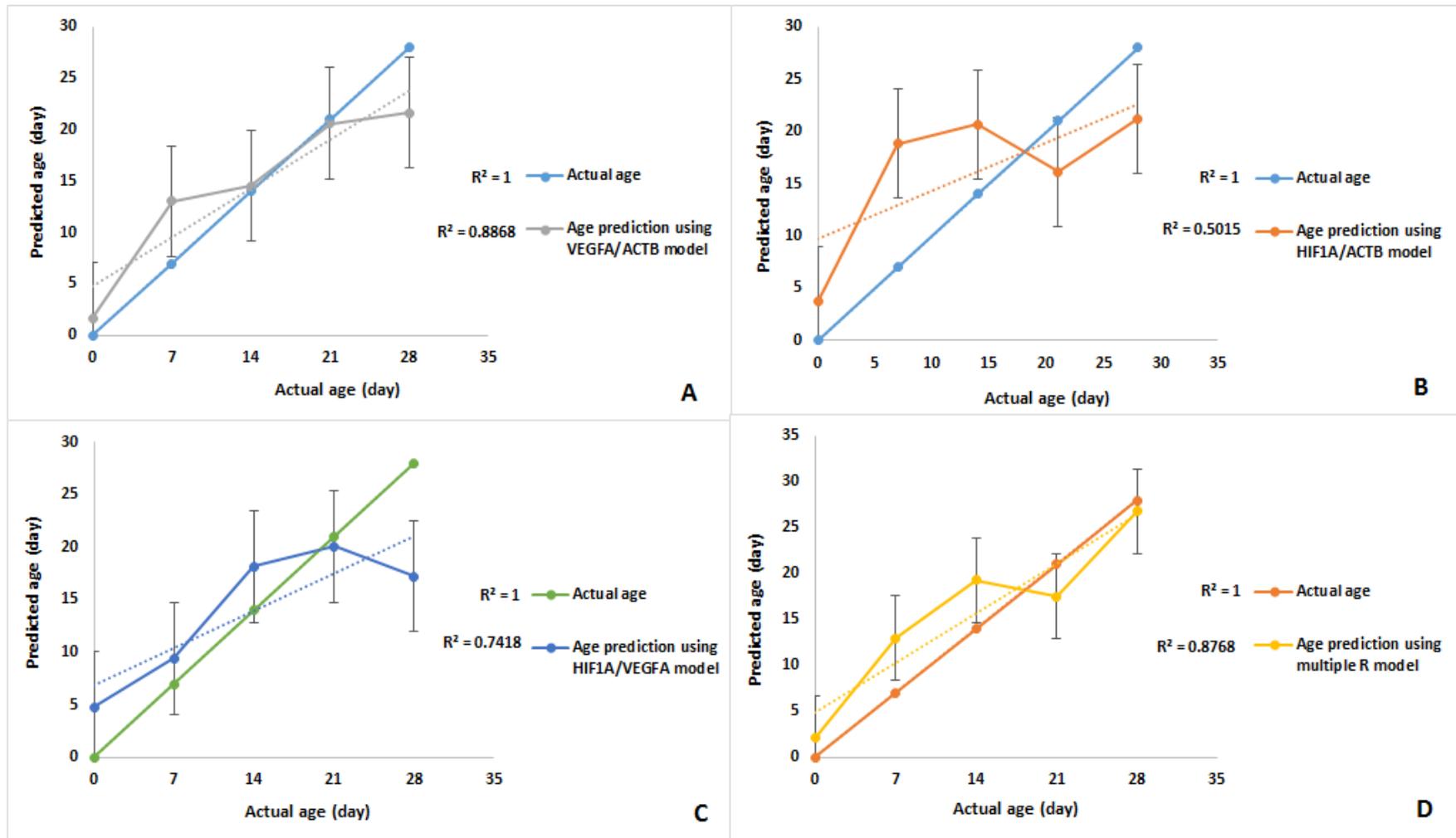


Figure 56. Blind semen samples using models obtained from hypoxia markers. Errors bars resent mean absolute deviation (MAD).

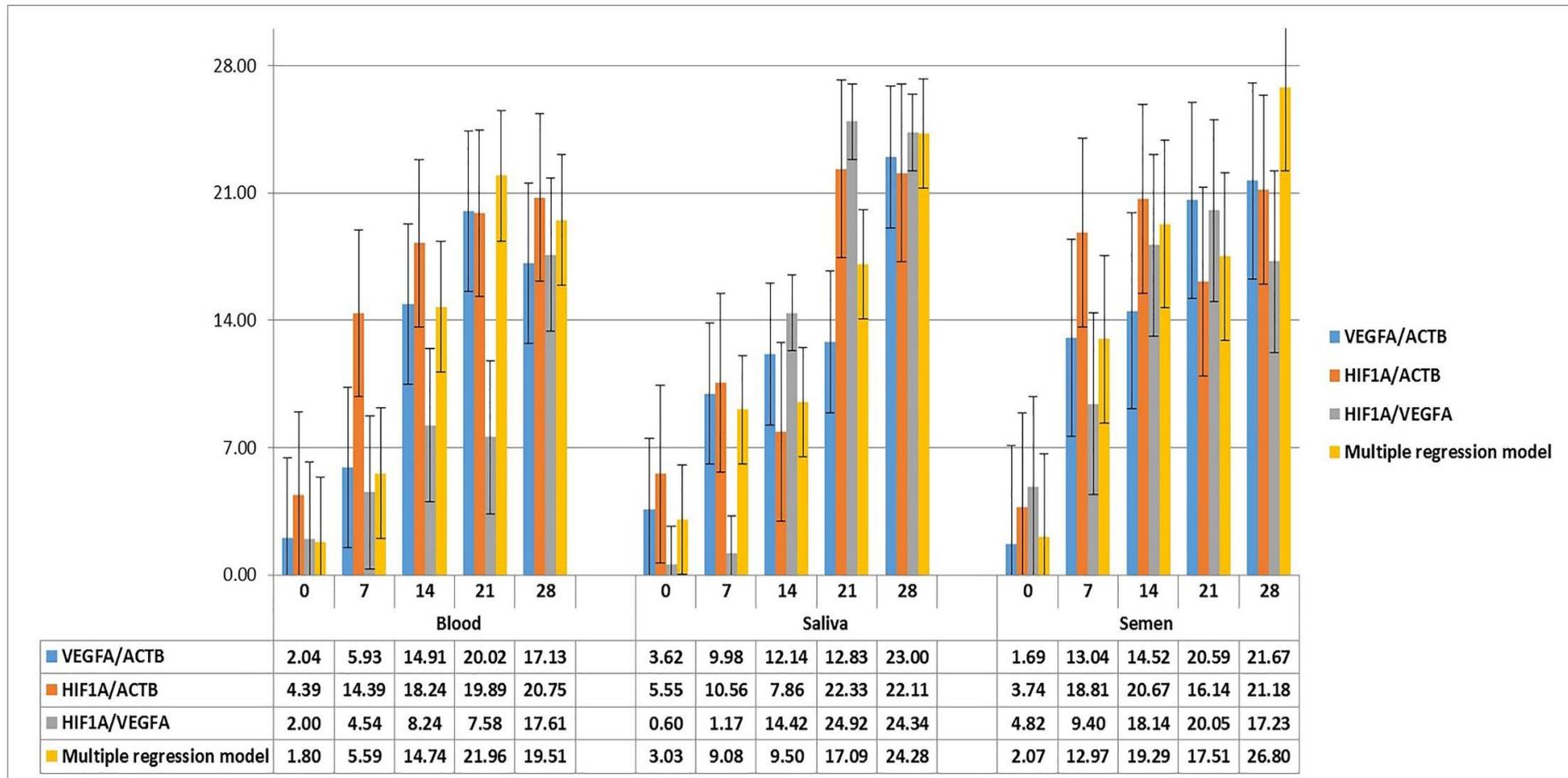


Figure 57. Age prediction using hypoxia models in blind blood, saliva, and semen samples. Errors bars represent mean absolute deviation (MAD) (N= 15).

## 5.5. Discussion

Since its development in 1985 by Sir Alec Jeffreys. DNA profiling has revolutionised the way in which criminal investigations are conducted. DNA profiling allow the identification of the individual or individuals who deposit the body fluid or other biological material at the crime scenes (Jeffreys, 1987). However, there is a limit to its application within the forensic investigations: DNA profiling can only identify the donor, it cannot confirm the source of DNA and, till now, the time of deposition.

RNA is less stable than DNA, and as such, through the measurement of its degradation patterns, it stands out as a potential candidate for determining the age of biological stains (Sampaio-Silva *et al.*, 2013, Bauer *et al.*, 2003a). The focus of this study was the degradation of primers over time, when compared to time zero. Therefore, the aim of this chapter was to explore the degradation pattern of chosen RNA markers at a one-month period, using the SYBR Green chemistry method. This was achieved by using the RQ method to obtain models that could be useful in forensic investigation to predict the unknown age of blood, saliva, and semen stains. A correlation value was obtained for selection markers as an indicator of the validity of the chosen markers to be used in the RQ method. Once all RQ of chosen markers were obtained, linear regression analysis was employed. The data then underwent simple and multiple regression analysis, using standard confidence and prediction intervals of 95%, in order to develop the basis for predictive modelling. Linear regression analysis revealed that certain primers could be used to predict the age of unknown stains. RQ and  $\Delta Cq$  of selected markers were established as indicators to predict the age of samples under investigation.

Under normal physiological conditions, cells in the adult human body are provided with an adequate supply of oxygen. On the other hand, oxygen becomes sharply depleted during fatal processes. Recent advances in molecular biology suggest that HIF1, a transcription factor that

responds to low oxygen concentrations, together with EPO and VEGF, increases vascularisation of hypoxic areas in an attempt to maintain oxygen homeostasis (Paulding and Czyzyk-Krzeska, 2002). It is not surprising, therefore, that the cellular response to hypoxic and ischemic conditions is to increase the concentrations of these factors on both mRNA and protein levels (Bernaudin *et al.*, 2002). Therefore, the aim was to explore whether hypoxia sensitive biomarkers could be viable candidates for stain age prediction based upon the change in oxygen concentrations in the surrounding environment (i.e. in the blood compared to the atmosphere).

Many studies have been conducted using oxygen-regulated factors, such as EPO, VEGF and HIF1A mRNAs, to estimate PMI (Thaik-Oo *et al.*, 2002a, Zhao *et al.*, 2006). In this chapter, hypoxia condition was investigated in order to assess any changes or trends in blood, saliva, and semen samples taken at zero days, 7 days, 14 days, 21 days, and 28 days. For hypoxia markers in blood, saliva, and semen, the data were analysed in three ways, targeting VEGFA and HIF1A with ACTB as the reference gene, and then targeting HIF1A with VEGFA as the reference gene. In addition, multiple regression model was also tested. Two parameters were used to determine the model with the best fit,  $R^2$  value, and MAD. The result indicates that the rate of degradation of hypoxia gene could be used as a good indicator for ageing blood, saliva, and semen, because this rate was reported to increase with time in storage *in vitro* (Almac *et al.*, 2014). A strong correlation was detected in saliva using the HIF1A/VEGFA model. While no trend could be observed when considering each marker on its own, a linear trend with a correlation value of 0.94 and 0.92 was observed in saliva with VEGFA and HIF1A, respectively, when ACTB was used as a reference gene. Finally, a correlation value of 0.99 was calculated for HIF1A as a target when VEGFA was a reference gene, thus indicating a linear decrease in marker expression over a one-month period (Figure 43).

Linear correlations were also observed in samples taken at the same intervals with, 0.93, 0.87 and 0.84 in blood and 0.66, 0.77, and 0.73 in semen of VEGFA/ACTB, HIF1A/ACTB, and HIF1A/VEGFA respectively. Higher linear correlation values in saliva samples could be due to it being a less dynamic fluid when compared to blood and semen samples. More ongoing intracellular processes that occur in blood and semen could introduce a wider range of variables, thereby making any decomposition processes less linear. All prediction models were developed with hypoxia markers for blood, saliva and semen, with the lowest MAD = 2.1 days for saliva detected with HIF1A/VEGFA model, and MAD = 4.6 days and MAD = 3.7 days with multiple regression model for semen and blood samples, respectively. Given the lower linear correlation value for the blood samples, it was expected that the MAD for blood and semen would be greater than that for saliva.

Hypoxia models were also tested in blind samples up to one-month old. The result showed the strongest  $R^2$  value of 0.88, 0.96, and 0.88 when using the multiple regression model in blood, saliva, and semen, respectively (Figures 54D, 55D, and 56D). Consequently, it has been demonstrated that hypoxia sensitive biomarkers could be used for predicting stain age up to one month, at least in saliva and semen stains, and to a lesser extent in blood stains. There is a substantial margin of error, however, and any calculated day could be within a time- frame ranging from ~2 days through to ~5 days. However, whilst this time frame is rather substantial, it is an improvement when compared to the existing capabilities, whereby the difference between fresh and one-month old stains cannot be determined. Therefore, the present study indicates the possibility of using relative quantification when it is necessary to avoid any externally influencing factors. Quality and integrity of RNA play a critical role in performing accurate study and the selection of an adequate reference gene is one of the most important aspects. An imprecise RNA measurement after extraction could also lead to erroneous results, because RNA levels *ex-vivo* are more unpredictable. Further, the

differences in the quality and integrity of the extracted mRNA between the samples, as well as the performance of reverse transcription reaction and PCR, and the normalisation with the reference gene can all influence the accuracy of the results (Huggett *et al.*, 2005). Finally, environmental factors can exert an influence on the samples, with oxygen from the air having a particularly potent effect through its denaturation of ribonucleases, the key proteins responsible for the oxidative damage on RNA (Lund *et al.*, 2011). It is possible for this to affect the poor correlation between RNA degradation and PMI previously reported (Heinrich *et al.*, 2007, Zubakov *et al.*, 2008).

## **5.6. Conclusion**

Stain age prediction models have been developed for blood, saliva, and semen stains up to one-month old. Despite the limitations described above, the proof of principle has been demonstrated, and it was shown that some markers, especially hypoxia sensitive biomarkers, could be used to predict stain age. The capability of differentiating between fresh stains and one-month old stains has also been demonstrated on selection samples. Further work should focus on identifying further markers, exploring longer time frames, and efforts to reduce the margin of error/mean absolute deviation.

**Chapter Six: Exploring the degradation pattern of  
miRNA in biological stains to estimate time since  
deposition**

## 6.1. Introduction

RNA forensic analysis has been used for many purposes, including identifying body fluids, estimating the age of biological stains, and determining PMI by focusing on RNA decay analysis (Bauer *et al.*, 2003a, Bauer and Patzelt, 2003). MicroRNA (miRNA) is short in length and is considered relatively stable compared to mRNA. Silva, *et al.* (2012), demonstrated that PMI evaluation based only on mRNA decay is limited by several physical and chemical factors, occurring at the time of death, or during PMI. It was suggested that miRNA profiling is the preferred method compared to mRNA quantification methods because it is more stable and sensitive. Furthermore, the potential of miRNA for PMI estimation is relatively unknown (Silva, 2012).

As the similarity between PMI and age estimation of biological stains, this chapter aims to explore the potential of miRNA in estimating the age of blood, saliva, and semen, as well as its stability as optimal reference genes. It is noted that the stability of reference genes has been investigated in many previous studies, focusing on specific tissues in several species (de Kok *et al.*, 2005, Ohl *et al.*, 2005, Silver *et al.*, 2006, Zhang *et al.*, 2005). These previous studies demonstrated the difficulty in finding a ‘universal’ reference gene that has a stable expression in all cell types and tissues, as well as remaining stable at different points of time, in varying experiments (Nygard *et al.*, 2007). Here, a combination of the most degraded elements (mRNA) and the most stable elements (miRNA) could be useful in generating models of age prediction of blood, saliva, and semen. Currently, no study has been conducted using geometric mean normalisation, with a combination of mRNA and miRNA for age prediction. A similar study explored the level of 18S rRNA and miRNA to estimate PMI up to 7 days (Li *et al.*, 2014).

## 6.2. Experimental design

Sixty samples based on availability were investigated, targeting miRNA markers 451, 205 and 891a for blood, saliva, and semen, respectively. The RNU 44 marker was also employed, together with three other markers as reference genes. Information related to these markers is listed in chapter Two (Table 4). The blood, saliva, and semen samples underwent total RNA extractions, DNA digestions, cDNA synthesis and triplicated qPCR, as described in chapter Two. Negative controls were also included in this study. The data were statistically processed using SPSS analysis software (v22). The normality was tested with the Shapiro and Wilk test (Shapiro and Wilk, 1965). The level of significance was set at  $p < 0.05$ , and an acceptable Pearson's correlation value was used, as previously described (Mukaka, 2012). The data were plotted using SPSS and Excel 2013. Data were normalised using a single reference gene, and geometric mean normalisation with multiple reference genes. The formula used for geometric mean normalisation was described by Vandesompele, *et al.* (2002).

## 6.3. Results

Ideally, an accurate age estimation of any biological sample requires the evaluation of parameters that potentially correlate with a deposition time. Furthermore, the normalisation into a stable reference gene is also an important factor. Here, miRNA analysis was employed to identify the stability of the chosen markers, over a period of time. The results showed all assays were successfully detected up to 28 days; this was expected as all miRNA markers were adopted from other research groups. Hanson, *et al.* (2009) reported that miRNA 451 is stable in dried bloodstains ranging from 15 months to 84 years old, whereas, miRNA 205 is stable in aged saliva samples from 26 to 58 years old (Hanson *et al.*, 2009).

Once all markers were detected, the  $\Delta Cq$  of each marker was calculated ( $Cq_{\max} - Cq_{\text{target}}$ ). In saliva, the  $\Delta Cq$  value indicated that miRNA 205 was more abundant than miRNA 44. This

unexpected result showed that both markers tend to decrease over all time points. Statistical analysis was performed (Shapiro and Wilk test followed by T-test) and revealed that the  $\Delta Cq$  of both markers was significantly lower than the calibrator sample (0 days), except at day 7 and day 28, for miRNA 205 and RNU 44, respectively (Figure 58A). The trend of both assays was investigated using  $\Delta Cq$  as an indicator with two relationships: linear and polynomial. This finding demonstrated that miRNA 205 has linear and polynomial curves, with  $R^2$  values of 0.93 and 0.97, respectively, whereas,  $\Delta Cq$  of RNU 44 has linear and polynomial curves, with  $R^2$  values of 0.73 and 0.91, respectively (Figures 58. B and C).

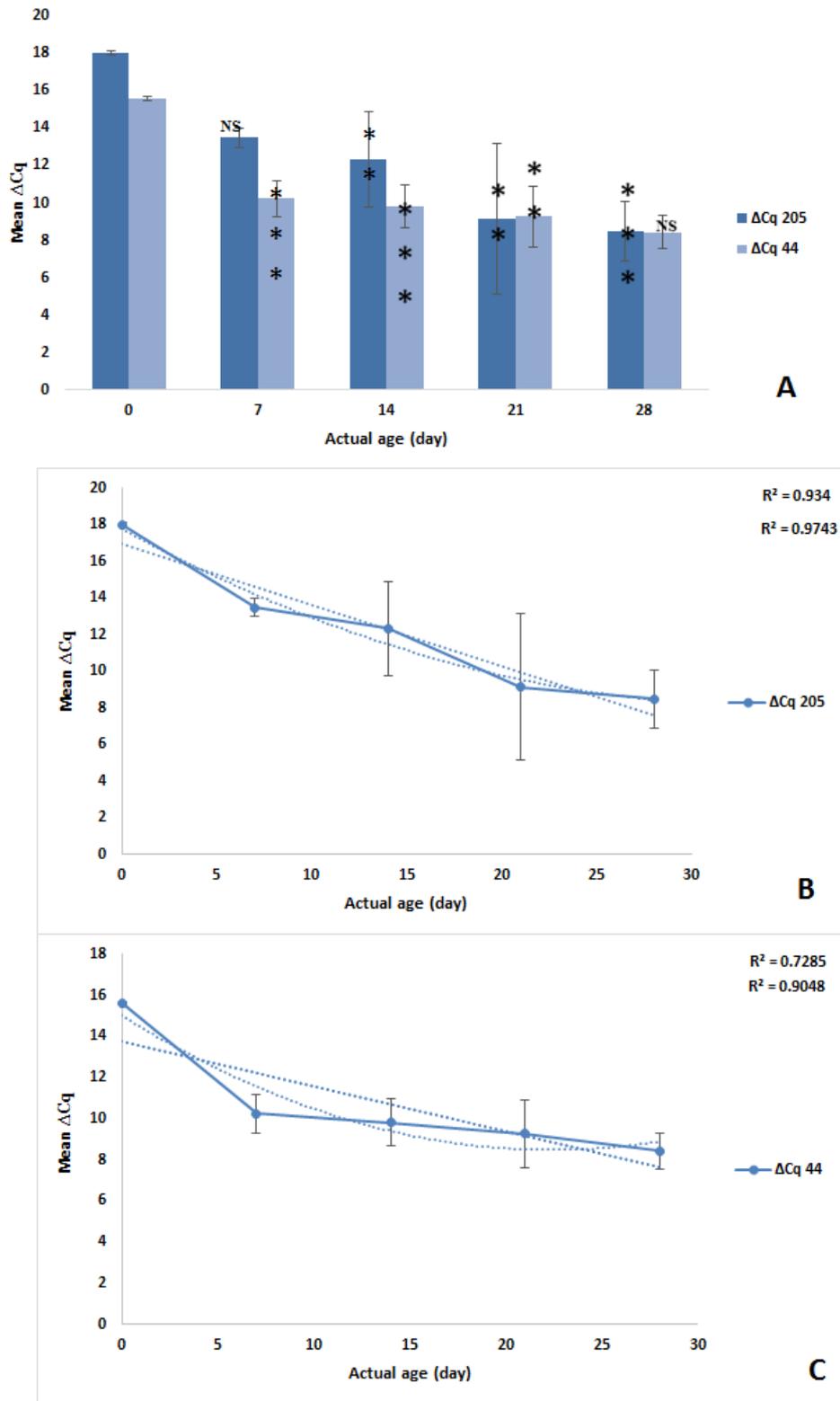


Figure 58. MiRNA in saliva samples. (A) Mean  $\Delta Cq$  of markers miRNA 205 and miRNA 44, up to 28 days, (B and C) the patterns of degradation of miRNA 205 and miRNA 44 in saliva, respectively. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=15).

In the blood samples,  $\Delta Cq$  was calculated in the same way as it was in the saliva samples. Whilst miRNA 451 was more abundant than the reference gene, no significant difference was detected in both markers, except in miRNA 451, at day 14 (Figure 59A). This finding suggests a minor variation was observed between fresh (0 days) and 14 days. The degradation patterns of miRNA in the blood samples was also investigated, and the results showed that miRNA 451 had the strongest linear and polynomial curves, with  $R^2$  values of 0.29 and 0.67, respectively. No correlation was observed with RNU 44 (Figures 59. B and C). Generally,  $\Delta Cq$  of blood markers showed a weak correlation when compared to the saliva markers. A similar study was conducted by Nakao, *et al.* (2013), and results showed that miRNA 451 and miRNA 16 significantly decreased over time. The authors of this study explained that this correlation was evident because of humidity rather than the increasing the age of samples (Nakao *et al.*, 2013).

Data relating to the semen samples are illustrated in Figures 60 and 61. Here, RNU 44 and miRNA 891a markers were detected in all semen samples, aged up to 28 days. As previously mentioned, two extraction methods were tested in the semen samples: with DTT (+DTT) and without DTT (-DTT). To show the influence of the DTT reagent on miRNA markers, the  $\Delta Cq$  of semen markers was calculated ( $Ct_{\max} - Ct_{\text{target}}$ ) in both experiments. The results indicate that DTT has a strong impact on RNA extraction because the  $\Delta Cq$  of both markers was high, in the absence of DTT (Figures 60A and 61A). In addition, miRNA 891a was more abundant than RNU 44 when DTT was added, and the opposite was observed when DTT was not included. Furthermore, both markers at all-time points were significantly differenced with a fresh sample when DTT was avoided. This finding suggests that miRNA 891a is internally abundant, or RNU 44 is strongly influenced by the DTT reagent. No clear explanation for this finding has been found in previous literature; however, it is possible that an extra amount of DTT works as an inhibitor with miRNA, resulting in a lower rate of amplification. A similar

explanation was reported by Haas *et al.* (2011b) in the Turbo DNA free kit which used for DNA digestion (Haas *et al.*, 2011b).

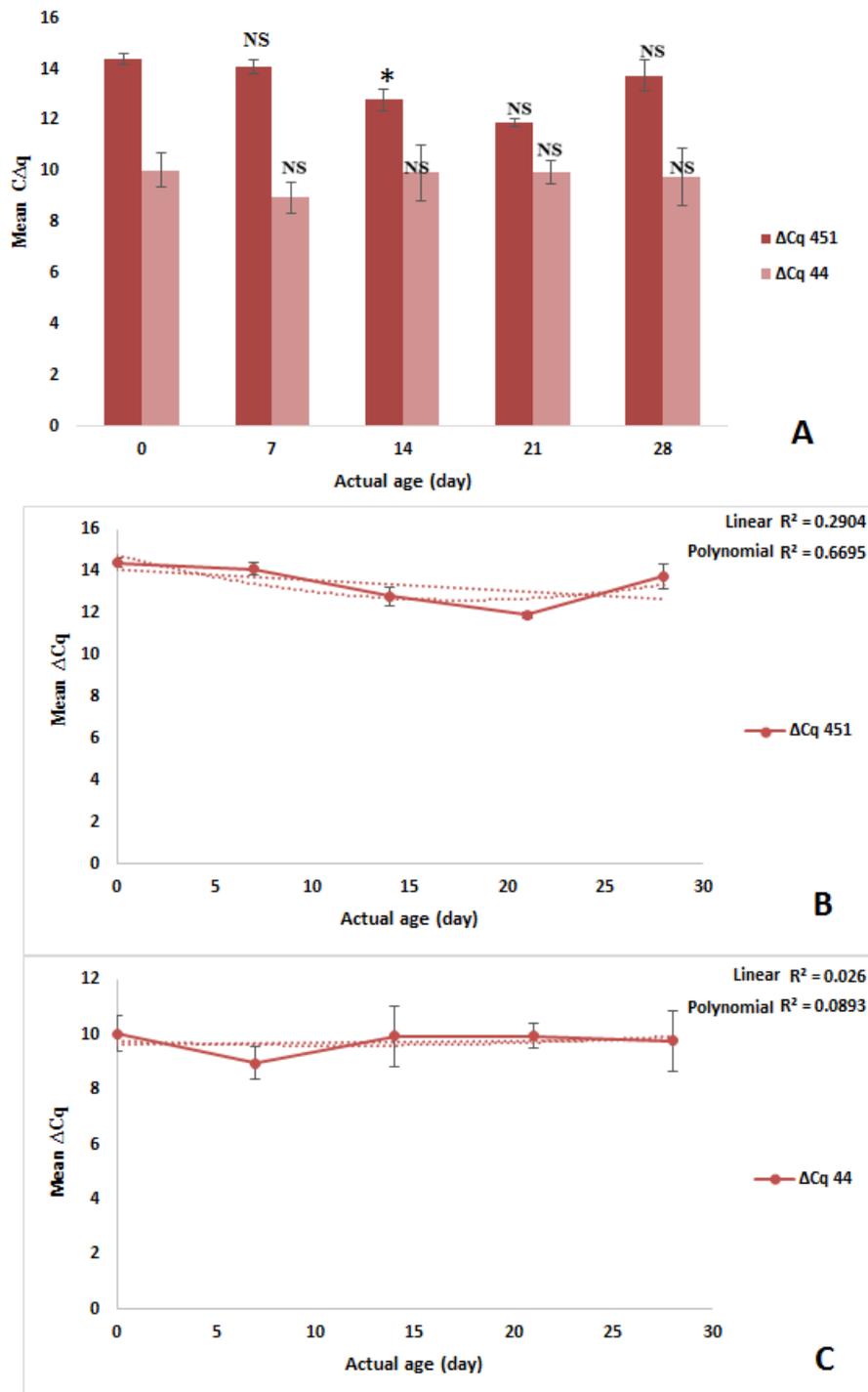
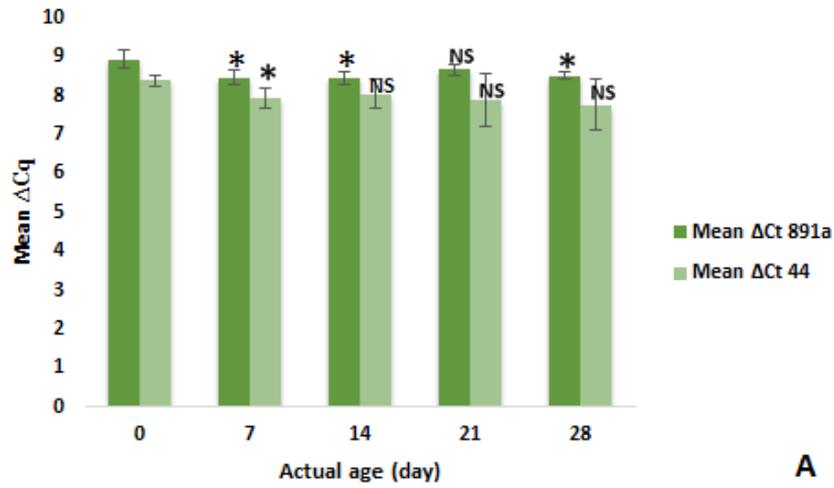
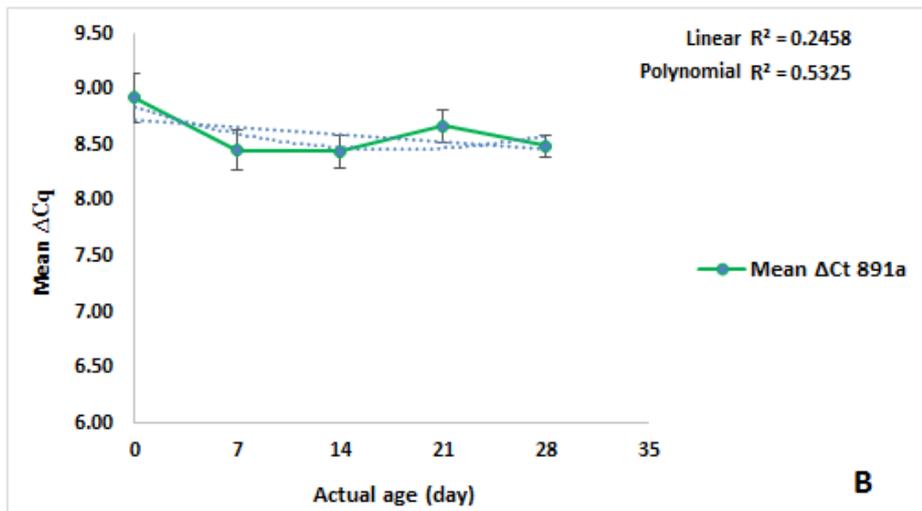


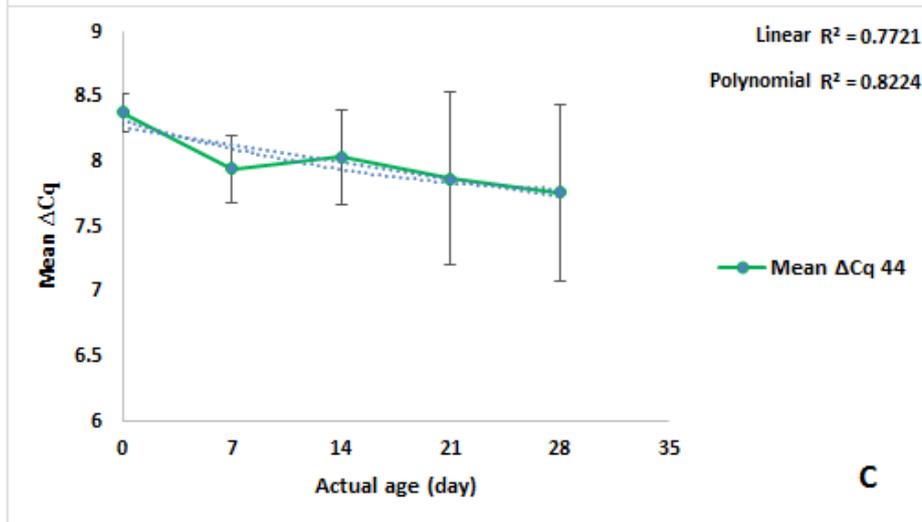
Figure 59. MiRNA in blood samples. (A) Mean  $\Delta Cq$  of miRNA 451 and 55, aged up to 28 days, (B and C) the degradation patterns of miRNA 451 and RNU 44 in blood, respectively. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=15).



**A**



**B**



**C**

Figure 60. MiRNA in semen samples with DTT (+DTT). (A) Mean  $\Delta Cq$  of miRNA 891a and miRNA 44, (B and C) the degradation pattern of miRNA 891a and miRNA 44, aged up to 28 days, respectively. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=15).

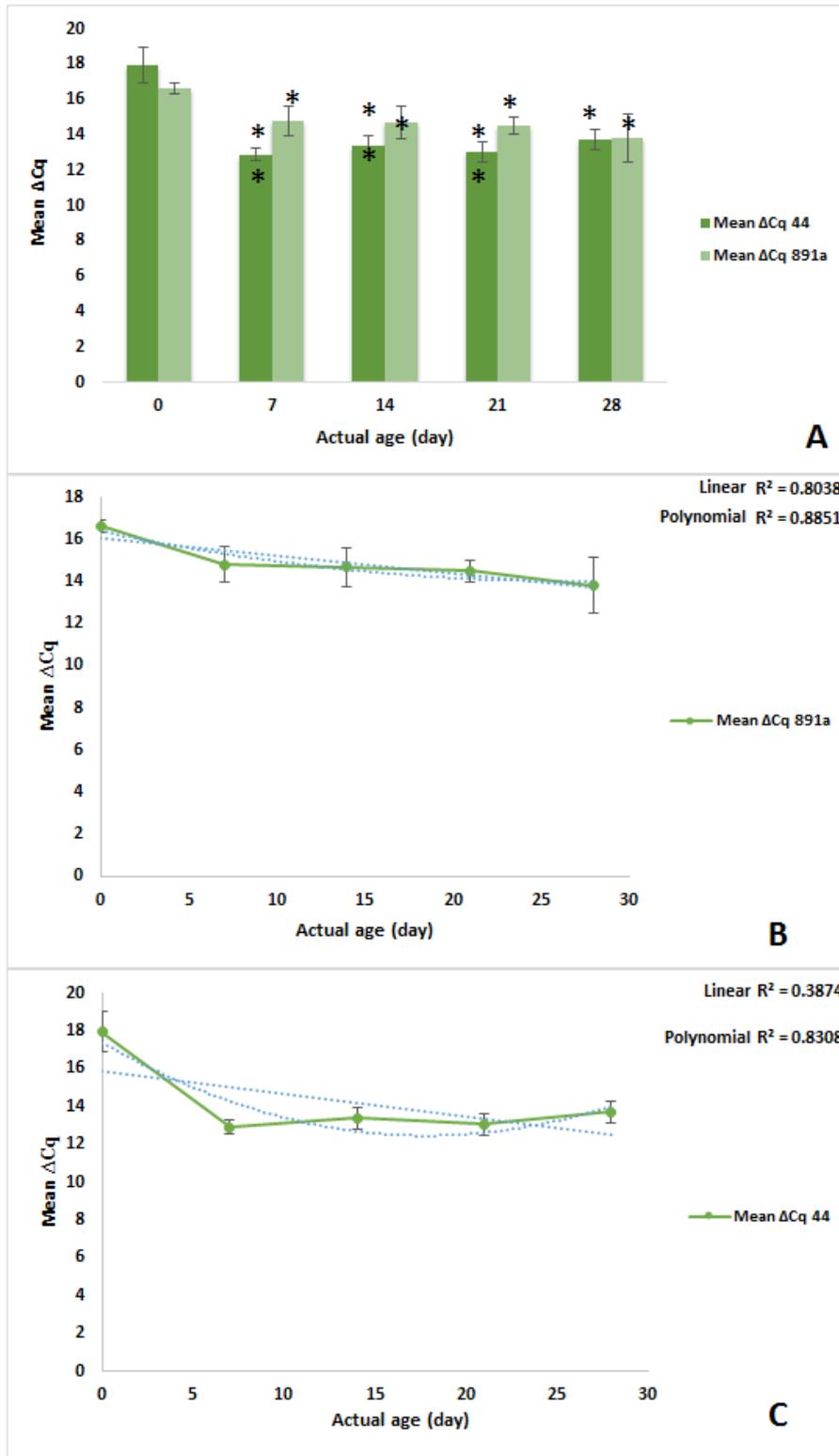


Figure 61. MiRNA in semen samples without DTT (-DTT), aged up to 28 days. (A) Mean  $\Delta Cq$  of miRNA 891a and miRNA 44. (B and C) the degradation pattern of miRNA 891a and miRNA 44. Column denoted with one asterisk (\*) indicates  $p < 0.05$ , column with two asterisks (\*\*) indicates  $p < 0.01$  and column with NS indicates no significant difference. Error bars represent one standard deviation (N=15).

As seen in Figures 60 and 61, an unexpected result was also detected in the semen samples. Here, it was demonstrated that both markers generally tend to correlate with actual age, and both have different patterns with and without DTT. MiRNA 891a decreased linearly with  $R^2$  values of 0.25 and 0.80, with and without DTT, respectively (Figures 60B and 61B). Whereas, RNU 44 had the strongest  $R^2$  value of 0.77, with DTT (Figures 60C and 61C). The stability and ability of markers to be degraded could be the main reason because the pattern of miRNA degradation is still poorly understood (Gutiérrez-Vázquez *et al.*, 2017, Zhang *et al.*, 2012). These results are likely to have been influenced by experimental factors or storage conditions. The observed correlation suggests that  $\Delta Cq$  could be used as an indicator to predict age, particularly in saliva, but less so in semen and blood.

### **6.3.1. The patterns of $\Delta Cq$ in microRNA markers as an indicator to predict the age of a stain**

A previous study concluded that high quality and intact RNA produced highly amplicon level (low Cq) and vice versa when integrity and quality are decreased (Silva, 2012). To explore the correlation patterns of Cq values, calculations, such as  $\Delta Cq$  ( $Cq_{Max} - Cq_{Target}$ ),  $\Delta Cq$  ( $Cq_{Target} - Cq_{Reference}$ ), and the relative ratio between markers ( $\Delta Cq_{Target} / \Delta Cq_{Reference}$ ) were performed. The Pearson's correlation analysis was performed on all Ct calculations (Appendix 4. Table 45), to explore which form was the most suitable in blood, saliva, and semen. The results of this analysis showed that all Cq forms in saliva samples were significantly correlated with actual age, except the relative ratio ( $\Delta Cq_{205} / \Delta Cq_{44}$ ). The strongest negative correlation was detected in the  $\Delta Cq$  (40 -  $Cq_{205}$ ), with a correlation coefficient (r) of -0.85. In the blood samples, no significant difference in age was observed in all Cq forms (Appendix 4. Table 47); the strongest correlation (0.51) was observed in the  $\Delta Cq$  of miRNA 451 (40 -  $Cq_{451}$ ). In the semen samples (+DTT), the strongest correlation (0.68) was observed in the  $\Delta Cq$  of RNU 44, whereas,  $\Delta Cq$  of miRNA 891a showed the highest

correlation (0.69) when DTT was avoided (Appendix 4. Tables 46 and 48). Only the two previous parameters ( $\Delta Cq$  of RNU 44 and  $\Delta Cq$  of miRNA 891a) in semen samples were significantly correlated with actual age. Regardless of the type of correlation, the possibility of using single and multiple regression analysis was still available, especially in the saliva samples. Therefore, all  $\Delta Cq$  forms were tested and the results showed that appropriate models for saliva age prediction were generated (Table 28).

Table 27. Models were generated in saliva using Ct forms as indicators.

Saliva samples	Pearson's correlation	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
$\Delta Cq$ of 205	0.85**	Quadratic	$y=28.8-0.03x-0.09x^2$	0.74	4.4
$\Delta Cq$ of 44	0.81**	Linear	$y=45.72-2.97x^2$	0.65	4.7
$\Delta Cq$ (Cq 2055-Cq44)	0.56**	Linear	$y=18.49-2.75x$	0.32	6.9
Relative ratio( $\Delta Cq205/\Delta Cq44$ )	0.45**	Linear	$y=39.42-22.21x$	0.20	7.4

Note: \* represented of  $p<0.05$  and \*\* represented that  $p<0.01$

As seen in the table above, the strongest correlation was obtained with  $\Delta Cq$  model (40 - Cq 205), and the lowest MAD value with the same model. Therefore,  $\Delta Cq$  of 205 could be the best model generated to predict the age of saliva samples. The multiple regression analysis showed no significant difference in all combinations, except in  $\Delta Cq$  (Cq 205 - Cq44), and in relative ratio (Cq 205/Cq 44), with a R<sup>2</sup> value of 0.61, and MAD value of 5.5. This equation is listed in Table 29. By using the three best models obtained in the saliva samples, the mean age prediction was calculated (Figure 62).

Table 28. Age prediction model using multiple regression analysis in saliva samples.

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-122.289	46.694		-2.619	.022
$\Delta Cq$ (Cq 205-Cq 44)	-16.600	4.674	-3.391	-3.552	.004
Relative Ratio ( $\Delta Cq$ 205/ $\Delta Cq$ 44)	142.799	47.308	2.881	3.019	.011

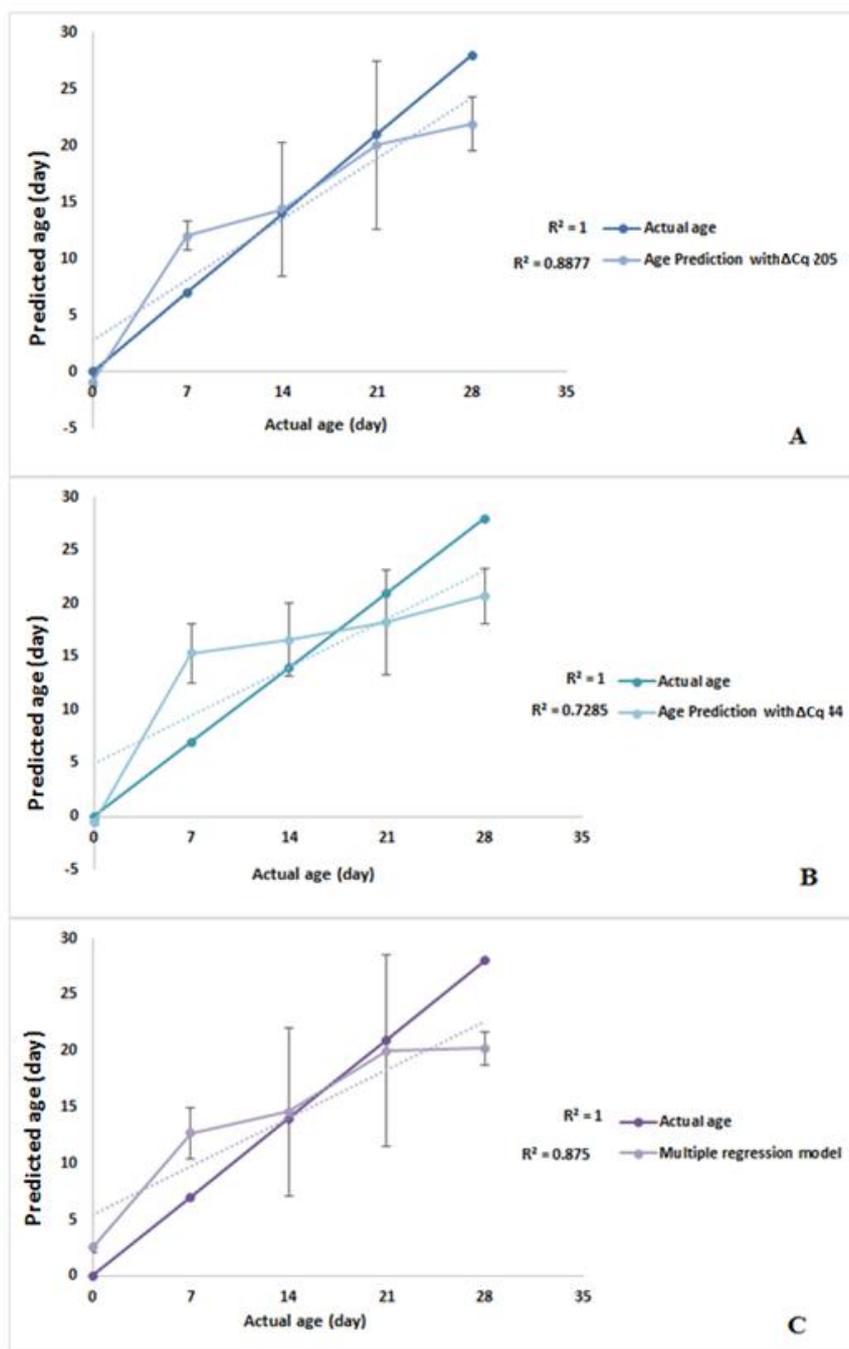


Figure 62. The mean age prediction using miRNAs markers in saliva. (A, B, and C) the mean age predictions using  $\Delta Cq$  of 205, 44, and multiple regression in saliva, aged up to 28 days, respectively. Error bars represent one standard deviation.

As illustrated in figures 62A and B, the saliva miRNA models were close to predicting the fresh samples (0 days). In addition,  $\Delta Cq$  miRNA 205 and multiple regression models were close to predicting the real points at 14 and 21 days, except high SD at these points. Furthermore,  $\Delta Cq$  of the RNU 44 model appropriately predicted age at previous points, with the lowest SD compared to other models. Therefore,  $\Delta Cq$  205 can be considered as the best model to predict the age of saliva, as three points were approximately predicted, with the highest  $R^2$  value.

### **6.3.2. The patterns of RQ of microRNA markers as an indicator to predict the age of a stain**

Relative transcripts of miRNAs were also applied to generate models that could be used to predict the age of blood, saliva and semen. In these experiments, miRNA 451, 205, and 891a were investigated, and all normalised with RNU 44. The mean value of the quantification cycle (Cq) was used to calculate RQ. The results showed that relative levels of miRNA markers tended to decrease over time (Figure 63).

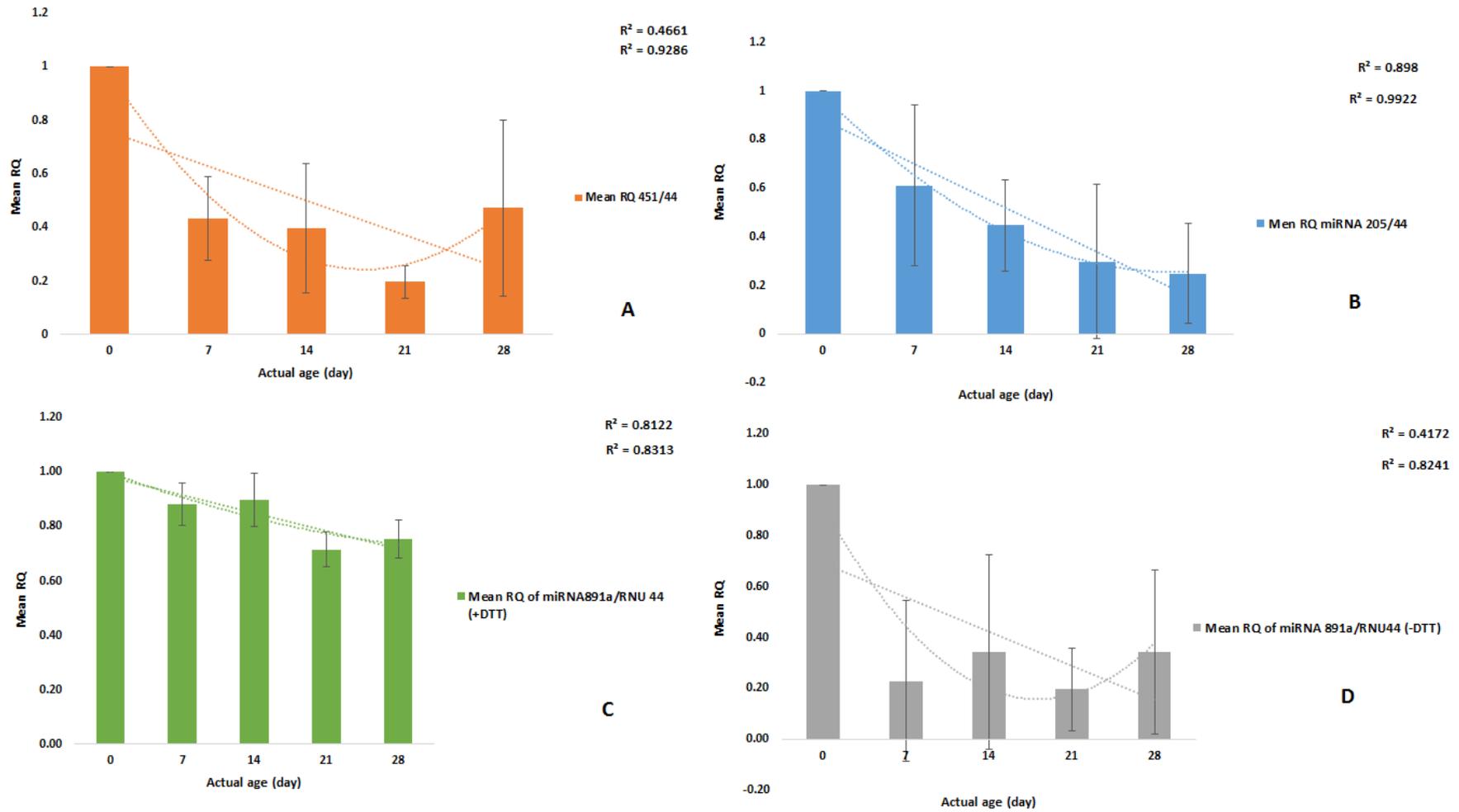


Figure 63. The mean RQ of selected miRNA aged to up 28 days. (A) The relative level of miRNA 451 in the blood (B) the relative level of miRNA 205 in saliva, (C and D) the relative level of miRNA 891a with and without DTT, respectively. Error bars represent one standard deviation.

As seen in Figure 63, the strongest linear and polynomial correlation was observed in miRNA 205 in saliva samples, with  $R^2$  values of 0.898 and 0.99, respectively (Figure 63B). As mentioned previously, the calibrator sample (Time 0) is normalised with itself and the resulting is always equal 1 without any error. For this reason, the standard deviation of the calibration sample is always 0 (Livak and Schmittgen, 2001, Rao *et al.*, 2013).

To generate the models, the Pearson's correlation analysis was performed using RQ as the parameter, and the results showed appropriate values were found. Consequently, a simple regression analysis with CI and PI at 95% was performed to explore the regression line of RQ in selected markers. Linear, quadratic and cubic curves were investigated. The results showed that blood and saliva markers have a linear curve, with  $R^2$  values of 0.34 and 0.59, respectively, whereas, the semen markers have two different curves: linear and cubic, with  $R^2$  values of 0.62 and 0.51, with and without DTT, respectively. Based on the results of this analysis, four equations were obtained and listed in Table 30.

Table 29. The simple regression analysis using RQ of miRNA markers in blood, saliva and semen samples.

Type of sample	Frame time	Number of samples	Pearson's correlation(r)	Target gene	Reference gene	Optimal curve	Estimated curve equation	R <sup>2</sup>	MAD (day)
Saliva	Month	15	0.77**	miRNA 205	RNU 44	Linear	$y=25.82 - 22.71x$	0.59	5.2
Blood	Month	15	0.58**	miRNA 451	RNU 44	Linear	$y=23.24 - 18.48x$	0.34	6.4
Semen(+DTT)	Month	15	0.79**	miRNA 891a	RNU 44	Linear	$y=70.16 - 66.13$	0.62	5.3
Semen(-DTT)	Month	15	0.51*	miRNA 891a	RNU 44	Cubic	$y=15.54 + 11.31x + 21.85x^2 - 48.72x^3$	0.51	5.1

Note: \* represented of  $p < 0.05$  and \*\* represented that  $p < 0.001$

Table 30 illustrates all models generated, using RQ of miRNA, and in each model, a unique optimal curve was observed. As mentioned previously, a strong correlation also generates strong R<sup>2</sup> value, and verse versa. All RQ obtained illustrated a significant correlation with actual age. The MAD in all models was high and this could be due to low R<sup>2</sup> values that were obtained. By using these models, age prediction was calculated in the saliva, blood and semen samples (Appendix 1. Table 37), and the data generated were plotted against actual age (Figure 64).

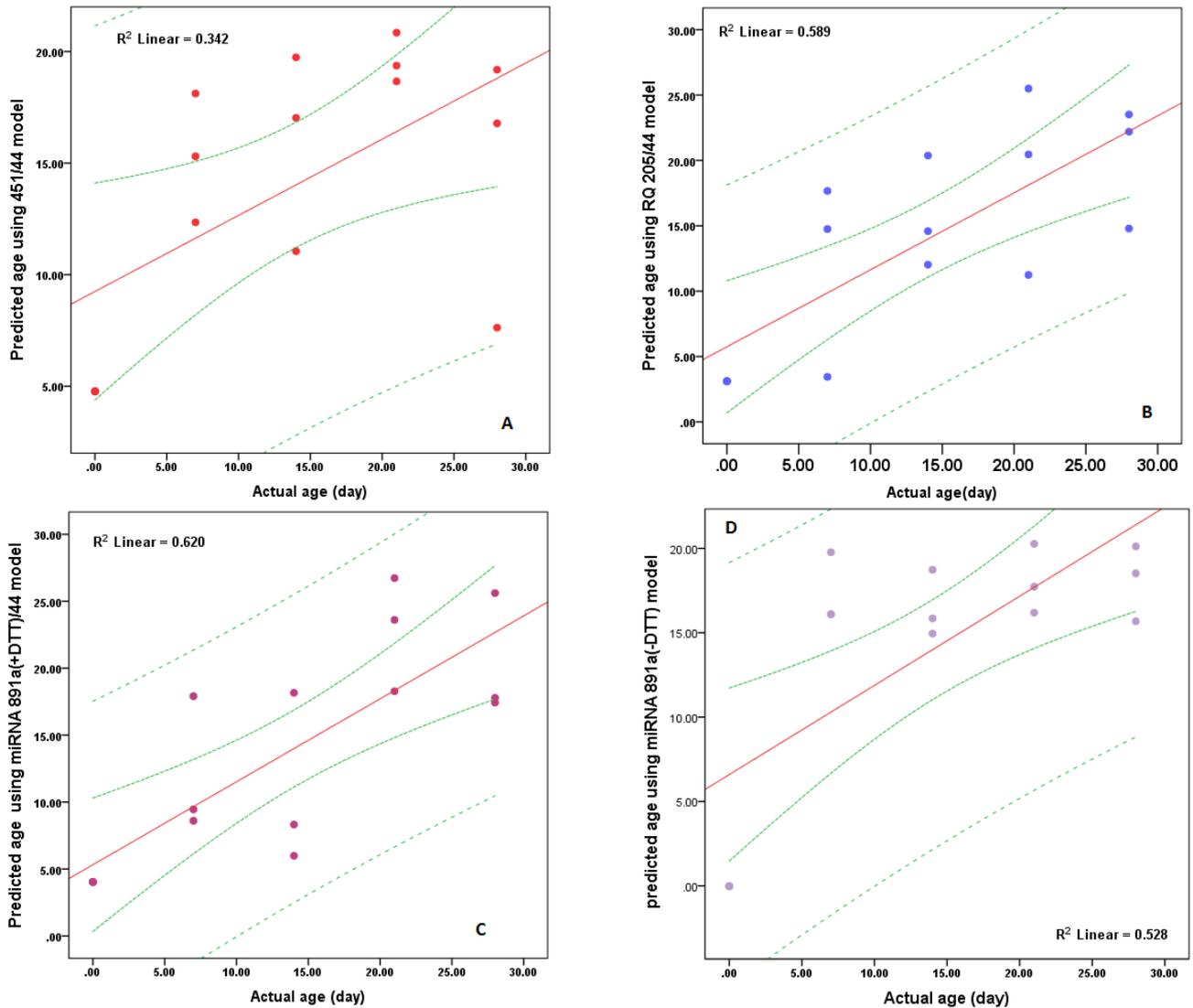


Figure 64. Age prediction obtained, using SRI. (A) Age prediction calculated with miRNA 451/44 in blood, (B) predicted age using miRNA 205/44 in saliva, (C and D) age prediction using the semen models, with and without DTT, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal, the small dashed green lines represent confidence limits, and the large dashed green lines represent prediction limits.

As illustrated in Figure 64, a wide confidence interval was observed, particularly in the blood samples. The main reason could be related to a high standard error, due to the small sample size and different Cq values obtained. As seen, no data were observed out the PI level (large dashed green line). To present the findings clearly, the predicted mean age was calculated and plotted against actual age (Figure 65).

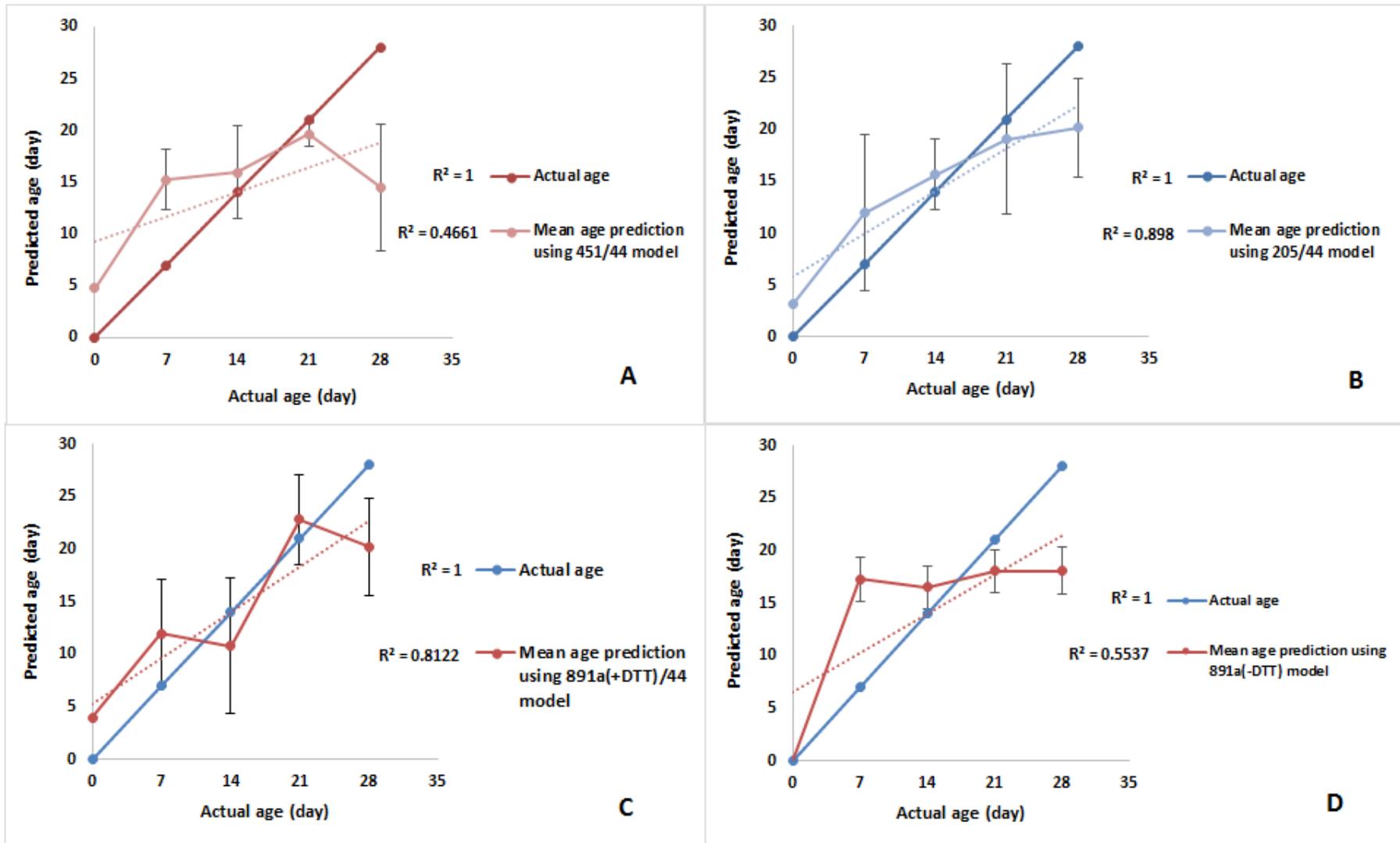


Figure 65. The mean age prediction using miRNA models. (A) Blood age prediction, using miRNA 451/44 model, (B) miRNA 205/44 model in saliva samples, (C and D) semen models, with and without DTT, respectively. The error bars represent one standard deviation.

Figure 65 shows the predicted mean age that was calculated in the blood, saliva and semen samples. The saliva model (RQ of miRNA 205/44) has the strongest correlation with actual age, whereas, the miRNA 451/44 model in blood has the lowest correlation (Figure 65A and B). Therefore, the accuracy of any model depends on the best choice of markers and reference genes. In time-based experiments, the stability of reference genes is the most important factor to generate accurate normalisation. To measure expression levels accurately and improve the obtained models, normalisation of multiple reference genes, rather than just one, is required.

### **6.3.3. Geometric mean normalisation strategy**

The instability of reference genes has two main sources: experimental variability including technology variation, and natural variability between tissues, individuals, and time. To minimise the likelihood of these variations occurring, normalisation with multiple reference genes is the most useful approach (Chervoneva *et al.*, 2010, Vandesompele *et al.*, 2002, Peltier and Latham, 2008, Imai *et al.*, 2014). A normalisation factor ( $NF_N$ ) using multiple genes was defined as the geometric mean of relative expression value in the N gene regions, in each sample (Imai *et al.*, 2014). Here, the geometric mean normalisation was employed using multiple reference genes, including miRNA and mRNA markers. This strategy could be used to improve the models, as well as reduce the MAD value.

Three models were used in the geometric mean normalisation strategy; the selection of these models was based on the lowest  $R^2$  obtained. In the saliva samples, the HTN3/GAPDH model ( $R^2 = 0.45$ ) was selected, and  $\Delta Ct$  of HTN3 was normalised geometrically. Firstly, this was achieved using a combination of miRNA 205, RNU 44, and GAPDH, and secondly, with only miRNAs 205 and 44. All obtained data underwent SRI in the same way as described earlier to generate a new prediction model (Figure 66).

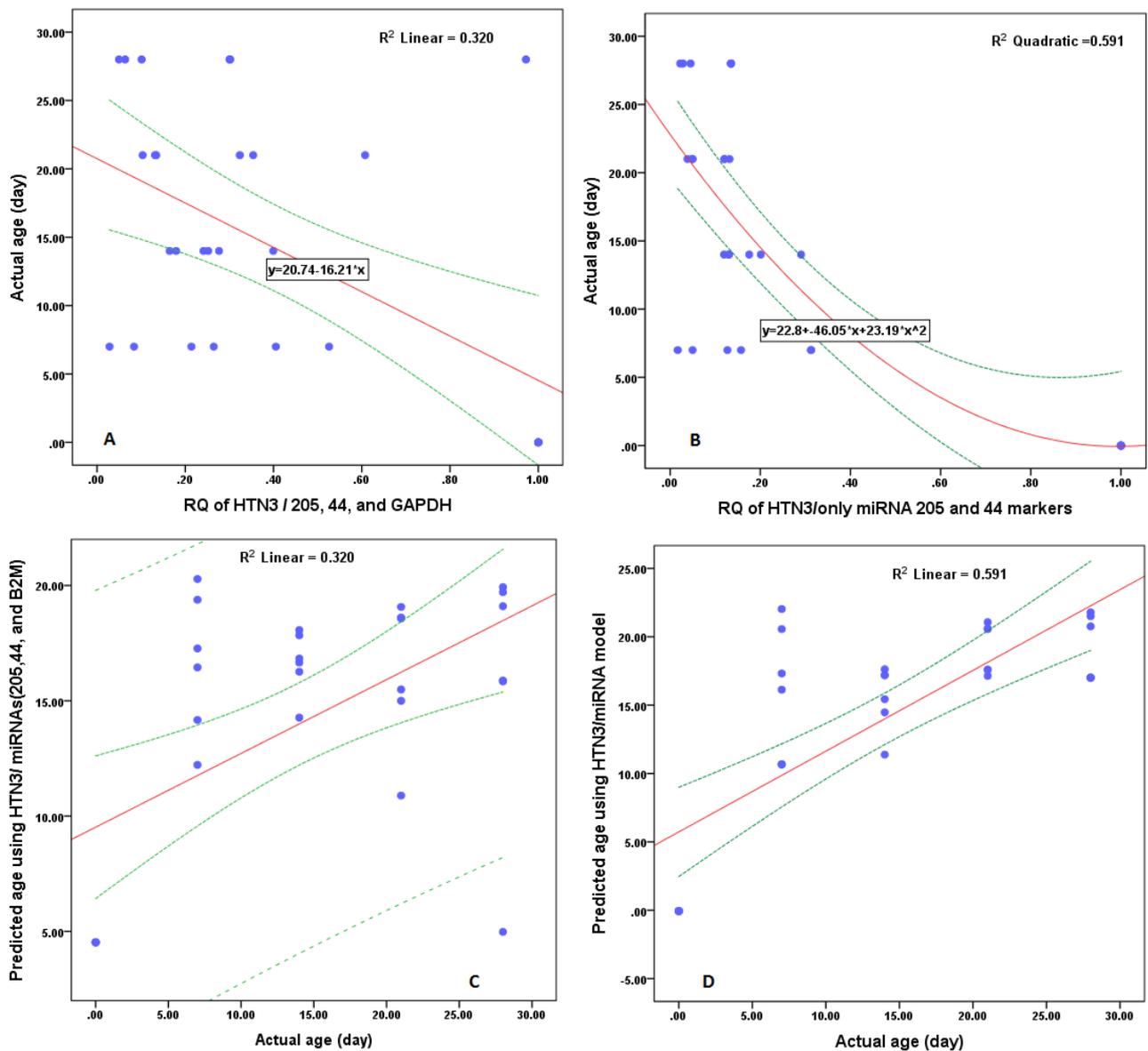


Figure 66. Geometric mean normalisation strategy in saliva samples. (A) The model that was generated using RQ of HTN3, normalised with miRNAs 205 and 44, and GAPDH, (B) the model generated with RQ of HTN3 that was normalised with miRNAs 205 and 44, (C and D) the age prediction using the previous models, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. The small dashed green lines indicate confidence limits and the large dashed green lines indicate prediction limits (N=30).

As seen in the figure above, there are two models with two different normalisation strategies. The first normalisation strategy was conducted using three reference genes, including miRNAs 205 and 44, and GAPDH (Figure 66A), and the second normalisation strategy included miRNAs 205 and 44 only (Figure 66 B). Two models with different  $R^2$  values were generated and illustrated in each graph; the highest  $R^2$  value was obtained by using only

miRNA as reference genes. The results indicate that GAPDH was unstable reference gene, and it could be the main reason for generating a low  $R^2$  value when it was used as individual reference gene with HTN3 marker in saliva samples (Table 13, chapter 4).

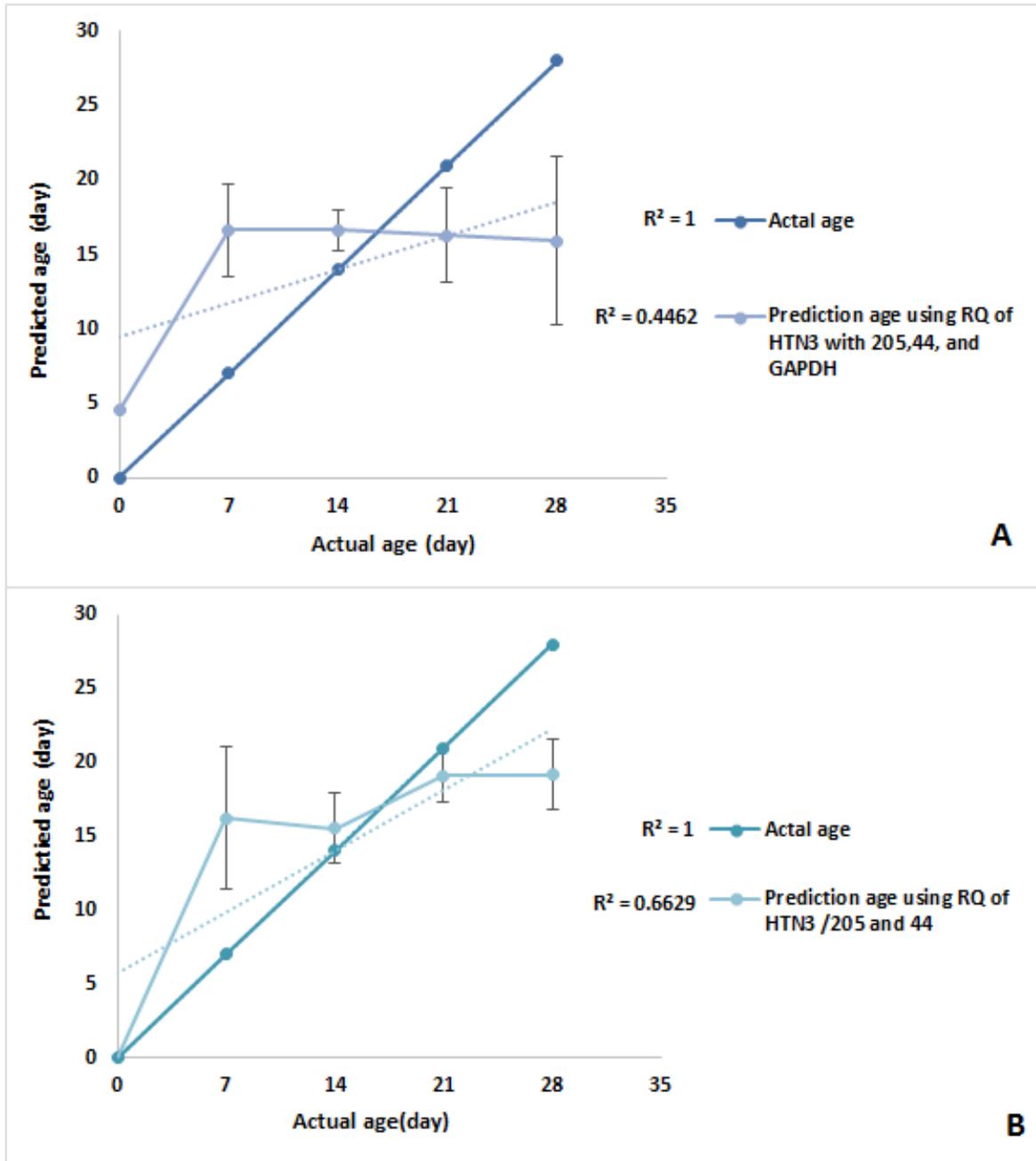


Figure 67. The mean age prediction in saliva samples. (A) The mean age prediction using RQ model of HTN3 normalised with GAPDH, miRNA205 and RNU 44 as reference genes, (B) the same model that was normalised with miRNA205 and RNU 44 only. The error bars represent one standard deviation.

By using the two equations as listed in graphs above, the mean age prediction for the saliva samples was calculated (Appendix 3. Table 43). The results obtained strongly indicate that the best model used was the normalised one with miRNA only, as reference genes. Three points were approximately predicted as same as the actual age, especially in a zero sample, whereas, another model generated low  $R^2$  value and was so far of almost of real points (Figures 67. A and B). Furthermore, the MAD value decreased from 5.8 to 4.5 in the HTN3/ GAPDH and HTN3/ (205+44) models, respectively.

Similarly, the same geometric mean normalisation strategy was employed in the semen samples. This was an attempt to improve the SEMG1/B2M model that obtained a  $R^2$  value of 0.64, when it was normalised with B2M (Table 14, chapter 4). This was achieved when RQ of SEMG1 was normalised with three reference genes, including B2M, miRNA 891a, and RNU 44, as well as with two reference genes, including miRNA assays only. Again, a single regression analysis (SRA) was performed, and two models were obtained (Figures 68. A and C). The same finding that was obtained in the saliva samples was also observed in the semen samples. The highest  $R^2$  value was obtained when the SEMG1 was normalised with geometric miRNA assays only (0.768). Once the two equations were generated, the age prediction was calculated (Appendix 3. Table 43) and plotted against actual age (Figures 68. B and D). The results obtained confirmed the advantage of using geometric mean, especially with the miRNA assays only. The mean age prediction was calculated to compare the two new models and plotted against actual age (Figures 69. A and B).

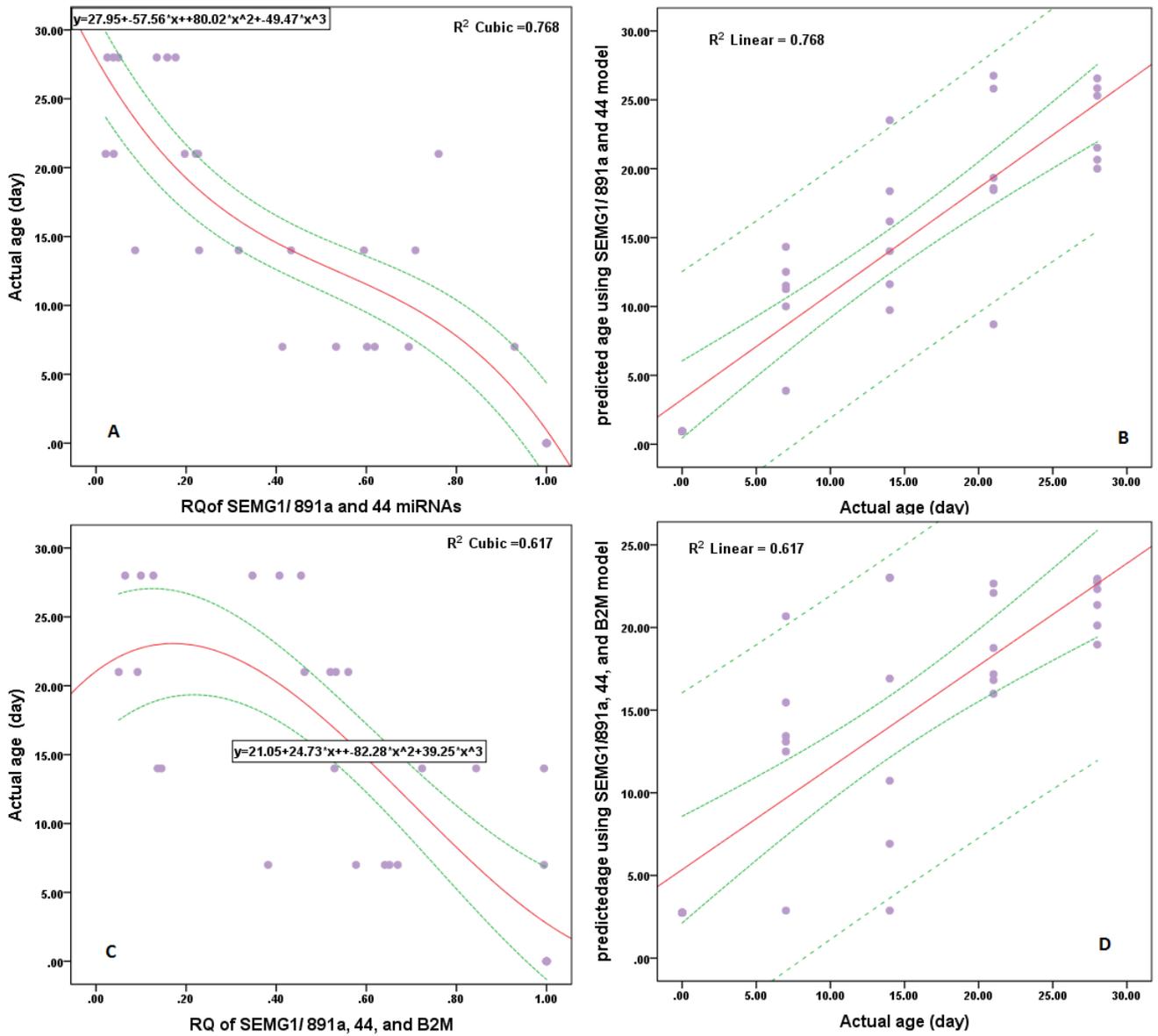


Figure 68. Geometric mean normalisation in semen samples. (A and C) models of RQ of SEMG1 that were normalised using a geometric mean of miRNA assays only, and accompanied with miRNA 891a, RNU 44, and B2M, respectively. (B and D) Age prediction using previous models, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. The small dashed green lines indicate confidence limits and the large dashed green lines indicate prediction limits.

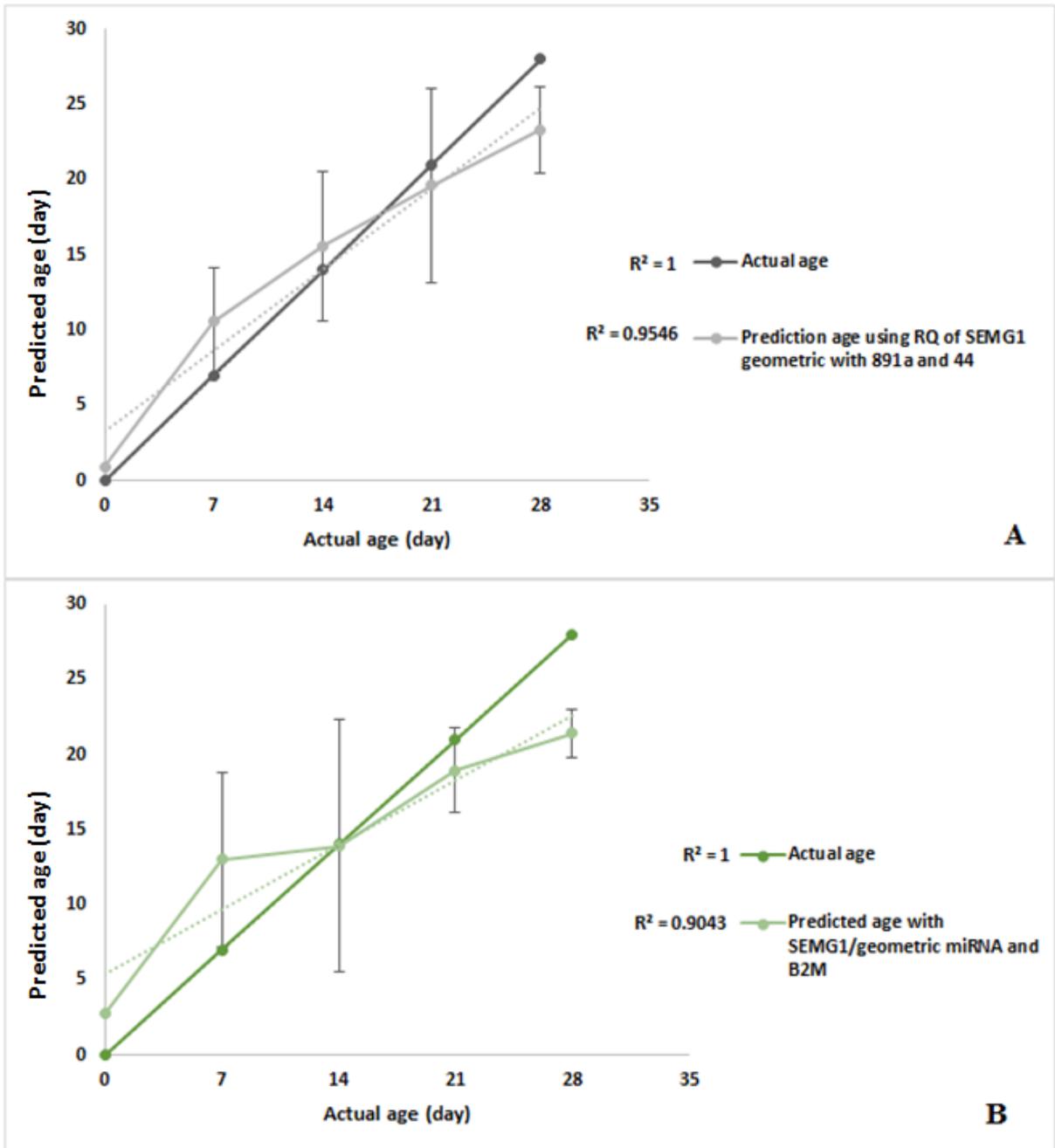


Figure 69. Mean age prediction in semen samples. (A and B) Mean age prediction calculated using SEMG1 models that was normalised using miRNA 891a and RNU 44, or normalised with miRNA 891a, RNU 44, and B2M, respectively. The error bars represent one standard deviation (N=30).

As illustrated in the figures above, a SEMG1 model using miRNA assays only generated strong  $R^2$  values, indicating a correlation with actual age. At least three-time points were approximately detected, whereas, other models predicted the same time-point at day 14, with a large error bar. Therefore, a new SEMG1 model is greatly improved when compared to previous models, when B2M was used for normalisation. The final  $R^2$  value for the mean prediction was 0.76 (Figure 33 B, chapter 4); however, the SEMG1 model with geometric normalisation generated the highest  $R^2$  value (Figures 69. A and B).

The GYPA model used in the blood samples was also normalised with a geometric mean, using miRNA 451, RNU 44, and B2M, as described in the saliva and semen examples. In fact, the GYPA marker obtained a low  $R^2$  value when it was normalised with B2M (0.11) (Table 12, chapter 4). For new normalisation, both groups obtained approximately the same  $R^2$  value (Figures 70A and B). The prediction age was calculated using two new formulas (Appendix 3. Table 43), and the results showed great confidence and prediction intervals due to using inaccurate models. Therefore, the mean age prediction for both new models was calculated and plotted against actual age. As seen in Figures 71A and B, there was no improvement to the new models, using two types of normalisation strategies. The main reason for this observation could be related to GYPA that showed a weak correlation with actual age. As previously mentioned, a good marker for age prediction should be correlated with an increase in time.

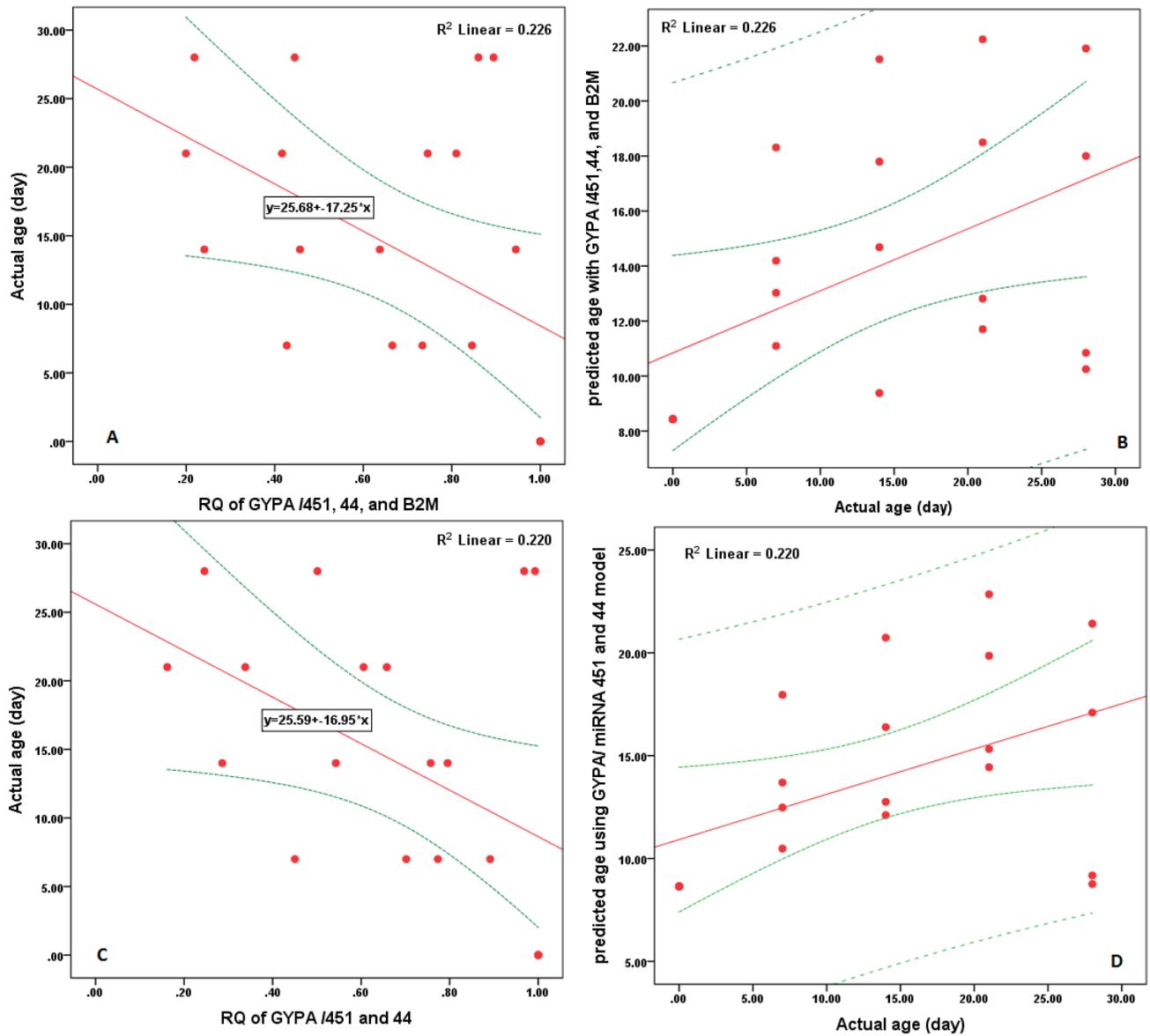


Figure 70. The new model generated with geometric mean in blood samples. (A and C) Models of RQ GYPA that were normalised with miRNA 451, RNU 44, and B2M, or normalised with only miRNA 451 and RNU 44, respectively, (B and D) age prediction generated by previous models, respectively. The solid red line in the graph represents the modelled space where predicted age and actual age are equal. The small dashed green lines indicate confidence limits and the large dashed green lines indicate prediction limits (N=20).

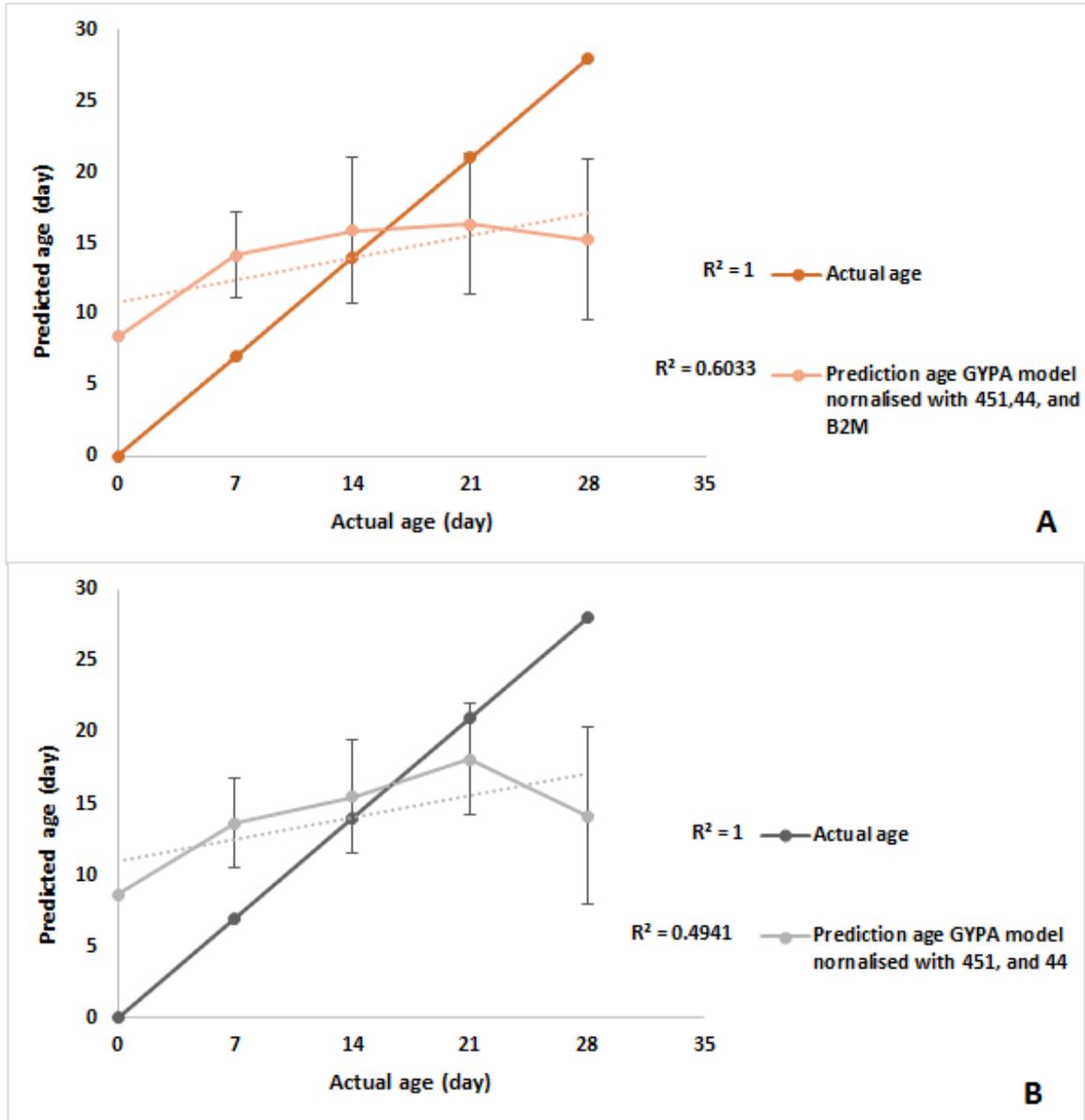


Figure 71. New normalisation with GYPA in the blood samples. (A and B) the mean age prediction calculation using GYPA models that were normalised with miRNA 451, RNU 44, and mRNA B2M, or only miRNA451 and RNU44, respectively. The error bars represent one standard deviation.

#### 6.4. Discussion

The prediction of age in any biological stains in forensic investigations plays a major role in determining the elapsed time between deposition and sample recovery. Age prediction for biological stains based on qPCR of RNA decay has precisely required the evaluation of parameters that correlated with the time after deposition. In previous chapters, it was illustrated that some mRNA markers have a strong correlation with actual age; however, in some cases, these markers generated models with low  $R^2$  values. As discussed previously, this could be due to inaccurate normalisation with unstable reference genes. Therefore, this chapter aimed to explore the patterns of miRNA degradation up to 28 days, and the possibility of using the most stable miRNA markers as reference genes to improve normalisation of the previous mRNA models, through geometric mean normalisation strategies. This was achieved by using the qPCR method, targeting miRNA 451, miRNA 205, and miRNA 891a in blood, saliva and semen samples, respectively. Furthermore, RNU 44 was also investigated as an additional reference gene alongside each marker, in all samples.

Once all markers were successfully detected,  $\Delta Cq$  was calculated using several methods in the blood, saliva and semen samples. Although miRNAs are among the most stable elements, an unexpected result was detected in the saliva samples that showed that  $\Delta Cq$  significantly decreased over time. The strongest linear and polynomial correlations were observed in saliva samples, with miRNA 205 and  $R^2$  values of 0.93 and 0.97, respectively (Figure 58B). The lowest linear and polynomial correlations were observed in blood samples, with  $R^2$  values of 0.026 and 0.089, respectively (Figure 59C). In the blood samples, a minor variation was also observed with miRNA 451, between fresh (0 days) and 14 days (Figures 59A and B). The study of Wang, *et al.* (2013a) also supported this finding where the minor fluctuation in expression levels between a 24-hour and 1 month was observed (Wang *et al.*, 2013a).

Furthermore, the study of Uchimoto (2014) concluded that a variation was found in a 24-hour versus 1 week (Uchimoto, 2014).

Another study was conducted to estimate the age of bloodstains based on biological and toxicological analyses. In this experiment, miRNA 451 significantly decreased over time, up to 28 days. The authors explained that degradation occurred as a result of the blood samples only partially drying, before being placed in a chamber at 25°C and 50% humidity (Nakao *et al.*, 2013).

The time-wise stability of miRNA markers was tested in the aged sample for 1 year, in laboratory conditions. In this study, three miRNA markers for blood samples, and four miRNA markers for semen samples, including miRNA 891a, were tested. The results indicated that no time-wise degradation was detected in the miRNA markers tested (Zubakov *et al.*, 2010). Therefore, the correlation observed in the current study is possibly related to extended storage conditions, over 1 year at -20°C. Unfortunately, the amount of RNA was limited and did not allow for quality and quantity testing of the samples. Although the stability of miRNA in degraded RNA was investigated, the results indicated robust miRNA measurements with RT-qPCR, even in degraded RNA (Jung *et al.*, 2010). In contrast, other studies reported that the expression profiling of miRNA is potentially influenced not only by the method used for isolation, but also by the RNA storage conditions and handling (Bravo *et al.*, 2007, Wang *et al.*, 2008, Doleshal *et al.*, 2008). The stability of isolated miRNA in stored RNA samples was also investigated, where Trizol/TRI-Reagent isolation methods were tested, and compared to the mirVana isolation kit (Ambion) and RNeasy kit (Qiagen) (Mraz *et al.*, 2009). The results showed that Trizol/TRI-Reagent based isolation methods are robust and can be reproduced and obtained miRNA samples without the tendency to degradation when the sample is stored and handled properly. The conclusion

of previous statements could be explained that RNase Mini kit which was used for miRNA extraction may influence the patterns of miRNA expression.

On the contrary, the observed correlation between miRNA and actual age, particularly in the saliva samples, could be related to reasons other than the degradation of miRNA. In addition, saliva samples were the least stable, when compared to blood samples, in a study that demonstrated blood-specific and saliva-specific mRNA markers can be amplified successfully and reliably, in old stains up to 16 years old and 6 years old, respectively (Zubakov *et al.*, 2009). The tendency to decrease was also observed in semen samples; the reason could be the same as discussed previously, as well as DTT reagent could be potentially affected on miRNA.

The correlation patterns of different forms of  $\Delta Cq$  were explored in the blood, saliva, and semen samples. The results showed that  $\Delta Cq$  in saliva has the strongest correlation value, therefore, those samples underwent single and multiple regression analysis to obtain models for saliva age prediction (Tables 28 and 29). The  $\Delta Cq$  of miRNA 205 and RNU 44 had  $R^2$  values of 0.888 and 0.73, respectively (Figures 62A and B). The  $\Delta Cq$  forms in blood samples showed no significant difference with actual age and were therefore excluded. In semen samples,  $\Delta Cq$  forms are significantly correlated with actual age, but with Low  $R^2$  values.

As was achieved in mRNA markers, RQ was also employed for miRNA markers to generate models that could be used for age prediction in blood, saliva and semen (+DDT and -DDT) samples. The average Cq of each marker in every sample was calculated, after normalisation with reference gene RNU 44 was achieved. Three curves were tested as regression lines, and the results showed that blood, saliva and semen (+DDT) samples have an optimal linear curve, with  $R^2$  values of 0.34, 0.59, and 0.62, respectively (Table 30). Although the mean RQ of markers used showed appropriate correlations with actual age, particularly in saliva

(Figure 63), the regression analysis showed this average decreased. This could be related to high SD, due to the variability of samples, and small number of samples used.

Once the equations were generated, age prediction in the blood, saliva, and semen samples was calculated, plotted against actual age, and the accuracy of models used was compared. The results indicated that the mean age predictions in saliva and semen (+DTT) samples have the highest correlation with actual age, with  $R^2$  values of 0.898 and 0.81, respectively (Figure 65. B and C).

MiRNAs with short sequences are more stable than mRNAs. In contrast, miRNAs showed limited stability over a short period in the Sethi and Lukiw (2009) study that investigated miRNA as an endogenous gene in human brain tissue (Sethi *et al.*, 2009). Another study was conducted to investigate the level of RNA, 18S rRNA and miRNA, to estimate PMI (Li, *et al.*, 2014). The findings of this study indicated that the degradation pattern of tissue 18S-rRNA and miRNA is useful in the determination of the interval, within 7 days (Li *et al.*, 2014). The pattern of five miRNA expressions was investigated within 48-hours, in room temperature, to determine PMI in adult mice (Wang, *et al.*, 2013a). The results indicated that miRNA remained stable in the first 24-hours, before some of miRNA started decreasing; the authors suggested to study degradation patterns of miRNA before employed miRNA for PMI (Wang *et al.*, 2013a). The previous statements could be supported the finding that miRNA can stay stable and degraded according to its function. A similar study was conducted using miRNA-142-5p and miRNA-541, for bloodstain deposition timing (Lech, *et al.*, 2014). The results concluded that two miRNA markers were not suitable for estimating time deposition of the bloodstain. miRNA-142-5p showed no statistical significant differences in expression during the 24-hour day/night period, and miRNA 541 was abundantly low (Lech *et al.*, 2014).

Normalisation, using a single reference gene, leads to inaccurate normalisation up to 6.4-fold, in 10% of cases (Leitao *et al.*, 2014, Vandesompele *et al.*, 2002). Therefore, to improve the data obtained in this study, the combination of mRNA and miRNA markers could be useful. This was achieved by employing the geometric mean normalisation strategy as a link between more stable candidates (miRNA), and the most degraded assays (mRNA markers). Three models with the lowest  $R^2$  values were selected, including HTN3/GAPDH, SEMG1/B2M, and GYPA/B2M for saliva, semen and blood, respectively. All three models underwent geometric mean normalisation strategy in two different ways. The results showed that new models are improved, particularly when using miRNA markers, and this suggests that miRNAs are the most useful candidates for normalisation, rather than for using them as parameters of age prediction.

## **6.5. Conclusion**

The expression patterns of miRNA could be dependent on different factors, including isolation methods, storage time, and natural of target marker and sample. MiRNA is more sensitive and stable compared to mRNA, and this could lead to improvements in quantification RT-PCR data, by using it as a reference gene. In some cases, miRNA tended to decrease over time, therefore, so this behaviour requires further investigations.

## **Chapter Seven: General Discussion**

Blood, saliva, and semen are the most common biological evidence found at a crime scene (Sirker *et al.*, 2016). However, identification of the person whose sample was found at the crime scene is not sufficient to ground their conviction or an exoneration as the perpetrator of the crime. This is because the deposition of the sample could have taken place before particular offence was committed. Consequently, determination of the TSD is the most important factor in linking relevant evidence to the crime scene and/or excluding those that happened outside of the time period in which the crime was committed. Some efforts have been made to overcome this dilemma using different techniques and methods. Amongst these, RNA decay rate measurement stands out as one of the most sensitive and accurate methods. Therefore, the aim of this thesis was to explore the forensic applications of RNA analysis for a stain age prediction. Several factors were examined, namely mRNA and miRNA in an attempt to determine the age of blood, saliva, and semen samples using RT-qPCR method.

A number of studies showed that level of RNA transcription generally decreased with a longer PMI (Sampaio-Silva *et al.*, 2013, Partemi *et al.*, 2010, Bauer *et al.*, 2003a). The stability of mRNA has been tested in many studies and showed that it is possible to detect mRNA expression in aged biological stains. For example, mRNA was found on seminal stains that were 33 and 56 years old (Nakanishi *et al.*, 2014), whereas, Lindenbergh *et al.* (2012a) detected mRNA in 28-years old samples (Lindenbergh *et al.*, 2012a). Bauer *et al.* (2003b) demonstrated that although RNA gradually degrades in dried bloodstains, mRNA is still suitable for RT-qPCR and can be extracted from blood samples stored up to 15 years (Bauer *et al.*, 2003b). Another study demonstrated that the blood and saliva –specific mRNA markers can be amplified successfully and reliably in old stains up to 16 and 6 years old, respectively, suggesting their suitability for tissue identification in forensic studies (Zubakov *et al.*, 2009).

In the present thesis, the amplification of all markers was successfully detected over 28 days period. Two common reporter systems were used, namely TaqMan probe system (Holland *et al.*, 1991, Livak *et al.*, 1995), and intercalating SYBR Green assay (Wittwer *et al.*, 1997). The most important step in RT-qPCR to ensure that a single strand cDNA is successfully generated. Therefore, good cDNA synthesis strategy is an imperative for successful method and generation of good quality data. Interestingly, SYBR Green assay proved the most valuable when evaluating RNA formation and decay (Schmittgen *et al.*, 2000). In Schmittgen and co-workers (2000) study, four quantitative RT-PCR methods, band densitometry and probe hybridisation as end-point, and SYBR Green and TaqMan as real-time methods, were compared in the ability to test the same cDNA. The linear regression analysis of the first-order mRNA decay plots was performed for different methods, and the  $R^2$  values obtained were the highest for SYBR Green. Unsurprisingly, real-time methods (SYBR Green and TaqMan probe) proved to be more precise and offer greater dynamic range than end point methods. Interestingly, SYBR Green detection was more precise and gave more linear decay plot than TaqMan chemistry. Previously, mRNA was used for BFI using SYBR Green, and the authors showed that SYBR Green based BFI test could be developed, and may even be more appropriate for commercial forensic science environment use when compared to TaqMan based BFI test. However, caution is warranted here since high background signal in RT negative controls may render SYBR Green based BFI unreliable for use in a court of law (Connolly and Williams, 2011). In summary, each chemistry has its advantages and limitations. For example, the limitation of SYBR Green dye is that it cannot be used in multiplex reactions, due to the fact that it non-specifically binds any double-stranded DNA. This will influence the specificity and the sensitivity of this method, making it difficult to generate accurate result when the template is at the low starting concentration. However, a post-PCR melting curve can help to improve the specificity of the reaction. The main

advantage of SYBR Green dye is that it is flexible and can be used to monitor amplification of any dsDNA sequence, as well as that it does not require probe and therefore reduces the steps in the assay set up and the running costs. Finally, SYBR Green chemistry is easily detectable by the most real-time PCR systems.

In current study, the comparison between both chemistries (TaqMan and SYBR Green) was performed, and for clearer representation of the data, the mean age prediction was calculated for all models. In blood, HBB/B2M model using TaqMan chemistry had the strongest  $R^2$  value (0.91) (Figure 25 B), while the best fit when using SYBR Green chemistry was observed for FGB/ACTB model ( $R^2 = 0.88$ ) (Figure 41E). In saliva, with TaqMan chemistry, the mean age prediction using  $\Delta Cq$  of HTN3 model had the strongest  $R^2$  value of 0.98 (Figure 28 A). With SYBR Green the strongest  $R^2$  value (0.99) was observed when HIF1A/VEGFA model was used (Figure 50B). Moreover, in blood samples HBB marker was used for both chemistries, and showed the strongest linearity with  $R^2$  value of 0.82 obtained when Taqman was applied (Figure 24 B).

The comparative Ct method has been widely used as a relative quantification strategy for RT-qPCR method (Rao *et al.*, 2013). This method is a convenient way of calculating relative target gene levels using directly the threshold cycles of different samples. However, this approach is mostly dependent on an invalid assumption of 100% PCR amplification efficiency across all samples under the investigation. Following the hypothesis that RNA decay should decrease over time, an RNA expression experiment was conducted to determine the age of the forensic hair sample that was up to 3 months old. The ratio between 18S rRNA and  $\beta$ -actin was explored, and the result showed both 18S rRNA and  $\beta$ -actin are suitable targets for age estimation of hair samples (Hampson *et al.*, 2011). The current study also examined the decay rate of mRNA and miRNA over a certain time. The single and multiple regression analysis were performed, and mathematical models for age prediction were

generated. In the first experiment, blood samples were left to decompose for one week. Here, B2M/GAPDH emerged as the best predictor of the age of blood stain for up to 7 days, with an  $R^2$  value of 0.43 and MAD of  $\pm 1.4$  day (Table 11). The reason for the weak correlation in this experiment could be related to inoperative selection markers in short period. The deposition time is another factor thought to influence the degradation patterns of genes (Simard *et al.*, 2011). Therefore, in order to investigate this further, the time-frame for sample decomposition was increased up to 1 month, with analyses performed at 7, 14, 21 and 28 days. The quality of all markers was acceptable, except for GYPA1 and MUC7. Delta Cq was obtained and showed that HBB and PRM1 were the most abundant markers detected in blood and semen, respectively. This finding was also supported by the study that revealed HBB and PRM1 as the most stable and reliable markers explored, and they were both detected up to 71 weeks of dry storage (Sirker *et al.*, 2016).

Oxygen-regulated factors such as HIF1A and VEGFA were also explored in blood, saliva, and semen. The RQ was calculated in three different ways, using VEGFA/ACTB, HIF1A/ACTB, and HIF1A/VEGFA as the models. Particularly strong correlation detected for saliva, strongly supports the advantage of using hypoxia marker in these analyses. This was further supported by Bai *et al.* (2017) in a study that showed that hypoxia genes are suitable markers for PMI (Bai *et al.*, 2017). Up to date, no studies were found that used hypoxia genes for stain age prediction, almost of efforts using hypoxia genes were only for PMI (Zhao *et al.*, 2006, Thaik-Oo *et al.*, 2002a).

Micro RNA levels are dictated by both biosynthetic and decay process, which can be presented by three events transcription, maturation, and degradation. To date, the pattern of miRNA degradation and the processing involved remains largely unclear (Marzi *et al.*, 2016). Therefore, miRNA was also explored in the same way as with mRNA markers, namely using qPCR targeting of miRNA 205, 451, and 891a for saliva, blood, and semen, respectively. The

strong correlation which was detected especially in saliva could be related to the hypothesis that RNU44 had affected the sensitivity range of miR-451 and miR-205 on the expression levels. This was also supported by the sensitivity study conducted on saliva and blood deposits using miRNA 205, 451, and RNU44. The result indicated that the sensitivity range of RNU44 was between 1 in 10 and 1 in 100, which was 10-fold higher than miR-205 (Uchimoto, 2014).

Once all models were obtained, age prediction for blood, saliva, and semen was calculated. Table 31 shows the linearity of mean age prediction using mRNA and miRNA models (miRNA models marked in grey). The strongest linearity with mRNA markers was detected using HIF1A/VEGFA model in saliva (0.99). On the other hand, the highest linearity using miRNA marker was detected with  $\Delta Cq$  miRNA 205 in saliva (0.89). In blood age prediction, HBB/B2M could be considered the best model obtained, while in semen the multiple regression model (PRM1/B2M and SEMG1/B2M) was the best model (0.90).

Dependent on the availability of stock reagents, some of the models were validated using blind samples for blood, saliva, and semen. Although all models obtained covered a time period from fresh samples (0 days) up to a month of decomposition, blind blood samples aged from fresh until one year were tested, targeting ALAS2/B2M, HBB/B2M, and HBB/ACTB models. The data obtained from this experiment has strongly supported the hypothesis mentioned earlier because the RQ of almost of models was gradually decreased as the age increased. Therefore, RQ method could be the best choice to quantify RNA degradation because strong correlation was detected in almost all of the markers investigated in the current study. This finding was also supported by Bauer *et al.* (2003b) who investigated the degradation of mRNA as a possible indicator of the age of bloodstains (Bauer *et al.*, 2003b).

Table 30. Mean age prediction calculated with all mRNA and miRNA models obtained using TaqMan and SYBR-GREEN chemistries.

Sample	Frame time	Number of sample	fluorescent chemistries	Model using (Single or multiple regression)	Linearity R <sup>2</sup>	MAD (days)
Blood	Week	56	TaqMan	ALAS2/GAPDH	0.54	±1.7
Blood	Week	56	TaqMan	ALAS2/B2M	0.48	±1.4
Blood	Week	56	TaqMan	GAPDH/B2M	0.53	±1.4
Blood	Week	64	TaqMan	ALAS2/GAPDH ALAS2/B2M	0.63	±1.4
Blood	Month	40	TaqMan	ALAS2/B2M	0.87	±3.8
Blood	Month	40	TaqMan	HBB/B2M	0.91	±3.2
Blood	Month	40	TaqMan	ALAS2/B2M HBB/B2M	0.9	±3.4
Blood	Month	25	SYBRGREEN	HBB/ACTB	0.87	±3.4
Blood	Month	20	SYBRGREEN	FN1/ACTB	0.78	±4.9
Blood	Month	20	SYBRGREEN	FN1/EPOS1	0.61	±4.9
Blood	Month	20	SYBRGREEN	FGB/ACTB	0.88	±3.9
Blood	Month	20	SYBRGREEN	EPOS1/ACTB	0.63	±5.0
Blood	Month	20	SYBRGREEN	CO A/ACTB	0.59	±5.1
Blood	Month	40	SYBRGREEN	VEGFA/ACTB	0.86	±4.4
Blood	Month	40	SYBRGREEN	HIF1A/ACTB	0.8	±4.6
Blood	Month	40	SYBRGREEN	HIF1A/VEGFA	0.84	±4.2
Blood	Month	40	SYBRGREEN	VEGFA/ACTB HIF1A/ACTB	0.91	±3.7
Blood	Month	15	TaqMan	miRNA 451/RNU44	0.47	±6.4
Saliva	Month	30	TaqMan	HTN3/GAPDH	0.55	±5.8
Saliva	Month	30	TaqMan	ΔCq of HTN3	0.98	±3.3
Saliva	Month	40	SYBRGREEN	VEGFA/ACTB	0.94	±3.9
Saliva	Month	40	SYBRGREEN	HIF1A/ACTB	0.89	±4.9
Saliva	Month	40	SYBRGREEN	HIF1A/VEGFA	0.99	±2.1
Saliva	Month	40	SYBRGREEN	VEGFA/ACTB HIF1A/ACTB	0.95	±3.0
Saliva	Month	15	TaqMan	ΔCq of miRNA 205	0.89	±4.4
Saliva	Month	15	TaqMan	ΔCq of RNU44	0.73	±4.7
Saliva	Month	15	TaqMan	Multiple R model	0.88	±5.5
Saliva	Month		TaqMan	miRNA 205/RNU44	0.9	±5.2
Semen	Month	30	TaqMan	PRM1/B2M	0.81	±4.7
Semen	Month	30	TaqMan	SEMG1/B2M	0.76	±4.4
Semen	Month	30	TaqMan	PRM1/SEMG1	0.87	±4.3
Semen	Month	30	TaqMan	PRM1/B2M SEMG1/B2M	0.9	±4.4
Semen	Month	30	SYBRGREEN	VEGFA/ACTB	0.66	±5.4
Semen	Month	30	SYBRGREEN	HIF1A/ACTB	0.67	±5.2
Semen	Month	30	SYBRGREEN	HIF1A/VEGFA	0.63	±5.3
Semen	Month	30	SYBRGREEN	HIF1A/VEGFA HIF1A/ACTB	0.66	±4.6
Semen(+DTT)	Month	15	TaqMan	miRNA891a/RNU44	0.81	±5.3
Semen(-DTT)	Month	15	TaqMan	miRNA891a/RNU44	0.55	±5.1

In order to obtain accurate results, some of the limitation factors need to be successfully overcome. Integrity and quality of samples are likely the most important factors, because both could be influenced the accuracy of normalisation strategy through reference genes. Consequently, the controlling for input amount of RNA makes the biggest difference. Quantification of extracted RNA via Nano-drop and checking of the quality of RNA samples by electrophoresis are the key initial steps in the process. Following this, all RNA samples should be diluted in the same concentration before cDNA synthesis is performed. Application of these steps would undoubtedly lead to more robust and reproducible RT-qPCR. In addition, the stability of the reference genes during all experiment stages is another potentially limiting factor in the procedure. There is no unique reference gene available; however, normalisation with more than one reference gene would increase the accuracy of normalisation. In this study, a geometric mean normalisation proved to be a useful procedure to improve the data obtaining.

The amplification efficiency of selected markers is another problem related to RQ method. This approach assumes a uniform PCR efficiency of 100% across chosen markers (Arocho *et al.*, 2006, Livak and Schmittgen, 2001). The presence of inhibitors or enhancers, the choice of RNA extraction methods, and the use of different reagents are the most common factors which that could influence the efficiency of PCR assay (Liu and Saint, 2002).

In the current study, some markers were successfully validated, and the slope of  $< 0.1$  allowed for the use of RQ experiment. However, the assumption for obtaining 100% PCR efficiency at all-time points became problematic, because variation in PCR efficiency could lead to the distortion of quantification results. For example, previous study showed that the difference in PCR efficiency as small as 0.04 would generated a 4-fold error in fold change (Ramakers *et al.*, 2003). Therefore, PCR efficiency between the target and the reference gene should be less than 5%, since increase in this value could result in miscalculation of

expression ratio by as high as 43.2% (Rao *et al.*, 2013). In general, different results obtained in the present study could be related to the difference in PCR efficiency at the different time points for the period of up to 28 days. In order to overcome this limitation, the validation experiment at different time points should be performed. Finally, in  $2^{-\Delta\Delta C_q}$  method, the subtraction of a background fluorescence is automatically performed by the qPCR software. This would mean that an imprecise background fluorescence calculation would lead to the distortion of the results. To improve RQ data analysis, a calculation of a corrected individual efficiency is required (Rao *et al.*, 2013).

Finally, the data obtained and the conclusions throughout the thesis are highly influenced by the statistical analysis methods applied. For example, many statistical analyses have been developed to test normality of the data. In present study, The Shapiro Wilk analysis was used, and the results showed that the symmetric bell-shaped distribution was not present in the observed data. As mentioned earlier, the normal distribution is not essential when dealing with a regression analysis for prediction models. However, the data must be normally distributed for the calculation of standard error (Guo *et al.*, 2010), therefore, big error bars presented as MAD in some models could be related to the non-normally distributed data. It is a common practice to use transformation, such as log-transformation in order to convert skewed distributed data into normally distributed. The problem with this approach in the present experiments is that, even if the transformation would be successful, it could be lead to incorrect interpretation of the experimental results. In fact, in some cases, applying the transformation made the distribution more skewed than the original data. This is why in some studies the authors recommended the use of the new analytic analysis that is not dependent on the distribution of that data such as generalised estimating equations (GEE) (Changyong *et al.*, 2014, Guo *et al.*, 2010).

### **7.1. Novel work**

The first novel aspect in this work was the generation of the mathematical models for age prediction of blood, saliva, and semen stains, based on the degradation patterns of mRNA markers using TaqMan and SYBR Green chemistries. The second novel aspect was the exploration of the patterns of miRNA markers as indicators for age prediction of blood, saliva, and semen samples. Further, the combination of mRNA and miRNA models were performed through geometric mean normalisation strategy.

### **7.2. Future work**

The sensitivity and stability of mRNA and miRNA markers are the most important factors in selecting the candidate for determining the age of the biological stains. This is because each candidate has a unique pattern of degradation. Consequently, continual mRNA and miRNA markers screenings are required. Also, the degradation patterns of mRNA and miRNA markers may differ dependent on several factors. The environmental factors such as temperature, light, humidity and the rain could influence the degradation patterns of selection markers. Therefore, further study under the different environmental conditions is warranted. Further, age, race, and the sex of the person from whom the sample was obtained could also influence the degradation pattern of chosen markers. Finally, the data presented in this thesis showed different degradation levels in the samples that were left to decompose for one week and one month, further supporting that the sensitivity and stability of selection markers requires more attention.

In this study, all samples under investigation were collected with sterile cotton swabs and filter paper which is currently not readily available at the real crime scene. Therefore, further exploration of the common deposition surfaces such as carpet, denim, and leather is required.

Geometric mean normalisation stood out as a method of choice to normalise multiple reference genes. Therefore, a combination of the ratio of mRNA and DNA at a different time points should be required to understand the relationship between more degraded and stable candidates.

All models in this study were generated using a singleplex RT-qPCR method. However, it would be interesting to develop multiplex PCR markers in one kit that would be able to predict the age and identification of blood, saliva, and semen in one reaction. This would reduce the amount of sample required and would also be more cost-effective and less time-consuming. Additionally, this multiplex PCR would be more adequate method for subsequent analysis by multiple regression models. However, the method for quantification of degraded mRNA markers via DNA analyser is now available, and this enables generation of DNA profiling, age prediction, and identification of body fluids, all in one run.

## **Chapter Eight: References**

- ACKERMANN, K., BALLANTYNE, K. N. & KAYSER, M. 2010. Estimating trace deposition time with circadian biomarkers: a prospective and versatile tool for crime scene reconstruction. *International Journal of Legal Medicine*, 124, 387-395.
- ADALSTEINSSON, B. T., GUDNASON, H., ASPELUND, T., HARRIS, T. B., LAUNER, L. J., EIRIKSDOTTIR, G., SMITH, A. V. & GUDNASON, V. 2012. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS one*, 7, e46705.
- ALMAC, E., BEZEMER, R., HILARIUS-STOKMAN, P. M., GOEDHART, P., KORTE, D., VERHOEVEN, A. J. & INCE, C. 2014. Red blood cell storage increases hypoxia-induced nitric oxide bioavailability and methemoglobin formation in vitro and *in vivo*. *Transfusion*, 54, 3178-3185.
- ALSHEHHI, S., MCCALLUM, N. A. & HADDRILL, P. R. 2017. Quantification of RNA degradation of blood-specific markers to indicate the age of bloodstains. *Forensic Science International: Genetics Supplement Series*, e453-e455.
- ALVAREZ, M., JUUSOLA, J. & BALLANTYNE, J. 2004. An mRNA and DNA co-isolation method for forensic casework samples. *Analytical Biochemistry*, 335, 289-298.
- ANDERSEN, C. L., JENSEN, J. L. & ØRNTOFT, T. F. 2004. Normalisation of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalisation, applied to bladder and colon cancer data sets. *Cancer Research*, 64, 5245-5250.
- ANDERSON, S., HOWARD, B., HOBBS, G. R. & BISHOP, C. P. 2005. A method for determining the age of a bloodstain. *Forensic Science International*, 148, 37-45.
- ANDERSON, S. E., HOBBS, G. R. & BISHOP, C. P. 2011. Multivariate analysis for estimating the age of a bloodstain. *Journal of Forensic Sciences*, 56, 186-193.
- ANDRASKO, J. 1997. The estimation of age of bloodstains by HPLC analysis. *Journal of Forensic Sciences*, 42, 601-607.
- ANTUNES, J., SILVA, D. S., BALAMURUGAN, K., DUNCAN, G., ALHO, C. S. & MCCORD, B. 2016. High-resolution melt analysis of DNA methylation to discriminate semen in biological stains. *Analytical Biochemistry*, 494, 40-45.
- ARANY, S. & OHTANI, S. 2011. Age estimation of bloodstains: A preliminary report based on aspartic acid racemization rate. *Forensic Science International*, 212, e36-e39.
- AROCHO, A., CHEN, B., LADANYI, M. & PAN, Q. 2006. Validation of the  $2^{-\Delta\Delta C_t}$  calculation as an alternate method of data analysis for quantitative PCR of BCR-ABL P210 transcripts. *Diagnostic Molecular Pathology*, 15, 56-61.
- BAI, X., PENG, D., LI, Z., TIAN, H., ZHANG, L., YANG, D., BAI, P. & LIANG, W. 2017. Postmortem interval (PMI) determination by profiling of HAF mRNA degradation using RT-qPCR. *Forensic Science International: Genetics Supplement Series*.
- BARBER, R. D., HARMER, D. W., COLEMAN, R. A. & CLARK, B. J. 2005. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics*, 21, 389-395.
- BATTAGLIA, M., PEDRAZZOLI, P., PALERMO, B., LANZA, A., BERTOLINI, F., GIBELLI, N., DA PRADA, G., ZAMBELLI, A., PEROTTI, C. & ROBUSTELLI, D. C. G. 1998. Epithelial tumour cell detection and the unsolved problems of nested RT-

- PCR: a new sensitive one step method without false positive results. *Bone Marrow Transplantation*, 22, 693-698.
- BAUER, M. 2007. RNA in forensic science. *Forensic Science International: Genetics*, 1, 69-74.
- BAUER, M., GRAMLICH, I., POLZIN, S. & PATZELT, D. 2003a. Quantification of mRNA degradation as possible indicator of postmortem interval—a pilot study. *Legal Medicine*, 5, 220-227.
- BAUER, M. & PATZELT, D. 2002. Evaluation of mRNA markers for the identification of menstrual blood. *Journal of Forensic Sciences*, 47, 1278-1282.
- BAUER, M. & PATZELT, D. 2003. A method for simultaneous RNA and DNA isolation from dried blood and semen stains. *Forensic Science International*, 136, 76-78.
- BAUER, M., POLZIN, S. & PATZELT, D. 2003b. Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Science International*, 138, 94-103.
- BENES, V. & CASTOLDI, M. 2010. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods*, 50, 244-249.
- BERNAUDIN, M., NEDELEC, A.-S., DIVOUX, D., MACKENZIE, E. T., PETIT, E. & SCHUMANN-BARD, P. 2002. Normobaric hypoxia induces tolerance to focal permanent cerebral ischemia in association with an increased expression of *hypoxia-inducible* factor-1 and its target genes, erythropoietin and VEGF, in the adult mouse brain. *Journal of Cerebral Blood Flow & Metabolism*, 22, 393-403.
- BIOSYSTEMS, A. 2004. Guide to performing relative quantitation of gene expression using real-time quantitative PCR. *Applied Biosystems, Foster City*, 28-30.
- BIOSYSTEMS, A. 2008. Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR.
- BRAVO, V., ROSETO, S., RICORDI, C. & PASTORI, R. L. 2007. Instability of miRNA and cDNAs derivatives in RNA preparations. *Biochemical and Biophysical Research Communications*, 353, 1052-1055.
- BREMMER, R. H., DE BRUIN, K. G., VAN GEMERT, M. J., VAN LEEUWEN, T. G. & AALDERS, M. C. 2012a. Forensic quest for age determination of bloodstains. *Forensic Science International*, 216, 1-11.
- BREMMER, R. H., DE BRUIN, K. G., VAN GEMERT, M. J. C., VAN LEEUWEN, T. G. & AALDERS, M. C. G. 2012b. Forensic quest for age determination of bloodstains. *Forensic Science International*, 216, 1-11.
- BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W. & SHIPLEY, G. L. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55, 611-622.
- BUSTIN, S. A. & NOLAN, T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of biomolecular techniques: JBT*, 15, 155.
- CHANGYONG, F., HONGYUE, W., NAIJI, L., TIAN, C., HUA, H. & YING, L. 2014. Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26, 105-109.

- CHAPMAN, R., BROWN, N. & KEATING, S. 1989. The isolation of spermatozoa from sexual assault swabs using Proteinase K. *Journal of the Forensic Science Society*, 29, 207-212.
- CHEN, C., RIDZON, D. A., BROOMER, A. J., ZHOU, Z., LEE, D. H., NGUYEN, J. T., BARBISIN, M., XU, N. L., MAHUVAKAR, V. R. & ANDERSEN, M. R. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33, e179-e179.
- CHEN, C. Y. A. & SHYU, A. B. 2011. Mechanisms of deadenylation-dependent decay. *Wiley Interdisciplinary Reviews: RNA*, 2, 167-183.
- CHERVONEVA, I., LI, Y., SCHULZ, S., CROKER, S., WILSON, C., WALDMAN, S. A. & HYSLOP, T. 2010. Selection of optimal reference genes for normalisation in quantitative RT-PCR. *BMC bioinformatics*, 11, 253.
- CONNOLLY, J.-A., CLARKE, D. & WILLIAMS, G. 2011. The use of forensic mRNA analysis to determine stain age. The Higher Education Academy UK Physical Sciences Centre in conjunction with Nottingham Trent University.
- CONNOLLY, J.-A. & WILLIAMS, G. 2011. Evaluating an mRNA based body fluid identification test using sybr green fluorescent dye and real-time PCR. *International Journal of Criminal Investigation*, 1, 177-185.
- COURTS, C. & MADEA, B. 2010. Micro-RNA—a potential for forensic science? *Forensic Science International*, 203, 106-111.
- COURTS, C. & MADEA, B. 2011. Specific Micro-RNA Signatures for the Detection of Saliva and Blood in Forensic Body-fluid Identification. *Journal of Forensic Sciences*, 56, 1464-1470.
- DE GANNES, M. K. 2014. Rapid method of processing sperm for nucleic acid extraction in clinical research. current. Masters Theses 9.
- DE KOK, J. B., ROELOFS, R. W., GIESENDORF, B. A., PENNING, J. L., WAAS, E. T., FEUTH, T., SWINKELS, D. W. & SPAN, P. N. 2005. Normalisation of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Laboratory Investigation*, 85, 154-159.
- DE PLANELL-SAGUER, M. & RODICIO, M. C. 2011. Analytical aspects of microRNA in diagnostics: a review. *Analytica Chimica Acta*, 699, 134-152.
- DHEDA, K., HUGGETT, J. F., BUSTIN, S. A., JOHNSON, M. A., ROOK, G. & ZUMLA, A. 2004. Validation of housekeeping genes for normalising RNA expression in real-time PCR. *Biotechniques*, 37, 112-119.
- DOLESHAL, M., MAGOTRA, A. A., CHOUDHURY, B., CANNON, B. D., LABOURIER, E. & SZAFRANSKA, A. E. 2008. Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues. *The Journal of Molecular Diagnostics*, 10, 203-211.
- DRAGAN, A., PAVLOVIC, R., MCGIVNEY, J., CASAS-FINET, J., BISHOP, E., STROUSE, R., SCHENERMAN, M. & GEDDES, C. 2012. SYBR Green I: fluorescence properties and interaction with DNA. *Journal of Fluorescence*, 22, 1189-1199.
- FANG, R., MANOHAR, C., SHULSE, C., BREVNOV, M., WONG, A., PETRAUSKENE, O., BRZOSKA, P. & FURTADO, M. Real-time PCR assays for the detection of

- tissue and body fluid specific mRNAs. International Congress Series, 2006. Elsevier, 685-687.
- FIORI, A. 1962. Detection and identification of bloodstains: methods of forensic sciences. *Interscience*, 1, 243-290.
- FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E. & MELLO, C. C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *nature*, 391, 806-811.
- FLEMING, J. T., YAO, W.-H. & SAYLER, G. S. 1998. Optimization of differential display of prokaryotic mRNA: application to pure culture and soil microcosms. *Applied and environmental microbiology*, 64, 3698-3706.
- FLEMING, R. I. & HARBISON, S. 2010. The development of a mRNA multiplex RT-PCR assay for the definitive identification of body fluids. *Forensic Science International: Genetics*, 4, 244-256.
- FUJITA, Y., TSUCHIYA, K., ABE, S., TAKIGUCHI, Y., KUBO, S.-I. & SAKURAI, H. 2005. Estimation of the age of human bloodstains by electron paramagnetic resonance spectroscopy: long-term controlled experiment on the effects of environmental factors. *Forensic Science International*, 152, 39-43.
- GILSBACH, R., KOUTA, M., BONISCH, H. & BRUSS, M. 2006. Comparison of *in vitro* and *in vivo* reference genes for internal standardization of real-time PCR data. *Biotechniques*, 40, 173-177.
- GINZINGER, D. G. 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Experimental hematology*, 30, 503-512.
- GÖRLACH, M., BURD, C. G. & DREYFUSS, G. 1994. The mRNA Poly(A)-Binding Protein: Localization, Abundance, and RNA-Binding Specificity. *Experimental Cell Research*, 211, 400-407.
- GRABMÜLLER, M., MADEA, B. & COURTS, C. 2015. Comparative evaluation of different extraction and quantification methods for forensic RNA analysis. *Forensic Science International: Genetics*, 16, 195-202.
- GREAT BRITAIN, O. O. N. S. 2014. Crime in England & Wales, Year Ending March 2014, United Kingdom. *Office of National Statistics*
- GRIMSON, A., FARH, K. K.-H., JOHNSTON, W. K., GARRETT-ENGELE, P., LIM, L. P. & BARTEL, D. P. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, 27, 91-105.
- GUHANIYOGI, J. & BREWER, G. 2001. Regulation of mRNA stability in mammalian cells. *Gene*, 265, 11-23.
- GUO, H., TAGUE, L., RAY, R. & TIGYI, G. 2010. A two-step approach to qPCR experimental design and software for data analysis. *BMC Bioinformatics*, 11, P18.
- GUO, K., ZHEGALOVA, N., ACHILEFU, S. & BEREZIN, M. Y. 2013. Bloodstain age analysis: toward solid state fluorescent lifetime measurements. Proceedings Volume 8572, Advanced Biomedical and Clinical Diagnostic Systems XI; 857214.
- GUTIÉRREZ-VÁZQUEZ, C., ENRIGHT, A. J., RODRÍGUEZ-GALÁN, A., PÉREZ-GARCÍA, A., COLLIER, P., JONES, M. R., BENES, V., MIZGERD, J. P.,

- MITTELBRUNN, M. & RAMIRO, A. R. 2017. 3' Uridylation controls mature microRNA turnover during CD4 T-cell activation. *RNA*, 23, 882-891.
- HA, M. & KIM, V. N. 2014. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15, 509-524.
- HAAS, C., HANSON, E., ANJOS, M., BANEMANN, R., BERTI, A., BORGES, E., CARRACEDO, A., CARVALHO, M., COURTS, C. & DE COCK, G. 2013. RNA/DNA co-analysis from human saliva and semen stains—results of a third collaborative EDNAP exercise. *Forensic Science International: Genetics*, 7, 230-239.
- HAAS, C., HANSON, E., BÄR, W., BANEMANN, R., BENTO, A., BERTI, A., BORGES, E., BOUAKAZE, C., CARRACEDO, A. & CARVALHO, M. 2011a. mRNA profiling for the identification of blood—results of a collaborative EDNAP exercise. *Forensic Science International: Genetics*, 5, 21-26.
- HAAS, C., HANSON, E., KRATZER, A., BÄR, W. & BALLANTYNE, J. 2011b. Selection of highly specific and sensitive mRNA biomarkers for the identification of blood. *Forensic Science International: Genetics*, 5, 449-458.
- HAAS, C., KLESSER, B., KRATZER, A. & BÄR, W. 2008. mRNA profiling for body fluid identification. *Forensic Science International: Genetics Supplement Series*, 1, 37-38.
- HAAS, C., KLESSER, B., MAAKE, C., BÄR, W. & KRATZER, A. 2009. mRNA profiling for body fluid identification by reverse transcription endpoint PCR and realtime PCR. *Forensic Science International: Genetics*, 3, 80-88.
- HAMPSON, C., LOUHELAINEN, J. & MCCOLL, S. 2011. An RNA Expression Method for Aging Forensic Hair Samples. *Journal of Forensic Sciences*, 56, 359-365.
- HAN, N., LU, H., ZHANG, Z., RUAN, M., YANG, W. & ZHANG, C. J. G. 2018. Comprehensive and in-depth analysis of microRNA and mRNA expression profile in salivary adenoid cystic carcinoma. *Gene*, 678, 349-360.
- HANSON, E. K. & BALLANTYNE, J. 2010. A blue spectral shift of the hemoglobin soret band correlates with the age (time since deposition) of dried bloodstains. *PLoS one*, 5, e12830.
- HANSON, E. K., LUBENOW, H. & BALLANTYNE, J. 2009. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Analytical Biochemistry*, 387, 303-314.
- HEID, C. A., STEVENS, J., LIVAK, K. J. & WILLIAMS, P. M. 1996. Real time quantitative PCR. *Genome Research*, 6, 986-994.
- HEINRICH, M., MATT, K., LUTZ-BONENGL, S. & SCHMIDT, U. 2007. Successful RNA extraction from various human postmortem tissues. *International Journal of Legal Medicine*, 121, 136-142.
- HOLLAND, P. M., ABRAMSON, R. D., WATSON, R. & GELFAND, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'---3'exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences*, 88, 7276-7280.
- HUANG, X., BAUMANN, M., NIKITINA, L., WENGER, F., SURBEK, D., KÖRNER, M. & ALBRECHT, C. 2013. RNA degradation differentially affects quantitative mRNA measurements of endogenous reference genes in human placenta. *Placenta*, 34, 544-547.

- HUGGETT, J., DHEDA, K., BUSTIN, S. & ZUMLA, A. 2005. Real-time RT-PCR normalisation; strategies and considerations. *Genes and Immunity*, 6, 279-284.
- HUTH, A., VENNEMANN, B., FRACASSO, T., LUTZ-BONENGL, S. & VENNEMANN, M. 2013. Apparent versus true gene expression changes of three hypoxia-related genes in autopsy derived tissue and the importance of normalisation. *International Journal of Legal Medicine*, 127, 335-344.
- HUTVÁGNER, G. & ZAMORE, P. D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 297, 2056-2060.
- IMAI, T., UBI, B. E., SAITO, T. & MORIGUCHI, T. 2014. Evaluation of reference genes for accurate normalisation of gene expression for real time-quantitative PCR in *Pyrus pyrifolia* using different tissue samples and seasonal conditions. *PLoS One*, 9, e86492.
- INOUE, H., TAKABE, F., IWASA, M. & MAENO, Y. 1991. Identification of fetal hemoglobin and simultaneous estimation of bloodstain age by high-performance liquid chromatography. *International Journal of Legal Medicine*, 104, 127-131.
- INOUE, H., TAKABE, F., IWASA, M., MAENO, Y. & SEKO, Y. 1992. A new marker for estimation of bloodstain age by high performance liquid chromatography. *Forensic Science International*, 57, 17-27.
- JEFFREYS, A. J. 1987. Highly variable minisatellites and DNA fingerprints. *Biochem. Soc. Trans*, 15, 309-317.
- JOBLING, M. A. & GILL, P. 2004. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, 5, 739-751.
- JUNG, M., SCHAEFER, A., STEINER, I., KEMPKENSTEFFEN, C., STEPHAN, C., ERBERSDOBLER, A. & JUNG, K. 2010. Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. *Clinical Chemistry*, 56, 998-1006.
- JUNG, U., JIANG, X., KAUFMANN, S. H. & PATZEL, V. 2013. A universal TaqMan-based RT-PCR protocol for cost-efficient detection of small noncoding RNA. *RNA*, 19, 1864-1873.
- JUUSOLA, J. & BALLANTYNE, J. 2003. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic Science International*, 135, 85-96.
- JUUSOLA, J. & BALLANTYNE, J. 2005. Multiplex mRNA profiling for the identification of body fluids. *Forensic Science International*, 152, 1-12.
- KANNAN, M. R., SRIDHARAN, J., LAKSHMI, N. A. & VELANGANNI, A. A. J. 2017. IDENTIFICATION AND CHARACTERIZATION OF EVOLUTIONARY CONSERVED miRNAs IN VARIOUS TYPES OF CANCER: IN SILICO AND EXPERIMENTAL APPROACH. *Life Science Informatics Publications*, 70.
- KAPITULNIK, J. 2004. Bilirubin: an endogenous product of heme degradation with both cytotoxic and cytoprotective properties. *Molecular pharmacology*, 66, 773-779.
- KIKUCHI, K., KAWAHARA, K.-I., BISWAS, K. K., ITO, T., TANCHAROEN, S., SHIOMI, N., KODA, Y., MATSUDA, F., MORIMOTO, Y. & OYAMA, Y. 2010. HMGB1: A new marker for estimation of the postmortem interval. *Experimental and Therapeutic Medicine*, 1, 109-111.

- KILDEMO, H. 2012. *RNA expression in sperm as markers of sperm-quality*. Master thesis in Toxicology.
- KIND, S., PATTERSON, D. & OWEN, G. 1972. Estimation of the age of dried blood stains by a spectrophotometric method. *Forensic Science*, 1, 27-54.
- KOHLMEIER, F. & SCHNEIDER, P. M. 2012. Successful mRNA profiling of 23 years old blood stains. *Forensic science international: Genetics*, 6, 274-276.
- KOZERA, B. & RAPACZ, M. 2013. Reference genes in real-time PCR. *Journal of Applied Genetics*, 54, 391-406.
- KOZOMARA, A. & GRIFFITHS-JONES, S. 2010. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*, 39, D152-D157.
- KOZOMARA, A. & GRIFFITHS-JONES, S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42, D68-D73.
- KROL, J., SOBCZAK, K., WILCZYNSKA, U., DRATH, M., JASINSKA, A., KACZYNSKA, D. & KRZYZOSIAK, W. J. 2004. Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *Journal of Biological Chemistry*, 279, 42230-42239.
- KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., FOROOTAN, A., JONÁK, J., LIND, K., SINDELKA, R., SJÖBACK, R., SJÖGREEN, B. & STRÖMBOM, L. 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27, 95-125.
- LAGOS-QUINTANA, M., RAUHUT, R., LENDECKEL, W. & TUSCHL, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science*, 294, 853-858.
- LANDGRAF, P., RUSU, M., SHERIDAN, R., SEWER, A., IOVINO, N., ARAVIN, A., PFEFFER, S., RICE, A., KAMPHORST, A. O. & LANDTHALER, M. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129, 1401-1414.
- LECH, K., ACKERMANN, K., WOLLSTEIN, A., REVELL, V. L., SKENE, D. J. & KAYSER, M. 2014. Assessing the suitability of miRNA-142-5p and miRNA-541 for bloodstain deposition timing. *Forensic Science International: Genetics*, 12, 181-184.
- LEE, R. C., FEINBAUM, R. L. & AMBROS, V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75, 843-854.
- LEE, Y., KIM, M., HAN, J., YEOM, K. H., LEE, S., BAEK, S. H. & KIM, V. N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, 23, 4051-4060.
- LEERS, O. 1910. Modern Medicine. *The Journal of the American Medical Association*, 55, 151.
- LEITAO, M. D. C. G., COIMBRA, E. C., DE LIMA, R. D. C. P., DE LIMA GUIMARAES, M., DE ANDRADE HERACLIO, S., NETO, J. D. C. S. & DE FREITAS, A. C. 2014. Quantifying mRNA and MicroRNA with qPCR in cervical carcinogenesis: a validation of reference genes to ensure accurate data. *PloS one*, 9, e111021.
- LI, C., ZHAO, S., ZHANG, N., ZHANG, S. & HOU, Y. 2013a. Differences of DNA methylation profiles between monozygotic twins' blood samples. *Molecular Biology Reports*, 40, 5275-5280.

- LI, W.-C., MA, K.-J., LV, Y.-H., ZHANG, P., PAN, H., ZHANG, H., WANG, H.-J., MA, D. & CHEN, L. 2014. Postmortem interval determination using 18S-rRNA and microRNA. *Science & Justice*, 54, 307-310.
- LI, X., WONG, W., LAMOUREUX, E. L. & WONG, T. Y. 2012. Are linear regression techniques appropriate for analysis when the dependent (outcome) variable is not normally distributed? *Investigative Ophthalmology and Visual science*, 53, 3082-3083.
- LI, Y., LI, Z., ZHOU, S., WEN, J., GENG, B., YANG, J. & CUI, Q. J. B. R. I. 2013b. Genome-wide analysis of human microRNA stability. *BioMed Research International*, Article ID 368975, 2013, 12 pages.
- LINDENBERGH, A., DE PAGTER, M., RAMDAYAL, G., VISSER, M., ZUBAKOV, D., KAYSER, M. & SIJEN, T. 2012b. A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Science International: Genetics*, 6, 565-577.
- LINDENBERGH, A., MAASKANT, P. & SIJEN, T. 2013. Implementation of RNA profiling in forensic casework. *Forensic Science International: Genetics*, 7, 159-166.
- LIU, L., SHU, X., REN, L., ZHOU, H., LI, Y., LIU, W., ZHU, C. & LIU, L. 2007. Determination of the early time of death by computerized image analysis of DNA degradation: which is the best quantitative indicator of DNA degradation? *Journal of Huazhong University of Science and Technology-Medical Sciences*, 27, 362-366.
- LIU, W. & SAINT, D. A. 2002. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Analytical Biochemistry*, 302, 52-59.
- LIVAK, K. J., FLOOD, S., MARMARO, J., GIUSTI, W. & DEETZ, K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Research*, 4, 357-362.
- LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta C_T}$  Method. *Methods*, 25, 402-408.
- LUND, M. N., HEINONEN, M., BARON, C. P. & ESTEVEZ, M. 2011. Protein oxidation in muscle foods: A review. *Molecular Nutrition & Food Research*, 55, 83-95.
- LV, Y.-H., MA, J.-L., PAN, H., ZHANG, H., LI, W.-C., XUE, A.-M., WANG, H.-J., MA, K.-J. & CHEN, L. 2016. RNA degradation as described by a mathematical model for postmortem interval determination. *Journal of Forensic and Legal Medicine*, 44, 43-52.
- LV, Y. H., MA, K. J., ZHANG, H., HE, M., ZHANG, P., SHEN, Y. W., JIANG, N., MA, D. & CHEN, L. 2014. A time course study demonstrating mRNA, microRNA, 18S rRNA, and U6 snRNA changes to estimate PMI in deceased rat's spleen. *Journal of Forensic Sciences*, 59, 1286-1294.
- MARZI, M. J., GHINI, F., CERRUTI, B., DE PRETIS, S., BONETTI, P., GIACOMELLI, C., GORSKI, M. M., KRESS, T., PELIZZOLA, M. & MULLER, H. 2016. Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *GenomeResearch*, 26, 554-565.

- MATSUOKA, T., TAGUCHI, T. & OKUDA, J. 1995. Estimation of bloodstain age by rapid determinations of oxyhemoglobin by use of oxygen electrode and total hemoglobin. *Biological and Pharmaceutical Bulletin*, 18, 1031-1035.
- MCMANUS, J., CHENG, Z. & VOGEL, C. 2015. Next-generation analysis of gene expression regulation—comparing the roles of synthesis and degradation. *Molecular BioSystems*, 11, 2680-2689.
- MEER, D., UCHIMOTO, M. L. & WILLIAMS, G. 2013. Simultaneous Analysis of Micro-RNA and DNA for Determining the Body Fluid Origin of DNA Profiles. *Journal of Forensic Sciences*, 58, 967-971.
- MEYER, S., TEMME, C. & WAHLE, E. 2004. Messenger RNA turnover in eukaryotes: pathways and enzymes. *Critical Reviews in Biochemistry and Molecular biology*, 39, 197-216.
- MIKI, T., KAI, A. & IKEYA, M. 1987. Electron spin resonance of bloodstains and its application to the estimation of time after bleeding. *Forensic Science International*, 35, 149-158.
- MONTGOMERY, M. K., XU, S. & FIRE, A. 1998. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 95, 15502-15507.
- MORENO, L. I., TATE, C. M., KNOTT, E. L., MCDANIEL, J. E., ROGERS, S. S., KOONS, B. W., KAVLICK, M. F., CRAIG, R. L. & ROBERTSON, J. M. 2012. Determination of an effective housekeeping gene for the quantification of mRNA for forensic applications. *Journal of Forensic Sciences*, 57, 1051-1058.
- MORRISON, T. B., WEIS, J. J. & WITWER, C. T. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques*, 24, 954-8, 960-962.
- MORSE, D. L., CARROLL, D., WEBER, L., BORGSTROM, M. C., RANGER-MOORE, J. & GILLIES, R. J. 2005. Determining suitable internal standards for mRNA quantification of increasing cancer progression in human breast cells by real-time reverse transcriptase polymerase chain reaction. *Analytical Biochemistry*, 342, 69-77.
- MORTA, W. 2012. A Study on Nucleic Acid Degradation in Drying and Dried Bloodstains as a Means to Determine the Time Since Deposition. PhD thesis. University of Auckland.
- MOSCHENROSS, D. M. 2011. Characterization of a liver specific microRNA: miR-122.
- MOTULSKY, H. & CHRISTOPOULOS, A. J. I., SAN DIEGO 2003. Fitting models to biological data using linear and nonlinear regression. GraphPad Software Inc, San Diego, CA. 351 pp.
- MRAZ, M., MALINOVA, K., MAYER, J. & POSPISILOVA, S. 2009. MicroRNA isolation and stability in stored RNA samples. *Biochemical and Biophysical Research Communications*, 390, 1-4.
- MUKAKA, M. 2012. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24, 69-71.
- MULLIS, K. B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American*, 262, 56-61.

- NAKANISHI, H., HARA, M., TAKAHASHI, S., TAKADA, A. & SAITO, K. 2014. Evaluation of forensic examination of extremely aged seminal stains. *Legal Medicine*, 16, 303-307.
- NAKAO, K.-I., SHIMADA, R., HARA, K. & KIBAYASHI, K. 2013. Experimental study on age estimation of bloodstains based on biological and toxicological analysis. *Open Forensic Science Journal*, 6, 6-11.
- NANGIA, V., JONAS, J. B., SINHA, A., GUPTA, R. & AGARWAL, S. 2011. Visual acuity and associated factors. The Central India eye and medical study. *PloS one*, 6, e22756.
- NAPOLI, C., LEMIEUX, C. & JORGENSEN, R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The plant cell*, 2, 279-289.
- NAVARRO, E., SERRANO-HERAS, G., CASTAÑO, M. J. & SOLERA, J. 2015. Real-time PCR detection chemistry. *Clinica Chimica Acta*, 439, 231-250.
- NUSSBAUMER, C., GHAREHBAGHI-SCHNELL, E. & KORSCHINECK, I. 2006. Messenger RNA profiling: a novel method for body fluid identification by real-time PCR. *Forensic Science International*, 157, 181-186.
- NYGARD, A.-B., JØRGENSEN, C. B., CIRERA, S. & FREDHOLM, M. 2007. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. *BMC molecular Biology*, 8, 67.
- OHL, F., JUNG, M., XU, C., STEPHAN, C., RABIEN, A., BURKHARDT, M., NITSCHKE, A., KRISTIANSEN, G., LOENING, S. A. & RADONIĆ, A. 2005. Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalisation? *Journal of Molecular Medicine*, 83, 1014-1024.
- OKELLO, J., RODRIGUEZ, L., POINAR, D., BOS, K., OKWI, A. L., BIMENYA, G. S., SEWANKAMBO, N. K., HENRY, K. R., KUCH, M. & POINAR, H. N. 2010. Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. *PLoS one*, 5, e13931-e13931.
- OMELIA, E. J., UCHIMOTO, M. L. & WILLIAMS, G. 2013. Quantitative PCR analysis of blood-and saliva-specific microRNA markers following solid-phase DNA extraction. *Analytical Biochemistry*, 435, 120-122.
- ORPHANOU, C. M. R. 2015. A bioanalytical approach to forensic body fluid identification & age determination. PhD Thesis. Staffordshire University.
- OWEN, D. H. & KATZ, D. F. 2005. A review of the physical and chemical properties of human semen and the formulation of a semen simulant. *Journal of Andrology*, 26, 459-469.
- PALMERO, E. I., DE CAMPOS, S. G. P., CAMPOS, M., SOUZA, N. C., GUERREIRO, I. D. C., CARVALHO, A. L. & MARQUES, M. M. C. 2011. Mechanisms and role of microRNA deregulation in cancer onset and progression. *Genetics and Molecular Biology*, 34, 363-370.
- PARKER, C. 2011. The Relative Recoverability Of DNA and RNA profiles from forensically relevant body fluid stains. Master thesis. University of Central Florida Orlando, Florida 165.

- PARRISH, S., FLEENOR, J., XU, S., MELLO, C. & FIRE, A. 2000. Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Molecular Cell*, 6, 1077-1087.
- PARTEMI, S., BERNE, P. M., BATLLE, M., BERRUEZO, A., MONT, L., RIURÓ, H., ORTIZ, J. T., ROIG, E., PASCALI, V. L. & BRUGADA, R. 2010. Analysis of mRNA from human heart tissue and putative applications in forensic molecular pathology. *Forensic Science International*, 203, 99-105.
- PATNAIK, S. K., MALLICK, R. & YENDAMURI, S. 2010. Detection of microRNAs in dried serum blots. *Analytical Biochemistry*, 407, 147-149.
- PATTERSON, D. J. N. 1960. Use of reflectance measurements in assessing the colour changes of ageing bloodstains. *Nature*, 187, 688-689.
- PAULDING, W. R. & CZYZYK-KRZESKA, M. F. 2002. Hypoxia-induced regulation of mRNA stability. *Oxygen Sensing*. Springer.
- PELTIER, H. J. & LATHAM, G. J. 2008. Normalisation of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *Rna*, 14, 844-852.
- PETRICCIONE, M., MASTROBUONI, F., ZAMPELLA, L. & SCORTICHINI, M. 2015. Reference gene selection for normalisation of RT-qPCR gene expression data from *Actinidia deliciosa* leaves infected with *Pseudomonas syringae* pv. *actinidiae*. *Scientific Reports*, 5, 16961.
- PFAFFL, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29, e45-e45.
- PFAFFL, M. W., TICHOPAD, A., PRGOMET, C. & NEUVIANS, T. P. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnology Letters*, 26, 509-515.
- PIECHACZYK, M., BLANCHARD, J., MARTY, L., DANI, C., PANABIERES, F., EL SABOUTY, S., FORT, P. & JEANTEUR, P. 1984. Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. *Nucleic Acids Research*, 12, 6951-6963.
- PORRUA, O., BOUDVILLAIN, M. & LIBRI, D. 2016. Transcription termination: variations on common themes. *Trends in Genetics*, 32, 508-522.
- QI, B., KONG, L. & LU, Y. 2013. Gender-related difference in bloodstain RNA ratio stored under uncontrolled room conditions for 28 days. *Journal of Forensic and Legal Medicine*, 20, 321-325.
- RAJAMANNAR, K. 1977. Determination of the age of bloodstains using immunoelectrophoresis. *Journal of Forensic Sciences*, 22, 159-164.
- RAMAKERS, C., RUIJTER, J. M., DEPREZ, R. H. L. & MOORMAN, A. F. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters*, 339, 62-66.
- RAO, X., HUANG, X., ZHOU, Z. & LIN, X. 2013. An improvement of the  $2^{-\Delta\Delta CT}$  method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, 3, 71-85.

- REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. & RUVKUN, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403, 901-906.
- RICHARD, M. L. L., HARPER, K. A., CRAIG, R. L., ONORATO, A. J., ROBERTSON, J. M. & DONFACK, J. 2012. Evaluation of mRNA marker specificity for the identification of five human body fluids by capillary electrophoresis. *Forensic Science International: Genetics*, 6, 452-460.
- RIEMER, A. B., KESKIN, D. B. & REINHERZ, E. L. 2012. Identification and validation of reference genes for expression studies in human keratinocyte cell lines treated with and without interferon- $\gamma$ —a method for qRT-PCR reference gene determination. *Experimental Dermatology*, 21, 625-629.
- ROEDER, A. D. & HAAS, C. 2013. mRNA profiling using a minimum of five mRNA markers per body fluid and a novel scoring method for body fluid identification. *International Journal of Legal Medicine*, 127, 707-721.
- ROEDER, A. D. & HAAS, C. 2016. Body Fluid Identification Using mRNA Profiling. *Forensic DNA Typing Protocols*, 13-31.
- ROSS, J. 1995. mRNA stability in mammalian cells. *Microbiological reviews*, 59, 423-450.
- ROY, B. & JACOBSON, A. 2013. The intimate relationships of mRNA decay and translation. *Trends in Genetics*, 29, 691-699.
- RÜEGGER, S. & GROßHANS, H. 2012. MicroRNA turnover: when, how, and why. *Trends in Biochemical Sciences*, 37, 436-446.
- SACHS, A. 1990. The role of poly(A) in the translation and stability of mRNA. *Current Opinion in Cell Biology*, 2, 1092-1098.
- SAKURADA, K., AKUTSU, T., WATANABE, K., FUJINAMI, Y. & YOSHINO, M. 2011. Expression of statherin mRNA and protein in nasal and vaginal secretions. *Legal Medicine*, 13, 309-313.
- SAMPAIO-SILVA, F., MAGALHÃES, T., CARVALHO, F., DINIS-OLIVEIRA, R. J. & SILVESTRE, R. 2013. Profiling of RNA degradation for estimation of post mortem interval. *PLoS One*, 8, e56507.
- SAND, M., GAMBICHLER, T., SAND, D., SKRYGAN, M., ALTMAYER, P. & BECHARA, F. G. 2009. MicroRNAs and the skin: tiny players in the body's largest organ. *Journal of Dermatological Science*, 53, 169-175.
- SANTOS, J. I., TEIXEIRA, A. L., DIAS, F., GOMES, M., NOGUEIRA, A., ASSIS, J. & MEDEIROS, R. 2014. Restoring TGF $\beta$ 1 pathway-related microRNAs: possible impact in metastatic prostate cancer development. *Tumor Biology*, 35, 6245-6253.
- SCHMITTGEN, T. D., ZAKRAJSEK, B. A., MILLS, A. G., GORN, V., SINGER, M. J. & REED, M. W. 2000. Quantitative reverse transcription–polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Analytical Biochemistry*, 285, 194-204.
- SCHOENBERG, D. R. & MAQUAT, L. E. 2012. Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics*, 13, 246-259.

- SCHOTSMANS, E. M., NICHOLAS, M. & FORBES, S. L. 2017. Taphonomy of human remains: forensic analysis of the dead and the depositional environment, John Wiley & Sons.
- SCHWARZ, F. 1937. Quantitative Untersuchungen der Katalase und Peroxydase im Blutfleck. *International Journal of Legal Medicine*, 27, 1-34.
- SCHWARZACHER 1930. Determination of the age of bloodstains. *The American Journal of Police Science* 1, 374-380.
- SENSABAUGH, G., WILSON, A. & KIRK, P. 1971. Protein Stability in Preserved Biological Remains. I. Survival of Biologically Active Proteins in a 8-year-old Sample of Dried Blood. *International Journal of Biochemistry*, 2, 545-557.
- SETHI, P. & LUKIW, W. J. 2009. Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex. *Neuroscience letters*, 459, 100-104.
- SETZER, M., JUUSOLA, J. & BALLANTYNE, J. 2008. Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. *Journal of Forensic Sciences*, 53, 296-305.
- SHANDILYA, J. & ROBERTS, S. G. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819, 391-400.
- SHAPIRO, S. S. & WILK, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591-611.
- SHERWIN, J. C., KELLY, J., HEWITT, A. W., KEARNS, L. S., GRIFFITHS, L. R. & MACKEY, D. A. 2011. Prevalence and predictors of refractive error in a genetically isolated population: the Norfolk Island Eye Study. *Clinical & Experimental Ophthalmology*, 39, 734-742.
- SILVA, F. P. S. 2012. RNA and microRNA profiling for the determination of post mortem interval. Thesis <https://repositorio.cespu.pt/handle/20.500.11816/201>.
- SILVA, S. S., LOPES, C., TEIXEIRA, A., DE SOUSA, M. C. & MEDEIROS, R. 2015. Forensic miRNA: potential biomarker for body fluids? *Forensic Science International: Genetics*, 14, 1-10.
- SILVER, N., BEST, S., JIANG, J. & THEIN, S. L. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, 7, 33.
- SIMARD, A.-M., DESGROSEILLERS, L. & SARAFIAN, V. 2011. Assessment of RNA stability for age determination of body fluid stains in forensic biology. *The FASEB Journal*, 25, 705.7-705.7.
- SIMARD, A.-M., DESGROSEILLERS, L. & SARAFIAN, V. 2012. Assessment of RNA Stability for Age Determination of Body Fluid Stains. *Canadian Society of Forensic Science Journal*, 45, 179-194.
- SIRKER, M., SCHNEIDER, P. M. & GOMES, I. 2016. A 17-month time course study of human RNA and DNA degradation in body fluids under dry and humid environmental conditions. *International Journal of Legal Medicine*, 130, 1431-1438.

- SKERN, R., FROST, P. & NILSEN, F. 2005. Relative transcript quantification by quantitative PCR: roughly right or precisely wrong? *BMC Molecular Biology*, 6, 10.
- SMITH, T. L. 2010. Investigating the potential of RNA to be used in forensic casework analysis, West Virginia University.
- STÅHLBERG, A., HÅKANSSON, J., XIAN, X., SEMB, H. & KUBISTA, M. 2004. Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry*, 50, 509-515.
- STEWART, L., EVANS, N., BEXON, K. J., VAN DER MEER, D. J. & WILLIAMS, G. A. 2015. Differentiating between monozygotic twins through DNA methylation-specific high-resolution melt curve analysis. *Analytical Biochemistry*, 476, 36-39.
- STRASSER, S., ZINK, A., KADA, G., HINTERDORFER, P., PESCHEL, O., HECKL, W. M., NERLICH, A. G. & THALHAMMER, S. 2007. Age determination of blood spots in forensic medicine by force spectroscopy. *Forensic Science International*, 170, 8-14.
- SU, C.-W. 2017. The applications of RNA analysis on identification of body fluids and dating plucked hairs. PhD thesis. King's College London.
- SUN, W., JULIE LI, Y.-S., HUANG, H.-D., SHYY, J. Y. & CHIEN, S. 2010. microRNA: a master regulator of cellular processes for bioengineering systems. *Annual Review of Biomedical Engineering*, 12, 1-27.
- SVINGEN, T., JØRGENSEN, A. & RAJPERT-DE MEYTS, E. 2014. Validation of endogenous normalising genes for expression analyses in adult human testis and germ cell neoplasms. *Molecular Human Reproduction*, 20, 709-718.
- TAN, S. C., CARR, C. A., YEOH, K. K., SCHOFIELD, C. J., DAVIES, K. E. & CLARKE, K. 2012. Identification of valid housekeeping genes for quantitative RT-PCR analysis of cardiosphere-derived cells preconditioned under hypoxia or with prolyl-4-hydroxylase inhibitors. *Molecular Biology Reports*, 39, 4857-4867.
- TANAKA, S., FURUKAWA, T. & PLOTKIN, S. A. 1975. Human cytomegalovirus stimulates host cell RNA synthesis. *Journal of Virology*, 15, 297-304.
- TEARE, J., ISLAM, R., FLANAGAN, R., GALLAGHER, S., DAVIES, M. & GRABAU, C. 1997. Measurement of nucleic acid concentrations using the DyNA Quant and the GeneQuant. *Biotechniques*, 22, 1170-1174.
- TECHNOLOGIES, L. 2011. TaqMan Small RNA Assays Protocol.
- THAIK-OO, M., TANAKA, E., TSUCHIYA, T., KOMINATO, Y., HONDA, K., YAMAZAKI, K. & MISAWA, S. 2002a. Estimation of postmortem interval from hypoxic inducible levels of vascular endothelial growth factor. *Journal of Forensic Science*, 47, 186-189.
- THAIK-OO, M., TANAKA, E., TSUCHIYA, T., KOMINATO, Y., HONDA, K., YAMAZAKI, K. & MISAWA, S. 2002b. Estimation of postmortem interval from hypoxic inducible levels of vascular endothelial growth factor. *Journal of Forensic Sciences*, 47, 186-189.
- THELLIN, O., ZORZI, W., LAKAYE, B., DE BORMAN, B., COUMANS, B., HENNEN, G., GRISAR, T., IGOUT, A. & HEINEN, E. 1999. Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology*, 75, 291-295.

- TIWARI, M. 2011. Science behind human saliva. *Journal of Natural Science, Biology, and Medicine*, 2, 53-58.
- TOMELLINI, L. 1907. De l'emploi d'une table chromatique pour les taches du sang. *Archives d'antropologie Criminelle de Criminologie*, 22, 580-582.
- UCHIMOTO, M. L. 2014. Developing a microRNA body fluid identification test for use in forensic casework. PhD thesis. University of Huddersfield.
- VAN DER KROL, A. R., MUR, L. A., BELD, M., MOL, J. & STUITJE, A. R. 1990. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *The Plant Cell*, 2, 291-299.
- VANDENBERG, N. & OORSCHOT, R. A. 2006. The Use of Polilight® in the Detection of Seminal Fluid, Saliva, and Bloodstains and Comparison with Conventional Chemical-Based Screening Tests. *Journal of Forensic Sciences*, 51, 361-370.
- VANDESOMPELE, J., DE PRETER, K., PATTYN, F., POPPE, B., VAN ROY, N., DE PAEPE, A. & SPELEMAN, F. 2002. Accurate normalisation of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3, research0034. 1.
- VISSER, M., ZUBAKOV, D., BALLANTYNE, K. N. & KAYSER, M. 2011. mRNA-based skin identification for forensic applications. *International Journal of Legal Medicine*, 125, 253-263.
- WAN, Y., KERTESZ, M., SPITALE, R. C., SEGAL, E. & CHANG, H. Y. 2011. Understanding the transcriptome through RNA structure. *Nature Reviews Genetics*, 12, 641.
- WANG, H., MAO, J., LI, Y., LUO, H., WU, J., LIAO, M., LIANG, W. & ZHANG, L. 2013a. 5 miRNA expression analyze in post-mortem interval (PMI) within 48h. *Forensic Science International: Genetics Supplement Series*, 4, e190-e191.
- WANG, W.-X., WILFRED, B. R., BALDWIN, D. A., ISETT, R. B., REN, N., STROMBERG, A. & NELSON, P. T. 2008. Focus on RNA isolation: obtaining RNA for microRNA (miRNA) expression profiling analyses of neural tissue. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1779, 749-757.
- WANG, Z., LUO, H., PAN, X., LIAO, M. & HOU, Y. 2012. A model for data analysis of microRNA expression in forensic body fluid identification. *Forensic Science International: Genetics*, 6, 419-423.
- WANG, Z., ZHANG, S., DI, Z., ZHAO, S. & LI, C. 2013b. Messenger RNA profiling for forensic body fluid identification: research and applications. *Fa yi xue za zhi*, 29, 368-374.
- WEBER, J. A., BAXTER, D. H., ZHANG, S., HUANG, D. Y., HUANG, K. H., LEE, M. J., GALAS, D. J. & WANG, K. 2010. The microRNA spectrum in 12 body fluids. *Clinical Chemistry*, 56, 1733-1741.
- WEI, Z., LUO, L., START, M. & GAO, L. 2016. Uncertainty Characterization and Quantification in Product Validation and Reliability Demonstration. SAE Technical Paper.
- WILLIAMS, G., CONNOLLY, J.-A., OMELIA, E., BEASLEY, E. & GADUZO, D. 2013. Characterising degradation profiles of RNA molecules within blood stains: Pilot studies. University of Huddersfield, scientific poster.

- WINTER, J. & DIEDERICHS, S. 2011. Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biology*, 8, 1149-1157.
- WITTEWER, C. T., HERRMANN, M. G., MOSS, A. A. & RASMUSSEN, R. P. 1997. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22, 130-139.
- XIA, X., HUO, W., WAN, R., XIA, X., DU, Q. & CHANG, Z. 2017. Identification of housekeeping genes as references for quantitative real-time RT-PCR analysis in *Misgurnus anguillicaudatus*. *Journal of Genetics*, 96, 895-904.
- XU, Y., XIE, J., CAO, Y., ZHOU, H., PING, Y., CHEN, L., GU, L., HU, W., BI, G. & GE, J. 2014. Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. *PloS one*, 9, e100123.
- YIN, J. L., SHACKEL, N. A., ZEKRY, A., MCGUINNESS, P. H., RICHARDS, C., VAN DER PUTTEN, K., MCCAUGHAN, G. W., ERIS, J. M. & BISHOP, G. A. 2001. Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I. *Immunology and Cell Biology*, 79, 213-221.
- ZHANG, J. & BYRNE, C. D. 1999. Differential priming of RNA templates during cDNA synthesis markedly affects both accuracy and reproducibility of quantitative competitive reverse-transcriptase PCR. *Biochemical Journal*, 337, 231-241.
- ZHANG, X., DING, L. & SANDFORD, A. J. 2005. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. *BMC Molecular Biology*, 6, 4.
- ZHANG, Z., QIN, Y. W., BREWER, G. & JING, Q. 2012. MicroRNA degradation and turnover: regulating the regulators. *Wiley Interdisciplinary Reviews: RNA*, 3, 593-600.
- ZHAO, D., ZHU, B.-L., ISHIKAWA, T., QUAN, L., LI, D.-R. & MAEDA, H. 2006. Real-time RT-PCR quantitative assays and postmortem degradation profiles of erythropoietin, vascular endothelial growth factor and hypoxia-inducible factor 1 alpha mRNA transcripts in forensic autopsy materials. *Legal Medicine*, 8, 132-136.
- ZHAO, H., WANG, C., YAO, L., LIN, Q., XU, X., HU, L. & LI, W. 2017. Identification of aged bloodstains through mRNA profiling: Experiments results on selected markers of 30-and 50-year-old samples. *Forensic Science International*, 272, e1-e6.
- ZHENG, Y., LAMOUREUX, E. L., CHIANG, P. P.-C., CHENG, C.-Y., ANUAR, A. R., SAW, S.-M., AUNG, T. & WONG, T. Y. 2011. Literacy is an independent risk factor for vision impairment and poor visual functioning. *Investigative Ophthalmology & Visual Science*, 52, 7634-7639.
- ZUBAKOV, D., BOERSMA, A. W., CHOI, Y., VAN KUIJK, P. F., WIEMER, E. A. & KAYSER, M. 2010. MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *International Journal of Legal Medicine*, 124, 217-226.
- ZUBAKOV, D., HANEKAMP, E., KOKSHOORN, M., VAN IJCKEN, W. & KAYSER, M. 2008. Stable RNA markers for identification of blood and saliva stains revealed from

whole genome expression analysis of time-wise degraded samples. *International Journal of Legal Med*, 122, 135-42.

ZUBAKOV, D., KOKSHOORN, M., KLOOSTERMAN, A. & KAYSER, M. 2009. New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. *International Journal of Legal Medicine*, 123, 71-74.

### Appendix 1. Age prediction

Table 31. The mean age prediction in blood using TaqMan chemistry among 1 week.

Actual age	ALAS2/GAPDH	SD	ALAS2/B2M	SD	B2M/GAPDH	SD	Multiple regression model	SD
0	1.86	0.00	0.23	0.00	0.24	0.00	0.54	0.00
1	3.19	0.91	3.85	0.38	3.78	0.45	3.33	0.70
2	3.41	0.85	3.10	1.14	2.56	1.48	2.72	1.00
3	4.15	0.97	3.96	0.89	3.96	0.92	4.20	1.44
4	3.82	0.76	4.18	0.60	4.46	0.38	4.16	0.60
5	3.92	1.00	4.21	0.65	4.41	0.55	4.18	0.95
6	3.62	1.11	4.56	0.32	4.47	0.34	4.67	0.99
7	4.05	1.07	3.91	0.57	4.09	0.37	4.20	0.86
MAD	±1.7		±1.4		±1.4		±1.4	

Table 32. The mean age prediction in semen (-DTT) using TaqMan.

Semen model (-DTT)						
Actual age	PRM1/B2M	SD	SEMG1/B2M	SD	PRM1/SEMG1	AS
0	0.01	0.00	2.52	0.00	8.83	0.00
7	17.63	0.06	14.98	5.43	16.54	1.55
14	17.23	0.77	13.12	6.62	15.36	3.47
21	17.55	0.23	19.95	3.60	15.06	3.10
28	17.57	0.10	19.52	3.59	14.24	3.93
MAD	±5.6		±5.4		±8.2	

Table 33. The mean age prediction using TaqMan chemistry in blood, saliva, and semen among 1 month.

Actual age	Blood models								Semen models						Saliva models					
	ALAS2/B2M	SD	HBB/B2M	SD	ALAS2/B2M HBB/B2M	SD	GYPA/B2M	SD	PRM1/B2M	SD	SEMG1/B2M	SD	PRM1/SEMG1	SD	PRM1/B2M SEMG1/B2M	SD	HTN3/GAPDH	SD	Ct HTN3	SD
	0.00	0.69	0.00	0.61	0.00	-0.79	0.18	9.18	0.00	1.04	0.00	0.29	0.00	1.46	0.00	1.75	0.00	2.19	0.00	2.32
7.00	9.41	5.51	7.94	4.58	7.60	3.68	16.36	0.27	12.42	5.70	12.94	4.87	9.52	6.96	10.12	4.92	17.03	5.16	8.09	4.63
14.00	18.95	3.04	18.52	4.15	17.62	3.16	16.25	1.95	17.69	2.76	17.56	4.87	17.82	4.04	17.07	6.08	16.27	2.29	15.66	6.81
21.00	19.07	3.51	20.68	2.66	19.00	1.82	12.75	3.51	18.72	5.46	20.00	1.25	20.65	2.82	20.26	4.13	16.52	2.95	20.13	1.77
28.00	21.97	2.98	22.84	2.77	21.68	1.60	14.78	3.05	20.22	3.44	19.23	3.21	20.56	2.57	20.86	4.38	17.97	2.98	23.93	3.77
MAD	±3.8		±3.2		±3.4		±8.1		±4.7		±4.4		±4.3		±4.4		±5.8		±3.3	

Table 34. The mean age prediction in blood using SYBRGREEN chemistry among 1 month.

Actual age	Blood models													
	HBB/ACTB	SD	EPOS1/ACTB	SD	FN1/ACTB	SD	FN1/EPOS1	SD	FGB/ACTB	SD	COA/ACTB	SD	FGB/COA	SD
0	0.34	0.00	-0.16	0.00	1.35	0.00	-0.49	0.00	0.21	0.00	0.20	0.00	4.17	0.00
7	10.28	6.00	15.76	3.76	13.69	5.42	16.00	4.28	12.19	4.55	15.84	3.90	15.70	3.76
14	15.83	3.40	17.82	2.85	17.75	5.58	17.20	0.62	16.03	6.28	17.65	2.87	13.48	3.44
21	22.75	1.99	17.82	2.85	17.62	5.79	19.20	0.19	19.91	3.80	18.53	2.10	14.56	8.76
28	20.81	2.75	18.85	1.26	20.41	0.30	18.11	0.96	21.77	3.00	17.74	1.49	22.07	0.79
MAD	±3.4		±5		±4.9		±4.9		±3.9		±5.1		±5.8	

Table 35. The mean age prediction using SYBRGREEN chemistry in blood, saliva, and semen among 1 month.

Actual age(day)	VEGFA/ACTB model						HIF1A/ACTB model						HIF1A/VEGFA model						Multiple regression model					
	Blood	SD	Saliva	SD	Semen	SD	Blood	SD	Saliva	SD	Semen	SD	Blood	SD	saliva	SD	Semen	SD	Blood	SD	Saliva	SD	Semen	SD
0	0.56	0.00	0.69	0.00	0.94	0.00	0.86	0.00	1.72	0.00	1.92	0.00	0.12	0.00	0.02	0.00	0.68	0.00	0.82	0.00	-0.15	0.00	0.34	0.00
7	11.77	6.03	11.65	5.51	15.67	5.07	12.03	6.28	12.54	6.54	13.91	5.07	11.98	6.58	8.44	4.35	15.57	5.58	10.01	5.35	11.32	5.33	15.57	3.58
14	17.24	4.20	14.33	4.97	16.81	4.35	17.78	4.43	16.20	6.54	18.04	2.55	17.98	3.83	14.14	3.19	17.26	3.25	17.18	4.88	14.14	4.33	17.80	3.38
21	19.29	3.80	20.10	3.32	17.59	2.30	19.68	1.65	17.51	3.95	18.25	4.53	18.60	3.91	21.80	2.01	18.16	1.23	20.17	3.97	19.58	2.80	17.80	5.32
28	21.11	2.53	23.26	3.94	19.07	2.68	19.63	1.95	22.02	2.30	17.91	4.61	21.39	2.08	25.47	1.61	18.23	1.59	21.79	2.79	24.28	2.73	19.46	2.99
<b>MAD</b>	<b>±4.4</b>		<b>±3.9</b>		<b>±5.4</b>		<b>±4.6</b>		<b>±4.9</b>		<b>±5.2</b>		<b>±4.2</b>		<b>±2.1</b>		<b>±5.0</b>		<b>±3.7</b>		<b>±3.0</b>		<b>±4.6</b>	

Table 36. The mean age prediction calculated by using miRNA models in blood, saliva, and semen among 1 month.

Actual age	Age prediction in blood model		Age prediction in saliva model		Age prediction in semen model			
	miRNA 451/44	SD	miRNA 205/44	SD	miRNA (+DTT) 891a/44	SD	miRNA (-DTT) 891a/44	SD
0	4.77	0.00	3.11	0.00	4.03	0.00	-0.02	0.00
7	15.26	2.89	11.96	7.51	11.99	5.15	17.32	2.12
14	15.94	4.44	15.67	3.36	10.82	6.46	16.51	1.97
21	19.63	1.11	19.07	7.23	22.87	4.28	18.07	2.06
28	14.53	6.10	20.17	4.70	20.28	4.62	18.11	2.25
<b>MAD</b>	<b>±6.40</b>		<b>±5.20</b>		<b>±5.3</b>		<b>±5.1</b>	

## Appendix 2. Blind samples

Table 37. Age predicted of blind sample using single and multiple regression models with ALAS2/B2M and HBB/B2M.

	ALAS2					
	Fresh	A	B	C	D	E
C <sub>q</sub>	33.26	29.9	28.24	37.03	32.07	33.76
	34.02	30.02	28.17	35.7	32.2	34.4
	32.57	29.88	28.15	36.9	31.94	33.27
<b>Mean C<sub>q</sub></b>	<b>33.28</b>	<b>29.93</b>	<b>28.19</b>	<b>36.54</b>	<b>32.07</b>	<b>33.81</b>
<b>SD</b>	0.00	0.11	0.06	0.19	0.01	0.19
R <sub>Q</sub>	1.00	0.34	0.23	0.10	0.06	0.85
	1.00	0.31	0.18	0.06	0.06	0.48
	1.00	0.51	0.31	0.41	0.08	0.66
<b>Mean R<sub>Q</sub></b>	<b>1.00</b>	<b>0.39</b>	<b>0.24</b>	<b>0.19</b>	<b>0.07</b>	<b>0.66</b>
<b>Age prediction</b>		<b>16.93</b>	<b>20.85</b>	<b>22.18</b>	<b>25.47</b>	<b>9.66</b>
<b>MAD</b>		<b>±3.8</b>	<b>±3.8</b>	<b>±3.8</b>	<b>±3.8</b>	<b>±3.8</b>
	HBB					
	Fresh	A	B	C	D	E
C <sub>q</sub>	26.7	21.85	19.89	27.52	24.67	26.63
	26.68	21.88	19.94	27.53	24.73	26.72
	26.65	21.92	19.96	27.56	24.71	26.72
<b>Mean C<sub>q</sub></b>	<b>26.68</b>	<b>21.88</b>	<b>19.93</b>	<b>27.54</b>	<b>24.70</b>	<b>26.69</b>
<b>SD</b>	0.03	0.04	0.04	0.02	0.03	0.05
R <sub>Q</sub>	1.00	0.12	0.07	0.01	0.03	0.79
	1.00	0.18	0.10	0.04	0.05	0.60
	1.00	0.12	0.06	0.04	0.03	0.99
<b>Mean R<sub>Q</sub></b>	<b>1.00</b>	<b>0.14</b>	<b>0.08</b>	<b>0.03</b>	<b>0.04</b>	<b>0.79</b>
<b>Age prediction</b>		<b>23.06</b>	<b>24.78</b>	<b>26.01</b>	<b>25.73</b>	<b>6.02</b>
<b>MAD</b>		<b>±3.2</b>	<b>±3.2</b>	<b>±3.2</b>	<b>±3.2</b>	<b>±3.2</b>
<b>Multiple regression</b>		<b>21.30</b>	<b>24.02</b>	<b>25.37</b>	<b>26.48</b>	<b>6.87</b>
<b>MAD</b>		<b>±3.4</b>	<b>±3.4</b>	<b>±3.4</b>	<b>±3.4</b>	<b>±3.4</b>
	B2M					
	Fresh	A	B	C	D	E
C <sub>q</sub>	27.97	26.17	25.04	35.11	30.93	28.24
	28.65	26.33	25.3	34.3	30.91	27.96
	28.01	26.29	25.3	33.64	31.03	28.1
<b>Mean C<sub>q</sub></b>	<b>28.21</b>	<b>26.26</b>	<b>25.21</b>	<b>34.35</b>	<b>30.96</b>	<b>28.10</b>
<b>SD</b>	0.38	0.08	0.15	0.74	0.06	0.14

Table 38. Blind blood samples using hypoxia models.

Blind blood samples with Hypoxia models						
	Calibration	C (0)	E (7)	D (14)	A(21)	B(28)
VEGFA Ct	26.3	27.01	27.56	28.97	27.4	29.91
HIF1A Ct	25.56	26.15	27.12	28.84	26.11	27.21
ACTB Ct	23.32	23.99	24.4	25.1	26.65	28.25
RQ of VEGFA/ACTB	1.00	0.97	0.88	0.54	0.21	0.4
Age prediction using VEGFA/ACTB model	0.56	2.04	5.93	14.91	20.02	17.13
RQ of HIF1A/ACTB	1.00	0.95	0.72	0.35	0.15	0.1
Age prediction using HIF1A/ACTB model	0.86	4.39	14.39	18.24	19.89	20.75
RQ of HIF1A/VEGFA	1.00	0.92	0.81	0.66	0.68	0.26
Age prediction using HIF1A/VEGFA model	0.12	2.00	4.54	8.24	7.58	17.61
Age prediction using multiple R model	0.82	1.80	5.59	14.74	21.96	19.51

Table 39. Blind saliva samples using hypoxia models.

Blind saliva						
	Calibration	C (0)	B (7)	A (14)	E(21)	D(28)
VEGFA Ct	27.21	27.45	28.78	29.94	31.54	32.77
HIF1A Ct	28.43	28.7	29.94	30.13	29.67	29.13
ACTB Ct	22.75	22.84	23.77	24.76	26.65	25.98
RQ of VEGFA/ACTB	1	0.90	0.68	0.61	0.58	0.20
Age prediction using VEGFA/ACTB model	0.69	3.62	9.98	12.14	12.83	23.00
RQ of HIF1A/ACTB	1	0.88	0.71	0.81	0.16	0.17
Age prediction using HIF1A/ACTB model	1.72	5.55	10.56	7.86	22.33	22.11
RQ of HIF1A/VEGFA	1	0.98	0.96	0.49	0.12	0.14
Age prediction using HIF1A/VEGFA model	0.02	0.60	1.17	14.42	24.92	24.34
Age prediction using multiple R model	-0.15	3.03	9.08	9.50	17.09	24.28

Table 40. Prediction the blind samples using TaqMan models in semen.

Blind sample in semen with Taqman models						
	Calibration	B (0)	E (7)	A (14)	E(21)	D(28)
PRM1 Ct	25.54	25.98	26.16	28.97	30.01	31.92
SEMG1 Ct	28.12	28.52	29.49	30.21	29.87	28.71
B2M Ct	26.15	26.47	27.36	28.89	29.65	30.85
RQ of PRM1/B2M	1	0.920188	0.664343	0.619854	0.510506	0.312083
Age prediction using PRM1/B2M model	1.04	5.583124	14.49046	15.38512	17.06671	19.19621
RQ of SEMG1/B2M	1	0.946058	0.895025	0.63728	0.297302	0.057912
Age prediction using SEMG1/B2M model	0.29	3.783576	6.65021	15.83645	20.02306	22.47416
RQ of PRM1/SEMG1	1	0.972655	0.594604	0.395021	0.151774	0.018073
Age prediction using PRM1/SEMG1 model	1.46	3.170616	15.93638	17.95782	20.41146	23.20066
Age prediction using multiple R model	1.745	3.585161	7.888117	11.89483	17.89967	23.853

Table 41. Age prediction using hypoxia models in semen.

Blind semen samples						
	Calibration	D (0)	A (7)	C (14)	B(21)	E(28)
VEGFA Ct	28.53	29.33	29.6	29.52	30.69	29.02
HIF1A Ct	24.67	25.59	25.45	24.72	29.05	32.25
ACTB Ct	21.65	22.5	23.86	24.01	26.96	26.1
RQ of VEGFA/ACTB	1.00	0.97	0.45	0.39	0.11	0.06
Age prediction using VEGFA/ACTB model	0.94	1.69	13.04	14.52	20.59	21.67
RQ of HIF1A/ACTB	1.00	0.95	0.37	0.20	0.52	0.11
Age prediction using HIF1A/ACTB model	1.92	3.74	18.81	20.67	16.14	21.18
RQ of HIF1A/VEGFA	1.00	0.92	0.82	0.52	0.21	0.01
Age prediction using HIF1A/VEGFA model	0.68	4.82	9.40	18.14	20.05	17.23
Age prediction using multiple R model	0.34	2.07	12.97	19.29	17.51	26.80

### Appendix 3. Models generate with geometric mean normalisation

Table 42. The mean age prediction using geometric mean in blood, saliva, and semen.

Actual age	Age prediction using geometric mean normalization in Blood				Age prediction using geometric mean normalization in saliva				Age prediction using geometric mean normalization in semen			
	GYP A/(B2M+miRNA451+44)	SD	GYP A/(miRNA451+44)	SD	HTN3/(GAPDH+miRNA205+44)	SD	HTN3/(miRNA 205+44)	SD	SEMG1/(B2M+miRNA891a+44)	SD	SEMG1/(miRNA 891a+44)	SD
0	8.43	0.00	8.64	0.00	4.53	0.00	-0.06	0.00	2.75	0.00	0.94	0.00
7	14.16	3.05	13.65	3.16	16.63	3.06	16.23	4.80	13.01	5.80	10.58	1.62
14	15.85	5.14	15.50	3.97	16.66	1.36	15.55	2.38	13.91	8.43	15.57	4.97
21	16.31	4.95	18.12	3.95	16.27	3.16	19.09	1.82	18.92	2.95	19.61	6.50
28	15.25	5.67	14.11	6.20	15.91	5.66	19.18	2.40	21.41	1.58	23.31	2.90
MAD	±7.6		±7.2		±6.7		±4.5		±5.4		±3.8	

### Appendix 4. Pearson's correlation

Table 43. Relationship between ALAS2, HBB and B2M in blood among one week

		Age	RQ of ALAS2/B2M	RQ of ALAS2/GAPDH	RQ of B2M /GAPDH
Age	Pearson's Correlation	1			
	Sig. (2-tailed)				
	N	56			
RQ of ALAS2/B2M	Pearson's Correlation	-.596**	1		
	Sig. (2-tailed)	.000			
	N	56	56		
RQ of ALAS2/ GAPDH	Pearson's Correlation	-.447**	.296*	1	
	Sig. (2-tailed)	.001	.027		
	N	56	56	56	
RQ of B2M/GAPDH	Pearson's Correlation	-.643**	.876**	.409**	1
	Sig. (2-tailed)	.000	.000	.002	
	N	56	56	56	56

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Table 44. Pearson's correlation calculation in miRNA 205 and 44 in saliva samples.

Correlations							
		Age	ΔCq of 205	ΔCq of 44	ΔCq (Cq 205-Cq44)	Ratio 205/44	RQ of 205/44
Age	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	15					
ΔCq of 205	Pearson Correlation	-.851**	1				
	Sig. (2-tailed)	.000					
	N	15	15				
ΔCq of 44	Pearson Correlation	-.807**	.872**	1			
	Sig. (2-tailed)	.000	.000				
	N	15	15	15			
ΔCq of (Cq of 205-Cq 44)	Pearson Correlation	-.561*	.760**	.345	1		
	Sig. (2-tailed)	.030	.001	.209			
	N	15	15	15	15		
Ratio 205 /44	Pearson Correlation	-.448	.634*	.176	.982**	1	
	Sig. (2-tailed)	.094	.011	.530	.000		
	N	15	15	15	15	15	
RQ of 205/44	Pearson Correlation	-.767**	.815**	.787**	.519*	.408	1
	Sig. (2-tailed)	.001	.000	.001	.047	.131	
	N	15	15	15	15	15	15

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Table 45. Correlation between miRNA 891a and 44 (+DTT) in semen samples.

Correlations							
Semen (+DDT)		Age	$\Delta$ Cq of 891a	$\Delta$ Cq of 44	$\Delta$ Cq (Cq 891a-Cq44)	Ratio of 891a /44	RQ of 891a /44
Age	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	15					
$\Delta$ Cq of 891a	Pearson Correlation	-0.38	1				
	Sig. (2-tailed)	.163					
	N	15	15				
$\Delta$ Cq of 44	Pearson Correlation	-.678**	0.228	1			
	Sig. (2-tailed)	.006	.414				
	N	15	15	15			
$\Delta$ Cq (Cq of 891a - Cq 44)	Pearson Correlation	-0.296	-.553*	.685**	1		
	Sig. (2-tailed)	0.29	.032	.005			
	N	15	15	15	15		
Ratio of 891a /44	Pearson Correlation	.331	0.511	-.720**	-.998**	1	
	Sig. (2-tailed)	.228	.051	.002	.000		
	N	15	15	15	15	15	
RQ of 891a /44	Pearson Correlation	-0.788**	0.22	.626*	0.372	-.410	1
	Sig. (2-tailed)	.000	.433	.013	.172	.129	
	N	15	15	15	15	15	15

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Table 46. Correlation in blood with miRNA markers.

Correlations							
		Age	ΔCq of 451	ΔCq of 44	ΔCq (Cq 451-Cq44)	Ratio of 451/44	RQ of 451/44
Age	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	15					
ΔCq of 451	Pearson Correlation	-.512	1				
	Sig. (2-tailed)	.051					
	N	15	15				
ΔCq of 44	Pearson Correlation	.081	-.364	1			
	Sig. (2-tailed)	.774	.183				
	N	15	15	15			
ΔCq (Cq 451 -Cq 44)	Pearson Correlation	.385	-.864**	.784**	1		
	Sig. (2-tailed)	.156	.000	.001			
	N	15	15	15	15		
Ratio of 451/44	Pearson Correlation	-.321	.789**	-.856**	-.989**	1	
	Sig. (2-tailed)	.244	.000	.000	.000		
	N	15	15	15	15	15	
RQ of 451/44	Pearson Correlation	-.584*	.656**	.071	-.399	.298	1
	Sig. (2-tailed)	.022	.008	.801	.141	.281	
	N	15	15	15	15	15	15

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\* . Correlation is significant at the 0.01 level (2-tailed).

Table 47. Semen markers without DTT (-DTT).

Correlations							
		Age	ΔCq of 891a	ΔCq of 44	ΔCq (Cq 891a-Cq44)	Ratio 891a/44	RQ of 891a /44
Age	Pearson Correlation	1					
	Sig. (2-tailed)						
	N	15					
ΔCq of 891a	Pearson Correlation	-.689**	1				
	Sig. (2-tailed)	.004					
	N	15	15				
ΔCq of 44	Pearson Correlation	-0.490	0.296	1			
	Sig. (2-tailed)	.064	.285				
	N	15	15	15			
ΔCq of (Cq of 891a -Cq 44)	Pearson Correlation	-0.144	-0.216	.869**	1		
	Sig. (2-tailed)	0.61	.439	.000			
	N	15	15	15	15		
Ratio of 891a /44	Pearson Correlation	.170	0.219	-.858**	-.990**	1	
	Sig. (2-tailed)	.545	.433	.000	.000		
	N	15	15	15	15	15	
RQ of 891a /44	Pearson Correlation	-.518*	0.27	.924**	.805**	-.775**	1
	Sig. (2-tailed)	.048	.330	.000	.000	.001	
	N	15	15	15	15	15	15

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

### Appendix 5. Normal disruption of RQ of ALAS2/B2M

Table 48. Multiple regression analysis in blood during one week.

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.817	.792		8.605	.000
	RQ of ALAS2/B2M	-1.484	1.551	-.203	-.957	.034
	RQ of ALAS2/GAPDH	-2.313	1.104	-.236	-2.095	.041
	RQ of B2M/GAPDH	-2.602	1.572	-.368	-1.655	<b>.104</b>

Table 49. Multiple regression analysis in semen samples during 28 days.

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	28.123	2.164		12.994	.000
	RQ of PRM1/B2M	-9.113	4.886	-.308	-1.865	.023
	RQ of SEMG1/B2M	-7.656	5.118	-.249	-1.496	.033
	RQ of PRM1/SEMG1	-9.676	5.051	-.375	-1.916	<b>.066</b>