

# cdr

## Telomere length analysis: A new tool for molecular photofitting

Item Type	Thesis or dissertation
Authors	Cargill, Stephen R.
Publisher	University of Chester
Download date	2026-05-20 19:56:57
Link to Item	<a href="http://hdl.handle.net/10034/126273">http://hdl.handle.net/10034/126273</a>



***Telomere Length Analysis***  
***A New Tool For Molecular Photofitting***

***Thesis submitted according to the requirements***  
***of the University of Chester for the degree***  
***of Master of Philosophy by***

***Stephen Richard Cargill***  
***January 2011***

Supervisors

Dr I McDowall

Dr D Commane

# Contents

---

	<b>List of Figures</b>	iii
	<b>Abbreviations</b>	vi
	<b>Declaration</b>	vii
	<b>Acknowledgements</b>	viii
	<b>Abstract</b>	<b>1</b>
<hr/>		
<b>One</b>	<b>Literature Review</b>	<b>2</b>
	1.1 Introduction	3
	1.2 The Science of DNA Profiling	4
	1.3 DNA Profiling and the Courts	8
	1.4 “Molecular Photofitting”	15
<hr/>		
<b>Two</b>	<b>Telomere Length Analysis</b>	<b>21</b>
	2.1 Telomeres – Structure & Function	22
	2.2 The Telomere Restriction Fragment Assay	31
	2.3 The Real-Time Quantitative PCR Assay	34
	2.4 The Telomeric Multiplication Factor Assay	42
	2.5 Aims & Objectives	46
<hr/>		
<b>Three</b>	<b>Materials &amp; Method</b>	<b>48</b>
	3.1 Materials & Method – Phase One	49
	3.2 Materials & Method – Phase Two	58
<hr/>		
<b>Four</b>	<b>Results &amp; Analysis</b>	<b>62</b>
	4.1 Validation of the TMF Assay	63
	4.2 Linear Regression Age Prediction	64
	4.3 Telomere Length Analysis	70
	4.4 Telomere Length Analysis & Alternative Tissues	87

4.5	TMF & Other Factors	96
<hr/>		
<b>Five</b>	<b><i>Discussion &amp; Conclusions</i></b>	<b>98</b>
5.1	The Legal Implications of Telomere Length Analysis	99
5.2	The Practical Implications of Telomere Length Analysis	103
5.3	The Technical Limitations of Telomere Length Analysis	107
5.4	Conclusion	111
<hr/>		
<b>Six</b>	<b><i>References &amp; Bibliography</i></b>	<b>112</b>
<hr/>		
<b>Seven</b>	<b><i>Appendices</i></b>	<b>119</b>

## ***List of Figures***

---

<i>Figure 1.2.1</i>	Illustration of restriction fragment length polymorphisms.
<i>Figure 1.2.2</i>	Slipped strand mispairing.
<hr/>	
<i>Figure 2.1.1</i>	Multiple repair options of a double-stranded break in a linear chromosome.
<i>Figure 2.1.2</i>	The role of telomeres in the maintenance of chromosomal structure.
<i>Figure 2.1.3</i>	Outline of linear DNA replication from multiple primer-binding sequences.
<i>Figure 2.1.4</i>	The 3'-end replication problem.
<hr/>	
<i>Figure 2.2.1</i>	Illustration of the TRF assay.
<i>Figure 2.2.2</i>	The effect of point mutations on the TRF assay.
<hr/>	
<i>Figure 2.3.1</i>	Primer-dimer formed by Cawthon's telomeric primers.
<i>Figure 2.3.2</i>	Alignment of telomeric primers to telomeric DNA.
<i>Figure 2.3.3</i>	Point mutations and telomeric primers.
<i>Figure 2.3.4</i>	Failure to consistently amplify repetitive DNA.
<i>Figure 2.3.5</i>	Partly complementary primers do not target telomere ends.
<i>Figure 2.3.6</i>	Multiple amplimers arising from telomeric PCR.
<hr/>	
<i>Table 3.1.1</i>	Age and gender distribution of Phase One participants.
<i>Table 3.1.2</i>	Height and gender distribution of Phase One participants.
<i>Table 3.1.3</i>	Body mass and gender distribution of Phase One participants.
<hr/>	
<i>Table 3.1.4</i>	Body Mass Index and gender distribution of Phase One participants.
<i>Figure 3.1.1</i>	Results of PCR optimisation.
<i>Table 3.1.5</i>	Telomeric PCR primers.
<i>Table 3.1.6</i>	Single-copy locus PCR primers
<hr/>	

<i>Table 3.2.1</i>	Age and gender distribution of Phase Two participants.
<i>Table 3.2.2</i>	Height and gender distribution of Phase Two participants.
<i>Table 3.2.3</i>	Body mass and gender distribution of Phase Two participants.
<i>Table 3.2.4</i>	Body Mass Index and gender distribution of Phase Two participants.
<hr/>	
<i>Graph 4.1.1</i>	The strong correlation between telomere lengths as measured by the real-time qPCR assay and by the TMF assay.
<hr/>	
<i>Graph 4.2.1</i>	The relationship between TMF and age.
<i>Graph 4.2.2</i>	Comparison of ages predicted using LRAP with known ages.
<i>Graph 4.2.3</i>	The relationship between TMF and age in males and females.
<i>Graph 4.2.4</i>	Comparison of ages predicted using GD-LRAP with known ages in males and females.
<i>Graph 4.2.5</i>	Prediction profiles of LRAP and GD-LRAP.
<i>Graph 4.2.6</i>	Comparison of ages predicted for test participants using LRAP and GD-LRAP of male and female participants.
<hr/>	
<i>Graph 4.3.1</i>	Mean age and mean TMF within 5-year age groups.
<i>Table 4.3.1</i>	Calculation of $TMF_A/TMF_C$ and $Age_A/Age_C$ .
<i>Graph 4.3.2</i>	The relationship between $TMF_A/TMF_C$ and $Age_A/Age_C$ .
<i>Table 4.3.2</i>	Illustration of the prediction of $Age_A$ for Participant 1.
<i>Graph 4.3.3</i>	Comparison of ages predicted using TLA with known ages.
<i>Graph 4.3.4</i>	Prediction profiles for TLA and LRAP.
<i>Table 4.3.3</i>	The effect of thresholds on database size.
<i>Table 4.3.4</i>	The effect of threshold values of $TMF_A/TMF_C$ on age prediction.
<i>Graph 4.3.5</i>	Prediction profiles for unrestricted TLA and TLA with restrictive thresholds at $\pm 0.5$ and $\pm 0.3$ .
<i>Graph 4.3.6</i>	Prediction profiles for TLA with restrictive thresholds at $\pm 0.25$ , $\pm 0.2$ and $\pm 0.15$ .
<i>Graph 4.3.7</i>	Prediction profiles for TLA with restrictive thresholds at $\pm 0.1$ and $\pm 0.05$ .
<i>Graph 4.3.8</i>	Comparison of ages predicted using restricted TLA with known ages.

<i>Graph 4.3.9</i>	Prediction profiles for TLA, restricted TLA and LRAP.
<i>Graph 4.3.10</i>	The relationship between $TMF_A/TMF_C$ and $Age_A/Age_C$ in males and females.
<i>Graph 4.3.11</i>	Comparison of ages using GD-TLA with known ages in males and females.
<i>Graph 4.3.12</i>	Prediction profiles for TLA, GD-TLA and GD-LRAP.
<i>Graph 4.3.13</i>	Comparison of ages predicted for test participants using restricted TLA and GD-TLA in men and women with known ages.
<i>Table 4.3.5</i>	Composition of databases.
<i>Graph 4.3.14</i>	Prediction profiles for TLA using databases of 40, 60 and 80 participants.

---

<i>Graph 4.4.1</i>	The lack of a relationship between $_B TMF$ and $_F TMF$ .
<i>Graph 4.4.2</i>	The relationship between $_B TMF$ and $_F TMF$ in males and females.
<i>Graph 4.4.3</i>	Comparison of the relationships between $_B TMF$ or $_F TMF$ and age.
<i>Graph 4.4.4</i>	Prediction profiles for TLA on Phase Two participants, using $_B TMF$ and $_F TMF$ .
<i>Graph 4.4.5</i>	The relationship between $_L TMF$ and $_B TMF$ in all participants.
<i>Graph 4.4.6</i>	The relationship between $_L TMF$ and $_B TMF$ in males and females.
<i>Graph 4.4.7</i>	Comparison of $_B TMF$ values derived from $_L TMF$ ( $_D TMF$ ) with known $_B TMF$ values.
<i>Graph 4.4.8</i>	Prediction profiles for TLA on Phase Two participants using $_B TMF$ and using $_L TMF$ to calculate $_D TMF$ values.
<i>Graph 4.4.9</i>	Comparison of ages predicted by restricted TLA with known ages, using $_B TMF$ and $_D TMF$ for Phase Two participants.

---

<i>Graph 4.5.1</i>	The absence of a relationship between TMF and height.
<i>Graph 4.5.2</i>	The absence of a relationship between TMF and body mass.
<i>Graph 4.5.3</i>	The absence of a relationship between TMF and BMI.

---

<i>Graph 5.1.1</i>	An illustration of the effect of puberty and old age on TMF.
--------------------	--

## Abbreviations

---

<i>BMI</i>	Body Mass Index.
<i>CODIS</i>	Combined DNA Index System.
<i>ECHR</i>	European Court of Human Rights.
<i>GD-LRAP</i>	Gender-Differentiated Linear Regression Age Prediction.
<i>GD-TLA</i>	Gender-Differentiated Telomere Length Analysis.
<i>LCN</i>	Low Copy Number [DNA profiling].
<i>LRAP</i>	Linear Regression Age Prediction.
<i>NDNAD</i>	National DNA Database.
<i>PACE</i>	Police and Criminal Evidence Act 1984.
<i>PBS</i>	Primer-Binding Sequence.
<i>PCR</i>	Polymerase Chain Reaction.
<i>qPCR</i>	Quantitative Polymerase Chain Reaction.
<i>RFLP</i>	Restriction Fragment Length Polymorphism.
<i>RMP</i>	Random Match Probability.
<i>ROR</i>	Random Occurrence Ratio.
<i>SGM+</i>	Second Generation Matrix Plus.
<i>SNP</i>	Single Nucleotide Polymorphism.
<i>STR</i>	Short Tandem Repeat.
<i>TLA</i>	Telomere Length Analysis.
<i>TMF</i>	Telomeric Multiplication Factor (also $\tau$ ). TMF can also be classified as: ${}_B$ TMF                      TMF measured in buccal epithelial cells (also ${}_B\tau$ ). ${}_D$ TMF                      Derived TMF (also ${}_D\tau$ ). ${}_F$ TMF                      TMF measured in hair follicle cells (also ${}_F\tau$ ). ${}_L$ TMF                      TMF measured in leucocytes (also ${}_L\tau$ ). TMF <sub>A</sub> TMF subject to analysis (also $\tau_A$ ). TMF <sub>C</sub> TMF acting as a comparison (also $\tau_C$ ). TMF <sub>X</sub> TMF of Participant X (also $\tau_X$ ).
<i>TRF</i>	Telomeric Restriction Fragment.
<i>VNTR</i>	Variable Number Tandem Repeat.

## ***Declaration***

---

I certify that this work is original in its entirety and has never before been submitted for any form of assessment.

The practical work, data analysis and presentation, and written work presented are all my own work unless otherwise stated.

Amendments requested by the *viva voce* panel of the 14<sup>th</sup> December 2010 have been made.

All sources have been properly referenced and attributed.

Signed: \_\_\_\_\_

Stephen Richard Cargill

Date: \_\_\_\_\_

## ***Acknowledgements***

---

I would like to express my gratitude to my supervisors, Dr Ian McDowall and Dr Daniel Commane, for their assistance and encouragement throughout.

I would also like to thank the University of Chester's technical staff, particularly Lesley Holland, for their peerless support.

I am particularly grateful to Luan Lunt of LGC Forensics, who gave up her time to offer me a professional's opinion on my research, and Dr Carmen Martin-Ruiz of the Institute of Ageing and Health at Newcastle University, for her assistance in the validation of our assay protocols.

I would like to thank my friends and family, who supported me throughout and gave their time freely to read draft of this thesis.

Most of all, I am grateful to all the volunteers who gave up their time to provide the samples and information I needed. They must remain nameless, but they know who they are.

## **Abstract**

---

*This thesis describes a new assay and analytical protocol to determine the telomere length of an individual, and its potential application in criminal investigations.*

*The most commonly used existing assay is based on real-time qPCR. The Telomeric Multiplication Factor (TMF) assay described here instead uses end-point PCR and densitometry. Because most existing forensic DNA techniques already use end-point PCR, the TMF assay can be more easily integrated into the suite of tests available.*

*Along with the TMF assay, a new procedure to use telomere length to determine age is presented. Previous attempts to do this rely on the calculation of a linear regression, and the interpolation of an age based on telomere length, which is not accurate enough for use in criminal investigations (with a covariance between known and predicted age of 0.2848,  $P=0.010564$ ). Telomere Length Analysis (TLA) uses a database of known individuals. An unknown individual's telomere length is compared to all telomere lengths on the database, and used to calculate the unknown individual's age as a multiple of each age on the database. This increases exponentially the number of comparisons that can be made, and improves the accuracy of age predictions (with a covariance between known and predicted age of 0.4561,  $P=0.000021$ ). The accuracy increases as the size of the database increases.*

*TLA was developed using buccal epithelial cell samples. However, this thesis demonstrates that TLA also works on hair and blood samples, although to a lesser degree of accuracy. This makes TLA applicable in more forensic scenarios.*

*TLA is intended to operate alongside the recent developments in molecular photofitting, to provide phenotypic and biographical information about an unknown offender to hasten his arrest and conviction, in the event that no matching DNA profile is recorded on investigating authorities' databases. The new TMF assay and TLA analytical profile are more accurate and more applicable to a forensic scenario than other, previous, attempts to use telomere length to determine age.*

## Chapter One: Literature Review

---

<b>1.1</b>	<b>Introduction</b>	<b>3</b>
1.1.1	The limitations of DNA profiling	3
<b>1.2</b>	<b>The Science of DNA Profiling</b>	<b>4</b>
1.2.1	Introduction	4
1.2.2	RFLPs as DNA profile markers	4
1.2.3	STRs and VNTRs as DNA profile markers	5
1.2.4	International standards of DNA profiling	7
<b>1.3</b>	<b>DNA Profiling and the Courts</b>	<b>8</b>
1.3.1	Police and Criminal Evidence Act 1984	8
1.3.2	The National DNA Database	8
1.3.3	The need for other evidence	9
1.3.4	Random Match Probabilities	10
1.3.5	The Prosecutor's Fallacy	11
1.3.6	LCN DNA profiling	13
1.3.7	Retention of DNA profiles	13
<b>1.4</b>	<b>"Molecular Photofitting"</b>	<b>15</b>
1.4.1	The principle of molecular photofitting	15
1.4.2	Gender	15
1.4.3	Ethnicity and physical appearance	16
1.4.4	Ethnicity and appearance in Western Europeans	17
1.4.5	Eye and hair colour	19
1.4.6	Height, weight and build	19
1.4.7	Age	20

## **1.1: Introduction**

---

### *1.1.1: The limitations of DNA profiling*

Before evaluating Telomere Length Analysis as a molecular photofitting assay, it is necessary to consider how it will best fit into criminal investigations. While DNA profiling is the ideal, it has several limitations. Except for the fact that degraded DNA is a less suitable substrate for DNA profiling, the majority of these limitations have been imposed by the courts. Degraded DNA samples, of course, have their uses, most notably in overturning miscarriages of justice by demonstrably showing that the individual convicted of a crime prior to the advent of DNA profiling is, in fact, innocent.

The relatively new science of molecular photofitting has developed as a response to the limitations of DNA profiling. It is as a part of this suite of assays that Telomere Length Analysis is expected to function.

## **1.2: The Science of DNA Profiling**

---

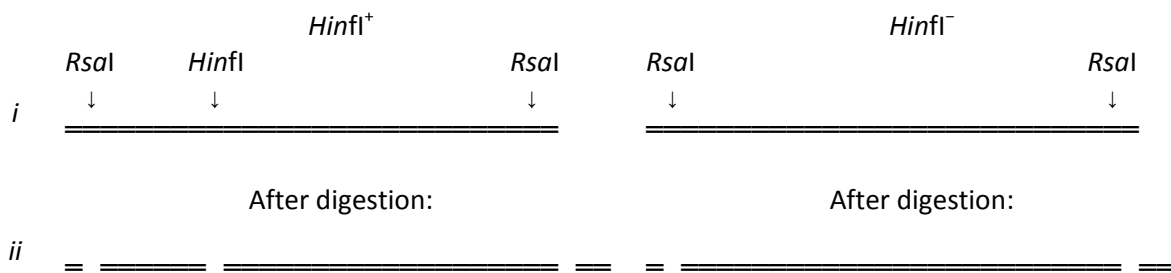
### *1.2.1: Introduction*

DNA profiling was first used in a criminal prosecution in January 1988, to convict Colin Pitchfork of the rape and murder of two women (Ridley, 1999; Watson, 2004). Since then, it has become a significant part of the armoury available to investigating authorities to identify offenders and their victims. From 2006 to 2009, positive identifications of individuals by DNA profiling numbered 106,097 (Travis, 2009a).

Although only 33,000, or 0.67%, of the 4,900,000 crimes recorded in the United Kingdom in the year 2008-2009 – the latest year for which figures are available – were solved as a result of DNA profiling, it plays a greater role in specific offences. For example, 40% of recorded burglaries were solved as a consequence of DNA profiling (Travis, 2010). Offences likely to yield a DNA profile, such as offences against the person or burglary, constitute only around 20% of recorded crime (Hoare, 2009). Other offences, such as drug offences or criminal damage, are unlikely to be solved by DNA profiling – the former because many drug convictions arise from searches of an individual or property, the latter because their very nature means that an offender is less likely to leave identifiable traces of DNA.

### *1.2.2: RFLPs as DNA profile markers*

DNA profiling relies on the fact that the overwhelming majority of the human genome does not encode any functional RNA or proteins. As a consequence of this, there is no selection pressure opposing mutation in these non-coding sequences and so mutation is common. Early DNA profiles, such as that used to convict Pitchfork, were based on restriction fragment length polymorphisms, or RFLPs (Watson, 2004). An RFLP occurs when the presence or absence of the recognition sequence for a restriction enzyme results in restriction fragments of noticeably different sizes (see Figure 1.2.1, overleaf).



*Figure 1.2.1:* Illustration of restriction fragment length polymorphisms. The DNA strand on the left has two restriction sites for *RsaI*, with the recognition sequence 5'-GTAC-3', and a single restriction site for *Hinfl*, with the recognition sequence 5'-GTNAC-3'. The DNA strand on the right has only the *RsaI* restriction sites (*i*). Following digestion, the strand on the left, with the genotype *Hinfl*<sup>+</sup>, has been cleaved in three places, giving two distinct fragments between the *RsaI* restriction sites. In contrast, the strand on the right, with the genotype *Hinfl*<sup>-</sup>, has only been cleaved twice, resulting in a single, longer fragment (*ii*) from between the *RsaI* restriction sites (adapted from Lewin, 2000).

However, it is clearly evident that an individual RFLP can only have two alleles – the presence or absence of a restriction site. A stretch of DNA may of course harbour several RFLPs, but even so, the number of alleles does not allow for great discrimination between individuals (Watson, 2004).

### 1.2.3: STRs and VNTRs as DNA profile markers

Current DNA profiling protocols use sequences of repetitive DNA found in non-coding parts of the genome. These can be either minisatellites or the smaller microsatellites (Butler, 2001). In either case, the principles remain the same. Repetitive sequences are particularly prone to mutation through a process called slipped strand mispairing (Krawczak and Schmidtke, 1998). Essentially, the DNA polymerase 'stutters' as it synthesises new DNA from a repetitive template strand (see Figure 1.2.2, overleaf).



marker. By labelling primers with particular fluorescent dyes, it is possible to amplify several markers in a single PCR tube – a multiplex PCR – and still be able to identify which STRs and VNTRs have been amplified, and measure their amplicon lengths individually (Watson, 2004).

#### *1.2.4: International standards of DNA profiling*

DNA profiling procedures use several STR markers: for example, the United Kingdom uses ten in a multiplex PCR reaction referred to as the Second Generation Matrix Plus (SGM+), while the federal bodies in the United States use thirteen in their Combined DNA Index System (CODIS). Interpol requires a minimum of seven STRs in a profile before it can be included in an international database. Interpol's requirement recognises that different countries use different markers, so it allows flexibility between nations, although SGM+ and CODIS include all seven of Interpol's markers (Watson, 2004). The use of multiple markers allows greater discrimination between individuals, as well as increasing the possibility of recovering at least some useful data from a degraded sample.

Interpol requires, and SGM+ and CODIS both include, an extra marker, from intron 1 of the amelogenin gene (Watson, 2004). The forensic significance of the amelogenin marker is discussed further in section 1.4.2. SGM+ therefore uses eleven markers and CODIS fourteen, while Interpol require eight markers in total.

By using so many STR markers, as well as amelogenin, a DNA profile is able to distinguish between two different individuals with a high degree of accuracy. Working on the assumption that each of the ten STR markers used in the SGM+ system has ten alleles occurring at equal frequencies, there are one trillion possible diploid genotypes –  $(10 \times 10)^{10}$ , or  $10^{12}$ . While this is a generalisation, as allele frequencies are not equal in a population, the probability of an individual not being the source of the DNA that was profiled is typically given in court as one in a billion or higher (McCartney, 2008) – this is referred to as the Random Match Probability, or RMP (Geddes, 2010).

## ***1.3: DNA Profiling and the Courts***

---

### *1.3.1: Police and Criminal Evidence Act 1984*

The powers of the police in the United Kingdom to take samples from individuals suspected of a crime are laid down in the Police and Criminal Evidence Act 1984 (PACE) and its various amendments and Codes of Practice. To summarise the legislation, the police are able to demand a non-intimate sample, usually taken by a buccal swab, from an individual arrested on reasonable suspicion of committing a recordable offence (PACE Code D paragraph 6.6). A recordable offence is defined as any offence that carries a potential custodial sentence (PACE Code D notes for guidance paragraph 4A).

Of course, the police are free to request a DNA sample from anyone – for example, in screening a population – but this requires written consent (PACE Code D paragraph 6.5). However, there have been incidents where individuals refusing to participate in screening have been arrested and their DNA profiles taken without their consent. In Operation Minstead, 1,000 black men were invited to give samples of their DNA to identify a serial rapist – 125 refused and received “intimidatory” letters from the police. 5 were subsequently arrested. This has been described as considering “non-compliance as a crime,” jeopardising the relationship between the public and the police (McCartney, 2006, p180).

### *1.3.2: The National DNA Database*

After the DNA sample is collected and processed, the resulting profile is loaded onto the National DNA Database (NDNAD). Here, it can be compared with profiles taken from the scene of the crime and from others still unsolved. Similarly, a sample taken from the scene of the crime can be compared with the samples of all individuals held on the NDNAD, to identify suspects, or with profiles from other crimes to establish links between them.

In December 2008, the NDNAD held the profiles of approximately 4.3 million people – more per head of population than any other country (Travis, 2008). Given that the NDNAD held only 2.9 million profiles in March 2005 (McCartney, 2006), this

suggests the addition of approximately 1,000 profiles per day between 2005 and late 2008. By October 2009, the NDNAD held the profiles of 4.89 million people, with the rate of addition of profiles at over 2,000 per day (Travis, 2009a).

An individual linked to a crime by DNA profiling can then be charged and tried in court. There, his or her DNA profile can be entered as evidence of his guilt by the Crown Prosecution Service. The manner in which the courts handle DNA evidence has changed little since it was first introduced in the 1980s, although the most complete summary can be found in *R v Doheny* [1997].

### 1.3.3: *The need for other evidence*

DNA evidence is not in itself proof of guilt. This is a consequence of the definitions of many crimes, which require both an *actus reus* and a *mens rea* (Ashworth, 2006) – literally, the guilty act and the guilty mind (Ivamy, 1993).

As an example, the simplest definition of rape is sexual intercourse without reasonable belief that the other person consents to this (s1 Sexual Offences Act 2003). Immediately, rape can be seen in terms of its *actus reus* – sexual intercourse – and its *mens rea* – the absence of reasonable belief in consent. Similarly, burglary requires trespass with the intent to steal, cause harm or to cause criminal damage (s9 Theft Act 1968). The act of trespass forms the *actus reus*, while the intent to steal, cause harm or to cause criminal damage comprises the *mens rea*.

Where an offence has both an *actus reus* and a *mens rea*, both must typically be proven to the satisfaction of a jury for a conviction. Clearly, DNA evidence may be extremely useful in proving the *actus reus*, but it is of little help in demonstrating the *mens rea*. To return to the example of rape, DNA evidence can easily show that sexual intercourse took place, but it simply cannot indicate whether or not there was a reasonable belief in consent.

Because of this, the courts in the United Kingdom must have other evidence of guilt besides a DNA profile before determining whether or not the accused is guilty (*R v Doheny* [1997]). This definitive criterion does not appear to apply to other forms of forensic evidence, although there is, in fact, an unwritten convention that these should not be the only evidence to convict an individual.

The lack of an explicit prohibition on reliance solely on a single piece of forensic evidence such as fingermarks is probably a consequence of the fact that fingermarks, footwear marks and toolmarks have been considered to be valid, admissible evidence for far longer than DNA profile evidence, and such a prohibition has become legal convention. Furthermore, fingermarks derive from the unique pattern of ridges in an individual's fingerprint, footwear marks from an individual's stance and posture, as well as wear and tear, and toolmarks from wear and tear. As such, they clearly connect a specific individual or tool to an event. In contrast to these other forms of forensic evidence, a DNA profile is a statistical product of marker allele frequencies based on population data.

#### *1.3.4: Random Match Probabilities*

An individual's DNA profile results from the frequency of alleles in his or her particular population of the markers used. As a consequence, while it is unlikely, it is not impossible that two unrelated people can share a matching DNA profile, while related individuals will share DNA profile marker alleles in proportion to the closeness of their relationship. Siblings would be expected to have half of their marker alleles in common, while monozygotic twins would have identical DNA profiles. This is not something that has been seen in conventional fingerprinting which, since its first use in criminal proceedings in 1902, has yet to produce a single recorded coincidental match (Barnett, 2004).

The second significant rule on DNA evidence is also stated in *Doheny*, and requires that both the prosecution, including their expert witnesses, and the judge stress to the jury the statistical probability of a coincidental DNA profile match, derived from the frequency of the particular alleles in the population and the Hardy-Weinberg equation (Krawczak and Schmidtke, 1998; Geddes, 2010). Based on a full profile, this Random Match Probability (RMP) is typically given as around one in a billion (McCartney, 2008). For a partial profile obtained from degraded DNA, the RMP reflects the greater probability of a coincidental match.

An inability to provide the RMP of a DNA profile renders the evidence unreliable (*R v Bamber* [2002]). Similarly, a failure to provide an accurate RMP to the court undermines the significance of the DNA profile – in *R v Tran* [1991], the Australian Supreme Court ruled DNA evidence to be inadmissible when they learned that the RMP between the Vietnamese defendant and samples recovered from the

crime was calculated using allele frequencies in Afro-Caribbean men (*R v Tran* [1991]). In either case, the DNA profile will likely be considered to be prejudicial without this RMP to set it in context, and therefore be excluded by the courts (s78 PACE) – the judge retains the power to refuse to admit evidence if he feels that its probative value is outweighed by the likelihood of it prejudicing the trial.

However, RMPs are calculated using population data from the early days of DNA profiling, and there is concern that they might not be accurate. In the Arizona State DNA Database, a small part of CODIS, 144 out of 65,493 profiles matched at 9 or more of the 13 loci tested – an RMP of 0.22% or 1 in 455, substantially higher than that predicted. Given that a 9-locus match is considered adequate cause for prosecution in the US courts, in order to allow for a certain extent of DNA degradation, this suggests that many identifications based on DNA profiles in Arizona may be flawed. Furthermore, a study of the Australian State of Victoria DNA database suggested that 1 in 1,000 profiles may have been incorrectly entered, leading to more potential miscarriages of justice. Researchers have not yet been granted access to national DNA databases such as CODIS to search for errors or amend the RMPs, due to data protection legislation (Geddes, 2010).

### *1.3.5: The Prosecutor's Fallacy*

Regardless of the accuracy, or inaccuracy, of a Random Match Probability, a DNA profile is presented to the court as a statistical construct, and is therefore prone to be misinterpreted and misrepresented, due to the Prosecutor's Fallacy.

The Prosecutor's Fallacy is, essentially, the lack of understanding of statistical evidence in the legal profession (Choo, 2006). The entire undergraduate Law class of 2007 at the University of Leeds – 214 students – contained only 39, or 18%, with any science qualifications post-16 and there is no specific training on the majority of law courses in the United Kingdom for dealing with forensic evidence. Furthermore, the postgraduate qualifications required to practice as a solicitor or barrister in England and Wales do not commonly feature training in presenting statistical evidence (McCartney, 2008).

Because of this lack of understanding, statistics can frequently be misrepresented to the jury – who cannot be expected to understand them any better than the lawyers involved – although there is no intention to suggest that this is

anything other than an innocent mistake. It has been argued that, when an individual's liberty and livelihood relies on the proper interpretation of statistical evidence, there is a responsibility on the part of those involved in the administration of justice to be sure that they understand it (Goldacre, 2009).

Consider, as a hypothetical example, a rape case. The man accused has been told that there is a 1 in 1,000,000 chance that someone else in the UK has the same DNA profile as him – in practice, the RMP is significantly smaller. The prosecution, and the jury, may assume that this RMP equates to a 1 in 1,000,000 chance that he did not leave the semen sampled, and therefore a 999,999 in 1,000,000 chance that he is guilty.

This is, in fact, incorrect. Given an RMP of 1 in 1,000,000 of a coincidental match, and an estimated UK population of 30,000,000 men – assuming a population of roughly 60,000,000 people and an even distribution of men and women – there are in fact 30 men in the UK with that DNA profile. Therefore, 29 other men could have left the semen recovered from the victim. This therefore means that there is, in the absence of any other evidence, a 29 in 30 chance that the accused is innocent – the Random Occurrence Ratio (ROR).

The job of the prosecution is therefore to provide compelling evidence that the accused, rather than any of these other 29 men, committed that particular alleged rape. The defence's task is to undermine the prosecution's evidence, suggest that any of the other 29 men committed the offence or, as discussed in section 1.3.3, suggest that it was in fact consensual sex, countering the *mens rea* required for the criminal offence of rape to be committed.

If the judge is unhappy with how the evidence has been presented – for example, if the RMP is given, rather than the ROR, as discussed above – he or she can ask the expert witness to clarify his or her evidence or, in extreme cases of statistical ambiguity, direct the jury in how to consider it. This is reliant on the judge adequately understanding statistics, as the Prosecutor's Fallacy also applies to the judiciary.

### 1.3.6: LCN DNA profiling

Nowhere is the *Doheny* ruling more important than in the correct handling of evidence derived from Low Copy Number (LCN) DNA profiling. In essence, LCN DNA profiling is able to recover DNA from a much smaller quantity of cells than conventional DNA profiling. Obviously, this means that freedom from even the slightest contamination is of far greater importance in LCN DNA profiling – a fact recognised by the courts in *R v Hoey* [2007]. As the prosecution were unable to show – and had indeed misled the court about – the quality of their LCN DNA evidence, the judge refused to admit it (*R v Hoey* [2007], paragraphs 46, 47, 50 and 60). Many would prefer to see LCN DNA profiling used solely for intelligence and as exculpatory evidence – that is, used to eliminate suspects from inquiries or to exonerate the accused – or require a higher burden of corroboration than conventional DNA profiling (Foster, 2008). This having been said, LCN DNA profiling remains admissible in the courts, provided that the prosecution can show that contamination has not taken place.

### 1.3.7: Retention of DNA profiles

The most significant judgement on DNA profiling, however, comes from the European Court of Human Rights (ECHR). Logically, the NDNAD is only of any use in identifying offenders if it has a substantial number of profiles to compare. Firstly, a large number of profiles increases the probability that an offender is already on the NDNAD. Secondly, a larger database allows more accurate calculation of allele frequencies to abide by the *Doheny* ruling. However, as the NDNAD contains the profiles of all individuals arrested on suspicion of committing a criminal offence, and not every individual arrested is charged, let alone convicted, there are a substantial number of people on the NDNAD who are, in the eyes of the law, innocent of any crime.

In December 2008, when the ECHR passed judgement, 857,000 of the 4.3 million people on the NDNAD – approximately one in five – had never been convicted of a recordable offence (Travis, 2008). By October 2009, the total stood at 986,000 (Travis, 2009a), suggesting the addition of over 400 technically innocent individuals' profiles every day.

The ECHR declared that the indefinite retention of an innocent person's DNA profile constituted an infringement of the right to respect for his or her private life that was disproportionate to the aim of detecting and preventing crime. As such, this was incompatible with Article 8 of the European Convention on Human Rights. Although the ECHR was primarily concerned with the potential to extract significant personal information – citing ethnic identity and health details – in the future, they also ruled that the mere ability to determine family relationships meant that indefinite retention breached Article 8. In their ruling, the ECHR noted that, of all member states of the Council of Europe, the UK – excluding Scotland, which has a separate legal system and different legislation regarding DNA retention – was the only one to retain indefinitely the DNA profile of an innocent person (*S and Another v The United Kingdom* [2008]).

Accordingly, the UK was declared to be in violation of the European Convention on Human Rights and ordered to remove the DNA profile of all individuals not convicted of a recordable offence from the NDNAD. It has yet to comply with this (Travis, 2009b).

Legislation passed by the Houses of Parliament allows for the indefinite retention of all DNA profiles taken from individuals arrested on suspicion of terrorism, or convicted of any recordable offence, and the retention for 6 years of all other profiles (s14 Crime and Security Bill 2009-10, amending s64 PACE), but is unlikely to satisfy the ECHR's compliance unit as it contravenes the ECHR's ruling (Travis, 2009b). This Bill was passed as part of the 'wash up' of legislation following the calling of the 2010 general election. Its repeal, and the full implementation of the ECHR's ruling, was a manifesto pledge of both of the governing coalition parties during the election, although – at the time of writing – neither has yet happened.

The ECHR ruling did, however, confirm that the UK's policy of indefinite retention of convicted individuals' DNA profiles was, except in some circumstances where the offender was a juvenile, a proportionate measure and therefore did not contravene Article 8 (*S and Another v The United Kingdom* [2008]). An individual who has been convicted of a recordable offence, therefore, remains on the NDNAD indefinitely.

Once the ECHR ruling is implemented, it is obvious that the NDNAD will become less able to identify an offender due to the reduced number of samples that can be retained. As such, the ability to derive other forms of data from a DNA sample, to provide information about an offender beyond his DNA profile, becomes more important.

## **1.4: “Molecular Photofitting”**

---

### *1.4.1: The principle of molecular photofitting*

The principle of molecular photofitting is very simple – given that an individual’s appearance is influenced significantly by his or her genes, it is possible to deduce aspects of his or her appearance from a DNA sample. Molecular photofitting is likely to become of more value to investigating authorities following the ruling of the ECHR and the attendant reduction of profiles on the NDNAD. Already, it has been used on several occasions in the US and the UK – most notably, in providing intelligence that aided in the identification and arrest of an alleged serial rapist as part of the 17-year-long Operation Minstead, discussed in section 1.3.1 (Laville, 2009; McCartney, 2006).

An important consideration for determining whether or not to incorporate a particular marker into molecular photofitting protocols is whether or not the relevant phenotype is apparent. There is little use in, for example, noting that an offender has a predisposition to coronary heart disease, as this is extremely unlikely to be noticed by any witnesses.

A simple guide as to whether or not a marker and its expressed phenotypes are useful is to consider the descriptions of suspects released by police in an appeal for witnesses: such an appeal often includes details of the suspect’s gender, ethnicity, hair and eye colour, height, build, age, clothing and accent.

### *1.4.2: Gender*

Gender is perhaps the simplest example of molecular photofitting, and is already in use. The amelogenin gene – encoding a protein found in teeth – exists as two distinct sex-linked alleles. The X chromosome allele has a 6bp deletion in intron 1 when compared to the Y chromosome allele (Butler, 2001). The incorporation of this amelogenin marker in the CODIS and SGM+ DNA profile protocols means that, in the overwhelming majority of cases, the gender of an offender can be determined. A profile homozygous for the amelogenin marker demonstrates the presence of only X chromosomes, which almost certainly indicates that the profile derives from a female. Heterozygosity for the amelogenin marker demonstrates the presence of

both an X and a Y chromosome, which generally indicates that the source of the DNA is male.

However, there are some cases where the SRY locus, which triggers masculinisation of the foetus, has been transposed from the Y chromosome to the X during gametogenesis. This can result in male offspring with ostensibly female karyotypes – a maternally inherited X chromosome and a paternal X chromosome with a transposed SRY locus. Alternatively, female offspring could possess an apparently male karyotype – a maternal X chromosome and a paternal Y chromosome lacking the SRY locus. Individuals such as these were instrumental in locating the SRY locus and determining its function (Koopman, Gubbay, Vivian, Goodfellow and Lovell-Badge, 1991). The amelogenin test for gender may accordingly be inaccurate in these admittedly rare cases, unless the amelogenin locus itself is also transposed along with the SRY locus.

#### *1.4.3: Ethnicity and physical appearance*

Ethnic variation in conventional DNA profile markers is already significant enough to require the use of ethnicity-specific databases to calculate RMPs, as was seen in *R v Tran* [1991]. Given that the frequency of marker alleles varies between ethnic groups to such an extent, the presence or absence of ethnicity-specific markers can be used to determine an individual's ethnicity.

The allele frequency of D1S80, a 16bp VNTR, has been shown to vary considerably between tribes of Aboriginal Australians (Walsh and Eckhoff, 2007), countries in Africa (Herrera, Adrien, Ruiz and Duncan, 2004) and different races (Budowle et al., 1995). Therefore, the presence or absence of particular alleles of D1S80, and other markers, can be used to determine an individual's ethnic origin. The UK Border Agency is to use ethnicity-specific DNA testing to determine an individual's country of origin within Africa to verify claims for asylum (Doward, 2009).

Ethnic typing is most accurate when there is a history of separation between populations. For example, the ability of D1S80 to determine which tribe an Aboriginal Australian belongs to is a consequence of the fact that, due to geographical isolation, these tribes have not significantly interbred (Diamond, 1998). Therefore, each tribe retains a distinct pool of DNA profile markers that allows the identification of an individual's tribal origin (Walsh and Eckhoff, 2007).

Some basic assumptions can be made regarding an individual's appearance based on ethnic typing. The skulls of East Asian and African individuals have distinct shapes, resulting in characteristic facial traits – the prognathoid jaw is common to African individuals, while East Asians have a rounder, flatter face than Europeans. Assumptions can also be made about skin, hair and eye colour in African and East Asian individuals, who generally have dark hair and dark eyes (Frudakis, 2008).

Skin colour can be predicted based on more specific ethnic typing: North Africans, often referred to as “whites”, have lighter skin and straighter hair than Central African “blacks” and “pygmies” – the terms, while slightly objectionable, are still commonly used in anthropology – while the Khoi and San of Southern Africa have yellowish skin (Diamond, 1998). Similarly, the Han majority have distinctive skin tones when compared with many of the other 55 ethnic groups that make up the population of China (Man, 2004).

Determination of aspects of physical appearance by ethnic typing struggles, however, in European populations and those of former European colonies such as the United States and Australia. Even leaving aside the fact that European and former colonial populations show the greatest diversity in hair, skin and eye colour (Frudakis, 2008), the differences in DNA marker allele frequencies closely mirror the differences in native languages (Cavalli-Sforza, 2000).

It is comparatively easy to distinguish between an Egyptian and a Cameroonian by their respective D1S80 allele frequencies (Herrera et al., 2004), but also because Egypt's native languages are primarily Afro-Asiatic and Nilo-Saharan in origin, while Cameroon's are Bantu (Diamond, 1998). Similarly, Western Russians can be differentiated from Eastern Russian's primarily Asiatic populations both on the basis of allele frequency (Verbenko et al., 2006) and language – Western Russians speak Slavic languages, while Eastern Russians retain Asiatic tongues of the Altaic and Uralic-Yukaghir families (Cavalli-Sforza, 2000).

#### *1.4.4: Ethnicity and appearance in Western Europeans*

A few European populations can be distinguished easily from their neighbours. People of Basque origin retain a language – Euskara – that does not fall into any extant linguistic family, and as a result of cultural isolation from the surrounding Spanish and French populations, possess a unique set of allele frequencies. The

Ashkenazim and Romani also tend to marry within their own populations, and retain a distinct genetic identity. Icelanders, due to centuries of geographical isolation, can be distinguished from the Scandinavians from which the Icelandic population originates by allele frequencies as a result of genetic drift. Icelandic is similarly different to other Scandinavian languages (Cavalli-Sforza, 2000; Diamond, 1998; Frudakis, 2008).

The main languages of Western Europe belong primarily to the Indo-European family, and demonstrate a history of interbreeding and population movements. French, Italian, Spanish and Portuguese are closely related, due to the influence of the Roman Empire and Latin, as are the Scandinavian languages. Attempts to determine the area of origin within Europe of an individual, despite the use of 320 markers, struggle to reliably distinguish between speakers of Latin-derived languages of different nationalities (Frudakis, 2008).

At best, an individual can be described as Iberian, Mediterranean, North European or Scandinavian. Iberia includes Spain and Portugal, while the Mediterranean group includes Southern France, Italy and parts of the Balkans. Northern Europe effectively incorporates the rest of Europe excluding Scandinavia, which retains a distinct genetic identity. More accurate determination of an individual's nationality is significantly more difficult (Frudakis, 2008).

The population of the United Kingdom – and, by extension, much of the Caucasian population of former colonies such as the United States, Canada, Australia and New Zealand – is significantly more homogeneous than most other European populations. English itself is primarily Germanic in origin, although it has been influenced by Norse and Latin dialects (Diamond 1998), and a study of surnames and place names shows the presence of Roman, Viking, Anglo-Saxon and Norman colonists (Davies, 1999).

Three hundred aristocratic families can prove descent from William the Conqueror (Jones, 1996), while the relatively small numbers of Norman invaders in 1066 secured power by marrying local Anglo-Saxons (Davies, 1999). The army brought to England by William of Orange during the Glorious Revolution of 1688 was around twice the size of the then population of Birmingham (Levenson, 2009). Many ethnic groups other than those that might be considered historically “British” have been present in enough numbers to interbreed, blurring the distinctions between them, even before the population movements that resulted from Empire, colonisation, immigration and emigration are considered. In addition, as a relatively small country, the UK has had a historically mobile population (Davies, 1999), and

internal population movements from the countryside to the cities during the Industrial Revolution will have increased the opportunities for interbreeding, reducing the distinctiveness of marker allele frequencies.

Determining the origin of a Western European individual is therefore more difficult and provides less useful information about appearance – aside from skin colour – than in most other ethnic groups.

#### *1.4.5: Eye and hair colour*

The OCA2 gene has 33 single nucleotide polymorphisms (SNPs) associated with eye colour. By studying these SNPs, investigators can predict to a high degree of accuracy the colour and shade of an individual's eyes (Frudakis, 2008; Liu et al., 2009). For the most common eye colours, such as blue or brown, the accuracy of these predictions can be as high as 95%. For other, intermediate eye colours such as hazel, the accuracy is approximately 75%.

Attempts to isolate mutations specific to a particular hair colour have not been as successful. So far, only the MC1R gene has been of any use – analysis of mutations in this gene have been able to determine whether or not the individual has red hair (Frudakis, 2008). However, given that red hair is not the commonest even in Western European populations, who show the most variety in hair and eye colour, a test based on the MC1R gene does not immediately appear to be of significant value in the majority of cases. Even if a reliable genetic test for hair colour could be developed, it would be of limited value since hair may be dyed, shaved off, turned grey or lost. Knowing that the offender is genetically predisposed to possess brown hair is of little use if the suspect is bald.

#### *1.4.6: Height, weight and build*

Other commonly reported characteristics, such as height or build, are affected to varying degrees by environmental as well as genetic factors (Ridley, 2003). While 80-90% of the variation in height between individuals is thought to be caused by genetic factors (Le Fanu, 2010), the environment still plays a significant role – for example,

the effects of a poor diet or maternal tobacco smoking during pregnancy, are both well known to limit growth in children (Scheur and Black, 2000). To date, less than 5% of the variation in height can be accounted for by 40 different gene alleles (Le Fanu, 2010) and an accurate prediction of height from a DNA sample is currently impossible.

Similarly, an individual's build is affected by diet and the amount of exercise that he or she does. Like height, 80-90% of the variation between individuals in weight is thought to be genetic in origin (Le Fanu, 2010), but – obviously – diet and lifestyle are significant factors. In common with height, many genetic markers have been identified, but do not account for much of the variation between individuals. This is in fact a common observation regarding polygenic traits – the level of heritability of variation in a trait has not, in many cases, been fully accounted for, even when dozens of markers have been investigated and identified (Maher, 2008).

The principle genetic indicators of obesity between races are the so-called “thrifty genes” which promote the efficient storage of excess dietary sugar and fat (Sapolsky, 1997), and are most common in people of African origin (Frudakis, 2008). More accurately, the Western European versions of these genes have been described as “sloppy”, as Western metabolisms appear more wasteful than those of Africans.

Even possession of “thrifty genes”, or the absence of “sloppy genes”, does not actually show that an individual is fat – rather, that exposure to an extravagant Western diet is likely to lead to obesity (Sapolsky, 1997). The key influence is diet, rather than an individual's genotype.

#### 1.4.7: Age

While an individual's genotype may give some information about life expectancy – indeed, this was one of the issues considered by the ECHR in its ruling on indefinite retention of DNA profiles (*S and Another v The United Kingdom* [2008]) – the presence or absence of particular alleles cannot indicate the age of an individual. It is, however, apparent that a reliable assay to determine age will be of use as a component of molecular photofitting in the context of criminal offences.

## Chapter Two: Telomere Length Analysis

---

<b>2.1</b>	<b><i>Telomeres – Structure &amp; Function</i></b>	<b>22</b>
2.1.1	Telomeric structure	22
2.1.2	Maintenance of chromosome structure	22
2.1.3	3'-end replication	24
2.1.4	Telomerase	27
2.1.5	Telomeres limit mitosis	27
2.1.6	Cell senescence	28
2.1.7	Telomere Length Analysis	29
2.1.8	Assays for telomere length	30
<b>2.2</b>	<b><i>The Telomere Restriction Fragment Assay</i></b>	<b>31</b>
2.2.1	Restriction digestion of telomeres	31
2.2.2	Limitations of the TRF assay	32
<b>2.3</b>	<b><i>The Real-Time Quantitative PCR Assay</i></b>	<b>34</b>
2.3.1	Primer-dimers	34
2.3.2	Amplifying long repetitive sequences	36
2.3.3	Telomeric PCR	37
2.3.4	The need for a control amplification	40
2.3.5	Real-time PCR	41
2.3.6	Problems with the qPCR assay	41
<b>2.4</b>	<b><i>The Telomeric Multiplication Factor Assay</i></b>	<b>42</b>
2.4.1	Introduction	42
2.4.2	The qPCR equation	42
2.4.3	Intercalating dyes	43
2.4.4	The telomeric multiplication factor	43
2.4.5	The TMF equation	44
<b>2.5</b>	<b><i>Aims &amp; Objectives</i></b>	<b>46</b>

## 2.1: Telomeres – Structure & Function

---

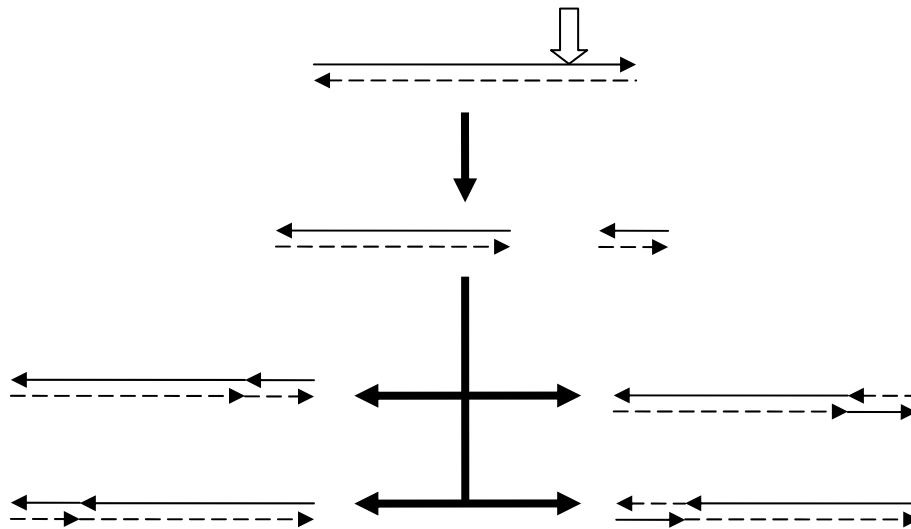
### 2.1.1: Telomeric structure

Telomeres are structures unique to linear eukaryotic chromosomes. In humans, they consist of between 4 and 15kbp of repeats of the hexameric sequence 5'-TTAGGG-3' (Pollard and Earnshaw, 2004) and many associated proteins – these include TRF1, TRF2 and POT1, which specifically recognise and bind to the telomeric repeat sequence, and TIN2, PTP/PIP1 and PINX1, which associate with the telomere-binding proteins (Rodier, Kim, Nijjar, Yaswen and Campisi, 2004). In addition, the nucleosomes of telomeres have a shorter repeat unit than conventional nucleosomes, resulting in densely packed heterochromatin (Wolffe, 1998). Furthermore, the 3'-end of the telomere is longer than its complementary strand, forming a single-stranded region of 100-200bp in length – the 3'-overhang. This overhang is thought to loop back on itself and invade the double-stranded region of the telomere, forming a short triple-stranded stretch known as the t loop (Lewin, 2000; Rodier et al., 2004).

This combination of the t loop and specific telomere-binding proteins, and the proteins that associate with them, serves to give the telomere a structure distinct from the bulk of the eukaryotic chromosome. It is this structure that allows the telomere to play a crucial role in the maintenance of genomic integrity (Rodier et al., 2004).

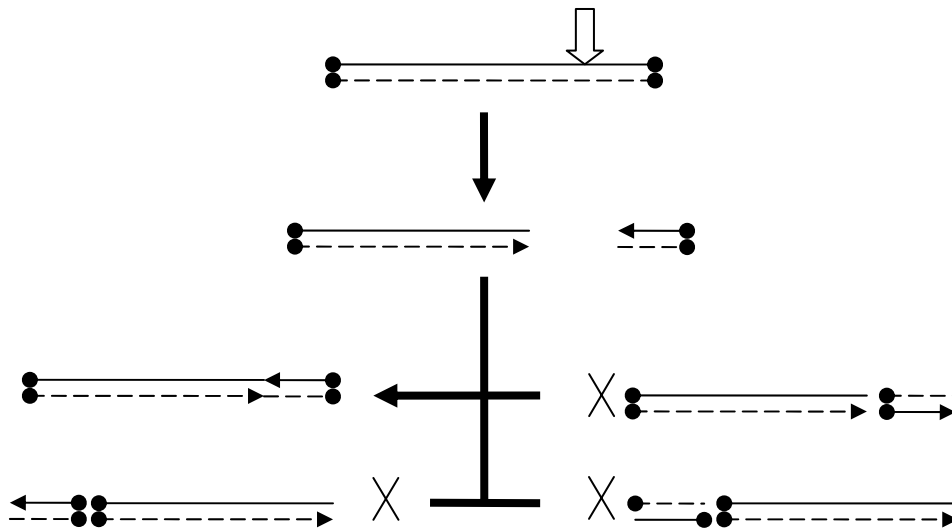
### 2.1.2: Maintenance of chromosomal structure

The linear chromosomes of eukaryotes are harder to maintain than the circular chromosomes of prokaryotes. Perhaps the simplest illustration of this is to consider what happens when a chromosome breaks. In a prokaryote, a double-stranded break in a circular chromosome produces two ends, which can be reconnected without the risk of disrupting the genetic information located in that region. In a linear chromosome, however, a double-stranded break produces *four* ends – consequently, there are four ways the fragments can be ligated together (see Figure 2.1.1, overleaf).



*Figure 2.1.1:* Multiple repair options of a double-stranded break in a linear chromosome. If the chromosome breaks (represented by the white arrow), the two fragments can be reassembled in four different ways. If a gene straddles the point at which the chromosome breaks, it stands a significant chance of being disrupted, leading to the loss of its function, or being fused to a different promoter, resulting in inappropriate activation.

The telomeric structure forms a 'cap' that the eukaryotic cell recognises as the 'normal' end of a chromosome. Therefore, a double-stranded break results in two capped ends and two uncapped ends. By inhibiting the DNA repair response, telomeres resist being ligated to other telomeres or uncapped ends (Chan and Blackburn, 2002), which results in the chromosome being reassembled correctly (see Figure 2.1.2, overleaf). As chromosomal abnormalities still arise, telomeric inhibition of the double-stranded break repair pathways is not absolutely effective, although it undoubtedly reduces the frequency of these abnormalities.



*Figure 2.1.2:* The role of telomeres in the maintenance of chromosome structure. As in Figure 2.1.1, a chromosome suffers a double-stranded break. But, in this case, the chromosome has telomeric caps (represented by circles) which inhibit DNA repair. Thus, the most frequent way chromosomal fragments can be reassembled is in the original order and orientation.

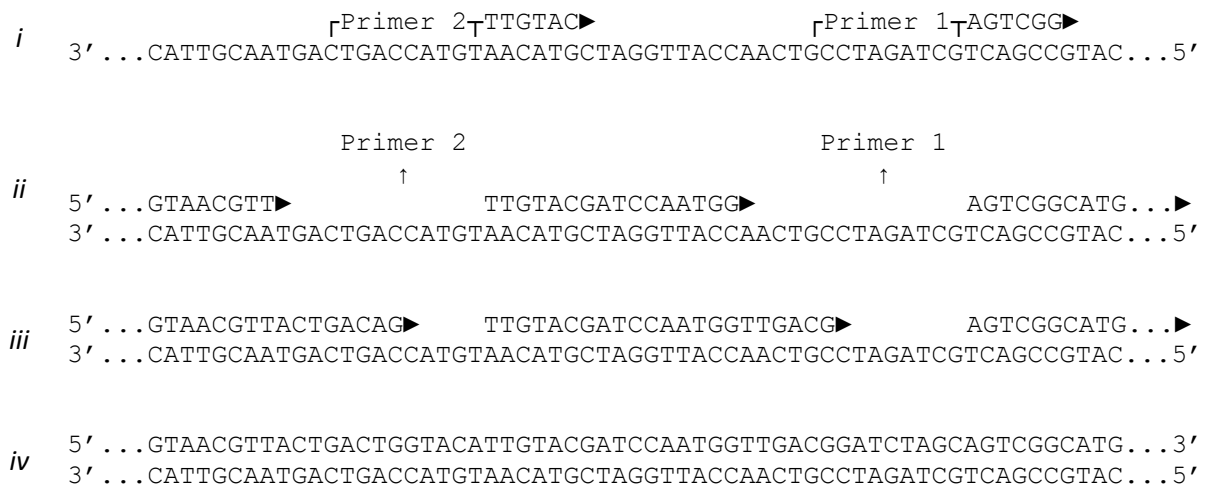
By indicating a normal end to the chromosome, telomeres also prevent multiple linear chromosomes being ligated together. The fusion of two chromosomes results in a structure with two active centromeres – a dicentric chromosome. At mitosis, a dicentric chromosome risks being pulled to both poles of the cell at the same time, resulting in chromosomal breakage (Gisselsson, 2004).

Therefore, telomeres maintain the integrity of linear chromosomes, which would otherwise be threatened by the cell's double-stranded DNA repair response.

### 2.1.3: 3'-end replication

A second issue that arises from linear chromosomes is a consequence of the way DNA is replicated during normal cell division. DNA polymerases can only replicate DNA by the addition of DNA nucleotides to the 3'-end of an RNA primer. The region of template DNA to which the primer binds cannot be replicated by that DNA polymerase, as they can only progress from the 3'-end of the template towards the 5'-end. Again, this does not present difficulties to a circular chromosome, because the DNA polymerase will eventually return to the primer-binding sequence (PBS),

after replicating the rest of the chromosome and arriving back where it started. Returning to the point at which DNA replication started is not, however, possible when the chromosome is linear. Instead, the PBS is replicated by another DNA polymerase that started replication further towards the 3'-end of the template DNA (see Figure 2.1.3).



*Figure 2.1.3:* Outline of linear DNA replication from multiple primer-binding sequences (PBSs). DNA polymerases (represented by arrowheads) replicate from the 5'-end of RNA primers (*i*). After the primers dissociate from the template DNA, the PBS remains unreplicated (*ii*). But DNA polymerases proceeding from further towards the 3'-end of the template DNA are able to use PBSs towards the 5'-end as templates, and so replicate these PBSs (*iii*). Following ligation of the newly synthesised DNA into a single strand (*iv*), replication is complete (adapted from Alberts et al., 2002).

However, DNA will be lost from the extreme 3'-end of the template (see Figure 2.1.4, overleaf). This is because there can be no additional DNA polymerase proceeding from the 3'-end to replicate the PBS – a phenomenon known as the 3'-end replication problem (Alberts et al., 2002).



although the fact that the mean rate of telomere attrition is 50bp per mitosis suggests that there is some mechanism preventing excessive loss of telomeric DNA.

#### *2.1.4: Telomerase*

Over time, the telomeres would be eroded to such an extent that they cease to offer any protection to functional DNA. A mechanism must therefore exist to restore telomeres to their full length from time to time. Telomerase is a reverse transcriptase that, in large multicellular organisms, is only active in a few tissues, rebuilding telomeres using an RNA template for the telomeric sequence (Pollard and Earnshaw, 2004). It is active during gametogenesis, thereby ensuring that the next generation start with telomeres sufficiently long to protect functional DNA from erosion. It is also active in some tissues that have notably high rates of cell attrition. In particular, telomerase is active in epithelial cells in the large intestine, where cells abraded away by the passage of food must be replaced frequently (Pollard and Earnshaw, 2004).

The importance of telomerase has been illustrated by knocking out the RNA template it uses as a template to reconstruct telomeres in mice. After six generations, the mice simply became sterile – gametogenesis was impossible as precursor cells became senescent due to telomere erosion (Pollard and Earnshaw, 2004). Obviously, in smaller multicellular organisms that reproduce by budding such as *Hydra*, or unicellular eukaryotes such as *Amoeba* or yeasts (Campbell, 1996), telomerase is constitutively active.

#### *2.1.5: Telomeres limit mitosis*

The reason that telomerase is not constitutively active in larger multicellular organisms is because the telomeres act as a brake on cell division. Cancer is, put simply, a disease of increased cell division, and is accompanied by greater loss of telomeric DNA. Once the telomeres drop below a critical length, or are disrupted by damage to the nucleosomes, they trigger checkpoint proteins that arrest cell division. These checkpoint proteins, such as p53, stimulate DNA repair or cause cell senescence or apoptosis. In order for cancer to develop, therefore, cells must be

immortalised by the activation of telomerase or the inactivation of these checkpoint proteins (Hubbard and Ozer, 1999). This allows cell division to continue uncontrolled and permits a cancer to develop (Rodier et al., 2004).

Telomeres also act as a barometer of chromosomal mutation. Mutation events are more likely to occur in the telomeres than individual coding sequences of DNA, because telomeres constitute a greater proportion of chromosomal DNA. At the upper length of 15kbp, the 92 telomeres in a human diploid genome – two per chromosome – constitute 1.38Mbp, approximately 0.02% of the total 6Gbp. Any single gene, when ignoring any introns present, represents less of the genome than the total telomeric content, but a significantly mutated telomere suggests that a substantial proportion of the coding DNA has been damaged. Similarly, telomeres shortened by erosion indicate DNA damage, which accrues at each cell division. Therefore, short or damaged telomeres suggest that functional DNA has probably been mutated (Nakamura et al., 2008). The triggering of checkpoint proteins in this situation serves to prevent defective genes having a deleterious effect on the organism, and leads to cell senescence. Notably, longer telomeres have been shown to inhibit p53 more effectively than shorter ones.

In order to protect the genome adequately, the telomere must not be extended too much during gametogenesis – a longer-than-normal telomere would take more time to degrade enough to trigger cell senescence, which increases the risk of cancer. Telomerase activity is therefore regulated during gametogenesis, probably by the interaction between TRF1 and POT1 (Lundblad, 2003).

#### *2.1.6: Cell senescence*

Cell senescence has been linked to aging. Senescent cells do not replicate to replace or repair worn out, damaged or defective tissues, and it is this failure to produce new cells that has been implicated in many of the effects of aging. The limit on cell divisions observed in non-immortalised cell lines – known as the Hayflick Number – has been shown to correlate to the lifespan of an organism, and cells cultured from older organisms reach this limit more quickly than cells taken from young organisms (Brookbank, 1990).

Some diseases of increased rates of aging, such as Werner syndrome and dyskeratosis congenita, have been shown to originate in defective proteins that

interact improperly with telomeres (Chang et al., 2004; Mitchell, Wood and Collins, 1999). In addition, degenerative defects similar to those present in dyskeratosis congenita can be induced by the artificial shortening of telomeres in mice (Armanios, Alder, Parry, Karim, Strong and Greider, 2009). Some symptoms of aging, such as vascular disease, immunosenescence and the reduction in ability to heal wounds, have also been associated with shortened or defective telomeres (Rodier et al., 2004). The artificial induction of telomerase has been shown to reverse several aspects of premature aging in mice born with defective telomeres (Jaskelioff et al., 2011).

These suggest a key role for telomeres in regulating the lifespan of an organism, although there is no obvious relationship between telomere length, telomerase regulation and life expectancy, other than the fact that longer telomeres stave off cell senescence for longer, and so slow the onset of age-related physical symptoms (Rodier et al., 2004).

Logically, however, cells with higher rates of mitosis must lose less of their telomeres at each division than cells with slower rates of mitosis potentially can – in order to retain a pool of cells to replace those lost through damage or attrition, these cells must remain mitotically active for longer, and so those with longer telomeres will be selected for as they do not become senescent as rapidly as those with shorter telomeres. Therefore, while telomere lengths may be highly heterogeneous in an organism (Wolffe, 1998), it is expected that, within mitotically active cells such as epithelial or white blood cells (Lewin, 2000), telomere length will in fact be able to estimate the number of cell divisions since gametogenesis. It is this that forms the basis of Telomere Length Analysis.

### *2.1.7: Telomere Length Analysis*

Given that telomere length has been shown to decrease by, typically, 50bp per cell division, this suggests that the length of an individual's telomeres in high turnover cell lines, such as epithelial cells or leucocytes, will be inversely proportionate to that individual's age. This presents the intriguing possibility that telomere length can be used as part of the molecular photofitting suite of assays to determine the age of an unknown offender. The use of telomere length to predict age can be referred to as Telomere Length Analysis (TLA).

Previous studies have variously succeeded (Tsuji, Ishiko, Takasaki and Ikeda, 2002) or failed to predict ages reliably from telomere length (Karlsson, Svensson, Marklund and Holmlund, 2008). Potentially, flaws in the assay used by Tsuji et al. (2002) and the methods of analysis used by Karlsson et al. (2008) explain these contradictory findings, although these will be considered in sections 2.2.2 and 4.2.1-4.2.6.

#### *2.1.8: Assays for telomere length*

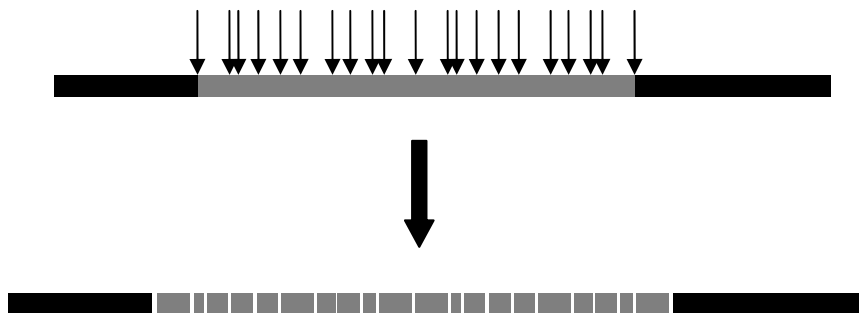
In order for TLA to become a viable forensic technique, there are a number of obstacles that must be overcome. Firstly, telomere measurement must be easy and straightforward, and preferably capable of being performed at the same time as conventional DNA profiling, to allow its easy integration into established forensic procedures. Secondly, there must be a method to predict age from telomere length that allows a reporting scientist to give the probability of the prediction being accurate, so that investigating authorities apply the appropriate weight to it.

There are two main techniques used to measure telomere length: the telomere restriction fragment (TRF) assay (Kill and Faragher, 2000) and the real-time quantitative PCR (qPCR) assay developed by Cawthon (2002).

## 2.2: The Telomere Restriction Fragment Assay

### 2.2.1: Restriction digestion of telomeres

The principle of the TRF assay is simple and straightforward, and relies on the fact that an individual's telomeres consist entirely of thousands of repeats of the sequence 5'-TTAGGG-3'. The majority of the rest of the genome can be digested by restriction enzymes that do not recognise this sequence, leaving telomere restriction fragments that can be sorted and measured by electrophoresis. For example, the restriction enzyme *RsaI* has the recognition sequence 5'-GTAC-3'. Based on a random distribution of DNA bases, this sequence would be expected to arise once in every 128bp – 4<sup>4</sup>bp – yielding restriction fragments of around this length. However, this sequence does not occur in the telomeres, so these would remain intact as TRFs of 4 to 15kbp in length. After digestion of a chromosome with a battery of restriction enzymes that do not recognise the telomeric sequence, the only restriction fragments of any substantial size to remain will be TRFs (see Figure 2.2.1).



*Figure 2.2.1:* Illustration of the TRF assay. A hypothetical chromosome carries 21 restriction sites (indicated by arrows) for a particular restriction enzyme in genomic DNA (shaded grey). However, no restriction sites are present in the telomeres (shaded black). Therefore, digestion of the chromosome results in two large TRFs, and 20 smaller fragments. In practice, several restriction enzymes are used, which cut genomic DNA frequently – there would in consequence be many more genomic restriction fragments, of much smaller size relative to the TRFs. Electrophoresis is used to separate restriction fragments according to size – many genomic fragments would be lost from the bottom of the gel, but the TRFs, being significantly larger, will not migrate as far. The sizes of the TRFs can then be determined.



radiolabelled probes or substantial exposure times. Cawthon (2002) reports the use of <sup>32</sup>P-labelled DNA probes with exposure times of up to 5 days. The TRF assay can therefore take appreciably more time than conventional DNA profiling procedures. This is unlikely to greatly impede its use as a part of molecular photofitting, however, as many markers must be identified in order to determine an offender's appearance. Nevertheless, the use of radioisotope labels introduces further problems of licences, training, safety and disposal, which is likely to increase the cost of TLA.

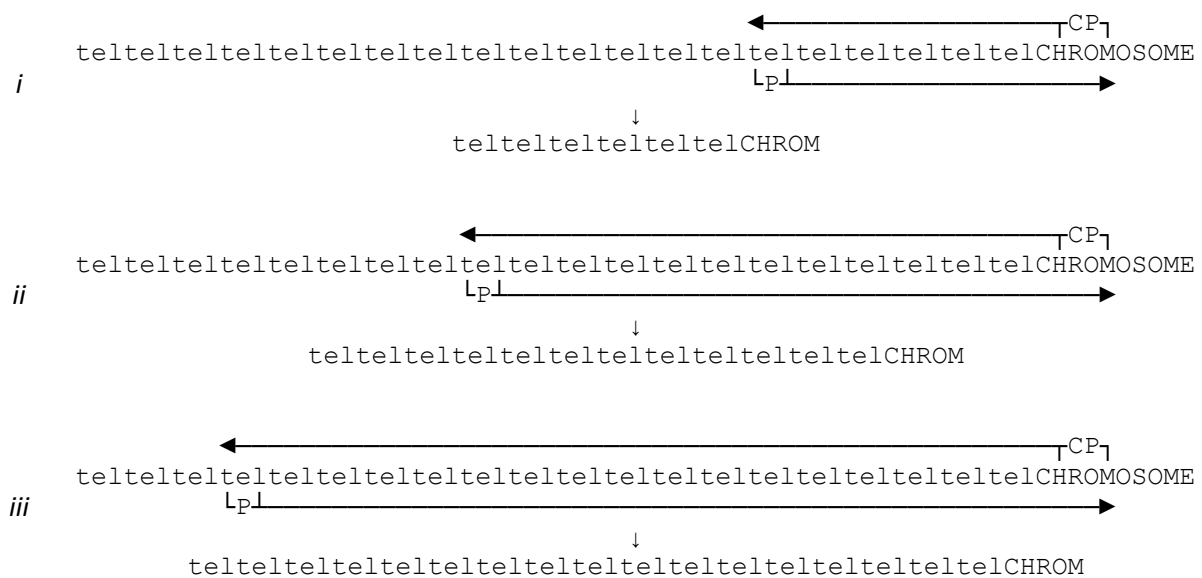




### 2.3.2: Amplifying long repetitive sequences

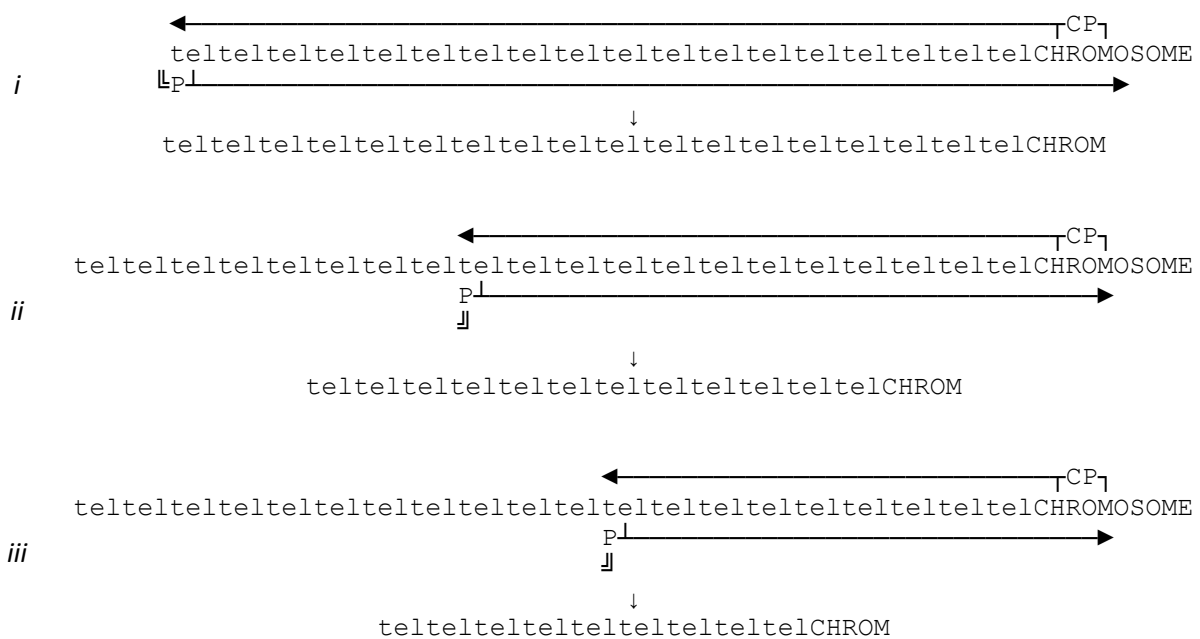
Resolving the primer-dimer amplification problem does not in itself alter the fact that amplifying by PCR a stretch of repetitive DNA several thousand kilobase pairs in length is not feasible. Amplifying such a lengthy sequence would require an extremely long extension phase in the PCR cycle, but DNA polymerases are known to amplify longer sequences less efficiently than shorter ones – a phenomenon known as allele drop-out (Butler, 2001).

In addition, when considering repetitive DNA such as a telomere, it is impossible to ensure that primers anneal to the extreme ends of the sequence. It is, of course, perfectly feasible to target one primer to the chromosomal end of the telomere, by making it complementary to unique chromosomal DNA sequences. However, the other primer will be able to anneal to any point of the telomere, because it is complementary to any point of the repetitive telomeric sequence, resulting in amplicons of wildly varying lengths (see Figure 2.3.4).



*Figure 2.3.4:* Failure to consistently amplify repetitive DNA. In this example, one primer (CP) is complementary to chromosomal DNA just downstream of the telomere (CHROMOSOME), while the other (P) is complementary to the repetitive telomeric DNA (indicated by repeats of tel). Therefore, the second primer can anneal at any point of the telomere's length (*i*, *ii* and *iii*). Consequently, PCR can produce amplicons of variable size.

Because nucleic acids are flexible molecules, it is not possible to design a primer that will target the extreme end of a telomere by making it only partly complementary to the telomere sequence in the hope that the non-complementary part means that the primer can only anneal at the telomere end. Instead, the primer can still anneal at any point, with the non-complementary bases bent away from the template DNA (see Figure 2.3.5).



*Figure 2.3.5:* Partly complementary primers do not target telomere ends. A primer that is both complementary (indicated by a single line) and not complementary (indicated by a double line) might be expected to target the extreme end of a telomere, as this allows base pairing by the complementary part while the non-complementary part does not interfere. The only place where this is possible is at the end, where the non-complementary part forms an overhang (i). However, RNA is flexible and the non-complementary part of the primer can curve away from the template, and therefore does not interfere with the primer binding at internal sequences (ii and iii). Hence, the amplicons produced will still be variable in length.

### 2.3.3: Telomeric PCR

The solution, developed by Cawthon (2002) is not to target chromosomal DNA, or to design primers that can only anneal at the end of the telomere, but to target only

repetitive telomeric DNA with both primers. Obviously, this produces amplimers of many and varied lengths, but the single most frequently produced amplimer is in fact 76bp in length, produced by the two primers annealing close together (see Figure 2.3.6, overleaf).

This amplimer comes to dominate the PCR reaction for two reasons. Firstly, as the smallest possible complete amplimer, it is the least likely to be affected by allele drop-out. Essentially, if a long sequence is amplified with 90% of the efficiency of a shorter sequence, then, after 20 cycles of PCR, only 12% of the total number of amplimers –  $0.9^{20}$  – will be derived from the longer sequence. After 30 cycles, 4% of the total number of amplimers –  $0.9^{30}$  – will be derived from the longer sequence. The shorter sequence therefore comes to dominate the reaction.

Secondly, telomeric primers can anneal to any point of the telomere. The 76bp amplimer results from the two primers annealing close together, and this configuration will occur more frequently than those of primers annealing further apart, because it occupies less of the telomere's length than longer configurations. This means that a single telomere can produce more of a shorter amplimer than of a longer one in each cycle of PCR.

A telomere of 7,600bp in length can generate 100 76bp amplimers in any single cycle of PCR, because that length allows 100 sets of primers annealing in the appropriate configuration. However, the same telomere can only generate a maximum of 50 152bp amplimers, even if allele drop-out is not considered, because each primer-template complex that results in these amplimers occupies twice as much template DNA as those resulting in 76bp amplimers. Even considering amplimers of 82bp – which result from the telomeric primers annealing separated by 6bp, or a whole telomeric repeat – a telomere of 7,600bp in length will only generate, at most, 92 amplimers per cycle of PCR. A telomere of 11,400bp in length can generate up to 150 76bp amplimers, but only 139 or 75 copies of the 82bp or 152bp amplimers respectively.

Because of the exponential nature of PCR, where each strand of every amplimer acts as a template for replication, the most commonly produced amplimer – in this case, the 76bp amplimer – will quickly dominate the reaction.

The number of copies of this 76bp amplimer produced by PCR is dependent on the number of loci from which it can be amplified. Logically, a longer telomere has more loci that give rise to 76bp amplimers than a shorter one, and so the quantity of 76bp amplimers produced by PCR is proportionate to the length of the telomere.



*Figure 2.3.6 (previous page):* Multiple amplimers arising from telomeric PCR. Amplimer *i*, at 76bp, is the most likely to arise, as the template-primer complexes required for this amplimer are the smallest possible. Logically, therefore, this amplimer will be produced more than amplimers *ii* or *iii*, formed by separation of primers by whole hexameric repeats of the telomeric sequence and which require longer template-primer complexes. Amplimer *iv*, the product of separation of the primers by an incomplete hexameric repeat, is less likely to arise because base pairing is reduced for one primer or the other, reducing the probability of this primer-template complex forming. Other primer-template complexes, as in *v*, result in the 3'-end of one primer or the other not stably annealing to the template (indicated by an arrow), preventing replication by DNA polymerase.

Comparisons of the TRF and qPCR assays for telomere length show a highly significant correlation between the two very different assays. The qPCR assay should be favoured in TLA and molecular photofitting, though, because it does not require radioisotope labelling or lengthy exposure times.

#### *2.3.4: The need for a control amplification*

If, as discussed in section 2.3.3, a 7,600bp long telomere can produce 100 76bp amplimers during each cycle of PCR, it is apparent that the same number of amplimers can be generated by two telomeres, each of 3,800bp in length. Therefore, without a method to determine the relative numbers of telomeres present, a 7,600bp telomere cannot be distinguished from two 3,800bp telomeres.

Accordingly, a control amplification of a single-copy locus marker is required. The most commonly used single-copy locus marker is 36B4, part of the gene encoding acidic ribosome phosphoprotein PO. 36B4, located on Chromosome 16, is an accepted single-copy marker for calculating gene dosage (Boulay, Reuter, Ritschard, Terracciano, Herrmann and Rochlitz, 1999) and so can be used to estimate the number of 76bp amplimers produced during telomeric PCR.

The quantity of the 36B4 amplimer depends on the number of cells present and can therefore be used to distinguish between one cell with telomeres averaging 7,600bp in length and two cells with average telomere lengths of 3,800bp. The ratio between the quantity of 36B4 and 76bp telomeric amplimers is proportionate to the average telomere length in a sample of cells.

### *2.3.5: Real-time PCR*

The method proposed by Cawthon (2002) uses real-time PCR to determine telomere length. Real-time PCR allows the quantity of amplimers produced at each cycle of the PCR reaction to be determined by fluorescent densitometry or spectrophotometry.

To determine telomere length, Cawthon first defined a critical intensity, above baseline fluorescence. Because fluorescence is proportionate to the quantity of DNA present, the intensity increased with each cycle of PCR. The number of PCR cycles at which the intensity equalled the previously defined critical intensity threshold was defined as  $C_t$ . As there are more loci available for the 76bp telomeric amplimer than for the single-copy locus control, telomeric PCR achieved this critical intensity in fewer cycles than the control PCR.

The difference between the two numbers of cycles – variously described as  $\Delta C_t$  (Cawthon, 2002), or the T/S (Telomeric and Single-copy) ratio (Karlsson et al., 2008) – was proportionate to telomere length as measured using the TRF assay.

### *2.3.6: Problems with the qPCR assay*

The real-time qPCR assay requires the use of a real-time PCR thermal cycler, which is both expensive and markedly different in its use to the techniques used in standard DNA profiling. A DNA profile, arising as it does from polymorphic STR alleles, requires electrophoresis following the end point of the PCR reaction to determine amplimer sizes. Similarly, many of the markers used in molecular photofitting are STRs, VNTRs or RFLPs, which require electrophoresis to determine amplimer or restriction fragment length. SNPs used in molecular photofitting require the use of sequencing procedures, which are different to both real-time PCR and conventional DNA profiling techniques. Therefore, a real-time PCR thermal cycler would be used by forensics bodies only to determine telomere length, representing an investment of money and training that might be considered disproportionate to any benefit.

## ***2.4: The Telomeric Multiplication Factor Assay***

---

### *2.4.1: Introduction*

The use of the real-time qPCR assay represents a considerable investment in equipment and training. An assay based on the end point of a PCR reaction represents the easiest and most efficient way to integrate TLA into a forensic investigation. To this end, the telomeric multiplication factor (TMF) assay is proposed, as a technique for determining telomere length using the tools and skills already in use in DNA profiling.

The TMF assay is based on the same principles as the qPCR assay. However, instead of determining the difference in the number of cycles of PCR required for telomeric and single-copy locus amplimers to reach a predetermined quantity, telomeric and single-copy locus PCRs are run for set numbers of cycles, and the relative quantities of telomeric and control amplimers used to determine telomere length.

### *2.4.2: The qPCR equation*

The TMF assay derives from the equations used to calculate copy number or cell quantity from a quantitative PCR reaction. The conventional qPCR equation describes how the relationship between the quantity of template DNA and amplimers ( $X_T$ ) increases with the number of cycles of PCR ( $C$ ), depending on the initial quantity of DNA templates ( $T_0$ ) – a product of the number of cells present and the copy number of the PCR loci – and the efficiency of amplification ( $E$ ), based on the kinetics of the PCR reaction. According to Bustin and Nolan (2004b), this relationship can be described by the equation:

$$X_T = T_0 (1+E)^C$$

### 2.4.3: Intercalating dyes

Intercalating dyes, such as ethidium bromide or GelRed, insert molecules into double-stranded DNA. This insertion alters the fluorescent properties of the dye molecules, which allows the visualisation of double-stranded DNA in a gel. However, the intensity of fluorescence of a band in a gel depends on not only the quantity of amplimers present – a greater number of amplimers allows more dye molecules to intercalate into DNA – but also the length of the amplimers – longer amplimers allow the intercalation of more dye molecules than shorter ones (Bustin and Nolan, 2004a). This explains why quantitative PCR ladders, such as the Norgen PCRSizer 100bp ladder, typically only allow the mass of DNA present in a band to be determined: the intensity of fluorescence is a product of both amplimer length and number.

Based on these observations, it is possible to express the relationship between intensity of fluorescence ( $I$ ), amplimer length ( $L$ ), amplimer quantity ( $n$ ) and the particular fluorescent properties of the intercalating dye used ( $k$ ) as an equation:

$$I = kLn$$

As the amplimer quantity in the densitometric equation ( $n$ ) is equal to the amplimer quantity in the qPCR equation ( $X_T$ ), these equations can be combined to give:

$$I = kLT_0(1+E)^C$$

For the single-copy locus PCR required as a control for any qPCR assay for telomere length, this can be written as:

$$I_S = kL_S T_0 (1+E_S)^{C_S}$$

### 2.4.4: The telomeric multiplication factor

For the equation to be applied to a telomeric PCR, there are two amendments that must be made. Firstly, there are 46 telomeres for every copy of the single-copy locus in humans – each haploid genome will contain 1 single-copy locus and 23

chromosomes, each with 2 telomeres. Therefore, in a telomeric PCR, the term  $T_0$  must be replaced with  $46T_0$ .

Secondly, a term must be introduced to take into account the fact that a single telomere can give rise to many 76bp amplicons – the telomeric multiplication factor, or TMF ( $\tau$ ). Based on the work of Cawthon (2002), as discussed above, it is apparent that TMF will be proportionate to mean telomere length. By incorporating TMF into the equation for telomeric PCR, it can therefore be expressed as:

$$I_T = \tau k L_T 46 T_0 (1 + E_T)^{C_T}$$

#### 2.4.5: The TMF equation

In both the single-copy locus and telomeric PCR equations,  $k$ ,  $T_0$ ,  $E_S$  and  $E_T$  remain unknown. If the assumptions that  $k$  remains constant for a specific dye at all times, and that  $E_S$  and  $E_T$  both equal 1 – as it is generally considered to be during the exponential phase of a PCR cycle (Bustin and Nolan, 2004b) – are made,  $k$ ,  $T_0$ ,  $E_S$  and  $E_T$  are equal in both equations.

In actual fact, even if  $E_S$  and  $E_T$  are not equal, or do not equal 1, the error introduced by the assumption that  $E_S$  and  $E_T$  are 1 will be constant throughout TLA so long as the same single-copy locus is amplified by the same enzyme. Application of TLA in a standardised form by single bodies on national levels – for example, the UK's Forensic Science Service (FSS) – or several bodies using the same procedures will ensure that the error introduced by this assumption that  $E$  is 1 can effectively be discounted.

As  $E$  is assumed to be 1,  $1 + E_T$  and  $1 + E_S$  both equal 2. Therefore, by replacing  $1 + E_T$  and  $1 + E_S$  with 2, and dividing  $I_T$  by  $I_S$ , the equations can be combined:

$$I_T/I_S = (\tau k L_T 46 T_0 2^{C_T}) / (k L_S T_0 2^{C_S})$$

If the assumption is made that  $k$  remains constant as long as the same dye is used, this can be removed from the equation. Similarly,  $T_0$  will be constant as long as both single-copy locus and telomeric PCRs arise from the same DNA extract, and so can be removed. This simplifies the equation:

$$I_T/I_S = (\tau 46 L_T 2^{C_T}) / (L_S 2^{C_S})$$

$2^{C_T}/2^{C_S}$  can be expressed as  $2^{(C_T-C_S)}$  or  $1/2^{(C_S-C_T)}$ . Therefore, the terms expressing the efficiencies of the two PCR reactions can be expressed as a single term:

$$I_T/I_S = (\tau 46 L_T) / (L_S 2^{(C_S-C_T)})$$

Because the aim of TMF qPCR is to determine the value of TMF, which is proportionate to telomere length, the equation can be rearranged accordingly:

$$\tau 46 L_T = (I_T/I_S) (L_S 2^{(C_S-C_T)})$$

$$\tau 46 L_T = (I_T L_S 2^{(C_S-C_T)}) / I_S$$

$$\tau = (I_T L_S 2^{(C_S-C_T)}) / (46 I_S L_T)$$

This final equation – the TMF equation – can therefore be used to determine the length of an individual's telomeres, based on the relative intensity of fluorescence at the end point of a telomeric and a control PCR.

## 2.5: Aims & Objectives

---

Other researchers have attempted to use telomere length to determine age.

Tsuji et al. (2002) used the TRF assay to accurately predict individuals' ages from their telomere lengths by determining the mean length of the TRFs present. This mean length, determined by X ray exposure of labelled probes, can be relatively easily under- or overestimated, if TRFs present in relatively small quantities are not visualised. Accordingly, exposure time is an important factor in the TRF assay. There are other additional factors, discussed in section 2.2.2, that limit the efficacy of the TRF in the context of a criminal investigation.

While Tsuji et al. (2002) report the ability to predict age, that other researchers, such as Karlsson et al. (2008) – using the qPCR assay, which is not affected by exposure time – have failed to do so suggests that the TRF assay is flawed, possibly because of the presence of shorter or longer TRFs that have not been visualised. As already discussed, in section 2.3.6, the qPCR assay is not ideal for use in forensic DNA profiling. Additionally, Karlsson et al. (2008) used only linear regression to attempt to predict age from telomere length, and were unable to do so with a suitable degree of accuracy.

This thesis proposes both a new assay for telomere length – the TMF assay, described in section 2.4 – and a new statistical model for predicting age – Telomere Length Analysis.

There are two principal aims for this thesis.

Firstly, it is intended to establish the presence of a quantifiable relationship between telomere length and age and to use this relationship to predict the age of an unknown individual. Accordingly, the hypotheses can be expressed as:

- $H_0$ : that an unknown individual's age cannot be predicted to any significant degree.
- $H_1$ : that an unknown individual's age can be predicted to any significant degree.

Secondly, it is intended to establish the presence of a quantifiable relationship between telomere length in buccal epithelial cells and in hair follicle and white blood cells. Accordingly, the hypotheses can be expressed as:

- $H_0$ : that there is no statistically significant relationship between telomere length in buccal epithelial cells and in hair follicle and white blood cells.
- $H_1$ : that there is a statistically significant relationship between telomere length in buccal epithelial cells and in hair follicle and white blood cells.

These aims will be met by developing the TMF assay and protocols for analysing telomere length, and by comparing telomere lengths in buccal epithelial, hair follicle and white blood cells.

## ***Chapter Three: Materials & Method***

---

<b>3.1</b>	<b><i>Materials &amp; Method – Phase One</i></b>	<b>49</b>
3.1.1	Ethical approval	49
3.1.2	Participant recruitment	49
3.1.3	Biographical information and DNA collection	49
3.1.4	BMI as a lifestyle measure	50
3.1.5	DNA extraction	52
3.1.6	Optimisation of telomeric PCR	52
3.1.7	Telomeric PCR	54
3.1.8	Single-copy locus PCR	55
3.1.9	Electrophoresis	56
3.1.10	Densitometry and the TMF assay	56
<b>3.2</b>	<b><i>Materials &amp; Method – Phase Two</i></b>	<b>58</b>
3.2.1	Ethical approval	58
3.2.2	Participant recruitment	58
3.2.3	Biographical information and DNA collection	58
3.2.4	DNA extraction – buccal epithelial cell samples	60
3.2.5	DNA extraction – hair follicle cell samples	60
3.2.6	DNA extraction – blood samples	60
3.2.7	PCR, electrophoresis, densitometry and the TMF assay	61

## **3.1: Materials & Method – Phase One**

---

### *3.1.1: Ethical approval*

Ethical approval for this research was granted by the University of Chester Faculty Research Ethics Committee. The reference for approval for Phase One research is 370/09/SC/BIOL.

### *3.1.2: Participant recruitment*

Participants were primarily selected from staff and students of the University of Chester. This ensured that all participants had access to an email account, so that each could withdraw from the study at any time without having to endure telephone calls or face-to-face meetings, both of which might be considered intimidating. A limited number were friends or family of the researchers. Informed consent was obtained from all participants.

Participants were approached with the intention of obtaining equal numbers of male and female participants, across an approximately even distribution of ages. It was also intended to recruit around 100 participants. Recruiting this number proved difficult, and the target was lowered to allow experimental work to begin.

For Phase One, a total of 89 participants were recruited. 80 were recruited by the lead researcher, and each was assigned a tracking number from 1 to 81 – due to a minor administrative error, the consent form numbered 58 was left out and so, consequently, there is no Participant 58. A further 9 – numbered T1 to T9 – were recruited by the researcher's supervisors, without the researcher's knowledge, to allow for blind testing.

### *3.1.3: Biographical information and DNA collection*

Each participant was asked to provide a sample of DNA, collected using sterile buccal swabs, along with a limited amount of biographical information – date of birth and

gender, as well as height and body mass, which were used to calculate each participant's Body Mass Index (BMI), shown in Tables 3.1.1 to 3.1.4, overleaf.

#### *3.1.4: BMI as a lifestyle measure*

Several lifestyle indicators have been associated with variation in telomere length, including exercise (Cherkas et al., 2008), and vitamin D or omega-3 fatty acid intake (Richards et al., 2007; Farzaneh-Far, Lin, Epel, Harris, Blackburn and Whooley, 2010).

Initially, these indicators were to be assessed using a diet and lifestyle questionnaire as a part of this research, in order to attempt to use telomere length to produce a 'biography' for each participant. However, objections to the proposed questionnaire were raised by the Faculty Research Ethics Committee, chiefly regarding the time required of participants to complete it. After reviewing the FREC's response, it was decided to discard this questionnaire, and use BMI as the sole lifestyle indicator, for several reasons.

Firstly, the proposed questionnaire had fifty questions, which represented a considerable time commitment likely to reduce the number of people willing to participate in this project. Secondly, the questionnaire would introduce a great number of variables, which would therefore require a significantly larger number of participants in order to properly analyse results. Finally, and perhaps most importantly, much of the information that could be obtained by the questionnaire – while important in considering factors affecting telomere length – would not be immediately of use to investigating authorities as part of a molecular photofit.

For the purposes of this study, it was decided that BMI was the most practical measure. Firstly, determining BMI caused the minimum of disruption for participants, given that it required only measuring height and body mass. Secondly, while it must be noted that BMI is not flawless as a lifestyle measure, as it does not take into account an individual's build or body shape, it can be considered to be generally indicative of a participant's quality of diet and exercise regime (National Health Service, 2009). Finally, if lifestyle were to prove to be a factor affecting telomere length, a broad indication of BMI was considered more likely to be of use to investigating authorities than a more detailed description of an offender's diet or lifestyle. However, it must be stressed that BMI can only ever represent an approximation of an individual's lifestyle.

Age (years)	Gender		Total
	Male	Female	
18-29	9	17	26
30-39	5	11	16
40-49	12	6	18
50-59	8	5	13
Over 60	4	3	7
<b>Total</b>	<b>38</b>	<b>42</b>	<b>80</b>

Table 3.1.1: Age and gender distribution of Phase One participants.

Height (metres)	Gender		Total
	Male	Female	
Under 1.50	0	0	0
1.50-1.59	0	10	10
1.60-1.69	4	22	26
1.70-1.79	18	9	27
1.80-1.89	15	1	16
Over 1.90	1	0	1
<b>Total</b>	<b>38</b>	<b>42</b>	<b>80</b>

Table 3.1.2: Height and gender distribution of Phase One participants.

Body Mass (kilograms)	Gender		Total
	Male	Female	
50-59	1	11	12
60-69	4	17	21
70-79	8	7	15
80-89	10	2	12
90-99	5	3	8
100-109	5	0	5
110-119	3	0	3
Over 120	2	2	4
<b>Total</b>	<b>38</b>	<b>42</b>	<b>80</b>

Table 3.1.3: Body mass and gender distribution of Phase One participants.

BMI	NHS weight range	Gender		Total
		Male	Female	
Under 18.4	Underweight	0	0	0
18.4-19.9	Ideal	1	3	4
20-24.9	Ideal	11	20	31
25-29.9	Overweight	13	11	24
30-34.9	Obese	8	5	13
35-39.9	Obese	3	1	4
Over 40	Morbidly Obese	2	2	4
<b>Total</b>		<b>38</b>	<b>42</b>	<b>80</b>

Table 3.1.4: Body Mass Index and gender distribution of Phase One participants. BMI=(body mass in kilograms)/(height in metres)<sup>2</sup>. Weight ranges are for illustrative purposes only (National Health Service, 2009).

### *3.1.5: DNA extraction*

DNA was extracted from buccal swabs using the protocol described by Walsh, Metzger and Higuchi (1991). Briefly, swabbed cells were resuspended in 1.5ml 0.85% saline in 2ml screwcap microcentrifuge tubes, and centrifuged at 1,000rpm for 10 minutes. The supernatant was discarded and 0.5ml 10% Chelex-100 (BioRad, United Kingdom) solution was added. The pellet was resuspended using a vortex mixer and the tubes incubated in boiling water for 10 minutes, before being cooled on ice for 2 minutes.

The lysed cells were centrifuged at 13,000 rpm for 30 seconds before the supernatant was transferred to sterile 1.5ml screwcap microcentrifuge tubes. Extracted DNA was then frozen until needed.

### *3.1.6: Optimisation of telomeric PCR*

Telomeric PCR using Cawthon's telomeric primers produces a range of amplicons of a variety of sizes. Given that the TMF assay is an end-point PCR amplification followed by electrophoresis and densitometry, Cawthon's telomeric primers produce a smear that would make densitometry more difficult. Because of this, the PCR protocol described by Cawthon (2002) needed to be tested on end-point PCRs and, if necessary, optimised. In particular, it was desirable to increase the rate of allele drop-out, discussed in sections 2.3.2 and 2.3.3.

Allele drop-out is a particular problem in the multiplex PCRs now commonly used in DNA profiling, because it is more difficult to develop a unified protocol that enhances PCR of longer alleles without impeding the amplification of shorter alleles. Higher relative concentrations of magnesium chloride increase the stability of primer-template-polymerase complexes, reducing the tendency of the polymerase to dissociate, while longer extension times mean that longer templates are more likely to be replicated fully (Henegariu, Heerema, Dlouhy, Vance and Vogt, 1997).

However, microsatellites are typically used as markers in current DNA profiling assays, rather than minisatellites, because the shorter repeat units of microsatellites means that the variation in length of microsatellite alleles is smaller than that of minisatellites, reducing the rate of allele drop-out (Butler, 2001).

Unlike conventional DNA profiling, the TMF assay benefits from an enhanced rate of allele drop-out, as it results in the majority of amplicons produced being the desired 76bp. As the real-time qPCR and TMF assays are, in effect, multiplex PCRs – because of the many possible amplicons that can be produced – allele drop-out can be enhanced by reversing the standard troubleshooting methods described by Henegariu et al. (1997).

A further complication was that the available densitometer has an upper limit to the amount of DNA it can detect. Individual pixels in the image used for densitometry can become ‘saturated’, which means that they are more fluorescent than it is within the machine’s capacity to measure. A ‘saturated’ pixel can therefore contain  $x$  DNA, where  $x$  is the quantity of DNA required to fluoresce sufficiently to ‘saturate’ the pixel, or  $2x$ ,  $3x$  or more. Densitometry of ‘saturated’ pixels can therefore produce misleading results.

Two PCR mixes were tested: one, described overleaf, was based on the Epicentre Failsafe Buffer (Eurofins, United Kingdom), which is the standard PCR buffer used by the University of Chester, while the other, based on the GoTaq Flexi Buffer (Promega, United Kingdom), contained less magnesium chloride than that used by Cawthon (2002). The addition of magnesium chloride reduces allele drop-out (Henegariu et al., 1997), so the reduction of the concentration of magnesium chloride in this custom buffer was expected to increase it. PCR was performed using both mixes with extension times of 1 minute or 30 seconds – while Cawthon (2002) ran extension times of 2 minutes, shorter extension times increase allele drop-out (Henegariu et al., 1997) – for 15, 16, 17 or 18 cycles. Cawthon (2002) used 18 cycles, but reducing the number of cycles was intended to limit the number of ‘saturated’ pixels (see Figure 3.1.1, overleaf).

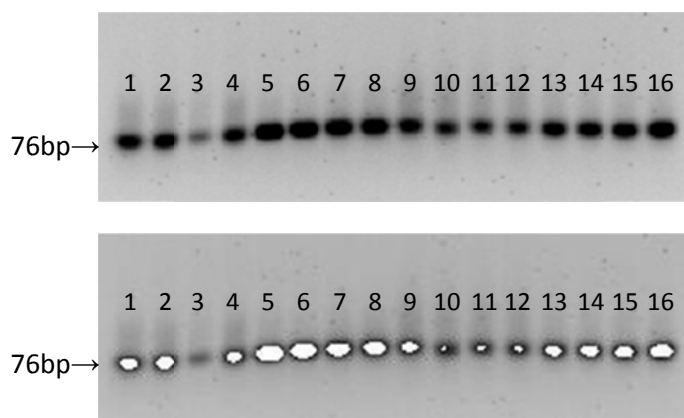


Figure 3.1.1: Results of PCR optimisation. The upper image is a straightforward capture of the gel. The lower image shows in white 'saturated' pixels. The densitometer records these in red, but the image was amended for clarity using Microsoft Paint. On the basis of these results, lane 3 (highlighted below) was the optimal telomeric PCR protocol. 16 different telomeric PCR protocols were assessed, with varying buffers, extension times and number of cycles:

Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Buffer (Standard or Custom)	S	S	S	S	C	C	C	C	S	S	S	S	C	C	C	C
Extension time (s)	60	60	60	60	60	60	60	60	30	30	30	30	30	30	30	30
Number of cycles	18	17	15	16	18	17	16	15	18	17	16	15	18	17	16	15

On this basis, it was decided that lane 3 – with the Epicentre Failsafe Buffer, and 15 cycles of PCR with an extension time of 1 minute – was the optimal PCR protocol for the TMF assay, with a relatively small and faint 'tail' to the 76bp amplicer band, and no 'saturated' pixels. The full details of this PCR protocol are described overleaf.

### 3.1.7: Telomeric PCR

Subject to the amendments already discussed in section 3.1.6, telomeric PCR was largely performed as described by Cawthon (2002). 5µl of DNA extract was added to PCR tubes containing 12.5µl 2× Epicentre Failsafe Premix Buffer E, 6µl distilled water, 0.5µl 100pmol/µl primer 1 (Eurofins, United Kingdom), 0.5µl 100pmol/µl primer 2 (Eurofins, United Kingdom) and 0.5µl 5 units/µl GoTaq Hotstart DNA polymerase (Promega, United Kingdom), making a total of 25µl in each PCR tube. The primer sequences used are shown in Table 3.1.5, overleaf.

Telomeric Primer 1	5'-GGT TTT TGA GGG TGA GGG TGA GGG TGA GGG TGA GGG T-3'
Telomeric Primer 2	5'- TCC CGA CTA TCC CTA TCC CTA TCC CTA TCC CTA TCC CTA-3'

Table 3.1.5: Telomeric PCR primers.

Telomeric PCR used the following profile: initial denaturation at 95°C for 10 minutes, followed by 15 cycles of 95°C for 15 seconds and 54°C for 1 minute, followed by final extension at 72°C for 5 minutes in a Techne (trading as Bibby Scientific, United Kingdom) TC-312 thermal cycler.

### 3.1.8: Single-copy locus PCR

36B4 PCR was largely performed as described by Cawthon (2002), with some amendments. 5µl of DNA extract was added to PCR tubes containing 12.5µl 2× Epicentre Failsafe Premix Buffer E, 6µl distilled water, 0.5µl 100pmol/µl primer 1 (Eurofins, United Kingdom), 0.5µl 100pmol/µl primer 2 (Eurofins, United Kingdom) and 0.5µl 5 units/µl GoTaq Hotstart DNA polymerase (Promega, United Kingdom), making a total of 25µl in each PCR tube. These primers produce a 36B4 amplicon 74bp in length (Cawthon, 2002). The primer sequences used are shown in Table 3.1.6.

36B4 Locus Primer 1	5'-CAG CAA GTG GGA AGG TGT AAT CC-3'
36B4 Locus Primer 2	5'-CCC ATT CTA TCA TCA ACG GGT ACA A-3'

Table 3.1.6: 36B4 locus PCR primers.

36B4 PCR used the following profile: initial denaturation at 95°C for 10 minutes, followed by 30 cycles of 95°C for 15 seconds and 58°C for 1 minute, followed by final

extension at 72°C for 5 minutes in a Techne (trading as Bibby Scientific, United Kingdom) TC-312 thermal cycler.

### *3.1.9: Electrophoresis*

2% agarose gels were prepared using 98ml distilled water, 2ml 50× TAE buffer (pH8), 2g genetic analysis grade agarose (Fisher Scientific, United Kingdom) and 10µl 10000× GelRed (Biotium, distributed by Cambridge BioScience, United Kingdom) and were immersed in 1×TAE buffer (pH8).

Samples were prepared by the addition of 5µl 6× loading dye (Promega, United Kingdom) to each PCR tube, giving a total volume of 30µl. Typically, 12µl of each amplified DNA solution was added to the relevant wells of the gel. However, a DNA extract made from a large number of cells has the potential to saturate the densitometer after PCR, as discussed above. Accordingly, some samples, where the pellet after Chelex extraction was particularly large – indicating that a high number of cells had been collected – had only 9µl or 6µl loaded onto a gel from both the telomeric and single-copy locus PCRs. While it was essential to load identical quantities from the telomeric and single-copy locus PCR amplifications for any single participant, changing this quantity from participant to participant does not affect the TMF equation.

Each gel was run at 125V for approximately 1 hour.

### *3.1.10: Densitometry and the TMF assay*

Densitometry was performed using the densitometric function of a BioRad (United Kingdom) GelDoc 2000 transilluminator and Quantity One (BioRad, United Kingdom). Densitometry is not a perfect way to quantify DNA, as it typically requires the operator to define an area for intensity to be measured within. Therefore, to minimise variation in the selection of bands for densitometry, the intensity of each amplicon for each participant was measured at least three times, until at least two values for intensity were within 100 units per mm<sup>2</sup> of each other. The greater the divergence in values for intensity, the greater the number of concordant readings

that were required. For example, if the first two readings were 1140 and 2390 units per mm<sup>2</sup>, and the third reading was 2400 units per mm<sup>2</sup>, further measurements were taken to confirm that the intensity of the band was indeed around 2395 units per mm<sup>2</sup>.

The means of these concordant values were taken and recorded as I<sub>T</sub> and I<sub>S</sub>. Because L<sub>T</sub> and L<sub>S</sub> are known to be 76bp and 74bp respectively, TMF was calculated using the equation:

$$\tau = (I_T L_S 2^{(C_S - C_T)}) / (46 I_S L_T)$$

Inserting the known values of L<sub>T</sub>, L<sub>S</sub>, C<sub>T</sub> and C<sub>S</sub>, the equation can be written as:

$$\tau = (I_T \times 74 \times 2^{(30 - 15)}) / (46 \times I_S \times 76)$$

$$\tau = (I_T \times 74 \times 2^{15}) / (I_S \times 3496).$$

## **3.2: Materials & Method – Phase Two**

---

### *3.2.1: Ethical approval*

Ethical approval for this research was granted by the University of Chester Faculty Research Ethics Committee. The reference for approval is 380/10/SC/BIOL.

### *3.2.2: Participant recruitment*

Phase One participants were asked if they would be willing to donate small samples of blood and hair should the research continue to this point. Prospective volunteers were emailed to request their participation at least a week before their arranged appointments, to allow them to read Participant Information Sheets and decide without any haste. A small number of participants at this phase were new to the research. 14 participants agreed to take part, and were numbered from 2.1 to 2.14. As with Phase One, this was a smaller number of participants than desired – the fact that giving blood was a more intrusive process than providing a buccal swab was cited on several occasions. However, given that Phase Two largely supplemented Phase One research, a large sample population was less necessary.

### *3.2.3: Biographical information and DNA collection*

Buccal swabs were collected from each participant. Small blood samples of 1-2 drops were taken using a sterile automatic lancet to prick the skin of each participant's off-hand middle finger. Blood was collected in microcentrifuge tubes. A few hairs were plucked from each participant's head using sterile tweezers and cut into 1 cm lengths. All fragments ending in a follicle were selected and transferred to microcentrifuge tubes.

In addition, each participant's date of birth, height and body mass was recorded and BMI calculated as described in section 3.1.3 (see Tables 3.2.1 to 3.2.4, overleaf).

Age (years)	Gender		Total
	Male	Female	
18-29	2	5	7
30-39	1	1	2
40-49	2	0	2
50-59	0	3	3
Over 60	0	0	0
<b>Total</b>	5	9	14

Table 3.2.1: Age and gender distribution Phase Two participants.

Height (metres)	Gender		Total
	Male	Female	
Under 1.50	0	0	0
1.50-1.59	0	2	2
1.60-1.69	0	5	5
1.70-1.79	2	2	4
1.80-1.89	3	0	3
Over 1.90	0	0	0
<b>Total</b>	5	9	14

Table 3.2.2: Height and gender distribution of Phase Two participants.

Body Mass (kilograms)	Gender		Total
	Male	Female	
50-59	0	3	3
60-69	0	2	2
70-79	1	1	2
80-89	1	1	2
90-99	2	1	3
100-109	1	0	1
110-119	0	0	0
Over 120	0	0	0
<b>Total</b>	5	9	14

Table 3.2.3: Body mass and gender distribution of Phase Two participants.

BMI	NHS weight range	Gender		Total
		Male	Female	
Under 18.4	Underweight	0	0	0
18.4-19.9	Ideal	0	1	1
20-24.9	Ideal	1	2	3
25-29.9	Overweight	2	3	5
30-34.9	Obese	1	1	2
35-39.9	Obese	1	1	2
Over 40	Morbidly Obese	0	1	1
<b>Total</b>		5	9	14

Table 3.2.4: Body Mass Index and gender distribution of Phase Two participants. Weight ranges are for illustrative purposes only (National Health Service, 2009).

#### *3.2.4: DNA extraction – buccal epithelial cell samples*

DNA was extracted from the buccal swabs as described in section 3.1.5.

#### *3.2.5: DNA extraction – hair follicle cell samples*

DNA was extracted using QuickExtract solution (Epicentre, distributed by Cambio, United Kingdom) and the manufacturer's protocols. 0.3ml QuickExtract solution was added to each microcentrifuge tube, and mixed, before being heated in water at 65°C for 1 hour. After a second mixing, the solution was heated in water at 98°C for 2 minutes, before being stored on ice.

#### *3.2.6: DNA extraction – blood samples*

DNA was extracted using the Qiagen (United Kingdom) DNeasy Blood and Tissue kit spin-column protocol for whole enucleated blood and the manufacturer's protocols. 20µl of the supplied proteinase K and the collected blood were added to microcentrifuge tubes and made up to 220µl with PBS buffer. 200µl of buffer AC was added and incubated at 56°C for 10 minutes. 200µl of pure ethanol was added and the mixture transferred to the supplied mini spin columns in collection tubes and centrifuged at 8,000 rpm for 1 minute. The spin columns were transferred to fresh collection tubes, before 500µl of buffer AW1 was added and the tubes spun again, at 8,000 rpm for 1 minute.

The spin columns were transferred to fresh collection tubes and 500µl of buffer AW2 was added, before being spun at 14,000 rpm for 3 minutes. The spin columns were then transferred to fresh microcentrifuge tubes, before 100µl of buffer AE was added and spun through at 8,000 rpm for 1 minute. This last step was repeated to maximise the quantity of DNA recovered.

### *3.2.7: PCR, electrophoresis, densitometry and the TMF assay*

The samples were then processed as described in sections 3.1.7 to 3.1.10.

## Chapter Four: Results & Analysis

---

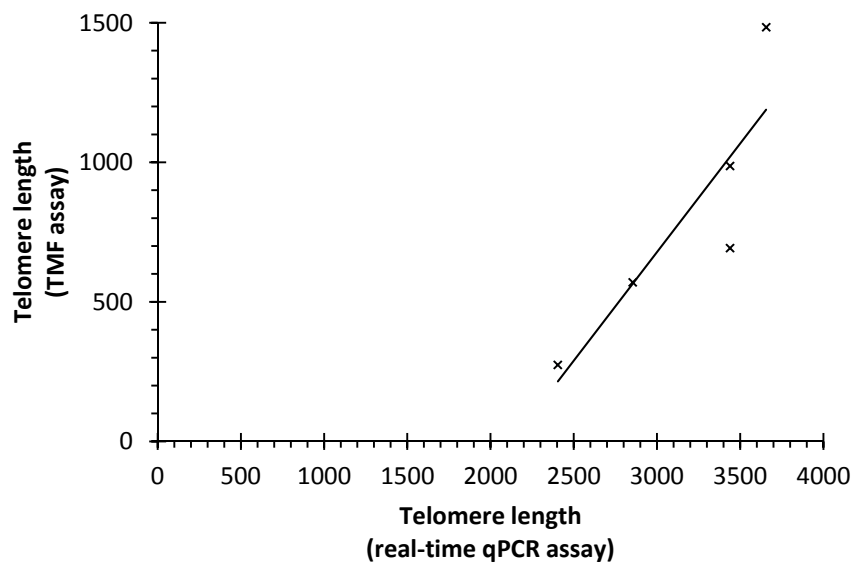
<b>4.1</b>	<b>Verification of the TMF Assay</b>	<b>63</b>
4.1.1	Summary	63
<b>4.2</b>	<b>Linear Regression Age Prediction</b>	<b>64</b>
4.2.1	Introduction	64
4.2.2	Linear Regression Age Prediction	64
4.2.3	Gender-differentiated LRAP	65
4.2.4	Prediction profiles of LRAP and GD-LRAP	67
4.2.5	Blind testing of LRAP and GD-LRAP	68
4.2.6	Evaluation of LRAP and GD-LRAP	69
<b>4.3</b>	<b>Telomere Length Analysis</b>	<b>70</b>
4.3.1	The limitations of LRAP	70
4.3.2	Telomere Length Analysis	71
4.3.3	Restriction of $TMF_A/TMF_C$ values	76
4.3.4	Gender-differentiated TLA	80
4.3.5	Blind testing of TLA and GD-TLA	83
4.3.6	Evaluation of TLA	84
4.3.7	The TLA database increases accuracy	85
<b>4.4</b>	<b>Telomere Length Analysis &amp; Alternative Tissues</b>	<b>87</b>
4.4.1	Introduction	87
4.4.2	Comparison of ${}_B TMF$ with ${}_F TMF$	88
4.4.3	Comparison of ${}_F TMF$ with age	89
4.4.4	${}_F TMF$ -based TLA	90
4.4.5	Comparison of ${}_B TMF$ with ${}_L TMF$	91
4.4.6	${}_L TMF$ -based ${}_D TMF$ TLA	93
<b>4.5</b>	<b>TMF &amp; Other Factors</b>	<b>96</b>
4.5.1	Introduction	96
4.5.2	TMF and height, body mass and BMI	96

## 4.1: Verification of the TMF Assay

### 4.1.1: Summary

Determination of telomere length by the real-time qPCR assay required 40-50µl of DNA solution at concentrations of at least 10ng/ml. Of 35 samples assessed using spectrophotometry, 24 – 69% – reached or exceeded this concentration. 31% did not, and so were not suitable for the real-time qPCR assay.

Of the appropriate samples, six were selected, covering a range of TMF values, and sent to Newcastle University to undergo the real-time qPCR assay. Of these samples, one failed to return any results, for reasons that are not clear. The remaining five were measured by the real-time qPCR assay and the results compared with those from the TMF assay (see Graph 4.1.1).



*Graph 4.1.1:* The strong correlation between telomere lengths as measured by the real-time qPCR assay and by the TMF assay ( $R^2=0.764$ ;  $n=5$ ). This correlation compares well with the correlation between telomere lengths as measured by the real-time qPCR assay and the TRF assay –  $R^2=0.677$  (Cawthon, 2002) – although with an appreciably smaller sample size.

These results allow the tentative conclusion that the TMF assay is a valid measure of telomere length.

## **4.2: Linear Regression Age Prediction**

---

### *4.2.1: Introduction*

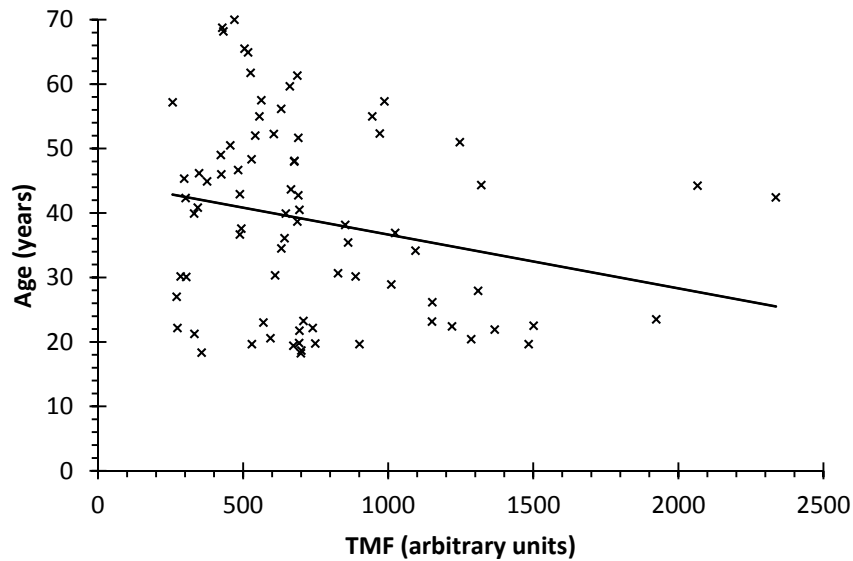
TMF was calculated as described previously. Calculations were performed using Microsoft Excel 2007 and SPSS v17.0. Statistical analysis was performed as described by Holmes, Moody and Dine (2006), and calculated using the VassarStats Statistical Computation website (Lowry, 2008).

The first and most straightforward method to predict age from telomere length is to determine a linear relationship between age and telomere length. It is this technique that has been previously attempted, and been found wanting (Karlsson et al., 2008), due to the poor accuracy of age predictions when compared to the known age. However, performing this technique – described here as Linear Regression Age Prediction (LRAP) – allows the comparison of the accuracy of other forms of analysis.

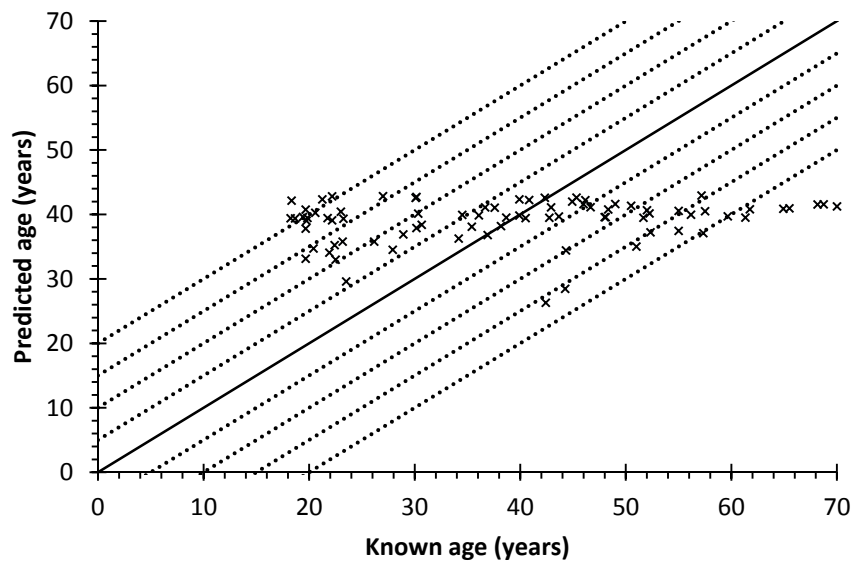
### *4.2.2: Linear Regression Age Prediction*

In order to perform LRAP, the calculated TMFs were plotted against the known ages of participants, and a linear regression determined (see Graph 4.2.1, overleaf). Using the linear regression equation  $\text{Age} = -0.008\tau + 45.01$ , ages could be predicted for all 80 participants and compared to their known ages (see Graph 4.2.2, overleaf).

Graph 4.2.1: The relationship between TMF and age. This demonstrates a weak correlation between telomere length and age ( $\text{Age} = -0.008\tau + 45.01$ ;  $R^2 = 0.053$ ;  $n = 80$ ).



Graph 4.2.2: Comparison of ages predicted using LRAP with known ages. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5$ , 10, 15 and 20 years ( $n = 80$ ). The standard deviation in predicted age using LRAP is 14.359.



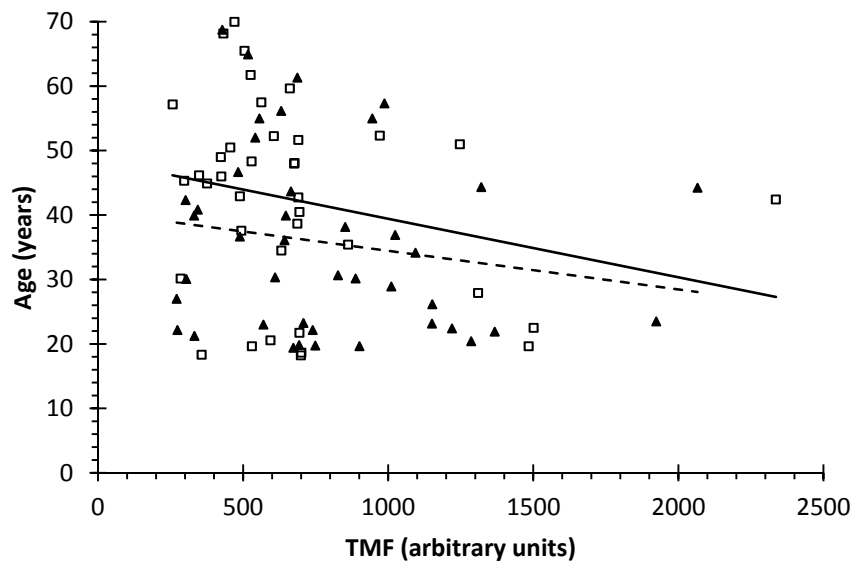
#### 4.2.3: Gender-differentiated LRAP

Given that a DNA profile can be used to determine gender using amelogenin, there remained the possibility that gender-differentiated LRAP (GD-LRAP) might, if more accurate than conventional LRAP, be of greater use to investigating authorities. This

is, of course, dependent on the DNA sample not being degraded so much as to prevent amplification of the amelogenin locus.

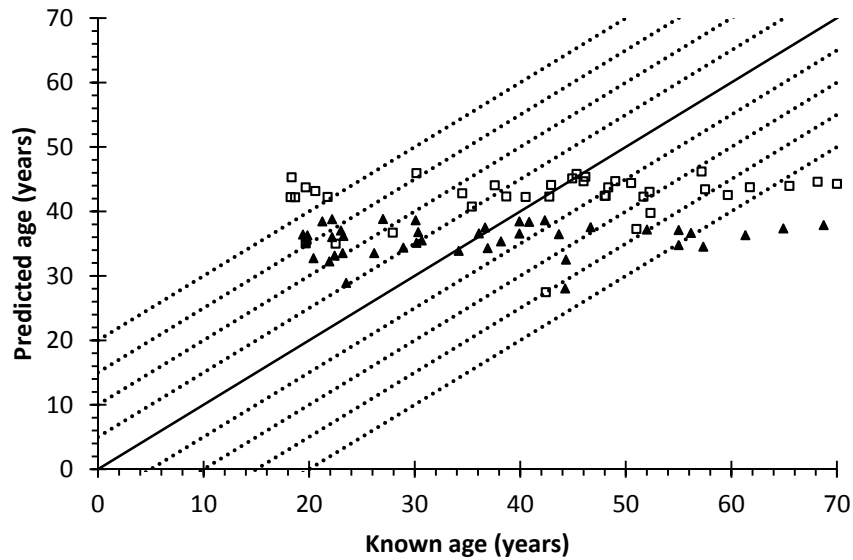
Accordingly, participants were divided according to gender and the process of LRAP repeated (see Graph 4.2.3).

*Graph 4.2.3:* The relationship between TMF and age in males and females. This shows a stronger correlation between TMF and age in males (open squares and solid line;  $\text{Age} = -0.009\tau + 48.51$ ;  $R^2 = 0.061$ ;  $n = 38$ ) than in females (filled triangles and dashed line;  $\text{Age} = -0.006\tau + 40.45$ ;  $R^2 = 0.030$ ;  $n = 42$ ). It also demonstrates a tendency for males to have longer telomeres than females of the same age.



Using these newly-derived linear regression equations, the age of each individual could be predicted and compared to the known age, as in Graph 4.2.2 (see Graph 4.2.4, overleaf).

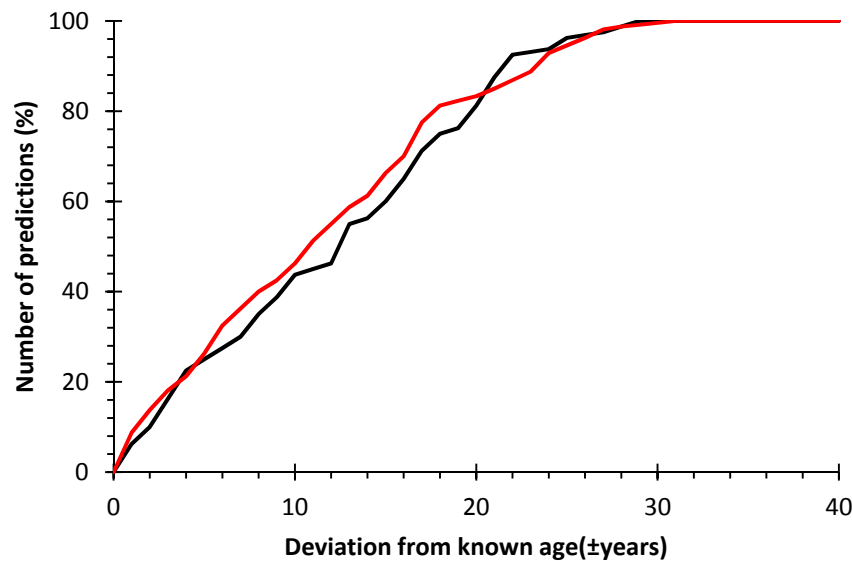
Graph 4.2.4: Comparison of ages predicted using GD-LRAP with known ages in males (open squares; n=38) and females (filled triangles; n=42). The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5$ , 10, 15 and 20 years. The standard deviation in predicted age using GD-LRAP is 14.052.



#### 4.2.4: Prediction profiles of LRAP and GD-LRAP

The accuracy of predictions made using LRAP and GD-LRAP were calculated and used to determine the probability of any single age prediction being within, for example,  $\pm 5$  years of the known age. For LRAP, 25.00% of all predictions were accurate to within  $\pm 5$  years, while GD-LRAP predicted ages to within  $\pm 5$  years 26.32% and 26.19% of the time for males and females respectively. These prediction profiles are summarised in Graph 4.2.5, overleaf.

Graph 4.2.5: Prediction profiles of LRAP and GD-LRAP. The deviation from the known age is on the horizontal axis, while the percentage of predictions that were accurate to this degree is on the vertical axis. LRAP is denoted by the black line, while GD-LRAP is denoted by the red line (n=80).

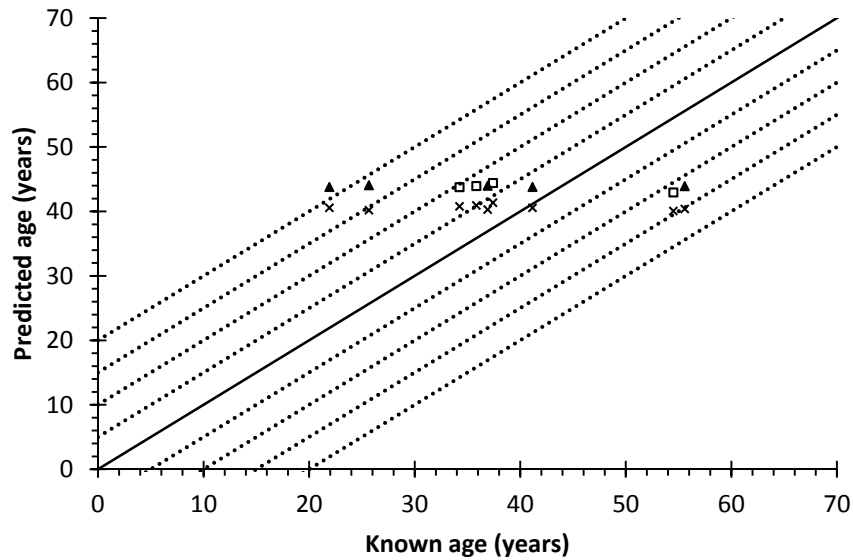


#### 4.2.5: Blind testing of LRAP and GD-LRAP

As discussed in section 3.1.2, a further nine participants were recruited without the lead researchers' knowledge. These participants were reserved to act as subjects for blind testing of any assays developed. Following DNA extraction and calculation of TMF, the age of each test participant was predicted using LRAP and GD-LRAP (see Graph 4.2.6, overleaf).

Taking as the null hypothesis ( $H_0$ ) that there is no significant difference between the expected accuracy of the predictions of the blind test participants' ages, based on Graph 4.2.5, and the observed accuracy, Z scores were found for LRAP and GD-LRAP. These supported the null hypothesis for LRAP ( $Z=0.456$ ) and GD-LRAP ( $Z=1.212$ ) –  $H_0$  is confirmed if  $-1.96 \leq Z \leq 1.96$ . Therefore, there is no significant difference between the predicted and observed accuracies of age predictions using LRAP and GD-LRAP. However, these predictions were not particularly accurate.

Graph 4.2.6: Comparison of ages predicted for test participants using LRAP (crosses; n=9) and GD-LRAP of male (open squares; n=4) and female (filled triangles, n=5) participants. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5, 10, 15$  and  $20$  years.



#### 4.2.6: Evaluation of LRAP

Using the predicted and known ages calculated by LRAP, Spearman's Rank Order Correlation Test was performed with the VassarStats Statistical Computation website (Lowry, 2008), as well as by hand. For LRAP, the covariance of predicted age with known age was 0.2848 ( $r_s=0.2848$ ;  $t=2.62$ ;  $n=80$ ;  $df=78$ ;  $P$  (two-tailed)=0.010564). This demonstrates a statistically significant relationship between ages predicted using LRAP and the known age of individuals. However, the accuracy of these predictions is such that LRAP is not a viable tool to predict age.

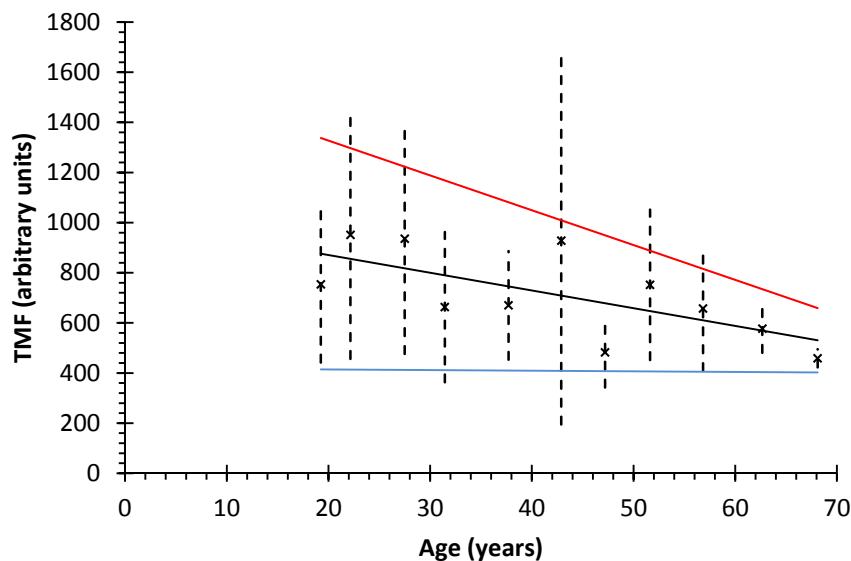
## 4.3: Telomere Length Analysis

### 4.3.1: The limitations of LRAP

LRAP fails to accurately predict age because individuals of the same age can have very different telomere lengths, and individuals with similar telomere lengths can be of different ages. For example, Participants 27 and 55 were both 55 and 0 months years old at the date of sampling, but had TMFs of 945 and 556 respectively. On the other hand, Participants 72 and 73 both had TMFs of 687, but were aged 38 years, 8 months and 61 years, 4 months respectively at the date of sampling.

To demonstrate this, participants were divided into age groups spanning 5 years, such as 20-25 and 35-30 years, and the mean age and TMF of each group determined. The standard deviation in TMF for each group was also calculated. While there was a greater correlation between a group's mean age and its mean TMF than that between each individual's age and TMF, the range of  $\pm 1$  standard deviation for each group overlapped with those of neighbouring groups (see Graph 4.3.1).

*Graph 4.3.1:* Mean age and mean TMF within 5-year age groups.  $Age_{mean}$  and  $TMF_{mean}$  are marked with crosses, and the range of  $\pm 1$  standard deviation indicated by the dashed lines. There are greater correlations between  $Age_{mean}$  and  $TMF_{mean}$  (black line;  $\tau_{mean} = -7.0729Age_{mean} + 1012.2$ ;  $R^2 = 0.450$ ), and  $Age_{mean}$  and  $TMF_{mean} + 1$  standard deviation (red line;  $\tau_{mean} + s = -13.898Age_{mean} + 1605.5$ ;  $R^2 = 0.393$ ) than between individuals' age and TMF (see Graph 4.2.1;  $R^2 = 0.053$ ).  $TMF_{mean} - 1$  standard deviation (blue line;  $\tau_{mean} - s = -0.2476Age_{mean} + 418.93$ ;  $R^2 = 0.002$ ) showed a weaker correlation.



This results in the observed inaccuracy of LRAP, demonstrated in section 4.2 and by Karlsson et al. (2008). Clearly, linear regression is not suitable for predicting age from telomere length. Therefore, in order to make assays based on telomere length a part of molecular photofitting procedures, a new protocol must be developed.

The greater correlations between  $Age_{mean}$  and  $TMF_{mean}$  or  $TMF_{mean} \pm 1$  standard deviation still cannot be used to determine age for two reasons. Firstly, a range of  $\pm 1$  standard deviation incorporates only 61% of all members of the test population (Swinscow and Campbell, 1997), so 39% of individuals within any particular age group lie outside the ranges shown in Graph 4.3.1. Secondly, the ranges shown overlap to such a degree that many individuals' TMF may fall into several age groups' ranges. For example, Participant 1 has a TMF of 632, which falls into the ranges of all groups except the 45-50 group. Linear regressions such as those shown in Graph 4.3.1 are therefore unlikely to be accurate predictors of age.

It can be argued that the closer an individual's TMF is to a particular group's  $TMF_{mean}$ , the more likely it is that the individual belongs to that group. Given that  $TMF_{mean}$  is a product of the individuals within a group, this suggests that an individual's TMF, relative to that of other individuals, may be a better predictor of age than LRAP.

#### 4.3.2: Telomere Length Analysis

The principle of Telomere Length Analysis is that the ratio between two individual's TMFs will be inversely proportionate to the ratio between their ages. The closer these individual's TMFs are to each other, the more similar their ages will be. This can be expressed as:

$$TMF_A/TMF_C \propto (Age_A/Age_C)^{-1}$$

where  $TMF_A$  and  $Age_A$  are those of the individual subject to analysis, and  $TMF_C$  and  $Age_C$  are those of a known individual acting as a comparison. Because  $Age_C$  is known, a value of  $Age_A$  can be calculated. Using multiple comparison individuals allows the determination of multiple values of  $Age_A$ .

With a total sample population of 80 participants, each individual's TMF and age were divided by every other individual's TMF and age, resulting in a total of

6,400 – 80<sup>2</sup> – data pairs of  $TMF_A/TMF_C:Age_A/Age_C$ . This is illustrated in Table 4.3.1, using Participant 1 as an example.

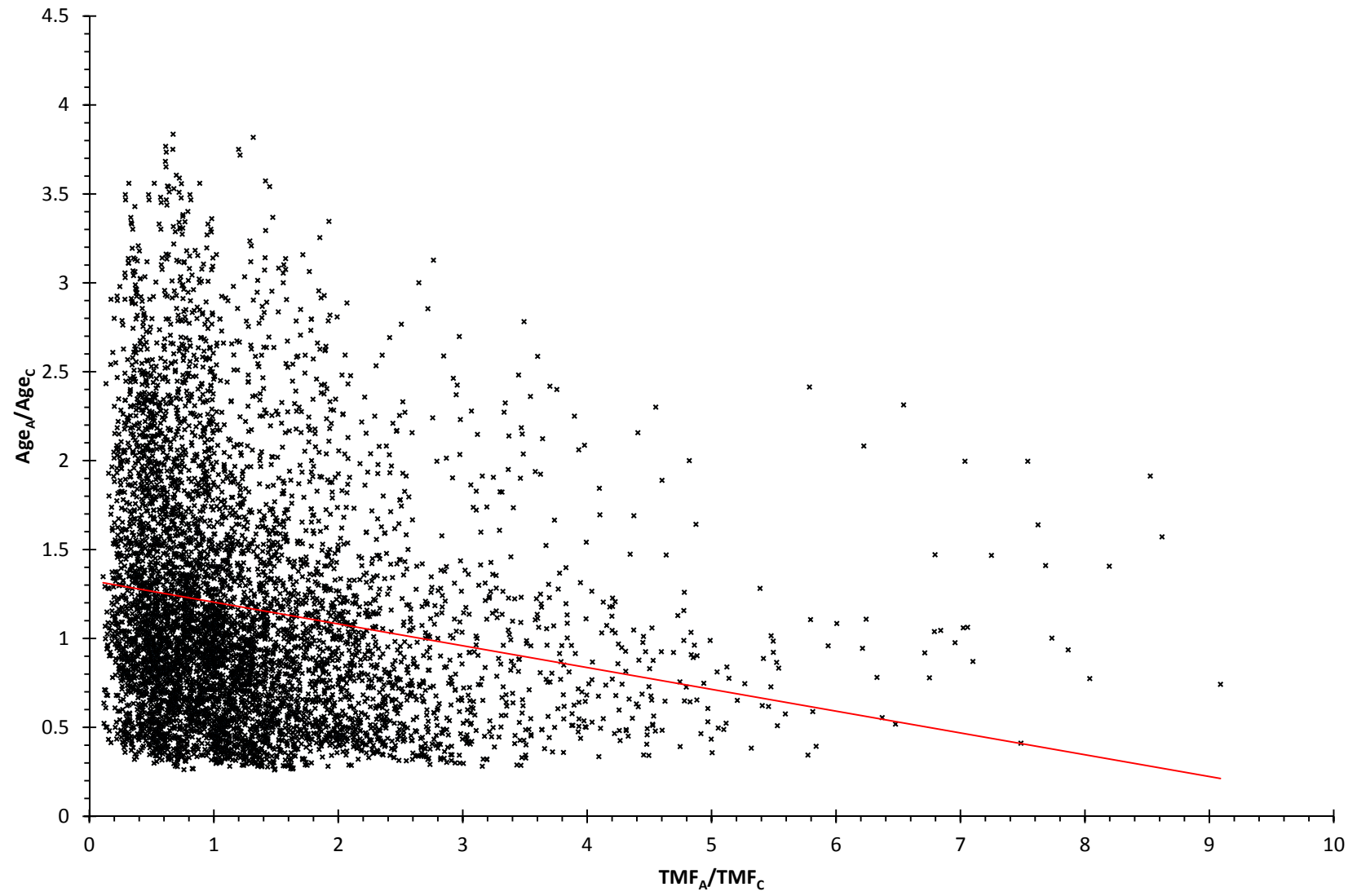
Participant <sub>A</sub>	TMF <sub>A</sub>	Age <sub>A</sub> (years)	Participant <sub>C</sub>	TMF <sub>C</sub>	Age <sub>C</sub> (years)	TMF <sub>A</sub> /TMF <sub>C</sub>	Age <sub>A</sub> /Age <sub>C</sub>
1	632	34.500	1	632	34.500	1.000	1.000
1	632	34.500	2	1321	44.333	0.478	0.778
1	632	34.500	3	1011	28.917	0.625	1.193
1	632	34.500	4	827	30.667	0.764	1.125
1	632	34.500	5	2336	42.417	0.271	0.813
1	632	34.500	6	708	23.250	0.892	1.484
...	...	...	...	...	...	...	...
1	632	34.500	80	563	57.500	1.123	0.6
1	632	34.500	81	1247	51.000	0.507	0.676

Table 4.3.1: Calculation of  $TMF_A/TMF_C$  and  $Age_A/Age_C$ , where Participant 1 is the subject of analysis and Participants 1-81, omitting number 58, are the comparisons.

The resulting 6,400 data pairs were plotted and a linear regression determined (see Graph 4.3.2, overleaf).

The equation of the linear regression was  $Age_A/Age_C = -0.122(\tau_A/\tau_C) + 1.327$ . Therefore, if  $TMF_A/TMF_C = 1$ , as it does when Participant<sub>A</sub> and Participant<sub>C</sub> are the same individual,  $Age_A/Age_C = -0.122 \times 1 + 1.327 = 1.205$ . It was therefore necessary to amend the equation to  $Age_A/Age_C = (-0.122(\tau_A/\tau_C) + 1.327) / 1.205$  for subsequent calculations.

Graph 4.3.2 (overleaf): The relationship between  $TMF_A/TMF_C$  and  $Age_A/Age_C$  (n=6,400). This shows a weak correlation between  $TMF_A/TMF_C$  and  $Age_A/Age_C$  ( $Age_A/Age_C = -0.122(\tau_A/\tau_C) + 1.327$ ;  $R^2 = 0.031$ , n=6,400). The standard deviation of  $TMF_A/TMF_C$  was 0.985 and the standard deviation of  $Age_A/Age_C$  was 0.676.



Taking Participant 1 as an example again, the 80 values of  $TMF_A/TMF_C$  were used to generate 80 values of  $Age_A/Age_C$ . Because all values of  $Age_C$  were known, this allowed the calculation of 80 values of  $Age_A$ . The mean of these values can be taken as the predicted age (see Table 4.3.2).

TMF <sub>A</sub>	Participant <sub>C</sub>	TMF <sub>C</sub>	Age <sub>C</sub> (years)	TMF <sub>A</sub> / TMF <sub>C</sub>	Unadjusted Equation		Adjusted Equation	
					Age <sub>A</sub> / Age <sub>C</sub>	Age <sub>A</sub> (years)	Age <sub>A</sub> / Age <sub>C</sub>	Age <sub>A</sub> (years)
632	1	632	34.500	1.000	1.205	41.573	1.000	34.5
632	2	1321	44.333	0.478	1.269	56.243	0.778	46.674
632	3	1011	28.917	0.625	1.251	36.167	1.193	30.014
632	4	827	30.667	0.764	1.234	37.836	1.125	31.399
632	5	2336	42.417	0.271	1.294	54.887	0.813	45.549
632	6	708	23.250	0.892	1.218	28.321	1.484	23.503
...	...	...	...	...	...	...	...	...
632	80	563	57.500	1.123	1.190	68.428	0.988	56.787
632	81	1247	51.000	0.507	1.265	64.524	1.045	53.547
					<b>Mean</b>	46.209	<b>Mean</b>	38.347

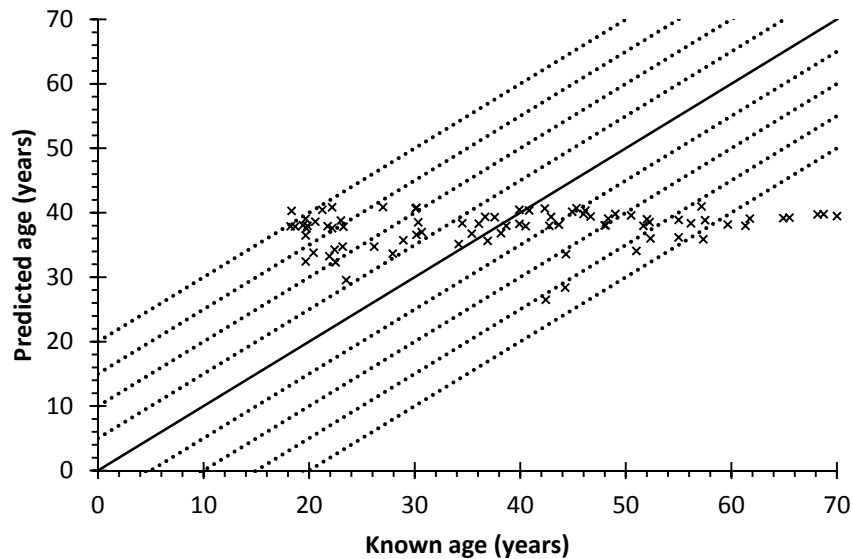
Table 4.3.2: Illustration of the prediction of  $Age_A$  for Participant 1. Please note that, at the time of sampling, Participant 1 was 34 years, 6 months old.  $Age_A/Age_C$  was calculated using the linear regression equation calculated in Graph 4.3.2 ( $Age_A/Age_C = -0.122(\tau_A/\tau_C) + 1.327$ ) and with the adjustment described previously ( $Age_A/Age_C = (-0.122(\tau_A/\tau_C) + 1.327)/1.205$ ). The amended equation produced values of  $Age_A/Age_C$  17% smaller than the original equation, and therefore returned values of  $Age_A$  and a lower mean  $Age_A$ .

TLA, using the amended equation  $Age_A/Age_C = (-0.122(\tau_A/\tau_C) + 1.327)/1.205$ , was performed in this manner for all 80 participants, and the accuracy of its predictions determined (see Graph 4.3.3, overleaf).

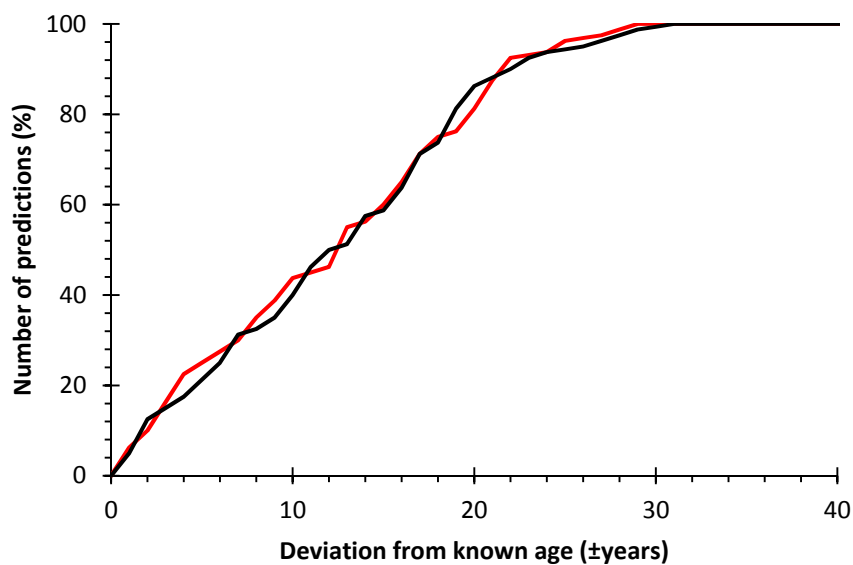
Predictions determined using TLA at this point were not noticeably more accurate than those predicted using LRAP. This was not unexpected, because neither can limit the effect of the variation in TMF of people of similar ages. The introduction of the amendment to the TLA equation, by dividing by 1.205, produced predicted ages consistently lower than those predicted by LRAP. However, while this reduced the number of predictions that were too high, it increased the number of predictions that were too low. Therefore, the prediction profiles of LRAP and TLA were broadly

similar at this point in the development of the TLA protocols (see Graph 4.3.4, below).

*Graph 4.3.3:* Comparison of ages predicted using TLA with known ages. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5$ , 10, 15 and 20 years ( $n=80$ ).



*Graph 4.3.4:* Prediction profiles for TLA and LRAP. The deviation from the known age is on the horizontal axis, while the percentage of predictions that were accurate to this degree is on the vertical axis ( $n=80$ ). TLA is denoted by the black line and LRAP by the dashed line. LRAP and, at this point, TLA have largely similar prediction profiles.



### 4.3.3: Restriction of $TMF_A/TMF_C$ values

As demonstrated in Graph 4.3.4, LRAP and, up to this point, TLA cannot accurately predict ages due to the variation in values of TMF in people of similar ages. However, the hypothesis was that, as  $TMF_A/TMF_C$  tends towards 1, so does the value of  $Age_A/Age_C$ . In other words, the more similar two individuals' TMFs, the more likely they are to be of similar age.

TLA as described compares an unknown individual's TMF to all those in a database, without selectively favouring TMFs similar to that of the unknown individual. In order to assess the accuracy of TLA if values of  $TMF_C$  similar to  $TMF_A$  were favoured, it was necessary to determine upper and lower thresholds of acceptable values of  $TMF_A/TMF_C$ . Given that the intention was to select values of  $TMF_C$  similar to  $TMF_A$ , the central value of  $TMF_A/TMF_C$  was taken to be 1.

The standard deviation of  $TMF_A/TMF_C$ , 0.985, suggests that 69% of values of  $TMF_A/TMF_C$  lie within 1 standard deviation of 1, which is between 0.015 and 1.985, and 38% within 0.5 standard deviations, or between 0.5075 and 1.4925. In practice, it was more straightforward to discount  $TMF_A/TMF_C$  values above or below specific thresholds that approximated particular standard deviation increments. Several upper and lower thresholds were assessed (see Table 4.3.3).

Threshold $TMF_A/TMF_C$		Standard Deviation Equivalent	Significance	n
Lower	Upper			
None	None	N/A	N/A	6,400
0.50	1.50	$\pm 0.508$	0.442	3,570
0.70	1.30	$\pm 0.305$	0.654	2,212
0.75	1.25	$\pm 0.254$	0.708	1,867
0.80	1.20	$\pm 0.203$	0.765	1,502
0.85	1.15	$\pm 0.152$	0.815	1,185
0.90	1.10	$\pm 0.102$	0.864	870
0.95	1.05	$\pm 0.051$	0.922	502

Table 4.3.3: The effect of thresholds on database size. Significance was calculated as  $1-(n/6400)$ , working on the principle that significance is inversely proportional to the number of values that fall within the predetermined thresholds.

The effect of thresholds in the calculation of a mean Age<sub>A</sub> can be illustrated using Participant 1 as an example (see Table 4.3.4).

Part. <sub>C</sub>	TMF <sub>A</sub> /TMF <sub>C</sub>	Upper & Lower Thresholds								Age <sub>C</sub>	Predicted	
		None	1±0.5	1±0.3	1±0.25	1±0.2	1±0.15	1±0.1	1±0.05		Age <sub>A</sub> /Age <sub>C</sub>	Age <sub>A</sub>
1	1.000	✓	✓	✓	✓	✓	✓	✓	✓	34.500	1.000	34.500
2	0.478	✓								44.333	1.053	46.674
3	0.625	✓	✓							28.917	1.040	30.014
4	0.764	✓	✓	✓	✓					30.667	1.024	31.399
5	0.271	✓								42.417	1.074	45.549
6	0.892	✓	✓	✓	✓	✓	✓			23.250	1.011	23.503
...	...	...	...	...	...	...	...	...	...	...	...	...
80	1.123	✓	✓	✓	✓	✓	✓			57.500	0.988	56.787
81	0.507	✓	✓							51.000	1.050	53.547
		38.347	40.831	38.612	38.978	37.445	37.903	38.368	44.110			
		<b>Mean Age<sub>A</sub></b>										

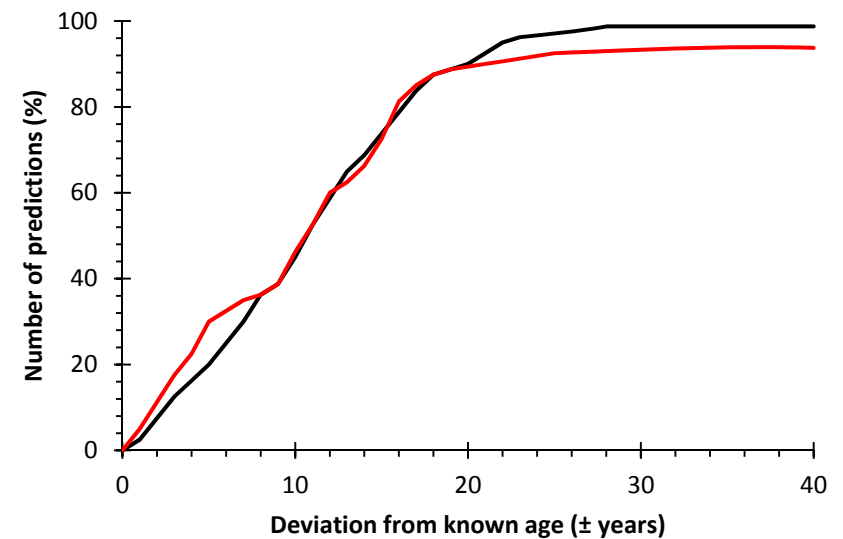
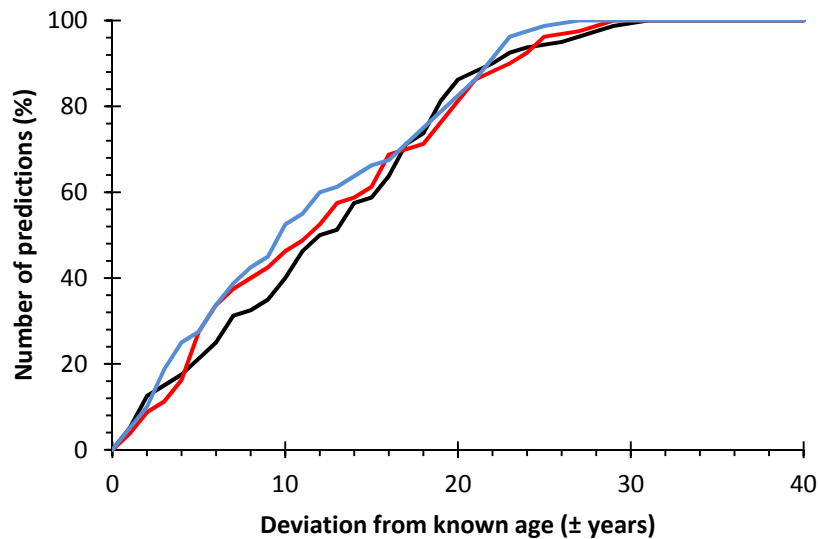
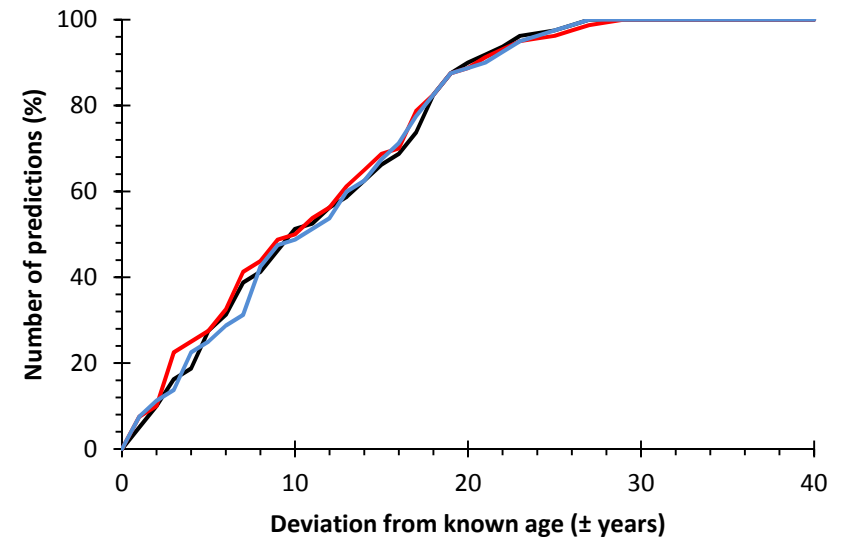
Table 4.3.4: The effect of threshold values of TMF<sub>A</sub>/TMF<sub>C</sub> on age prediction. Using Participant 1, aged 34 years, 6 months at the time of sampling, with a TMF of 632, as an example, values of TMF<sub>A</sub>/TMF<sub>C</sub> can be included in (indicated by ticked cells) or excluded from (indicated by shaded cells) those used to calculate Age<sub>A</sub>. For example, TMF<sub>A</sub>/TMF<sub>4</sub>, 0.764, is included in the analysis if there are no thresholds or if the accepted ranges are ±0.5, ±0.3 or ±0.25. 1.024, the predicted value of Age<sub>A</sub>/Age<sub>4</sub>, results in a predicted Age<sub>A</sub> of 31.399 years and can therefore be incorporated in the calculation of mean Age<sub>A</sub> in these cases. However, if the TMF<sub>A</sub>/TMF<sub>C</sub> threshold ranges are ±0.2, ±0.15, ±0.1 or ±0.05, Age<sub>A</sub>/Age<sub>4</sub> cannot be used to determine the mean Age<sub>A</sub>. TMF<sub>A</sub>/TMF<sub>80</sub>, 1.123, is included in the analysis if there are no thresholds, or thresholds of ±0.5, ±0.3, ±0.25, ±0.2 or ±0.15. 0.988, the predicted value of Age<sub>A</sub>/Age<sub>80</sub>, results in a predicted Age<sub>A</sub> of 56.787 years, which can be used to calculate mean Age<sub>A</sub> in these cases. If the TMF<sub>A</sub>/TMF<sub>C</sub> thresholds were set at ±0.1, or ±0.05, TMF<sub>A</sub>/TMF<sub>80</sub> would be excluded, and the predicted Age<sub>A</sub>/Age<sub>80</sub> and resultant Age<sub>A</sub> would not be incorporated into calculation of the mean Age<sub>A</sub>.

Using these threshold values for TMF<sub>A</sub>/TMF<sub>C</sub>, the process of determining the mean value of Age<sub>A</sub> for all 80 participants was repeated and used to generate prediction profiles for all threshold values (see Graphs 4.3.5 to 4.3.7, overleaf).

Graph 4.3.5 (below): Prediction profiles for unrestricted TLA (black line) and TLA with restrictive thresholds at  $\pm 0.5$  (red line) and  $\pm 0.3$  (blue line).

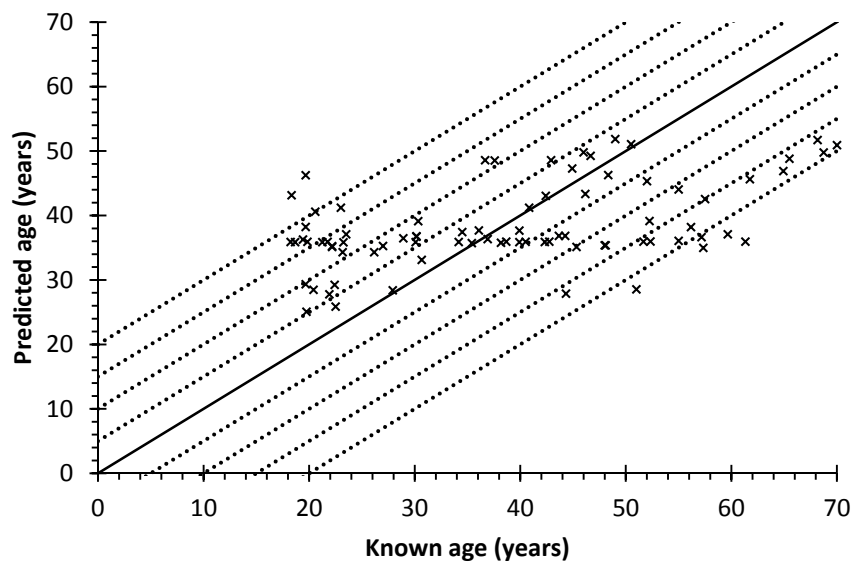
Graph 4.3.6 (upper right): Prediction profiles for TLA with restrictive thresholds at  $\pm 0.25$  (black line),  $\pm 0.2$  (red line) and  $\pm 0.15$  (blue line).

Graph 4.3.7 (lower right): Prediction profiles for TLA with restrictive thresholds at  $\pm 0.1$  (black line) and  $\pm 0.05$  (red line). With thresholds at  $\pm 0.1$ , Participant 5 (1.25% of participants) had no included  $TMF_A/TMF_C$  values other than the one resulting from  $TMF_5/TMF_5$ , and so could not have an age predicted. With thresholds at  $\pm 0.05$ , Participants 5, 7, 19, 62 and 63 (6.25% of participants) could not have their ages predicted for the same reason.

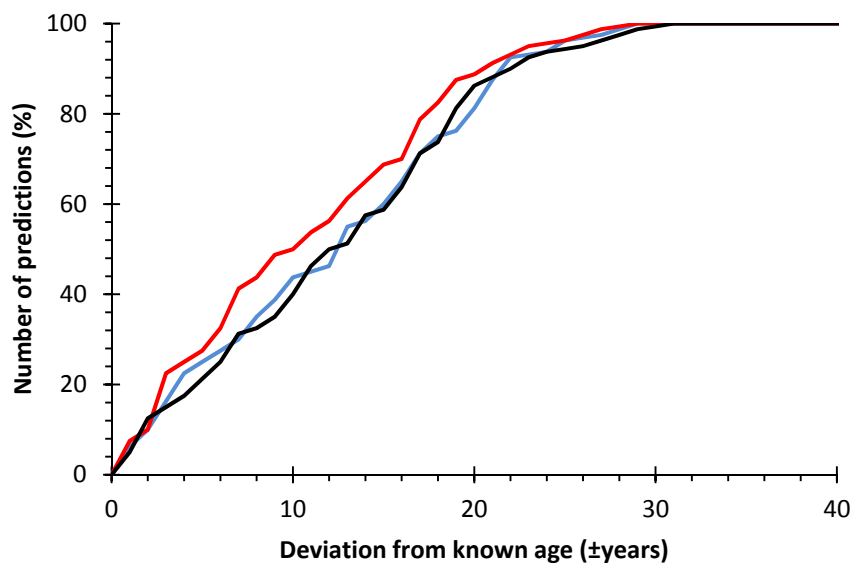


Thresholds at  $\pm 0.2$ , approximately  $\pm 0.2$  standard deviations, resulted in the greatest accuracy, while still producing predictions for all participants. Accordingly, thresholds at  $\pm 0.2$  were accepted as the standard for TLA (see Graphs 4.3.8 and 4.3.9).

*Graph 4.3.8:* Comparison of ages predicted using restricted TLA with known ages. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5, 10, 15$  and  $20$  years ( $n=80$ ). The standard deviation in age predicted using restricted TLA is 12.793.



*Graph 4.3.9:* Prediction profiles for TLA (black line), restricted TLA (red line) and LRAP (blue line). The deviation from the known age is on the horizontal axis, while the percentage of predictions that were accurate to this degree is on the vertical axis ( $n=80$ ).



#### 4.3.4: Gender-differentiated TLA

Based on the fact that GD-LRAP has a higher predictive power than simple LRAP, a similar process was applied to TLA.

Gender-differentiated TLA (GD-TLA) was not expected to be substantially more accurate than TLA, because the tendency for males to have higher TMFs in buccal epithelial samples than females would be compensated for by the inclusion of restrictions at approximately  $\pm 0.2$  standard deviations. More values of  $TMF_A/TMF_C$ , where the subject of analysis is male, would result from comparisons with other males, because females, with typically lower TMFs, would be more likely to generate  $TMF_A/TMF_C$  outside the restrictive range.

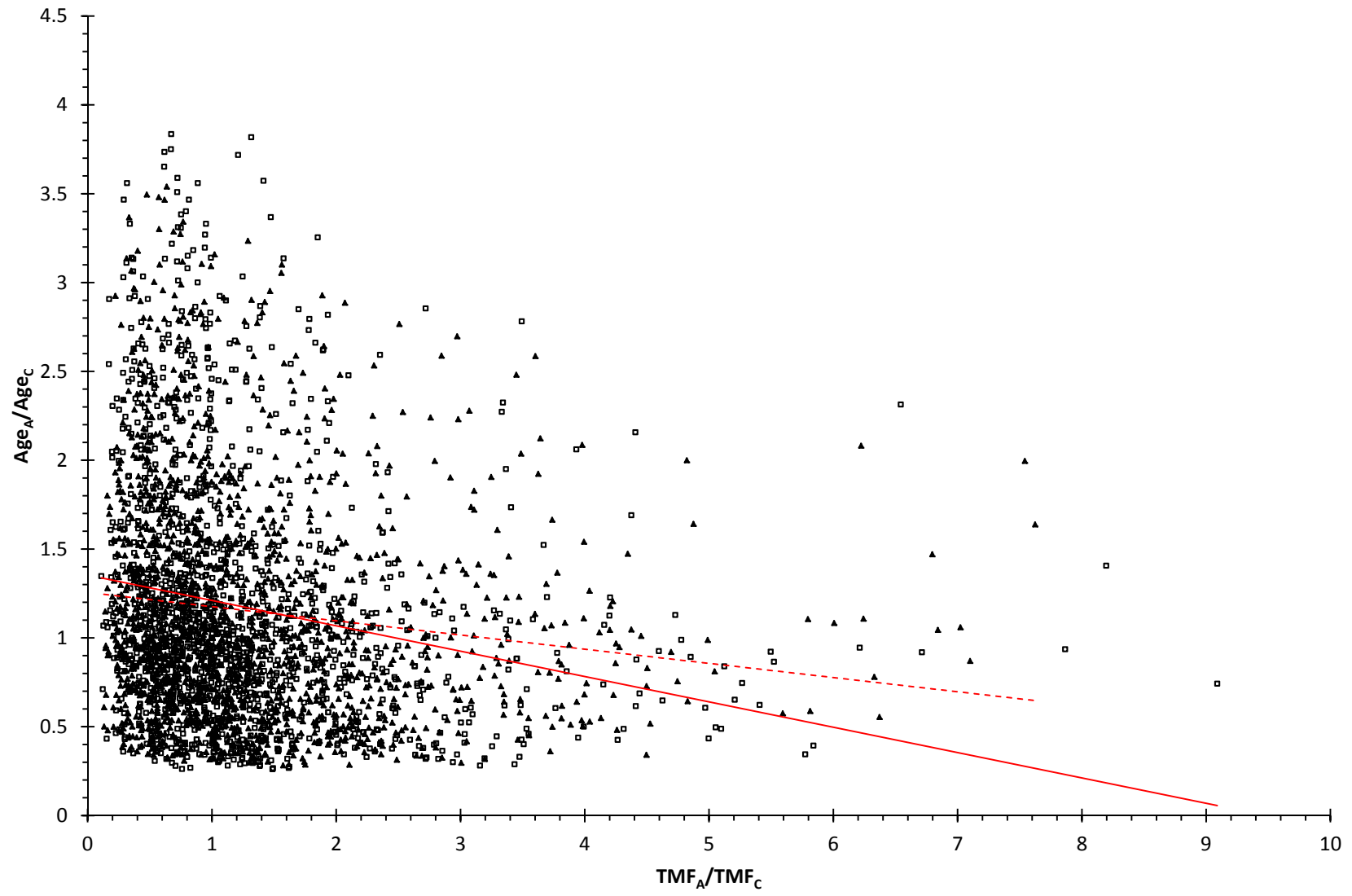
Participants were divided according to gender and the whole process of TLA repeated with gender-specific databases (see Graph 4.3.10, overleaf).

For male participants, the equation of the linear regression was  $Age_A/Age_C = -0.142(\tau_A/\tau_C) + 1.353$ . If  $TMF_A/TMF_C = 1$ ,  $Age_A/Age_C = -0.142 \times 1 + 1.353 = 1.495$ . Therefore, the equation was amended to  $Age_A/Age_C = (-0.142(\tau_A/\tau_C) + 1.353) / 1.495$  for subsequent calculations for male participants only.

For female participants, the equation of the linear regression was  $Age_A/Age_C = -0.079(\tau_A/\tau_C) + 1.255$ . If  $TMF_A/TMF_C = 1$ ,  $Age_A/Age_C = -0.079 \times 1 + 1.255 = 1.176$ . Therefore, the equation was amended to  $Age_A/Age_C = (-0.079(\tau_A/\tau_C) + 1.255) / 1.176$  for subsequent calculations for female participants only.

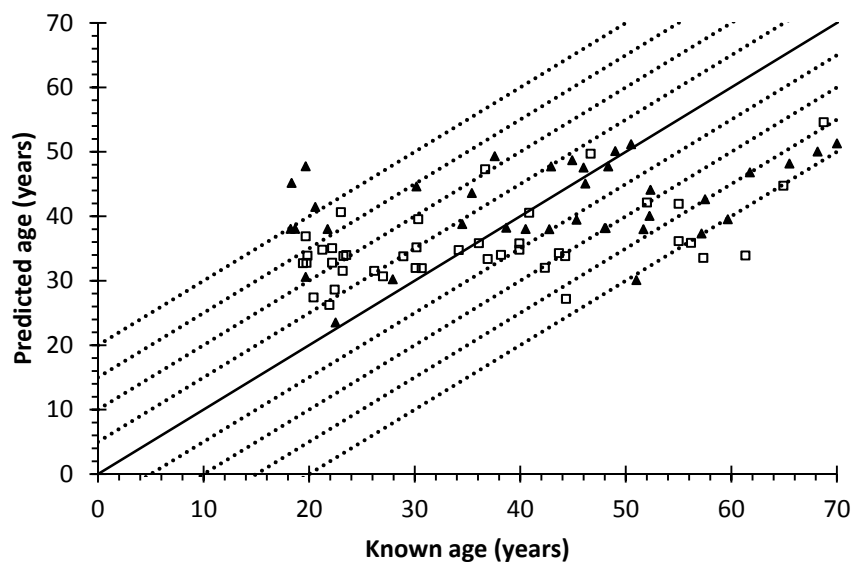
The standard deviation of  $TMF_A/TMF_C$  was 0.957 for males and 1.002 for females, which set the threshold values of  $TMF_A/TMF_C$  at  $\pm 0.2$  for both male and female participants, working on an approximate threshold of  $\pm 0.2$  standard deviations. These thresholds were equivalent to 0.191 and 0.2 standard deviations in males and females respectively.

*Graph 4.3.10 (overleaf):* The relationship between  $TMF_A/TMF_C$  and  $Age_A/Age_C$  in males and females. These show weak correlations between  $TMF_A/TMF_C$  and  $Age_A/Age_C$  in males (open squares and solid line;  $Age_A/Age_C = -0.142(\tau_A/\tau_C) + 1.353$ ;  $R^2 = 0.038$ ;  $n = 1,440$ ) and in females (filled triangles and dashed line;  $Age_A/Age_C = -0.079(\tau_A/\tau_C) + 1.255$ ;  $R^2 = 0.016$ ;  $n = 1,764$ ). The standard deviations of  $TMF_A/TMF_C$  were 0.957 for males and 1.002 for females.



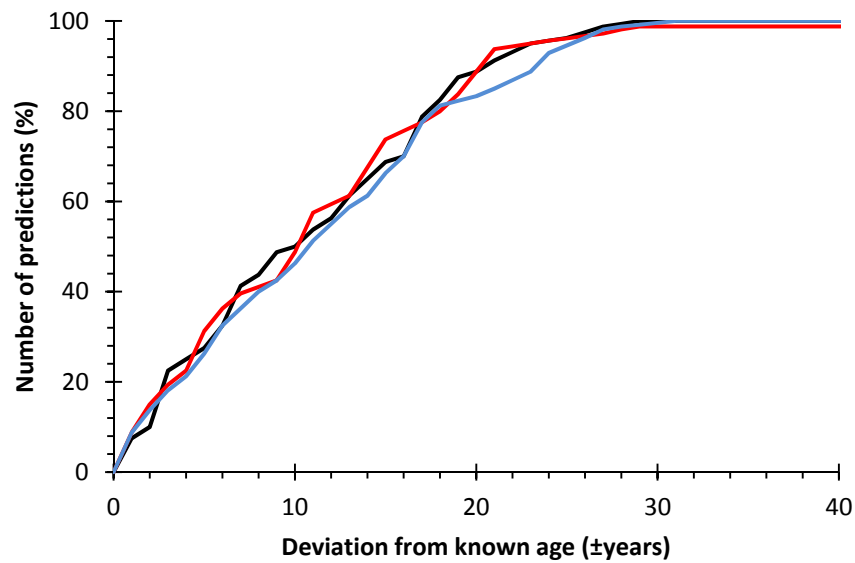
GD-TLA was performed for all 80 participants, using the gender-specific equations developed previously and gender-specific databases (see Graph 4.3.11). Because, when using gender-specific databases, there were no values of  $TMF_5/TMF_C$  within the permitted range of  $1 \pm 0.2$  other than  $TMF_5/TMF_5$ , there were no predicted values of  $Age_5/Age_C$  other than  $Age_5/Age_5$ , or 1. Accordingly, it was impossible to predict the age of Participant 5, and so only 79 participants could have their ages predicted using GD-TLA.

*Graph 4.3.11:* Comparison of ages predicted using GD-TLA with known ages in males (open squares; n=37) and females (filled triangles; n=42). Participant 5, male, had no included values other than  $TMF_5/TMF_5$ . The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5, 10, 15$  and 20 years. The standard deviation was 12.830.



Prediction profiles for TLA, GD-TLA and GD-LRAP are shown in Graph 4.3.12. TLA and GD-TLA both returned more accurate predictions of age than GD-LRAP. As predicted, however, the prediction profiles of TLA and GD-TLA were broadly similar, but the inability of GD-TLA to determine the age of Participant 5 makes GD-TLA less effective than TLA.

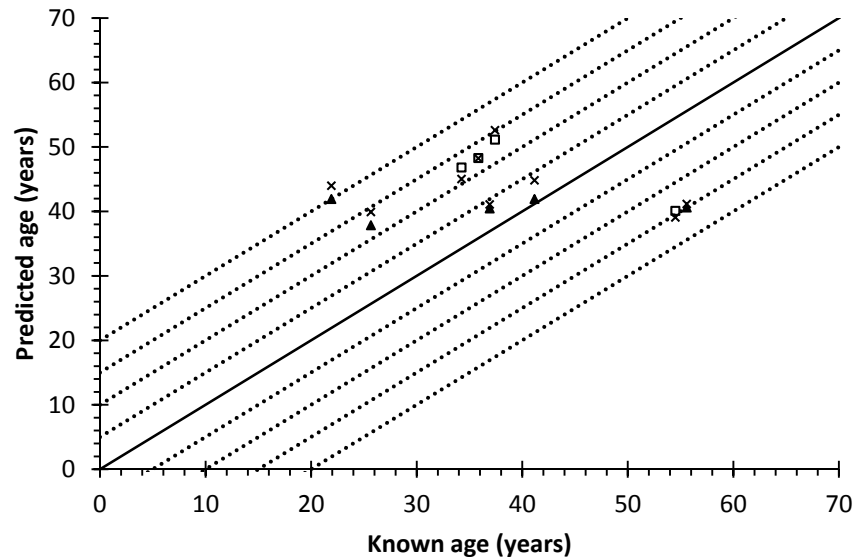
Graph 4.3.12: Prediction profiles for TLA (black line; n=80), GD-TLA (red line; n=79, as described in Graph 4.3.11) and GD-LRAP (blue line; n=80). The deviation from the known age is on the horizontal axis, while the percentage of predictions that were accurate to this degree is on the vertical axis.



#### 4.3.5: Blind testing of TLA and GD-TLA

Using the same blind test participants as in section 4.2.5, TLA and GD-TLA, both using the  $TMF_A/TMF_C=1\pm0.2$  restrictions, were performed (see Graph 4.3.13, overleaf).

Graph 4.3.13: Comparison of ages predicted for test participants using restricted TLA (crosses; n=9) and GD-TLA of male (open squares; n=4) and female (filled triangles; n=5) participants with known ages. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5, 10, 15$  and  $20$  years.



Taking as the null hypothesis ( $H_0$ ) that there is no significant difference between the expected accuracy of the predictions of the blind test participants' ages, based on Graph 4.3.9, and the observed accuracy, Z scores were found for TLA and GD-TLA. These supported the null hypothesis for TLA ( $Z=1.443$ ) and GD-TLA ( $Z=1.268$ ) –  $H_0$  is confirmed if  $-1.96 \leq Z \leq 1.96$ . Therefore, there is no significant difference between the predicted and observed accuracies of age predictions using TLA and GD-TLA, and they remain viable tools to predict age.

#### 4.3.6: Evaluation of TLA

Using the predicted and known ages calculated by TLA, Spearman's Rank Order Correlation Test was performed with the VassarStats Statistical Computation website (Lowry, 2008) and by hand. For TLA, the covariance of predicted age with known age was 0.4561 ( $r_s=0.4561$ ;  $t=4.53$ ;  $n=80$ ;  $df=78$ ;  $P$  (two-tailed)=0.000021). This demonstrates a statistically significant relationship between ages predicted using TLA and the known age of individuals. The correlation, 0.4561, is substantially higher than that shown by LRAP, which was 0.2848. In addition, the prediction profile of restricted TLA, in Graph 4.3.9, suggests that it is substantially more accurate than LRAP and unrestricted TLA, with the majority of age predictions closer to the known age of participants.

This indicates that TLA is a substantially better predictor of the age of an unknown individual than LRAP.

#### 4.3.7: The TLA database increases accuracy

As has been shown in section 4.3.6, the database nature of TLA improves the accuracy of predictions, because upper and lower threshold values can be imposed on a set of 6,400 – 80<sup>2</sup> – data pairs more easily than on a set of only 80 data pairs. TLA is not, at this point, as accurate in its predictions as was desired. It is anticipated that, as the database increases in size, age predictions will become more accurate.

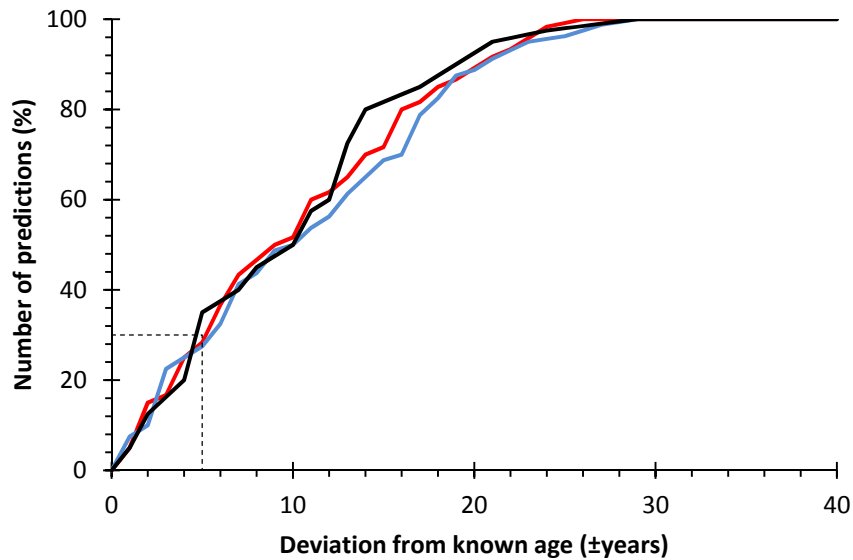
To demonstrate how the prediction profile improves as the database increases in size, smaller subsets of the 80-strong TLA database were created (see Table 4.3.5).

Database	Number of Participants	Number of Data Pairs
1	40	1,600
2	60	3,600
3	80	6,400

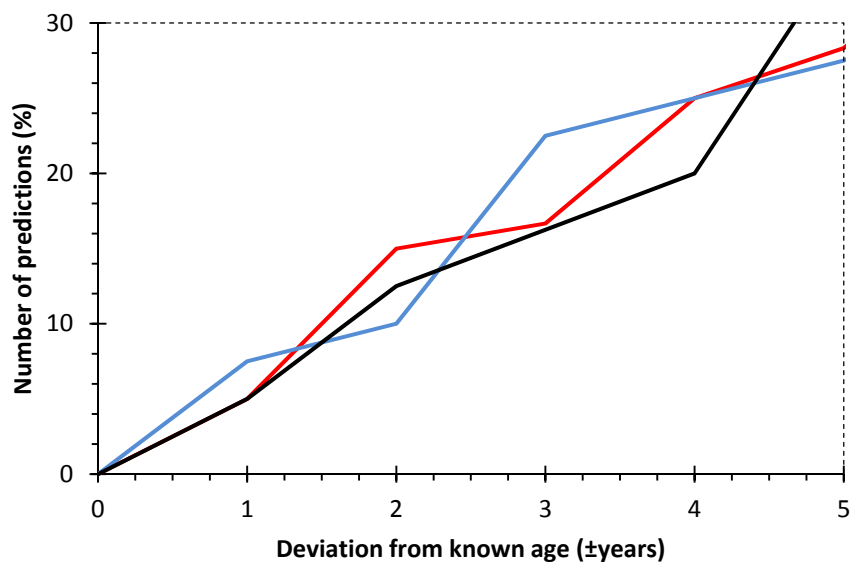
Table 4.3.5: Composition of subset databases.

Using these smaller databases, TLA was performed on all participants in each subset, and the prediction profile of each database determined. Graph 4.3.14 (overleaf) demonstrates an improvement in accuracy as the size of the database increases. Given that, by October 2009, roughly 2,000 profiles were being added per week to the National DNA Database (Travis, 2009a) – although 20% could be expected to belong to people not convicted of any offence (Travis, 2008) – the use of only 1% of incoming DNA samples for TLA would result in a database of over 1,000 individuals inside a year.

Graph 4.3.14: Prediction profiles for TLA using databases of 40 (black line), 60 (red line) or 80 (blue line) participants. Although the 40-strong database returns more predictions accurate to within 15-25 years, both the 60- and 80-strong databases return more predictions accurate to within 0-4 years. Within this range, the 80-strong database is more accurate than the 60-strong database (see detail of area enclosed by dashed line, below).



Detail of Graph 4.3.14: The frequency of predictions accurate to within 0-4 years increases with database size. An 80-strong database (blue line) returns more predictions within 0-4 years than databases of 60 (red line) or 40 (black line) individuals. This is an enlargement of the section of Graph 4.3.14 enclosed by the dashed line.



## **4.4: Telomere Length Analysis & Alternative Tissues**

---

### *4.4.1: Introduction*

For ease of recruiting participants, the majority of samples used were buccal epithelial cells, collected by non-invasive and painless swabs. In the same vein, DNA samples taken from people – as opposed to those collected at the scene of a crime – are usually collected by buccal swabs. Given that these samples are most likely to be accompanied by biographical information such as age, a TLA database will probably consist largely of buccal epithelial cell samples.

However, crime scene DNA samples can be from any of a number of tissues, typically those easily shed from a person – skin, hair or saliva – or those left behind as a consequence of the commission of particular offences – spilt blood from violence and self-defence, or semen from sexual assaults, for example. Semen is only included as a source of crime scene DNA here for completeness: due to telomerase activity during gametogenesis, TLA is entirely useless on semen samples. With the exception of semen, there are two ways in which TLA, as described here, can be of use on other alternative tissues.

Firstly, if the TMF in the alternative tissue varies with age in a similar – although not necessarily proportionate – manner to that in buccal epithelial cells, the principles of TLA can be applied to that alternative tissue. This will necessitate setting up several databases, so that each crime scene sample can undergo TLA by comparison with similar samples.

Secondly, if the TMF in the alternative tissue varies consistently and proportionately with the TMF of buccal epithelial cells, the TMF of a crime scene sample can be converted into a derived buccal TMF. The derived TMF could then undergo TLA using the buccal cell database, which is likely to be larger than any other tissue's database, for the reasons described previously.

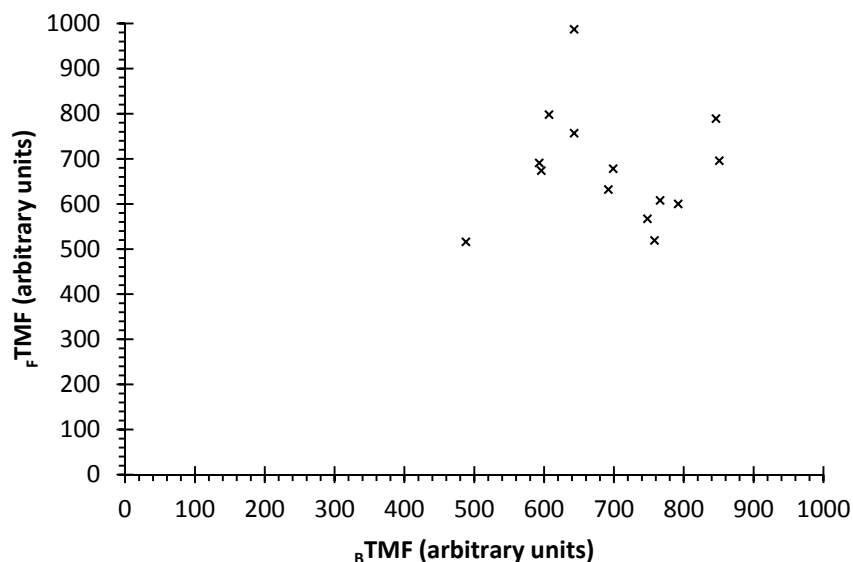
Use of a derived TMF ( $_D$ TMF) to predict a buccal TMF ( $_B$ TMF) is the better of these two options. Of course, the act of calculating  $_D$ TMF is likely to introduce more uncertainty and error into TLA. However, this may be preferable to, and possibly more accurate than, relying on the smaller databases of, for example, hair samples, for TLA. As previously demonstrated in section 4.3.7, TLA accuracy increases with database size.

Accordingly, hair and blood samples were collected from a small group of 14 participants, along with buccal swabs. This allowed the determination of buccal TMF ( $_B$ TMF), hair follicle cell TMF ( $_F$ TMF) and leucocyte TMF ( $_L$ TMF) and the investigation into the suitability of  $_F$ TMF and  $_L$ TMF for TLA as Phase Two of research. While this was a smaller sample size than was hoped for, the prospect of donating a small amount of blood appeared to discourage many potential participants.

#### 4.4.2: Comparison of $_B$ TMF with $_F$ TMF

The intention was to determine a significant relationship between  $_B$ TMF and  $_F$ TMF. This would allow a value of  $_B$ TMF to be determined from a sample of hair, and would allow TLA using the larger database with this  $_D$ TMF value. However, no such relationship was observed (see Graph 4.4.1).

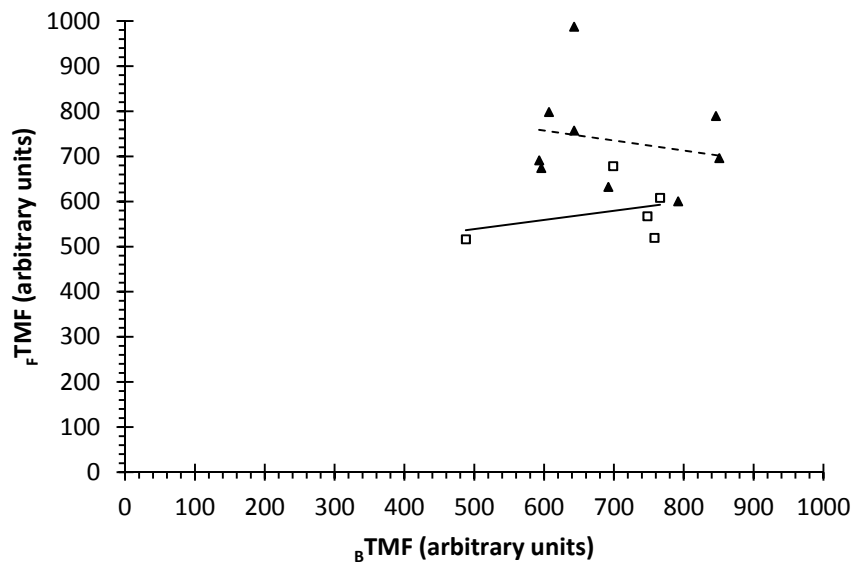
*Graph 4.4.1:* The lack of a relationship between  $_B$ TMF and  $_F$ TMF in all participants. There is a negligible correlation between  $_B$ TMF and  $_F$ TMF ( $R^2=0.001$ ), but this is not significant enough to suggest the presence of any relationship ( $n=14$ ).



Controlling for gender produces weak correlations between  $_B$ TMF and  $_F$ TMF in males and females separately (see Graph 4.4.2, overleaf). However, the fact that the linear regression in males has a positive gradient –  $m=0.202$  – while the same regression in females has a negative gradient –  $m=-0.222$  – suggests that these correlations are an

artefact. While different regressions might be acceptable, they would at least be expected to tend in the same direction. With this in mind, it was concluded that  $f$ TMF is unlikely to be able to predict  $b$ TMF to any degree of accuracy.

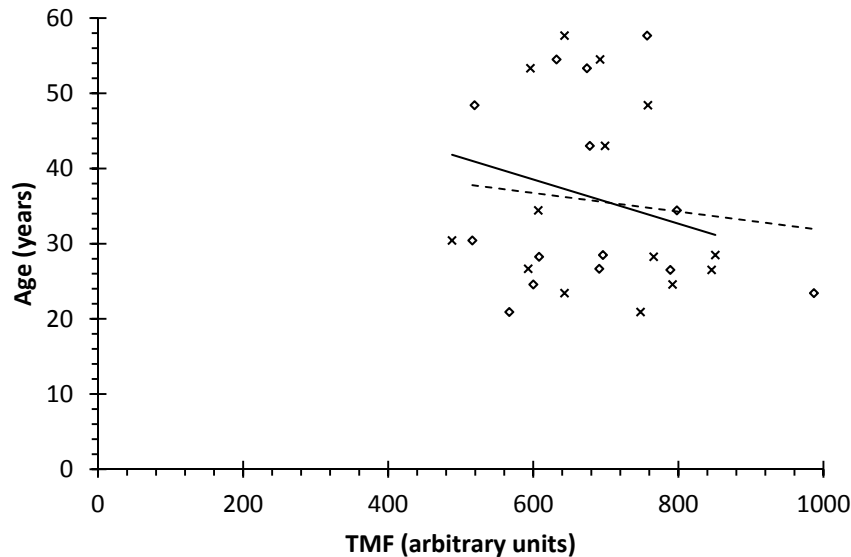
Graph 4.4.2: The relationship between  $b$ TMF and  $f$ TMF in males (open squares and solid line;  $f\tau=0.202b\tau+437.5$ ;  $R^2=0.122$ ;  $n=5$ ) and females (filled triangles and dashed line;  $f\tau=-0.222b\tau+890.7$ ;  $R^2=0.041$ ;  $n=9$ ).



#### 4.4.3: Comparison of $f$ TMF with age

Because  $f$ TMF cannot be used to predict  $b$ TMF with any accuracy, a hair sample cannot be subjected to TLA using a database developed from cheek epithelial cells. However, this does not necessarily mean that hair follicle cells can never be used as a substrate for TLA. Comparing  $f$ TMF and  $b$ TMF with age shows that the correlation between  $f$ TMF and age is approximately 25% of that between  $b$ TMF and age (see Graph 4.4.3, overleaf).

Graph 4.4.3: Comparison of the relationships between  $_B\text{TMF}$  or  $_F\text{TMF}$  and age.  $_B\text{TMF}$  shows a correlation comparable to that observed earlier in, for example, Graph 4.2.1 (crosses and solid line;  $\text{Age} = -0.029_B\tau + 56.17$ ;  $R^2 = 0.057$ ;  $n = 14$ ), while  $_F\text{TMF}$  shows a weaker correlation (open diamonds and dashed line;  $\text{Age} = -0.012_F\tau + 44.18$ ;  $R^2 = 0.014$ ;  $n = 14$ ).

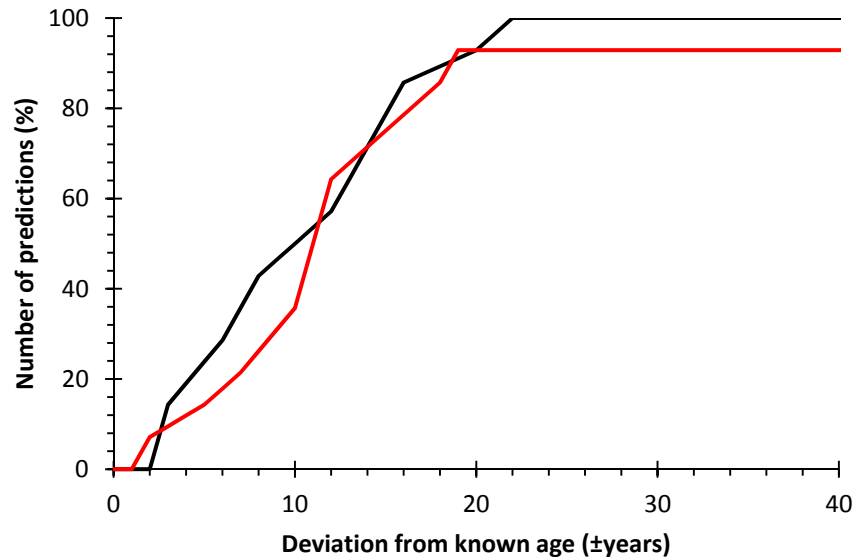


#### 4.4.4: $_F\text{TMF}$ -based TLA

The presence of a correlation, albeit a weak one, between  $_F\text{TMF}$  and age suggested that TLA, using a database derived from values of  $_F\text{TMF}$ , will have some predictive power. To test this, the whole process of TLA was performed for all 14 Phase Two participants for hair follicle cells and cheek epithelial cells. The latter were included to ensure that a comparison between the two prediction profiles would be as fair as possible (see Graph 4.4.4, overleaf).

Although based on a small sample population, TLA based on hair follicle cells still provided some reasonable predictions of age. At smaller deviations from known ages – up to  $\pm 10$  years – the percentage of predictions based on  $_F\text{TMF}$  within this range were roughly two thirds of those based on  $_B\text{TMF}$ . Given that, as has been demonstrated in section 4.3.7, TLA’s prediction profile improves as the sample population increases, it is not impossible that the accuracy of  $_F\text{TMF}$ -based TLA will improve to a point where it provides useful information to prosecuting authorities.

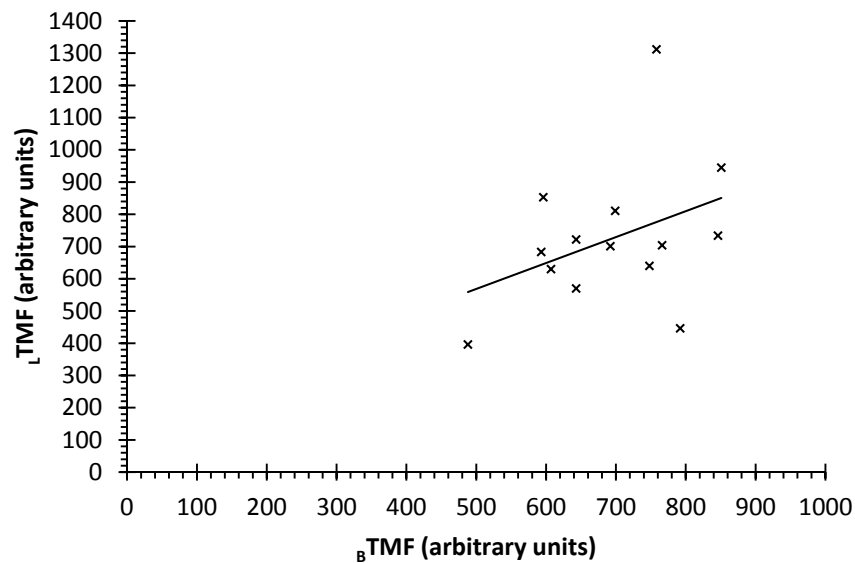
Graph 4.4.4: Prediction profiles for TLA on Phase Two participants, using  $_B$ TMF (black line; n=14) and  $_F$ TMF (red line; n=14). Please note that, because no acceptable values of  $_F$ TMF<sub>A</sub>/ $_F$ TMF<sub>C</sub> within the range  $_F$ TMF<sub>A</sub>/ $_F$ TMF<sub>C</sub>=1±0.2 were returned for Participant 2.4, other than  $_F$ TMF<sub>2.4</sub>/ $_F$ TMF<sub>2.4</sub>, no age could be predicted for this participant using  $_F$ TMF.



#### 4.4.5: Comparison of $_B$ TMF with $_L$ TMF

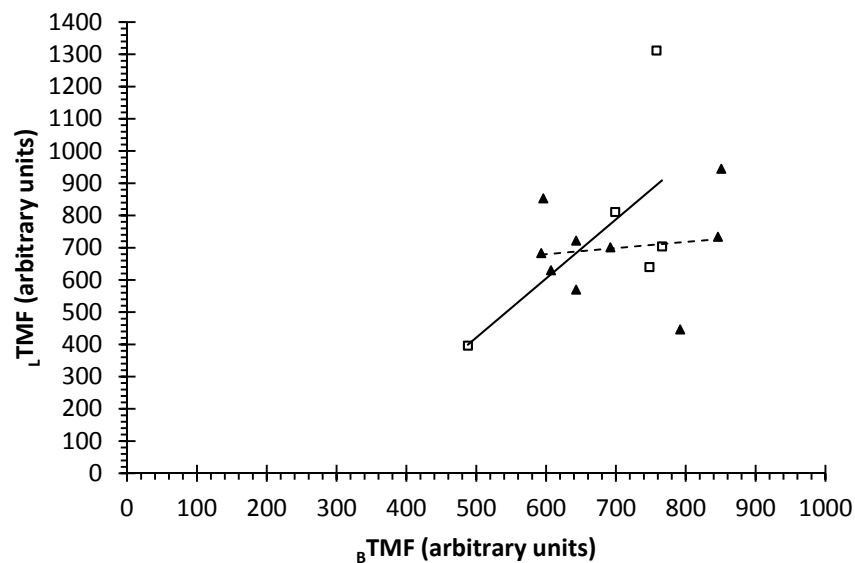
In much the same way as was intended for  $_F$ TMF, the first aim was to determine the presence of any relationship between  $_B$ TMF and  $_L$ TMF. As before, this would allow the calculation of a value of  $_D$ TMF based on  $_L$ TMF, and the use of the larger cheek epithelial TLA database for age prediction. In contrast to  $_F$ TMF, a weak correlation between  $_L$ TMF and  $_B$ TMF was found (see Graph 4.4.5 overleaf).

Graph 4.4.5: The relationship between  ${}_L\text{TMF}$  and  ${}_B\text{TMF}$  in all participants. This shows a correlation between  ${}_L\text{TMF}$  and  ${}_B\text{TMF}$  ( ${}_L\tau=0.804{}_B\tau+166.2$ ;  $R^2=0.144$ ;  $n=14$ ), which is substantially stronger than that observed between  ${}_F\text{TMF}$  and  ${}_B\text{TMF}$ .



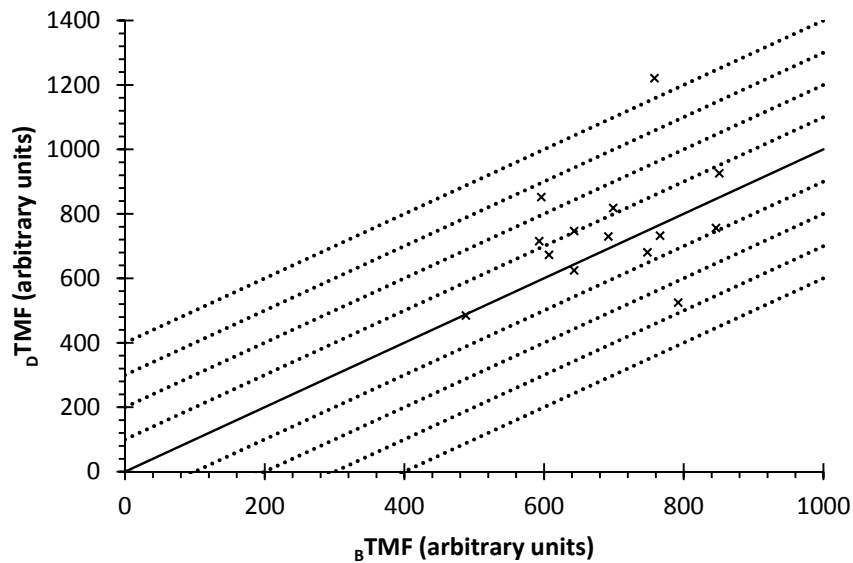
Controlling for gender shows a substantially stronger correlation between  ${}_L\text{TMF}$  and  ${}_B\text{TMF}$  in males than in females (see Graph 4.4.6). In addition, both the male and female linear regressions have a positive gradient, which suggests that these correlations are less likely to be artefacts than those between  ${}_F\text{TMF}$  and  ${}_B\text{TMF}$ .

Graph 4.4.6: The relationship between  ${}_L\text{TMF}$  and  ${}_B\text{TMF}$  in males (open squares and solid line;  ${}_L\tau=1.833{}_B\tau-496.1$ ;  $R^2=0.402$ ;  $n=5$ ) and females (filled triangles and dashed line;  ${}_L\tau=0.190{}_B\tau+565.9$ ;  $R^2=0.018$ ;  $n=9$ ).



Although  ${}_L\text{TMF}$  is able to predict  ${}_B\text{TMF}$ , any such prediction is not entirely accurate, with a standard deviation in the error of prediction of 169.873 (see Graph 4.4.7).

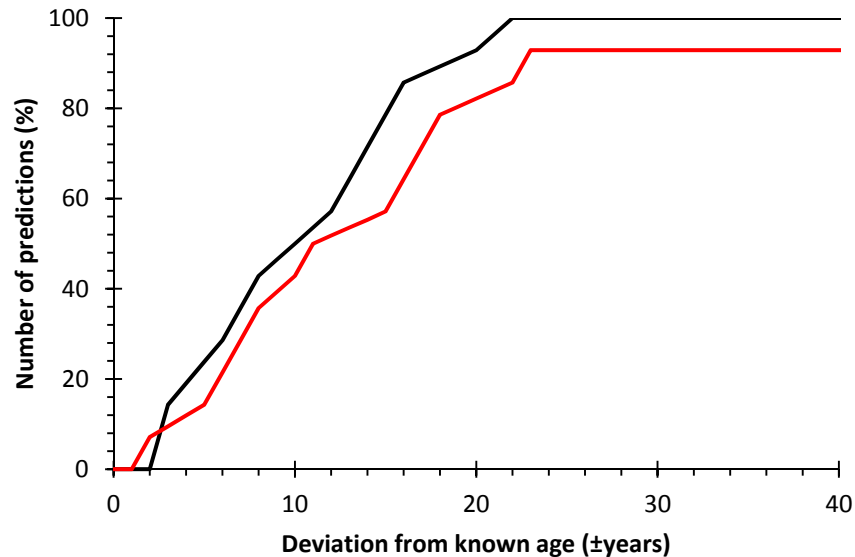
Graph 4.4.7: Comparison of  ${}_B\text{TMF}$  values derived from  ${}_L\text{TMF}$  ( ${}_D\text{TMF}$ ) with known  ${}_B\text{TMF}$  values. The solid line indicates where  ${}_D\text{TMF}$ , using the relationship described in Graph 4.4.5 ( ${}_L\tau=0.804{}_B\tau+166.2$ ), is equal to the known  ${}_B\text{TMF}$ . The dotted lines denote errors of  $\pm 100$ ,  $\pm 200$ ,  $\pm 300$ , and  $\pm 400$  arbitrary units ( $n=14$ ). The standard deviation in error of prediction was 169.873.



#### 4.4.6: ${}_L\text{TMF}$ -based ${}_D\text{TMF}$ TLA

TLA was performed for Phase Two participants based on  ${}_B\text{TMF}$  and – by using the relationship described in Graph 4.4.5 to calculate  ${}_D\text{TMF}$  – on  ${}_L\text{TMF}$ . Using only the Phase Two participants, as before, allowed a fairer comparison between the two prediction profiles (see Graph 4.4.8, overleaf). However,  ${}_L\text{TMF}$ -based TLA consisted of using  ${}_L\text{TMF}$  to calculate a value for  ${}_D\text{TMF}$ , which was then used as the  $\text{TMF}_A$ , using values of  ${}_B\text{TMF}$  as the comparative  $\text{TMF}_C$  values.

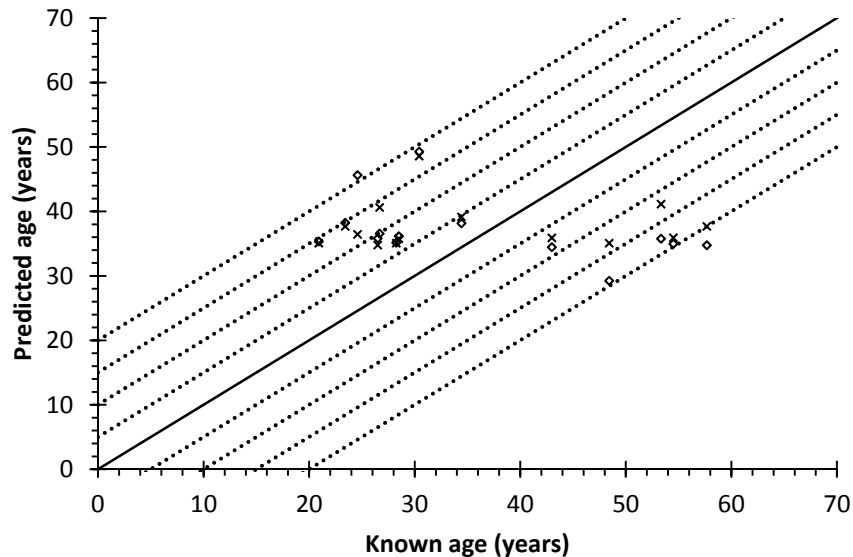
Graph 4.4.8: Prediction profiles for TLA on Phase Two participants using  $_B$ TMF (black line, n=14) and using  $_L$ TMF to calculate  $_D$ TMF values (red line, n=14). Please note that, for the latter, Participant 2.5's age could not be predicted, as the  $_D$ TMF value did not produce any acceptable values of  $_D$ TMF $_A$ / $_D$ TMF $_C$  where  $_D$ TMF $_A$ / $_D$ TMF $_C$ = $1\pm 0.2$ , other than  $_D$ TMF $_{2.5}$ / $_D$ TMF $_{2.5}$ .



As with  $_F$ TMF-based TLA,  $_L$ TMF-based TLA, using  $_D$ TMF, produced some reasonable predictions of age. Again,  $_L$ TMF-based TLA produced deviations within  $\pm 10$  years approximately two thirds as often as  $_B$ TMF-based TLA. And again, given the trend for increasing accuracy as the size of the TLA database increases, this suggests that the accuracy of  $_L$ TMF-based TLA will improve.

To confirm this,  $_L$ TMF-based  $_D$ TMF TLA was performed on all 14 Phase Two participants using the much larger 80-participant database developed earlier. As before, the 14 values of  $_L$ TMF were converted into  $_D$ TMF values, and used as TMF $_A$ , using the larger database. As a comparison, the  $_B$ TMF of the Phase Two participants was used for conventional TLA (see Graph 4.4.9, overleaf).

Graph 4.4.9: Comparison of ages predicted by restricted TLA with known ages, using  $_B$ TMF (crosses) and  $_D$ TMF (open diamonds) for Phase Two participants. The solid line indicates where the known age is equal to the predicted age, while the dotted lines denote errors of  $\pm 5, 10, 15$  and  $20$  years ( $n=14$ ). The standard deviation in age predicted using  $_B$ TMF was 13.371, and 15.593 using  $_L$ TMF-based  $_D$ TMF.



This suggested that  $_L$ TMF-based  $_D$ TMF TLA is not as accurate a predictor of age as  $_B$ TMF TLA. This was not altogether surprising, given that  $_L$ TMF has a correlation of only 0.244 with  $_B$ TMF.

However, taking as the null hypothesis ( $H_0$ ) that there is no significant difference between the expected accuracy of predictions of Phase Two participants' ages using  $_L$ TMF-based  $_D$ TMF TLA – which was the same as with conventional TLA – and the observed accuracy, the Z test was performed. This supported the null hypothesis for  $_L$ TMF-based  $_D$ TMF TLA ( $Z=0.441$ ) –  $H_0$  is confirmed if  $-1.96 \leq Z \leq 1.96$ . There is, therefore, no statistically significant difference in the accuracies of  $_B$ TMF TLA and  $_L$ TMF-based  $_D$ TMF TLA.

## **4.5: TMF & Other Factors**

---

### *4.5.1: Introduction*

As well as identified lifestyle factors that affect telomere length (Cherkas et al., 2008; Richards et al., 2007; Farzaneh-Far et al., 2010), it was possible that height and body mass might also have an effect on TMF. Both height and body mass are products of a history of greater cell division. Comparing a tall individual with a shorter one, the former will be made up of more cells than the latter. Similarly, although body mass is partly due to the size of fat droplets stored in adipose cells as well as the number of those cells (Campbell, 1996), a heavier individual may well contain more cells than a lighter person.

The identified lifestyle factors shown to affect telomere length, such as exercise (Cherkas et al., 2008) and diet (Richards et al., 2007; Farzaneh-Far et al., 2010), are unlikely to be of much use as part of a molecular photofit. Height, body mass and build, as roughly indicated by the Body Mass Index, are more obvious in an individual than his or her diet, and so are of more potential use.

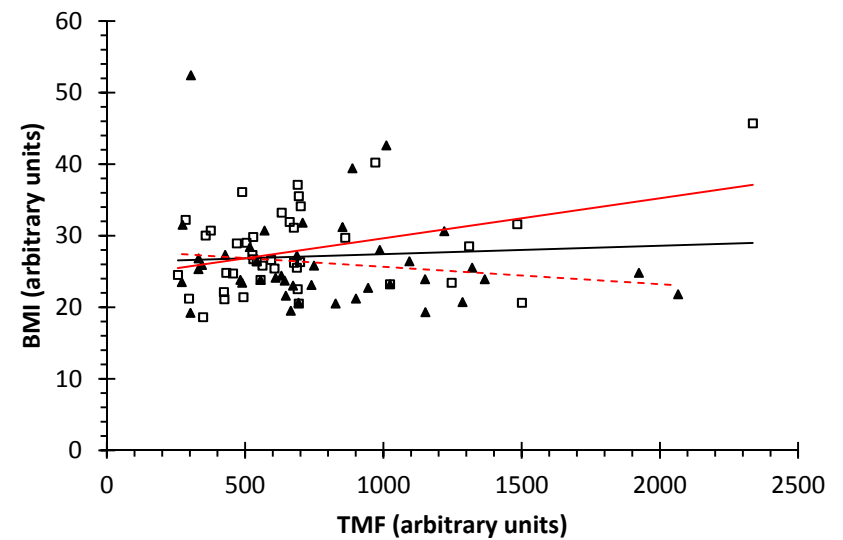
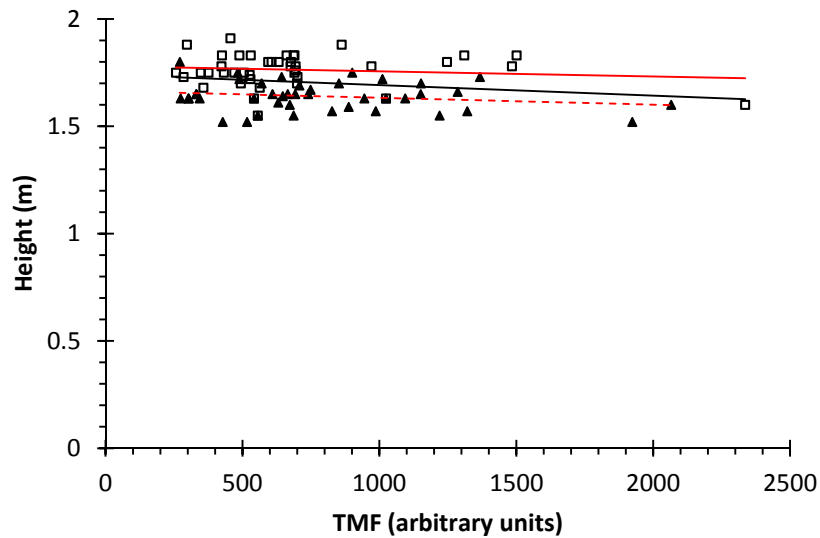
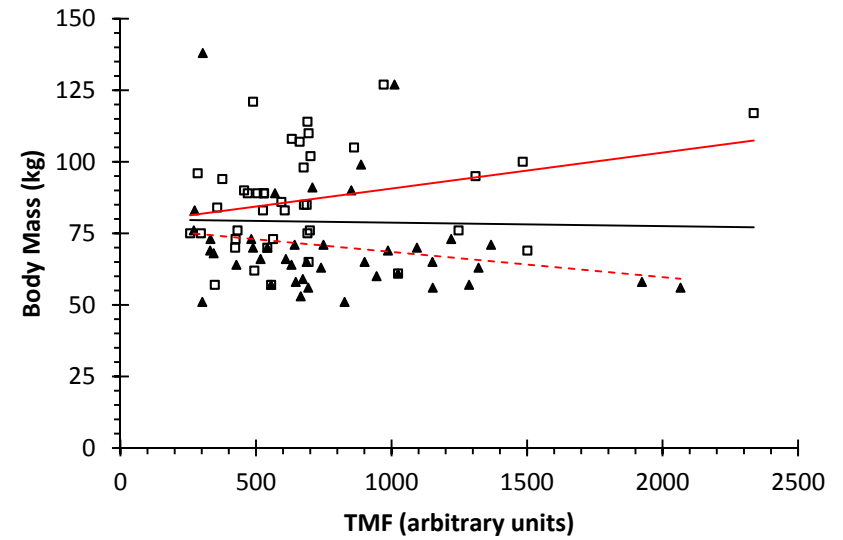
### *4.5.2: TMF and height, body mass and BMI*

There were no significant correlations between TMF and height, body mass or BMI, even after controlling for gender (see Graphs 4.5.1 to 4.5.3, overleaf). Accordingly, it was concluded that TMF is unlikely to yield an accurate prediction of height, weight or BMI as part of a molecular photofit. Given that most studies linking lifestyle to telomere length involved the testing of blood samples, it was not surprising that buccal epithelial samples did not show any relationship between telomere length and lifestyle. Blood, as a tissue required to react to the health of an individual by increasing the rate of leukocyte production to defend against disease, is more likely to show such a relationship.

Graph 4.5.1 (below): The absence of a relationship between TMF and height in all participants (black line;  $R^2=0.046$ ;  $n=80$ ) or in males (open squares and red solid line;  $R^2=0.016$ ;  $n=38$ ) and females (filled triangles and red dashed line;  $R^2=0.038$ ;  $n=42$ ) separately.

Graph 4.5.2 (upper right): The absence of a relationship between TMF and body mass in all participants (black line;  $R^2=0.000$ ;  $n=80$ ) or in males (open squares and red solid line;  $R^2=0.080$ ;  $n=38$ ) and females (filled triangles and red dashed line;  $R^2=0.041$ ;  $n=42$ ) separately.

Graph 4.5.3 (lower right): The absence of a relationship between TMF and BMI in all participants (black line;  $R^2=0.006$ ;  $n=80$ ) or in males (open squares and red solid line;  $R^2=0.149$ ;  $n=38$ ) and females (filled triangles and red dashed line;  $R^2=0.024$ ;  $n=42$ ) separately.



## Chapter Five: Discussion & Conclusions

---

<b>5.1</b>	<b><i>The Legal Implications of Telomere Length Analysis</i></b>	<b>99</b>
5.1.1	Introduction	99
5.1.2	Operation Minstead	99
5.1.3	Molecular photofitting as intelligence	100
5.1.4	Molecular photofitting and the defence	100
5.1.5	Molecular photofitting and s78 exclusions	101
<b>5.2</b>	<b><i>The Practical Implications of Telomere Length Analysis</i></b>	<b>103</b>
5.2.1	Expense	103
5.2.2	Integration into existing protocols	103
5.2.3	TLA's inaccuracy	104
5.2.4	TLA and random matches	104
5.2.5	TLA and forensic osteology	105
5.2.6	Conclusion	106
<b>5.3</b>	<b><i>The Technical Limitations of Telomere Length Analysis</i></b>	<b>107</b>
5.3.1	TLA and sexual offences	107
5.3.2	TLA and mixed samples	107
5.3.3	Further research	108
5.3.4	Telomere degradation	108
5.3.5	TLA, adolescence and old age	109
<b>5.4</b>	<b><i>Conclusion</i></b>	<b>111</b>

## ***5.1: The Legal Implications of Telomere Length Analysis***

---

### *5.1.1: Introduction*

When considering Telomere Length Analysis, or any new scientific technique for the investigation of crimes, it is perhaps appropriate to break the discussion down into three parts: how the courts are likely to handle evidence arising from the new technique, how investigating authorities are likely to make use of it, and its limitations and what further research is required.

### *5.1.2: Operation Minstead*

It is perhaps difficult to predict how molecular photofitting, which would incorporate TLA, would be handled by the UK courts. The overwhelming majority of cases discussed by Frudakis (2008) are American. While the American and British legal systems share common ancestry, both have diverged significantly since Independence. Using American examples to illustrate how the UK courts might consider molecular photofitting is therefore unwise. Furthermore, there is a paucity of information regarding molecular photofitting and the courts in available law reports.

The one British case that Frudakis discusses is that of the 17-year investigation to identify the ‘night stalker’, suspected of the rape or sexual assault of over 100 elderly people – Operation Minstead. In 2008, when Frudakis was writing, no individual had been charged with these offences, but his company was instrumental in identifying the likely ethnic background of the ‘night stalker’. Working on the basis that the ‘night stalker’ was of Caribbean ancestry, specifically the Windward Islands, police launched the mass screening of 1000 black men, as discussed in section 1.3.1 (Laville, 2009; McCartney, 2006).

Eventually, in November 2009, an individual was arrested and charged. The case has yet to be brought to court, and so details are not available, but the offences were linked by DNA profiling, so it is almost certain that one of the principal pieces of evidence to be introduced will be the defendant’s DNA profile (Laville, 2009).

### *5.1.3: Molecular photofitting as intelligence*

Operation Minstead in fact illustrates how molecular photofitting could be used: as a source of intelligence for the police in identifying suspects. Because a molecular photofit requires the analysis of many individual markers, a relatively undamaged sample of DNA is required. Given that several of the markers used to identify ethnic origin are common to both the SGM+ and CODIS protocols, it is obvious that any sample producing a molecular photofit will also produce a conventional DNA profile.

There would be no requirement for prosecuting authorities to enter into evidence a molecular photofit, because it will almost certainly be accompanied by a DNA profile match. Obviously, this profile match will be bound by the caveats of a Random Match Probability, but the prosecution will claim that it adequately identifies the defendant. Similarly, there is little need to enter as evidence a conventional photofit or e-fit image of a suspect, if there is fingerprint evidence available – the former is subjective, requiring the estimation of how well the defendant fits the picture, while the latter is objective and categorical, with definitive matches.

This is why case law on molecular photofitting is sparse, and likely to remain so: the courts have never been required to consider it, and are unlikely to ever have to. Carrying on the comparison with conventional photofitting, most case law and legislation surrounds its proper administration. An identification – by fingerprint, DNA profile or an eyewitness in an identity parade – is the preferred evidence of a suspect's guilt. Such case law and legislation has yet to arise for molecular photofitting but, given that a DNA profile is typically also produced, it is likely to develop in much the same way.

### *5.1.4: Molecular photofitting and the defence*

The results of molecular photofitting, undertaken at the request of investigating authorities, must be passed to the defence as part of the standard process of disclosure of prosecution evidence. If the DNA profile matched to the defendant is partially degraded, it logically increases the likelihood of that match being coincidental. The defence may therefore make this point to the jury, especially if the

defendant does not perfectly match the predictions made in the molecular photofitting. As has been seen, TLA predicts a range of possible ages for an individual – within 5 years just over a quarter of the time – and so the defendant will not necessarily match this predicted age. Other aspects of the molecular photofit – for example, hair colour, where only the presence of red hair can be predicted to any degree of accuracy – may well not match the defendant either. As such, the defence themselves may well present to the jury the molecular photofit, to discredit the prosecution evidence, if the defendant varies from the predicted appearance or age.

#### *5.1.5: Molecular photofitting and s78 exclusions*

Having concluded that molecular photofitting is unlikely to have to be put before a jury by the prosecution, it could still be challenged in the courts.

The judiciary in the United Kingdom have discretionary power to exclude any prosecution evidence if it is likely to undermine the right of the defendant to a fair trial using s78 of PACE – hence the term ‘s78 exclusion’. Before making a decision, usually on the application of the defence, the judge is required to consider, in addition to the relative probative and prejudicial value of the evidence, the circumstances in which it was obtained (s78(1) PACE). Applications for s78 exclusions are made in a special hearing called a *voir dire*, usually as part of the pre-trial process, although the judge has the power to remove the jury to hear a s78 application should the defence request one during the trial (Choo, 2006).

In general, the case law surrounding a s78 exclusion considers undercover operations by the police and improperly obtained confessions – not forced or coerced confessions, which are already automatically excluded (s76(2) PACE). For example, evidence obtained by an undercover police officer during Operation Edzell was excluded as prejudicial in the trial of Colin Stagg for the murder of Rachel Nickell (*R v Stagg* [1994]). Entrapment has been found to be reasonable grounds to grant a s78 exclusion (*Looseley; Attorney-General’s Reference No. 3 of 1999* [2001]).

However, challenges have been made to evidence arising from unlawful arrests under s78. Generally, a police officer is only authorised to make an arrest if he has reasonable suspicion that the individual he is arresting has committed a specific offence (s24(2) PACE). Given that the police are only authorised to demand a DNA sample from an individual if he has been arrested on suspicion of committing a

recordable offence, an unlawful arrest – not on the basis of reasonable suspicion – would mean that the DNA sample was unlawfully taken. On this basis, a defence application to exclude a DNA profile match using s78 would have to be considered by the judge.

Reasonable suspicion of guilt is a poorly defined legal concept – there must be some evidence of that individual having committed an offence, but this need not be admissible in court (Ward and Wragg, 2005). However, the clearest guidance comes from Code of Practice A of PACE, which in fact regulates the police’s statutory powers of stop and search – there is a general consensus that the definition of reasonable suspicion for a stop and search also applies to that required to arrest an individual.

The information obtained from a molecular photofit, incorporating TLA – race, age and appearance – is specifically excluded as the sole basis for suspicion (PACE Code A paragraph 2.2). Instead, suspicion must be based on facts, information or intelligence. Given the degree of information that can be taken from a molecular photofit, this might be considered to be suitable intelligence for reasonable suspicion. In practice, however, it would make sense for investigating authorities to have additional grounds for suspicion, or DNA profiles resulting from arrests made on the basis of a molecular photofit are likely to be challenged under s78.

This having been said, there is precedent for the inclusion of unlawfully obtained evidence, provided that its probative value outweighs its prejudicial value (*Allan v The United Kingdom* [2002]). So long as the evidence is reliable – and a DNA profile match would generally be considered as such – it is typically likely to be admitted, regardless of its provenance, unless it was obtained through violence, intimidation (s76(2) PACE), coercion (*Perry v The United Kingdom* [2002]) or entrapment (*Looseley; Attorney-General’s Reference No. 3 of 1999* [2001]).

Accordingly, a molecular photofit is unlikely to be challenged under s78 of PACE.

## ***5.2: The Practical Implications of Telomere Length Analysis***

---

### *5.2.1: Expense*

Molecular photofitting is in fact an expensive process. Frudakis (2008) reports charging American law enforcement agencies \$1000 for determining only an individual's ethnic origin within the American population – black, white, Hispanic or Native American, among others. While this cost can be justified, if the crime is severe enough – for example, the 'night stalker' attacks – molecular photofitting is unlikely to become a standard part of a criminal investigation.

If the molecular photofit is expanded to include the 320 markers used to determine an individual's origin from within Europe and the 33 SNPs used to determine eye colour, the cost of a molecular photofit that would be of benefit to investigators in Europe is likely to become prohibitive in all but the most serious of cases.

### *5.2.2: Integration into existing protocols*

TLA itself is comparatively cheap, and was designed to be easily integrated into the process of determining a DNA profile – it requires only the use of two lanes in a standard gel. As such, an approximate age can be predicted alongside the DNA profiling procedure. Additionally, processing a sample for TLA, by extracting and amplifying DNA – assuming that both telomeric and single-copy locus PCR can be performed simultaneously on two thermal cyclers – and subsequent electrophoresis, takes approximately the same amount of time as conventional DNA profiling – from start to finish, approximately a single working day. Data analysis for TLA may take more time than for DNA profiling, although this process could probably be automated by specialist computer programmers. Using Microsoft Excel, as here, to determine age takes 15-20 minutes, mainly in calculating  $TMF_A/TMF_C$  values with every participant in the database and applying restrictive thresholds. Automation will make this significantly easier.

Given that many DNA profiling techniques do not use intercalating dyes, as used here, but instead fluorescently labelled primers, the equation for calculating

TMF from the end point of PCR is not actually applicable. Because the labelled primers give an intensity that is proportionate to the number of molecules present, as opposed to the intercalating dye's quantity of DNA, the equation must be amended if TLA is to be added to the standard processes commonly used. Essentially, this amounts to nothing more than removing the  $L_T$  and  $L_S$  terms from the equation, making it:

$$\tau = (I_T 2^{(C_S - C_T)}) / (46 I_S)$$

This amendment to the equation will not affect the measurement of TMF, as the labelled primers allow the measurement of the number of molecules, rather than the total mass, which is a product of the number of molecules and their length.

### *5.2.3: TLA's inaccuracy*

It is probable that, like the other molecular photofitting assays, TLA will mainly be used in extremely serious cases where there is little other evidence. While it is most likely to be only used as intelligence, the relatively low accuracy runs the risk of misleading investigating authorities if not bound by many caveats – if they become too focused on individuals aged between 28 and 33, they may neglect the 34-year-old offender.

Currently, its accuracy is not satisfactory enough for it to be a permanent addition to DNA profiling. However, the demonstration that TLA accuracy improves as the size of its database increases suggests that a TLA database created on a national scale, even taking only 1% of the thousands of samples submitted for DNA profiling every week, may become more accurate, and therefore more useful.

### *5.2.4: TLA and random matches*

While the RMP of a DNA profile is typically given as less than one in a billion, there is evidence to suggest that this is not entirely accurate (Geddes, 2010). Accordingly, it may in fact be more likely that an individual facing trial as a result of a DNA profile match does so as a consequence of an inaccurate RMP. TLA, and indeed other molecular photofitting techniques that do not rely on the standard SGM+ and CODIS

STR markers, can therefore be used to distinguish between two people with matching DNA profiles. If the defence relies on a claim that the profile match is in fact a random one, TLA or molecular photofitting could provide further evidence that links the DNA sample recovered from the scene of the crime to the defendant. Alternatively, if the defendant differs notably from the predicted molecular photofit, this may serve as evidence to support his innocence.

#### *5.2.5: TLA and forensic osteology*

Where TLA may prove to be more useful is in identifying unknown victims of crimes. There have been several cases reported in recent years of bodies being dismembered and scattered over wide areas. When faced with an unidentifiable body part, an age predicted using TLA may help to reduce the number of reported missing persons that must be considered before the victim can be identified.

Most estimates of age of an unknown body rely on forensic osteology and the analysis of age-related changes to an individual's skeleton. For example, the iliac crest of the pelvis fuses completely by around 18 years of age, but the superior and inferior vertebral rings fuse completely by around 25 years of age (Cox, 2000). Logically, therefore, a skeleton presenting a fused iliac crest but unfused vertebral rings is approximately between 18 and 25 years old. The last main skeletal change to occur is ossification of the plastron, which is typically complete by the age of 35 (Cox, 2000). Therefore, a skeleton with fused vertebral rings but an incompletely ossified plastron is between 25 and 35 years old, although the degree of ossification can be used to reduce this range.

However, there is considerable variation in the age of completion of some skeletal processes. For example, fusion of the petroexoccipital articulation is usually complete at the age of 34, but has been observed to be incomplete in individuals aged 50 and above. In some cases, incomplete fusion has been observed in individuals who have died of old age (Cox, 2000).

The last skeletal process that can be used to reliably predict the age of skeleton is plastron ossification, which therefore means that ages above 35 cannot be reliably estimated by forensic osteology. This, in turn, means that a substantial number of bodies cannot have their ages determined by forensic osteology (Black, 2000).

TLA may be able to resolve this, as it relies upon the fact that telomeres shorten throughout an individual's life. It should be remarked that telomeres have been successfully recovered and measured from samples left at room temperature for several months (Tsuji et al., 2002). TLA based on blood samples is nearly as accurate as buccal TLA, and likely to become more accurate as database sizes increase. Given that blood cells are produced in the marrow of long bones, this suggests that bone marrow may be a suitable substrate for TLA. These bones are typically the sources of DNA for profiling, and TLA could therefore easily be included into the standard measures for identifying an unknown set of remains, where DNA is recoverable.

#### *5.2.6: Conclusion*

Accordingly, as it presently stands, TLA is not accurate enough to be considered to be a standard part of a criminal investigation. Instead, its use is likely to be limited to identifying unknown bodies, or in a crime severe enough to justify the expense of molecular photofitting.

This section was prepared following consultation with Luan Lunt of LGC Forensics. It is not inappropriate to thank her once more for her time, advice and encouragement.

## ***5.3: The Technical Limitations of Telomere Length Analysis***

---

### *5.3.1: TLA and sexual offences*

Firstly, as discussed in section 4.4.1, TLA cannot work on semen, which is a commonly recovered material following rape or sexual assault. While it would be desirable to be able to use any molecular photofitting technique on any DNA source, the fact that telomerase is activated during gametogenesis and reconstructs eroded telomeres means that TLA will not work on semen.

This is not to say that TLA is useless in sexual offences – frequently, other materials are recovered, especially if the victim struggles and, for example, scratches the attacker. With the evidence reported here that TLA functions to some extent on hair, there is no reason not to imagine that it will not work on pubic hair as well as hair from an individual's head – collection of pubic hair samples would have presented difficulties in participant recruitment for this study, and so was not attempted. Secondly, the recovery of materials suitable for TLA following a struggle – skin under the victim's nails, for example – will serve not only to identify the attacker, but reduce the likelihood of success of any defence of a reasonable belief in consent, by demonstrating the victim's efforts of resistance.

### *5.3.2: TLA and mixed samples*

TLA will be less accurate for samples containing cells from more than one person. A traditional DNA profile can easily be produced from a mixture of cells from an unknown offender and a known victim, simply by taking an uncontaminated DNA sample from the victim and subtracting his or her profile from that arising from the mixture. What remains belongs to the unknown attacker.

However, TLA of a mixed sample would produce an age based on the average of the offender's TMF and the victim's. The relative quantities of DNA profile markers from the offender and the victim might be used to determine the proportion of the mixed sample that belongs to the offender. The victim's TMF, calculated separately, can then be used to determine the proportion of single-copy and telomeric

amplimers following TLA of the mixed sample that belong to the offender, which would then permit the prediction of the offender's age.

### *5.3.3: Further research*

While TLA holds promise as a component of molecular photofitting, there remains a significant amount of further research to be done before it can be universally applied.

Firstly, more individuals must be added to the database, to improve accuracy of predictions. In addition, while  ${}_F$ TMF and  ${}_D$ TMF show some suitability for TLA, these need to be studied to the same depth as TLA using  ${}_B$ TMF, including increasing the sample sizes. Given the difficulty of recruitment of volunteers to provide blood – around 15% of Phase One volunteers agreed to participate in Phase Two – this needs further investigation.

### *5.3.4: Telomere degradation*

Research should be done into degradation of telomeres by the environment after a DNA sample is left. The samples used in this research were kept frozen for the majority of the time, but Tsuji et al. (2005) reports the successful recovery of telomeres after 5 months at room temperature with the loss of only 500bp from TRFs. While damage to the sequence of a telomere would disrupt telomere length measurement by the TRF assay as well as by any qPCR assay, this should be investigated. Furthermore, Tsuji et al. (2005) do not specify how exactly these samples were kept – being kept at 22°C constantly implies that they were kept indoors. Without research into how exposure to the elements degrades telomeres, TLA can only be used on fresh samples.

There is a potential technique to adjust values of TMF to compensate for environmental degradation of telomeres, although it works on the assumption that DNA degrades at a constant rate across its whole length. A rough estimate of the degree of degradation might be made by considering the number of DNA profile markers that can be amplified by PCR. The quantity of markers degraded to the point

that they cannot be amplified might indicate the amount of degradation undergone by the telomeres. Therefore, the observed TMF could be adjusted accordingly, to allow TLA to be performed.

Research into the rate of telomere degradation compared to the degradation of DNA profile marker sequences would therefore be required, although it would allow the use of TLA on older samples. This is of particular importance if TLA is to be used to supplement forensic osteology – a forensic osteologist might be required to examine older bodies of which few identifying features remain, so it is not unreasonable to expect that DNA would be highly degraded.

### *5.3.5: TLA, adolescence and old age*

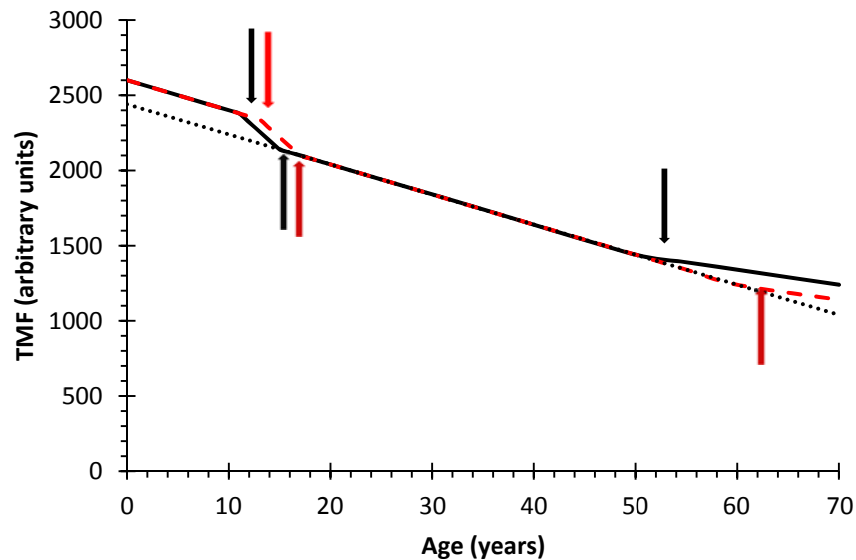
Research must be done into how the rate of telomere attrition varies with age. While TLA works on the principle that cell division in particular tissues is at an approximately constant rate, there are points in an individual's life when this rate changes, such as puberty and old age.

The age of onset of puberty varies with social, environmental and genetics, as does its duration, and neither can easily be predicted (Sapolsky, 1997). Because puberty is accompanied by a surge in the rate of an individual's growth, an increase in the rate of telomere attrition will also be seen.

During old age, the rate of cell division drops as more cells become senescent due to their telomeres dropping below the critical length. This reduction prevents cells being replaced in tissues and so causes many of the effects of old age (Brookbank, 1990; Jaskelioff et al., 2011). Because the rate of cell division falls, it would not be unreasonable to expect the rate of telomere attrition to also fall.

The rate of telomere attrition will therefore vary at particular points in an individual's life. The effect of this is illustrated in Graph 5.1.1, overleaf. While this graph is entirely hypothetical, it illustrates how TLA is only likely, at this point, to be accurate within the range of ages studied.

*Graph 5.1.1:* An illustration of the effect of puberty and old age on TMF. The expected rate of telomere attrition is indicated by the dashed line. One individual (represented by the black line) enters puberty at around 11 years old, and exits around the age of 15 (marked by the black arrows). The other individual (represented by the red dashed line) enters puberty at around 13 and exits at around 17 (marked by the red arrows). The first individual enters old age at around 50 (marked by the black arrow), while the second enters old age at around 60 (marked by the red arrow). TLA is currently only likely to be accurate between the exit from puberty and the entry into old age – extrapolating TLA into old age or childhood (indicated by the dotted line) from a constant rate of telomere degradation during adulthood is likely to be inaccurate.



Because of the terms of the ethical approval granted for this project, it was not possible to collect samples from children, or significantly older people. This omission must be addressed before TLA can be applied to children or elderly people.

## ***5.4: Conclusion***

---

In conclusion, Telomere Length Analysis is capable at this time of providing only limited information about an individual's appearance as part of the process of molecular photofitting. As a consequence of the cost of molecular photofitting, TLA is likely only to be used in extremely serious cases, those where little useful evidence has been recovered, or to provide information to aid the identification of a body.

While there is a substantial amount of research to be undertaken, TLA is not currently accurate enough in its predictions of age to be integrated into conventional DNA profiling. The database nature of TLA makes it a more accurate predictor of age than conventional comparison of telomere lengths, and it is likely that this accuracy will improve as the database expands. However, currently, TLA is not ready for use in criminal investigations.

## References & Bibliography

---

Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P (2002). *Molecular Biology of the Cell* 4<sup>th</sup> edition. Garland Science.

*Allan v The United Kingdom* [2002] 36 EHRR 12 (p143) (judgement of 2002).

Armanios M, Alder JK, Parry EM, Karim B, Strong MA and Greider CW (2009). Short telomeres are sufficient to cause the degenerative defects associated with aging. *American Journal of Human Genetics* doi 10.1016.ajhg.2009.10.028.

Ashworth A (2006). *Principles of Criminal Law* 5<sup>th</sup> edition. Oxford University Press.

Barnett K (2004). Marks and Impressions. In PC White (Ed), *Crime Scene to Court: The Essentials of Forensic Science* (pp82-114). Royal Society of Chemistry.

Black S (2000). Forensic Osteology in the United Kingdom. In M Cox and S Mays (Eds), *Human Osteology in Archaeology and Forensic Science* (pp491-503). Greenwich Medical Media.

Brookbank JW (1990). *The Biology of Aging*. Harper & Row.

Boulay JL, Reuter J, Ritschard R, Terracciano L, Herrmann R and Rochlitz C (1999). Gene dosage by quantitative real-time PCR. *BioTechniques* 27 pp 228-232.

Budowle B, Baechtel FS, Smerick JB, Presley KW, Giusti AM, Parsons G, Alevy MC and Chakraborty R (1995). D1S80 population data in African-Americans, Caucasians, South-eastern Hispanics, South-western Hispanics, and Orientals. *Journal of Forensic Sciences* 40 (1) pp38-44.

Bustin SA and Nolan T (2004a). Chemistries. In SA Bustin (Ed), *A-Z of Quantitative PCR* (pp215-278). International University Line.

Bustin SA and Nolan T (2004b). DNA Analysis and Interpretation. In SA Bustin (Ed), *A-Z of Quantitative PCR* (pp439-492). International University Line.

Butler JM (2001). *Forensic DNA Typing: Biology and Technology Behind STR Markers*. Academic Press.

Campbell NA (1996). *Biology* 4<sup>th</sup> edition. Benjamin/Cummings.

Cavalli-Sforza LL (2000). *Genes, People and Languages*. Penguin.

Cawthon RM (2002). Telomere measurement by quantitative PCR. *Nucleic Acids Research* 30 e47.

Chan SW and Blackburn EH (2002). New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin. *Oncogene* 21 pp 553-563.

Chang S, Multani AS, Cabrera NG, Naylor ML, Laud P, Lombard D, Pathak S, Guarente L and De Pinho RA (2004). Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nature Genetics* 36 pp877-882.

Cherkas LF, Hankin JL, Kato BS, Richards B, Gardner JP, Surdulescu GL, Kimura M, Lu X, Spector TD and Aviv A (2008). The association between physical activity in leisure time and leukocyte telomere lengths. *Archives of Internal Medicine* 168 (2) pp154-158.

Choo AL-T (2006). *Evidence*. Oxford University Press.

Cox M (2000). Aging Adults from the Skeleton. In M Cox and S Mays (Eds), *Human Osteology in Archaeology and Forensic Science* (pp61-81). Greenwich Medical Media.

Crime and Security Bill 2009-2010.

Davies N (1999). *The Isles: A History*. Macmillan.

Diamond J (1998). *Guns, Germs and Steel: A Short History of Everybody For The Last 13,000 Years*. Vintage.

Doward J (2009, September 20). Anger over DNA tests on asylum seekers. *The Observer* p18.

Farzaneh-Far R, Lin J, Epel ES, Harris WS, Blackburn EH and Whooley MA (2010). Association of marine omega-3 fatty acid levels with telomeric aging in patients with coronary heart disease. *Journal of the American Medical Association* 303 (3) pp250-257.

Foster C (2008). Untwining the strands. *New Law Journal* 158 p157.

Frudakis TN (2008). *Molecular Photofitting: Predicting Ancestry and Phenotype using DNA*. Elsevier.

Geddes L (2010). For justice, share DNA databases. *New Scientist* 2742 pp10-11.

Gisselsson D (2004). Chromosome instability in cancer: how, when and why? *Advances in Cancer Research* 87 pp1-29.

Goldacre B (2009). *Bad Science* 2<sup>nd</sup> edition. Fourth Estate.

Harley CB, Futcher AB and Greider CW (1990). Telomeres shorten during aging of human fibroblasts. *Nature* 345 pp458-460.

Henegariu O, Heerema NA, Dlouhy SR, Vance GH and Vogt PH (1997). Multiplex PCR: critical parameters and step-by-step protocol. *BioTechniques* 23 pp504-511.

Herrera RJ, Adrien LR, Ruiz LM, Sanabria NY and Duncan G (2004). D1S80 single-locus discrimination among African populations. *Human Biology* 76 (1) pp87-108.

Hoare J (2009). Extent and Trends. In A Walker, J Flatley, C Kershaw and D Moon (Ed), *Home Office Statistical Bulletin: Crimes in England and Wales 2008/09* (pp13-42). Crown Copyright.

Holmes D, Moody P and Dine D (2006). *Research Methods for the Biosciences*. Oxford University Press.

Hubbard K and Ozer HL (1999). Senescence and Immortalisation in Human Cells. In GP Studzinski (Ed), *Cell Growth, Differentiation and Senescence: A Practical Approach* (pp281-300). Oxford University Press.

Ivamy ERH (ed.) (1993). *Mozley and Whiteley's Law Dictionary* 11<sup>th</sup> edition. Butterworths.

Jaskelioff M, Muller FM, Paik J-H, Thomas E, Jiang S, Adams AC, Sahin E, Kost-Alimova M, Protopopov A, Cadiñanos J, Horner JW, Maratos-Flier E, DePinho RA (2011). Telomerase reactivation reverses tissue degeneration in aged telomerase-deficient mice. *Nature* 469 pp102-106.

Jones S (1996). *In The Blood: God, Genes And Destiny*. HarperCollins.

Karlsson AO, Svensson A, Marklund A and Holmlund G (2008). Estimating human age in forensic samples by analysis of telomere repeats. *Forensic Science International Genetics Supplement Series* 1 pp569-571.

Kasai K, Nakamura Y and White R (1991). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by Polymerase Chain Reaction (PCR) and its application to forensic science. *Journal of Forensic Sciences* 35(5) pp1196-1200.

Kill IR and Faragher RGA (2000). Senescence, Apoptosis and Necrosis. In JRW Masters (Ed), *Animal Cell Culture: A Practical Approach* 3<sup>rd</sup> edition (pp 281-302). Oxford University Press.

Krawczak M and Schmidtke J (1998). *DNA Fingerprinting* 2<sup>nd</sup> edition. BIOS Scientific Publishers.

Koopman P, Gubbay J, Vivian N, Goodfellow P and Lovell-Badge R (1991). Male development of chromosomally female mice transgenic for Sry. *Nature* 351 pp117-121.

Laville S (2009, November 16). Suspect arrested in hunt for 'night stalker' rapist. *The Guardian*, p5.

Le Fanu J (2010). The disappointments of the double helix: a master theory. *Journal of the Royal Society of Medicine* 103(2) pp43-45.

Levenson T (2009). *Newton And The Counterfeiter*. Faber and Faber.

Lewin B (2000). *Genes* 7<sup>th</sup> edition. Oxford University Press.

Liu F, van Duijn K, Vingerling J, Hofman A, Uitterlinden A, Janssens A and Kayser M (2009). Eye color and the prediction of complex phenotypes from genotypes. *Current Biology* 19 (5) pp192-193.

*Looseley; Attorney-General's Reference No. 3 of 1999* [2001] 4 All ER 897.

Lowry D (2008). *VassarStats Statistical Computation Website*. Retrieved from <http://faculty.vassar.edu/lowry/VassarStats.html>.

Lundblad V (2003). Telomeres: taking the measure. *Nature* 423 pp926-927.

Maher B (2008). The case of the missing heritability. *Nature* 456 pp18-21.

Man J (2004). *Genghis Khan: Life, Death and Resurrection*. Bantam Press.

McCartney C (2006). The DNA expansion program and criminal investigation. *British Journal of Criminology* 46 (2) pp175-192.

McCartney C (2008). When science doesn't meet the law. *New Law Journal* 158 p192.

Mitchell JR, Wood E and Collins K (1999). A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402 pp551-555.

Nakamura AJ, Chiang YJ, Hathcock KS, Horikawa I, Sedelnikova OA, Hodes RJ and Bonner WM (2008). Both telomeric and non-telomeric DNA damage are determinants of mammalian cellular senescence. *Epigenetics and Chromatin* 1:6 doi 10.1186/1756-8935-1-6.

National Health Service (2009, 20<sup>th</sup> November). *What is the Body Mass Index (BMI)?*  
Retrieved from  
<http://www.nhs.uk/chq/Pages/850.aspx?CategoryID=51&SubCategoryID=165>.

*Perry v The United Kingdom* [2002] (Admissibility), Application number 63737/00,  
26<sup>th</sup> September 2002.

Police and Criminal Evidence Act 1984.

Pollard TD and Earnshaw WC (2004). *Cell Biology*. Saunders.

*R v Bamber* [2002] All ER (D) 165 (Dec).

*R v Doheny* [1997] 1 Cr App R 369.

*R v Hoey* [2007] NICC 49.

*R v Stagg* [1994]. Unreported.

*R v Tran* [1991] Case No. 91/0162/89.

Richards JB, Valdes AM, Gardner JP, Paximades D, Kimura M, Nessa A, Lu X, Surdulescu GL, Swaminathan R, Spector TD and Aviv A (2007). Higher serum Vitamin D concentrations are associated with longer leukocyte telomere length in women. *American Journal of Clinical Nutrition* 86 (5) pp1420-1425.

Ridley M (2003). *Nature Via Nurture*. Harper Perennial.

Rodier F, Kim S, Nijjar T, Yaswen P and Campisi J (2004). Cancer and aging: The importance of telomeres in genome maintenance. *eScholarship Repository*, University of California LBNL-56548. Retrieved from  
<http://repositories.cdlib.org/lbnl-56548>.

*S and Another v The United Kingdom* [2008] All ER (D) 56 (Dec).

Sapolsky RM (1997). *Junk Food Monkeys And Other Essays On The Biology Of The Human Predicament*. Headline.

Scheur L and Black S (2000). Development and Aging of the Juvenile Skeleton. In M Cox and S Mays (Eds), *Human Osteology in Archaeology and Forensic Science* (pp9-21). Greenwich Medical Media.

Sexual Offences Act 2003.

Swinscow TDV and Campbell MJ (1997). *Statistics At Square One* 9<sup>th</sup> edition. BMJ Publishing Group.

Theft Act 1968.

Travis A (2008, December 5). 17 judges, one ruling – and 857,000 records must now be wiped clear. *The Guardian* pp6-7.

Travis A (2009a, October 22). Detections using DNA database fall despite huge rise in profiles. *The Guardian* p20.

Travis A (2009b, October 20). Plans to retain DNA records of innocent people dropped. *The Guardian* p12.

Travis A (2010a, January 6). DNA matches ‘solve few crimes’. *The Guardian* p9.

Travis A (2010b, March 17). Plan to limit police DNA storage may be shelved. *The Guardian* p2.

Tsuji A, Ishiko A, Takasaki T and Ikeda N (2002). Estimating age of humans based on telomere shortening. *Forensic Science International* 126 pp197-199.

Tsuji A, Ishiko A and Ikeda N (2005). Telomere shortening and age estimation in forensic medicine. *Gerontology* 51 p416.

Verbenko DA, Slominsky PA, Spitsyn VA, Bebyakova NA, Khusnutdinova EK, Mikulich AI, Tarskaia LA, Sorensen MV, Ivanov IP, Bets LV and Limborska SA (2006). Polymorphisms at locus D1S80 and other hypervariable regions in the analysis of Eastern European ethnic group relationships. *Annals of Human Biology* 33 (5-6) pp 570-584.

Walsh S, Metzger D and Higuchi R (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10(4) pp506-513.

Walsh SJ and Eckhoff C (2007). Australian Aboriginal population genetics at the D1S80 VNTR locus. *Annals of Human Biology* 34 (5) pp557-565.

Ward R and Wragg A (2005). *Walker & Walker’s English Legal System* 9<sup>th</sup> edition. Oxford University Press.

Watson N (2004). The Analysis of Body Fluids. In PC White (Ed), *Crime Scene to Court: The Essentials of Forensic Science* (pp377-413). Royal Society of Chemistry.

Wolffe A (1998). *Chromatin: Structure and Function* 3<sup>rd</sup> edition. Academic Press.

## ***Appendices***

---

A1	Sample Phase One Participant Information Sheet & Informed Consent Form	120
A2	Phase One Basic Participant Data	124
A3	Phase One TMF Calculation	126
A4	Blind Test TMF Calculation	128
A5	Sample Phase Two Participant Information Sheet & Informed Consent Form	129
A6	Phase Two Basic Participant Data	133
A7	<sub>B</sub> TMF Calculation	134
A8	<sub>H</sub> TMF Calculation	135
A9	<sub>L</sub> TMF Calculation	136

# ***A1: Sample Phase One Participant Information Sheet & Informed Consent Form***

---



University of  
Chester

Tracking Number: \_\_\_\_\_

## **University of Chester**

### **Department of Biological Sciences**

#### **Telomere Length Analysis: a new forensic technique**

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please, take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear, or if you would like more information. Take time to decide whether or not you wish to take part.

*Thank you*

#### **What is the purpose of this study?**

DNA profiling has been a powerful tool in solving crimes. However, it is unable to help the police when a sample taken from the scene of a crime does not match any individual stored in the National DNA Database. This study is investigating what other information can be obtained from DNA left at the scene of a crime. Specifically, we are looking at parts of the DNA called *telomeres*. These are found at the ends of every chromosome, and they help protect the rest of the chromosome from damage. But, they wear away over time, whenever a cell divides. By looking at the size of an individual's telomeres, we hope to be able to predict that person's age, height and weight, because these are associated with cell division.

#### **Why have I been chosen?**

You have been chosen because you are a member of staff or a student at the University of Chester. We need to take a small amount of DNA, which is most conveniently done on campus. We also need to ask you your height and weight – we can measure these easily on campus if you're not sure.

### **Do I have to take part?**

*It is up to you to decide whether or not to take part. If you decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time without giving a reason. Simply contact us at the email address below.*

### **What will happen to me if I take part?**

If you decide to take part, you will be given this information sheet to keep and asked to sign the consent form. This will give your consent for a researcher from the Department of Biological Sciences at the University of Chester to take a small sample of DNA from the inside of your mouth using a cheek swab – these are sterile. You will be asked your height and weight. Don't worry if you're not sure about these – we can easily measure them.

We may also ask you if we can contact you again, to give us a small sample of blood and hair. If we obtain good results from cheek swabs, we would like to repeat the experiment on blood and hair, which are often left at the scene of a crime. If you are willing to be contacted again, you can tick the box on the consent form. Agreeing to be contacted again does not mean that you *have* to give blood or hair – you can change your mind whenever you like.

### **What are the possible disadvantages and risks of taking part?**

There are no disadvantages or risks foreseen in taking part in this study.

### **What are the possible benefits of taking part?**

Unfortunately, although your contribution will be greatly valued, you will not personally benefit from taking part.

### **What if something goes wrong?**

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, please contact *[Dean of the School of Applied and Health Sciences' contact details]*.

. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed through someone's negligence (but not otherwise), you may have grounds for legal action. You may have to pay for this.

### **Will my taking part in the study be kept confidential?**

Throughout this study, you will be referred to by your tracking number. Only the researcher carrying out the research will have access to information that could be used to connect your tracking number to you.

**What will happen to the results of the research study?**

The results will be written up into a report for the University. They may also be published in an academic journal as a paper. Individuals who participate will not be identifiable in any report or publication.

**Who is organising and funding the research?**

This study is funded and organised by the University of Chester.

**Who may I contact for further information?**

If you would like more information about the research, at any time, please contact:

Steve Cargill (Lead Researcher): *[researcher's email address]*

Dr I McDowall (Supervisor): *[supervisor's email address]*

**Thank you for your time and interest in this study**



Tracking Number: \_\_\_\_\_

### University of Chester

### Department of Biological Sciences

### Telomere Length Analysis: a new forensic technique

### Informed Consent Form

Name of Researcher: Steve Cargill

Please initial box

1. I confirm that I have read and understood the participant information sheet, dated ....., for the above study and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

3. I agree to take part in the above study.

4. I am/am not [delete as applicable] willing to be approached at a later point to give additional samples of blood and hair. If I am willing, I would prefer to be contacted by phone/email [delete as applicable] on \_\_\_\_\_

Date of Birth: \_\_\_\_\_.

Height: \_\_\_\_\_ [feet and inches/metres].

Weight: \_\_\_\_\_ [stones and pounds/kilograms].

_____	_____	_____
Name of Participant	Date	Signature

_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature

_____	_____	_____
Name of Researcher	Date	Signature

## ***A2: Phase One Basic Participant Data***

<b>Participant</b>	<b>Age (years)</b>	<b>Height (m)</b>	<b>Mass (kg)</b>	<b>BMI</b>	<b>Gender</b>
1	34.5	1.8	108	33.2	M
2	44.333	1.57	63	25.5	F
3	28.917	1.72	127	42.6	F
4	30.667	1.57	51	20.5	F
5	42.417	1.6	117	45.7	M
6	23.25	1.69	91	31.8	F
7	23.5	1.52	58	24.8	F
8	22.5	1.83	69	20.6	M
9	21.917	1.73	71	23.9	F
10	34.167	1.63	70	26.4	F
11	42.75	1.75	114	37.1	M
12	18.25	1.7	76	26.3	M
13	18.667	1.73	102	34.1	M
14	19.667	1.78	100	31.6	M
15	26.167	1.7	56	19.3	F
16	57.333	1.57	69	28	F
17	20.417	1.66	57	20.7	F
18	20.583	1.8	86	26.6	M
19	44.25	1.6	56	21.8	F
20	27.917	1.83	95	28.5	M
21	43.667	1.65	53	19.5	F
22	38.167	1.7	90	31.2	F
23	48.083	1.8	85	26.2	M
24	59.667	1.83	107	31.9	M
25	48	1.78	98	31.1	M
26	51.667	1.83	75	22.5	M
27	55	1.63	60	22.7	F
28	22.417	1.55	73	30.6	F
29	23.167	1.65	65	23.9	F
30	19.417	1.6	59	23	F
31	19.833	1.65	56	20.6	F
32	21.75	1.78	65	20.5	M
33	23	1.7	89	30.7	F
34	30.167	1.59	99	39.4	F
35	30.333	1.65	66	24.1	F
36	36.083	1.73	71	23.7	F
37	19.667	1.83	89	26.7	M
38	22.167	1.65	63	23.1	F
39	19.667	1.75	65	21.2	F

Continued overleaf.

<b>Participant</b>	<b>Age (years)</b>	<b>Height (m)</b>	<b>Mass (kg)</b>	<b>BMI</b>	<b>Gender</b>
40	19.75	1.67	71	25.8	F
41	22.167	1.63	83	31.5	F
42	21.25	1.65	73	26.8	F
43	40.833	1.63	68	25.9	F
44	42.333	1.63	51	19.2	F
45	30.167	1.73	96	32.2	M
46	30.083	1.63	138	52.4	F
47	39.917	1.65	69	25.3	F
48	27	1.8	76	23.5	F
49	65.5	1.75	89	29	M
50	64.917	1.52	66	28.4	F
51	35.417	1.88	105	29.7	M
52	42.917	1.83	121	36.1	M
53	52	1.63	70	26.4	F
54	36.917	1.63	61	23.2	F
55	55	1.55	57	23.8	F
56	40.5	1.76	110	35.5	M
57	18.333	1.68	84	30	M
59	46.167	1.75	57	18.6	M
60	68.75	1.52	64	27.3	F
61	52.25	1.8	83	25.4	M
62	44.917	1.75	94	30.7	M
63	57.167	1.75	75	24.5	M
64	45.333	1.88	75	21.2	M
65	68.167	1.75	76	24.8	M
66	50.5	1.91	90	24.7	M
67	49	1.78	70	22.1	M
68	46	1.83	73	21.1	M
69	70	1.75	89	28.9	M
70	52.333	1.78	127	40.2	M
71	56.167	1.61	64	24.4	F
72	38.667	1.83	85	25.5	M
73	61.333	1.55	65	27.2	F
74	46.667	1.75	73	23.8	F
75	61.75	1.74	83	27.3	M
76	36.667	1.72	70	23.4	F
77	48.333	1.72	89	29.8	M
78	39.917	1.64	58	21.6	F
79	37.583	1.7	62	21.4	M
80	57.5	1.68	73	25.8	M
81	51	1.8	76	23.4	M

### **A3: Phase One TMF Calculation**

<b>Participant</b>	<b>I<sub>T</sub>1</b>	<b>I<sub>T</sub>2</b>	<b>Mean I<sub>T</sub></b>	<b>I<sub>S</sub>1</b>	<b>I<sub>S</sub>2</b>	<b>Mean I<sub>S</sub></b>	<b>TMF</b>
1	3318	3352	3335	3667	3654	3660.5	632
2	3516	3550	3533	1830	1879	1854.5	1321
3	3686	3618	3652	2489	2521	2505	1011
4	3752	3722	3737	3102	3168	3135	827
5	3685	3677	3681	1102	1084	1093	2336
6	3692	3758	3725	3757	3539	3648	708
7	3760	3680	3720	1324	1358	1341	1924
8	3760	3712	3736	1684	1768	1726	1501
9	2340	2284	2312	1170	1176	1173	1367
10	3385	3442	3413.5	2119	2209	2164	1094
11	3720	3760	3740	3760	3756	3758	690
12	3746	3758	3752	3725	3725	3725	699
13	3653	3750	3701.5	3570	3760	3665	701
14	3370	3416	3393	1585	1586	1585.5	1484
15	3300	3243	3271.5	2003	1935	1969	1152
16	3751	3728	3739.5	2668	2587	2627.5	987
17	2207	2207	2207	1184	1197	1190.5	1286
18	3107	3046	3076.5	3616	3568	3592	594
19	3597	3577	3587	1211	1197	1204	2066
20	2998	2910	2954	1579	1548	1563.5	1310
21	3496	3524	3510	3637	3687	3662	665
22	3569	3541	3555	2850	2939	2894.5	852
23	3664	3662	3663	3760	3750	3755	677
24	3531	3613	3572	3758	3744	3751	661
25	3616	3652	3634	3739	3714	3726.5	676
26	3746	3740	3743	3760	3760	3760	690
27	3709	3760	3734.5	2772	2699	2735.5	945
28	3624	3683	3653.5	2100	2055	2077.5	1220
29	3654	3710	3682	2216	2223	2219.5	1151
30	3549	3650	3599.5	3690	3731	3710.5	673
31	3754	3754	3754	3760	3758	3759	693
32	3748	3760	3754	3749	3760	3754.5	694
33	2002	2069	2035.5	2503	2448	2475.5	570
34	3499	3499	3499	2748	2718	2733	888
35	2387	2382	2384.5	2665	2762	2713.5	610
36	3298	3272	3285	3509	3583	3546	643
37	2899	2797	2848	3754	3694	3724	530
38	2944	2944	2944	2790	2732	2761	740
39	3760	3760	3760	2940	2852	2896	901

Continued overleaf.

<b>Participant</b>	<b>I<sub>r1</sub></b>	<b>I<sub>r2</sub></b>	<b>Mean I<sub>r</sub></b>	<b>I<sub>s1</sub></b>	<b>I<sub>s2</sub></b>	<b>Mean I<sub>s</sub></b>	<b>TMF</b>
40	2639	2639	2639	2480	2409	2444.5	749
41	1450	1464	1457	3686	3703	3694.5	274
42	1846	1753	1799.5	3760	3757	3758.5	332
43	1830	1817	1823.5	3724	3630	3677	344
44	1663	1607	1635	3760	3759	3759.5	302
45	1482	1382	1432	3488	3488	3488	285
46	1629	1659	1644	3744	3760	3752	304
47	1687	1634	1660.5	3480	3480	3480	331
48	1413	1497	1455	3691	3750	3720.5	271
49	2576	2602	2589	3513	3598	3555.5	505
50	2702	2719	2710.5	3624	3651	3637.5	517
51	2645	2554	2599.5	2098	2085	2091.5	862
52	2478	2554	2516	3584	3557	3570.5	489
53	2963	2891	2927	3752	3733	3742.5	542
54	2733	2690	2711.5	1806	1866	1836	1024
55	2968	2891	2929.5	3654	3653	3653.5	556
56	1995	2028	2011.5	1995	2028	2011.5	694
57	1283	1331	1307	2539	2539	2539	357
59	1345	1389	1367	2712	2742	2727	348
60	1410	1312	1361	2224	2187	2205.5	428
61	1360	1394	1377	1549	1605	1577	606
62	1223	1241	1232	2270	2270	2270	376
63	1285	1266	1275.5	3409	3470	3439.5	257
64	1316	1395	1355.5	3150	3173	3161.5	297
65	2140	2141	2140.5	3461	3411	3436	432
66	2504	2437	2470.5	3751	3760	3755.5	456
67	2123	2232	2177.5	3591	3518	3554.5	423
68	2168	2187	2177.5	3568	3537	3552.5	425
69	2480	2534	2507	3667	3726	3696.5	470
70	2258	2290	2274	1641	1608	1624.5	971
71	3379	3335	3357	3735	3640	3687.5	631
72	3555	3609	3582	3598	3639	3618.5	687
73	3092	3160	3126	2672	3644	3158	687
74	2631	2588	2609.5	3759	3736	3747.5	483
75	2816	2861	2838.5	3755	3730	3742.5	526
76	2568	2568	2568	3594	3689	3641.5	489
77	2705	2655	2680	3474	3549	3511.5	529
78	3349	3250	3299.5	3539	3539	3539	647
79	2640	2571	2605.5	3648	3665	3656.5	494
80	2793	2805	2799	3464	3434	3449	563
81	2015	1974	1994.5	1109	1109	1109	1247

#### ***A4: Blind Test TMF Calculation***

<b>Participant</b>	<b>Age (years)</b>	<b>Gender</b>	<b>I<sub>T</sub>1</b>	<b>I<sub>T</sub>2</b>	<b>Mean I<sub>T</sub></b>	<b>I<sub>S</sub>1</b>	<b>I<sub>S</sub>2</b>	<b>Mean I<sub>S</sub></b>	<b>TMF</b>
T1	21.917	F	2668	2759	2713.5	3343	3431	3387	556
T2	35.833	M	2642	2701	2671.5	3649	3649	3649	508
T3	25.667	F	2774	2752	2763	3130	3222	3176	603
T4	41.167	F	2590	2637	2613.5	3249	3287	3268	555
T5	36.917	F	2761	2761	2761	3257	3268	3262.5	589
T6	54.5	M	2767	2750	2758.5	3147	3051	3099	617
T7	55.583	F	2609	2521	2565	3057	3105	3081	577
T8	37.417	M	2152	2227	2189.5	3334	3346	3340	455
T9	34.25	M	2674	2630	2652	3522	3449	3485.5	528

## ***A5: Sample Phase Two Participant Information Sheet & Informed Consent Form***

---



University of  
Chester

Tracking Number: \_\_\_\_\_

### **University of Chester**

#### **Department of Biological Sciences**

##### **Telomere Length Analysis: a new forensic technique**

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please, take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear, or if you would like more information. Take time to decide whether or not you wish to take part.

*Thank you*

##### **What is the purpose of this study?**

DNA profiling has been a powerful tool in solving crimes. However, it is unable to help the police when a sample taken from the scene of a crime does not match any individual stored in the National DNA Database. This study is investigating what other information can be obtained from DNA left at the scene of a crime. Specifically, we are looking at parts of the DNA called *telomeres*. These are found at the ends of every chromosome, and they help protect the rest of the chromosome from damage. But, they wear away over time, whenever a cell divides. By looking at the size of an individual's telomeres, we hope to be able to predict that person's age, height and weight, because these are associated with cell division.

##### **Why have I been chosen?**

You have been chosen because you've previously participated in the early stages of this project, and agreed to allow us to contact you again to see if you'd be interested in giving us a small sample of blood and hair.

### **Do I have to take part?**

*It is up to you to decide whether or not to take part.* If you decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. *If you decide to take part, you are still free to withdraw at any time without giving a reason.* Simply contact us at the email address below. The fact that you've agreed to allow us to contact you does not mean that you're obliged to consent now.

### **What will happen to me if I take part?**

If you decide to take part, you will be given this information sheet to keep and asked to sign the consent form. This will give your consent for a researcher from the Department of Biological Sciences at the University of Chester to take a small sample of hair – two or three hairs with the roots still attached. We will also ask you to use a sterile lancet to prick your own finger and collect a small amount of blood – we only need a drop or two. We ask you to do this yourself to reduce the risk of cross-infection to you. Throughout the whole process, every effort to minimise the risk to you has been taken

### **What are the possible disadvantages and risks of taking part?**

You will have a sore finger for an hour or so – if you're a blood donor, you'll be familiar with the finger-prick test for measuring your haemoglobin levels. Plucking the hairs from your head will also sting. By asking you to draw and collect your own blood, we aim to keep the risk of you picking anything up from us to the absolute minimum.

### **What are the possible benefits of taking part?**

Unfortunately, although your contribution will be greatly valued, you will not personally benefit from taking part.

### **What if something goes wrong?**

If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, please contact [*Dean of the School of Applied and Health Sciences' contact details*].

If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed through someone's negligence (but not otherwise), you may have grounds for legal action. You may have to pay for this.

**Will my taking part in the study be kept confidential?**

Throughout this study, you will be referred to by your tracking number. Only the researcher carrying out the research will have access to information that could be used to connect your tracking number to you.

**What will happen to the results of the research study?**

The results will be written up into a report for the University. They may also be published in an academic journal as a paper. Individuals who participate will not be identifiable in any report or publication.

**Who is organising and funding the research?**

This study is funded and organised by the University of Chester.

**Who may I contact for further information?**

If you would like more information about the research, at any time, please contact:

Steve Cargill (Lead Researcher): *[researcher's email address]*

Dr I McDowall (Supervisor): *[supervisor's email address]*

**Thank you for your time and interest in this study**



Tracking Number: \_\_\_\_\_

**University of Chester**

**Department of Biological Sciences**

**Telomere Length Analysis: a new forensic technique**

**Informed Consent Form**

**Name of Researcher: Steve Cargill**

Please initial box

- 1. I confirm that I have read and understood the participant information sheet, dated ....., for the above study and have had the opportunity to ask questions.
- 2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
- 3. I agree to take part in the above study.

Date of Birth: \_\_\_\_\_.

\_\_\_\_\_  
Name of Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent  
(if different from researcher)

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Researcher

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

### ***A6: Phase Two Basic Participant Data***

<b>Participant</b>	<b>Age (years)</b>	<b>Height (m)</b>	<b>Mass (kg)</b>	<b>BMI</b>	<b>Gender</b>
2.1	54.5	1.60	56	21.8	Female
2.2	28.25	1.83	95	28.5	Male
2.3	24.583	1.52	59	25.4	Female
2.4	23.417	1.69	91	31.8	Female
2.5	48.417	1.80	85	26.2	Male
2.6	30.417	1.73	96	32.2	Male
2.7	34.417	1.63	70	26.4	Female
2.8	43	1.75	114	37.1	Male
2.9	26.5	1.70	56	19.3	Female
2.10	57.667	1.57	65	26.2	Female
2.11	28.5	1.73	127	42.6	Female
2.12	53.333	1.64	89	33.1	Female
2.13	26.667	1.64	62	23.1	Female
2.14	20.917	1.80	77	23.8	Male

## **A7: $B$ TMF Calculation**

<b>Participant</b>	<b><math>I_T1</math></b>	<b><math>I_T2</math></b>	<b>Mean <math>I_T</math></b>	<b><math>I_S1</math></b>	<b><math>I_S2</math></b>	<b>Mean <math>I_S</math></b>	<b>TMF</b>
2.1	3056	3092	3074	3111	3052	3081.5	692
2.2	3151	3197	3174	2903	2844	2873.5	766
2.3	3184	3273	3228.5	2829	2829	2829	792
2.4	3136	3169	3152.5	3412	3385	3398.5	643
2.5	3571	2571	3071	2806	2811	2808.5	758
2.6	2595	2595	2595	3681	3690	3685.5	488
2.7	2594	2587	2590.5	2943	2975	2959	607
2.8	2430	2454	2442	2424	2424	2424	699
2.9	2958	2958	2958	2399	2453	2426	846
2.10	2683	2683	2683	2846	2943	2894.5	643
2.11	2746	2802	2774	2254	2266	2260	851
2.12	2734	2669	2701.5	3145	3145	3145	596
2.13	3026	3045	3035.5	3548	3548	3548	593
2.14	3364	3364	3364	3107	3130	3118.5	748

## **A8: $f$ TMF Calculation**

<b>Participant</b>	<b><math>I_T1</math></b>	<b><math>I_T2</math></b>	<b>Mean <math>I_T</math></b>	<b><math>I_S1</math></b>	<b><math>I_S2</math></b>	<b>Mean <math>I_S</math></b>	<b>TMF</b>
2.1	2593	2602	2597.5	2867	2827	2847	633
2.2	2873	2937	2905	3291	3341	3316	608
2.3	2768	2778	2773	3176	3231	3203.5	600
2.4	3150	3107	3128.5	2182	2213	2197.5	987
2.5	2378	2345	2361.5	3140	3169	3154.5	519
2.6	2803	2741	2772	3760	3686	3723	516
2.7	3167	3115	3141	2729	2729	2729	798
2.8	2084	3084	2584	2597	2692	2644.5	678
2.9	3359	3458	3408.5	3026	2968	2997	789
2.10	3240	3258	3249	2989	2967	2978	757
2.11	3003	3037	3020	2973	3049	3011	696
2.12	2932	2953	2942.5	3029	3029	3029	674
2.13	2817	2838	2827.5	2798	2880	2839	691
2.14	2758	2835	2796.5	3424	3404	3414	568

### **A9: $I_T$ TMF Calculation**

<b>Participant</b>	<b><math>I_{T1}</math></b>	<b><math>I_{T2}</math></b>	<b>Mean <math>I_T</math></b>	<b><math>I_{S1}</math></b>	<b><math>I_{S2}</math></b>	<b>Mean <math>I_S</math></b>	<b>TMF</b>
2.1	1609	1577	1593	1577	1577	1577	701
2.2	694	726	710	679	720	699.5	704
2.3	368	408	388	573	635	604	446
2.4	938	938	938	1185	1097	1141	570
2.5	622	687	654.5	325	367	346	1312
2.6	716	717	716.5	1256	1257	1256.5	396
2.7	411	354	382.5	464	378	421	630
2.8	1322	1302	1312	1074	1170	1122	811
2.9	904	981	942.5	925	857	891	734
2.10	1182	1173	1177.5	1133	1130	1131.5	722
2.11	1384	1390	1387	996	1040	1018	945
2.12	1387	1288	1337.5	1085	1090	1087.5	853
2.13	1010	1105	1057.5	1115	1032	1073.5	683
2.14	954	1027	990.5	1028	1118	1073	640