International Journal of Methodology

Volume 1, Issue 1, pp. 44-57, 2022
DOI: https://doi.org/10.21467/ijm.1.1.5784

Impact of DNA Extraction Methods on Quantitative PCR Telomere Length Assay Precision in Human Saliva Samples

Dana L. Smith[1*], Calvin Wu[1], Steve Gregorich[2], Guorui Dai[1] Jue Lin[1*]

[1] Department of Biochemistry and Biophysics, University of California, San Francisco
[2] Department of Medicine, University of California, San Francisco

Corresponding authors email: dana.smith@ucsf.edu, jue.lin@ucsf.edu
Received: 22 July 2022 / Revised: 14 December 2022 / Accepted: 16 December 2022 / Published: 27 December 2022

ABSTRACT

Telomere length (TL) has emerged as a promising replicative cellular aging marker that reflects both genetic and non-genetic influences. Quantitative PCR (qPCR) TL measurement has been favored as a cost-effective method that can be easily implemented, especially in population studies with limited quantities of source material. However, several recent reports have revealed inconsistencies in telomere length measurements when applying different DNA extraction methods to the same source material. In this study we tested three DNA extraction methods on saliva samples from 48 participants of the National Growth and Health Study (NGHS) collected with DNA Genotek's Oragene kit. The chosen extraction kits represent three distinct approaches to genomic DNA extraction from lysed cells and we employed two different operators to carry out all assays on the same samples. We measured DNA yield and quality and calculated the between-operator agreement of qPCR TL measurements (intraclass correlation, ICC). Our analyses showed that while both QIAamp and Agencourt DNAdvance had higher agreement between the 2 operators (ICC=0.937, CI [0.891, 0.965] and ICC=0.95, CI [0.911, 0.972] respectively), compared to PrepIT kit (ICC=0.809, CI [0.678, 0.889]), QIAamp extracted DNA samples were notably degraded. Using generalizability theory, we found that the participant-by-extraction-method interaction explained about 10% of total variation in TL, suggesting that TL differences across methods are somewhat participant-specific. Therefore, our results suggest that the among the three DNA extraction methods tested, Agencourt (magnetic bead purification) is the preferred kit, and we also strongly recommend against combining different extraction methods within a study population.

Keywords: qPCR telomere length measurement, DNA extraction, Assay precision

1 Introduction

Telomeres, the protective complexes at the ends of eukaryotic chromosomes, are made up of short tandem DNA repeats and their associated proteins (Blackburn, 2001). Telomeres naturally shorten with age through successive rounds of DNA replication, which leads to cell cycle arrest and senescence (Chan & Blackburn, 2002) when telomeres are critically short. This process can be accelerated due to reactive oxidative stress (Ahmed & Lingner, 2018; Fouquerel et.al., 2019) or inflammation (Jose et.al., 2017).

Telomere length (TL) has emerged as a useful biomarker of cellular aging and associated diseases. TL variation over time is shaped by both genetic and non-genetic influences throughout life (Broer et al., 2013; Factor-Litvak et al., 2017; Lin & Epel, 2022; Shalev et al., 2013). Therefore, TL values reflect both inheritance and the cumulative effects of environmental exposures and lived experience, allowing them to serve as a risk factor for disease susceptibility (Aviv & Shay, 2018; Blackburn et al., 2015; Cheng et al., 2021; de Meyer et al., 2018; Haycock, 2017; Ridout et al., 2017; Ruiz et al., 2021; Schneider et al., 2022; Wang et al., 2021; Willis et al., 2018).

As molecular biology has advanced, three main approaches have been developed to measure TL (Lai et.al., 2018; Lin et.al., 2019; Lindrose et.al., 2021). Traditionally, telomeres have been measured by telomere restriction fragment (TRF) length analysis, a type of Southern blot, in which genomic DNA is digested by restriction enzymes and incubated with a telomere-specific probe. The size of the terminal telomere-specific fragment reflects the sum of telomere length as well as some subtelomeric DNA. Fluorescent *in situ* hybridization (FISH) is another technique developed to quantify telomere repeats in individual cells or chromosomes, after flow cytometry of cell populations (Flow-FISH) (Baerlocher et.al., 2006) or in metaphase cells (Q-FISH) (Poon & Lansdorp, 2001). A third approach emerged with quantitative PCR (qPCR) technology, making it possible to measure telomere length as amplification of telomere repeats relative to amplification of a single copy gene (Cawthon, 2002, 2009). This represents another way to measure bulk telomere length, starting with a small DNA sample (Figure 1).

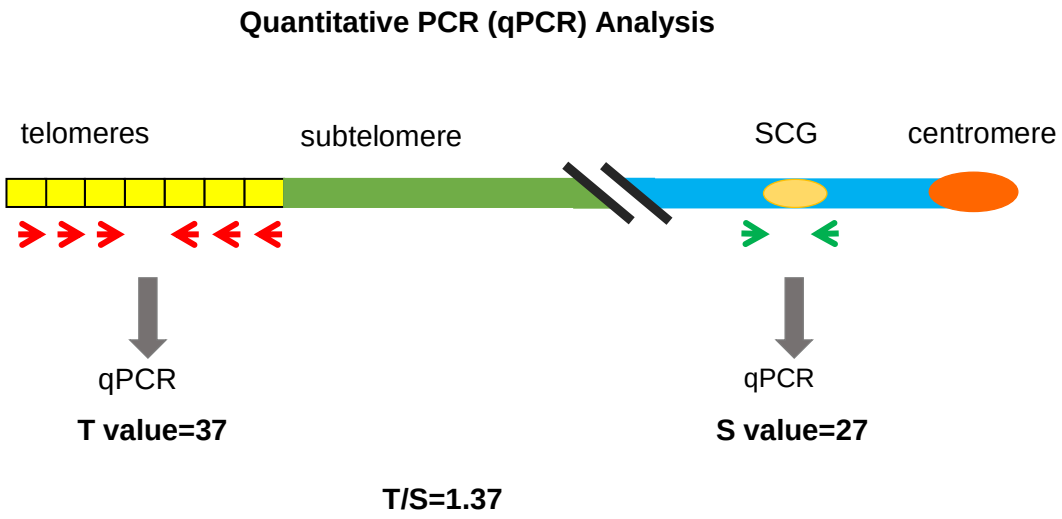


Figure 1: *qPCR TL measurement* (Cawthon, 2002) *involves PCR amplification and quantification of telomeric DNA sequence (T) and a Single Copy Gene DNA sequence (S) with appropriate primers. The Single Copy Gene product serves as an internal control and the ratio T/S reflects relative telomere length.*

Telomere length measurements, like all bioassays, are prone to measurement variation. Telomere length varies widely from person to person and can change within individuals. The environmental effects that influence TL can be small compared to the measurement variation, making it necessary to sample many individuals in order to make powerful comparisons. The epidemiological studies that link TL to disease states and risk factors are founded on TL measurements made from population studies, and in some cases, from multiple timepoints gathered over extended periods of time. The TRF and Flow-FISH methods pose logistical challenges for large epidemiological and population studies. TRF analysis requires large quantities of high-molecular weight genomic DNA and Flow-FISH requires immediate processing of freshly collected blood, which is not compatible for analysis of archived samples. qPCR is relatively low-cost and can be carried out with small quantities of DNA, which is obtainable from most archived samples. In this regard, qPCR TL measurement can facilitate large epidemiological and population-based studies that typically require hundreds or thousands of samples. The large-sample requirement has motivated developing the most reliable, standardized qPCR TL measurement techniques possible.

While qPCR assays are straightforward and relatively inexpensive, many factors can contribute to variability in qPCR measurements. These include tissue type used for DNA extraction, the method of sample collection, sample storage conditions, DNA extraction technique, variation in assay procedures and between-operator variation at each step. There is a need to systematically study how each of these factors

impacts the qPCR TL assay, and where variation might introduce inconsistencies in final TL measurement. (Lin et.al., 2019; Lindrose et.al., 2021).

Toward this goal, this manuscript focuses on qPCR-based TL measurements of DNA acquired from saliva samples collected passively, using DNA Genotek's Oragene kit (DNA Genotek). Saliva presents an attractive sample source for a number of reasons. Relative to blood, saliva collection is less invasive and can be gathered passively or collected with swabs, making it ideal for children or animal subjects. Saliva collection poses no risk of blood-borne pathogen exposure and allows collection from participants living in relatively remote places, where venous blood draw is less feasible. Travel restrictions imposed during the COVID pandemic highlight the advantages of home collection, with participants shipping the collected saliva directly to research sites. Once on site, saliva samples are easier and less expensive to store, usually at room temperature.

2 Methods

We started with saliva samples from 48 participants and estimated the reliability of qPCR TL data assayed from DNA extracted using three different extraction kits, by two different operators. Our extraction kit choices represent three distinct approaches to genomic DNA extraction from lysed cells: precipitation by salt exclusion (PrepIT), binding to and elution from a silica membrane column (QIAamp), or binding to and elution from magnetic beads (Agencourt). We employed 2 different laboratory technicians (operators) to complete all three extraction methods on the same samples. We measured DNA yield, DNA quality, and estimated reliability of qPCR telomere length measurements (TL). Reliability estimates included intraclass correlations (ICC) as well as two agreement coefficients from generalizability theory. Intraclass correlation (ICC) measures the reproducibility of quantitative measurements and is a more appropriate method to assess TL assay precision (Eisenberg, 2016; Verhulst et.al., 2015). Estimated intraclass correlations (ICC) of qPCR TL included within-operator ICCs (across two runs per DNA sample) stratified by DNA extraction method, between-operator ICCs stratified by DNA extraction method, and between-extraction-method ICCs. Generalizability theory models decomposed sources of TL variation and estimated coefficients of *(i) relative agreement*, i.e., the degree to which two operators or two DNA extraction methods agreed on the rank ordering of participants with respect to TL; and *(ii) absolute agreement*, i.e., the degree to which two operators agreed with respect to TL in an absolute sense.

2.1 Sample collection, DNA extraction from Human Saliva, and TL Length measurement

DNA purification of saliva lysates was carried out, using three different kits representing three unique extraction techniques: PrepIT, through salting-out; QIAamp, using silicon spin columns; and Agencourt, using magnetic bead preps (Figure S1). The saliva samples for this study were originally collected for the National Growth and Health Study (NGHS, nghstoday.com, PIs Barbara Laraia and Elissa Epel). Part of a large set, the first 48 samples with at least 3 ml of saliva left were used for further analysis in this study. The samples from this subset (Table S1) were collected between November 2015-April 2016, using DNA Genotek Oragene tubes (cat # ORG-500) and stored at room temperature, in the supplied collection kit lysing solution until extraction in late 2019. Just after purification, DNA quantity of each sample was measured through two or three independent approaches: through photometry (using the UV-Vis Nanodrop function), fluorescence (using Picogreen), and amplification and measurement of the human genomic, single-copy RNAse P gene (Agencourt preps only). The quality of each DNA sample was inferred, using the UV-Vis Nanodrop function and samples with OD260/OD280 >1.7 and < 2.0 and OD260/OD230 >1.0 were judged to be relatively free of contaminants. DNA quality was further characterized by visual inspection after running samples normalized for DNA concentration on standard 0.8% agarose gels. Samples were independently judged to be intact or 'degraded' by both operators. DNA preparations were stored at -80˚C until telomere length measurements were made, within 2 weeks of extraction.

DOI: 10.2147/ijm
Available online at Journals.aijr.org

2.2 qPCR (TL) Assay

The telomere length measurement assay (qPCR, TL) was adapted from the original method published by Cawthon (2002) and represents a ratio of two qPCR reactions: **T**elomere over **S**ingle copy gene (T/S, Figure 1). Monoplex qPCR assays were carried out on a Roche 480 Lightcycler PCR machine. The telomere thermal cycling profile consisted of:

Cycling for T(telomere) PCR: Denature at 96°C for 1 minute, one cycle; denature at 96°C for 1 second, anneal/extend at 54°C for 60 seconds, with fluorescence data collection, 30 cycles.

Cycling for S (single copy gene) PCR: Denature at 96°C for 1 minute, one cycle; denature at 95°C for 15 seconds, anneal at 58°C for 1 second, extend at 72°C for 20 seconds, 8 cycles; followed by denature at 96°C for 1 second, anneal at 58°C for 1 second, extend at 72°C for 20 seconds, hold at 83°C for 5 seconds with data collection, 35 cycles.

The primers for the telomere PCR are *tel1b* [5'-CGGTTT(GTTTGG)$_5$GTT-3'], used at a final concentration of 100 nM, and *tel2b* [5'-GGCTTG(CCTTAC)$_5$CCT-3'], used at a final concentration of 900 nM. The primers for the single-copy gene (human beta-globin) PCR are *hbg1* [5'-GCTTCTGACACAACTGTGTTCACTAGC-3'], used at a final concentration of 300 nM, and *hbg2* [5'-CACCAACTTCATCCACGTTCACC-3'], used at a final concentration of 700 nM.

House-made master mixes were utilized, and the final reaction mix contained 20 mM Tris-HCl, pH 8.4; 50 mM KCl; 200 mM each dNTP; 1% DMSO; 0.4x Syber Green I; 22 ng E. coli DNA; 0.4 Units of Platinum Taq DNA polymerase (Invitrogen Inc.); and approximately 10 ng of genomic DNA per 11 microliter reaction. Tubes containing 26, 8.75, 2.9, 0.97, 0.324 and 0.108ng of a reference DNA were included in each PCR run so that the quantity of targeted templates in each research sample could be determined relative to the reference DNA sample by the standard curve method. The same reference DNA was used for all PCR runs, Human Genomic DNA, from buffy coat (Sigma Aldrich, Cat#11691112001). Control DNA was extracted Genomic DNA from cancer cell lines and represented a wide range of telomere length: Hela, 293T, H1299, UMUC3, UMUC3 transfected by a lentivirus that overexpresses hTER to extend telomeres. The transfected cells were cultured and harvested after 8 and 12 population doublings post transfection. The PCR efficiency of single copy gene and telomere primers was > 95% and 90% respectively.

2.3 Data analysis

The DNA stock samples for this study were normalized to 20ng/μl in three 96-well plates (3 plates). Each sample was measured in triplicate on 384-well T & S plates. The T/S ratio for each sample triplicate was measured twice (2 runs). When the duplicate T/S value and the initial value varied by more than 7% for any sample, it was run a third time and the two closest values were reported. The repeat plate was a mixture from all study plates and was used to make batch adjustments. All assays for the study were performed using the same lots of reagents. Lab personnel who performed the assays were provided with de-identified samples and were blind to all demographic and clinical data.

For this study, TL of all DNA samples was measured at least twice. 116 extract samples (40%) were assayed a 3rd time; 20 samples (7%) were assayed a 4th time and 6 samples (2%) were assayed 5 times. There is partial data for a few extracts derived from the 48 starting saliva samples: due to insufficient starting material, there are no QIAamp extract measurements and only half of the Agencourt measurements for saliva #39; Agencourt DNA from one operator was low or missing for saliva samples #47 and #48, and PrepIT DNA was missing for saliva #8.

2.4 Methods for ICC and Generalizability (G) Theory

The ICC calculations were made using the rptR package supplemented by the Telomere Research Network (www.trn.tulane.edu). ICC coefficients between operators were calculated using the average of two TL runs by each operator. ICC coefficients between methods were calculated using the average TL

measurements by both operators. Samples with missing T/S value were not included in the calculation. The confidence intervals were estimated by parametric bootstrapping (n=1000).

Agreement coefficients. We used generalizability theory to estimate telomere length agreement across operators, stratified by extraction method (Model 1) and across both operators and extraction methods (Model 2). Application of generalizability theory proceeds in two steps (Brennan, 2001; Shavelson & Webb, 1991). In the first step, a generalizability (G) study estimates variance components of outcome response (telomere lengths) that are attributable to the sources under investigation (e.g., study participants, operators, extraction methods). In the second step, a decision (D) study uses the G study variance component estimates to calculate agreement coefficients.

Model 1 targeted between-operator, within-extraction-method agreement. As such, Model 1 analyses were stratified by extraction method and corresponding G studies estimated variance components for the following sources: participants (*p*), operators (*o*), and residual (*r*). In Model 1, the residual confounds the participants-by-operator (*po*) and random error (*e*) sources of variation. Model 2 targeted between-operator-and-method agreement and the G study estimated variance components for the following sources: participants (*p*), operators (*o*), extraction methods (*m*), participants-by-operator (*po*), participants-by-method (*pm*), operator-by-method (*om*), and residual (*r*). In Model 2, the residual confounds participant-by-operator-by-method (*pom*) and random error (*e*) sources of variation. Initially, Model 2 was fit to data from all three extraction methods. Subsequently, Model 2 was fit to data from each pair of extraction methods. All G study models were fit using SAS PROC MIXED with restricted maximum likelihood. For each G study, we descriptively report the percentage of total variation attributable to each estimated variance component.

In G theory parlance, participants, operators, and extraction methods are considered facets of measurement. Participants are also considered the objects of measurement. In this study, participants and operators were regarded as random facets because the goal of the analyses is to generalize to the population of potential participants and operators. On the other hand, extraction methods were considered a fixed facet because there was no intention to generalize the results to other existing or potential extraction methods.

Two types of agreement coefficients are reported: absolute and relative. Absolute agreement directly compares telomere length values across levels of one or more facets, i.e., to obtain a high level of agreement, two operators would require highly similar length assessments. Relative agreement compares rank ordering of participants (in terms of telomere length) across levels of one or more facets, e.g., one operator may systematically generate shorter telomere lengths, but both operators agree on the rank ordering of participants with respect to telomere lengths. Both absolute and relative agreement coefficients have a possible range of [0,1]. Absolute agreement represents the proportion of total variation that is exclusively attributable to the objects of measurement, here participants. For Model 1, absolute agreement (ϕ) was estimated via Equation 1, where, e.g., $\hat{\sigma}_p^2$ represents the variance component estimate for participants (Table 2 in main text). For relative agreement ($\mathrm{E}\rho^2$), the denominator includes all sources of variation attributable to participants (including participant-related interaction effects) plus random errors. For Models 1 and 2, relative agreement was estimated via Equations 2 (Table 2 in main text) and 3 (Table 3 in main text), respectively.

$$\hat{\phi}_{\mathrm{M1}} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_o^2 + \hat{\sigma}_r^2} \quad \text{Eq. 1}$$

$$\widehat{\mathrm{E}\rho}^2_{\mathrm{M1}} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_r^2} \quad \text{Eq. 2}$$

$$\widehat{\mathrm{E}\rho}^2_{\mathrm{M2}} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_{po}^2 + \hat{\sigma}_{pm}^2 + \hat{\sigma}_r^2} \quad \text{Eq. 3}$$

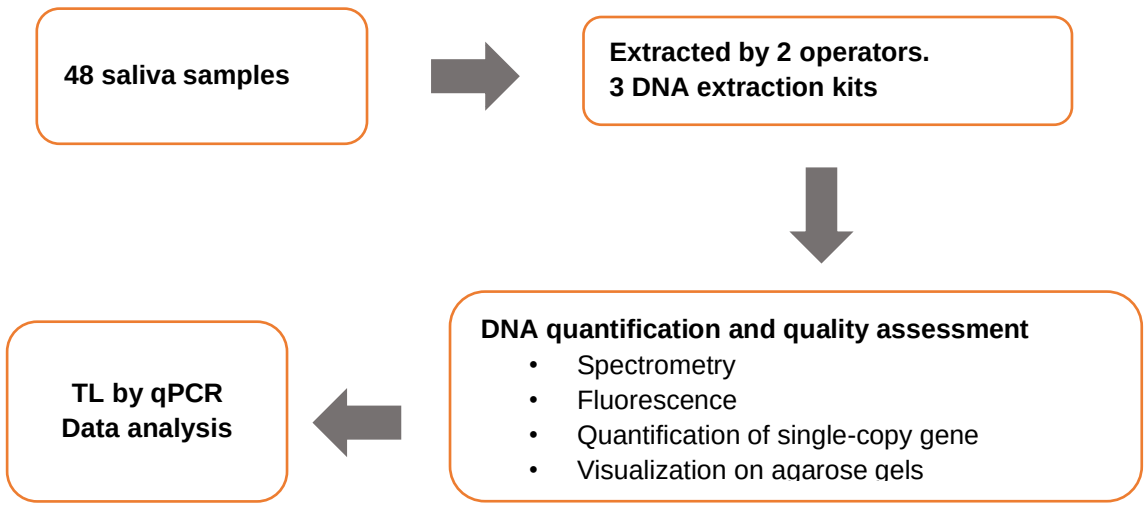
DOI: 10.2147/ijm
Available online at Journals.aijr.org

3 Results

Saliva samples for this study were a subset from a larger collection, originally gathered for the National Growth and Health Study (NGHS, nghstoday.com). Our samples came from mothers and children of different races and include 14 mother-child pairs (Table S1). The three approaches routinely used to extract DNA from whole cells are: salting out, binding to and elution from a silicon spin column, and binding to and release from magnetic beads. Three kits were chosen to represent each of these approaches (Figure S1). The overall study design is illustrated in Figure 2.

Figure S1 summarizes the three approaches to DNA extraction, all starting with lysis of cells. The PrepIT Kit, depicted in the left column, is based on the principle of salt exclusion. DNA from the cell lysate is made insoluble in a very high ionic strength solution before being precipitated with isopropanol, washed with ethanol, and then hydrated. In the center, with the QIAamp kit, lysed cells are bound to a proprietary silica membrane column. After successive washes, DNA is eluted from the column. The Agencourt Kit, on the right, works by binding DNA to magnetic beads which are attracted to and released from a magnetic plate, with successive washes before final elution from the magnetic beads.

A. Study Overview



B. DNA extraction methods included in this study.

Saliva Volume (x2 operators)	Extract principle	Kit name	Manufacturer
0.4ml	Salting out	PrepIT.L2P	DNA Genotek
0.4ml	Silicone membrane column	QIAamp DNA Mini	QIAGEN
0.5ml	Magnetic beads	Agencourt DNAdvance	Beckman Coulter

Figure 2: ***A.*** *Study overview. DNA was extracted from 48 saliva samples by two operators, using three different DNA extraction kits, based on three distinct extraction methods. DNA quantity was determined with light spectrometry, Picogreen fluorescence and amplification of RNAse P, a single-copy gene. DNA integrity was measured by observation after electrophoresis on 0.8% agarose gels.* ***B.*** *DNA extraction kit summary.*

DNA from each saliva sample was extracted by two operators (on 2 different days) using the three different techniques, such that each saliva sample yielded 6 individual extractions. Both operators used identical lots of kit reagents. DNA quantity was measured three ways: through spectrometry (using the UV-Vis Nanodrop function), fluorescence (using Picogreen) and amplification and measurement of the human genomic, single-copy RNAse P gene. The quantity of RNase P gene within a sample represents the quantity of human genome and can thus be used to quantify human DNA samples (Figure 2).

Apparent DNA yield varied widely, according to both extraction method and measurement technique. The PrepIT, salting-out treatment gave the highest, most variable extract yields (average 174.1 ng/µl +/- 19.2). QIAamp (silicon spin columns) gave consistently lower, less variable results (average 16.7 ng/µl +/- 1.2), and Agencourt magnetic bead treatment delivered intermediate amounts of DNA (average 32.2 ng/µl +/- 2.4), Figure 3). Given the different volumes of saliva used for each kit following the instructions of a specific kit, we normalized the DNA yields to represent the total yields/ml of saliva. The raw DNA concentration data for every DNA sample are listed in Table S2.

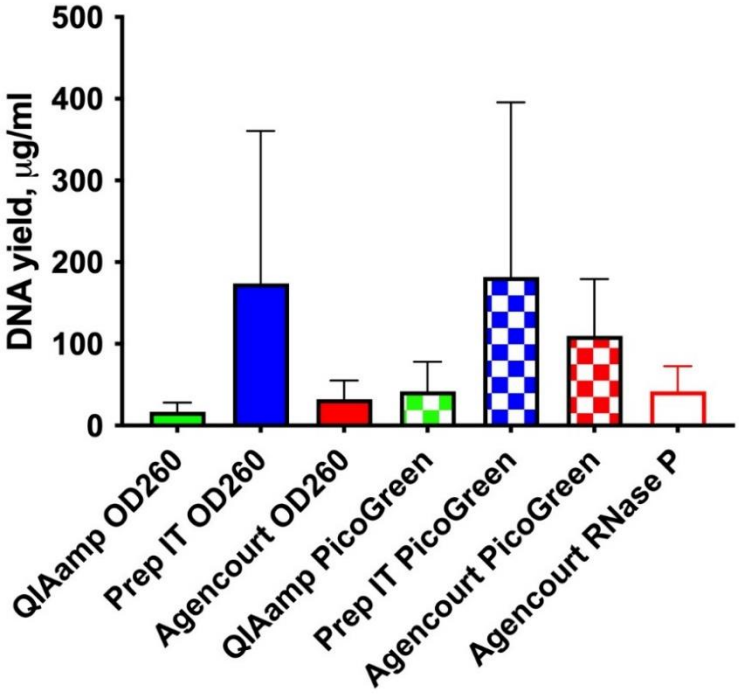


Figure 3.: *Quantification of saliva DNA (mg/µl of starting saliva sample) with Nanodrop UV-VIS spectrometry (solid), Picogreen fluorescence (checkered) or PCR amplification of the single copy RNAseP gene (no fill) extracted with QIAmp columns (green) PrepIT salting-out (blue) or Agencourt magnetic beads (red).*

Linear regressions found Agencourt DNA levels to be highly correlated across all pairings of the three measurement techniques, despite differences in DNA concentration across techniques (Figure 4). These characteristic differences may stem from inherent properties of each measurement method. Spectrometric DNA quantification was determined using the formula, dsDNA concentration = 50 µg/mL × OD260, without normalization to a standard curve. DNA quantity by both Picogreen and the RNase P assay was carried out using relative measurement to a standard curve made by serial dilution of the same commercial human genomic DNA (Sigma Aldrich cat# 11691112001). However, Picogreen measures total dsDNA including both human gDNA and bacterial gDNA, which is known to be present in saliva samples. This could explain why Picogreen values were consistently greater than RNase P values.

The quality of each DNA sample was inferred, using the UV-Vis Nanodrop function. Although pure DNA has an OD260/OD280 ratio of 1.8, conventionally, a ratio between 1.7 and 2.0 is considered acceptable, and we therefore used this criterion. Low OD260/OD230 readings indicate the presence of impurities, however, there is no consensus as to what value should be used as the cutoff. We empirically considered DNA with OD260/OD230 >1.0 to be relatively free of contaminants and found a number of QIAamp extracted DNA preps with OD260/230 values <1.0 (Table S2). DNA quality was further characterized by visual inspection after running samples normalized for DNA concentration on standard 0.8% agarose gels. Samples were independently judged to be intact or 'degraded' by both operators. Gel visualization revealed the QIAamp preps to be consistently degraded to a small degree, whereas PrepIT DNA and Agencourt preps appeared intact (Representative gel images of 4 saliva samples extracted by the 3 kits are shown in Figure 4B and images for all samples in Figure S2). The lower yields for QIAamp preps are likely attributable to this persistent DNA degradation factor.

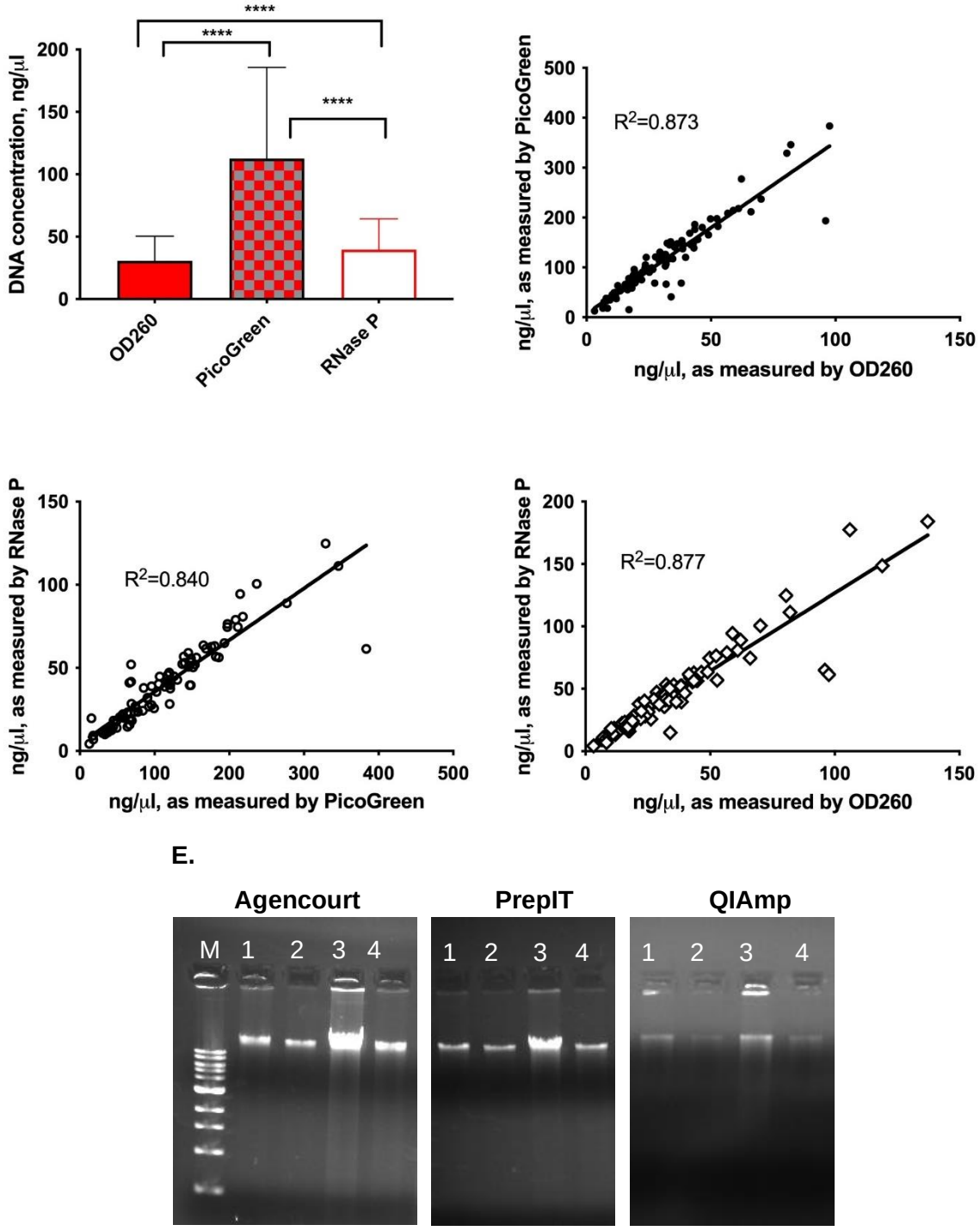


Figure 4:***(A)*** *Comparison of Agencourt (magnetic bead prep) DNA extract yield measurements. DNA concentration is expressed as ng/µl and yields are normalized to represent total yield/ml of starting saliva sample. The three measurement methods are highly correlated (P, **** <0.0001):* ***(B)*** *PicoGreen fluorescence with Nanodrop UV-Vis OD260;* ***(C)*** *RNAse P amplification with PicoGreen fluorescence; and* ***(D)*** *RNAse P amplification with OD260.* ***(E)*** *Saliva DNA extracted by QIAamp mini kit exhibits partial degradation compared with Agencourt or PrepIT.*

To evaluate the telomere length data, comparisons were first made between extraction methods by averaging the data from both operators for analysis (See Table S3 for complete set of telomere length values.). Significant differences in TL (T/S qPCR ratios) were found, dependent on extraction technique.

T/S ratios obtained with QIAamp were greater than PrepIT which were greater than Agencourt (Figure 5A). Further, Agencourt and QIAamp each had strong between-operator correlations, whereas the between-operator correlation for PrepIT was less substantial (Figure 5 B, C, D).

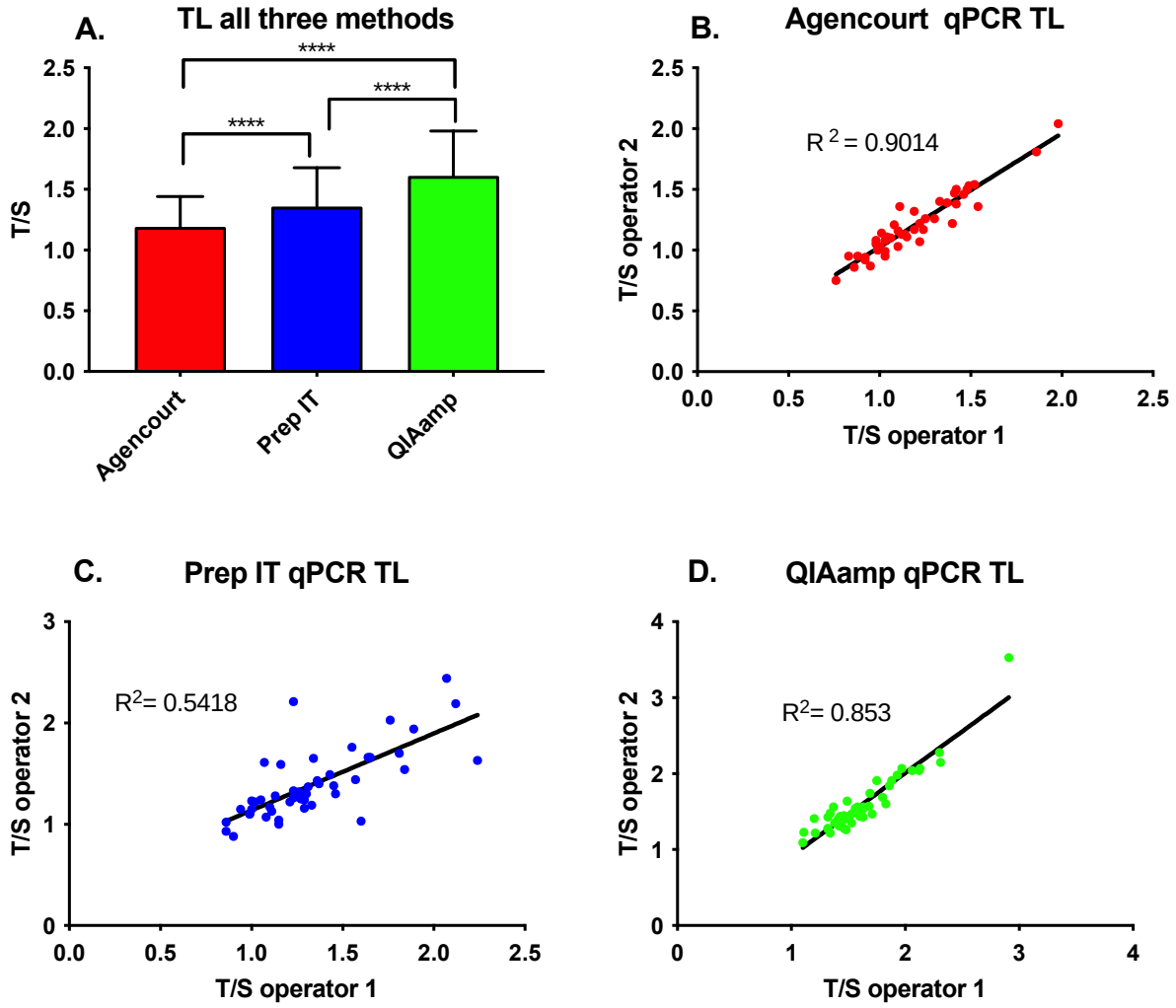


Figure 5: *TL qPCR correlation.* ***A.*** *Systematic differences in TL, with different DNA extraction methods (P, **** < 0.0001).* ***B-D.*** *TL correlation between two operators, based on extraction technique:* ***B.*** *Agencourt qPCR TL, operator 1 vs. operator 2.* ***C.*** *PrepIT qPCR TL, operator 1 vs. operator 2.* ***D.*** *QIAamp qPCR TL, operator 1 vs. operator 2.*

To further characterize these comparisons, we estimated intraclass correlations (ICC) to assess agreement between T/S values across extraction techniques, between operators, and between runs. Consistent with linear regression correlations depicted in Figure 5, ICCs between two operators for Agencourt and QIAamp DNA were high but lower for PrepIT. Agencourt is better correlated with both QIAamp and PrepIT, compared with QIAamp versus PrepIT. In addition, ICC between two runs of the same DNA sample was very high, and no difference between the three kits was present. This is likely because the between run ICC is a reflection of the precision of the analytical components of the assay, not of inherent DNA quality or operator handling (Table 1).

DOI: 10.2147/ijm
Available online at Journals.aijr.org

Table 1: *ICC coefficients for within-operator between-run, between-operator, and between-extraction method TL measurements.*

Method / Operator	Agencourt / Op 1	Agencourt / Op 2	Prep IT / Op 1	Prep IT / Op 2	QIAamp / Op 1	QIAamp / Op 2
Agencourt / Op 1	0.985 CI = [0.972, 0.991]					
Agencourt / Op 2	0.950 CI = [0.911, 0.972]	0.991 CI = [0.985, 0.995]				
Prep IT / Op 1	Agencourt vs. Prep IT 0.839 CI = [0.732, 0.905]		0.993 CI = [0.988, 0.996]			
Prep IT / Op 2			0.809 CI = [0.678, 0.889]	0.989 CI = [0.979, 0.994]		
QIAamp / Op 1	Agencourt vs. QIAamp 0.876 CI = [0.793, 0.932]		Prep IT vs. QIAamp 0.755 CI = [0.606, 0.861]		0.990 CI = [0.981, 0.994]	
QIAamp / Op 2					0.937 CI = [0.891, 0.965]	0.985 CI = [0.974, 0.992]

We applied generalizability theory to decompose sources of TL variation and estimate agreement coefficients. The first set of generalizability theory analyses focused on between-operator agreement of TL measures, stratified by DNA extraction method. Absolute between-operator agreement was high for Agencourt and QIAamp, whereas absolute agreement was lower for PrepIT in this study (Table 2). Corresponding absolute and relative agreement coefficients had similar values because variation attributable exclusively to operators was negligible.

Table 2: *Between-operator agreement of TL measures, stratified by DNA extraction method: Variance component estimates and agreement coefficients.*

	Percentage of TL measure variation by source			Agreement coefficients	
Method	participant	operator	residual	Absolute	Relative
Agencourt	94.63	0.04	5.33	0.946	0.947
PrepIT	81.80	0.33	17.87	0.818	0.821
QIAamp	92.35	0.06	7.59	0.924	0.924

Note: In Model 1, the residual confounds the participant-by-operator and random error sources.

The second set of generalizability theory analyses focused on between-operator-and-method agreement of TL measures. Because telomere lengths generated from the three extraction methods had differing mean levels (Figure 5A), we emphasized relative agreement for these comparisons. Considering all three extraction methods simultaneously, the relative agreement coefficient equaled 0.742 (Table 3). Relative agreement levels for each pair of extraction methods were roughly similar, ranging from 0.696 to 0.779. Not surprisingly, the percentage of total variation attributable to extraction methods was substantial

(29.6% when considering all three extraction methods). The percentage of variation attributable to extraction method across the three pairwise analyses are consistent with what we know about the relative performance of the three methods: Agencourt tends to yield the lowest T/S ratios, PrepIT middling, and QIAamp the highest. The participant-by-method interaction explained about 10% of total variation in telomere lengths, suggesting that differences in relative length estimates across methods are somewhat participant-specific.

Table 3.: *Between-operator-and-method agreement of TL measures:Variance component estimates and agreement coefficients***.**

Extraction Methods	Percentage of total telomere length variation by source							Relative Agreement
	participant	operator	method	po	pm	om	residual	
All three Methods	52.34	0	29.26	1.29	10.17	0.18	6.77	0.742
Agencourt v PrepIT	65.25	0.30	14.09	1.08	7.96	0	11.31	0.762
Agencourt v QIAamp	41.58	0	46.57	0	8.06	0.03	3.75	0.779
PrepIT v QIAamp	56.54	0	18.45	3.01	14.49	0.29	7.22	0 .696

Abbreviations for variance sources: po, participant-by-operator; pm, participant-by-method, om, operator-by-method. In Model 2, the residual confounds the participant-by-operator-by method (pom) and random error sources.

4 Discussion

Over the 20 years that qPCR telomere length methods have been developed, it has become clear that individual labs can generate different TL qPCR values for the same set of biological samples. The reasons for this are complicated, with variation arising from many possible sources. In addition to DNA extraction kits and quantification methods, other contributing factors (not addressed in this study) include: the use of different reference DNAs for standardization of TL values, selection of different master mixes for the PCR reaction, different single copy gene assays used in calculating T/S ratios, and different approaches to data analysis. In this single-lab analysis of 3 contributing factors, DNA extraction method, DNA quantification technique and human operator, several trends emerged.

Agencourt kit gave the best quality DNA and most consistent TL results

We have shown that qPCR TL measurements vary depending on which kit is used to extract DNA. Often, DNA yield was higher with the PrepIT (salting out) kit, but there was only modest correlation of PrepIT TL values between two operators. Based on both linear regression and ICC, both Agencourt (magnetic bead separation) and QIAamp (silicon spin column separation) kits resulted in high between-operator TL measure agreement. Nevertheless, since the QIAamp DNA was consistently partially degraded, while Agencourt DNA remained intact, we recommend Agencourt as the better DNA extraction technique.

TL data gathered from DNA extracted by different methods cannot be combined.

Despite starting from the same saliva samples, the TL values obtained from the three extraction techniques varied, but not in a systematic way that would make it appropriate to apply a 'correction factor' to arrive at similar TL measurements from DNA extracted different ways. This is demonstrated by the between-operator-and-method agreement and variance component estimates (Table 3). Variation due to the extraction method main effect, while substantial, is not considered in the estimation of relative agreement. Yet, the agreement coefficients, ranging from 0.70 to 0.78, were not high. We also found that the participant-by-method interaction explained about 10% of total variation in telomere lengths, suggesting

that differences in relative length estimates across methods are somewhat participant-specific. Therefore, we would strongly recommend applying the same DNA purification kit to all samples within a given study. It is not appropriate to combine samples extracted with different kits.

Comparison to another TL method needed to estimate qPCR TL assay accuracy in saliva DNA.

We found that despite absolute differences in in DNA yield, DNA quality and TL values, Agencourt and QIAamp data were highly correlated. Unfortunately, most of our DNA, extracted from saliva, was not concentrated enough to carry out TRF Southern blot analysis. Furthermore, TRF analysis of a subset of highly concentrated Agencourt DNA extracts (executed as described in (Kimura et al., 2010), and shown in Figure S3), revealed an average telomere length of 3.8 +/- 0.1Kb, considerably shorter than the expected telomere length seen with TRF analysis in human DNA extracted from blood, which is about 9.5 +/- 0.7 in newborns and 7.8 +/- 0.7 in adults (Factor-Litvak et.al., 2016), suggesting telomere degradation in DNA extracted from saliva. This apparent degradation of our telomere-specific DNA on a Southern blot gel did not correlate with the clean migration of bulk DNA we observed on 0.8% agarose gels when assessing DNA integrity. Since we were unable to directly compare qPCR TL values to a different telomere measurement technique, our differences in TL values remain highly correlated but relative. Our study measured assay reliability (agreement) but not absolute TL validity (accuracy), as we cannot compare qPCR TL with another method expressed as basepairs. The relative nature of qPCR TL measurement is an inherent limitation and suggests a strong need to develop other DNA-based methods to measure absolute telomere length.

Some new approaches involve whole genome sequencing followed by analysis of the repetitive telomeric sequences(Lee et.al., 2017). Nanopore technology, in which long sequencing reads enable the measurement of the entire telomere sequence is especially promising (Fang et al., 2022; Tham et al., 2023). Perhaps, with further characterization, high-throughput qPCR TL results could be reliably and accurately validated.

5 Conclusions

As telomere length continues to evolve as a replicative aging marker in population studies, qPCR remains a more practical technique for studies with limited available source material. Saliva is a cost-effective, non-invasive choice for sample collection. In this work we examined the impacts of DNA extraction methods and DNA quality on assay precision and identified potential sources of assay variation and ways to improve precision. From this work, we provide some recommendations:

Extraction method matters. We found that the Agencourt kit, which involves purification of DNA through binding to magnetic beads, gave the best quality DNA and most consistent TL results. Alternative purification techniques led to inconsistent yield and/or a greater tendency for degradation.

TL data from DNA extracted by different methods cannot be combined. We found that TL values varied with extraction technique and a significant percentage of variation is participant specific. Whatever extraction technique is employed, we recommend that DNA extracted through different methods never be combined for analysis.

Comparison to another TL method is needed to estimate qPCR TL assay accuracy in saliva DNA. Our extractions did not yield DNA of sufficient quantity and quality for comparison to TRF analysis. New approaches and technologies may reveal new ways to measure telomeres with base-pair precision.

6 Declarations

6.1 Acknowledgments

We thank Elissa Epel, Department of Psychiatry, University of California, San Franciscofor providing access to the saliva samples for this study and Ethan Dutcher for assisting with Table S1. The

authors assert that approval for use of human saliva samples was issued under University of California Federalwide Assurance #00006252 by the Berkeley Committee for Protection of Human Subjects (CPHS).

6.2 Funding Source

National Institutes of Health, Grant number U01AG064785

6.3 Competing Interests

The authors report no conflicts of interest in this publication.

6.4 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published institutional affiliations.

How to Cite this Article:

D. L. Smith, C. Wu, S. Gregorich, G. Dai, and J. Lin, "Impact of DNA Extraction Methods on Quantitative PCR Telomere Length Assay Precision in Human Saliva Samples", *Int. J. Methodol.*, vol. 1, no. 1, pp. 44–57, Dec. 2022. https://doi.org/10.21467/ijm.1.1.5784

Supplementary File

Figures & Tables (S1-S3) are available at URL https://journals.aijr.org/index.php/ijm/article/view/5784/535

References

Ahmed, W., & Lingner, J. (2018). PRDX1 and MTH1 cooperate to prevent ROS-mediated inhibition of telomerase. *Genes & Development*, *32*(9–10), 658–669. https://doi.org/10.1101/GAD.313460.118

Aviv, A., & Shay, J. W. (2018). Reflections on telomere dynamics and ageing-related diseases in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1741). https://doi.org/10.1098/RSTB.2016.0436

Baerlocher, G. M., Vulto, I., de Jong, G., & Lansdorp, P. M. (2006). Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nature Protocols 2006 1:5*, *1*(5), 2365–2376. https://doi.org/10.1038/nprot.2006.263

Blackburn, E. H. (2001). Switching and Signaling at the Telomere. *Cell*, *106*(6), 661–673. https://doi.org/10.1016/S0092-8674(01)00492-5

Blackburn, E. H., Epel, E. S., & Lin, J. (2015). Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science*, *350*(6265), 1193–1198. https://doi.org/10.1126/SCIENCE.AAB3389

Brennan, R. L. (2001). *Generalizability Theory*. Springer New York. https://doi.org/10.1007/978-1-4757-3456-0

Broer, L., Codd, V., Nyholt, D. R., Deelen, J., Mangino, M., Willemsen, G., Albrecht, E., Amin, N., Beekman, M., de Geus, E. J. C., Henders, A., Nelson, C. P., Steves, C. J., Wright, M. J., de Craen, A. J. M., Isaacs, A., Matthews, M., Moayyeri, A., Montgomery, G. W., … Boomsma, D. I. (2013). Meta-analysis of telomere length in 19 713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *European Journal of Human Genetics 2013 21:10*, *21*(10), 1163–1168. https://doi.org/10.1038/ejhg.2012.303

Cawthon, R. M. (2002). Telomere measurement by quantitative PCR. *Nucleic Acids Research*, *30*(10), e47–e47. https://doi.org/10.1093/NAR/30.10.E47

Cawthon, R. M. (2009). Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Research*, *37*(3), e21–e21. https://doi.org/10.1093/NAR/GKN1027

Chan, S. W.-L., & Blackburn, E. H. (2002). New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin. *Oncogene 2002 21:4*, *21*(4), 553–563. https://doi.org/10.1038/sj.onc.1205082

Cheng, F., Carroll, L., Joglekar, M. v., Januszewski, A. S., Wong, K. K., Hardikar, A. A., Jenkins, A. J., & Ma, R. C. W. (2021). Diabetes, metabolic disease, and telomere length. *The Lancet Diabetes & Endocrinology*, *9*(2), 117–126. https://doi.org/10.1016/S2213-8587(20)30365-X

de Meyer, T., Nawrot, T., Bekaert, S., de Buyzere, M. L., Rietzschel, E. R., & Andrés, V. (2018). Telomere Length as Cardiovascular Aging Biomarker: JACC Review Topic of the Week. *Journal of the American College of Cardiology*, *72*(7), 805–813. https://doi.org/10.1016/J.JACC.2018.06.014

DNA Genotek. *Saliva collection kits for high-quality DNA saliva samples enabling genetic diagnostics*. DNA Genotek's Oragene Kit . Retrieved January 27, 2023, from https://www.dnagenotek.com/US/products/collection-human/collection-kits-for-diagnostics.html

Eisenberg, D. T. A. (2016). Telomere length measurement validity: the coefficient of variation is invalid and cannot be used to compare quantitative polymerase chain reaction and Southern blot telomere length measurement techniques. *International Journal of Epidemiology*, *45*(4), 1295–1298. https://doi.org/10.1093/IJE/DYW191

Factor-Litvak, P., Susser, E., & Aviv, A. (2017). Environmental Exposures, Telomere Length at Birth, and Disease Susceptibility in Later Life. *JAMA Pediatrics*, *171*(12), 1143–1144. https://doi.org/10.1001/JAMAPEDIATRICS.2017.3562

Factor-Litvak, P., Susser, E., Kezios, K., McKeague, I., Kark, J. D., Hoffman, M., Kimura, M., Wapner, R., & Aviv, A. (2016). Leukocyte telomere length in newborns: Implications for the role of telomeres in human disease. *Pediatrics*, *137*(4). https://doi.org/10.1542/PEDS.2015-3927

Fang, L., Liu, Q., Monteys, A. M., Gonzalez-Alegre, P., Davidson, B. L., & Wang, K. (2022). DeepRepeat: direct quantification of short tandem repeats on signal data from nanopore sequencing. *Genome Biology*, *23*(1), 1–27. https://doi.org/10.1186/S13059-022-02670-6

Fouquerel, E., Barnes, R. P., Uttam, S., Watkins, S. C., Bruchez, M. P., & Opresko, P. L. (2019). Targeted and Persistent 8-Oxoguanine Base Damage at Telomeres Promotes Telomere Loss and Crisis. *Molecular Cell*, *75*(1), 117-130.e6. https://doi.org/10.1016/J.MOLCEL.2019.04.024

Haycock, J. (2017). Estimating Ten-Year Trends in Septic Shock Incidence and Mortality in United States Academic Medical Centers Using Clinical Data: Kadri SS, Rhee C, Strich JR, et al. Chest 2017;151(2):278-285. *The Journal of Emergency Medicine*, *53*(1), 153. https://doi.org/10.1016/J.JEMERMED.2017.05.010

Jose, S. S., Bendickova, K., Kepak, T., Krenova, Z., & Fric, J. (2017). Chronic inflammation in immune aging: Role of pattern recognition receptor crosstalk with the telomere complex? *Frontiers in Immunology*, *8*(SEP), 1078. https://doi.org/10.3389/FIMMU.2017.01078

Kimura, M., Stone, R. C., Hunt, S. C., Skurnick, J., Lu, X., Cao, X., Harley, C. B., & Aviv, A. (2010). Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nature Protocols 2010 5:9*, *5*(9), 1596–1607. https://doi.org/10.1038/nprot.2010.124

Lai, T. P., Wright, W. E., & Shay, J. W. (2018). Comparison of telomere length measurement methods. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1741). https://doi.org/10.1098/RSTB.2016.0451

Lee, M., Napier, C. E., Yang, S. F., Arthur, J. W., Reddel, R. R., & Pickett, H. A. (2017). Comparative analysis of whole genome sequencing-based telomere length measurement techniques. *Methods*, *114*, 4–15. https://doi.org/10.1016/J.YMETH.2016.08.008

Lin, J., & Epel, E. (2022). Stress and telomere shortening: Insights from cellular mechanisms. *Ageing Research Reviews*, *73*, 101507. https://doi.org/10.1016/J.ARR.2021.101507

Lin, J., Smith, D. L., Esteves, K., & Drury, S. (2019). Telomere length measurement by qPCR – Summary of critical factors and recommendations for assay design. *Psychoneuroendocrinology*, *99*, 271–278. https://doi.org/10.1016/J.PSYNEUEN.2018.10.005

Lindrose, A. R., McLester-Davis, L. W. Y., Tristano, R. I., Kataria, L., Gadalla, S. M., Eisenberg, D. T. A., Verhulst, S., & Drury, S. (2021). Method comparison studies of telomere length measurement using qPCR approaches: A critical appraisal of the literature. *PLOS ONE*, *16*(1), e0245582. https://doi.org/10.1371/JOURNAL.PONE.0245582

Poon, S. S., & Lansdorp, P. M. (2001). Quantitative Fluorescence In Situ Hybridization (Q-FISH). *Current Protocols in Cell Biology*, *12*(1), 18.4.1-18.4.21. https://doi.org/10.1002/0471143030.CB1804S12

Ridout, K. K., Levandowski, M., Ridout, S. J., Gantz, L., Goonan, K., Palermo, D., Price, L. H., & Tyrka, A. R. (2017). Early life adversity and telomere length: a meta-analysis. *Molecular Psychiatry 2018 23:4*, *23*(4), 858–871. https://doi.org/10.1038/mp.2017.26

Ruiz, A. ;, Flores-Gonzalez, J. ;, Buendia-Roldan, I. ;, Chavez-Galan, L., Ruiz, A., Flores-Gonzalez, J., Buendia-Roldan, I., & Chavez-Galan, L. (2021). Telomere Shortening and Its Association with Cell Dysfunction in Lung Diseases. *International Journal of Molecular Sciences 2022, Vol. 23, Page 425*, *23*(1), 425. https://doi.org/10.3390/IJMS23010425

Schneider, C. v., Schneider, K. M., Teumer, A., Rudolph, K. L., Hartmann, D., Rader, D. J., & Strnad, P. (2022). Association of Telomere Length With Risk of Disease and Mortality. *JAMA Internal Medicine*, *182*(3), 291–300. https://doi.org/10.1001/JAMAINTERNMED.2021.7804

Shalev, I., Entringer, S., Wadhwa, P. D., Wolkowitz, O. M., Puterman, E., Lin, J., & Epel, E. S. (2013). Stress and telomere biology: A lifespan perspective. *Psychoneuroendocrinology*, *38*(9), 1835–1842. https://doi.org/10.1016/J.PSYNEUEN.2013.03.010

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory : a primer* (Vol. 1). Sage Publications.

Tham, C. Y., Poon, L. F., Yan, T. D., Koh, J. Y. P., Ramlee, M. K., Teoh, V. S. I., Zhang, S., Cai, Y., Hong, Z., Lee, G. S., Liu, J., Song, H. W., Hwang, W. Y. K., Teh, B. T., Tan, P., Xu, L., Koh, A. S., Osato, M., & Li, S. (2023). High-throughput telomere length measurement at nucleotide resolution using the PacBio high fidelity sequencing platform. *Nature Communications 2023 14:1*, *14*(1), 1–13. https://doi.org/10.1038/s41467-023-35823-7

Verhulst, S., Susser, E., Factor-Litvak, P. R., Simons, M. J. P., Benetos, A., Steenstrup, T., Kark, J. D., & Aviv, A. (2015). Commentary: The reliability of telomere length measurements. *International Journal of Epidemiology*, *44*(5), 1683–1686. https://doi.org/10.1093/IJE/DYV166

Wang, Q., Codd, V., Raisi-Estabragh, Z., Musicha, C., Bountziouka, V., Kaptoge, S., Allara, E., Angelantonio, E. di, Butterworth, A. S., Wood, A. M., Thompson, J. R., Petersen, S. E., Harvey, N. C., Danesh, J. N., Samani, N. J., & Nelson, C. P. (2021). Shorter leukocyte telomere length is associated with adverse COVID-19 outcomes: A cohort study in UK Biobank. *EBioMedicine*, *70*, 103485. https://doi.org/10.1016/J.EBIOM.2021.103485

Willis, M., Reid, S. N., Calvo, E., Staudinger, U. M., & Factor-Litvak, P. (2018). A scoping systematic review of social stressors and various measures of telomere length across the life course. *Ageing Research Reviews*, *47*, 89–104. https://doi.org/10.1016/J.ARR.2018.07.006