

Chapter

Current Technologies for Measuring or Predicting Telomere Length from Genomic Datasets

Ting Zhai and Zachary D. Nagel

Abstract

The gold standard for measuring telomere length is technically challenging, which limits its use in large population studies. Numerous bioinformatics tools have recently been developed to estimate telomere length using high-throughput sequencing data. This allows for scaling up telomere length estimates in large datasets. Telomere length depends substantially on genetics, and many genetic studies have looked at this relationship, which provides an opportunity to predict telomere length from genotyping data. However, in part because environment also significantly affects telomere length, the accuracy of telomere length predictions and estimates made from genomic data remains uncertain. In this chapter, we will summarize currently available bioinformatics tools for predicting or measuring telomere length from genomics datasets, and we will discuss each method's limitations and advantages.

Keywords: DNA sequencing, genomics, genetics, population sciences, bioinformatics

1. Introduction

Telomere length, as a biomarker of aging and genome integrity, has frequently been used in aging and health research. Telomere length in blood leukocytes or peripheral blood mononuclear cells (PBMCs) reflects the progressive shortening of telomeres in hematopoietic stem and progenitor cells (HSPCs) and correlates with telomere length in other tissues [1]. Among leukocyte types, naïve T-cells and B-cells have the longest telomeres, whereas NK- and memory T-cells have shorter telomeres within the same individual [2]. However, consistent with blood telomere length as a representative marker for telomere attrition in other tissues, telomere length in all cell types shows an inverse correlation with participant age. Many age related diseases are associated with diminished telomere length in leukocytes or PBMCs, including cardiovascular disease [3] and cancer [4]. Accurate and accessible telomere length techniques are increasingly needed in population studies to evaluate the biological aging process and genome maintenance, and may represent a target for the development of strategies for early detection and prevention of disease.

The gold standard measurement of telomere length is by Terminal Restriction Fragment (TRF) analysis. This approach involves frequent-cutting restriction enzymes that do not recognize the telomeric repeats, followed by Southern

hybridization [5]. The entire process of TRF can take up to one week, making it infeasible to apply on a large population scale. Flow cytometry-based fluorescent in situ hybridization (flow-FISH) is frequently used in clinical settings for patient diagnoses of telomere related disease [6]. An advantage of this approach is that it allows for determination of telomere length for specific cell population, but it requires expertise in flow cytometry and still labor intensive. Currently, the monochrome multiplex-quantitative polymerase chain reaction (MM-qPCR) is widely used in population-based studies because it is less labor intensive and relevantly fast in the turnaround time [7]. However, even this approach requires careful attention to details that may affect the results of the analysis and places demands on sample processing that may not be compatible with population studies [8].

Recently, genomics datasets have been produced from many large population cohorts, enabling new approaches to estimating telomere length through bioinformatics tools. These approaches afford an opportunity to estimate telomere length in large-scale population datasets. In this chapter, we summarize currently available tools for predicting or measuring telomere length from genomics datasets, and we will discuss each method's limitations and advantages.

2. Predicting telomere length from genotyping datasets

2.1 Genome-wide association studies of telomere length

Telomere length variation between individuals can be explained at least in part by genetics. Telomeropathies are Mendelian diseases characterized by impaired telomere maintenance and caused by defects in genes involved in telomere maintenance [9]. Genome-Wide Association Studies (GWAS) of leukocyte telomere length have consistently identified genetic loci within genes with key roles in telomere length regulation. These studies have shed light on the connection between genetic variants and telomere length, allowing for the prediction of telomere length using an individual's genetic information.

In 2021, Codd et al. reported the largest GWAS on telomere length to date, using genetic data from over 472,000 participants in the UK Biobank [10]. The study identified 197 independent genetic variants associated with leukocyte telomere length at 138 genomic loci, with 108 being newly discovered. Genes involved in regulating telomeres were found in 44 loci, including those that encode components of the Shelterin and CTC1-STN1-TEN1 (CST) complexes. The newly discovered loci also included genes involved in the alternative lengthening of telomeres (ALT) pathway and factors that modify key telomere proteins post-translationally. Additionally, genes that regulate telomerase such as TERC and TERT were reported. This study further established a method for predicting telomere length from the identified loci and revealed its relationship to age-related disease outcomes involving multiple biological traits and chronic pathologies.

Also in 2021, Chang et al. reported the largest GWAS of leukocyte telomere length in an Asian population, among 25,533 Chinese Singaporean individuals [11]. The study identified three variants in or near the POT1, TERF1 and STN1 genes that were associated with telomere length and specific to East Asians. Additionally, the authors reported a significantly increased risk of incident lung cancer with increased genetic telomere length. Upon further analysis stratifying on subtypes, this association was found only in lung adenocarcinoma.

These studies provide valuable insights into the genetic determinants of telomere length and its association with various health outcomes. Further research is needed to fully understand the mechanisms underlying these associations and to develop potential interventions to improve health outcomes.

2.2 Predicting telomere length using Mendelian randomization

In this section, we discuss the emerging use of Mendelian randomization (MR) to predict telomere length from germline genotyping data. MR is an approach that uses genetic variants to infer causal relationships between genotype and phenotype. It is based on Mendel's laws of inheritance and causal inference theory, using instrumental variables to account for unmeasured confounding [12]. MR has recently gained popularity, with a growing number of methodologies and applied studies being published, enabled by an increasing availability of genetic data [13].

There are three basic assumptions of MR in telomere length studies: (1) the relevance assumption - the genetic variant(s) used as instruments must be strongly associated with telomere length; (2) the independence assumption - the genetic variant(s) should be independent of the outcome, given telomere length and all confounders; and (3) the exclusion assumption - the genetic variant(s) should only influence the outcome through their effect on telomere length. The first assumption is directly related to our topic of predicting telomere length from germline variants.

To identify germline variants that can be used as instrumental variables for telomere length, a set of screening and filtering criteria can be applied to available genetic datasets. In this section, we outline our approach for selecting genetic instruments for telomere length in European populations. We have chosen to focus on European populations because they currently have the largest available databases among all populations in published genetic datasets.

In January 2023, we conducted a search of telomere length-related studies on the GWAS Catalog (https://www.ebi.ac.uk/gwas/efotraits/EFO_0004505), and identified 303 SNPs from 20 published GWAS studies. A detailed flowchart has been provided to illustrate each step of our screening and filtering process (**Figure 1**). We then systematically screened these studies using the following exclusion criteria: (1) non-European ethnicity; (2) telomere length measured in patient populations; (3) telomere length measured in cell types other than leukocytes or PBMCs; and (4) inconsistent units of telomere length measurement. After applying these criteria, 9 out of the 20 studies were selected for full-text screening.

During the full-text screening process, we eliminated 4 out of the 9 studies because they were conducted earlier with smaller sample sizes and shared the same populations as the included studies. We also expanded our list of potential genetic instrument candidates by adding more SNPs from the study results and supplementary lists along with the published manuscript. This resulted in a candidate list of 143 SNPs from 5 studies [14–18], which was reduced to a final list of 138 SNPs after meta-analyzing for duplicated SNPs.

From this final candidate list, we applied SNP-specific filtering criteria to identify strong SNPs for telomere length: (1) genome-wide significance with p -value $< 5 \times 10^{-8}$, as defined by the relevance assumption; (2) minor allele frequency (MAF) > 0.01 , to avoid potential statistical bias from SNPs with low MAF; and (3) pruning for linkage disequilibrium (LD) at an R^2 coefficient of correlation < 0.01 , to ensure that the included genetic instruments are independent of each other, and that we have only one representative SNP per region of LD.

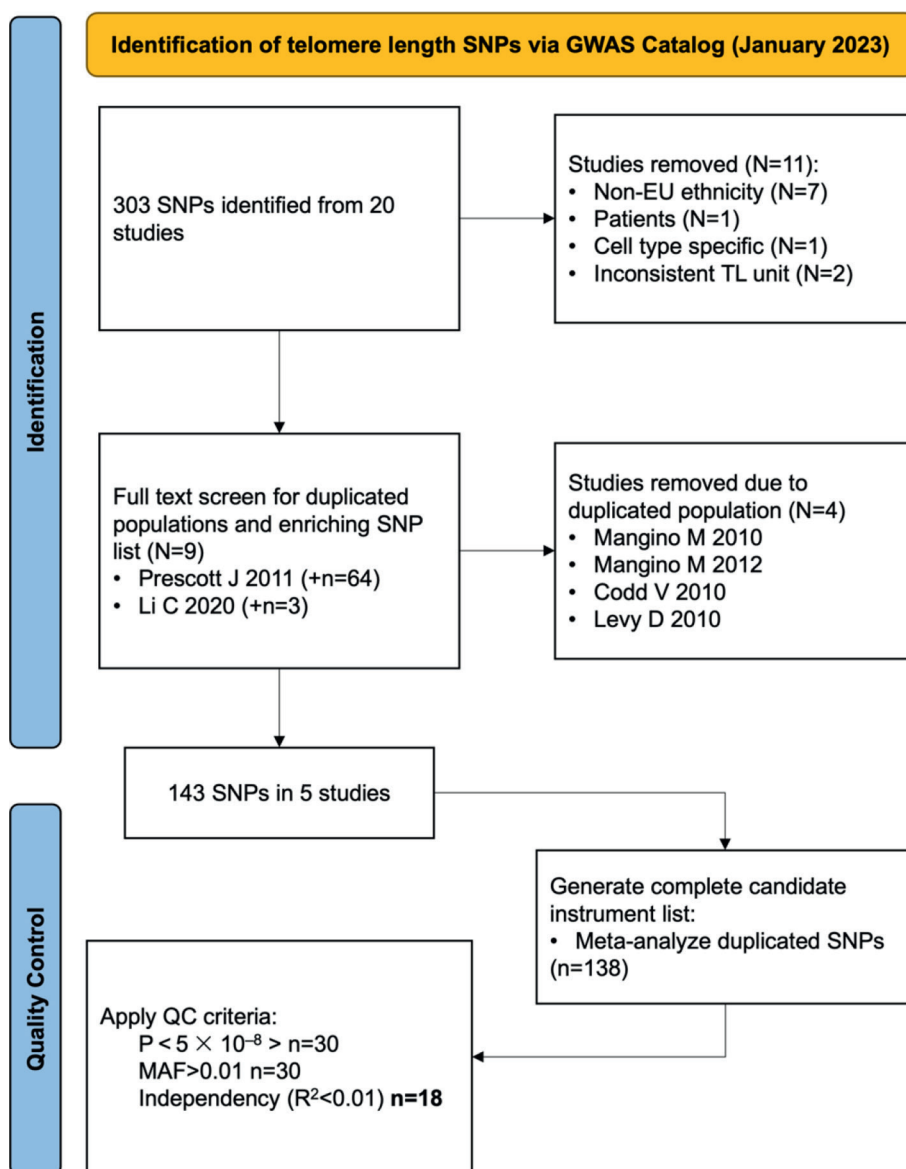


Figure 1.

Identification of genetic instruments of telomere length from GWAS catalog. In January 2023, the GWAS Catalog was searched for the trait “telomere length” (EFO_0004505), resulting in 303 SNPs from 20 published studies. After screening the abstracts, seven studies were removed due to non-European populations, one due to being conducted on patients, one due to reporting cell-type specific telomere length, and two due to reporting associations in different units. Of the remaining nine studies, four were removed after full-text screening because they used the same study populations as other studies but had smaller sample sizes. This left five studies with a total of 143 SNPs. After meta-analysis to remove duplicate SNPs and SNP-specific filtering, 30 SNPs had genome-wide significance ($P < 5 \times 10^{-8}$) and minor allele frequency > 0.01 . Of these, 18 were independent ($R^2 < 0.01$) within a 10 Mb region. The final list consisted of 18 independent SNPs as genetic instruments for telomere length, eligible for Mendelian randomization analysis. Abbreviation: SNP, single-nucleotide polymorphism; GWAS, genome-wide association study; EU, European; TL, telomere length; QC, quality control; MAF, minor allele frequency. N denotes the sample size for studies, and n denotes the sample size for SNPs.

After applying these criteria, we identified 30 SNPs with genome-wide significant p value $<5 \times 10^{-8}$ and MAF > 0.01 . We then performed LD clumping with $R^2 < 0.01$ to identify the top SNP per 10 Mb region, resulting in a total of 18 SNPs as our final genetic instruments for telomere length (**Table 1**).

With our final genetic instruments for predicting telomere length through genetic variants in hand, we can use these SNPs to conduct Mendelian randomization analyses. This approach allows us to infer causal relationships between telomere length and other phenotypes by using the genetic instruments as proxies for telomere length.

CHR	POS	GENE	SNPS	PVAL	BETA	SE	REF	EFT
1	226,562,621	PARP1	rs3219104	9.31E-11	0.0417	0.0064	A	C
2	54,248,729	ACYP2	rs11125529	8.00E-10	-0.0560	0.0100	A	C
3	169,514,585	LRRC34 (TERC)	rs10936600	6.42E-51	-0.0858	0.0057	A	T
3	101,232,093	SENP7	rs55749605	2.38E-08	-0.0373	0.0067	C	A
4	164,048,199	NAF1	rs4691895	1.47E-21	0.0577	0.0061	G	C
4	71,774,347	MOB1B	rs13137667	2.37E-08	0.0765	0.0137	T	C
5	1,285,974	TERT	rs7705526	4.82E-45	0.0820	0.0058	C	A
6	31,587,561	PRRC2A	rs2736176	3.41E-10	0.0345	0.0055	G	C
7	124,554,267	POT1	rs59294613	1.12E-13	-0.0407	0.0055	C	A
10	103,916,188	OBFC1	rs9419958	1.56E-19	0.0523	0.0058	C	T
11	108,105,593	ATM	rs228595	1.39E-08	-0.0285	0.0050	G	A
14	73,404,752	DCAF4	rs2302588	1.64E-08	0.0476	0.0084	G	C
16	69,406,986	TERF2	rs3785074	4.50E-10	0.0351	0.0056	A	G
16	82,199,980	MPHOSPH6	rs7194734	6.72E-10	-0.0369	0.0060	C	T
16	74,680,074	RFWD3	rs62053580	3.96E-08	-0.0389	0.0071	A	G
19	22,032,639	ZNF208	rs8105767	7.01E-21	-0.0420	0.0045	G	A
20	62,269,750	STMN3	rs75691080	5.75E-14	-0.0671	0.0089	C	T
20	62,291,599	RTEL1	rs34978822	7.04E-10	-0.1397	0.0227	C	G

Abbreviation: CHR, chromosome; POS, position; SNPS, single-nucleotide polymorphisms; PVAL, p-value; SE, standard error; REF, reference allele; EFT, effect allele.

Table 1.
 Summary statistics of genetic instruments for telomere length.

3. Estimating telomere length from sequencing reads

3.1 Telomeres in the era of next-generation sequencing

Telomeres are repetitive sequences of $(TTAGGG)_n$ on chromosome ends. In humans, the length of telomeres can vary from 10 to 15 kilobases [19]. With advancements in genome sequencing technology, it is now possible to measure telomere length by applying computational algorithms to sequencing reads, as sequence data from these reads contains information about telomeres, just as they do for other regions of the genome.

DNA sequencing is the process of determining the order of nucleotide bases (T, C, A, G). To sequence a genome, the DNA is typically broken down into small fragments and sequenced in parallel [20]. The resulting sequence reads are then assembled to reconstruct the genome. High-throughput next-generation sequencing (NGS) allows for rapid and scalable sequencing of millions of DNA fragments. NGS can be used to sequence specific regions of the genome, such as protein-coding regions, or the entire genome to identify DNA sequences and somatic mutations. It can also be applied to RNA to measure gene expression.

Whole-genome sequencing (WGS) is a type of next-generation sequencing (NGS) that provides a comprehensive view of the entire genome, including the ends of chromosomes where telomeres are located. Several large-scale sequencing initiatives have been undertaken. These include the Trans-Omics for Precision Medicine (TOPMed) program, which includes over 130,000 whole-genome sequences from more than 80 studies [21], and the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, an international effort to analyze over 2600 cancer whole genomes from the International Cancer Genome Consortium [22]. These projects have generated extensive WGS data integrated with clinical health outcomes on an individual level. Although telomere length was not the primary focus of these initiatives, the wealth of biomarker and health outcome data available makes these studies valuable resources for investigating the potential health impacts of telomere length variation on a population scale.

3.2 Computational tools for estimating telomere length and content

Numerous bioinformatic tools are currently available for estimating telomere length from WGS data. In 2010, Castle et al. first described their approach for estimating telomere content as a proxy for telomere length by counting sequencing reads that contained the repetitive $(TTAGGG)_4$ motif [23]. Since then, several algorithms have been developed and extended to determine telomere content, including Motif_counter [24], TelSeq [25], Computel [26], Telomerecat [27], TelomereHunter [28], Telogator [29], Qmotif [30]. Each tool is based on different algorithms and utilizes different methods to estimate telomere length/content from WGS data. In this section, we will introduce each tool and discuss their validation and application.

Motif_counter (<https://sourceforge.net/projects/motifcounter>) is a bash script that quantifies telomere content by counting sequencing reads that contain telomere motif. It takes Binary Alignment/Map (BAM) formatted files as input and thus requires sequence reads to be aligned to a reference genome prior to running. The motif to be searched for, i.e., the character string of the telomere repetitive region, and the threshold for classifying a read as a telomeric read must be specified. Despite its simplicity and user-friendly command interface, this tool is unable to control

for variations in genome coverage or sequence depth, so additional normalization is needed. The telomere sequence read counts estimated by `motif_counter` correlated well with telomere length measured by TRF (Pearson's $r = 0.855$), after being normalized to genome coverage [31].

TelSeq (<https://github.com/zd1/telseq>) is a C++ software that takes BAM files as input and is specifically designed for estimating telomere length from WGS and could be extended to whole-exome sequencing (WES) data. It searches for telomeric reads of TTAGGG repeats and calculates the mean telomere length from telomeric content. One advantage of TelSeq is that it controls for sequencing biases by normalizing telomere counts based on the percentage of total reads that have a GC-content similar to telomere sequences (48–52%). This is important because a high GC value favors more amplification during PCR, potentially leading to biased estimates. TelSeq has been validated using 260 leukocyte samples from the TwinsUK cohort, where its estimated mean telomere length was compared to TRF estimates [25]. Although TelSeq estimates were consistently shorter than TRF estimates (mean 5.63 kb compared to 6.97 kb), their correlation remained stable across a range of pre-defined numbers of telomeric repeats (Spearman's $\rho = 0.6$). Additionally, TelSeq and TRF estimates had a correlation of 0.78 on exome data, providing promising evidence for its extended application. TelSeq is the most widely used tool and have been applied in several large WGS datasets [32–34].

Computel (<https://github.com/lilit-nersisyan/computel>) is an R program that operates in the Linux environment and takes FASTQ files as input. It identifies telomeric reads by aligning raw sequencing reads to a specially designed telomeric reference, distinguishing it from previous methods that were based on pattern matching. Mean telomere length is then calculated based on the ratio of coverage at the telomeric reference and genomic reference, read length, telomeric pattern length, and the number of chromosomes in a haploid genome. It also allows for telomeric repeat variant analysis to estimate the relative abundance of canonical and variant telomeric repeat patterns. This is important because telomeres may not always contain canonical repeat patterns (TTAGGG)_n if there are variants within telomeric regions. Variant analysis provides necessary information about the distribution of telomeric repeat variants in samples. Computel has been validated with simulated data, where strong and linear correlation was observed between actual and estimated telomere length, and the results suggested that Computel could outperform TelSeq but consistently generated lower telomere length estimates [26].

Telomerecat (<https://github.com/cancerit/telomerecat>) is written in Python and operates in both the Linux and MacOSX environments. It is specifically designed to operate independently of the number of telomeres present in a cell by normalizing telomeric content against subtelomeric regions instead of the entire genome, making it applicable to WGS data from cancer cells. Telomerecat takes BAM files as input and extracts read pairs with at least two instances of telomeric hexamer. The software then classifies read pairs based on their sequence composition and orientation. Telomere length is calculated using the ratio of complete to boundary read-pairs, along with the insert length distribution. Telomerecat has been validated in 260 adult females from the TwinsUK10K study, showing a significant correlation with TRF estimates (Spearman's $\rho = 0.618$) and TelSeq ($\rho = 0.631$) [27].

TelomereHunter (<https://pypi.org/project/telomerehunter>) is a Python software designed to estimate telomeric content from WGS data of matched tumor and normal tissue control pairs within the same individual. The program accepts BAM files as input, selects reads with high telomere repeats, and organizes them by mapping their

position into categories such as intrachromosomal, subtelomeric, junction spanning, and intratelomeric reads. Telomere content can be derived from the intratelomeric reads. Telomere content can be calculated from intratelomeric reads. TelomereHunter has been validated by strong correlations between its estimated telomere content and telomere length measures from qPCR (Pearson's $r = 0.94$) and TRF (Pearson's $r = 0.72$) after GC correction [28].

Telogator (<https://github.com/zstephens/telogator>) is a Python software designed to estimate chromosome-specific telomere length from long reads. This recently developed tool is built on long-read telomere analysis and the newest human reference genome by the Telomere-to-Telomere (T2T) consortium [35]. Telogator takes long reads in FASTA or FASTQ format, performs alignment, and extracts reads mapped to subtelomeres and telomeres. It then identifies telomere regions, clusters reads by their telomere-subtelomere boundaries, and reports telomere length for specific chromosome arms. High correlation (Pearson's $r = 0.91$) has been reported between Telogator and TelomereHunter when comparing the averaged chromosome-specific telomere lengths against the telomere content from TelomereHunter [29].

Qmotif (<https://github.com/AdamaJava/adamajava>) is a Java-based software designed to estimate telomere content from WGS data in a fast and efficient manner. It takes BAM files as input and searches for user-defined motifs using a two-pass matching system. In the first stage, a quick string matching is performed to filter strings into the second stage for regular expression (regex) matching. For telomere quantification, the first stage matches 3 consecutive repeats of the canonical telomere motif (TTAGGG), while the second stage uses regex to match any 2 adjacent repeats of the motif with variation allowed in the first 3 base pairs. The runtime can be further sped up by instructing the algorithm to search for telomere repeats in regions of the genome most likely to contain them. Qmotif has been validated by comparison with qPCR (Spearman's $\rho = 0.69$) and other computational tools such as TelSeq (Spearman's $\rho = 0.99$) and TelomereHunter (Spearman's $\rho = 0.85$), with a much faster runtime of under 1 minute on the same set of samples (compared to 1–7 hours for TelSeq and 4–19 hours for TelomereHunter) [30].

4. Sequencing based estimates for population genetic studies of telomere length

TelSeq has been used in several large WGS datasets, including the TOPMed study, which is one of the largest population-based studies with available WGS information. Telomere length was estimated using TelSeq in 109,122 TOPMed participants of diverse ancestry, including European, African, Asian, and Hispanic/Latino [33]. A strong correlation was observed between TelSeq estimates and TRF and flow-FISH in a subset of samples ($r = 0.68$ and 0.80 , respectively). The results of GWAS on TelSeq-estimated telomere length were compared to those of GWAS on qPCR estimates and showed great consistency in effect estimates of the identified genetic predictors: Pearson's $r = 0.92$ for 37 overlapping variants with 23,096 Chinese Singaporeans and $r = 0.86$ for 43 overlapping variants with 78,592 Europeans. A genetic correlation of 0.81 between TelSeq and qPCR-estimated telomere length suggests that these methods share a high degree of genetic determinants.

In another study, TelSeq was applied to WES data from the UK Biobank, a population cohort of more than 500,000 UK adult residents [36]. Telomere length was estimated using TelSeq in 49,738 participants and compared to qPCR estimates

from 472,594 participants within the UK Biobank and WGS estimates from 63,302 TOPMed participants using TelSeq. The WES-based telomere length (mean of 0.83 kb) was found to be shorter than the qPCR measures within the same population and even shorter than the WGS estimates in TOPMed (mean of 3.27 kb). Additionally, when estimating the effect sizes of SNPs in predicting telomere length, WES data showed the lowest correlation with SNPs and a deflation in effect size estimates. Since WES restricts sequencing to targeted regions of the genome, estimating telomere length from WES data may require further correction and adjustment.

5. Conclusions

In conclusion, genetic variants play an important role in determining leukocyte telomere length, enabling the estimation of an individual's telomere length through genotyping data. Furthermore, next-generation sequencing data offers an alternative to qPCR for measuring telomere length in large-scale population studies. The ability to estimate absolute telomere length in base pairs also facilitates the comparison of results across different studies. The use of sequencing-based telomere length estimates holds great promise for genetic studies, particularly given the likely overlap between sources of genetic variant and sequencing data.

Acknowledgements

This work was supported by National Cancer Institute grant U01ES029520.

Conflict of interest


The authors declare no conflict of interest.

Author details

Ting Zhai and Zachary D. Nagel*
John B. Little Center for Radiation Sciences, Department of Environmental Health,
Harvard T.H. Chan School of Public Health, Boston, USA

*Address all correspondence to: znagel@hsph.harvard.edu

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Demanelis K, Jasmine F, Chen LS, Chernoff M, Tong L, Delgado D, et al. Determinants of telomere length across human tissues. *Science*. 2020;**369**(6509):eaaz6876
- [2] Andreu-Sanchez S, Aubert G, Ripoll-Cladellas A, Henkelman S, Zhernakova DV, Sinha T, et al. Genetic, parental and lifestyle factors influence telomere length. *Communications Biology*. 2022;**5**(1):565
- [3] Haycock PC, Heydon EE, Kaptoge S, Butterworth AS, Thompson A, Willeit P. Leucocyte telomere length and risk of cardiovascular disease: Systematic review and meta-analysis. *BMJ*. 2014;**349**:g4227
- [4] Telomeres Mendelian Randomization C, Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, et al. Association between telomere length and risk of cancer and non-neoplastic Diseases: A Mendelian randomization study. *JAMA Oncology*. 2017;**3**(5):636-651
- [5] Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, et al. Measurement of telomere length by the southern blot analysis of terminal restriction fragment lengths. *Nature Protocols*. 2010;**5**(9):1596-1607
- [6] Baerlocher GM, Vulto I, de Jong G, Lansdorp PM. Flow cytometry and FISH to measure the average length of telomeres (flow FISH). *Nature Protocols*. 2006;**1**(5):2365-2376
- [7] Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Research*. 2009;**37**(3):e21
- [8] Telomere Research Network. Available from: <https://trn.tulane.edu/>
- [9] Holohan B, Wright WE, Shay JW. Cell biology of disease: Telomeropathies: An emerging spectrum disorder. *The Journal of Cell Biology*. 2014;**205**(3):289-299
- [10] Codd V, Wang Q, Allara E, Musicha C, Kaptoge S, Stoma S, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nature Genetics*. 2021;**53**(10):1425-1433
- [11] Chang X, Gurung RL, Wang L, Jin A, Li Z, Wang R, et al. Low frequency variants associated with leukocyte telomere length in the Singapore Chinese population. *Communications Biology*. 2021;**4**(1):519
- [12] Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ*. 2018;**362**:k601
- [13] de Leeuw C, Savage J, Bucur IG, Heskes T, Posthuma D. Understanding the assumptions underlying Mendelian randomization. *European Journal of Human Genetics*. 2022;**30**(6):653-660
- [14] Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nature Genetics*. 2013;**45**(4):422-427
- [15] Gu J, Chen M, Shete S, Amos CI, Kamat A, Ye Y, et al. A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. *Cancer Prevention Research (Philadelphia, Pa.)*. 2011;**4**(4):514-521
- [16] Lee JH, Cheng R, Honig LS, Feitosa M, Kammerer CM, Kang MS,

- et al. Genome wide association and linkage analyses identified three loci-4q25, 17q23.2, and 10q11.21-associated with variation in leukocyte telomere length: The long life family study. *Frontiers in Genetics*. 2013;**4**:310
- [17] Li C, Stoma S, Lotta LA, Warner S, Albrecht E, Allione A, et al. Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *American Journal of Human Genetics*. 2020;**106**(3):389-404
- [18] Prescott J, Kraft P, Chasman DI, Savage SA, Mirabello L, Berndt SI, et al. Genome-wide association study of relative telomere length. *PLoS One*. 2011;**6**(5):e19635
- [19] Srinivas N, Rachakonda S, Kumar R. Telomeres and telomere length: A general overview. *Cancers (Basel)*. 2020;**12**(3):558
- [20] van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends in Genetics*. 2014;**30**(9):418-426
- [21] Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature*. 2021;**590**(7845):290-299
- [22] Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature*. 2020;**578**(7793):82-93
- [23] Castle JC, Biery M, Bouzek H, Xie T, Chen R, Misura K, et al. DNA copy number, including telomeres and mitochondria, assayed using next-generation sequencing. *BMC Genomics*. 2010;**11**:244
- [24] Conomos D, Stutz MD, Hills M, Neumann AA, Bryan TM, Reddel RR, et al. Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *The Journal of Cell Biology*. 2012;**199**(6):893-906
- [25] Ding Z, Mangino M, Aviv A, Spector T, Durbin R, Consortium UK. Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*. 2014;**42**(9):e75
- [26] Nersisyan L, Arakelyan A. Computel: Computation of mean telomere length from whole-genome next-generation sequencing data. *PLoS One*. 2015;**10**(4):e0125201
- [27] Farmery JHR, Smith ML, Diseases NB-R, Lynch AG. Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Scientific Reports*. 2018;**8**(1):1300
- [28] Feuerbach L, Sieverling L, Deeg KI, Ginsbach P, Hutter B, Buchhalter I, et al. TelomereHunter - in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics*. 2019;**20**(1):272
- [29] Stephens Z, Ferrer A, Boardman L, Iyer RK, Kocher JA. Telogator: A method for reporting chromosome-specific telomere lengths from long reads. *Bioinformatics*. 2022;**38**(7):1788-1793
- [30] Holmes O, Nones K, Tang YH, Loffler KA, Lee M, Patch AM, et al. Qmotif: Determination of telomere content from whole-genome sequence data. *Bioinformatics Advances*. 2022;**2**(1):vbac005
- [31] Lee M, Hills M, Conomos D, Stutz MD, Dagg RA, Lau LM, et al. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Research*. 2014;**42**(3):1733-1746

[32] Zheng S, Cherniack AD, Dewal N, Moffitt RA, Danilova L, Murray BA, et al. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell*. 2016;**29**(5):723-736

[33] Taub MA, Conomos MP, Keener R, Iyer KR, Weinstock JS, Yanek LR, et al. Genetic determinants of telomere length from 109,122 ancestrally diverse whole-genome sequences in TOPMed. *Cell Genomics*. 2022;**2**(1):100084

[34] Barthel FP, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nature Genetics*. 2017;**49**(3):349-357

[35] Nurk S, Koren S, Rhie A, Rautiainen M, Bizikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. 2022;**376**(6588):44-53

[36] Nakao T, Bick AG, Taub MA, Zekavat SM, Uddin MM, Niroula A, et al. Mendelian randomization supports bidirectional causality between telomere length and clonal hematopoiesis of indeterminate potential. *Science Advances*. 2022;**8**(14):eabl6579