TelOMpy enables single-molecule resolution of telomere length from optical genome mapping data

Ivan Pokrovac, Sunčica Stipoljev and Željka Pezer

Abstract

Existing methods for assessing telomere length are limited in various ways, including resolution, scope, throughput, and reproducibility. These limitations stem from the applicability of laborious and complicated indirect approaches. We developed TelOMpy, a bioinformatic tool for determining lengths of individual telomeres from optical genome mapping data. TelOMpy directly measures the absolute lengths of thousands of telomeres within a single sample, eliminating errors associated with indirect methodologies and ensuring high analytical power and reproducibility of results. We use optical genome mapping data generated from mice in a high-fat diet experiment to demonstrate the power of TelOMpy performance at a single-molecule level. Moreover, using our newly developed protocol for high-molecular weight DNA extraction, we produced the first optical genome maps from sperm and compared telomere lengths between sperm and somatic tissue in overweight mice. TelOMpy reveals immense variation in length of individual telomeres and the effect of high-fat diet on telomere shortening and loss.

Keywords: telomere, telomere shortening, optical genome mapping, high-fat diet, overweight

Introduction

Telomeres are specialized structures on chromosome ends that ensure chromosome integrity. At a nucleotide level, they are formed by stretches of tandemly repeated short DNA motif. The lengths of these stretches are variable and relevant for aging-related processes and diseases (Blackburn et al., 2015). Multiple methods for assessing telomere length are available. One of the most used techniques is quantitative PCR (qPCR; Cawthon, 2002), which requires small DNA quantity and a simple and cheap procedure for sample preparation. Specifically, the qPCR-based methods determine the average telomere length between DNA molecules present in the qPCR reaction, relative to a reference sample. A major downside of this approach is low reproducibility between studies (Aviv et al., 2011; Martin-Ruiz et al., 2015). Terminal restriction fragment (TRF) is another method for measuring average telomere length in a population of cells that is insensitive to very short telomeres (Montpetit et al., 2014). On the other hand,

STELA (Single-telomere length analysis) and Q-FISH (quantitative fluorescence in situ hybridization) are methods of choice for assessing the length of single telomeres with a resolution of 0.1 kilobase pairs (kbp). However, while the former is usually limited to telomeres of a few chromosomes, only cells in metaphase can be assayed with the latter (Aubert et al., 2012; Montpetit et al., 2014). While qPCR-based methods can be scaled up to screen hundreds of samples, essentially all these methods are low-throughput.

Optical genome mapping (OGM, Bionano Genomics) is a high-throughput genome analysis technology that uses imaging to construct whole genome maps. High molecular weight DNA (HMW DNA) is fluorescently labeled using the DLE-1 enzyme on the CTTAAG motif, which is repeated in the human and mouse genome on average every 10-15 kbp. Long DNA molecules are stretched through nanochannels under an electric field, and fluorescent signals are detected at labeled positions (Yuan et al., 2020). The imaged molecules are digitalized, and the alignment of molecules is performed based on the pattern of DLE-1 motif positions (*labels*) to create consensus maps (CMAP) of genomic regions. These maps can be further used to construct genomes *de novo* and to detect genomic structural variants based on comparison with the *in-silico* labeled reference genome. Being a single-molecule technique, OGM enables the direct study of individual DNA molecules without the need for fragmentation or amplification, which eliminates errors related to these procedures. However, given that the result of OGM are genome maps of labeled motif, information on nucleotide sequence is not available from the obtained data.

We present TelOMpy, a bioinformatic tool that measures absolute lengths of individual telomeres from optical genome mapping data. Within a single sample, TelOMpy directly calculates lengths of thousands of telomeres, from long high-quality molecules aligned to chromosome ends of a reference genome. TelOMpy can be applied to any species for which chromosome-level genome assembly is available. We demonstrate its usage on mice in a high-fat experiment. Overweight and obesity are known to be associated with telomere shortening in somatic and reproductive cells (Valdes et al., 2005; Gielen et al., 2018; Clemente et al., 2019; Yang et al., 2016; Raee et al., 2023), and this association may have harmful consequences for offspring, through maternal or paternal effect (Yang et al., 2015; Yang et al., 2016; Wei et al.; 2021). It is therefore of interest to study the effect of diet on telomere length in both somatic and reproductive cells. As part of this research, we developed a protocol for the extraction of high molecular weight (HMW) DNA from mouse sperm. It allowed us to perform the first OGM on sperm cells and apply TelOMpy to calculate lengths of thousands of telomeres and to compare them with telomeres from kidney as a representative of somatic cells. Our results illustrate the ability of TelOMpy to perform high-throughput measurements of telomere lengths at an unprecedented resolution, enabling the detection of subtle differences. Our study highlights the tremendous variation of telomere length between and within individuals and chromosomes and the effect of high-fat diet on telomere shortening and loss.

Results

TelOMPy overview

Optical mapping is a method based on single molecules, which entails the capacity to determine the lengths of individual telomeres. These are calculated from the molecules that are mapped to the end label on a chromosome arm in the reference genome, as the distance between it and the end of the molecule. This calculation of telomere length is based on the absence of a DLE-1 motif in telomeres, as they typically consist of a tandemly repeated TTAGGG motif. In addition to the telomeric sequence, such measurement includes a part of the subtelomeric region with no DLE-1 motif. Hence, the length of non-telomeric part can be calculated as the distance from the last aligned label on the reference genome to the start position of the annotated telomere (Figure 1).

TelOMpy takes as input the results of *de novo* genome assembly generated by the OGM and determines telomere length in four steps (Figure 1): I) It first identifies contig(s) aligned to the ends of a chromosome and II) identifies *telomeric molecules*, i.e. molecules that are aligned to the telomere end of these contigs. Next, III) the alignments are checked for suitability by counting the number of the most distally positioned labels that are continuously unaligned and by measuring the distance between the annotated chromosome end and the closest label. Ideally, the most distal label on the reference chromosome should also be the last aligned label. One unaligned label position is tolerated on contigs or molecules to account for the possibility of unspecific labeling or mutations, which may create a DLE-1 motif in telomere repeats. In addition, the larger the distance between the chromosome end and the closest label on the reference genome, the more unlikely it will be bridged by the molecules in OGM. Therefore, in such cases, the most distally aligned molecules are not those of telomeres but rather represent the last genomic region where alignment was still possible, adjacent to a genomic “desert” without labels. Chromosome arms that are suitable for analysis are retained, and in the last step, IV) their telomere lengths are calculated by applying one of the two equations on every individual telomeric molecule, depending on the chromosome arm and the orientations of the alignments (see Methods).

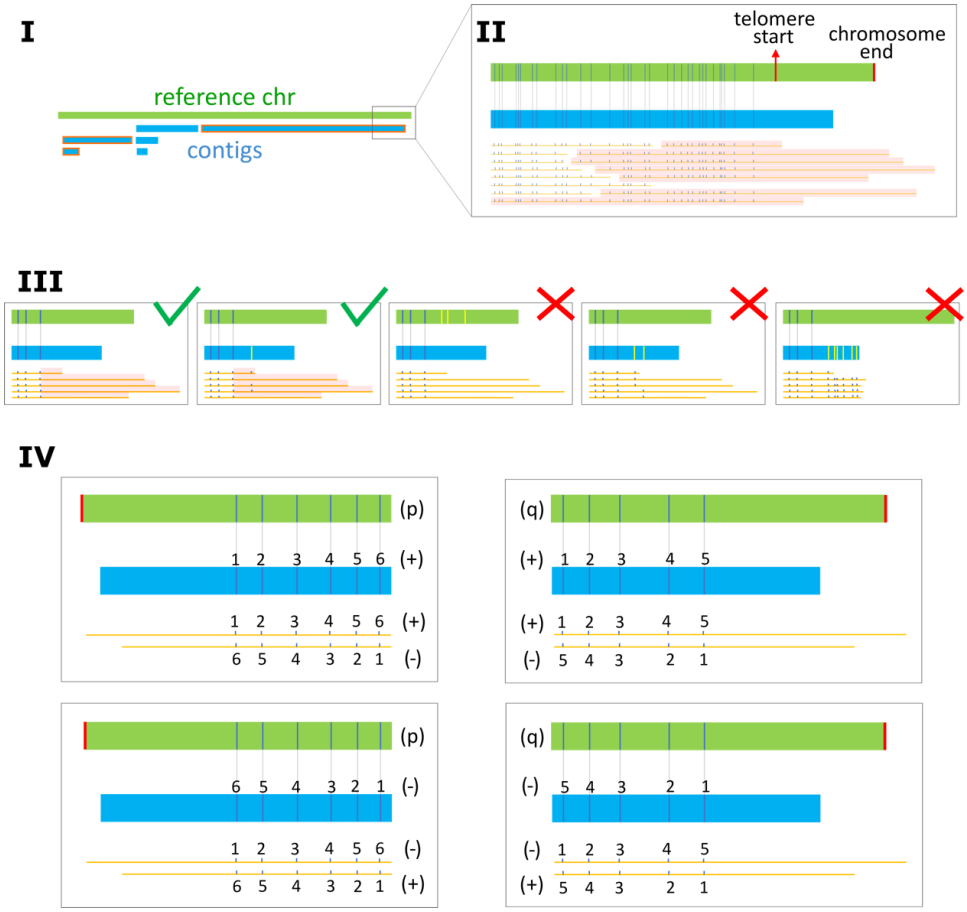

Figure 1. TelOMpy workflow. Throughout the figure, reference chromosome is depicted by green and contigs by blue rectangles; molecules are shown as orange horizontal lines with dark blue ticks as labels, labels on contigs and reference are shown by vertical dark blue (aligned) and yellow (unaligned) lines; alignments between labels on reference and contig are indicated by connecting light gray vertical lines. I) Contigs aligned to chromosome ends are identified (outlined in red) from the OGM *de novo* assembly. II) Telomeric molecules (highlighted in light red) are identified as molecules that are aligned to the telomere end of these contigs. Only a long chromosome arm telomere is illustrated, but telomeres on both chromosome arms are considered. III) Alignments on chromosome ends are checked, and only those of the highest quality (see Methods) are retained for analysis. Individual telomeres are highlighted in light red as part of telomeric molecule after the outermost aligned label. IV) Telomere lengths are calculated by applying a proper equation on each telomeric molecule, based on the chromosome arm and orientations of molecule and contig. The most distal parts of p-arm (left) and q-arm (right) are sketched with all possible combinations of orientations of molecules and contigs. Position of chromosome end on the reference genome is shown by a red vertical line. Orientation of molecule and contig is indicated by "+" and "-" sign, and illustrated by ascending or descending label number in telomeric molecules.

Newly established protocol yields high-quality DNA from sperm cells suitable for optical genome mapping

To test the performance of TelOMpy and further examine the link between overweight/obesity and telomere length in somatic and sperm cells, we set up a high-fat diet experiment on male mice. At a prepubertal age, mice were put on a diet rich in fat, cholesterol, and sugar (Western diet group, WD) for 80 days. In parallel, mice of the same age were kept on a standard diet during the whole experiment (control group, C). Although there was no significant difference in weight between the two groups at the beginning of the experiment (Mann Whitney test, p value 0.58), mice in the WD group were expectedly heavier at the end of the experiment (Mann Whitney test, p value 0.0086, Cohen's D of 1.9, CLES 0.97), indicating the effect of high-fat diet (Figure 2A).

Kidney tissue and seminal fluid were dissected from each mouse. High-molecular weight (HMW) DNA was extracted from kidney cells by using standard Bionano protocol (see Methods). For the extraction of HMW DNA from sperm cells, we modified the procedure by adding dithiothreitol (DTT), which acts as a reducing agent by breaking the disulfide bonds in protamines, proteins that are specifically present in sperm chromatin (Balhorn, 2007). A portion of the isolated HMW DNA was digested with a restriction enzyme to establish the efficiency of protein removal. The success of digestion and the integrity of isolated DNA were checked by pulsed-field gel electrophoresis (PFGE). The size of the DNA before digestion was mainly above 1 Mbp (Figure 2B). After digestion, the length of DNA did not exceed 1 Mbp, and a significant amount was below 48 kbp. We concluded that the HMW DNA obtained from sperm cells with the established protocol is in the size range that meets the requirements for OGM (minimum size of 150 kb), and free from proteins, thus accessible for labeling with the DLE-1 enzyme.

The optical genome mapping was performed on HMW DNA isolated from kidney and sperm cells of five mice from the experimental group and four mice from the control group. DNA concentration after DLS labeling ranged within the Bionano's recommendations (Table S1, Supplementary Material).

Quality of optical genome mapping data

To establish how the newly developed protocol for HMW DNA extraction affects the quality of molecules and the *de novo* genome assembly, we compared the data between sperm and kidney samples. Multiple metrics indicate high quality of the resulting OGM data after the *de novo* assembly (Table S2, Supplementary Material). All samples meet the recommended quality criteria for the average length of molecules, effective coverage of the reference, and the fraction of molecules aligned. Moreover, in all samples, diploid genome map N50 was much above the recommended 50 Mbp minimum, indicating high continuity of the assembly.

DNA molecules that enter OGM are filtered during the *de novo* genome assembly pipeline: those shorter than 150 kbp and containing less than 9 labels are considered low-quality and are thus discarded. We compared the number of molecules before and after the removal of low-quality molecules. Although the initial number of molecules is similar in both tissues, more are filtered out in sperm samples, resulting in approximately three times more molecules aligned to the reference genome in kidney than in sperm (Figure 2C). The more extensive removal of initial molecules in sperm samples resulted in differences in the effective coverage of the reference genome, which is significantly lower for sperm than for kidney samples (paired t-test, p value 0.0007, Cohen's D 2.53); however, still above the recommended coverage minimum (Table S2, Supplementary Material). Although the average length of molecules isolated from sperm was shorter than that of kidneys, this difference disappeared after the removal of low-quality molecules (Figure 2D).

In conclusion, compared to the standard Bionano protocol for HMW DNA isolation from somatic tissue, the newly developed procedure for isolation from sperm yields initially lower quality of molecules on average, yet after filtering, the remaining number of molecules is sufficient to achieve high-quality *de novo* genome assembly.

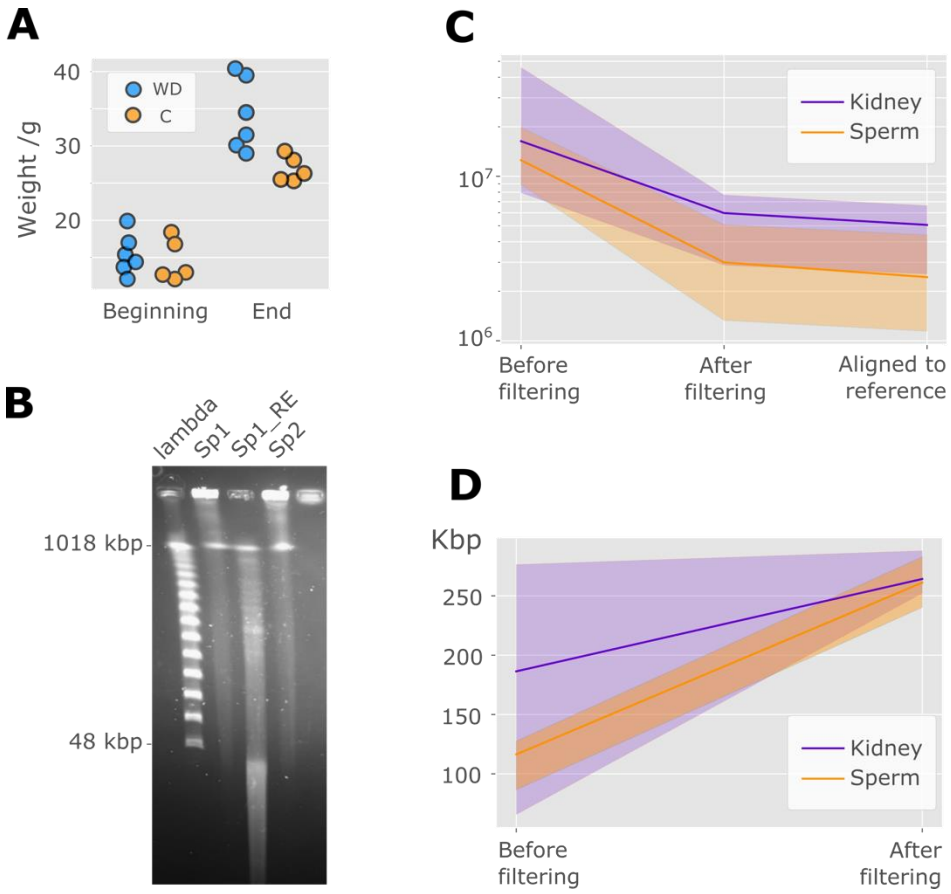

Figure 2. DNA and data quality indicators. A) Weight of mice at the beginning and end of the experiment. B) An example of PFGE results. Lambda is PFG Ladder with fragment range from 48.5 kbp to 1018 kbp. Lanes labeled *Sp1* and *Sp2* contain HMW DNA isolated from approximately 2.5 million sperm cells each. Lane *Sp1_RE* contains *Sp1* digested with the EcoR1 enzyme. C) Number of molecules produced by OGM at various stages of *de novo* genome assembly process. D) Average length of molecules before and after filtering. In both C) and D), lines represent mean values, and shading shows the range of values between samples.

Data suitability

TelOMpy was applied on OGM data generated from kidney and sperm of mice in the Western diet experiment. As described above and in the Methods section, information on telomere length is only available from high-quality alignments of molecules and contigs with the reference genome at chromosome ends (Figure 1). Due to this requirement, it was not possible to determine the telomere lengths on the short (p) arm of almost every chromosome in all samples (Figure 3A). Mouse chromosomes are acrocentric, containing regions of unknown sequence and length between the telomere on the p-arm and the centromere. These regions are annotated in the reference genome as gaps of arbitrarily chosen length of 10 kbp. In our samples, they are devoid of labels corresponding to DLE-1 enzyme, and molecules obtained by OGM are not sufficiently long to bridge the p-arm telomere and centromere regions and provide information on telomere length. We encountered a similar problem on the long (q) arm of the sex chromosomes and chromosome 4. Additionally, chromosomes 9 and 14 of some samples had one or more unaligned labels on the reference after the last aligned label (Figure 3A). Finally, for these reasons, all short chromosome arms as well as long arms of sex chromosomes and chromosomes 4, 9, and 14 were excluded from subsequent analyses of telomere length in all samples. We extracted telomeric molecules for the remaining chromosome arms. To estimate the degree of their alignment in subtelomeric regions, we calculated the proportion of molecules with different numbers of unaligned labels at distal positions. About 90% of telomeric molecules support contigs whose outermost label is aligned with the outermost label on the reference genome, and approximately 10% are aligned to contigs with single unaligned label at the most distal position. In more than 80% of all telomeric molecules, the outermost label was aligned with the outermost label on the reference, and about 98% had less than four continuously unaligned labels at distal positions (Figure 3B). These data suggest that the degree of alignment of molecules at the end of the q-arm of most chromosomes is suitable for telomere length analyses.

The calculated lengths of telomeres in our dataset may not reflect true lengths but rather shorter artefacts that stem from degradation processes that may have occurred at chromosome ends during cell lysis. To control for this possibility, we compare the lengths of telomeric molecules with the lengths of all molecules that enter the alignment process. We found no significant correlation between telomere length and the length of molecules from which telomere length is calculated (Pearson's $R=0.24$; Figure 3C). Similarly, we found no significant difference between the length of molecules carrying telomeres and the average length of all molecules aligned to the reference genome (Cohen's $D < 0.3$; Figure 3D). These results indicate that the molecule length in our data set does not have a significant effect on the determination of telomere length.

Considering that the telomere length from the optical mapping data is calculated from the last DLE-1 motif to the end of the molecule, each measured telomere length also contains a part of the subtelomeric region between the last motif DLE-1 and the beginning of the annotated telomere on the reference genome. This part varies between chromosomes on the mouse reference genome and amounts between tens and thousands of base pairs (Supplementary Figure S1) and can be considered negligible compared to the average mouse telomere length of 50 kbp (Hemann, 2000). These values can be subtracted from the calculated telomere lengths to obtain corrected absolute lengths. However, their inclusion does not affect the comparison of telomere lengths between samples since this factor is equally present in calculations for all samples.

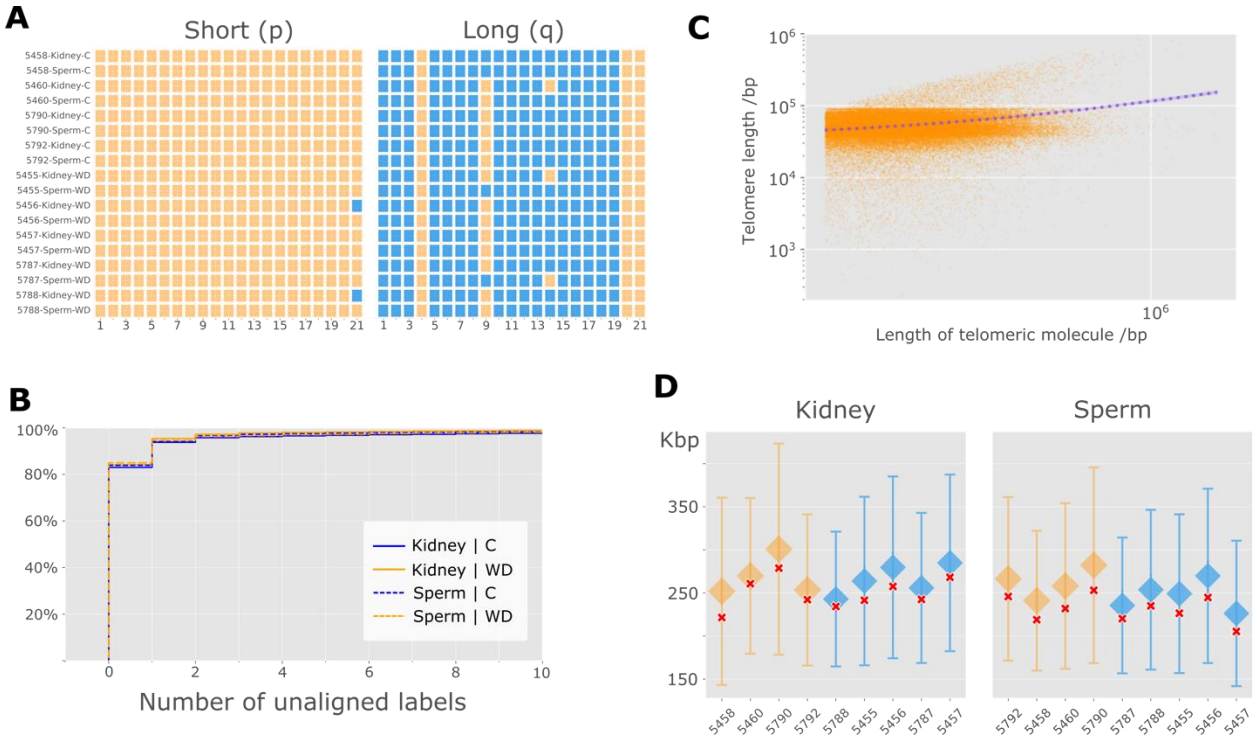

Figure 3. Data suitability for analysis of telomere length. A) Overview of chromosome arms suitable for telomere length determination by sample. Chromosome arms at which alignments do not meet the quality criteria are marked in orange, and chromosome arms at which alignments are suitable for analysis are marked in blue. Information for autosomes 1-19 and sex chromosomes (designated 20 and 21) are shown by columns for each sample per row, on p and q chromosome arms. B) Cumulative empirical distribution for the number of continuously unaligned labels at distal positions on the molecule. C) Correlation between the length of molecules carrying telomeres and the calculated telomere lengths. D) Distribution of the length of telomeric molecules per mouse in kidney (left) and sperm (right) for control (orange) and experimental group (blue). Mouse identification numbers are shown on the x-axis. Diamond shapes indicate the average length of telomeric molecules, and caps indicate one standard deviation of the average length. The red sign x indicates the average length of all molecules aligned to the reference genome.

Differences between experimental and control group

A total of 70,215 telomeres were compared in our data set, on average 3,900 per sample (range 1,179-7,869). On average, we find a three-fold lower number of telomeres in sperm than in kidneys. This can be only partially explained by the overall lower genome coverage in sperm samples, given that the normalization for the number of aligned molecules only somewhat reduces the difference (Figure 4A). In addition, within both tissues, we found a smaller number of telomeres in mice from the experimental group. This difference is even more pronounced when we normalize the number of telomeres by dividing it by the total number of molecules. These results suggest that lower telomere count in the WD group is a result of the experimental effect rather than a technical artefact (Figure 4A).

The measured length of individual telomeres varies between 236 bp and 840 kbp, with an average of 55 kbp in the dataset. Telomeres are on average shorter in mice from the experimental group (Figure 4B), by 2.6 kbp (median 1.3 kbp) in kidney tissue (T test, p value 4×10^{-25}) and by 1.2 kbp (median 2.5 kbp) in sperm (T test, p value 4×10^{-19}). However, given the large variation, this effect is small for both kidney tissue (Cohen's D = 0.08) and sperm (Cohen's D = 0.04). Despite the wide range of telomere lengths within the same chromosome, we observe significant differences in average telomere length between chromosomes, with 43 kbp as the shortest in chromosome 10 and 74 kbp as the longest in chromosome 11 (Figure S2).

To determine whether the high-fat diet induces specific telomere shortening in sperm cells, we compared the distribution of telomere lengths from sperm with the distribution of telomere lengths from the kidneys of the same individual. The distribution is shifted towards lower values in sperm compared to kidneys in all individuals, which indicates a general shortening of telomeres in sperm and/or their lengthening in kidney in both groups. There was no significant difference in this effect between the experimental and control groups: in the control group (N=4) the average length difference was 7.8 ± 4.4 kbp (median difference 6.8 ± 4.2 kbp), and in the experimental group (N=5) 6.1 ± 3.1 kbp (median difference 6.0 ± 2.8 kbp) (Figure S3).

It is known that critically short telomeres can affect the replicative potential of the cell and cause genomic instability, apoptosis, premature aging, and a shorter lifespan (Hemann et al., 2001). Hence, while the average telomere length may not differ significantly, the shortest telomeres may cause dysfunction and are thus of higher relevance for the study of telomere length. To analyze in more detail the differences in the length of critically short telomeres, we extracted the first 25% of the shortest telomeres (first quartile of length, Q1) within each chromosome of each individual to account for the significant variation in telomere length between individuals and between chromosomes. Within the first quartile of length, telomeres are shorter in the sperm of the experimental group compared to the control group (T test, p-value 9×10^{-26}; Cohen's D = 0.31) by 4.3 kbp on average (median 4.4 kbp). This difference in kidney tissue amounts to 1 kb and is not statistically significant (Table 1, Figure 4C). The difference in average telomere length in the first quartile of length between kidney tissue and sperm in the control group (N=4) was 10.1 ± 7.5 kbp (median 10.0 ± 7.2 kbp), and in the experimental group (N=5) by 12.1 ± 6.6 kbp (median 12.5 ± 7.4 kbp) (Figure S4).

Table 1. Descriptive statistics of telomere length (in Kbp).

Group	Tissue	Average	Average STDEV**	Median	Average IQR**
C	Kidney	38.7	16.0	38.4	17.8
WD	Kidney	37.7	15.5	38.8	18.1
C	Sperm	30.0 *	14.9	30.0	19.4
WD	Sperm	25.7 *	13.1	25.6	17.4

* Statistically significant difference between WD and C (Mann-Whitney test, P value < 0.05; Cohen D > 0.3)
** IQR - interquartile range; STDEV - standard deviation

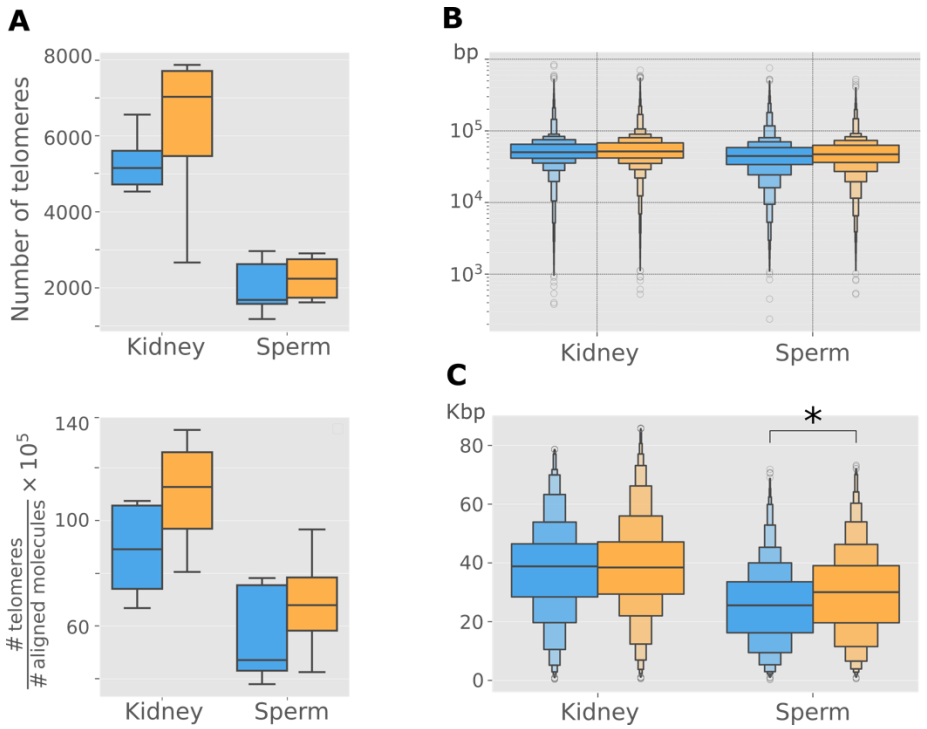

Figure 4. Comparison of telomere length between experimental and control group. Throughout the figure, the WD group is shown in blue and the C group in orange. A) Telomere count (top) and the normalized number of telomeres (bottom). This normalized number corresponds to the number of telomeric molecules per 100,000 molecules aligned in the genome. B) Distribution of telomere lengths shown by "boxen" diagram by tissue. C) Distribution of telomere lengths in Q1. The horizontal line in the widest part of each of the shown distributions in B) and C) represents the median value.

Validation of TelOMpy calculated telomere lengths

The wide range of telomere sizes and the average length determined by TelOMpy in our dataset are consistent with previous studies which used other methods to show comparable average and variation of telomere lengths in the C57BL/6 strain (Zijlmans et al. 1997; Hemann, 2000; Zhang et al., 2025), indicating a general reliability of TelOMpy.

To further validate our results obtained by TelOMpy, we performed qPCR assessment of telomere length (Cawthon, 2002). For this purpose, a commercial kit was used to extract genomic DNA from samples of kidney and sperm from the same samples of mice in the described high-fat diet experiment, and additional 12 mice from an extended experiment (see Methods). For each biological sample, the isolated genomic DNA was used as a template in singleplex qPCR reactions with primer pairs specific either to telomere (T) or albumin gene, as a single-copy reference gene (S). Relative telomere length was expressed as the ratio of telomere copy number to the reference gene number (T/S). Our qPCR experiments confirmed the shorter telomeres in sperm compared to kidney (two-tailed paired t-test; p-val = 2.3×10^{-3}) and no difference in telomere length between the control and experimental group (two-tailed unpaired t-test; pval = 0.84) (Figure 5).

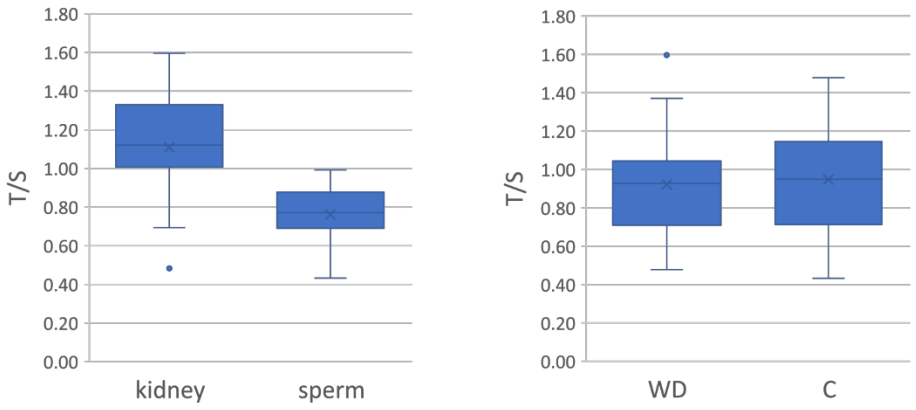

Figure 5. Validation of TelOMpy results by qPCR. Telomere lengths relative to a control sample are presented as T/S ratio in comparison between tissues (left) and groups (right).

Discussion

Telomere length at a single-molecule level

In this study, we present a newly developed tool TelOMpy, for calculating individual telomere lengths from optical genome mapping data. TelOMpy calculates the absolute lengths of individual telomeres on individual chromosomes, which is a major advantage compared to existing methods for determining telomere length. Two methods for determining individual telomere lengths from optical mapping data have been described so far. One of them requires the CRISPR/Cas9 system to mark the telomeric motif, which creates additional complications and costs in implementation (McCaffrey et al., 2017). The other is not described in sufficient detail to be reproducible, nor has it been implemented in publicly available software (Young et al., 2017). Unlike the mentioned methods, TelOMpy is a publicly available tool for determining a wide range of telomere lengths from optical genome maps, which does not require the use of additional technology. TelOMpy can be implemented at no additional cost in any research involving standard *de novo* genome assembly procedures and structural variant detection from OGM data. For example, OGM is often used to investigate genomic rearrangements in tumor tissues, in which changes in telomere length are part of cancerous processes (Fan et al., 2021).

We demonstrate TelOMpy usage in a high-fat diet experiment on mice, on OGM data produced from kidney and sperm. We find that telomeres vary in length over three orders of magnitude - from only a few hundred to several hundred thousand base pairs. Such variability exists between and within individuals and tissues and between and within chromosomes and is consistent with telomere length polymorphism previously demonstrated in human somatic cells (Lansdorp et al., 1996; Schmidt et al., 2024). Our results show that, on average, sperm telomeres are significantly shorter in the experimental group than in the control group, which is the detection power at the level of the qPCR measurement of telomere length. This result would suggest that a high-fat diet is associated with telomere shortening in sperm. However, TelOMpy enabled more detailed comparisons - at the level of individual chromosomes between kidney and sperm within a single mouse, which do not support this conclusion because no consistent shortening of telomeres exist exclusively in the sperm of experimental mice. These results demonstrate the unparalleled power and resolution of TelOMpy in analyses of telomere length.

Numerous studies have witnessed longer telomeres in sperm than somatic cells, due to the activity of telomerase in spermatogenic cells, which lengthens telomeres (Fice and Robaire, 2019). However, in all analyzed mice in our study, we find that telomeres are consistently shorter in sperm than in kidney. Our control analyses suggest that these results are not technical artefacts but rather represent a true biological state. We further confirm this finding by qPCR on an extended sample set, using a different method for genomic DNA extraction. Although rarely, telomere shortening in sperm has been reported previously as a result of aging (de Frutos et al., 2016) or infertility (Tahamtan et al., 2019). In addition, it has been shown that mutations in various genes can lead to a decrease in telomerase activity and consequent telomere shortening in sperm (Margalef et al., 2018; Mirabello et al., 2010; Soerensen et al., 2012; Bojesen et al., 2013; Grill and Nandakumar, 2021). The relatively shorter telomere length in sperm compared to kidneys in our samples may be a result of mutation in one or more genes among the many involved in telomere maintenance. Since OGM data does not provide information on nucleotide sequence, it was not possible to perform analyses at a sequence level. Nonetheless, mutations that would diminish or prevent

telomerase activity are possible and would not necessarily be detrimental, as telomerase-deficient mice can be viable, fertile, and without visible morphological abnormalities (Blasco et al., 1997). Alternatively, the relatively longer telomeres in kidney than in sperm could result from telomerase activity in kidney cells. Mice are known to express telomerase in some somatic tissues (Prowse and Greider, 1995), and telomerase activity in various somatic tissues - including kidney - has also been reported in chickens (Venkatesan and Price, 1998).

Surprisingly, we find that in both analyzed tissues, the number of telomeres is significantly lower in the experimental group than in the control group, indicating the loss of entire telomeres due to a high-fat diet. Complete loss of telomeres on chromosomes is a phenomenon previously reported in tumor cells (Fouladi et al., 2000; Lo et al., 2002; Bai and Murnane, 2003; Ferreira et al., 2004; Lin et al., 2010; Raseley et al., 2023) and mutants (Vannier et al., 2012). Although a firm confirmation of this finding would require validations by alternative methods, our results suggest that diet can also induce complete telomere loss in both somatic and reproductive cells.

Optical genome mapping on sperm cells

As part of this study, a protocol was developed for the isolation of HMW DNA from sperm for the purpose of optical genome mapping. The existing commercial protocol for the isolation of HMW DNA (Bionano Genomics) is optimized for somatic cells and is inefficient in breaking down the protein component of sperm chromatin, which is a crucial requirement for successful DNA labeling in OGM. We developed a procedure in which chromatin proteins are efficiently removed from sperm genome while long HMW DNA molecules are isolated. This enabled us, for the very first time, to apply OGM technology to sperm genomes. We show that the developed protocol successfully yields HMW DNA of sufficient amount and quality and results in high-quality OGM data that is comparable to somatic tissue.

OGM is commonly used for the detection of large structural variants (SVs). The protocol for HMW DNA extraction that we present here enables the investigation of structural variation in sperm genomes as representatives of the genetic form that gets directly transmitted to the next generation. In OGM performed on sperm genomes, each molecule originates from a single sperm cell. Therefore, each cell is represented by one allele of a structural variant at a particular locus. This enables two important breakthroughs in the field. First, the direct detection of *de novo* SVs in the germline can be studied at unmatched resolution among a population of sperm cells in a single individual. Second, shifts in the frequency of existing variants can be monitored within a single generation. This opens the possibility of investigating the environmental effects at the level of structural variations, which can have direct consequences on offspring phenotype.

Methods

High-fat diet experiment

Before the experiment, mice were kept in groups of three mice per cage under standard laboratory conditions. Males of the C57BL/6 strain aged 4-5 weeks were randomly selected for the experiment. Six mice in the control group were fed a standard laboratory diet (4RF 21 Mucedola srl, Italy). Six mice in the experimental group were fed Western diet (WD, ssniff EF R/M acc TD88137 mod, catalog # E15721-34, ssniff Spezialdiaten GmbH, Germany), characterized by a high content of fat (21.2 %), sugar (33.2 %) and cholesterol (2.071 mg/kg). Food and water were provided *ad libitum* in both groups. The experiment lasted 80 days.

In an extended experiment, performed at a different time, additional six males of the C57BL/6 strain aged 8-9 weeks were randomly selected to repeat the high-fat diet experiment as described above. Six males of the same age were kept as a control group under standard conditions in parallel. Mice from the extended experiment were used solely for validations of telomere lengths by qPCR.

Tissue preparation

One mouse from the control group died before the end of the experiment. At the end of the experiment, the remaining mice were euthanized by cervical dislocation. Kidneys and *cauda epididymis* were isolated from each mouse. Each kidney was cut into 2-3 pieces, corresponding to ~30 mg per piece, frozen by immersion in liquid nitrogen, and stored at -80 °C. *Cauda epididymis* was separated from testis and transferred to an Eppendorf tube with 150 µL of DPBS buffer (Dulbecco's Phosphate Buffered Saline, Sigma-Aldrich, catalog # D8537) for brief storage before further processing. *Cauda epididymis* was placed on a parafilm sheet and cleaned from adipose tissue. *Epididymis* was punctured approximately ten times with a needle (16-21G), and after the viscous contents was drained from it, it was transferred back into the Eppendorf tube with 1,000 µL of DPBS, placed horizontally in the hybridization oven and incubated for 1 hour at 37 °C, 6 rotations per minutes (rpm). After incubation, the suspension was filtered through a 40 µm filter. Another 500 µL of DPBS was added to the tube and filtered again. The filtrates were collected and combined. 10 µL of the combined filtrate was transferred to a Neubar chamber for sperm cell counting. The rest of the filtrate was centrifuged for 10 minutes at 1,000 g at 4 °C, and the supernatant was discarded by careful pipetting. The remaining pellet in the Eppendorf tube was immersed directly into liquid nitrogen and stored at -80 °C until DNA isolation.

Isolation of high-molecular-weight (HMW) DNA and shipment

Isolation of HMW DNA from kidney tissue was performed by the Institute of Applied Biotechnologies a.s. (IAB, Olomouc, Czech Republic) by using *Bionano Prep SP Tissue and Tumor DNA Isolation Protocol* (Document Number: 30339). Approximately 30 mg of kidney tissue per mouse was placed in a cryotube and shipped to IAB on dry ice.

Isolation of HMW DNA from sperm was performed according to the protocol in the Supplementary Data (Text S1). Briefly, 3-4 million sperm cells were enclosed per agarose plug before cell lysis, and dithiothreitol (DTT) was used as a reducing agent during lysis to break down the protein component of the chromatin. A portion of DNA was digested with a restriction enzyme and compared with undigested DNA by PFGE to check the accessibility and integrity of isolated DNA (Supplementary Text S1). Each agarose block with HMW DNA isolated from sperm was inserted into a separate 5 mL tube. The tube was filled to the top with Wash Buffer (10 mM Tris-Hcl, 50 mM EDTA), closed, sealed with parafilm, and shipped to IAB at room temperature.

Optical genome mapping

Optical genome mapping was performed by IAB on the Saphyr platform (Bionano Genomics). Genome optical maps were generated for a total of 18 samples from nine mice: five from the experimental group and four from the control group. From each of them, OGM was performed on HMW DNA isolated from sperm and kidney tissue. Enzymatic labeling of HMW DNA from kidneys and sperm was performed at the IAB according to the BNG protocol described in the document *Bionano Prep Direct Label and Stain (DLS) Protocol* (Document Number: 30206). DNA concentration after DLS labeling was taken as a measure of initial DNA quality, and the coefficient of variation of DNA concentration was calculated from three measurements. Initial data processing, including de novo assembly of the genome and detection of SVs, was performed by the IAB using the set of bioinformatics tools BioNano Solve, according to the BNG *Guidelines for Running Bionano Solve on the Command Line* (Document Number: 30205).

TelOMpy workflow

In order to extract molecules from the OGM data that contain telomeres, TelOMpy first identifies contigs that are aligned to the ends of chromosomes. This information is in the "exp_refineFinal1_merged.xmap" file, under the output of the de novo assembly pipeline results (directory /output/annotation). By using contig identification number (ContigID), information about the alignment of molecules to the specified contig is extracted from XMAP file labeled "EXP_REFINEFINAL_contig{X}.xmap", where X is the unique ContigID for the given data set. Similarly, the same ContigID is used on corresponding CMAP files labeled "EXP_REFINEFINAL_contig{X}_q.cmap" to extract information on the molecule length and label positions of the aligned molecules. Depending on the chromosome arm, orientation of alignment between contig and the reference chromosome, and the orientation of alignment between molecule and contig, one of eight different ways of calculating the telomere length is applied (Figure 1), and can be reduced to two equations:

$$telomere\ length = \begin{cases} LP & if\ sign = +\\ ML - LP & if\ sign = -\end{cases}$$

where *LP* represents label position on the molecule that is aligned to the first (p-arm) or last (q-arm) label on the chromosome of reference genome; and *ML* represents total length of the molecule. If we assign the value -1 to long chromosome arm (q-arm) and the value 1 to short chromosome arm (p-arm), *sign* in the above equation is defined as:

$$sign = chromosome\ arm \times MOLtoCON \times CONtoREF$$

where *MOLtoCON* refers to alignment orientation of molecule to contig; and *CONtoREF* refers to alignment orientation of contig to reference chromosome.

Cases where the most distal label on the reference genome is not aligned (paired) are controlled through parameter *number of unpaired labels on the reference*. In the default settings of the TelOMpy tool, the value of this parameter is 0. However, this parameter can be set to values greater than zero if the distance to the most proximally aligned label on the reference is short enough for a molecule to fully encompass a telomere. The user is directed to check such cases visually in Bionano Access browser and to account for possible species-specificity in telomere length, label density, and average molecule length.

Similarly, unaligned labels on contig or molecule, distally from the last aligned label on the reference genome, are controlled by introducing the parameter *number of unpaired labels on the molecule* and *number of unpaired labels on the contig*. Some molecules or contigs may have one unaligned label in telomeric regions due to mutation or non-specific labeling. It is important to emphasize that these parameters refer to the number of continuously unaligned (*unpaired*) labels to the left (on p-arm) or the right (on q-arm) from the outermost aligned label on the reference genome and should not be confused with the total number of unaligned labels on a molecule or contig, which may, or may not be higher.

If the outermost label on the reference chromosome is aligned with a label on contig, but its distance from the annotated chromosome end (in the reference genome) is significantly greater than the average length of molecules, it is unlikely that the most distally aligned molecules will contain telomeres. For example, the first label position on chromosome 1 of the mm10 assembly is located 3 Mbp after the chromosome start position. To control for this, we introduced the *distance from the end* parameter as the maximum distance (in nucleotide base pairs) between the first/last aligned label and the end of a reference chromosome. This value, in addition to depending on the average length of the molecules, also depends on the length of the annotated telomeres for a particular organism. Telomeres are annotated on the reference genome as gaps of an arbitrarily set size, which is 100 kbp in mice and 10 kbp in humans (UCSC Table Browser), and the user should take this into account when defining the value of the *distance from the end* parameter. In the case of mouse mm10 genome assembly, a value of 200,000 (200 kbp) was used, which is twice the length of annotated telomeres but less than the average length of molecules mapped to the reference genome.

Validation by quantitative PCR

To perform quantitative PCR, we isolated genomic DNA by using QIAamp DNA Mini Kit (QIAGEN). Frozen kidney tissue was ground in liquid nitrogen and proceeded to DNA extraction by following the *Protocol: DNA Purification from Tissues* (QIAamp DNA Mini Kit). For the DNA isolation from sperm cells, DPBS buffer was added to frozen sperm pellets to a volume of 100 µL, followed by the procedure described in user-developed protocol 2: *Isolation of genomic DNA from sperm* (QA04; QIAamp DNA Mini Kit). DNA concentration was measured with QUBIT fluorometer using the Qubit dsDNA BR Assay Kit (Invitrogen).

Telomere length assessment was performed in triplicates for all the samples on Bio-Rad CFX96. Two separate singleplex reactions were performed for each sample: one with telomere (T) primers (forward: 5'-CGG TTT GTT TGG GTT TGG GTT TGG GTT TGG GTT TGG GTT-3'; reverse: 5'-GGC TTG CCT TAC CCT TAC

CCT TAC CCT TAC CCT TAC CCT-3') and one with primers that amplify a part of albumin gene (forward: 5'-CGG CGG CGG GCG GCG CGG GCT GGG CGG AAG TGC TGT AGT GGA TCC CTG-3'; reverse: 5'-GCC CGG CCC GCC GCG CCC GTC CCG CCG GAG AAG CAT GGC CGC CTT T-3') as a single-copy reference (S). Reactions were performed in 10 µl volume, containing 5 µl of iTaq Universal SYBR Green Supermix (Bio-Rad), 20 ng of DNA, nuclease-free water, and forward and reverse primers at a final concentration of 0.3 µM each for the telomere target and 0.8 µM each for the albumin target. Samples were run in 96-well plates, and each plate contained three negative controls with nuclease-free water instead of DNA. The thermal profile for amplification of the albumin target was 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C, 10 s at 62 °C, and 15 s at 74 °C (data collection). For the telomere target, the thermal profile was 10 min at 95 °C, followed by 30 cycles of 15 s at 95 °C and 1 min at 56 °C (data collection). We used CFX Maestro 2.2 software (Bio-Rad) to extract the Ct values for telomere and albumin amplicons from each sample. Average Ct values from triplicates were used to calculate the relative telomere length (T/S ratio) according to the formula: $2^{-\Delta\Delta Ct}$, where ΔΔCt = ΔCt(sample) – ΔCt(reference sample), and ΔCt = Ct(T) - Ct(S).

Declarations

Ethics approval

All procedures on mice were approved by the Ministry of Agriculture, Forestry and Water Management of the Republic of Croatia, Class: UP/I-322-01/19-01/59 and Case number: 525-10/0543-20-4, and carried out in compliance with relevant guidelines: ILAR Guide for the Care and Use of Laboratory Animals, EU Directive 2010/63/EU, the Croatian Animal Welfare Act (NN 102/17) and the Ruđer Bošković Institute's policy on animal use and ethics.

Consent for publication

Not applicable.

Availability of data and materials

TelOMpy is written in Python programming language. The package, along with all usage information, is available on the github repository - https://github.com/ivanp1994/telompy. The OGM dataset used in the current study is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Croatian Science Foundation (grant UIP-2019-04-7898).

Authors' contributions

ŽP designed and led the research. IP developed protocol for HMW DNA extraction from sperm cells, performed experiments and dissections on mice, HMW DNA extractions, and PFGE. IP and ŽP developed TelOMpy, analysed OGM data, and interpreted the results. IP wrote the code for TelOMpy. SS performed and analyzed qPCR experiments. ŽP wrote the manuscript draft. All authors contributed to the final manuscript.

Acknowledgements

Data analysis was performed on the high-performance computing cluster at the University Computing Centre (SRCE), University of Zagreb. We thank Kristina Sporiš for technical support and Petra Mikolčević for help with the preparation of sperm cells.

References

Aviv A, Hunt SC, Lin J, Cao X, Kimura M, Blackburn E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. Nucleic Acids Res. 2011;39(20):e134.

Bai Y, & Murnane JP. Telomere instability in a human tumor cell line expressing a dominant-negative WRN protein. Human Genetics. 2003;113(4):337–347.

Balhorn R. The protamine family of sperm nuclear proteins. Genome Biology. 2007;8(9):227.

Blackburn EH, Epel ES, Lin J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. Science. 2015;350(6265):1193-8.

Blasco MA, Lee H-W, Hande MP, Samper E, Lansdorp PM, DePinho RA, & Greider CW. Telomere Shortening and Tumor Formation by Mouse Cells Lacking Telomerase RNA. Cell. 1997;91(1):25–34.

Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature Genetics. 2013;45(4):371–384.

Cawthon RM. Telomere measurement by quantitative PCR. Nucleic Acids Res. 2002;30(10):e47.

Clemente DBP, Maitre L, Bustamante M, Chatzi L, Roumeliotaki T, Fossati S, et al. Obesity is associated with shorter telomeres in 8 year-old children. Sci Rep. 2019;9(1):18739.

de Frutos C, López-Cardona AP, Fonseca Balvís N, Laguna-Barraza R, Rizos D, Gutierrez-Adán A, & Bermejo-Álvarez P. Spermatozoa telomeres determine telomere length in early embryos and offspring. REPRODUCTION. 2016;151(1):1–7.

Fan H-C, Chang F-W, Tsai J-D, Lin K-M, Chen C-M, Lin S-Z, et al. Telomeres and Cancer. Life. 2021;11(12):1405.

Ferreira MG, Miller KM, & Cooper JP. Indecent Exposure. Molecular Cell. 2004;13(1):7–18.

Fice H, & Robaire B. Telomere Dynamics Throughout Spermatogenesis. Genes. 2019;10(7):525.

Fouladi B, Sabatier L, Miller D, Pottier, G & Murnane, JP. The Relationship Between Spontaneous Telomere Loss and Chromosome Instability in a Human Tumor Cell Line. Neoplasia. 2000;2(6):540–554.

Gielen M, Hageman GJ, Antoniou EE, Nordfjall K, Mangino M, Balasubramanyam M, et al. Body mass index is negatively associated with telomere length: a collaborative cross-sectional meta-analysis of 87 observational studies. Am J Clin Nutr. 2018;108(3):453-475.

Grill S, & Nandakumar J. Molecular mechanisms of telomere biology disorders. Journal of Biological Chemistry. 2021;296:100064.

Hemann MT, Strong MA, Hao LY, Greider CW. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability Cell. 2001;107(1):67-77.

Hemann, MT. Wild-derived inbred mouse strains have short telomeres. Nucleic Acids Research. 2000;28(22):4474–4478.

Lansdorp PM, Verwoerd NP, van de Rijke FM, Dragowska V, Little MT, Dirks RW, et al. Heterogeneity in telomere length of human chromosomes. Human Molecular Genetics. 1996;5(5):685–691.

Lin TT, Letsolo BT, Jones RE, Rowson J, Pratt G, Hewamana S, et al. Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. Blood. 2010;116(11):1899–1907.

Lo AWl, Sabatier L, Fouladi B, Pottier G, Ricoul M, & Mumane JP. DNA Amplification by Breakage/Fusion/Bridge Cycles Initiated by Spontaneous Telomere Loss in a Human Cancer Cell Line. Neoplasia. 2002;4(6):531–538.

Margalef P, Kotsantis P, Borel V, Bellelli R, Panier S, & Boulton SJ. Stabilization of Reversed Replication Forks by Telomerase Drives Telomere Catastrophe. Cell. 2018;172(3):439-453e14.

Martin-Ruiz CM, Baird D, Roger L, Boukamp P, Krunic D, Cawthon R, Dokter MM, van der Harst P, Bekaert S, de Meyer T, Roos G, Svenson U, Codd V, Samani NJ, McGlynn L, Shiels PG, Pooley KA, Dunning AM, Cooper R, Wong A, Kingston A, von Zglinicki T. Reproducibility of telomere length assessment: an international collaborative study. Int J Epidemiol.. 2015;44(5):1673-83.

McCaffrey J, Young E, Lassahn K, Sibert J, Pastor S, Riethman H, & Xiao M. High-throughput single-molecule telomere characterization. Genome Research. 2017;27(11):1904–1915.

Mirabello L, Yu K, Kraft P, De Vivo I, Hunter DJ, Prescott J, et al. The association of telomere length and genetic variation in telomere biology genesa. Human Mutation. 2010;31(9):1050–1058.

Montpetit AJ, Alhareeri AA, Montpetit M, Starkweather AR, Elmore LW, Filler K, et al. Telomere Length. Nursing Research. 2014;63(4):289–299.

Prowse KR, Greider CW Developmental and tissue-specific regulation of mouse telomerase and telomere length. Proc Natl Acad Sci U S A. 1995;92:4818–4822.

Raee P, Shams Mofarahe Z, Nazarian H, Abdollahifar MA, Ghaffari Novin M, Aghamiri S, Ghaffari Novin M. Male obesity is associated with sperm telomere shortening and aberrant mRNA expression of autophagy-related genes. Basic Clin Androl. 2023;33(1):13.

Raseley K, Jinwala Z, Zhang D, & Xiao M. Single-Molecule Telomere Assay via Optical Mapping (SMTA-OM) Can Potentially Define the ALT Positivity of Cancer. Genes. 2023;14(6):1278.

Schmidt TT, Tyer C, Rughani P, Haggblom C, Jones JR, Dai X, et al. High resolution long-read telomere sequencing reveals dynamic mechanisms in aging and cancer. Nature Communications. 2024;15(1):.

Soerensen M, Thinggaard M, Nygaard M, Dato S, Tan Q, Hjelmborg J, et al. Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. Aging Cell. 2012;11(2):223–227.

Tahamtan S, Tavalaee M, Izadi T, Barikrow N, Zakeri Z, Lockshin RA, et al. Reduced sperm telomere length in individuals with varicocele is associated with reduced genomic integrity. Scientific Reports. 2019;9(1):.

Valdes AM, Andrew T, Gardner JP, Kimura M, Oelsner E, Cherkas LF, et al. Obesity, cigarette smoking, and telomere length in women. Lancet. 2005;366(9486):662-4.

Vannier J-B, Pavicic-Kaltenbrunner V, Petalcorin MIR, Ding H, & Boulton SJ. RTEL1 Dismantles T Loops and Counteracts Telomeric G4-DNA to Maintain Telomere Integrity. Cell. 2012;149(4):795–806.

Venkatesan RN, & Price C. Telomerase expression in chickens: Constitutive activity in somatic tissues and down-regulation in culture. Proceedings of the National Academy of Sciences. 1998;95(25):14763–14768.

Wei B, Shao Y, Liang J, Tang P, Mo M, Liu B, et al. Maternal overweight but not paternal overweight before pregnancy is associated with shorter newborn telomere length: evidence from Guangxi Zhuang birth cohort in China. BMC Pregnancy Childbirth. 2021;21(1):283.

Yang Q, Zhao F, Hu L, Bai R, Zhang N, Yao G, Sun Y. Effect of paternal overweight or obesity on IVF treatment outcomes and the possible mechanisms involved. Sci Rep. 2016;6:29787.

Yang Q, et al. Sperm telomere length is positively associated with the quality of early embryonic development. Human reproduction. 2015;30:1876–1881.

Young E, Pastor S, Rajagopalan R, McCaffrey J, Sibert J, Mak ACY, et al. High-throughput single-molecule mapping links subtelomeric variants and long-range haplotypes with specific telomeres. Nucleic Acids Research. 2017;45(9):e73–e73.

Yuan Y, Chung CY-L, & Chan T-F. Advances in optical mapping for genomic research. Computational and Structural Biotechnology Journal. 2020;18:2051–2062.

Zhang F, Cheng D, Porter KI, Heck EA, Wang S, Zhang H, et al. Modification of the telomerase gene with human regulatory sequences resets mouse telomeres to human length. Nat Commun. 2025;16(1):1211.

Zijlmans JM, Martens UM, Poon SS, Raap AK, Tanke HJ, Ward RK, Lansdorp PM. Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. Proc Natl Acad Sci U S A. 1997;94(14):7423-8.