Вавиловский журнал генетики и селекции. 2023;27(3):283-292
DOI 10.18699/VJGB-23-34

Original Russian text https://vavilovj-icg.ru/

Compilation and functional classification of telomere length-associated genes in humans and other animal species

E.V. Ignatieva ✉, N.S. Yudin, D.M. Larkin ✉

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ eignat@bionet.nsc.ru; dmlarkin@gmail.com

Abstract. Telomeres are the terminal regions of chromosomes that ensure their stability while cell division. Telomere shortening initiates cellular senescence, which can lead to degeneration and atrophy of tissues, so the process is associated with a reduction in life expectancy and predisposition to a number of diseases. An accelerated rate of telomere attrition can serve as a predictor of life expectancy and health status of an individual. Telomere length is a complex phenotypic trait that is determined by many factors, including the genetic ones. Numerous studies (including genome-wide association studies, GWAS) indicate the polygenic nature of telomere length control. The objective of the present study was to characterize the genetic basis of the telomere length regulation using the GWAS data obtained during the studies of various human and other animal populations. To do so, a compilation of the genes associated with telomere length in GWAS experiments was collected, which included information on 270 human genes, as well as 23, 22, and 9 genes identified in the cattle, sparrow, and nematode, respectively. Among them were two orthologous genes encoding a shelterin protein (*POT1* in humans and *pot-2* in *C. elegans*). Functional analysis has shown that telomere length can be influenced by genetic variants in the genes encoding: (1) structural components of telomerase; (2) the protein components of telomeric regions (shelterin and CST complexes); (3) the proteins involved in telomerase biogenesis and regulating its activity; (4) the proteins that regulate the functional activity of the shelterin components; (5) the proteins involved in telomere replication and/or capping; (6) the proteins involved in the alternative telomere lengthening; (7) the proteins that respond to DNA damage and are responsible for DNA repair; (8) RNA-exosome components. The human genes identified by several research groups in populations of different ethnic origins are the genes encoding telomerase components such as *TERC* and *TERT* as well as *STN1* encoding the CST complex component. Apparently, the polymorphic loci affecting the functions of these genes may be the most reliable susceptibility markers for telomere-related diseases. The systematized data about the genes and their functions can serve as a basis for the development of prognostic criteria for telomere length-associated diseases in humans. Information about the genes and processes that control telomere length can be used for marker-assisted and genomic selection in the farm animals, aimed at increasing the duration of their productive lifetime.
Key words: telomere length; candidate genes; genome-wide association study; functional classification.

For citation: Ignatieva E.V., Yudin N.S., Larkin D.M. Compilation and functional classification of telomere length-associated genes in humans and other animal species. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(3):283-292. DOI 10.18699/VJGB-23-34

Компиляция и функциональная классификация генов, ассоциированных с длиной теломер, у человека и других видов животных

Е.В. Игнатьева ✉, Н.С. Юдин, Д.М. Ларкин ✉

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ eignat@bionet.nsc.ru; dmlarkin@gmail.com

Аннотация. Теломеры – это концевые участки хромосом, обеспечивающие их стабильность в ходе клеточного деления. Укорочение теломер инициирует процесс старения клеток, что может приводить к дегенерации и атрофии тканей. Укорочение теломер связано с сокращением продолжительности жизни и с предрасположенностью к ряду заболеваний, поэтому данный показатель может быть использован в качестве предиктора продолжительности жизни и состояния здоровья отдельного индивида. Длина теломер – сложный фенотипический признак, который определяется многими факторами, в том числе генетическими. Многочисленные исследования (включая полногеномный анализ ассоциаций, ПГАА) свидетельствуют о полигенном характере контроля длины теломер. Цель работы – охарактеризовать генетические основы регуляции длины теломер на основе данных ПГАА, полученных при исследовании различных популяционных выборок человека и других животных. Для этого авторами была собрана компиляция генов, ассоциированных с длиной теломер по дан-

ным ПГАА, которая включала сведения о 270 генах человека, а также 23, 22 и 9 генах, выявленных у крупного рогатого скота, домового воробья и нематоды соответственно. Среди них присутствовали два гена-ортолога, кодирующих белок шелтеринового комплекса (*POT1* у человека и *pot-2* у *C. elegans*). Функциональный анализ показал, что на длину теломер могут влиять генетические варианты в генах, кодирующих: 1) структурные компоненты теломеразы; 2) белковые компоненты теломерных участков хромосом (шелтериновый комплекс и CST комплекс); 3) белки, участвующие в биогенезе теломеразы и регулирующие ее активность; 4) белки, регулирующие функциональную активность компонентов шелтеринового комплекса; 5) белки, участвующие в репликации и/или кэпировании теломер; 6) белки, контролирующие альтернативный путь удлинения теломер; 7) белки, реагирующие на повреждения ДНК и отвечающие за репарацию; 8) компоненты РНК экзосом. В работе выявлены гены человека, идентифицированные несколькими исследовательскими группами в популяциях различного этнического происхождения. Это гены, кодирующие компоненты теломеразы (*TERC* и *TERT*), а также ген *STN1*, кодирующий белок CST комплекса. По-видимому, полиморфные локусы, затрагивающие функции этих генов, могут быть наиболее надежными маркерами предрасположенности к заболеваниям, связанным с длиной теломер. Систематизированные нами данные о генах и их функциях будут полезны при разработке прогностических критериев заболеваний человека, для которых показана связь с длиной теломер. Сведения о генах и процессах, контролирующих длину теломер, могут быть востребованы для маркер-ориентированной и геномной селекции сельскохозяйственных животных, направленной на повышение продолжительности их хозяйственного использования.
Ключевые слова: длина теломер; гены-кандидаты; полногеномный анализ ассоциаций; функциональный анализ.

Introduction

Telomeres are the terminal regions of chromosomes that ensure their stability and represented by evolutionary conserved tandemly repeated DNA sequences (e. g., a hexanucleotide TTAGGG repeat in vertebrates) of several kb in length (Podlevsky, 2008; Monaghan, Ozanne, 2018). For example, their lengths in humans at birth are 10–15 kb (Jafri et al., 2016). 3′ terminal end of a telomere is a single-stranded guanine-rich DNA region (150–200 nucleotides), whose end interacts with the double-stranded region to form the so-called T-loop at the telomere end. T-loop formation and stabilization are ensured by a shelterin complex (Fig. 1). This structure prevents recognition of a chromosome terminal region by repair proteins (de Lange, 2018).

DNA polymerase is unable to fully replicate the 3′-end of a linear DNA during cell division, which leads to a loss of 50–200 nucleotides of the telomeric sequence at each cell division (Fan et al., 2021). Telomere shortening can also be facilitated by other factors and processes (Suppl. Material 1)[1], such as oxidative stress, inflammation, UV irradiation, effects of toxic agents, DNA replication errors, etc. (Aviv, Shay, 2018; Monaghan, Ozanne, 2018). These factors are likely to produce different effects depending on cell types and the organism's development stage and species (Monaghan, Ozanne, 2018).

Telomere shortening initiates cellular senescence. Activation of DNA damage response signaling pathways results in a cell cycle arrest, which may in turn lead to apoptosis and eventually – to progressive tissue degeneration (Jafri et al., 2016; Aviv, Shay, 2018; Monaghan, Ozanne, 2018). The data collected at the cellular level *in vitro* and in model organisms lay the foundation for using the telomere length as a predictor of life expectancy and health status of an individual. Indeed, studies in humans (Crocco et al., 2021), mice (Vera et al., 2012), sheep (Wilbourn et al., 2018), cattle (Seeker et al., 2021), wild birds (Bichet et al., 2020), and other animals have shown that shorter telomere length may be associated with reduced life expectancy. The studies in humans discovered an association between the telomere length and cardiovascular diseases, cancer, diabetes, inflammation, and other pathological states (Kong et al., 2013; Jafri et al., 2016; Aviv, Shay, 2018).

Telomere shortening is prevented by telomerase, a specialized ribonucleoprotein complex acting as a reverse transcriptase. In humans, telomerase is active in almost all the cancer cells studied (Jafri et al., 2016), in blastocyst, in most somatic tissues at 16–20 weeks of development (except for brain cells), and ovary and sperm cells at all ontogenetic stages (except for mature spermatozoids and oocytes) (Wright et al., 1996).

Telomerase activity is controlled by the proteins regulating expression of telomerase components, their movement to various cell compartments, processing, and assembly as well as by the proteins maintaining stability of the telomerase complex or, on the contrary, activating its degradation (Egan, Collins, 2012; Tseng et al., 2015; Schrumpfová, Fajkus, 2020). The main stages of telomerase biogenesis are presented in Figure 2. The examples of proteins affecting the telomerase activity are presented in Suppl. Material 2. In addition, the telomerase activity is also affected by the shelterin (Diotti, Loayza, 2011; de Lange, 2018) and CST complex (Fig. 3) (Chen et al., 2012).

Telomere length is a complex phenotypic trait determined by multiple factors including genetic ones. The meta-analysis of heritability data for this trait performed in eighteen vertebrate species showed the averaged heritability index of 45 %. The studies showed the value of 52 % in humans, 42 % – in Holstein cattle breed, 35 % – in hamadryas baboons, and 5 % – in sheep (Chik et al., 2022).

The problem of genetic basis of telomere length regulation is of interest for many researchers. The information on telo-

[1] Supplementary Materials 1–11 are available in the online version of the paper: http://vavilov.elpub.ru/jour/manager/files/Suppl_Ignatieva_Engl_27_3.pdf.

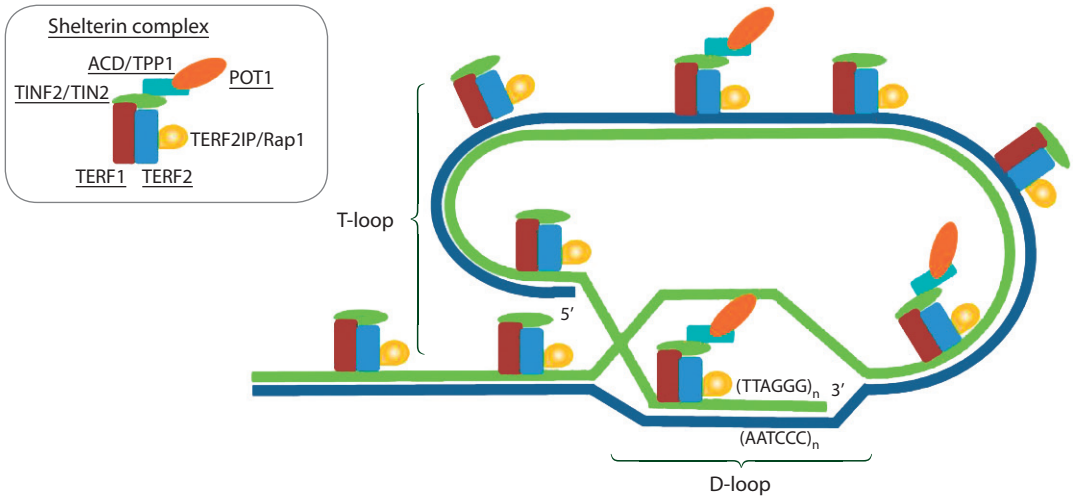

Fig. 1. Structure of a chromosome's telomeric region.

Telomeric DNA is presented as a T-loop reconstructed following the black-and-white illustration from Fan and co-workers (2021); nucleotide sequences in the DNA strands are not shown (Fan et al., 2021). The top left is simplified illustration of the relative positions of shelterin subunits, following the description from Jafri and co-workers (2016). Since the ACD/TPP1 and POT1 are much less abundant in nuclei (de Lange, 2018), some shelterin complexes are depicted without these subunits. D-loop is a structure, where two strands of the double-stranded DNA are separated, and one of them connects with the third DNA strand (a single-stranded 3′-end of the telomeric DNA region). The names of proteins corresponding to the human genes associated with telomere length according to the GWAS data are underlined.

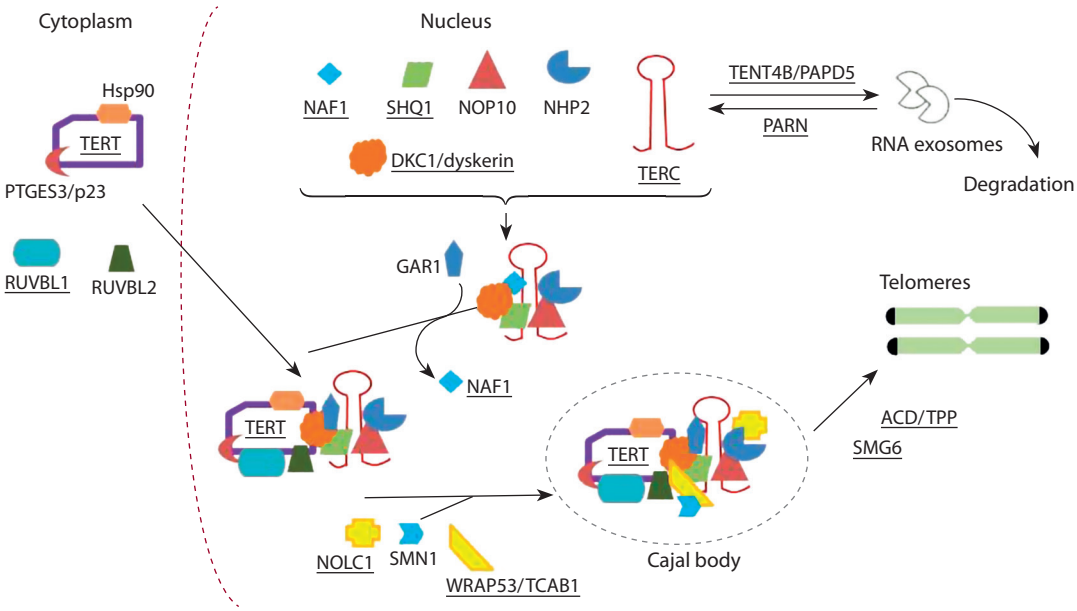

Fig. 2. Simplified presentation of the main stages of telomerase biogenesis.

The names of proteins and RNAs corresponding to the human genes associated with telomere length (as per GWAS results) are underlined. The scheme is based on the data on protein functions from the research articles cited in Suppl. Material 2.

merase components and proteins involved in telomere length regulation (including 20 proteins identified in mammals) can be found in The Telomerase Database (http://telomerase.asu.edu/) (Podlevsky et al., 2008). Joyce and co-workers (2018) presented a set of 80 human genes with telomere-related functions (Joyce et al., 2018).

The GWAS data also indicate a polygenic nature of telomere length heritability. For instance, the GWAS Catalog (https://www.ebi.ac.uk/gwas/) cites 99 human genes that either include or neighbor the telomere length-associated allelic variants. One of the largest GWA studies presents the data on 138 human genomic loci, whose allelic variants are associated with telomere length (Codd et al., 2021).

In addition to the telomere length association data gathered using GWA studies in various human population samples, the data obtained in other animal species also appear to be

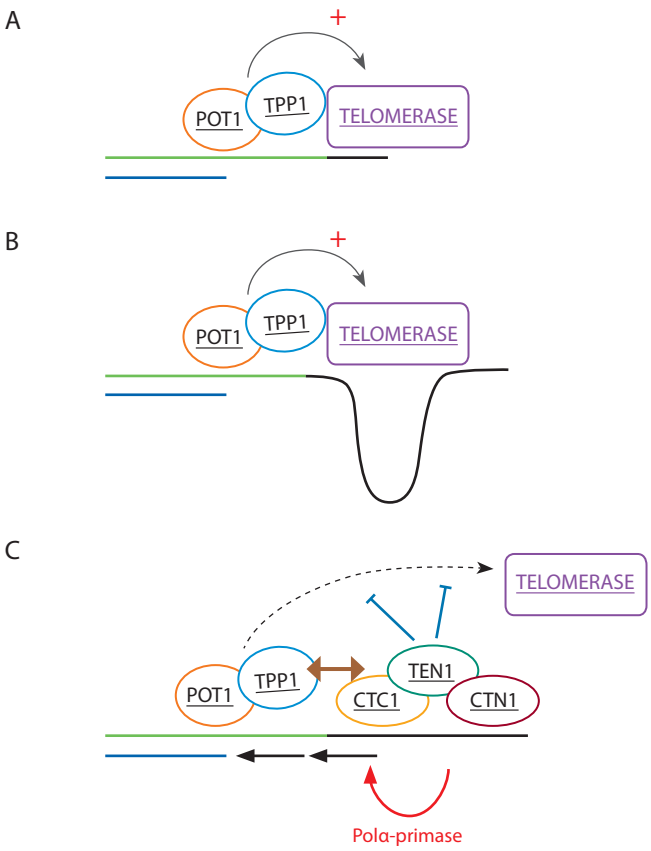

Fig. 3. The role of CST proteins in telomerase activity regulation at the late S/G2 phase.

The first step of the five-step mechanism described by Chen et al. (2012) is the recruitment of telomerase and additional ACD/TPP1 and POT1 by shelterin complex (Step 1, *Recruitment*, not shown in the figure). Then, telomerase starts extending the single-stranded region of the DNA molecule (Step 2, *Extention I*) (shown in Panel A, the newly synthesized DNA region is represented by a black line). After that (Step 3, *Extention II*) the single-stranded region of the DNA molecule is further extended (see Panel B). CST proteins interact with the newly synthesized single-stranded DNA region (~60 nucleotides) hindering the stimulating effect of ACD/TPP1 and POT1 on telomerase (Step 4, *Termination*) and initiating C-strand synthesis by DNA polymerase alpha-primase (Polα-primase) (Step 5, *Fill-in*). Steps 4 and 5 are presented in Panel C. The names of proteins corresponding to the human genes associated with telomere length (as per GWAS results) are underlined.

of interest. However, these studies are rather scarce and are available only for Holstein–Friesian cattle (Ilska-Warner et al., 2019), house sparrow nestlings (Pepke et al., 2021), and *C. elegans* (Cook et al., 2016).

The objective of this review was to characterize the genetic basis of telomere length regulation using the GWAS data collected in various human populations and to compare them with the results of similar experiments in other animal species. For this purpose, (1) the data on genes identified in GWAS telomere length experiments were systematized; (2) functional annotation of genes was performed, and the set of biological processes affecting telomere length was identified.

Materials and methods

The data on telomere length-associated genes were obtained from the papers available in the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) using such keywords as 'telomere length' and 'GWAS'. Functional annotation of genes was performed using information obtained from the papers presenting GWAS data, PubMed, The Telomerase Database (http://telomerase.asu.edu/) queries, and the DAVID knowledgebase (https://david.ncifcrf.gov/) (Sherman et al., 2022).

Results and discussion

Human genes identified through GWA studies

PubMed queries produced 18 scientific papers presenting the results of identifying telomere length-associated polymorphic loci in human genome based on GWAS data. These papers were analyzed, and the data on 270 telomere length-associated genes were collected (Suppl. Material 3). Most genes (262) were identified in European-ancestry population samples, the data on 15 genes were obtained from the studies in Southeast Asian population samples (natives of China, Bangladesh, and India), five genes were identified as a result of trans-ethnic meta-analysis of Singaporean Chinese and European ancestry data (Dorajoo et al., 2019), and one gene was found in African Americans (Zeiger et al., 2018).

The data on functional significance in the context of telomere length regulation were presented by the authors of GWA studies for 52 genes out of 270 (see Suppl. Material 3). The fact that the data on gene significance in the context of telomere length regulation were unavailable for a number of loci reflects the capabilities and limitations of GWAS methodology. Most loci identified by GWAS and associated with the trait of interest are located in intergenic regions. As a rule, in these cases, the set of candidate genes includes the nearest genes, whose functional significance is often difficult to interpret. To identify the mechanisms and genes, through which intergenic variants affect the studied traits, additional experiments are required. For example, it was shown that T-to-C substitution of rs1421085 in the intron of *FTO* gene affects the expression of *IRX3* and *IRX5*, whose transcription start sites are far away (~520 and ~1160 kb) from rs1421085 (Claussnitzer et al., 2015).

Main functional groups of human telomere length-associated genes

A functional classification was performed for a set of 52 human genes for which there was information about their functional significance in the context of telomere length regulation (Suppl. Material 4). As a result, several functional groups of genes have been identified (Fig. 4):

Genes encoding telomerase components: TERC is the telomerase RNA component acting as a matrix for DNA strand extension at the telomere end and TERT is a reverse transcriptase enzyme subunit (Egan, Collins, 2012; Tseng et al., 2015).

Genes encoding shelterin proteins: components of this complex (TERF1/TRF1, TERF2/TRF2, POT1, TERF2IP/Rap1/DRIP5, TINF2/TIN2, and ACD/TPP1/TINT1) can bind to both double-stranded and single-stranded telomeric DNA regions (see Fig. 1), stabilize them, protect them from exo-

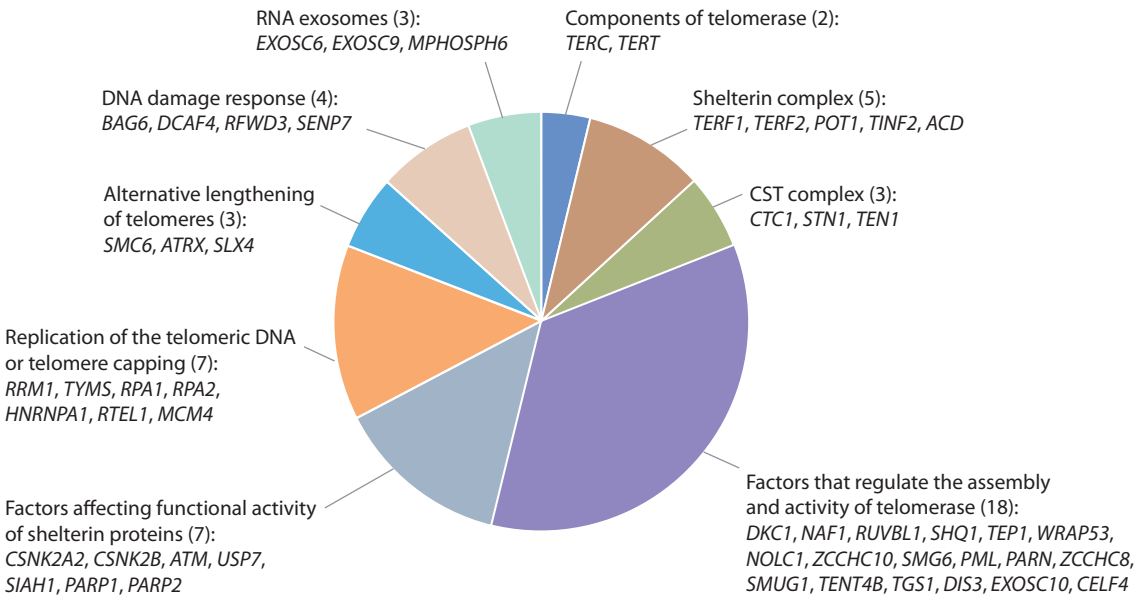

Fig. 4. Functional groups of human telomere length-associated genes.

The classification is presented for 52 genes, whose role in telomere length regulation is characterized in Suppl. Material 4. The numbers given in parentheses indicate the numbers of genes in groups.

nucleases, reduce telomerase access, and inhibit the proteins activated by damaged DNA and involved in double-stranded break repair (Diotti, Loayza, 2011; de Lange, 2018). The GWAS data on telomere length association were obtained for the genes coding for five out of six shelterin proteins (TERF1, TERF2, POT1, TINF2 and ACD/TPP1/TINT1) (see Fig. 4, Suppl. Material 4).

Genes encoding CST proteins: CTC1, STN1, TEN1. CST complex acts as a telomerase negative regulator at the late S to G2 phase of the cell cycle (see Fig. 3) (Chen et al., 2012).

Genes encoding proteins involved in telomerase biogenesis and regulating its activity. One of these genes, *ZCCHC10*, encodes a protein regulating telomerase synthesis at transcriptional level: ZCCHC10 suppresses *TERT* transcription (Ohira et al., 2019). Processing and assembly of a telomerase RNA subunit involves DKC1, NAF1, and SHQ1 (Egan, Collins, 2012), ribonuclease PARN, exoribonuclease DIS3, the component of a nuclear exosome targeting (NEXT) complex, ZCCHC8 (Tseng et al., 2015), SMUG1 (Kroustallaki et al., 2019), and CELF4/BRUNOL4 (Mangino et al., 2009). Non-canonical polymerase TENT4B/PAPD5 (Nagpal et al., 2020), trimethylguanosine synthetase TGS1 (Chen et al., 2020), and EXOSC10 RNA exosome component (Stuparević et al., 2021) cause a decrease in the level of active TERC. The assembly of telomerase nucleoprotein complex involves ATPase RUVBL1/pontin (Jafri et al., 2016) and telomerase-associated protein TEP1 (Codd et al., 2021). Two proteins (WRAP53/WDR79/TCAB1 and NOLC1/NOPP140) provide telomerase accumulation in Cajal bodies, the small nuclear organelles where processing of small nuclear and nucleolar RNAs and assembly or ribonucleoprotein complexes occur (Bizarro et al., 2019; Schrumpfová, Fajkus, 2020). Telomerase activity is modulated by activator protein SMG6/EST1A, which also binds to a single-stranded DNA (Snow et al., 2003), and PML protein, whose isoform PML-IV suppresses telomerase activity (Oh et al., 2009).

Genes encoding proteins regulating functional activity of shelterin proteins. CSNK2A2 and CSNK2B are the subunits of casein kinase which phosphorylates TERF1, increasing its binding to telomeres (Saxena et al., 2014; Li et al., 2020). ATM serine/threonine kinase, on the contrary, decreases TERF1 binding to the telomeric DNA (Li et al., 2020). Peptidase USP7 and ubiquitin ligase SIAH1 activate proteasomal degradation of POT1 and TERF2, respectively (Codd et al., 2021). ADP ribosylases PARP1 and PARP2 reduce the DNA binding activity of TERF2 (Dorajoo et al., 2019; Codd et al., 2021).

Genes encoding proteins involved in telomere replication and/or capping: (1) enzymes RRM1 and TYMS involved in synthesis of deoxynucleoside triphosphates (dNTP) and thymidylates required for DNA synthesis (Dorajoo et al., 2019; Nersisyan et al., 2019); (2) helicases RTEL1 and MCM4 (Codd et al., 2013, 2021); (3) RPA1 and RPA2, the subunits of the RPA complex capable to unfold G-quadruplex structures that may block DNA replication (Codd et al., 2021); (4) HNRNPA1 promoting telomere capping after DNA replication (Codd et al., 2021).

Genes encoding proteins affecting the alternative telomere lengthening pathway. This telomerase-independent mechanism (ALT or Alternative Lengthening of Telomere, see the description in Suppl. Material 5) includes the recombination between telomeric regions of two DNA molecules (Sobinoff, Pickett, 2017, 2020). Three genes identified in GWA studies were attributed to this group (see Fig. 4, Suppl.

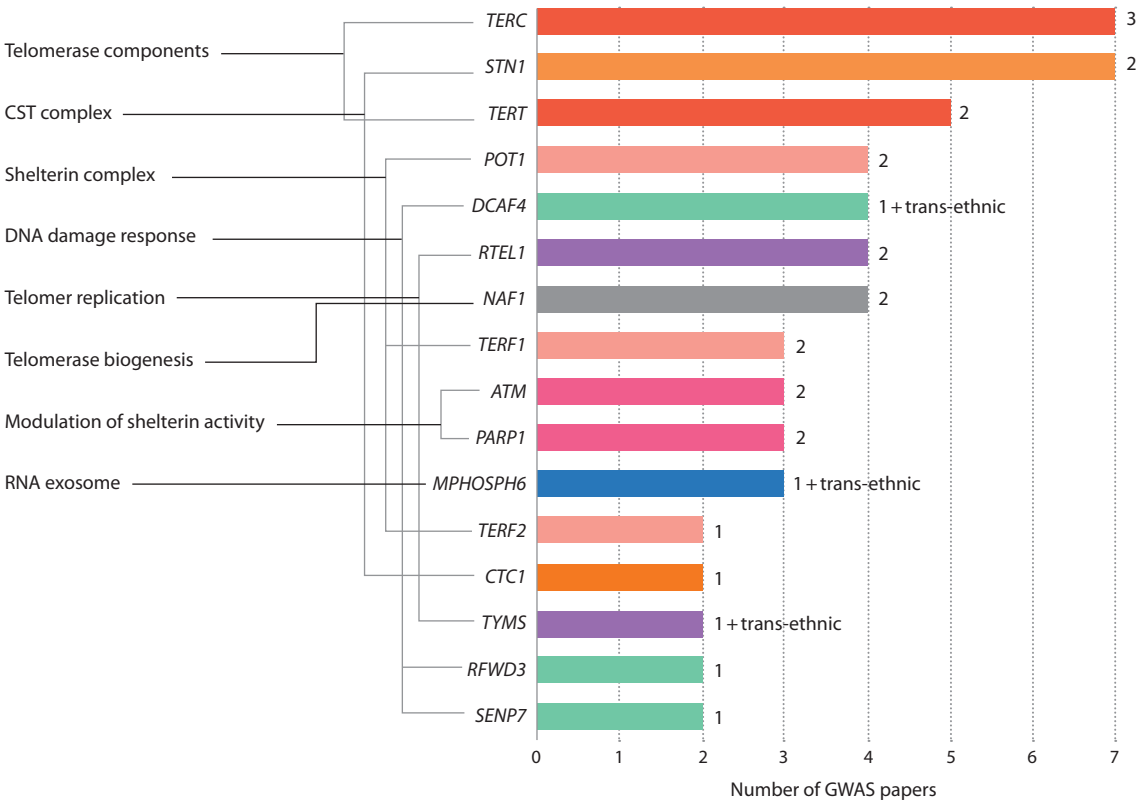

Fig. 5. Genes revealed in at least two GWAS papers.

Colors of columns indicate functional groups the genes belong to. The numbers on the right of columns represent the number of ethnically diverse GWAS population samples, in which the genes were identified. Trans-ethnic is the group consisting of Singaporean Chinese and Europeans.

Material 4). These genes encode SMC6 which activates ALT (Potts, Yu, 2007) and its two inhibiting proteins: ATRX with chromatin remodeling activity and SLX4 endonuclease (Sobinoff, Pickett, 2017).

Genes encoding DNA damage response proteins: (1) peptidase SENP7 (Li et al., 2020); (2) chaperone protein BAG6 (Li et al., 2020); (3) DCAF4 interacting with CUL4-DDB1 ligase (Mangino et al., 2015); (4) RFWD3 interacting with RPA protein (replication protein A) (Li et al., 2020).

Genes encoding subunits of RNA exosomes: EXOSC6, EXOSC9 (Codd et al., 2021) and MPHOSPH6 (Dorajoo et al., 2019). These proteins are functionally significant, because it is known that TERC may be subjected to 3′-processing, and the RNA-exosomes are involved in this process (Tseng et al., 2015).

Human candidate genes identified in more than one study

As mentioned above, we have analyzed 18 papers on identifying telomere length-associated human genome loci based on GWA studies and collected the data on 270 such genes (see Suppl. Material 3). Notably, only 16 genes were identified in at least two studies (Fig. 5).

The most frequently identified genes were the ones encoding both telomerase components (*TERC* and *TERT*) and *STN1* encoding a component of the CST complex (revealed in 7, 5, and 7 studies, respectively). Three genes *POT1*, *TERF1*, and *TERF2* encoding components of the shelterin complex were mentioned in 4, 3, and 2 publications, respectively. Three more genes *DCAF4*, *RTEL1*, and *NAF1* controlling the DNA damage response, telomere replication, and telomerase biogenesis were identified in four studies. *ATM*, *PARP1*, *MPHOSPH6*, *RFWD3*, *SENP7*, and *TYMS* were identified in 3 or 2 papers.

Most of 16 genes listed above were identified in population samples of different ethnic origin: (1) *TERC* in three ethnic groups, namely Europeans, Bangladeshis, and Singaporean Chinese; (2) *DCAF4*, *MPHOSPH6*, and *TYMS* in Europeans and as a result of the trans-ethnic meta-analyses (Singaporean Chinese+Europeans); (3) *TERT*, *STN1*, *POT1*, *RTEL1*, *NAF1*, *TERF1*, *ATM*, *PARP1* in two ethnic groups, namely Europeans and Singaporean Chinese.

Identification of the genes related to telomere length regulation according to DAVID

Using DAVID, we found the terms from the GOTERM_BP_DIRECT dictionary that were significantly (FDR < 0.05) associated with the list of 270 human genes presented in Suppl. Material 3. Sixteen terms indicating biological processes that directly control telomere length are presented in Suppl. Material 6, and the remaining fifteen terms are listed in Suppl. Material 7. There were 30 genes associated with the terms from the first group (see Suppl. Material 6), with two of them

(*SIRT6* and *TP53*) previously not recognized as biologically interpretable (these genes were presented in (Codd et al., 2021) without comment on their functional significance in the context of telomere length regulation). The analysis of scientific papers showed that proteins encoded by both genes can function in the subtelomeric regions of chromosomes (Tennen et al., 2021; Tutton et al., 2016), which means they could be indirectly involved in telomere length regulation.

Then, the genes associated with the second group of GOTERM_BP_DIRECT terms identified at FDR < 0.05 (see Suppl. Material 7) were analyzed. Among them, 29 genes were found that had no biological interpretation (highlighted in red in Suppl. Material 7, and listed in Suppl. Material 8). This group of 29 genes included the above mentioned *SIRT6* and *TP53*, as well as *BRCA1*, *SAMHD1*, and *BRCC3* associated with the maximum number of GO terms (six, four, and four, respectively). Apparently, the genes from the list thus obtained may also be of interest in the context of telomere length regulation.

Telomere length-associated genes found in other animal species

Genome-wide search for telomere length-associated loci and genes was carried out in three animal species: cattle (*Bos taurus*), sparrows (*Passer domesticus*), and nematodes (*Caenorhabditis elegans*).

A GWA study to investigate the species *Bos taurus* was carried out on 702 animals of the Holstein–Friesian breed (Ilska-Warner et al., 2019). The study of the DNA isolated from the whole blood of cows sampled at birth showed six telomere length-associated polymorphic loci, and three additional loci were identified when analyzing the DNA from blood samples at the first lactation. An analysis of the quantitative trait loci (QTL) corresponding to the identified genetic variants revealed 14 candidate genes at birth and 9 at the first lactation (see the Table and Suppl. Material 9). The authors were unable to find any data on direct involvement of the identified genes in processes associated with telomere length regulation. *NUP93* nucleoporin gene encoding a nuclear pore component was considered a potential regulator, because it was shown earlier in yeast that nucleoporins facilitated silencing of genes in proximity of telomeric regions (Van de Vosse et al., 2013).

The recently published results of GWA study (Pepke et al., 2021) in house sparrow (*Passer domesticus*) nestlings made it possible to identify 22 candidate genes (see the Table and Suppl. Material 10). According to the authors, the genes of interest in the context of telomere length regulation seem to be as follows: (1) *WNT9B* encoding a protein component of Wnt/β-catenin signaling pathway due to β-catenin involvement in *Tert* activation in embryonic stem cells of mice; (2) *CDCA4*, *GH*, and *GHRHR* regulating cell proliferation, apoptosis, and body growth; (3) *RHOF* involved in cytoskeletal organization; (4) *RNF34* (E3 ubiquitin-protein ligase RNF34) regulating ubiquitination; (5) *AQP1* due to involvement of aquaporin protein in transport of nitrogen oxide and active forms of oxygen, which increases oxidative stress which can in turn affect telomerase activity; (6) *SCN4A*, because its expression in human stem cells correlates with telomere length.

Our analysis showed that none of the candidate genes identified in cattle (23 genes) and house sparrows (22 genes) (see the Table) had orthologs among the 270 genes identified based on GWA studies in humans and presented in Suppl. Material 3.

The study in *C. elegans* (Cook et al., 2016) produced 9 candidate genes (see the Table and Suppl. Material 11). One out of nine genes, *pot-2*, is orthologous to *POT1* encoding a shelterin complex component in humans. The authors assume that another gene *ZK1127.4* may also be involved in telomere length regulation, because BCCIP encoded by an orthologous human gene interacts with BRCA2 involved in DNA replication.

In general, when comparing sets of candidate genes identified in humans and three other animal species, almost no orthologous genes are detected, which may be due to species-specific features of telomere length regulation, some peculiarities of regulation at various ontogenetic stages, and differences in sampled tissues or gender of the studied individuals.

Telomere length-associated animal genes (see Suppl. Materials 9–11 for additional data)

Species	Method / DNA source / Reference	Candidate genes
Bos taurus (female Holstein–Friesian cattle)	GWAS / whole blood of female cattle / (Ilska-Warner et al., 2019)	At birth: *NUP93, CCSER1, MMRN1, SNCA, GPRIN3, HDGFL1, RF00026, DOK6, RF00001, CCDC102B, TMX3, DSEL, bta-mir-138-2, bta-mir-2284c* In first lactation: *PTPRD, CYTL1, MSX1, STX18, NSG1, ACOX3, TRMT44, CPZ, HMX1*
Passer domesticus (house sparrow)	GWAS / whole blood of nestlings aged 5–14 days / (Pepke et al., 2021)	*FRMD4B, LMOD3, ARL6IP5, UBA3, TMF1, EOGT, AQP1, GHRHR, OXR1, ORAI1, MORN3, KDM2B, RNF34, TMEM120B, RHOF, ANAPC5, SHCBP1, CDCA4, SCN4A, GH, GOSR2, WNT9B*
Caenorhabditis elegans (soil-inhabiting nematode)	GWAS using genome-wide sequencing data / cells of whole-body nematodes / (Cook et al., 2016)	*pot-2* (*POT1*)*, *mms-19* (*MMS19*), *ZK1127.4* (*BCCIP*), *ZC487.2, srd-35, T06D8.3* (*PLPPR1, PLPPR5*), *ZK783.5, F58F6.3, C12D5.10*

* *C. elegans* genes are cited with human orthologs in parentheses.

Conclusions

In the present paper, a compilation of telomere length-associated genes identified based on GWA studies and including the data on 270 human genes (see Suppl. Material 3), as well as 23, 22, and nine genes identified in cattle, house sparrow, and nematode (see the Table) is presented. The analysis of functions of 52 human genes with functional interpretation available (see Fig. 4, Suppl. Material 4) showed that telomere length may be affected by variants of genes encoding: (1) the structural components of telomerase; (2) the protein components of telomeric chromosome regions (shelterin complex and CST complex); (3) the proteins involved in telomerase biogenesis and regulating its activity; (4) the proteins regulating functional activity of shelterin subunits; (5) the proteins involved in telomere replication and/or capping; (6) the proteins controlling the alternative telomere lengthening pathway; (7) DNA damage response and repair proteins; (8) RNA exosome components.

Candidate human genes identified by several research groups in population samples of different ethnic origin are determined: genes encoding telomerase components (*TERC* and *TERT*) and *STN1* encoding a subunit of CST complex (see Fig. 5). It seems that polymorphic loci that affect the functions of these genes can potentially be the most reliable predisposition markers for telomere length-associated diseases.

Comparison of the data obtained from GWA studies in humans (see Suppl. Material 3) with the results of similar experiments obtained for other animal species (see the Table and Suppl. Materials 9–11) confirmed and expanded the understanding of the complex polygenic nature of telomere length regulation. In addition, a pair of orthologous genes encoding a shelterin protein (*POT1* in humans and *pot-2* in *C. elegans*) was identified; this finding demonstrates the high biological significance of this gene in various species.

Systematized data on genes and their functions may lay the foundation for development of prognostic criteria for human pathologies explicitly associated with telomere length. In addition, the data on biological processes affecting telomere length and genes regulating these processes may be used for marker-assisted and genomic selection of the farm animals aimed at increasing the duration of their productive lifetime.

References

Aviv A., Shay J.W. Reflections on telomere dynamics and ageing-related diseases in humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160436. DOI 10.1098/rstb.2016.0436.

Bichet C., Bouwhuis S., Bauch C., Verhulst S., Becker P.H., Vedder O. Telomere length is repeatable, shortens with age and reproductive success, and predicts remaining lifespan in a long-lived seabird. *Mol. Ecol*. 2020;29(2):429-441. DOI 10.1111/mec.15331.

Bizarro J., Bhardwaj A., Smith S., Meier U.T. Nopp140-mediated concentration of telomerase in Cajal bodies regulates telomere length. *Mol. Biol. Cell*. 2019;30(26):3136-3150. DOI 10.1091/mbc.E19-08-0429.

Chen L., Roake C.M., Galati A., Bavasso F., Micheli E., Saggio I., Schoeftner S., Cacchione S., Gatti M., Artandi S.E., Raffa G.D. Loss of human TGS1 hypermethylase promotes increased telomerase RNA and telomere elongation. *Cell Rep*. 2020;30(5):1358-1372.e5. DOI 10.1016/j.celrep.2020.01.004.

Chen L.Y., Redon S., Lingner J. The human CST complex is a terminator of telomerase activity. *Nature.* 2012;488(7412):540-544. DOI 10.1038/nature11269.

Chik H.Y.J., Sparks A.M., Schroeder J., Dugdale H.L. A meta-analysis on the heritability of vertebrate telomere length. *J. Evol. Biol.* 2022;35(10):1283-1295. DOI 10.1111/jeb.14071.

Claussnitzer M., Dankel S.N., Kim K.-H., Quon G., Meuleman W., Haugen C., Glunk V., Sousa I.S., Beaudry J.L., Puviindran V., Abdennur N.A., Liu J., Svensson P.-A., Hsu Y.-H., Drucker D.J., Mellgren G., Hui C.-Ch., Hauner H., Kellis M. FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med*. 2015; 373:895-907. DOI 10.1056/NEJMoa1502214.

Codd V., Nelson C.P., Albrecht E., Mangino M., Deelen J., Buxton J.L., Hottenga J.J., Fischer K., Esko T., Surakka I., … Boomsma D.I., Jarvelin M.R., Slagboom P.E., Thompson J.R., Spector T.D., van der Harst P., Samani N.J. Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet*. 2013; 45(4):422-427. DOI 10.1038/ng.2528.

Codd V., Wang Q., Allara E., Musicha C., Kaptoge S., Stoma S., Jiang T., Hamby S.E., Braund P.S., Bountziouka V., Budgeon C.A., Denniff M., Swinfield C., Papakonstantinou M., Sheth S., Nanus D.E., Warner S.C., Wang M., Khera A.V., Eales J., Ouwehand W.H., Thompson J.R., Di Angelantonio E., Wood A.M., Butterworth A.S., Danesh J.N., Nelson C.P., Samani N.J. Polygenic basis and biomedical consequences of telomere length variation. *Nat. Genet.* 2021;53(10):1425-1433. DOI 10.1038/s41588-021-00944-6.

Cook D.E., Zdraljevic S., Tanny R.E., Seo B., Riccardi D.D., Noble L.M., Rockman M.V., Alkema M.J., Braendle C., Kammenga J.E., Wang J., Kruglyak L., Félix M.A., Lee J., Andersen E.C. The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics*. 2016;204(1):371-383. DOI 10.1534/genetics.116.191148.

Crocco P., De Rango F., Dato S., Rose G., Passarino G. Telomere length as a function of age at population level parallels human survival curves. *Aging* (*Albany NY*). 2021;13(1):204-218. DOI 10.18632/aging.202498.

de Lange T. Shelterin-mediated telomere protection. *Annu. Rev. Genet.* 2018;52:223-247. DOI 10.1146/annurev-genet-032918-021921.

Diotti R., Loayza D. Shelterin complex and associated factors at human telomeres. *Nucleus.* 2011;2(2):119-135. DOI 10.4161/nucl.2.2.15135.

Dorajoo R., Chang X., Gurung R.L., Li Z., Wang L., Wang R., Beckman K.B., Adams-Haduch J., M Y., Liu S., Meah W.Y., Sim K.S., Lim S.C., Friedlander Y., Liu J., van Dam R.M., Yuan J.M., Koh W.P., Khor C.C., Heng C.K. Loci for human leukocyte telomere length in the Singaporean Chinese population and trans-ethnic genetic studies. *Nat. Commun.* 2019;10(1):2491. DOI 10.1038/s41467-019-10443-2.

Egan E.D., Collins K. Biogenesis of telomerase ribonucleoproteins. *RNA*. 2012;18(10):1747-1759. DOI 10.1261/rna.034629.112.

Fan H.C., Chang F.W., Tsai J.D., Lin K.M., Chen C.M., Lin S.Z., Liu C.A., Harn H.J. Telomeres and cancer. *Life* (*Basel*). 2021; 11(12):1405. DOI 10.3390/life11121405.

Ilska-Warner J.J., Psifidi A., Seeker L.A., Wilbourn R.V., Underwood S.L., Fairlie J., Whitelaw B., Nussey D.H., Coffey M.P., Banos G. The genetic architecture of bovine telomere length in early life and association with animal fitness. *Front. Genet*. 2019;10: 1048. DOI 10.3389/fgene.2019.01048.

Jafri M.A., Ansari S.A., Alqahtani M.H., Shay J.W. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* 2016;8(1):69. DOI 10.1186/s13073-016-0324-x.

Joyce B.T., Zheng Y., Nannini D., Zhang Z., Liu L., Gao T., Kocherginsky M., Murphy R., Yang H., Achenbach C.J., Roberts L.R., Hoxha M., Shen J., Vokonas P., Schwartz J., Baccarelli A., Hou L. DNA methylation of telomere-related genes and cancer risk. *Cancer Prev. Res.* (*Phila*). 2018;11(8):511-522. DOI 10.1158/1940-6207.CAPR-17-0413.

Kong C.M., Lee X.W., Wang X. Telomere shortening in human diseases. *FEBS J*. 2013;280(14):3180-3193. DOI 10.1111/febs.12326.

Kroustallaki P., Lirussi L., Carracedo S., You P., Esbensen Q.Y., Götz A., Jobert L., Alsøe L., Sætrom P., Gagos S., Nilsen H. SMUG1 promotes telomere maintenance through telomerase RNA processing. *Cell Rep.* 2019;28(7):1690-1702.e10. DOI 10.1016/j.celrep.2019.07.040.

Li C., Stoma S., Lotta L.A., Warner S., Albrecht E., Allione A., Arp P.P., Broer L., Buxton J.L., Da Silva Couto Alves A., … Butterworth A.S., Danesh J., Samani N.J., Wareham N.J., Nelson C.P., Langenberg C., Codd V. Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *Am. J. Hum. Genet.* 2020;106(3):389-404. DOI 10.1016/j.ajhg.2020.02.006.

Mangino M., Christiansen L., Stone R., Hunt S.C., Horvath K., Eisenberg D.T., Kimura M., Petersen I., Kark J.D., Herbig U., Reiner A.P., Benetos A., Codd V., Nyholt D.R., Sinnreich R., Christensen K., Nassar H., Hwang S.J., Levy D., Bataille V., Fitzpatrick A.L., Chen W., Berenson G.S., Samani N.J., Martin N.G., Tishkoff S., Schork N.J., Kyvik K.O., Dalgård C., Spector T.D., Aviv A. *DCAF4*, a novel gene associated with leucocyte telomere length. *J. Med. Genet.* 2015;52(3):157-162. DOI 10.1136/jmedgenet-2014-102681.

Mangino M., Richards J.B., Soranzo N., Zhai G., Aviv A., Valdes A.M., Samani N.J., Deloukas P., Spector T.D. A genome-wide association study identifies a novel locus on chromosome 18q12.2 influencing white cell telomere length. *J. Med. Genet.* 2009;46(7):451-454. DOI 10.1136/jmg.2008.064956.

Monaghan P., Ozanne S.E. Somatic growth and telomere dynamics in vertebrates: relationships, mechanisms and consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160446. DOI 10.1098/rstb.2016.0446.

Nagpal N., Wang J., Zeng J., Lo E., Moon D.H., Luk K., Braun R.O., Burroughs L.M., Keel S.B., Reilly C., Lindsley R.C., Wolfe S.A., Tai A.K., Cahan P., Bauer D.E., Fong Y.W., Agarwal S. Small-molecule PAPD5 inhibitors restore telomerase activity in patient stem cells. *Cell Stem Cell*. 2020;26(6):896-909.e8. DOI 10.1016/j.stem.2020.03.016.

Nersisyan L., Nikoghosyan M., Genome of the Netherlands consortium, Arakelyan A. WGS-based telomere length analysis in Dutch family trios implicates stronger maternal inheritance and a role for *RRM1* gene. *Sci. Rep*. 2019;9(1):18758. DOI 10.1038/s41598-019-55109-7.

Oh W., Ghim J., Lee E.W., Yang M.R., Kim E.T., Ahn J.H., Song J. PML-IV functions as a negative regulator of telomerase by interacting with TERT. *J. Cell Sci.* 2009;122(Pt.15):2613-2622. DOI 10.1242/jcs.048066.

Ohira T., Kojima H., Kuroda Y., Aoki S., Inaoka D., Osaki M., Wanibuchi H., Okada F., Oshimura M., Kugoh H. PITX1 protein interacts with ZCCHC10 to regulate hTERT mRNA transcription. *PLoS One.* 2019;14(8):e0217605. DOI 10.1371/journal.pone.0217605.

Pepke M.L., Kvalnes T., Lundregan S., Boner W., Monaghan P., Saether B.E., Jensen H., Ringsby T.H. Genetic architecture and heritability of early-life telomere length in a wild passerine. *Mol. Ecol.* 2021. DOI 10.1111/mec.16288.

Podlevsky J.D., Bley C.J., Omana R.V., Qi X., Chen J.J. The telomerase database. *Nucleic Acids Res.* 2008;36(Database issue):D339-D343. DOI 10.1093/nar/gkm700.

Potts P.R., Yu H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat. Struct. Mol. Biol.* 2007;14(7):581-590. DOI 10.1038/nsmb1259.

Saxena R., Bjonnes A., Prescott J., Dib P., Natt P., Lane J., Lerner M., Cooper J.A., Ye Y., Li K.W., Maubaret C.G., Codd V., Brackett D., Mirabello L., Kraft P., Dinney C.P., Stowell D., Peyton M., Ralhan S., Wander G.S., Mehra N.K., Salpea K.D., Gu J., Wu X., Mangino M., Hunter D.J., De Vivo I., Humphries S.E., Samani N.J., Spector T.D., Savage S.A., Sanghera D.K. Genome-wide association study identifies variants in casein kinase II (*CSNK2A2*) to be associated with leukocyte telomere length in a Punjabi Sikh diabetic cohort. *Circ. Cardiovasc. Genet.* 2014;7(3):287-295. DOI 10.1161/CIRCGENETICS.113.000412.

Schrumpfová P.P., Fajkus J. Composition and function of telomerase-A polymerase associated with the origin of eukaryotes. *Biomolecules.* 2020;10(10):1425. DOI 10.3390/biom10101425.

Seeker L.A., Underwood S.L., Wilbourn R.V., Dorrens J., Froy H., Holland R., Ilska J.J., Psifidi A., Bagnall A., Whitelaw B., Coffey M., Banos G., Nussey D.H. Telomere attrition rates are associated with weather conditions and predict productive lifespan in dairy cattle. *Sci. Rep.* 2021;11(1):5589. DOI 10.1038/s41598-021-84984-2.

Sherman B.T., Hao M., Qiu J., Jiao X., Baseler M.W., Lane H.C., Imamichi T., Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* 2022;50(W1):W216-221. DOI 10.1093/nar/gkac194.

Snow B.E., Erdmann N., Cruickshank J., Goldman H., Gill R.M., Robinson M.O., Harrington L. Functional conservation of the telomerase protein Est1p in humans. *Curr. Biol.* 2003;13(8):698-704. DOI 10.1016/s0960-9822(03)00210-0.

Sobinoff A.P., Pickett H.A. Alternative lengthening of telomeres: DNA repair pathways converge. *Trends Genet.* 2017;33(12):921-932. DOI 10.1016/j.tig.2017.09.003.

Sobinoff A.P., Pickett H.A. Mechanisms that drive telomere maintenance and recombination in human cancers. *Curr. Opin. Genet. Dev.* 2020;60:25-30. DOI 10.1016/j.gde.2020.02.006.

Stuparević I., Novačić A., Rahmouni A.R., Fernandez A., Lamb N., Primig M. Regulation of the conserved 3′-5′ exoribonuclease EXOSC10/Rrp6 during cell division, development and cancer. *Biol. Rev. Camb. Philos. Soc.* 2021;96(4):1092-1113. DOI 10.1111/brv.12693.

Tennen R.I., Bua D.J., Wright W.E., Chua K.F. SIRT6 is required for maintenance of telomere position effect in human cells. *Nat. Commun*. 2011;2:433. DOI 10.1038/ncomms1443.

Tseng C.K., Wang H.F., Burns A.M., Schroeder M.R., Gaspari M., Baumann P. Human telomerase RNA processing and quality control. *Cell Rep.* 2015;13(10):2232-2243. DOI 10.1016/j.celrep.2015.10.075.

Tutton S., Azzam G.A., Stong N., Vladimirova O., Wiedmer A., Monteith J.A., Beishline K., Wang Z., Deng Z., Riethman H., McMahon S.B., Murphy M., Lieberman P.M. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J.* 2016;35(2):193-207. DOI 10.15252/embj.201490880.

Van de Vosse D.W., Wan Y., Lapetina D.L., Chen W.M., Chiang J.H., Aitchison J.D., Wozniak R.W. A role for the nucleoporin Nup170p in chromatin structure and gene silencing. *Cell.* 2013;152(5):969-983. DOI 10.1016/j.cell.2013.01.049.

Vera E., Bernardes de Jesus B., Foronda M., Flores J.M., Blasco M.A. The rate of increase of short telomeres predicts longevity in mammals. *Cell Rep.* 2012;2(4):732-737. DOI 10.1016/j.celrep.2012.08.023.

Wilbourn R.V., Moatt J.P., Froy H., Walling C.A., Nussey D.H., Boonekamp J.J. The relationship between telomere length and mortality risk in non-model vertebrate systems: a meta-analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2018;373(1741):20160447. DOI 10.1098/rstb.2016.0447.

Wright W.E., Piatyszek M.A., Rainey W.E., Byrd W., Shay J.W. Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.* 1996;18(2):173-179. DOI 10.1002/(SICI)1520-6408(1996)18:2<173::AID-DVG10>3.0.CO;2-3.

Zeiger A.M., White M.J., Eng C., Oh S.S., Witonsky J., Goddard P.C., Contreras M.G., Elhawary J.R., Hu D., Mak A.C.Y., Lee E.Y., Keys K.L., Samedy L.A., Risse-Adams O., Magaña J., Huntsman S., Salazar S., Davis A., Meade K., Brigino-Buenaventura E., LeNoir M.A., Farber H.J., Bibbins-Domingo K., Borrell L.N., Burchard E.G. Genetic determinants of telomere length in African American youth. *Sci. Rep*. 2018;8(1):13265. DOI 10.1038/s41598-018-31238-3.

ORCID ID

E.V. Ignatieva orcid.org/0000-0002-8588-6511
N.S. Yudin orcid.org/0000-0002-1947-5554
D.M. Larkin orcid.org/0000-0001-7859-6201

Acknowledgements. The work was supported by the Russian Scientific Foundation, Project No. 22-26-00143, https://rscf.ru/project/22-26-00143/.

Conflict of interest. The authors declare no conflict of interest.

Received November 17, 2022. Revised December 4, 2022. Accepted December 5, 2022.