Beginning at the ends: telomeres and human disease [version 1; referees: 4 approved]

Sharon A. Savage
Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

Abstract
Studies of rare and common illnesses have led to remarkable progress in the understanding of the role of telomeres (nucleoprotein complexes at chromosome ends essential for chromosomal integrity) in human disease. Telomere biology disorders encompass a growing spectrum of conditions caused by rare pathogenic germline variants in genes encoding essential aspects of telomere function. Dyskeratosis congenita, a disorder at the severe end of this spectrum, typically presents in childhood with the classic triad of abnormal skin pigmentation, nail dystrophy, and oral leukoplakia, accompanied by a very high risk of bone marrow failure, cancer, pulmonary fibrosis, and other medical problems. In contrast, the less severe end of the telomere biology disorder spectrum consists of middle-age or older adults with just one feature typically seen in dyskeratosis congenita, such as pulmonary fibrosis or bone marrow failure. In the common disease realm, large-scale molecular epidemiology studies have discovered novel associations between illnesses, such as cancer, heart disease, and mental health, and both telomere length and common genetic variants in telomere biology genes. This review highlights recent findings of telomere biology in human disease from both the rare and common disease perspectives. Multi-disciplinary collaborations between clinicians, basic scientists, and epidemiologist are essential as we seek to incorporate new telomere biology discoveries to improve health outcomes.

Keywords
telomere, dyskeratosis congenita, telomere biology disorder, cancer, Coats plus, Hoyeraal Hreidsarsson syndrome, epidemiology
Introduction

Telomeres are the nucleoprotein complex at chromosome ends with essential roles in maintaining chromosomal integrity. They shorten with each cell division because of incomplete replication of the 3’ ends of DNA and thus are markers of cellular aging. Over the last decade, there has been remarkable growth in the breadth and depth of understanding the multiple roles of telomere biology in human disease. At one end of the spectrum, very rare pathogenic germline genetic variants in telomere biology genes cause exceedingly short telomeres, resulting in dyskeratosis congenita (DC) and its related telomere biology disorders. The other end of the spectrum consists of large-scale population-based studies seeking to determine associations between telomere length, human disease, environmental exposures, or common genetic variants as well as the interactions between these factors.

The complexity of these interactions requires an integrated understanding of telomere basic science, clinical medicine, and epidemiology (Figure 1). Each of these topics is worthy of an in-depth critical review beyond the scope of this article. Instead, I will highlight some key findings and methodologic considerations and discuss where additional research is needed to aid in understanding the contribution of telomere biology to both rare and common human diseases.

Dyskeratosis congenita – a direct connection between germline telomere biology and human disease

DC was first described in a 1906 case report of males with the mucocutaneous triad of abnormal skin pigmentation, nail dystrophy, and oral leukoplakia (Figure 2). Additional similar cases were reported, including the first female case in 1963. Patients with DC also have very high rates of bone marrow failure; stenosis of the esophagus, urethra, or lacrimal ducts (or a combination of these); head and neck squamous cell carcinoma (HNSCC); myelodysplastic syndrome (MDS); acute myeloid leukemia (AML); pulmonary fibrosis; liver disease; avascular necrosis of the hips; and other medical problems (Table 1). DC is inherited in X-linked recessive, autosomal dominant, or autosomal recessive patterns.

Figure 1. Factors associated with human disease are integrally connected to telomere biology. This schematic illustrates the complex relationships between telomere biology, disease, aging, genetics, and environmental exposures, all of which should be considered in studies of telomeres and human disease.

Figure 2. Mucocutaneous features of dyskeratosis congenita of an adult male whose disease is due to a DKC1 mutation. (A) Dystrophic and ridged fingernails. (B) Hyper- and hypo-pigmented skin of neck and upper chest. (C) Irregular leukoplakia of the tongue.
Table 1. Clinical features of telomere biology disorders.

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Key clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyskeratosis congenita (DC)</td>
<td>Mucocutaneous triad of nail dysplasia, abnormal skin pigmentation (hyper/hypopigmented, lacy, reticular pigmentation), and oral leukoplakia. BMF, PF, PAVM, liver disease, avascular necrosis of hips or shoulders (or both), urethral stenosis, lacrimal duct stenosis, esophageal stenosis, HNSCC, MDS, AML, and/or developmental delay. Traditional diagnosis of DC: classic triad or one of the triad, BMF, and two other findings.</td>
</tr>
<tr>
<td>Revesz syndrome</td>
<td>Features of DC plus bilateral exudative retinopathy. Intracranial calcifications and developmental delay also reported.</td>
</tr>
<tr>
<td>Hoyeraal Hreidarsson syndrome</td>
<td>Features of DC plus cerebellar hypoplasia. Immunodeficiency has been reported as presenting problem.</td>
</tr>
<tr>
<td>Coats plus</td>
<td>Bilateral retinopathy, intracranial calcifications, leukodystrophy, anemia, osteopenia, and poor bone healing</td>
</tr>
<tr>
<td>DC-like</td>
<td>Progressive multi-lineage cytopenias, non-immune mediated. May occur in the absence of DC-associated features.</td>
</tr>
<tr>
<td>Aplastic anemia</td>
<td>Cytopenias with cellular dysplasia or clonal chromosomal translocations or both. May occur in the absence of DC-associated features.</td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td>May progress from MDS or aplastic anemia. May occur in the absence of DC-associated features.</td>
</tr>
<tr>
<td>Pulmonary fibrosis</td>
<td>May occur in the absence of DC-associated features.</td>
</tr>
<tr>
<td>Liver fibrosis</td>
<td>Non-alcoholic, non-infectious liver disease. May occur in the absence of DC-associated features.</td>
</tr>
<tr>
<td>Familial melanoma</td>
<td>Multiple family members with melanoma, often early age at onset.</td>
</tr>
<tr>
<td>Familial lymphoproliferative disease</td>
<td>Multiple-affected family members with chronic lymphocytic leukemia, or non-Hodgkin lymphoma</td>
</tr>
<tr>
<td>Li-Fraumeni-like syndrome</td>
<td>Cancer family history notable for angiosarcoma and other cancers</td>
</tr>
</tbody>
</table>

*Key references are noted in Table 2. AA, aplastic anemia; AML, acute myeloid leukemia; BMF, bone marrow failure; DC, dyskeratosis congenital; HNSCC, head and neck squamous cell carcinoma; MDS, myelodysplastic syndrome; PAVM, pulmonary arteriovenous malformation; PF, pulmonary fibrosis.

The first DC genetic locus was mapped to Xq28 in 1996 and specifically to mutations in dyskerin (encoded by DKC1) in 1999–10. The seminal work by Mitchell and Collins was the first to show a connection between telomere biology and human disease through aberrant dyskerin function and the resultant very short telomeres now well known in patients with DC. Currently, DKC1 mutations account for about 25% of classic DC cases. A combination of candidate gene sequencing, genetic linkage studies, and whole exome sequencing occurring over the last 15 years has since identified at least 14 telomere biology genes associated with DC or DC-like phenotypes: telomerase holoenzyme complex (DKC1, TERC, TERT, NOP10, and NHP2), shelterin telomere protection complex (ACD, TINF2, and POT1), telomere capping proteins (CTC1 and STN1), and other proteins that directly or indirectly interact with these key cellular processes (RTF1, NAF1, WRAP53, and PARN) (Figure 3 and Table 2) (reviewed in 6).

The germline mutations in DC-associated telomere biology genes result in very short telomere lengths for age (Figure 4). This knowledge made it possible to develop a diagnostic test for DC, flow cytometry with fluorescent in situ hybridization (flow FISH), in leukocyte subsets. Lymphocyte telomeres less than the first percentile for age are more than 95% sensitive and specific for differentiating patients with DC from their unaffected relatives or patients with other inherited bone marrow failure syndromes.13,14 In addition to aiding diagnosis, using telomeres less than the first percentile for age has greatly added in discovering the genetic causes of DC.15

**Telomere biology disorders – many names connected by pathophysiology**

The discovery of the multiple genetic causes and modes of inheritance has led to a growing appreciation that there is a wide range of clinical phenotypes associated with mutations in telomere biology genes. This started with the identification of TERT and TERC mutations in patients with apparently isolated aplastic anemia or pulmonary fibrosis. As defined above, classic DC is a complex multi-system disorder, but variable penetrance and expressivity of the clinical manifestations have identified a growing number of patients with one or a few features of DC as well as germline mutations in telomere biology genes and short telomeres (Table 1 and Table 2). This spectrum of illnesses has been termed telomeropathy, short telomere syndromes, or telomere biology disorders (TBDs). The last of these, TBD, was proposed and is favored because it is more descriptive and reflective of the underlying biology that unites these disorders.22
Table 2. Germline genetics of telomere biology disorders.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein name(s)</th>
<th>Disorder(s)</th>
<th>Mode of inheritance</th>
<th>Year gene first associated with any disease, relevant reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DKC1</td>
<td>DKC1, dyskerin</td>
<td>DC, HH</td>
<td>XLR</td>
<td>1998&lt;sup&gt;13&lt;/sup&gt;</td>
</tr>
<tr>
<td>TERC</td>
<td>hTr, telomerase RNA component (encodes an RNA)</td>
<td>DC, AA, PF</td>
<td>AD</td>
<td>2001&lt;sup&gt;16,18,24&lt;/sup&gt;</td>
</tr>
<tr>
<td>TERT</td>
<td>TERT, telomerase</td>
<td>DC, AA, MDS, AML, PF, LD, FM</td>
<td>AD</td>
<td>2005&lt;sup&gt;16,17,25–27&lt;/sup&gt;</td>
</tr>
<tr>
<td>NOP10</td>
<td>NOP10, NOLA3, nucleolar protein family A, member 3</td>
<td>DC</td>
<td>AR</td>
<td>2007&lt;sup&gt;28&lt;/sup&gt;</td>
</tr>
<tr>
<td>NHP2</td>
<td>NHP2, NOLA2 nucleolar protein family A, member 2</td>
<td>DC</td>
<td>AR</td>
<td>2008&lt;sup&gt;39&lt;/sup&gt;</td>
</tr>
<tr>
<td>TINF2</td>
<td>TIN2, TERF1 (TRF1)-interacting nuclear factor 2</td>
<td>DC, HH, RS</td>
<td>AD</td>
<td>2008&lt;sup&gt;35&lt;/sup&gt;</td>
</tr>
<tr>
<td>WRAP53</td>
<td>TCAB1, telomere cajal body associated protein 1</td>
<td>DC, HH</td>
<td>AR</td>
<td>2011&lt;sup&gt;30,31&lt;/sup&gt;</td>
</tr>
<tr>
<td>CTC1</td>
<td>CTC1, conserved telomere maintenance component 1</td>
<td>CP, DC</td>
<td>AR</td>
<td>2012&lt;sup&gt;32–34&lt;/sup&gt;</td>
</tr>
<tr>
<td>RTEL1</td>
<td>RTEL1, regulator of telomere elongation helicase 1</td>
<td>PF, AA</td>
<td>AD</td>
<td>2013&lt;sup&gt;35–40&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DC, HH</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td>TERF2IP</td>
<td>RAP1, TERF2-interacting protein</td>
<td>FM</td>
<td>AD</td>
<td>2015&lt;sup&gt;41&lt;/sup&gt;</td>
</tr>
<tr>
<td>PARN</td>
<td>PARN, poly(A)-specific ribonuclease</td>
<td>PF</td>
<td>AD</td>
<td>2015&lt;sup&gt;42–45&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DC, HH</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td>ACD</td>
<td>TPP1, telomere protection protein 1</td>
<td>AA, FM, FLPD</td>
<td>AD</td>
<td>2014&lt;sup&gt;47,48,49&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HH</td>
<td>AR</td>
<td>2016&lt;sup&gt;48&lt;/sup&gt;</td>
</tr>
<tr>
<td>STN1</td>
<td>STN1, CST-complex subunit</td>
<td>CP</td>
<td>AR</td>
<td>2016&lt;sup&gt;49&lt;/sup&gt;</td>
</tr>
<tr>
<td>POT1</td>
<td>POT1, protection of telomeres 1</td>
<td>FM, FLPD, LFL</td>
<td>AD</td>
<td>2014&lt;sup&gt;50–52&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td>AR</td>
<td>2016&lt;sup&gt;49&lt;/sup&gt;</td>
</tr>
<tr>
<td>NAF1</td>
<td>NAF1, nuclear assembly factor 1 ribonucleoprotein</td>
<td>PF</td>
<td>AD</td>
<td>2016&lt;sup&gt;44&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

AA, aplastic anemia; AD, autosomal dominant; AML, acute myeloid leukemia; AR, autosomal recessive; CP, Coats plus; DC, dyskeratosis congenita; FLPD, familial lymphoproliferative disease; FM, familial melanoma; HH, Hoyeraal-Hreidarsson syndrome; LD, liver disease; LFL, Li-Fraumeni-like syndrome; MDS, myelodysplastic syndrome; PF, pulmonary fibrosis; RS, Revesz syndrome; XLR, X-linked recessive.

Figure 3. Proteins encoding key components of telomere biology associated with disease. Protein names are noted in the figure. Associated disease and mode(s) of inheritance are shown in Table 1. The asterisk indicates proteins encoded by genes with single-nucleotide polymorphisms associated with cancer. The pound sign indicates proteins encoded by genes with single-nucleotide polymorphisms associated with telomere length.
The most complex TBDs are those disorders presenting very early in childhood, namely Hoyeraal Hreidarsson syndrome (HH), Revesz syndrome (RS), and Coats plus.\textsuperscript{55,56} In addition to having features of DC, patients with HH have cerebellar hypoplasia and immunodeficiency, whereas those with RS also have bilateral exudative retinopathy. Coats plus, a disorder characterized by retinal and gastrointestinal vascular abnormalities, poor bone healing, leukodystrophy, and cerebellar calcifications, joined the TBD spectrum when its primary cause was identified as autosomal recessive CTC1 mutations.\textsuperscript{32,33,59}

The other end of the clinical spectrum includes patients with middle or later age at onset of pulmonary fibrosis, liver disease, or bone marrow failure and heterozygous germline mutations in NAF1, TERT, TERC, PARN, or RTEL1.\textsuperscript{17,18,42,60} Additionally, it is important to note that most patients do not have all of the DC-associated medical complications. The mucocutaneous triad is diagnostic but varies with the age of onset and is usually progressive over time. Many patients, even members of the same family, may present with just one feature but develop more over time because of variable penetrance and expressivity of germline telomere biology defects.

There is also a growing role of mutations in the shelterin complex and cancer-prone families without DC-related clinical manifestations. Heterozygous rare, pathogenic variants in POT1 resulting in longer telomeres have been reported in familial melanoma, familial chronic lymphocytic leukemia (CLL), and a Li-Fraumeni-like syndrome family.\textsuperscript{50–52} POT1 somatic mutations in CLL have also been associated with CLL outcomes.\textsuperscript{62–65} Familial melanoma has also been associated with germline mutations in ACD (TPP1) and TERF2IP (RAP1).\textsuperscript{41} These studies suggest an interesting dichotomy in clinical phenotypes resulting from long versus short telomeres.

**Telomeres and cancer**

Telomeres are closely connected to cancer biology because of the role they play in chromosomal stability. There is a
detailed body of work in this realm and only a few of the key features will be highlighted herein as they have been reviewed in detail elsewhere.66–73.

The primary hypotheses connecting telomeres and cancer are based on the fact that telomeres shorten with each cell division. In general, cellular senescence or apoptosis is triggered when telomeres reach a critically short state. It likely takes just one critically short telomere on one chromosome arm to trigger these events as suggested in a TERC mouse model by Hemann et al.74. A cellular survival advantage is gained through bypassing apoptosis or senescence through the upregulation of telomerase, inactivation of TP53 or RB or both, initiation of alternative lengthening of telomeres (ALT), and other key biological pathways.76–80. The continued division of cells originally destined for death is hypothesized to lead to continued accumulation of mutations, and sticky chromosome ends due to abnormal telomeres can contribute to chromosomal aneuploidy. Unchecked cellular growth can occur if these genetic aberrations result in a growth advantage.

Activating somatic mutations in the TERT promoter have been described in melanoma, bladder, thyroid, and some central nervous system cancers.81–83. These somatic mutations in the TERT promoter result in increased telomerase expression and suggest that this activation could convey a growth advantage as cancer cells continue to divide despite the presence of aberrant telomeres.

Patients with DC/TBD have significantly increased risks of MDS, AML, HNSCC, and other malignancies.14–16. The 2017 update of cancer in the National Cancer Institute Inherited Bone Marrow Failure Syndrome cohort reported that cancer in patients with DC occurs at an approximately four fold higher incidence and a younger age than the general population.87. MDS and AML occurred at 578- and 24-fold greater incidence, respectively, than the general population. There was also an excess of solid cancers in patients with DC with observed/expected ratios of 74 for any HNSCC and 216 for tongue HNSCC.97. The mechanisms by which cancer develops in patients with DC/TBDs is unknown and represents an important research opportunity.

Cells of patients with TBD already have a “first hit” in a key component of telomere biology. Studies of the next steps in carcinogenesis in patient-derived cells could lead to important insights into the carcinogenic process.

**The advent of telomere molecular epidemiology**

Telomere molecular epidemiology has emerged with the development of high-throughput telomere length measurement methods, genome-wide genotyping platforms, and keen interest in the role of telomere biology in human disease.88,89. These large, often population-based studies seek to determine (1) whether telomere length is associated with disease, (2) whether common genetic variants (that is, single-nucleotide polymorphisms, or SNPs) are associated with telomere length, (3) the degree to which SNPs contribute to telomere biology, and (4) interactions between telomere length, SNPs, and disease or phenotypes (Figure 1 and Table 3). Although a great deal of excitement has been generated by these studies, it is important to point out that differences in telomere length between cases and controls in large population-based studies may be statistically significant but not clinically relevant. “Short telomeres” in a large case-control or cohort study are still within the clinically “normal” range and not nearly as short as telomeres of patients with TBDs (Figure 4).

Robust and accurate telomere length measurement is at the crux of all telomere length association studies. Blood or buccal cell DNA telomere length has been evaluated in a wide array of association studies, including cancer, cardiovascular disease, mental health, inflammatory diseases, environmental exposures, and many other settings. There are numerous methods to determine telomere length in tissues, single cells, and DNA preparations, each appropriate for different applications and reviewed extensively.90,91. Quantitative polymerase chain reaction (qPCR) is amenable to large studies because it uses very small amounts of DNA and can be scaled up rapidly.92,93. However, qPCR telomere assays generate a relative telomere length and are very sensitive to DNA extraction methods and storage.94. These challenges have led to significant challenges in reproducing data in case-control or cohort studies of qPCR relative telomere length and phenotypes.95,96. The telomere restriction fragment

<table>
<thead>
<tr>
<th>Table 3. Features of robust telomere length association studies.</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Strong <em>a priori</em> hypothesis of why telomere biology might be important in disease of interest</td>
</tr>
<tr>
<td>• Comprehensive clinical phenotyping</td>
</tr>
<tr>
<td>• Accurately measured exposure of interest</td>
</tr>
<tr>
<td>• Large sample size with power calculations reported</td>
</tr>
<tr>
<td>• Collection of samples prior to disease onset</td>
</tr>
<tr>
<td>• Detailed information on how samples were collected, processed, and stored</td>
</tr>
<tr>
<td>• Telomere length measurement methods described in detail, especially if any adaptations to published methods</td>
</tr>
<tr>
<td>• Accurate and reproducible telomere length measurement</td>
</tr>
<tr>
<td>• Strong statistical justification of association findings</td>
</tr>
</tbody>
</table>
method uses restriction enzymes to cut the subtelomeric ends of chromosomes in a DNA preparation and is most widely used in basic science laboratories, although a few groups use it in population-based studies\(^ {80,91}\).

**Blood or buccal cell telomere length association studies**

This section highlights just a few key topics within the growing literature of telomere length association studies. For example, individuals of African ancestry have longer telomeres than those of European ancestry and thus ancestry should be accounted for in analyses\(^ {87,98}\). Since self-reported ancestry can be quite variable, genomic approaches may be helpful in classifying cases and controls in order to appropriately adjust for ancestry.

There is also a growing understanding of associations between environmental exposures and telomere length. Smokers have shorter telomeres than non-smokers and thus it is important to adjust for smoking in association analyses\(^ {89}\). Associations between prenatal exposures to smoking and air pollution as well as exposure to certain occupational chemicals have also been explored but with varying results\(^ {94,95}\). In each of these studies, it is essential to precisely quantify the exposure of interest in addition to using a robust and reproducible telomere length measurement.

Early cancer-telomere length association studies suggested shorter telomeres as a cancer risk factor\(^ {98,101}\) but studies of other cancers were not consistent\(^ {108-112}\). Meta-analyses found that most studies with blood or buccal cell DNA collected prior to cancer diagnosis were null but that case-control studies were more likely to find associations\(^ {113,114}\). Similarly, a direct comparison of prospectively and retrospectively collected DNA samples from patients with breast or colorectal cancer reported that telomere shortening occurred primarily after cancer diagnosis\(^ {115}\). Many of these inconsistencies have been attributed to reverse causation bias due to the presence of cancer, underlying inflammation, or prior therapy at the time of sample collection\(^ {115,116}\). Currently, the most consistent studies are those of longer telomeres in pre-diagnostic samples of patients with lung cancer and melanoma\(^ {117,118}\). Interestingly, shorter leukocyte telomeres were associated with overall cancer mortality but not with cancer in a large prospective study of 64,637 individuals who developed 2,420 cancers\(^ {120}\).

There is a growing body of telomere length association studies and different aspects of mental health, including measures of perceived stress in caregivers, exposure to early life adversity, and in patients with schizophrenia, bipolar disorder, and depression\(^ {111-118}\). The biological mechanisms underlying these findings are unknown but current hypotheses include stress responses inducing oxidative stress, resulting in DNA damage and telomere shortening. Notably, abnormalities in brain development are present in patients with HH (cerebellar hypoplasia), RS (intracranial calcifications), and Coats plus (leukodystrophy and intracranial calcifications)\(^ {95,96}\). The only study to date of neuropsychiatric complications in DC found higher-than-expected occurrence of developmental delay and psychiatric disorders\(^ {22}\). Studies of DC/TBD patients by psychiatrists and neurobiologists constitute an unstudied area highly likely to generate important insights into telomere biology and brain development.

There is also a great deal of interest in using telomere length as a measure of biological age and even in modulating telomere length through lifestyle interventions. Numerous studies suggest associations between lifestyle, exercise, and telomere lengths\(^ {128-133}\). However, a recent review suggests that telomere length in and of itself is not sufficient as a specific aging biomarker\(^ {134}\).

Current data consistently report shorter leukocyte telomeres in individuals with atherosclerotic cardiovascular disease than in unaffected controls\(^ {135,137}\). The biology underlying this association is thought to be related to chronic inflammation and oxidative stress coupled with aging of the vasculature. This prompted Aviv et al. to propose a model whereby age-dependent telomere shortening varies on the basis of the replicative needs of the specific tissue\(^ {136,139}\). They hypothesize that skeletal muscle (a minimally replicative cell type) may represent telomere length closer to the time of birth and that the gap between skeletal muscle and leukocyte (a rapidly dividing cell type) telomere length attrition could serve to aid understanding of the associations between telomere length and human disease with each patient, in effect, serving as their own control. The first such study testing this hypothesis showed that increased attrition of telomeres in leukocytes was associated with atherosclerotic cardiovascular disease\(^ {132}\).

**Single-nucleotide polymorphisms, telomere biology genes, and disease**

The advent of genome-wide association studies (GWAS) opened the door to understanding associations between common genetic variants (that is, greater than 1% minor allele frequency, SNPs) and human disease or phenotypes (Figure 5) (reviewed in 141). GWAS genotype hundreds of thousands of SNPs in thousands of cases and controls and use methods to fine-tune risk estimates through large-scale replication studies and polygenic risk score computation\(^ {142}\).

Numerous GWAS of cancer etiology have identified variants in telomere biology genes as being associated with cancer risk or outcomes. SNPs in the **TERT-CLPTM1L** locus on chromosome 5p15.33 are associated with multiple cancer types, including lung, pancreatic, breast, bladder, ovarian, prostate, and testicular germ cell cancers as well as glioma, melanoma, and non-melanoma skin cancers\(^ {143,144}\). There are specific regions of this locus associated with different cancers, but these variants do not specifically encode deleterious coding alleles in **TERT**. They do, however, appear to be connected to telomere length through long-range regulation of this locus\(^ {145}\).

SNPs in **RTEL1** are associated with glioma in large GWAS of this rare brain cancer. The glioma-associated **RTEL1** SNPs are intronic, but functional studies have not yet been completed to understand their potential functions\(^ {146-152}\). These findings are intriguing because patients with DC or HH due to autosomal recessive **RTEL1** mutations often have abnormal brain development.
in the form of cerebellar hypoplasia\textsuperscript{15–17,153}. Although the specific genetic loci are different, it is intriguing to speculate that there could be an important biological connection between these findings.

In addition to GWAS of cancer or other illnesses, several GWAS have been conducted to identify novel loci associated with telomere length. SNPs in known telomere biology genes, including \textit{TERT}, \textit{OBFC1} (encodes STN1), \textit{CTC1}, \textit{TERC}, \textit{NAF1}, and \textit{RTEL1}, as well as genes not previously known to be associated with telomere biology have been discovered\textsuperscript{154–157}. These studies illustrate the complexities of telomere length regulation by showing that even common genetic variants, and especially combinations of common genetic variants, are associated with telomere length in the general population.

The existence of telomere length GWAS in various populations set the stage for even larger studies using Mendelian randomization methods in which telomere length–associated SNPs serve as surrogates for telomere length\textsuperscript{158}. One such study used nine telomere length–associated SNPs to create a telomere length surrogate score and found longer telomere length scores associated with lung adenocarcinoma but not the other cancers\textsuperscript{159}. Renal cell carcinoma, one of the cancers with reproducible telomere length association data, was also studied using nine telomere length surrogate SNPs and it was found that genetically longer telomerres were associated with renal cell carcinoma\textsuperscript{160}.

In 2017, a Mendelian randomization study of 16 telomere length–associated SNPs from 103 GWAS with summary data on 35 cancers and 48 non-neoplastic diseases found that genetically longer telomeres associated with elevated risk of many cancers, including glioma, ovarian cancer, lung cancer, neuroblastoma, bladder, skin, testicular germ cell cancer and kidney cancer, and endometrial cancer\textsuperscript{161}. That study also found an association between genetically shorter telomeres and the risk of interstitial lung disease, celiac disease, abdominal aortic aneurysm, and coronary heart disease but not of other inflammatory or psychiatric diseases\textsuperscript{161}.

Although several studies suggest that telomere length is associated with depression, one study using Mendelian randomization and three SNPs—one each in \textit{TERT}, \textit{TERC}, and \textit{OBFC1}—as surrogates for telomere length in 67,000 individuals did not find an association between depression and genetically shorter telomeres\textsuperscript{162}. These investigators used the same three SNPs to investigate genetically predicted telomere length and risk of

**Figure 5. Relationship between telomere length, variant allele frequency, and human disease.** The majority of genetic variants associated with common disease have a minor allele frequency (MAF) greater than 1% and telomeres in the “normal” range (that is, between the 1st and 99th percentiles for age). In contrast, genetic variants associated with rare and more highly penetrant disease are rare with MAF often much less than 1% and the extremes of telomere length. AA, aplastic anemia; CLL, chronic lymphocytic leukemia; CP, Coats plus; DC, dyskeratosis congenita; FM, familial melanoma; HH, Hoyeraal Hreidarsson syndrome; LD, liver disease; LFL, Li-Fraumeni-like syndrome; MDS, myelodysplastic syndrome; PF, pulmonary fibrosis; RS, Revesz syndrome.
ischemic heart disease\textsuperscript{13}. They found small but statistically significant associations in a dataset of 60,837 ischemic heart disease cases compared with controls.

The studies briefly described above have generated a great deal of enthusiasm but are not without limitations. In many instances, qPCR was used to measure the telomere lengths in GWAS and this assay can be variable between studies. The sensitivity of the assay telomere length measurement and relatively small contributions of SNPs to telomere length should be considered in interpreting large-scale telomere length Mendelian randomization studies.

**The way forward**

The connections between telomere biology and human disease are complex and myriad and require a multi-disciplinary approach to truly understand the clinically relevant data, important basic science questions, and implications of epidemiologic analyses (Figure 1). As protectors of chromosome ends, telomeres are clearly integral to all aspects of cell biology. They are markers of biological aging and are regulated by a wide range of proteins.

Both very rare and very common germline genetic variants in telomere biology genes are associated with human disease, although the specific clinical phenotypes comprise a wide-ranging disease spectrum. Inheritance of telomere length inheritance and epigenetic regulation are also important aspects of telomere biology and should be incorporated into collaborative studies of rare and common telomere phenotypes. Additionally, optimization of telomere length measurement methods and improved understanding of environmental factors contributing to telomere biology will be essential in order to thoroughly understand these complexities. It is of the utmost importance for clinicians, epidemiologists, and basic scientists, all of whom study telomeres for a wide variety of different yet important reasons, to work together to build upon the expertise they each possess and incorporate that into improved understanding of telomere biology in human disease. This multi-disciplinary approach will enable the discovery of therapeutics and disease prevention modalities effective for patients with TBDs and for the general population.

**Abbreviations**

AML, acute myeloid leukemia; CLL, chronic lymphocytic leukemia; DC, dyskeratosis congenita; GWAS, genome-wide association study; HH, Hoyeraal Hreidarsson syndrome; HNSCC, head and neck squamous cell carcinoma; MDS, myelodysplastic syndrome; qPCR, quantitative polymerase chain reaction; RS, Revesz syndrome; SNP, single-nucleotide polymorphism; TBD, telomere biology disorder.

**Competing interests**

The author declares that she has no competing interests.

**Grant information**

This work was supported by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgments**

The author acknowledges the invaluable contributions of all patients and their families who have taught us so much about the complexities of telomere biology disorders in all their forms.

---

**References**


74.  Harley CB, Futcher AB, Greider CW:
69.  Reddel RR:
65.  Pepper C, Baird D, Fegan C:
63.  Véronèse L, Tournilhac O, Callanan M,
62.  Ramsay AJ, Quesada V, Foronda M,
61.  Savage SA:
59.  Savage SA:
55.  Savage SA:

therapeutics in cancer.
PubMed Abstract
PubMed Abstract
PubMed Abstract
PubMed Abstract

44
bone marrow and integument.
leukoencephalopathy, slow pre- and post-natal linear growth and defects of
dysfunction in chronic lymphocytic leukemia.
PubMed Abstract
PubMed Abstract

43

et al.

18

Nature.
Publisher Full Text

17

Publisher Full Text

16

Publisher Full Text

15

Publisher Full Text

14

Publisher Full Text

13

Publisher Full Text

12

Publisher Full Text

11

Publisher Full Text

10

Publisher Full Text

9

Publisher Full Text

8

Publisher Full Text

7

Publisher Full Text

6

Publisher Full Text

5

Publisher Full Text

4

Publisher Full Text

3

Publisher Full Text

2

Publisher Full Text

1

Publisher Full Text


Open Peer Review

Current Referee Status: ✔ ✔ ✔ ✔

Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

**Version 1**

1. **Fadi Charchar** School of Applied and Biomedical Sciences, Faculty of Science and Technology, Federation University, Ballarat, Victoria, Australia  
   **Competing Interests:** No competing interests were disclosed.

2. **Steven E. Artandi** Department of Medicine, Stanford University School of Medicine, Stanford, California, USA; Cancer Biology Program, Stanford University School of Medicine, Stanford, California, USA; Department of Biochemistry, Stanford University School of Medicine, Stanford, California, USA  
   **Competing Interests:** No competing interests were disclosed.

3. **Tracy Bryan** Children’s Medical Research Institute, University of Sydney, Sydney, New South Wales, Australia  
   **Competing Interests:** No competing interests were disclosed.

4. **Dirk Hockemeyer** Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California, USA  
   **Competing Interests:** No competing interests were disclosed.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com