UMEÅ UNIVERSITY

EPIGENETIC AND TELOMERE ANALYSES IN DIFFUSE LARGE B-CELL LYMPHOMA AND TELOMERE BIOLOGY DISORDERS

Olivia Carlund

Department of Medical Biosciences
Umeå 2024

Responsible publisher under Swedish law: the Dean of the Medical Faculty

Dissertation for PhD
ISBN: 978-91-8070-431-1 (print)
ISBN: 978-91-8070-432-8 (pdf)
ISSN: 0346-6612
New Series Number 2315
Information about cover photo: Acrylic painting by my friend Karolina Eriksson
Electronic version available at: http://umu.diva-portal.org/
Printed by: Cityprint i Norr AB
Umeå, Sweden 2024

Ever since I was a child I've had the ability of going into hyperfocus. To stay there, I listen to music. It has to be the exact same song as I was listening to when the hyperfocus started, which means the song can be on repeat for hours. I wasn't aware of this habit until my flatmate, many years ago, came in to my room and asked me to *please* put on something else because he was going insane. Nowadays I use headphones. One of the songs I've had on repeat is "Gethsemane (I only want to say)" from Jesus Christ Superstar, in which Jesus is filled with internal struggle over his destiny. Or, if you will, it is the story of being a PhD student.

Then, I was inspired
Now, I'm sad and tired
Listen, surely I've exceeded expectations
Tried for three years, seems like thirty
Could you ask as much from any other man

-Andrew Lloyd Webber & Tim Rice

Table of contents

Abstract

Telomere biology and epigenetics are critical in cellular aging and malignant transformation. Telomeres at the end of chromosomes shorten during cellular replication which eventually induces cellular senescence. The telomeres can partially be restored by the telomerase enzyme, expressed by a few normal and most malignant cells. Telomere biology disorders (TBD) are caused by mutations (variants) in telomere-associated genes. TBD is characterized by short telomeres and heterogeneous clinical presentation, ranging from mild anemia to bone marrow failure and pulmonary fibrosis. However, several genetic variants in suspected TBD are classified as variants of unknown significance (VUS). VUS presents a dilemma since they are not actionable in clinical practice. Epigenetics involves chemical modifications of DNA, RNA, and proteins that alter the cellular phenotype without changing the DNA sequence. DNA methylation (DNAm) alterations are crucial in disease progression and malignant transformation. DNAm in large B-cell lymphoma is characterized by promoter CpG island (CGI) hypermethylation and global hypomethylation. DNAm is less studied in TBD. The overall aim of this thesis was to investigate alterations in telomere biology and epigenetics to improve the understanding of underlying factors contributing to TBD and large B-cell lymphoma.

We identified altered DNAm profiles in blood cells from TBD patients (n=35) compared to healthy controls (n=20). These changes were most prominent in symptomatic patients, regardless of telomere length, suggesting that DNAm alterations in blood cells could be involved in disease progression. However, we cannot exclude the possibility that severe telomere shortening lead to these DNAm changes, supported by the findings of altered DNAm in patients with extremely short telomere length and no hematological symptoms. Furthermore, seven genes of interest were identified. *PRDM8*, *SMC4, VARS*, and *WNT6* have previously been linked to TBD or telomere length. *MAS1L*, *NAV2*, and *TM4FS1* were novel in TBD. The functional relevance of these genes in terms of gene expression, telomere maintenance, and disease progression in TBD requires further evaluation.

We identified extensive DNAm alterations in tumor samples from patients with diffuse large B-cell lymphoma not otherwise specified (DLBCL, n=66), high-grade B-cell lymphoma (n=7), primary CNS lymphoma (n=8), and transformation from an indolent B-cell lymphoma to DLBCL (n=12). These entities had extensive semimethylation that was absent in normal cells and other B-cell neoplasms. Short telomere length and a high percentage of global hypomethylation were both independent prognostic factors for disease-specific survival in our cohort. The subpopulation with the highest percentage of global

hypomethylation also had a high percentage of hypermethylated CGIs. These methylation alterations could potentially be targets for epigenetic therapy, including hypomethylating agents.

We developed a functional analysis of telomerase activity (TA) based on the telomeric repeat amplification protocol combined with qPCR to evaluate genetic variants in TBD. TA was measured in activated T-cells from controls (n=100) and TBD patients (n=6). The genetic variants were classified as pathogenic (*TERT*, n=1) or likely-pathogenic (*TERT*, n=4 and *TERC*, n=1) following consensus guidelines from the American College of Medical Genetics and Genomics. TA was reduced in activated T-cells from TBD patients. Pathogenicity was supported for variants with a TA of more than 3 SD below the mean TA of controls (*TERT*, n=3). Functional analysis of TA in patient-derived cells could support pathogenicity evaluation and reduce the number of reported VUS in TBD.

Abbreviations

ACMG	The American College of Medical Genetics and Genomics
AID	Activation-induced cytidine deaminase
ALT	Alternative lengthening of telomeres
AML	Acute myeloid leukemia
A-TBD	Asymptomatic telomere biology disorder
CNS	Central nervous system
CSR	Class-switch recombination
CGI	CpG island
CIMP	CpG island methylator phenotype
Ct	Cycle threshold
CpG	Cytosine-phosphate-guanine
DM-CpGs	Differentially methylated CpGs
DLBCL	Diffuse large B-cell lymphoma (not otherwise specified)
DLBCL-ABC	Activated B-cell-like diffuse large B-cell lymphoma
DLBCL-GC	Germinal center B-cell-like diffuse large B-cell lymphoma
DLBCL-nonGC	Activated B-cell-like diffuse large B-cell lymphoma + unclassifiable diffuse large B-cell lymphoma
DSS	Disease-specific survival
DNAm	DNA methylation
ES-RTL	Extremely short relative telomere length
FISH	Fluorescence *in situ* hybridization
GC	Germinal center

H-TBD	Hematological telomere biology disorder
HSC	Hematopoietic stem cell
HGBL	High-grade B-cell lymphoma
IPF	Idiopathic pulmonary fibrosis
Ig	Immunoglobulin
LBCL	Large B-cell lymphoma
MNC	Mononuclear cells
MDS	Myelodysplastic syndrome
MVS	Methylation variability score
NH-TBD	Non-hematological telomere biology disorder
OS	Overall survival
BCP-ALL	Precursor B-cell acute lymphoblastic leukemia
PCNSL	Primary CNS lymphoma
PFS	Progression-free survival
qPCR	Quantitative polymerase chain reaction
RTL	Relative telomere length
R-CHOP	Rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone
S-RTL	Short relative telomere length
SHM	Somatic hypermutation
T-ALL	T-cell acute lymphoblastic leukemia
T-LBCL	T-cell acute lymphoblastic lymphoma
TA	Telomerase activity

TBD	Telomere biology disorder
TERT	Telomerase reverse transcriptase
TERC	Telomerase RNA component
t-DLBCL	Transformation from indolent lymphoma to diffuse large B-cell lymphoma
TRAP	Telomeric repeat amplification protocol
VUS	Variant of unknown significance

Original Papers

This thesis was based on the following papers, referred to in the text by their roman numerals.

Paper I
Carlund O, Norberg A, Osterman P, Landfors M, Degerman S, Hultdin M. DNA methylation variations and epigenetic aging in telomere biology disorders. Sci Rep. 2023;13(1):7955.

Paper II
Carlund O, Thörn E, Osterman P, Fors M, Dernstedt A, Forsell MNE, Erlanson M, Landfors M, Degerman S, Hultdin M. Semimethylation is a feature of diffuse large B-cell lymphoma, and subgroups with poor prognosis are characterized by global hypomethylation and short telomere length. Clin Epigenetics. 2024;16(1):68.

Paper III
Carlund O, Norberg A, Osterman P, Andersson I, Eriksson A, Degerman S, Hultdin M. Functional analysis of telomerase activity in T-lymphocytes as a diagnostic tool for pathogenicity assessment of novel genetic variants in telomere biology disorders.
Manuscript

Enkel sammanfattning på svenska

I den här avhandlingen undersökdes två grundläggande kännetecken för cellulärt åldrande och cancerutveckling: telomerbiologi och epigenetik. *Telomerer* är sekvenser längst ut på DNA:t som skyddar det från att brytas ner genom att interagera med proteinkomplexet shelterin. Telomererna kan inte kopieras fullt ut vid celldelning vilket leder till att de gradvis förkortas. När telomererna nått en kritiskt kort längd kan cellen inte längre dela sig och hamnar i ett vilande tillstånd. Enzymet *telomeras* kan återställa delar av telomersekvensen efter celldelning och på så sätt sakta ner cellens åldrande. Telomeras uttrycks endast av ett fåtal celler, bland annat könsceller, stamceller och aktiverade lymfocyter (T- och B-celler). Telomeras uttrycks också av majoriteten av cancerceller vilket leder till att deras celldelning inte stannar av. *Epigenetik* beskriver kemiska förändringar i DNA:t som påverkar genuttryck och fenotyp utan att ändra DNA-sekvensen. En av dessa förändringar är DNA metylering (DNAm) där en metylgrupp binder till kvävebaskombinationen cytosin-guanin (CpG). CpG sekvenserna är i regel metylerade och utspridda med långa avstånd över hela DNA-strängen. Undantaget är täta regioner av CpGs (CpG-öar (CGI)) som oftast är ometylerade och återfinns i promotorregioner (sekvens som reglerar genuttryck). Cellernas DNAm-mönster är dynamiskt och påverkas av bland annat genetik, miljö- och livsstilsfaktorer. Cancercellers DNAm-mönster skiljer sig kraftigt från normala celler, tillexempel är deras CGI ofta metylerade medan utspridda CpGs är ometylerade. Detta leder till förändrat genuttryck och fenotyp vilket bidrar till cancercellernas överlevnad.

Mutationer i telomerrelaterade gener (till exempel gener som kodar för proteinerna i telomeras och shelterin) kan leda till telomerbiologiska sjukdomar (TBD). TBD är ovanliga sjukdomar som generellt kännetecknas av extremt korta telomerer och åldersrelaterade symptom. Symptomspektrumet är brett, från tidigt grånande hår till hud- och slemmhinnepåverkan, benmärgssvikt och lungfibros. År 2023 hade mutationer (eller varianter) i 17 gener kopplats till TBD. Det finns dock ett stort antal patienter med varianter av oklar signifikans (VUS). VUS:ar är problematiska då det fortfarande är osäkert om de är sjukdomsorsakande eller inte.

Storcelliga B-cellslymfom (LBCL) härstammar från B-celler som utgör en del av det adaptiva (specifika) immunförsvaret. B-celler är de enda cellerna i kroppen som uttrycker antikroppar vars uppgift är att identifiera kroppsfrämmande ämnen. LBCL är aggressiva cancerformer som främst drabbar äldre. Majoriteten av LBCL är av typen diffust storcelligt B-cells lymfom utan närmare specifikation (DLBCL). DLBCL har förstorade B-celler med ett diffust tillväxtmönster och en brokig genetisk, fenotypisk och klinisk presentation. DLBCL delas in i tre

undergrupper baserat på genuttryck: DLBCL-ABC, DLBCL-GC och oklassificerbar DLBCL. Prognosen är generellt god och ungefär 2/3 av patienterna blir friska efter primärbehandling. Däremot är prognosen ogynnsam för de patienter som inte svarar på behandling. Andra, mer ovanliga, undergrupper av LBCL är högmaligna B-cellslymfom med *MYC*, *BCL2*, och/eller *BCL6* rearrangemang (HGBL), primärt storcelligt CNS lymfom (PCNSL) och indolenta (långsamt växande) lymfom som transformerar till DLBCL (t-DLBCL). Nya prediktiva biologiska markörer behövs för att kunna stratifiera patienter i prognostiskt relevanta grupper.

Det övergripande målet för den här avhandlingen var att undersöka förändringar i telomerbiologi och epigenetik för att få en djupare förståelse kring hur de bidrar till TBD och LBCL. Specifikt så undersöktes om DNAm-mönster kunde bidra med förståelse kring den varierande symptombilden i TBD (artikel I). Om DNAm-mönster och telomerlängd var relevanta för subklassificering och prognos av LBCL (artikel II). Samt om mätning av telomerasaktivitet var relevant vid undersökning av telomer-relaterade gener vid misstänkt TBD (artikel III).

Vi identifierade DNAm-förändringar i blodceller från TBD patienter (n=35) jämfört med friska kontroller (n=20). Dessa förändringar var speciellt markanta i symptomatiska patienter, oavsett telomerlängd, och kan styrka att DNAm-förändringar i blod är en del av sjukdomsutvecklingen. DNAm förändringar fanns också i patienter med extremkort telomerlängd utan hematologiska symptom. Vi kan därför inte utesluta att förändringarna uppkom till följd av allvarlig telomerförkortning. Vi identifierade också sju gener av intresse. Fyra av dessa (*PRDM8*, *SMC4, VARS* och *WNT6*) har tidigare kopplats till TBD eller telomerlängd. De andra tre generna (*MAS1L*, *NAV2* och *TM4FS1*) har ingen tidigare känd koppling till TBD. Om DNAm-förändringar i dessa gener påverkar genuttryck, mekanismer för att bibehålla telomerer samt sjukdomsutveckling i TBD bör undersökas i framtida studier.

Vi identifierade DNAm-förändringar i tumörvävnad från patienter med DLBCL (n=66), HGBL (n=7), PCNSL (n=8) och t-DLBCL (n=12). LBCL patienterna hade omfattande semimetylering (det vill säga metylerade i endast en fraktion av cellerna från samma tumör) jämfört med normala B-celler och andra B-cells maligniteter. Kort telomerlängd och en hög procent av global hypometylering (ometylerade CpGs) var självständiga prognostiska markörer för sjukdomsspecifik överlevnad. Patienter med en hög procent av global hypometylering hade också en hög procent av hypermetylerade CGIs. Dessa metyleringsförändringar skulle potentiellt kunna återställas med läkemedel som riktar sig mot epigenetiska förändringar.

Vi utvecklade en funktionell analys av telomerasaktivitet (TA) för att utvärdera genetiska varianter i TBD. TA och celldelningsförmånga (S-fas %) mättes i aktiverade T-celler från friska kontroller (n=100) och TBD patienter (n=6). De genetiska varianterna var klassade som sjukdomsorsakande (*TERT*, n=1) eller troligtvis sjukdomsorsakande (*TERT*, n=4 and *TERC*, n=1) enligt rekommendationer från the American College of Medical Genetics and Genomics. TA var nedsatt i aktiverade T-celler från TBD patienter. Att en variant var sjukdomsorsakande kunde stödjas för varianter med ett TA som avvek med 3 standardavvikelser eller mer från kontrollernas medelvärde (*TERT*, n=3). Vi föreslår att funktionell analys av TA i celler från patienter kan stödja utredningen av genetiska varianter i TBD vilket skulle kunna reducera antalet varianter av oklar signifikans.

Background

Aging is an inevitable part of life and a major risk factor for disease. As we age, our cells accumulate damage which eventually leads to cellular senescence and apoptosis [1]. Age is most commonly described as time since birth, known as *chronological age*. However, it can also refer to an individual's overall health and physical condition based on biological markers, known as *biological age*. Two individuals with identical chronological age can have widely different biological ages. A young biological age is beneficial while an increased biological age is associated with several health risks [2,3]. Importantly, since biological age is dynamic, it can be reversed, making it an interesting research subject. Several hallmarks of aging collectively contribute to the progressive decline in cellular function and vulnerability to age-associated diseases [4]. A number of the hallmarks of aging are similar to the hallmarks of cancer [5]. Genomic instability, epigenetic alterations, chronic inflammation, and dysbiosis promote both aging and malignant transformation. In contrast, telomere attrition and stem cell exhaustion promote cellular aging but suppress malignant transformation [6]. In this thesis, the research mainly focused on two of the hallmarks involved in aging and malignant transformation: telomeres and epigenetics.

Telomere biology

The telomere structure and its role in replicative senescence

Human DNA is linear, and thus, the DNA strand has ends. The ends are called *telomeres* and consist of a repetitive TTAGGG pattern that stretches between 5-15 kilobases [7]. The telomeres end in a 3' single-strand overhang that folds into a major telomeric (T) and a minor displacement (D) loop. The loop formation is assisted by the shelterin complex, which consists of six proteins (TRF1, TRF2, TINF2, RAP1, POT1, and TPP1) that interact with either the double-stranded or single-stranded telomere sequence. The interaction between the telomeres and the shelterin complex is crucial, since it prevents the telomere ends from being recognized as double-stranded breaks by the DNA damage repair machinery (Figure 1) [8-12].

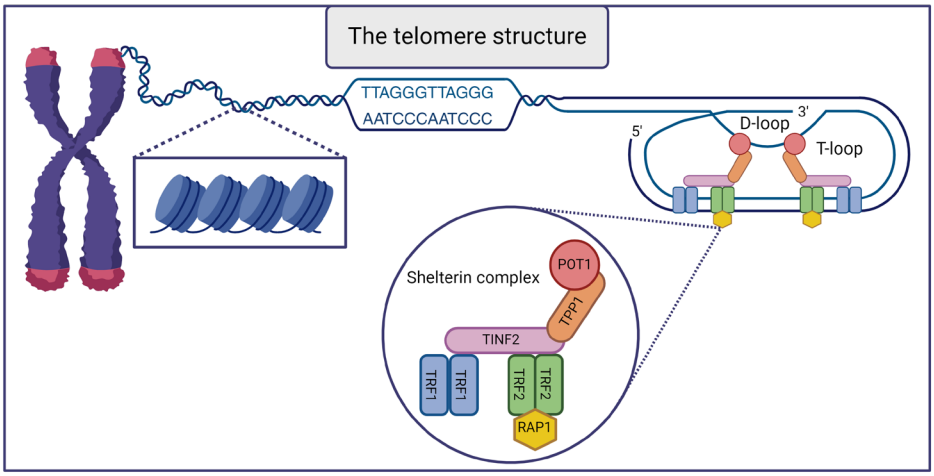


Figure 1. The telomeres are located at the end of chromosomes. They consist of a repetitive TTAGGG pattern in the 5' to 3' direction that stretches between 5-15 kilobases. The telomeres end in a 3' single-strand overhang that folds into a major telomeric (T) and a minor displacement (D) loop. The loop formation is assisted by the shelterin complex, which consists of six proteins (TRF1, TRF2, TINF2, RAP1, POT1, and TPP1) that interact with either the double-stranded or single-stranded telomere sequence. Created with BioRender.com.

The DNA double helix unwinds during cell replication and each DNA strand acts as a template for the synthesis of a new complementary strand. The complementary strand is copied continuously at the leading strand and in fragmented pieces at the lagging strand (Figure 2). The telomeric 3' single-strand overhang is a result of incomplete replication of the lagging strand, a phenomenon known as *the end-replication problem* [11,13,14]. Due to the end-replication problem, the telomeres lose around 50-200 base pairs per cell division. When the telomeres become critically short, the protective loop structure cannot be maintained, and the cell enters senescence or apoptosis [15]. The replicative inhibition of cells is an essential function since it prevents chromosomal end-to-end fusion that can cause genomic rearrangements, acquired mutations, and genomic instability [16]. Thus, replicative senescence is a tumor-suppressor mechanism. Maintaining telomeres at a length that does not induce cellular senescence or apoptosis, i.e. immortalization, is one of the hallmarks of cancer [17].

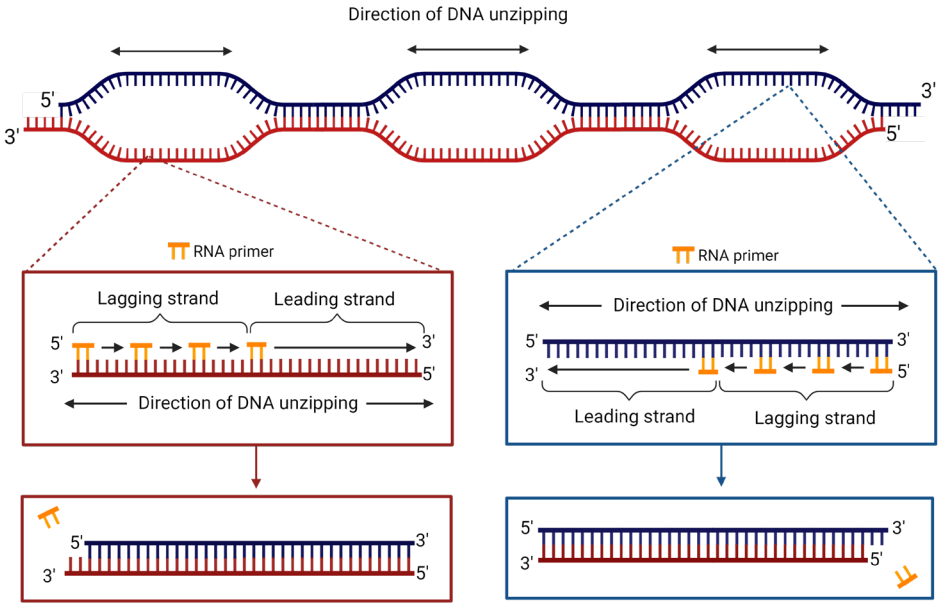


Figure 2. The DNA double helix unwinds during cell replication. The unwinding occurs at several places simultaneously and the DNA is unzipped in two opposite directions. RNA primers bind to the single-stranded DNA and recruit DNA polymerase, which synthesizes complementary DNA strands using the old strands as templates. DNA polymerase works in the 5' to 3' direction; thus, it can only add nucleotides at the 3' end of the chain. The DNA sequence will be synthesized continuously in one direction (the leading strand, the polymerase can keep adding nucleotides to the telomeric sequence as it unwinds). However, the DNA sequence will be synthesized in sections in the other direction (the lagging strand, the polymerase has to wait for a new primer to bind to the DNA sequence before it can add nucleotides). When the last primer is removed from the lagging strand, the polymerase cannot fill in the gap. This phenomenon is known as the end-replication problem, resulting in the telomeric 3' overhang [18]. During telomere replication, the newly synthesized lagging strand will have a 3' overhang (bottom left red box) while the newly synthesized leading strand will have blunt ends (bottom right blue box). The 5' end at the leading strand is digested by nucleases and DNA helicases to generate the essential 3' overhang [19]. Created with BioRender.com.

Telomere dynamics in hematological cells

Hematological cells (i.e. blood cells) serve as surrogates for tissues in situations where invasive sampling methods are practically and ethically challenging. This includes studies of telomere dynamics, such as telomere length measurements [20-24]. Although there are tissue-specific differences in telomere dynamics, there is a positive correlation between telomere length in hematological cells and telomere length in other tissues, e.g. lung, colon, pancreas, skin, and kidney [24]. Telomere

dynamics in hematological cells are studied in leukocytes (granulocytes, monocytes, and lymphocytes), peripheral blood mononuclear cells (monocytes and lymphocytes), or sorted cells.

Telomere length and attrition rate are highly individual and influenced by inheritance, gender, and ethnicity. Telomere length at birth is partially inherited from the parents, but whether the influence is mainly paternal [25-27] or maternal [28,29] is not established. Further, there are reported gender and ethnic differences in telomere length, although some results are conflicting [24,25,27,29-32]. Several studies have reported longer leukocyte telomere length in women compared to age-matched men [27,30-32], while other studies have not identified gender-specific differences [25,29,30]. Longer telomere length has also been observed in non-white populations in both adults and children, which could be influenced by genetic differences [24,30]. The telomere attrition rate in hematological cells is non-linear and accelerated from birth to adulthood, with the most rapid attrition rate observed during the very first years of life [29,32-37]. Additionally, there is evidence that the attrition rate, and thus the telomere length, in hematological cells are cell-type specific [32,38,39]. Taken together, there are many factors influencing telomere length. On top of the genetic influence, telomere dynamics are also affected by environmental and life style factors such as air pollution, stress, dietary habits, and smoking [40-42].

The telomerase enzyme

Is it possible to slow down the telomere attrition and thus prevent cellular aging? There are some cells that can counteract the telomere attrition rate by expressing the enzyme *telomerase*, which adds nucleotides to the telomere sequence after cell division [43,44]. The telomerase enzyme is a reverse transcriptase protein that consists of the core components telomerase reverse transcriptase (TERT) and telomerase RNA component (TERC). TERT and TERC interacts with several accessory proteins, e.g. dyskerin (*DKC1*), TCAB1 (*WRAP53*), NHP2, NOP10, and GAR1 (Figure 3) [45]. TERT has four domains that are highly conserved throughout eukaryotic species: the TERT amino-terminal domain (TEN), the TERT RNA-binding domain (TRBD), the reverse transcriptase domain (RT), and the carboxyl-terminal extension domain (CTE). TERC has three motifs that are shared by all eukaryotic species: the template, the pseudoknot, and a stem terminus element, although the size and sequence of *TERC* differ between species. Additionally, vertebrate TERC has an H/ACA motif, which interacts with small non-coding RNAs [46,47]. The telomerase enzyme is activated through the interaction of the TRBD with TERCs pseudoknot and stem terminus element. TERC further interacts with the H/ACA-motif binding proteins dyskerin, NHP2, NOP10, and GAR1 [47,48]. Lastly, TCAB1 is recruited to the telomerase complex to locate the enzyme to the telomeric ends [49].

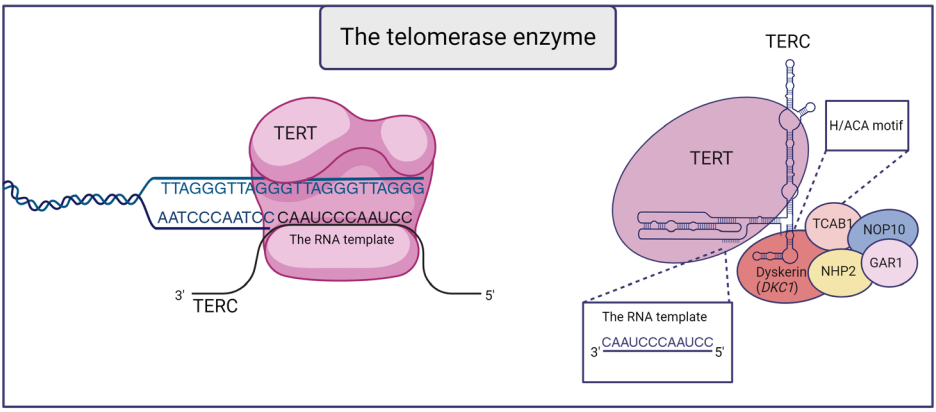


Figure 3. The telomerase enzyme is a reverse transcriptase protein that consists of the core components telomerase reverse transcriptase (TERT) and telomerase RNA component (TERC). TERC associates with the H/ACA-motif binding proteins dyskerin, NHP2, NOP10, and GAR1. TCAB1 is recruited to the telomerase complex to locate the enzyme to the telomeric ends. Created with BioRender.com.

Telomerase is expressed during early human embryogenesis but becomes silenced in most somatic cells throughout development, except for germ cells, stem cells, and activated lymphocytes [50-52]. Thus, telomerase activity in normal cells is usually measured in activated lymphocytes, which express telomerase as long as they are presented to a mitogen (a substance that stimulates mitosis) [39,51,53,54]. It is worth noticing that although telomerase is expressed in some cell types, this is not enough to counteract telomere attrition [55].

Telomere biology disorders

Pathogenic mutations in telomere-associated genes can cause telomere dysfunction and lead to pathological telomere shortening. *Telomere biology disorders* (TBD) is the common name for a spectrum of very rare conditions caused by these mutations. The symptomatology of TBD is broad and resembles symptoms of natural aging, e.g. premature hair graying, degenerative organ-failure, and myeloid cancer susceptibility [26,56-62]. Fast-dividing cells are especially affected by telomere maintenance dysfunction and patients with TBD often present with severe hematological symptoms including bone marrow failure, aplastic anemia, and acute myeloid leukemia (AML) [56,57,63,64]. However, the most common manifestation of TBD is idiopathic pulmonary fibrosis (IPF). IPF rarely presents before the 4th decade of life and therefore seems to be an age-dependent disorder [59,65,66]. The exact connection between telomere dysfunction and IPF is still unknown, but it has been suggested that senescence of alveolar epithelial type 2 cells promotes fibrosis [59,67].

TBD is presented differently in adults and children. Children often have severe manifestations of hematological symptoms accompanied by extremely short telomere length. Dyskeratosis congenita, a severe form of TBD, is associated with telomere length below the 1st percentile of age-matched controls [68,69]. Due to the disease severity it is usually, but not exclusively, manifested during childhood. Dyskeratosis congenita patients have a high risk of developing bone marrow failure, IPF, and mucocutaneous manifestations such as abnormal skin pigmentation, nail dystrophy, and oral leukoplakia [68]. There is also a cryptic form of the disease with a gradual onset, where symptoms can be absent until adult age [70]. The most severe form of TBD is presented very early during infancy or childhood. Hoyeraal-Hreidarsson syndrome, Coats plus, and Revesz syndrome all have features of dyskeratosis congenita plus some additional disease-specific symptoms. It has been suggested that these three are the same disorder within a spectrum [66,71].

TBD patients have an increased cancer risk compared to the normal aging population but the spectrum of malignancies is more restricted. Solid tumors are rare and mostly derived from skin and mucosal epithelium cells [64]. The most common malignancy in TBD is myelodysplastic syndrome (MDS) followed by AML [64,72]. Interestingly, MDS/AML is usually absent in children (with extremely short telomere length) and restricted to adults (with relatively longer telomere length). A proposed theory is that extremely short telomere length cause stem cell exhaustion, leading to stem cell dropout and loss of clonal mutations [64]. This will restrict MDS/AML transformation but instead promote aplastic anemia. In contrast, individuals with relatively longer telomere length will acquire clonal mutations over time, which increases the risk of myeloid malignancies [64].

The first connection between genetic variants and TBD was identified in 1998 when variants in *DKC1* were linked to dyskeratosis congenita [73]. In 2023, variants in 17 genes had been identified in TBD: *DKC1* (encoding dyskerin), *WRAP53* (encoding TCAB1), *NHP2*, *NOP10, TERC,* and *TERT* in the telomerase enzyme, *ACD1* (encoding *TPP1*), *POT1,* and *TINF2* in the shelterin complex, and *CTC1*, *DCLRE1B* (encoding apollo), *NAF1, PARN*, *RPA1, RTEL1*, *STN1,* and *ZCCH8* in other telomere-associated pathways [7]. Although the number of genes linked to TBD is growing, there are still many variants, including variants in these genes, with unknown significance. Assessing the pathogenicity of novel genetic telomere-associated variants presents several challenges. TBD often shows genetic anticipation, which is an earlier onset and more severe manifestation of the disease across generations. The genetic anticipation leads to successive telomere shortening in each generation and can therefore result in earlier disease onset and symptoms that have not been previously present within a family [32,58]. Another challenge is that telomere length is not strictly correlated to symptoms, meaning that an individual with extremely short telomere length could be

asymptomatic while another individual with short telomere length could have multisystem presentation [74]. Pathogenicity assessment of genetic variants follows the consensus guidelines from the American College of Medical Genetics and Genomics (ACMG) and includes: an absence of the variant in the general population, in silico prediction algorithms indicating an effect of the variant, and co-segregation of the variant with disease in the family [75]. Genetic variants in telomere-associated genes are evaluated according to the ACMG guidelines, with additional telomere length analysis [58,62,65,68,76]. However, many variants are classified as variants of unknown significance (VUS), i.e. there is not enough information to classify the variant as either benign or pathogenic. The recommendations are that VUS should not be actionable in clinical practice.

There is no effective cure for TBD. Instead, patients with TBD should be carefully monitored and followed-up. Eventually, organ transplantation (bone marrow or lung) might be necessary. However, since TBD patients often have multisystem presentation, the transplantation will not cure symptoms present in other tissues. In addition, immunosuppressive drugs can be fatal for TBD patients due to their inherited vulnerability to cellular damage. Myeloablative conditioning regimens, i.e. high doses of chemotherapy and/or radiation before the transplantation, are especially toxic to dyskeratosis congenita patients. Therefore, reduced-intensity conditioning transplants, i.e. reduced doses of chemotherapy and radiation, could be a better regimen for these patients [77]. There is also always a risk of graft-versus-host disease during transplantation. A systematic review of dyskeratosis congenita patients who underwent hematopoietic stem cell transplantation after year 2000 showed that the estimated 5-year survival was 70% but dropped to 28% for the 10-year survival. They also concluded that patients who had received a hematopoietic stem cell transplantation from a matched sibling donor had significantly increased survival compared to patients with a non-related donor [78]. This is important since family members might have asymptomatic TBD and thus should be evaluated before becoming donors.

Telomere maintenance and telomere length in malignant cells

Telomere maintenance is one of the hallmarks of cancer [17]. The telomere shortening acts as a tumor-suppressive mechanism and malignant cells must bypass the senescence and crisis state to become immortalized. Exactly how they succeed with this is not fully understood and it most likely involves association with several of the other hallmarks of cancer. However, the majority of malignant cells (85-90%) increase telomerase expression to stabilize and elongate their telomeres. A minority (10-15%) activates the *alternative lengthening of telomeres* (ALT) pathway [52].

The ALT pathway extends the telomeres by using homologous recombination where the single-stranded 3' overhang invades the double-stranded telomeric DNA on the same or an adjacent chromosome. The invading telomere then uses the invaded telomere sequence as a template for telomere elongation [79]. The exact molecular mechanism behind ALT is not fully understood, and why some malignant cells utilize ALT instead of telomerase is unknown. So far, ALT has been described in tumors originating from mesenchymal and neuroepithelial cells but not in hematological malignancies [80,81]. The general view used to be that malignant cells activated either telomerase or ALT but it has been shown that the two mechanisms can coexist within the same tumor [82]. Hu et al. showed that the telomere maintenance mechanism could switch from telomerase to ALT in a telomerase-positive cell line by inducing DNA damage in the ATRX/DAXX complex, which is frequently mutated in ALT+ cells [83]. Several telomerase-inhibitors have been evaluated in preclinical trials [81]. Imetelstat, an oligonucleotide that inhibit telomerase by binding to TERC, is FDA approved for treatment of myelofibrosis [81,84]. Furthermore, vaccines that induces T-cell responses against TERT-presenting cancer cells are evaluated in clinical trials. However, there are currently no ALT based therapeutics available [81]. There is some concern that malignant cells treated with telomerase-inhibitors could switch to ALT or that telomerase-inhibitors could lead to a selective advantage of ALT+ cells [80].

Telomere length in malignant cells shows inter- and intratumor variability [85-88]. Many tumor cells have short telomere length compared to their paired normal tissue but telomere elongation has also been observed [86]. Short telomere length in blood has been associated with increased cancer risk but there are conflicting results [20,21]. For example, long telomere length in blood has been associated with an increased risk of developing B-cell lymphoma, including diffuse large B-cell lymphoma and follicular lymphoma [89].

Hematopoiesis, B-cell maturation, and malignant transformation

The immune system protects the body from infectious pathogens like bacteria, viruses, parasites, and fungi. The immune system has two primary subclasses: the innate and the adaptive. The innate immune system provides a nonspecific immune response when it encounters an antigen (presented by a pathogen) while the adaptive immune system provides a specific response. The essential cells during immune responses are the B-lymphocytes (B-cells), T-lymphocytes (T-cells), dendritic cells, macrophages, and natural killer cells, which all originate from hematopoietic stem cells (HSCs). The HSCs reside in the bone marrow which, together with the thymus, are part of the primary lymphoid organs

responsible for B- and T-cell maturation. The secondary lymphoid organs are the lymph nodes, spleen, and the cutaneous and mucosal lymphoid system, which are responsible for the initiation of adaptive immune responses [90].

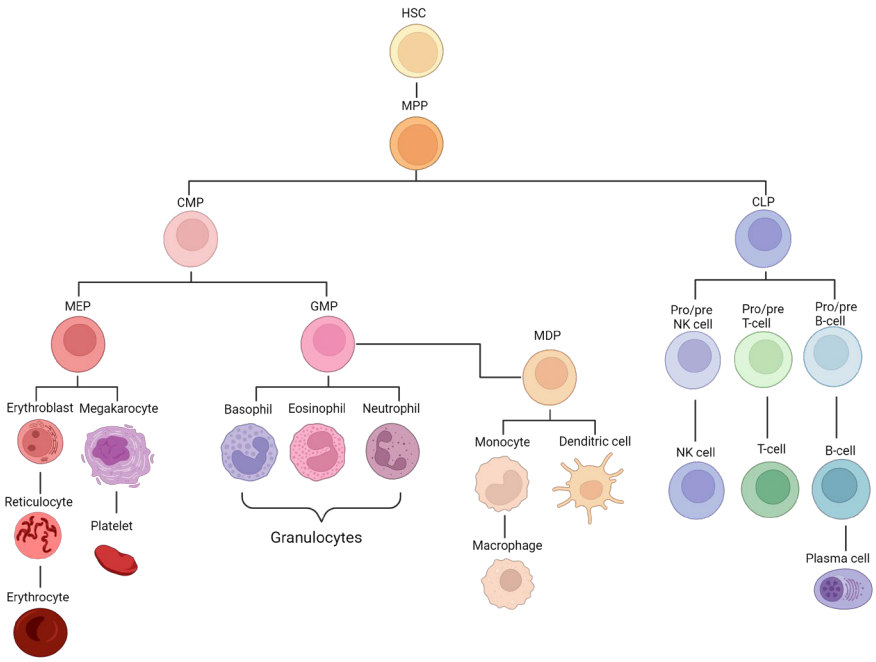


Figure 4. Hematopoietic stem cells (HSC) differentiate into multipotent progenitor (MPP) cells that can further differentiate into common myeloid progenitor (CMP) cells or common lymphoid progenitor (CLP) cells. CMPs gives rise to megakaryocyte erythrocyte progenitor (MEP) cells and granulocyte macrophage progenitor (GMP) cells. GMPs can differentiate into macrophage dendritic progenitor (MDP) cells. NK cells stand for natural killer cells. The figure was adapted from Robbins & Cotran's Pathologic Basis of Disease, chapter 13, Diseases of white blood cells, lymph nodes, spleen, and thymus, page 585 [90] and Wintrobe's Clinical Hematology, chapter 5, Origin and development of blood cells, page 239 [91]. Created with BioRender.com.

Hematopoiesis

The process of producing all possible blood cells from HSCs is called *hematopoiesis*. Hematopoiesis involves several steps where the cells become more lineage-specific for each differentiation stage (Figure 4). Briefly, HSCs differentiate into multipotent progenitor (MPP) cells which further differentiate into common myeloid progenitor (CMP) or common lymphoid progenitor (CLP) cells. The CMPs give rise to erythrocytes, platelets, granulocytes (basophils,

eosinophils, and neutrophils), monocytes, macrophages, and dendritic cells. The CLPs differentiate into natural killer cells, T-cells, and B-cells, also called lymphocytes [90,91]. The T- and B-cells are naïve before antigen interaction and become specialized after [90].

B-cell maturation

The B-cells produce antibodies that specifically bind to antigens. The antibodies are incorporated into the B-cell membrane as part of the B-cell receptor or secreted by plasma cells. Before the B-cells can produce functional and highly specific antibodies, they must go through the B-cell maturation program (Figure 5). The first stage of B-cell development is antigen-independent and takes place in the bone marrow. During this maturation step, the B-cell goes through somatic recombination of the immunoglobulin (Ig) genes for the first time. The Ig genes encode the antigen-recognizing component of the B-cell receptor which, together with a signal-transducer molecule (CD79A/CD79B heterodimer), forms the functional receptor. The Ig molecule consists of two heavy and two light chains, each with a constant and a variable region. The first initiation of antibody diversity is rearrangement of the genes encoding the variable region (the V(D)J segments) and at the end of this step, the B-cells express surface IgM/IgD and are non-specific. These cells migrate from the bone marrow to the blood stream and lymphoid organs where they circulate as naïve B-cells [91-93].

When a naïve B-cell encounters an antigen, it either directly matures into an antibody-secreting plasma cell (T-cell independent maturation) or migrates to the secondary lymphoid organs (T-cell dependent maturation). Germinal centers (GCs) are formed within the secondary lymphoid organs and the antigen-dependent stage of B-cell development is initiated. GCs are temporary micro-environments where the B-cells become specialized by further mutating and rearranging their Ig genes. The B-cells first enter the GC dark zone, where they undergo somatic hypermutation (SHM) of the variable regions to increase antigen specificity. They then migrate to the GC light zone and present their specialized Ig molecules to follicular dendritic cells and T-cells. During this process, some B-cells will undergo apoptosis and some will be recycled back to the GC dark zone to generate higher Ig affinity. The B-cells with highly specific Ig molecules go through class-switch recombination (CSR) where the constant regions of the Ig genes are mutated and rearranged. This rearrangement enables the Ig molecule to switch from surface IgM molecules to any of the four other molecules (IgA, IgD, IgE, or IgG). Both SHM and CSR are mediated by the activation-induced cytidine deaminase (AID) enzyme. The last step of the maturation program is differentiation into memory B-cells or antigen-secreting plasma cells [91,93,94].

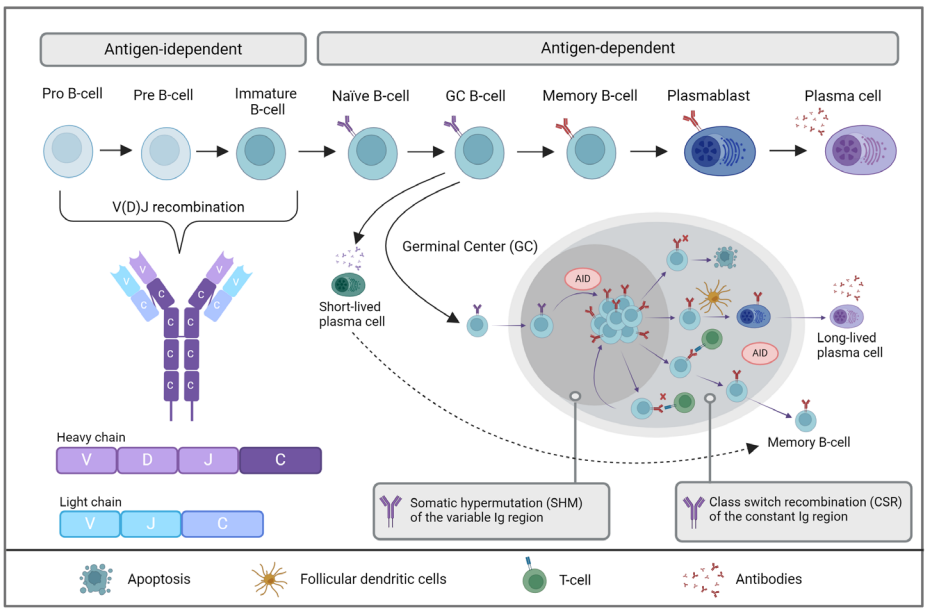


Figure 5. Schematic figure of the B-cell maturation program. The antigen-independent maturation step takes place in the bone marrow where the B-cells goes through somatic recombination of the immunoglobulin (Ig) genes. The Ig molecule consists of two heavy and two light chains, each with a constant and a variable region. The variable region consists of the V(D)J segments. When a circulating naïve B-cell encounters an antigen, it either directly matures into an antibody-secreting plasma cell (and probably to a memory-type B-cell) (T-cell independent maturation) or migrates to the secondary lymphoid organs (T-cell dependent maturation) where the antigen-dependent stage is initiated. Germinal centers (GCs) are formed within secondary lymphoid organs and consists of a dark zone and a light zone. The B-cells undergo somatic hypermutation and class-switch recombination, mediated by the activation-induced cytidine deaminase (AID) enzyme, to increase the antigen specificity. The last step of the maturation program is differentiation into memory B-cells or antigen-secreting plasma cells. The figure was adapted from The WHO classification of Tumors of Haematopoietic and Lymphoid Tissues 4th edition, chapter 11, Introduction and overview of the classification of lymphoid neoplasms, page 191 [95]. Created with BioRender.com.

Large B-cell lymphoma

Malignant tumors that originate from the lymphocytes are characterized by abnormal lymphocyte growth. Lymphoid neoplasms that primarily appear in the blood and bone marrow are usually called *leukemias*. Lymphoid neoplasms in secondary lymphoid organs and other extramedullary tissues are usually called *lymphomas* [95]. The two main categories of lymphoma are Hodgkin's and non-Hodgkin's lymphoma. Non-Hodgkin's lymphoma is the most common of these two and ranked as the 10th most commonly diagnosed cancer worldwide with an absolute number of incidence of around 550 000 in 2022 [96]. The subtypes of non-

Hodgkin's lymphoma, over 60 in number, can arise during any stage of lymphocyte development and be indolent (slow-growing) or aggressive (fast-growing) [95]. In this thesis, the focus was on four aggressive large B-cell lymphoma entities: *de novo* diffuse large B-cell lymphoma not otherwise specified (DLBCL NOS), high-grade B-cell lymphoma (HGBL) with *MYC* and *BCL2* and/or *BCL6* rearrangements, primary central nervous system (CNS) lymphoma (PCNSL), and transformation from an indolent lymphoma to DLBCL NOS (t-DLBCL) (Table 1).

DLBCL NOS (hereafter referred to as DLBCL) is the most common lymphoma, accounting for 30-40% of the non-Hodgkin's lymphomas and more than 80% of the large B-cell lymphomas [97]. It is a mature, aggressive B-cell lymphoma characterized by large B-cells with a diffuse growth pattern. DLBCL is most common in individuals over 60 years and typically arises *de novo*. In some cases, less aggressive lymphomas, such as chronic lymphocytic leukemia, follicular lymphoma, or marginal zone lymphoma, can transform into DLBCL. DLBCL is heterogeneous in terms of genetic, phenotypic, and clinical features. However, it can be classified into two major subgroups based on gene expression profiling: germinal center B-cell-like DLBCL (DLBCL-GC) or activated B-cell-like DLBCL (DLBCL-ABC) [98,99]. As the names imply, the lymphoma has either GC origin (DLBCL-GC) or GC exit/early plasmablastic or post-GC origin (DLBCL-ABC). Around 10-15% of DLBCL have an unclassifiable phenotype [95]. Besides the differences in gene expression patterns, the GC/ABC subtypes also have differences in mutation frequency [100,101]. The ABC subtype is characterized by chronic B-cell receptor signaling due to constitutive activation of the NF-κβ transcription complex (e.g. mutations in *CD79A/DC79B*, *TNFAIP3, MYD88,* and *CARD11*) [100,101]. The GC subtype is characterized by mutations in histone modifiers (see section titled Epigenetics, page 14), including the acetyltransferase genes *CREBBP/EP30* and the methyltransferase genes *KMT2D* and *EZH2*. EZH2 is part of the polycomb group protein family. Additional mutations in the GC subtype include chromosomal translocations of *BCL2* and *MYC* [100].

The standard treatment regimen for DLBCL is R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) and although DLBCL is an aggressive lymphoma, around 2/3 of the patients are cured by the primary treatment. However, the remaining 1/3 of patients that have primary refractory disease or relapse have a poor prognosis [101]. Cell of origin is an important factor for disease development in DLBCL where, at least before the introduction of rituximab, the ABC subtype has had an inferior outcome [99,102,103]. The heterogeneity of large B-cell lymphomas is likely affecting the treatment outcome and targeted agents may only benefit subgroups of patients within the same entity. Gene expression profiling of DLBCL is not a standard clinical practice today. Instead, immunohistochemistry-based algorithms are used, for

Table 1. Large B-cell lymphomas. The entities included in paper II are marked with an asterisk (*). The table was adapted from the WHO classification of Tumors of Haematopoietic and Lymphoid Tissues 4th edition, chapter 13: Mature B-cell neoplasms, Diffuse large B-cell lymphoma NOS, page 291 [95].

Diffuse large B-cell lymphoma not otherwise specified * - Activated B-cell-like - Germinal center B-cell-like
T-cell/histiocyte-rich large B-cell lymphoma
Primary CNS lymphoma *
Primary cutaneous diffuse large B-cell lymphoma, leg type
EBV-positive diffuse large B-cell lymphoma not otherwise specified
Diffuse large B-cell lymphoma associated with chronic inflammation
Lymphomatoid granulomatosis
Large B-cell lymphoma with *IRF* rearrangement
Primary mediastinal (thymic) large B-cell lymphoma
Intravascular large B-cell lymphoma
ALK-positive large B-cell lymphoma
Plasmablastic lymphoma
HHV8-positive diffuse large B-cell lymphoma (provisional entity)
Primary effusion lymphoma
High-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements *
High-grade B-cell lymphoma not otherwise specified
B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and classic Hodgkin lymphoma

Abbreviations: CNS = central nervous system, EBV = Epstein-Barr virus, ALK = anaplastic lymphoma kinase. HHV8 = human herpes virus type 8.

example Hans algorithm which classifies DLBCL into GC and nonGC (ABC + unclassifiable) subtypes [104]. In 2016, a new large B-cell lymphoma entity was defined: HGBL with rearrangement of *BCL2* and/or *BCL6,* also called double/triple hit lymphoma [95,105]. The new classification led to the implementation of fluorescence *in situ* hybridization (FISH) analysis of DLBCL as a standard routine to identify the entity. HGBLs have poor prognosis when treated with R-CHOP and are recommended intensified treatment [95,106,107]. PCNSL, also called primary diffuse large B-cell lymphoma of the CNS, is a rare, highly aggressive form of DLBCL with poor prognosis [108]. Standard R-CHOP regimens are insufficient in PCNSL since they cannot penetrate the blood-brain barrier. Therefore, a different treatment approach is required, such as high-dose methotrexate-based polychemotherapy, which effectively crosses the blood-brain barrier [108,109].

New classification schemes to further subgroup DLBCL are under investigation, most recently the LymphGen classification algorithm that subgroups DLBCL entities based on genetic alterations [110]. Epigenetic subgrouping could also be relevant due to the extent of epigenetic aberrations in DLBCL [111-113].

Epigenetics

Every cell within us contains the same genetic information and has the ability to develop into any given cell in the body. Still, they do not. So, how do stem cells know which progenitor cells they should develop into? The answer is epigenetic mechanisms, which can be described as chemical modifications of DNA, RNA, and proteins that alter the cellular phenotype without changing the DNA sequence [114]. Epigenetic mechanisms are stable during mitosis and meiosis and include histone modifications, non-coding RNA modifications, and DNA methylation (DNAm). Together, these modifications contribute to the specific gene expression pattern of specialized cells. This thesis focused on DNAm, but a general overview is also given for histone modifications since there is a dynamic interaction between these two mechanisms.

Histone modifications

Human DNA is condensed and packed in the cell nucleus by associating with a complex of eight histone proteins: two each of H2A, H2B, H3, and H4. This DNA-histone complex resembles “beads-on-a-string” when observed under a microscope. The “beads” are called nucleosomes and consist of a section of DNA wrapped around the eight histone proteins. The “string” consists of DNA that interacts with the histone protein H1 and links the nucleosomes together [115]. The complete DNA-histone complex is called *chromatin* (Figure 6). Tightly condensed chromatin, called *heterochromatin*, is associated with gene silencing

while the less condensed form, *euchromatin*, is associated with active gene transcription [116]. Posttranslational modifications on the histone tails are fundamental for chromatin structure and function. Histone acetylation is a posttranslational histone modification associated with gene activation by neutralizing the binding affinity of DNA and histones. The histone acetylation is mediated by histone acetyltransferases and histone deacetylases [114]. Another posttranslational modification is histone methylation which is associated with both gene activation and repression, depending on which residue that is methylated. The methylation is monitored by histone methyltransferases and histone demethylases [117].

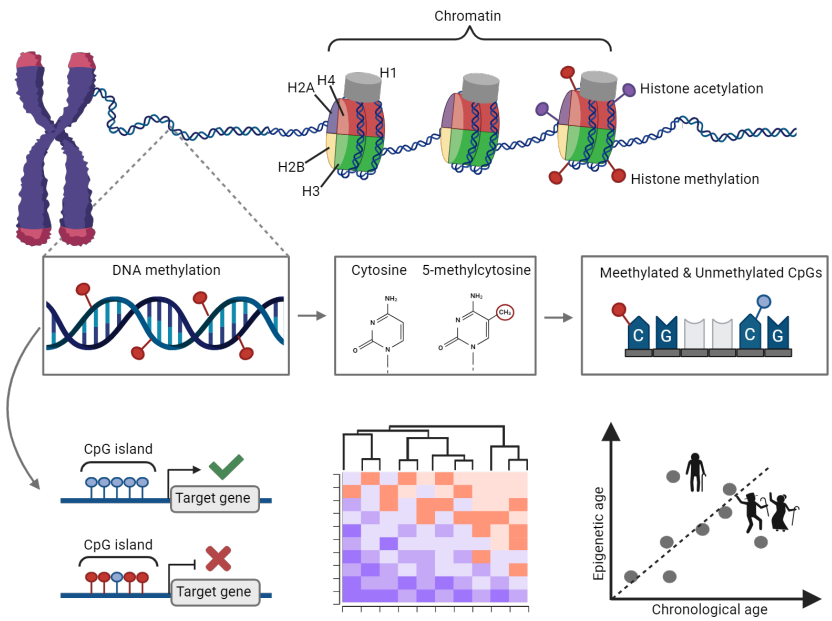


Figure 6. The nucleosomes consisting of DNA wrapped around the histone complex (two each of H2A, H2B, H3, and H4). The DNA that link the nucleosomes together interact with the histone protein H1. The complete DNA-histone complex is called chromatin. Histone acetylation and histone methylation are posttranslational modifications. DNAm predominantly occurs at cytosine-phosphate-guanine (CpG) sites and is characterized as a covalently bound methyl group to the fifth position of the cytosine. CpG islands (CGIs) are CpG-rich sections of around 1kb on average and their methylation status can alter gene expression. DNAm can be used to estimate epigenetic (biological) aging. Created with BioRender.com.

The DNA-histone chromatin structure is the first level of chromosome 3D organization. The chromatin can fold into loop-structures and further into

topologically associating domains (TADs). TADs are large, self-interacting genomic regions that are fundamental to the 3D structure of the genome. Multiple TADs form two main types of chromosome compartments: A and B. Compartment A corresponds to active chromatin (euchromatin) while compartment B corresponds to silent chromatin (heterochromatin). Finally, the chromosomes are positioned at specific locations within the nucleus, known as chromosome territories. Euchromatic chromosomal regions are often found within the interior of the nucleus while heterochromatic chromosomal regions are located near the nuclear envelope [118,119].

DNA methylation

The human genome consists of around 3 billion base pairs. Of those, around 28 million (1%) are cytosine-phosphate-guanine (CpG) sites, i.e. a cytosine followed by a guanine in the 5' direction. DNAm predominantly occurs at CpG sites and is characterized as a covalently bound methyl group to the fifth position of the cytosine (5mC) (Figure 6) [120]. DNAm regulates alternative gene splicing, retrotransposon silencing, X-chromosome inactivation, genomic imprinting, cryptic intragenic promoter silencing, and gene expression [121-124]. The DNAm pattern in cells is dynamic throughout life [123]. The initial changes occur immediately after fertilization, where the maternal and paternal genome undergoes extensive demethylation to erase most of the parental methylation profiles. The genome then becomes *de novo* methylated during implantation, i.e. when the fertilized egg attaches to the uterus, by the DNA methyltransferase (DNMT) 3A and DNMT3B enzymes. During cell replication, the DNAm pattern of the parental strand is copied to the daughter strand through maintenance DNAm, mediated by DNMT1 [125-127]. DNA demethylation has two mechanisms: *passive* (cell replication-dependent) and *active* (cell replication-independent). Passive DNAm occurs over time when DNMT1 is absent or unable to recognize a 5mC during replication. During active DNA demethylation, the 5mC is converted to 5'-hydroxymethylated cytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by the ten-eleven translocation protein (TET) family [128-131]. Lastly, 5caC is converted back to a cytosine by thymine-DNA glycosylase [132].

Assuming that the four nucleotides in the human genome (adenine, thymine, cytosine, and guanine) have an equal probability of occurring, we would expect the CpG sites to constitute 6.25% of the genome ($P(C_pG) = P(C) \times P(G) = 0.25 \times 0.25 = 0.0625$). The real frequency is around 1%, making the human genome CpG poor. The CpGs are dispersed throughout the genome, typically separated by long distances, and are often methylated [120]. The exception is CpG-rich sections of around 1kb on average, called CpG islands (CGIs), that are often unmethylated. These regions are common in promoters, specifically of housekeeping and developmental genes (Figure 6) [121]. CGIs can avoid *de novo*

methylation by an interplay with posttranslational histone modifications. Certain proteins preferentially bind to unmethylated CGIs and recruit histone methylation marks, e.g. H3K4me3, associated with gene activation [133]. CGIs can also alter their chromatin state by recruiting the lysine demethylase enzyme KDM2A, which specifically demethylates H3K36me2, resulting in gene activation [134]. Although the majority of CGIs are unmethylated, CGI methylation is induced in imprinted genes, inactivated X-chromosomes, and during cell differentiation to establish lineage-specific cells. A family of proteins that prefer binding to methylated CpGs is the methyl-CpG binding domain (MBD) family. The MBDs mediate transcriptional repression by associating with histone modification markers that promote heterochromatin formation [135,136]. Transcriptional repression can also be mediated by the polycomb group proteins that form multiple polycomb repressive complexes. These complexes repress genes that are key regulators of embryonic development and mediate H3K27 methylation, which is associated with gene silencing [121,137,138]. Several promoters marked by polycomb group proteins become methylated with age and during malignant transformation [139,140]. These examples are just a brief summary of how DNAm and histone modifications co-interact to regulate chromatin state, and thus, gene expression. Disruption of any of these epigenetic modifiers will affect the other and could lead to cellular aging and malignant transformation [4,5].

DNA methylation in normal B-cell development

The specification of HSCs begins during early embryonic development. The differentiation from HSCs to any type of hematological cell is highly mediated by lineage-specific methylation and the initial involvement of DNMT1 [141,142]. Differentiated hematological cells have cell type-specific DNAm patterns and can be distinguished based on DNA methylation profiles [143,144]. B-cells undergo excessive methylation changes during their maturation program, affecting up to 30% of their CpG sites [145,146]. Early B-cell differentiation stages are characterized by enhancer hypomethylation (low methylation levels) that affects key B-cell transcription factors. Late B-cell differentiation stages are characterized by extensive hypomethylation of heterochromatic regions and hypermethylation (high methylation levels) of polycomb-repressed regions. B-cells both retain DNAm profiles from previous differentiation stages and gain new unique maturation-specific DNAm profiles.[93,146]. The most pronounced methylation changes during B-cell maturation have been seen in GC B-cells, memory B-cells, and bone marrow plasmablasts [93,145,146]. Massive demethylation occurs during the transition from naïve B-cells to GC B-cells and it is thought that the B-cells in the GC dark-zone, called centroblasts, are the main target of epigenetic modifications [93]. AID, which is essential during Ig gene rearrangements in the GC, could be involved in active demethylation and increased methylation heterogeneity in GC

B-cells [93,147,148]. The end-stage of B-cell differentiation, i.e. memory B-cells and plasma cells, have excessive hypomethylation in heterochromatic regions and hypermethylation in polycomb-repressed regions [93,146]. Memory B-cells and plasmablasts have upregulated DNMT3A expression compared to GC B-cells, which could explain the hypermethylation of polycomb-repressed regions. The hypomethylation of heterochromatic regions could result from passive demethylation (assuming they are still proliferating) since these cells have very low expression of DNMT1, or from active demethylation [146].

DNA methylation in telomere biology disorders and large B-cell lymphoma

DNAm alterations are heavily linked to disease and can be found in for example atherosclerosis, osteoporosis, neurological, and immunological diseases [149]. However, DNAm alterations are probably mostly associated with malignant transformation [117,120,150,151]. Altered DNAm is also seen in aging cells and although aging itself is not a disease, an aging cell could be an indicator of disease development (see section titled Epigenetic clocks, page 19).

DNAm alterations in TBD have been rather unexplored and focused on dyskeratosis congenita [152,153]. Several studies, however, have investigated DNAm alterations in other premature-aging syndromes, e.g. Werner syndrome, Hutchinson-Gilford progeria, and Down's syndrome, and identified DNAm alterations to various extents [154-157] In general, up to a few thousand significant CpG sites have been identified in these studies using a linear modeling approach. A disadvantage with this method though is that very small methylation differences are captured and the biological relevance and clinical impact of these differences have not been evaluated.

DLBCL is a heterogeneous disease with large inter- and intratumor variability. Mutations in genes involved in epigenetic processes are frequent and a study of 1000 DLBCL patients showed that 18% (11/60) of the most common recurrent mutations were located in epigenetic modifiers [100,158]. Several studies have investigated DNAm alterations in DLBCL using panels targeting selected cancer-associated genes [159-163]. A few studies have examined global DNAm profiles at a large scale in DLBCL [146,164,165]. The inter- and intratumor variability in DLBCL is reflected in the methylation landscape and although one study identified entity-specific DNAm profiles in DLBCL [161], many others did not [159,162,166,167]. However, De et al. identified that intratumor DNAm variability progressively increased from normal GC B-cells to follicular lymphoma, DLBCL-GC, and DLBCL-ABC, reflecting a relationship between heterogeneity and disease aggressiveness [162]. DNAm in DLBCL has been linked to relapse and survival [162,163,167,168]. For example, Pan et al. identified that patients with relapsed DLBCL had increased intratumor

DNAm heterogeneity at diagnosis compared to non-relapsed patients [163]. Increased DNAm heterogeneity in promoter regions has been associated with worse progression-free survival (PFS) [167] and patients with global hypomethylation had inferior overall survival (OS) [168].

Richter et al. identified increased promoter hypermethylation in PCNSL compared to normal hematopoietic cells. [160]. The promoter hypermethylation was associated with CGIs and genes targeted by polycomb group proteins. However, they did not identify any entity-specific DNAm profiles between PCNSL and DLBCL. In contrast, Nakamura et al. identified enriched hypermethylation in PCNSL and a separation between DLBCL and PCNSL during cluster analysis of the most variable CGIs [169]. A study of follicular lymphomas which transformed to DLBCL found that DNAm profiles were preserved in the pre- and post-transformed samples, suggesting that DNAm alterations are an early event in lymphoma development [170]. In addition, genes targeted by polycomb group proteins were hypermethylated in follicular lymphoma compared to normal cells (mainly benign lymph nodes) [170]. To our knowledge, no published study has examined DNAm in human HGBL.

Epigenetic clocks

Biological age refers to an individual's overall health and physical condition based on biological markers. One of these markers is DNAm, and several clock models based on DNAm profiles have been developed during the last decade [2,171-174]. Since biological age is a broad concept that can be estimated by several biomarkers, the DNAm clock models will hereafter be referred to as epigenetic clocks and their output as epigenetic age. Two of the first and most well-established epigenetic clocks are Hannum's clock and Horvath's pan-tissue clock [2,174]. Hannum's clock was developed using blood cells, and it has a high correlation between DNAm and chronological aging. Horvath's pan-tissue clock estimates epigenetic aging in multiple tissues and is independent of cell divisions. Inspired by these, several additional epigenetic clocks have been developed. The epiTOC clock estimates the replicative history of stem cells, i.e. mitotic aging, based on methylation levels in promoter regions of genes targeted by polycomb group proteins [172]. The PhenoAge clock was developed using clinical parameters (e.g. immune cell characteristics) and chronological age [173]. The MiAge clock and the epiCMIT clock estimate proliferative history of normal and malignant cells [164,171]. Altered DNAm profiles and increased epigenetic age compared to chronological age have been associated with various types of cancer, all-cause mortality, and neurological disorders [3,164,175-178].

The CIMP classifiers

A CpG island methylator phenotype (CIMP) was first described in colorectal cancer by Toyota et al. [179]. Two types of promoter CGI methylation profiles were identified: one age-dependent profile present in normal colon samples and one exclusive to colorectal cancer. Further evaluation of the colorectal cancer-associated profile showed that the CGIs either had high methylation levels (CIMP+) or low methylation levels (CIMP-) [179]. Since then, CIMP profiles have been identified in other malignancies with large variations in the number of CpGs and genes included in each CIMP panel [180-182]. In our lab, a CIMP panel for pediatric T-cell acute lymphoblastic leukemia (T-ALL) was identified in 2013, which divided the T-ALL samples into CIMP+ or CIMP- groups. The CIMP groups had prognostic value, where CIMP- T-ALL patients had a worse OS (45% vs 86%, p=0.02) and event-free survival (36% vs 86%, p=0.001) compared to CIMP+ patients [183]. The prognostic value of CIMP was confirmed in a larger cohort [184]. The panel was also relevant in T-cell acute lymphoblastic lymphoma (T-LBL) and precursor B-cell acute lymphoblastic leukemia (BCP-ALL) [185,186].

Aim

The overall aim of this thesis was to investigate alterations in telomere biology and epigenetics to improve the understanding of underlying factors contributing to telomere biology disorders and large B-cell lymphoma.

Specific aims

- Investigate if global DNA methylation profiles and epigenetic aging were associated with phenotypic differences in telomere biology disorders (paper I)

- Investigate if global DNA methylation profiles and telomere length were relevant for the subclassification and prognosis of large B-cell lymphoma (paper II)

- Evaluate if telomerase activity measurements could provide additional support for pathogenicity of telomere-associated variants in telomere biology disorders (paper III)

Methods

Ethical approval (I-III)

The studies were approved by the regional ethical review board in Umeå (Dnr 2016/258-31 (paper I-III) and 2016/53-31 (paper II)) and in Uppsala (Dnr 2014/233 (paper II)) and were performed in accordance with the Declaration of Helsinki [187].

Study population (I)

TBD patients

Individuals with suspected TBD have been referred to the Department of Clinical Genetics, Umeå University Hospital, Sweden, for evaluation since 2011. By the end of 2020, a total of 54 index patients and relatives from 28 families with a known pathogenic mutation/variant in a telomere-related gene had been identified. The patients included in the study were selected from the TBD cohort based on age (16 years or older) and age-adjusted relative telomere length (RTL) in blood cells (see section titled Telomere length measurement, page 25). Briefly, we fitted a local regression model to the RTL-values of a normal cohort (n=174, 0-84 years, peripheral blood cells) and calculated the mean and standard deviations (SD) with respect to age. Then, the deviation from the age-adjusted mean RTL of the normal cohort was calculated for each TBD patient. Patients with RTL-values 0.5 to 2.5 SD below the mean (the "short (S) RTL" group, n=16) and with 3 SD or more below the mean (the "extremely-short (ES) RTL" group, n=19) were included in the study. In total, 35 patients from 18 families with mutations in *TERT* (eight families, 12 individuals), *TERC* (nine families, 22 individuals), and *DKC1* (one individual) were included. Clinical symptoms were classified as hematological if the patient presented TBD-associated symptoms and at least one of them was hematological (n=17), or as non-hematological if the patient only presented other TBD-associated symptoms (n=7), or as asymptomatic (n=11).

Controls

Twenty healthy blood donors were age-matched with the TBD patients and included as controls. The controls were collected at the Transfusion Medicine, Umeå University Hospital, Sweden.

Study population (II)

LBCL patients

The study cohort consisted of adult patients diagnosed with LBCL at Umeå University Hospital, Sweden. The biopsies were collected between year 2005 and 2018, and treatment guidelines for DLBCL remained mostly unchanged during this period. Patients with fresh frozen tumor tissue samples were selected (n=96). The LBCL patients were subtyped based on morphology, immunophenotype, and cytogenic aberrations performed by clinicians. The entity distributions were DLBCL-GC (n=38), DLBCL-nonGC (n=31), HGBL with *MYC* and *BCL2* and/or *BCL6* rearrangements (hereafter referred to as HGBL, n=7), PCNSL (n=8), and t-DLBCL (n=12).

Hans algorithm was used to subclass DLBCL into GC and nonGC based on immunohistochemical and/or flow cytometric expression of MUM-1, CD10, and BCL-6 with a cutoff of 30% positive tumor cells [104]. HGBL patients were identified through FISH analysis (rearrangements of *MYC*, *BCL2*, and/or *BCL6*), which was performed retrospectively on all LBCL patients. The samples' tumor cell content was microscopically analyzed and divided into subgroups (<20%, 20-39%, 40-59%, 60-79%, and 80-100% tumor cells). Three DLBCL samples, two GC and one nonGC, had a tumor cell content <40% and were excluded from the study. Clinical data was collected from medical records and the national lymphoma register (informationsnätverk för cancervården (INCA)) between November 2021 and March 2022. The information extracted from the medical records were: date of diagnosis, age, sex, lymphoma stage (according to Ann Arbor) [188,189], lactate dehydrogenase level, performance status (according to WHO) [190], treatment, response rate (complete remission, partial remission, or progressive disease), relapse rate, date of death (if applicable), and follow-up time. An age-adjusted international prognostic index (aaIPI) was scored based on lymphoma stage, lactate dehydrogenase level, and performance status for every patient [191].

Treatment regimens were divided into *R-CHOP-like* or *other* regimens based on selected oncological treatment and whether it was successfully concluded or not. The R-CHOP-like regimens included rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (and in some cases etoposide) which were given and completed with curative intent. The other regimens included CNS treatment, relapse treatment, radiotherapy, or only part of the planned treatment.

Controls and publicly available DNA methylation data sets

- Normal germinal center B-cells

A tonsillar sample was collected at the Ear-Nose-Throat clinic at Umeå University Hospital, Sweden. The cell suspension protocol has been described previously by Dernstedt et al. [192]. The cells were stained with Zombie aqua (1:1000) (BioLegend, San Diego, CA, US), anti-human IgD-BV421 (clone IA6-2, BD Biosciences, Franklin Lakes, NJ, US), CD20-FITC (clone 2H7, BioLegend), CD38-PE (clone HIT2, BioLegend), and CD19-PE-CF594 (clone HIB19, BD Biosciences) for 30 minutes at room temperature. The cells were sorted on a BD FACSMelody instrument (BD Biosciences, Franklin Lakes, NJ, US) and viable GC B-cells were defined as CD19+CD20+CD38+IgD-, with the sorting rate <1500 events/second. In total, 864,000 cells were sorted into RPMI-1640 medium (Gibco, Thermo Fisher Scientific, Waltham, MA, US), supplemented with 1% penicillin-streptomycin, 20mM HEPES, and 50% FBS.

- Normal B-cells, normal plasmablasts, and B-cell neoplasms

Publicly available DNAm data sets were downloaded from the Gene Expression Omnibus repository. Sorted normal B-cells from peripheral blood (n=28, GSE103541) were included as controls in the study. Additionally, BCP-ALL (n=663, GSE49031), chronic lymphocytic leukemia (n=75, GSE237299), mantle cell lymphoma (n=72, GSE237299), multiple myeloma (primary plasma cell leukemia, n=14, GSE104770), DLBCL (n=18, GSE237299 and n=31, GSE37362), and normal plasmablast samples (n=8, GSE72498) were included for methylation heterogeneity comparisons.

Study population (III)

Peripheral blood was collected from healthy blood donors (n=100, 18-73 years old, 51% men) in sodium heparin tubes at the Transfusion Medicine, Umeå University Hospital, Sweden. Peripheral blood was also collected in EDTA tubes from 90 of the donors for telomere length analysis. The control cohort had a balanced distribution of age and gender to capture age-related variation in TA.

Patients with suspected TBD (n=6) were referred to the Department of Clinical Genetics, Umeå University Hospital, Sweden, for genetic testing and/or telomere length analysis between year 2020 and 2023. The patients had a genetic variant in either *TERT* (n=5) or *TERC* (n=1) that was classified as pathogenic (*TERT* p.Arg865His (n=1)) or likely-pathogenic (*TERT* p.Leu1017_Leu1019del (n=1), p.Asp684Gly (n=1), and p.Tyr1002Cys (n=2), and *TERC* n.64G>A (n=1)) according to the ACMG criteria during the diagnostic evaluation in Umeå [75] .

Genomic DNA extraction

TBD patients and controls (I & III)

Genomic DNA from the controls (n=20 in paper I and n=100 in paper III) were extracted from whole blood cells using the Gentra Puregene blood kit (Qiagen, Hilden, DE) according to the manufacturer's instructions. Genomic DNA from the TBD patients in paper I was extracted at Umeå University Hospital (n=19) or by clinical genetic laboratories in Helsinki, Finland (n=1), Lund, Sweden (n=1), Oulu, Finland (n=4), Stockholm, Sweden (n=2), Uppsala, Sweden (n=7), and Århus, Denmark (n=1), according to the standard protocols at the respective center. Genomic DNA from the TBD patients in paper III was extracted at Umeå University Hospital (n=6). The DNA concentration and quality of all patients and controls were measured by UV- spectrometry using the DeNovix DS-11 instrument (DeNovix, Wilmington, DE, US) at Umeå University.

LBCL patients (II)

Genomic DNA from the LBCL patients was extracted using the AllPrep DNA/RNA mini kit (Qiagen, Hilden, DE) according to the manufacturer's instruction with some minor modifications: the addition of a four minutes incubation time to the washing steps, an additional washing step with buffer AW2, and a five minutes incubation time to the DNA elution step. DNA concentration and quality were measured by UV- spectrometry using the DeNovix DS-11 instrument. Additionally, DNA fragmentation was measured using the Agilent 2200 TapeStation System and the Agilent Genomic DNA ScreenTape assay (Agilent Technologies, Santa Clara, CA, US) according to the manufacturer's instruction. The output is a DNA integrity number (DIN) ranging from 0-10, where a high DIN number corresponds to highly intact DNA and a low DIN number to fragmented DNA. DIN could not be obtained for one LBCL sample due to low sample volume. The median DIN number of the LBCL cohort was 7.8 (range: 5.5-9.5), which was considered acceptable.

Telomere length measurements (I-III)

RTL was measured based on Cawthon's qPCR protocol from 2002 [193,194]. Each DNA sample was measured in triplicates in separate telomere and single-copy gene (the hemoglobin subunit beta gene) reactions in either a 96-well format on a BioRad CFX96 real-time PCR instrument (Bio-Rad, Hercules, CA) (paper I and III) or a 386-well format on a QuantStudio™ 6 Flex System instrument (ThermoFisher Scientific, Waltham, MA, US) (paper II). The PCR master mixes used were the QuantiFast Multiplex PCR kit (Qiagen, Hilden, DE) (paper I and paper III (n=30)), QuantiNova SYBR Green PCR kit (Qiagen, Hilden, DE) (paper

III (n=60)), and SYBER Green Master Mix for qPCR (ThermoFisher Scientific, Waltham, MA, US) (paper II). A telomere/single-copy gene (T/S) ratio was calculated for every sample by the formula: $2^{-(Ct_S - Ct_T)}$, where Ct_S stands for the cycle threshold of the single-copy gene and Ct_T for the cycle threshold of the telomere gene. CCRF-CEM cell line DNA was included as a positive control and standard curve in every run to monitor the PCR efficiency. The RTL values were obtained by dividing the T/S value of a given control or sample with the T/S value from the positive control. All DNA samples were analyzed in two separate runs, generating a mean RTL from both runs.

The RTL residual (age-adjusted RTL) of each patient was calculated to compensate for the natural loss of telomere length with age [32] (paper I and II). RTL-values from a normal cohort (n=174, 0–84 years, peripheral blood cells) from the Department of Clinical Genetics, Umeå University Hospital, Sweden, was included as a reference. The residuals were obtained by first fitting a local regression model (LOESS) to the RTL values of the normal cohort. Then, a LOESS was fitted to the squared residuals modeled as a function of age (squared residuals ~ age) to obtain the estimated local standard deviation of the normal cohort. The standardized residuals of RTL were generated by dividing the residuals of the patients with the local standard deviations of the normal cohort. A negative residual indicated shorter RTL than expected.

DNA methylation analysis (I and II)

The EZ DNA Methylation Kit (Zymo Research, Irvine, CA, US) was used to bisulfite convert DNA (750-1000 ng) according to the manufacturer's instructions. Methylight *ALU* PCR with the ALU-C4 primer/probe set confirmed efficient bisulfite conversion [195]. Global DNA methylation profiles was assessed using the MethylationEPIC BeadChip v.1 (Illumina, San Diego, CA, US) with a coverage of >850,000 CpG sites. The methylation level for each CpG on the EPIC array are stated as β-values, where 0 is completely unmethylated and 1 is fully methylated, by the equation:

$$\beta = \frac{M}{M + U + \varepsilon}$$

where M and U are the proportion of methylated/unmethylated intensity signal at each specific CpG (Figure 7). The standard error (ε) is usually set to 100 and is added to avoid zero in the denominator [196]. The Genome Studio software V2011.1 (Illumina) and R statistical software v 4.0.4 to 4.2.2 (R core team) were used for pre-processing and R for downstream analysis. The methylation data were imported into R using the Minfi package [197] and normalized with the background

normalization tool *Noob* [198] (paper II). Adjustments for the two different probe designs used by the EPIC array were performed with *BMIQ* (paper I and II) [196]. Omitted sites during data pre-processing were CpGs with a detection p-value >0.05, non-specific probes aligning to multiple loci [199], CpGs located ≤5 bp from a known SNP in the European population [199], methylation quantitative trait loci (mQTLs, which are genetic variants that could affect DNAm) [200,201], and duplicated CpGs mapping to the same gene. Additionally, CpGs located on the X and Y chromosomes were omitted to avoid gender-related biases (paper I and II).

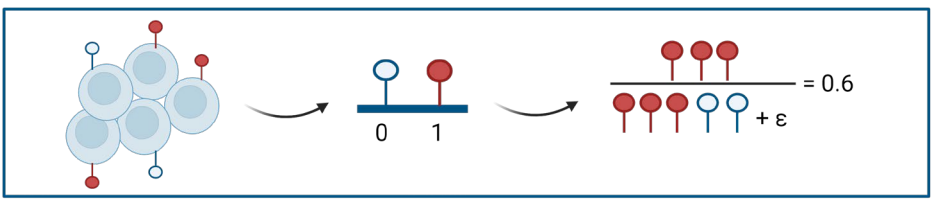


Figure 7. Example of β-value calculation. A CpG site in a single cell can either be methylated (score 1) or unmethylated (score 0). However, the samples used for DNAm analysis with the MethylationEPIC BeadChip v.1. consists of several cells. A score is given for each CpG site in each cell. The β-value is obtained from the proportion of methylated/total signal at each specific CpG. The standard error (ε) is usually set to 100 and is added to avoid zero in the denominator. Created with BioRender.com.

Adjusting for leukocyte cell composition (I)

The methylation pattern of blood cells are influenced by cell composition [143] and hence the β-values in paper I were adjusted for leukocyte cell composition prior to analysis. The leukocyte cell composition of TBD patients and controls were estimated by the FlowSorted.Blood.EPIC package in R, which uses DNAm profiles of sorted neutrophils, B-cells, monocytes, natural killer cells, CD4+ T-cells, and CD8+ T-cells as a reference library [202,203]. Methylation data from sorted granulocytes, B-cells, monocytes, CD4+ T-cells, and CD8+ T-cells were downloaded from the Gene Expression Omnibus (GSE103541) and included as a reference. The methylation data from the downloaded sorted blood cells was normalized and pre-processed in the same way as the TBD and control samples.

Adjusted β-values were calculated for each TBD patient and control by the equation:

$$\beta_j = \sum_i^n Estimated\ cell\ proportion\ _i\ of\ sample \times mean\ \beta\ of\ sorted\ blood\ cells_{ij}$$

where i is the cell type (granulocytes, B-cells, monocytes, CD4+ T-cells, or CD8+ T-cells) and j is the selected CpG site.

Adjusting epigenetic age for confounders (I)

The epigenetic ages were adjusted for chronological age and granulocyte proportion (specifically neutrophil proportion, estimated by the FlowSorted.Blood.EPIC package). A linear regression model was fitted to the control data and the adjusted epigenetic ages were obtained by the equation:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where Y is the adjusted epigenetic age, β_0 is the intercept, β_1 the slope coefficient of chronological age, β_2 the slope coefficient of the estimated neutrophil proportion, and ε the model's prediction error. The intercept and slope coefficients were used to calculate the adjusted epigenetic ages of the TBD patients. The residual (difference between the predicted and the adjusted age) were the stated delta epigenetic age.

CpG island methylator phenotype (II)

The CIMP panel was previously defined based on pediatric T-ALL samples analyzed with the Infinium HumanMethylation27 BeadChip (Illumina, San Diego, CA, US) [183]. The original panel included 1347 CpG sites with a β-value standard deviation of ≥0.3 across T-ALL samples. Of those, 1099 were present on the MethylationEPIC BeadChip v.1. The LBCL patients were classified as CIMP- if ≤25% of the CpGs in the panel had a β>0.4, otherwise they were classified as CIMP+ [185].

T-cell activation, proliferative capacity estimation, and protein extraction (III)

T-cell activation with phytohemagglutinin

Peripheral blood mononuclear cells from the sodium heparin tubes were isolated within 6 hours (controls) or within 32 hours (TBD patients) of blood withdrawal, using the standard Lymphoprep™ protocol (Stem Cell Technologies, Cambridge, UK). The cell concentration was measured on a HemoCue® WBC DIFF instrument (HemoCue, Ängelholm, SE) and 250×10^3 cells/ml were suspended in RPMI-1640 medium with 20% fetal calf serum, 200 mM L-glutamine, 5000 IU penicillin, 5000 µg/ml streptomycin, and 2.7% phytohemagglutinin. The cell suspensions were cultured at 37°C and 5% CO_2 for three days (72-77 hours) since the maximum telomerase activity of T-cells peaks around three days of stimulation with phytohemagglutinin [51,204]. The cells were then washed in PBS, and one fourth of the total cells were used for proliferative capacity estimation (Figure 8).

Proliferative capacity estimation

The proliferative capacity of the activated T-cells was evaluated by flow cytometry to obtain the S-phase fraction. The cell lysates were diluted to 10^6 cells/ml and incubated in trypsin solution with Nonidet P40 and spermine tetrachloride to isolate the nuclei. The nuclei were then stained using a propidium iodide solution for at least 30 minutes at 4°C in the dark [205,206]. The samples were analyzed by the FACSVia™ instrument (BD Biosciences, Franklin Lakes, NJ, US). The ModFit LT™ software v3.0 (Verity Software House, San Diego, CA, USA) was used for the S-phase fraction analysis and the creation of DNA histograms. The cut-off for acceptable runs was set to more than 10,000 modeled events in the G0/G1, G2/M, and S-phase components of all cycling populations.

Protein extraction

The cells that was not used for proliferative capacity estimation were lysed in CHAPS lysis buffer (Sigma-Aldrich, Saint Louis, MO, US) (Figure 8). An aliquot of each lysate was used for measuring protein concentration with the Pierce™ BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA, US) using the standard protocol, but with a reduction of the total volume by one fourth to 56,25 µl instead of 225 µl. The optical density was measured on a DeNovix DS-11 instrument at 562 nm. A standard curve from bovine serum albumin with a working range of 20-2000 µg/ml was included in every run. All samples were measured in triplicates and the mean value was stated as the total protein concentration. The remaining lysates was treated with dithiothreitol (DTT)

(Invitrogen, Waltham, MA, US) and Recombinant RNasin® Ribonuclease Inhibitor (Promega, Madison, WI, US) with a working volume of 1 mM DTT and 1 U/µl RNasin.

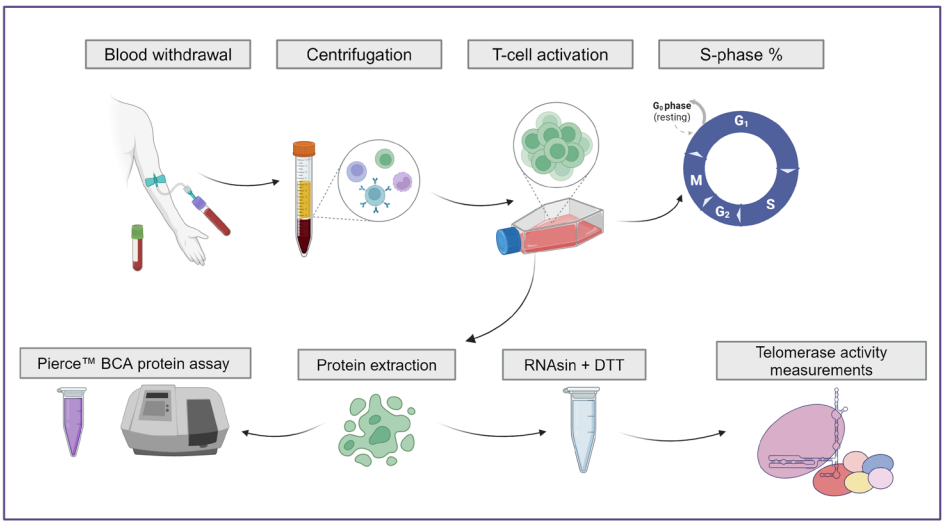


Figure 8. Flowchart over the laboratory steps in paper III. Peripheral blood was collected in sodium heparin tubes (purple cap) for the telomerase activity assay and EDTA tubes (green cap) for telomere length measurements. Peripheral blood mononuclear cells from the sodium heparin tubes were isolated and 250×10^3 cells/ml were cultured in cell medium with phytohemagglutinin at 37°C and 5% CO_2 for three days to activate the T-cells. The cells were then washed in PBS and an aliquot of the cells was evaluated by flow cytometry to obtain the S-phase fraction. The rest of the cells were lysed in CHAPs lysis buffer and protein concentration was determined in an aliquot of the sample. The remaining cells were treated with RNAsin + DTT and included in the telomerase activity measurements. Created with BioRender.com.

Telomerase activity measurements (III)

Telomerase activity was measured with the telomeric repeat amplification protocol (TRAP) [207] using the Telomerase Activity Quantification (TAQ) qPCR assay kit (ScienCell Research Laboratories, Carlsbad, CA, US). TRAP includes three steps: telomerase extension where the telomerase enzyme adds nucleotides to an artificial telomere sequence, PCR amplification, and product quantification. The TAQ qPCR protocol was optimized for adding a fixed protein concentration instead of a fixed number of cells to the telomerase extension step using a T-lymphoblastoid cell line (CCRF-CEM).

The telomerase extension step was carried out in separate test tubes for all samples. The input concentration was 100 ng/µl (2000 ng) for CCRF-CEM (standard) and 25 ng/µl (500 ng) for the patient samples, control samples, positive control (CCRF-CEM), negative control (heat-inactivated CCRF-CEM), and no template control (CHAPS lysis buffer). The telomerase extension reaction was run on a Veriti™ 96-Well Fast Thermal Cycler (Applied Biosystems, Waltham, MA, US) at 37°C for 3 hours and then heated to 85°C for 10 minutes to stop the reaction. The CCRF-CEM (standard) was diluted 1:5 to create a standard curve to monitor the qPCR.

The PCR reactions were run in a 96-well format in duplicates on a QuantStudio™ 6 Flex System instrument (ThermoFisher Scientific, Waltham, MA, US). The telomerase activity was quantified with the QuantStudio™ Real-Time PCR software v1.3 (ThermoFisher). An automatic cycle threshold (Ct) and baseline were set by the software. The samples were measured on one occation (contros, n=90) or on two occations (controls, n=10 and TBD, n=6). The telomerase activity (TA) was obtained by the formula: $TA = 2^{-(Ct_X - Ct_P)}$, where Ct_X corresponds to the mean Ct-value of a control or a TBD sample, and Ct_P corresponds to the mean Ct-value of the positive control.

Statistical analysis (I-III)

The statistical analyses were performed in R v.4.0.4 to v.4.2.2. A statistical test was considered significant if the p-value was ≤0.05. Categorical variables were compared using the Mann-Whitney U test (two groups) or the Kruskal-Wallis test (more than two groups). The Kruskal-Wallis test was followed by an ad hoc test with Bonferroni correction to account for multiple testing. Continuous variables were analyzed by linear modeling. Fischer's exact test, followed by an ad hoc test with Bonferroni correction, was used to compare the chromosomal distribution of differentially methylated (DM)-CpGs (paper I) and gene annotation distribution (paper II). Heatmaps were created using the ComplexHeatmap package [208] and principal component analysis (PCA) using the FactoMineR package [209]. A publicly available B-cell neoplasm classifier, based on DNAm profiles, was used to classify DLBCL into subtypes [164].

The gene ontology analysis of DM-CpGs in paper I and II was performed with the GeneGO MetaCore™ software (Tomson Reuters, New York, NY, US), using the enrichment analysis workflow tool. The results were presented as gene ontology processes (paper I) or process networks (paper II). The epigenetic regulation of gene expression pathways (paper I) was identified through the Reactome pathway knowledgebase (Epigenetic regulation of gene expression (Homo sapiens), R-HSA-212165) [210].

The epigenetic ages in paper I were estimated using three different clock models: Horvath's pan-tissue [2], PhenoAge [173], and epiTOC [172]. The epigenetic ages in paper II was estimated using nine different clock models: Hannum [174], Horvath's pan-tissue, PhenoAge, epiTOC, MiAge [171], epiCMIT, epiCMIT-hypo, epiCMIT-hyper [164], and DNAmTL [211].

The survival analysis in paper II included the DLBCL (n=56) and LBCL (n=68) patients that were treated with R-CHOP-like regimens. Disease-specific survival was estimated from the day of diagnosis until death caused by lymphoma or until the last follow-up date (March 2022). Progression-free survival was estimated from the day of diagnosis until progression, relapse, or last follow-up date. Kaplan Meier curves and the Cox proportional hazards model were used for uni- and multivariable analysis and the median follow-up time was analyzed by reverse Kaplan-Meier plots. Survival between groups were compared using the log-rank test or Wald's test. Factors included in the survival analysis were: age, aaIPI, entity, age-adjusted RTL, percentage of hypo-, semi, and hypermethylated CpGs, mean β-value, global methylation variability score (MVS) [167], promoter MVS, epiTOC age, and CIMP classes. The factors were divided based on median value (age, global MVS, promoter MVS, and epiTOC) or quartiles (mean β-value and percentage of hypo-, semi, and hypermethylated CpGs) based on the entire LBCL cohort (n=93). The age-adjusted RTL was considered short if it was more than 3 SD below the mean values of the control cohort and long if it was more than 3 SD above the mean values of the control cohort. aaIPI was stated as 0-1 and 2-3.

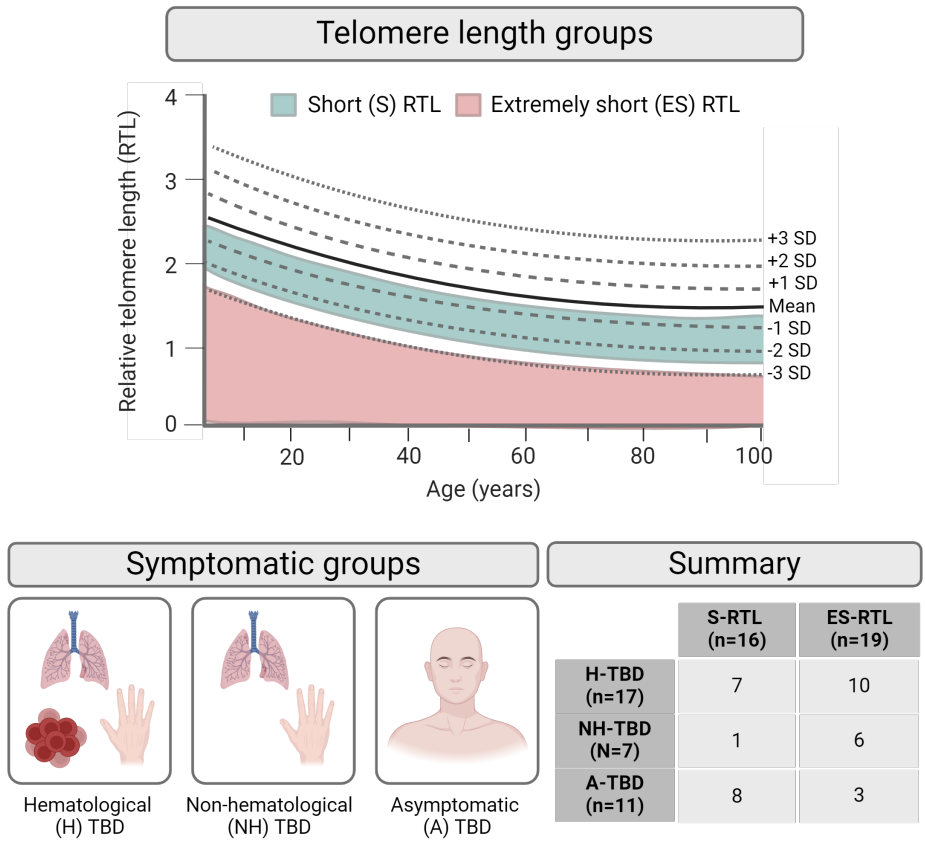
Results & Discussion

DNA methylation variations and epigenetic aging in telomere biology disorders (I)

TBD patients have a heterogeneous clinical presentation that cannot fully be explained by mutation status, indicating that telomere attrition is not the only mechanism that contributes to the disease phenotype [74]. Altered DNAm profiles have been identified in various diseases, including premature-aging syndromes [152-155]. However, studies of DNAm in TBD have been limited to dyskeratosis congenita [152,153,212]. In this study, we included TBD patients with various clinical presentations to examine if DNAm contributed to the observed phenotypic differences. Global DNAm profiles were analyzed in peripheral blood cells from 35 TBD patients with a pathogenic mutation in *TERC* (n=22), *TERT* (n=12), or *DKC1* (n=1), and 20 age-matched controls. The DNAm analysis was performed in relation to telomere length (two groups) and in relation to symptoms (three groups) (Figure 9). The telomere length groups were selected based on age-adjusted relative telomere length (RTL): short (S)-RTL and extremely short (ES)-RTL. The symptomatic groups were selected based on clinical presentation: asymptomatic (A)-TBD, hematological (H)-TBD, and non-hematological (NH)-TBD. Patients with H-TBD could present several symptoms, but at least one of them had to be hematological. Patients with NH-TBD did not have any reported hematological symptoms.

Differentially methylated CpGs in telomere biology disorders (I)

CpG sites that differed by at least 20% ($|\Delta\beta| \geq 0.2$) between the TBD groups and controls were considered differentially methylated (DM)-CpGs. We identified DM-CpGs in all TBD subgroups to various extents. In the telomere length groups, 73 DM-CpGs were identified in ES-RTL and five DM-CpGs in S-RTL compared to controls (Table 2). The genes associated with DM-CpGs were not typically involved in TBD or other telomere-related disorders [213]. Further, these genes were not associated with epigenetic regulation of gene expression [210]. Unsupervised cluster analysis of the 73 DM-CpGs in the ES-RTL group separated the methylation profiles into three clusters: one with only symptomatic patients (cluster A), one with a mixture of symptomatic and asymptomatic patients (cluster B), and one with controls and half of the asymptomatic patients (cluster C) (Figure 10). The cluster analysis indicated that symptomatic patients, regardless of telomere length, had similar DNAm profiles. When we compared only the symptomatic S-RTL patients (n=8) to controls we identified 42 DM-



	S-RTL (n=16)	ES-RTL (n=19)
H-TBD (n=17)	7	10
NH-TBD (N=7)	1	6
A-TBD (n=11)	8	3

Figure 9. The TBD patient groups included in the study. The telomere length groups were selected based on age-adjusted RTL. RTL-values between 0.5 to 2.5 SD below the mean RTL of controls were considered short (S). RTL-values of 3 SD or more below the mean RTL of controls were considered extremely short (ES). Symptoms were classified as hematological (H) if the patient presented TBD-associated symptoms where at least one of them was hematological or as non-hematological (NH) if the patient only presented other TBD-associated symptoms, or as asymptomatic (A). The table to the right shows the distribution of symptoms in the S-RTL and ES-RTL groups. Created with BioRender.com.

CpGs (55% overlap with the DM-CpGs in ES-RTL) (Table 2). We proposed two explanations for the DNAm alterations observed in symptomatic TBD patients. Firstly, bone marrow failure and cytopenia in TBD are caused by the loss of HSC and progenitor cells [7,213]. Thus, the methylation alterations could be involved in these disease processes. Secondly, the methylation alterations could be markers for subpopulations of cells with greater proliferation and survival potential within a depleted stem cell pool.

Table 2. DM-CpGs in the TBD patient groups compared to controls. Hypermethylated and hypomethylated CpGs in TBD are indicated with $^{+}$ and $^{-}$, respectively.

Group	Total number of DM-CpGs	Genes with two or more DM-CpGs
S-RTL	5	
ES-RTL	73	*PRDM8*$^{+}$, *VARS*$^{+}$, *SMC4*$^{-}$, *WNT6*$^{+}$, *MAS1L*$^{+}$
Symptomatic S-RTL	42	*SMC4*$^{-}$, *WN6*$^{+}$, *NAV2*$^{-}$
H-TBD	96	*PRDM8*$^{+}$, *VARS*$^{+}$, *SMC4*$^{-}$, *WNT6*$^{+}$, *NAV2*$^{-}$, *TM4FS1*$^{-}$
NH-TBD	58	*SMC4*$^{-}$, *WNT6*$^{+}$
A-TBD	4	

We further examined the symptomatic groups and identified 96 DM-CpGs in H-TBD, 58 in NH-TBD, and four in A-TBD (Table 2). As expected, there was an overlap between the DM-CpGs identified in the H-TBD and NH-TBD groups (n=34), confirming that symptomatic patients had similar, although not identical DNAm profiles. All but one of the NH-TBD patients had ES-RTL, which could support that the methylation alterations were markers of hematological cells with better survival potential. However, it could also have been that the altered methylation profiles were early signs of transition toward hematological symptom development in NH-TBD. We did not have enough samples to evaluate if NH-TBDs with S-RTL (n=1) or A-TBDs with ES-RTL (n=3) had altered DNAm profiles. This could be of interest in future studies, especially since the asymptomatic patients with ES-RTL did not cluster with controls.

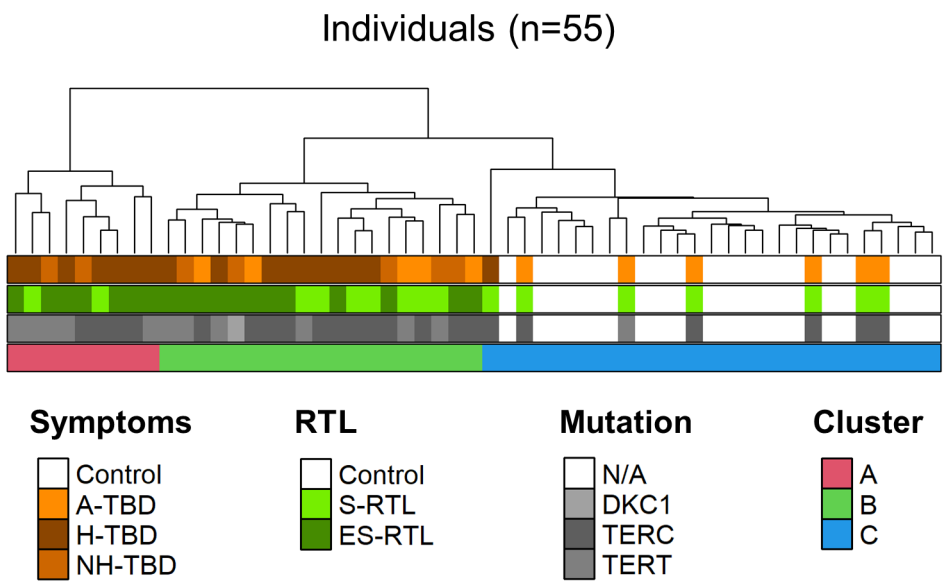


Figure 10. Unsupervised cluster analysis of the 73 DM-CpGs identified in the ES-RTL group. Three clusters were identified. Cluster A with symptomatic TBD, cluster B with a mixture of symptomatic and asymptomatic patients, and cluster C with controls and asymptomatic patients. The figure has been adapted from Carlund O. *et al.* DNA methylation variations and epigenetic aging in telomere biology disorders. Sci Rep. 2023;13(1):7955.

Genes associated with differentially methylated CpGs in telomere biology disorders (I)

Genes with two or more DM-CpGs were filtered out for the telomere length and symptomatic groups, respectively, resulting in seven interesting candidate genes (Table 2). *MAS1L*, *NAV2*, and *TM4FS1* had no previous known connection to TBD or telomere length while *PRDM8*, *SMC4*, *VARS*, and *WNT6* did.

The PRDM8 protein belongs to the highly conserved PRDM-family that acts as direct or indirect histone methyltransferases. The PRDM-family also acts in stem cells and is involved in cell differentiation, maturation, and developmental processes [214]. DNAm alterations in *PRDM8* have previously been linked to dyskeratosis congenita, Werner's syndrome, Hutchinson-Gilford progeria, and Down's syndrome [212]. In our cohort, methylation alterations in *PRDM8* were identified in the ES-RTL and H-TBD groups. However, it is unknown whether *PRDM8* is involved in the disease process of TBD or if it is a biomarker of cellular aging.

The *VARS* gene encodes Valyl-tRNA synthetase, one of the aminoacyl tRNA synthases essential for correct protein synthesis [215]. Methylation variabilities in *VARS* have previously been correlated to leukocyte telomere length in normal blood cells [216] but not with premature-aging syndromes. Methylation alterations in *VARS* were identified in the ES-RTL and H-TBD groups in our cohort, but not in the S-RTL patients with hematological symptoms. Hence, DNAm alterations in *VARS* could be a marker of extremely short telomere length and not associated with the disease process in TBD.

The *SMC4* gene encodes a protein involved in chromosome condensation and plays an essential role in embryonic cell division [217]. DNAm alterations in *SMC4* have previously been identified in Werner's syndrome [156]. It has also been associated with leukocyte telomere length and telomere dysfunction [218,219]. The *WNT6* gene is part of the WNT-family which is involved in several developmental processes [220]. WNT activity is important for stem cell homeostasis and WNT pathway mutations are frequent in cancer and degenerative diseases [221,222]. It has been suggested that a positive feedback loop exists between WNT signaling and telomere capping [220]. Extremely short telomeres trigger DNA damage responses, leading to WNT signaling pathway downregulation. The downregulation reduces the expression of the shelterin components. Pharmacological upregulation of the WNT signaling pathway in mice increased the expression of the shelterin proteins and recapped the telomeres [220,223]. We observed altered DNAm in *SMC4* and *WNT6* in the H-TBD, NH-TBD, ES-RTL, and symptomatic S-RTL groups. Hence, methylation variability in these genes could be early markers of TBD progression in hematological cells, not (yet) observed in asymptomatic patients. Future studies should evaluate the functional relevance of *PRDM8*, *VARS*, *SMC4*, *WNT6, MAS1L*, *NAV2*, and *TM4FS1* in terms of gene expression, telomere maintenance, and disease progression in TBD.

Increased DNAm in the *TERT* hypermethylated oncological region (THOR, located upstream of the *TERT* promoter) has been associated with increased TERT expression in malignant cells [224]. Furthermore, hypermethylation in *TERC* has been associated with decreased TERC expression in breast cancer [225]. We speculated that DNAm levels in *TERT* and *TERC* might be altered in TBD patients. However, DNAm in *TERT* did not differ between patients with a *TERT* mutation compared to either controls or patients without a *TERT* mutation. The same result was obtained when we investigated DNAm in *TERC*. These results were in concordance with a previous study that analyzed *TERT* and *TERC* methylation in dyskeratosis congenita using the Illumina Infinium HumanMethylation450 BeadChip [153]. A limitation with the arrays is that they do not provide full coverage of genomic CpG sites. To reach full coverage, whole genome bisulfite or enzymatic sequencing is required, which will allow for differentially methylated region analysis in future studies [226].

Epigenetic age in telomere biology disorders (I)

Increased epigenetic age has been associated with several diseases, including premature-aging syndromes [154,177,227,228]. We estimated the epigenetic age in TBD using three different clock models: Horvath's pan-tissue clock, the epiTOC clock, and the PhenoAge clock [2,172,173]. These clock models target slightly different DNAm alterations. Horvath's pan-tissue clock has multi-tissue prediction potential and is highly correlated to chronological age. The epiTOC and PhenoAge clock were developed for blood cells and captured the cells' mitotic and physical/physiological conditions rather than chronological age-related DNAm changes.

Table 3. Delta epigenetic age in controls and TBDs presented as mean value and standard deviation.

	ΔPhenoAge	**ΔepiTOC**	**ΔHorvath's pan-tissue**
Control (n=20)	0±2.4	0±0.006	0±3.4
S-RTL (n=16)	6.3±9.0	0.011±0.019	-3.0±5.6
ES-RTL (n=19)	10.2±9.9	0.018±0.026	-1.5±10.0
H-TBD (n=17)	13.7±8.5	0.020±0.025	-1.5±9.9
NH-TBD (n=7)	6.2±9.3	0.020±0.031	-0.7±8.5
A-TBD (n=11)	1.8±6.8	0.003±0.005	-4.2±4.8

For the telomere length groups, increased epigenetic age was identified in ES-RTL (ΔPhenoAge=10.2±9.9, $p<0.001$ and ΔepiTOC=0.018±0.026, $p=0.011$) and S-RTL (ΔPhenoAge=6.3±9.0, $p=0.044$) compared to controls (ΔPhenoAge=0±2.4 and ΔepiTOC=0±0.006) (Table 3). The S-RTL group showed a tendency towards increased ΔepiTOC age compared to controls, although not significant (0.011±0.019, $p=0.055$). For the symptomatic groups, increased epigenetic age was identified in the H-TBD group (ΔPhenoAge=13.7±8.5, $p<0.001$ and ΔepiTOC=0.020±0.025, $p=0.005$) compared to controls. Additionally, increased ΔPhenoAge was identified in H-TBD compared to A-TBD (1.8±6.8, $p=0.004$). A short or extremely short telomere length may induce stem cell exhaustion [55], compelling the remaining stem cell pool to undergo more frequent divisions to maintain the homeostasis of hematological cells. The increased ΔPhenoAge and ΔepiTOC ages in S-RTL, ES-RTL, and H-TBD could

reflect methylation alterations resulting from exhausted stem cells. Horvath's pan-tissue clock does not capture cell replication or cellular senescence, which could explain why there was no significant difference between the TBD patients and controls. This is consistent with previous studies that did not identify an increased epigenetic age in premature-aging syndromes with Horvath's pan-tissue clock [2,153]. Importantly, the CpGs constituting the clock models did not overlap with the DM-CpGs we identified using the 20% difference cutoff, indicating that the methylation profiles of several biological pathways were altered in TBD.

Conclusion (I)

We identified altered DNAm profiles in blood cells from TBD patients. These alterations were most prominent in symptomatic patients, regardless of telomere length, suggesting that DNAm alterations in blood cells could be involved in disease progression. However, we cannot exclude the possibility that severe telomere shortening lead to these DNAm changes, supported by the findings of altered DNAm in patients with extremely short telomere length and no hematological symptoms. Furthermore, the functional relevance of the DNAm alterations in *MAS1L*, *NAV2*, *PRDM8*, *SMC4*, *TM4FS1*, *VARS*, and *WNT6* in terms of gene expression, telomere maintenance, and disease progression in TBD requires further evaluation.

Semimethylation is a feature of diffuse large B-cell lymphoma, and subgroups with poor prognosis are characterized by global hypomethylation and short telomere length (II)

DLBCL is an aggressive B-cell lymphoma with a poor prognosis in the one-third of patients that have primary refractory or relapsing disease [101,229]. New predictive markers are needed to stratify patients into high- and low-risk groups. DNAm and telomere maintenance alterations are both hallmarks of cancer [5]. DNAm alterations in DLBCL have been identified previously [146,159-168,230]. However, there have been conflicting results regarding the prognostic significance of DNAm alterations in DLBCL and the relevance of DNAm-based subtyping. Most of these studies have focused on promoter regions or selected genes relevant for tumorigenesis [159-163,166-168,230], while only a few have focused on global DNAm alterations [146,164,165]. Additionally, both short and long telomere length in blood have been associated with increased cancer risk [21,89]. Therefore, we examined if global DNAm alterations and telomere length had prognostic significance in large B-cell lymphoma (LBCL) and were relevant for DLBCL subtyping. A total of 93 LBCL patients were included: DLBCL-GC (n=36), DLBCL-nonGC (n=30), HGBL (n=7), PCNSL (n=8), and t-DLBCL (n=12). The median age of the study cohort was 69 years (range: 25-89 years), 59% (55/93) were men and 73% (68/93) received R-CHOP-like treatment. One normal tonsillar GC B-sample and publicly available DNAm data from sorted normal peripheral blood B-cells (n=28) were included as controls. The methylation profiles of the LBCL patients were further compared to methylation profiles from previously published B-cell neoplasms and normal plasmablasts.

Global DNAm in large B-cell lymphoma and normal cells (II)

The percentage of hypomethylated ($\beta<0.15$), semimethylated ($0.15\leq\beta\leq0.80$), and hypermethylated ($\beta>0.80$) CpGs was calculated for each individual (Figure 11). There was a significant difference in median methylation levels between LBCL, normal GC B-cells, and normal B-cells ($\chi^2=19.696$, df=4, $p<0.001$). The LBCL patients had an increased percentage of semimethylated CpGs (mean=46%, range:25-61%) compared to the normal GC B-cells (mean=27%) and normal B-cells (mean=20%, range:15-30%), which have been described previously [146,162]. Density plots of the β-value distribution within each sample confirmed that all LBCL entities had an aberrant semimethylated pattern. The LBCL patients also had a decreased percentage of hypermethylated CpGs (mean=29%, range:12-44%) compared to normal GC B-cells (mean=45%) and normal B-cells (mean=53%, range:45-58%).

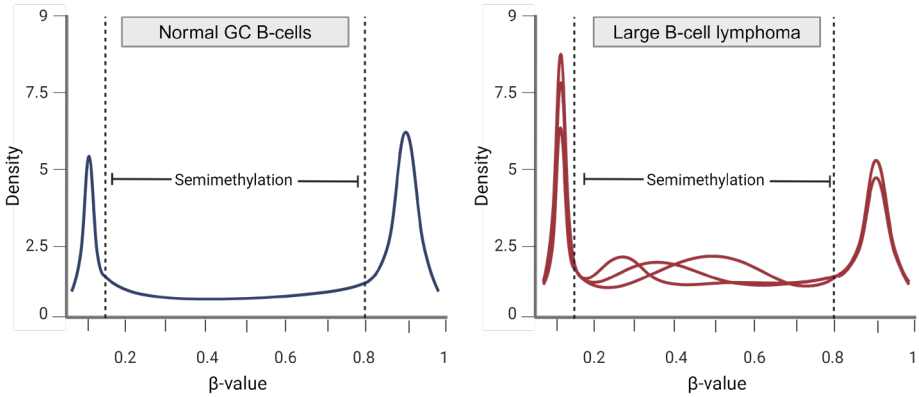


Figure 11. Example of β-value distribution in normal GC B-cells (left) and LBCL (right). The vertical dashed lines represent β=0.15 and β=0.8. CpG sites with β-values less than 0.15 were considered hypomethylated, β-values between 0.15 and 0.8 semimethylated, and β-values of more than 0.8 hypermethylated. Created with BioRender.com.

We compared the β-value distribution of LBCL with publicly available DNAm data from B-cell precursor acute lymphoblastic leukemia (BCP-ALL, n=663), chronic lymphocytic leukemia (CLL, n=75), mantle cell lymphoma (MCL, n=72), multiple myeloma (primary plasma cell leukemia, PCL, n=14), DLBCL (n=49), and normal plasmablasts (n=8). Overall, the β-value distribution in BCP-ALL, CLL, MCL, PCL, and normal plasmablasts was homogenous and had low levels of semimethylation. In contrast, the DLBCL validation cohorts (from two separate studies) had heterogeneous methylation patterns and abundant semimethylation. These results suggest that semimethylation is a characteristic of DLBCL and other LBCL entities. LBCL originates from the GC maturation stage, which has active mechanisms for DNAm alterations [94,146]. Additionally, GC B-cell-derived lymphomas, including DLBCL, often have mutations in epigenetic regulators [94,158]. Thus, the GC B-cell-derived lymphomas may render tumor subclones that retain or increase the methylation heterogeneity from their cell of origin.

Entity-specific DNAm alterations in LBCL (II)

The majority of DLBCL can be classified into two subgroups based on gene expression profiling [98,99]. However, subclassification based on DNAm profiles has been inconsistent [159,162,166]. We identified 1148 DM-CpGs (cutoff = mean $|\Delta\beta| \geq 0.2$) located in 639 genes between DLBCL-GC (n=36) and DLBCL-nonGC (n=30). Cluster analysis of these DM-CpGs could not differentiate between the entities. Gene ontology analysis of the 639 genes associated with the DM-CpGs

did not identify significant entity-specific networks. These results indicated that there were few entity-specific methylation differences between DLBCL-GC and DLBCL-nonGC, which is in concordance with other studies [159,162,166,167]. In contrast, Duran-Ferrer et al. have published a pan B-cell neoplasm classifier that can distinguish between several B-cell neoplasms and corresponding entities, including DLBCL-GC and DLBCL-ABC [164]. In our cohort, the classifier correctly identified 58/66 samples as DLBCL (sensitivity = 88%), but it was not applicable for DLBCL subclassification (GC: sensitivity = 42% and specificity = 87%. nonGC: sensitivity = 87% and specificity = 42%). DLBCL-GC and DLBCL-nonGC originate from slightly different GC B-cell stages but these maturation differences do not seem to be reflected in their DNAm profiles. The GC B-cells undergo extensive epigenetic alterations, somatic hypermutation, and class-switch recombination, resulting in cellular heterogeneity [94]. The extensive remodeling of the epigenome within GCs may obstruct attempts to separate DLBCL entities by DNAm profiles.

HGBL and t-DLBCL did not have entity-specific methylation profiles. However, the cohorts were small. HGBL has features of both DLBCL and Burkitt lymphoma [231], and the methylation profiles of HGBL and DLBCL overlapped in our cohort. The lack of entity-specific DNAm alteration in t-DLBCL was surprising. The t-DLBCLs originated from indolent lymphomas that could have been present for many years before transforming into DLBCL. Thus, we expected an increased epigenetic age in t-DLBCL compared to *de novo* DLBCL, but there was no significant difference. This could indicate that the methylation profiles of the indolent lymphomas were stable during progression, which has been suggested in follicular lymphoma [170,232]. Another explanation could be that the subclones that developed into aggressive lymphoma had a short replicative history.

PCNSL had an entity-specific DNAm profile. PCNSL had significantly increased CGI methylation (mean β=0.28±0.02) compared to DLBCL-GC (mean β=0.24±0.05, p=0.018) and DLBCL-nonGC (mean β=0.24±0.04, p=0.012), which was in line with a previous study [169]. All PCNSL patients were CIMP+ and had increased epigenetic age compared to the other entities. Further, they had the highest frequency of unique DM-CpGs compared to normal GC B-cells (n=111,072, cutoff = mean $|\Delta\beta| \geq 0.4$) but were not significantly more heterogenous than the other entities, based on the MVS heterogeneity score [167]. PCNSL also had significantly reduced global methylation (mean β=0.43±0.04) compared to DLBCL-GC (mean β=0.50±0.09, p=0.037) and t-DLBCL (mean β=0.53±0.04, p=0.014). Global loss of methylation in combination with increased methylation of CGIs and genes targeted by the polycomb group proteins has been associated with cellular aging and malignancies [4,139,233]. Whether the methylation alterations in PNCSL reflected replicative history or whether they were lymphoma-specific alterations needs further evaluation.

Prognostic relevance of telomere length in large B-cell lymphoma (II)

Normal telomere length was defined as age-adjusted RTL values within ±3 SD of the mean RTL of normal peripheral blood cells (n=174 samples). LBCL had heterogeneous telomere length, and there was no significant difference between any of the entities. The largest variation in telomere length was observed in DLBCL-GC, which had a subset of cases with extremely long telomeres (n=5, SD range: 8.5-18.9). Long telomeres in a subset of DLBCL-GC patients have previously been identified [85], but the biological processes leading to this phenotype are unknown. Heterogenous telomere length, including very long telomeres, are associated with the ALT phenotype [234]. ALT has been described in tumors originating from mesenchymal and neuroepithelial cells, such as osteosarcoma and glioma, but not in hematological malignancies [80,81]. Additional evaluation of the DLBCL-GC with long telomeres is necessary to understand and verify these findings. The ALT phenotype can be assessed by the C-circle assay and telomerase activity through the telomeric repeat amplification protocol [207,235]. Furthermore, genetic subgrouping, for example with the LymphGen classification algorithm, could give further insights into altered telomere maintenance mechanisms in DLBCL [110].

Short telomere length can cause genetic instability that promotes malignant transformation [16,52]. Several studies have investigated the relevance of telomere length as a prognostic marker in cancer, but the results have been inconsistent [85,87,236]. A meta-analysis of 45 independent studies found that short telomere length was associated with worse overall survival and progression in CLL, increased mortality risk in colorectal cancer, and decreased overall survival in esophageal cancer [236]. A study of B-cell lymphomas/leukemias found that short telomere length in CLL, but not in MCL or DLBCL, was associated with worse overall survival [85]. It is possible that abnormalities in telomere length might contribute to the onset of malignancy but may not be crucial for prognosis. Alternatively, telomere length could have prognostic value in certain cancer types but not others. We analyzed disease-specific survival (DSS) and progression-free survival (PFS) in patients treated with R-CHOP-like regimens (DLBCL, n=56 and LBCL, n=68). Short telomere length was independently associated with worse DSS in DLBCL (HR: 5.077, 95%CI:1.121-22.997, p =0.035) (Table 4) and LBCL (HR: 6.011, 95%CI: 1.319-27.397, p=0.020) in the multivariable analysis. Short telomere length was also independently associated with worse PFS in LBCL (4.689, 95%CI: 1.102-19.963, p=0.037) in the multivariable survival analysis. The number of events was insufficient to evaluate the LBCL entities separately, but it could be interesting in future studies.

Table 4. Multivariable survival analysis of disease-specific survival (DSS) in DLBCL-GC and DLBCL-nonGC treated with R-CHOP-like regimens. Quartile- and median classification were based on the entire LBCL cohort (n=93). Significant p-values (p<0.05, Wald's test) are indicated as bold text.

		DSS	
Variables (n=56)	**Events (n=12)**	**HR (95%CI)**	**p-value**
Age			
<median (n=31)	7	Reference	
≥median (n=25)	5	1.064 (0.278-4.068)	0.927
aaIPI			
0-1 (n=33)	3	Reference	
2-3 (n=23)	9	6.366 (1.558-26.024)	**0.010**
Entity			
GC (n=31)	5	Reference	
nonGC (n=25)	7	2.401 (0.566-10.185)	0.235
Age-adjusted RTL			
Normal (n=39)	4	Reference	
Short (n=8)	5	5.077 (1.121-22.997)	**0.035**
Long (n=9)	3	1.911 (0.358-10.192)	0.449
Hypomethylated CpGs %			
Q1-Q3 (n=33)	4	Reference	
<Q1 (n=12)	2	1.162 (0.196-6.887)	0.869
>Q3 (n=11)	6	6.920 (1.499-31.943)	**0.013**

Prognostic relevance of DNAm in large B-cell lymphoma (II)

Increased promoter methylation variability in DLBCL compared to normal GC B-cells has been associated with worse PFS [167]. In our R-CHOP-like treated LBCL cohort, neither promoter methylation variability nor global methylation variability (based on the MVS heterogeneity score) were significant prognostic markers. Additionally, the CIMP panel did not hold prognostic relevance in our cohort. CIMP classification, along with minimal residual disease, is a strong prognostic marker in T-ALL and also has prognostic value in relapsed BCP-ALL [184,185]. However, due to the extensive remodeling of the DNAm landscape in B-

cells during maturation, there might not be the same CGI alterations that hold prognostic value in mature B-cells. Instead, a high percentage of global hypomethylation was independently associated with worse DSS in DLBCL (HR:6.920, 95%CI: 1.499-31.943, p=0.013) (Table 4) and LBCL (HR: 5.147, 95%CI: 1.239-21.388, p=0.024), and with worse PFS in DLBCL (HR: 4.923, 95%CI: 1.286-18.849, p=0.020) (Table 5) in the multivariable survival analysis. The prognostic relevance of global hypomethylation is in concordance with a previous study [168].

We observed that patients with a high percentage of hypomethylated CpGs (Hypo>Q3) also had a high percentage of hypermethylated CGIs. We filtered out the CGIs that were hypermethylated in at least 70% of the Hypo>Q3 patients and in no more than 30% of the other R-CHOP-like treated DLBCL patients (n=550 GGIs). Gene ontology analysis of the genes (n=402) associated with these CpGs showed that the significant networks included signal transduction, reproduction, neurophysiological processes, development, cell adhesion, and cardiac development. We performed the analysis again, but this time on the hypermethylated promoter-associated CGIs in Hypo>Q3. Significant networks associated with these genes included signal transduction, neurophysiological processes, development, and cell adhesion. These categories are broad, and the relevance of these networks needs further evaluation. Unfortunately, we did not have any gene expression data, so whether these methylation differences (global hypomethylation and CGI hypermethylation) affected gene expression in our cohort is unknown but could be evaluated in future studies. We speculated that tumor subclones with global hypomethylation and CGI hypermethylation could have survival advantages by inactivating tumor suppressor genes through CGI promoter-methylation [120]. As discussed previously, global hypomethylation and CGI hypermethylation increase in normal B-cells throughout their maturation program [146]. Thus, it could be that the Hypo>Q3 DLBCL tumors had acquired these methylation changes due to an increased replicative history, or that cancer-specific epigenetic mechanisms were dysregulated to a greater extent in these patients. Evaluation of mutation status in relation to DNAm alterations in DLBCL would be an interesting subject in upcoming studies.

Table 5. Multivariable survival analysis of progression-free survival (PFS) in DLBCL-GC and DLBCL-nonGC treated with R-CHOP-like regimens. Quartile- and median classification were based on the entire LBCL cohort (n=93). Significant p-values (p<0.05, Wald's test) are indicated as bold text.

		PFS	
Variables (n=56)	**Events (n=14)**	**HR (95%CI)**	**p-value**
Age			
<median (n=31)	9	Reference	
≥median (n=25)	5	0.699 (0.199-2.456)	0.576
aaIPI			
0-1 (n=33)	4	Reference	
2-3 (n=23)	10	6.866 (1.746-27.003)	**0.006**
Entity			
GC (n=31)	5	Reference	
nonGC (n=25)	9	3.013 (0.815-11.135)	0.098
Age-adjusted RTL			
Normal (n=39)	5	Reference	
Short (n=8)	6	3.585 (0.860-14.945)	0.080
Long (n=9)	3	1.414 (0.297-6.736)	0.664
Hypomethylated CpGs %			
Q1-Q3 (n=33)	6	Reference	
<Q1 (n=12)	2	0.732 (0.132-4.060)	0.722
>Q3 (n=11)	6	4.923 (1.286-18.849)	**0.020**

Conclusion (II)

We have shown that semimethylation is a characteristic of DLBCL and other LBCL entities. Furthermore, we identified a subgroup of DLBCL-GC cases with extremely long telomere length that requires further investigation. Short telomere length and a high percentage of global hypomethylation were both independent prognostic factors for disease-specific survival in our cohort. The subpopulation with the highest percentage of global hypomethylation also had a high percentage of hypermethylated CGIs. These methylation alterations could potentially be targets for epigenetic therapy, including hypomethylating agents.

Functional analysis of telomerase activity in T-lymphocytes as a diagnostic tool for pathogenicity assessment of novel genetic variants in telomere biology disorders (III)

The first connection between genetic alterations and TBD was identified in 1998 when variants in *DKC1* were linked to dyskeratosis congenita [73]. Since then, 16 additional telomere-related genes have been identified [7]. Pathogenicity assessment of genetic variants follows the consensus guidelines from the American College of Medical Genetics and Genomics (ACMG). Support for pathogenicity includes an absence of the variant in the general population, in silico prediction algorithms indicating an effect of the variant, and co-segregation of the variant with disease in the family [75]. Genetic variants in telomere-associated genes are evaluated according to the ACMG criteria, with the addition of telomere length analysis [65,76]. Pathogenicity assessment of novel genetic variants in suspected TBD presents several challenges, for example heterogenous clinical presentation and genetic anticipation [58]. Additional evaluation tools could facilitate the classification and reduce the number of genetic variants of unknown significance (VUS). Therefore, we developed a functional assay of telomerase activity (TA) in patient-derived cells to evaluate genetic variants in TBD. The assay was based on the telomeric repeat amplification protocol (TRAP) combined with qPCR. We included 100 controls with an even distribution of age and gender (range:18-73 years, 51% men) and six TBD patients with a pathogenic (*TERT*, n=1) or likely-pathogenic (*TERT*, n=4 and *TERC*, n=1) variant. TA and S-phase fraction (%) were measured in activated T-cells in all patients and controls. RTL was measured in 90 controls and all TBD patients.

Table 6. Mean value and standard deviation of age, telomerase activity (TA), relative telomere length (RTL), and proliferative capacity (S-phase %) in controls (n=100) and TBD patients (n=6). The statistical analysis was performed with the Mann-Whitney U test and p-values <0.05 were considered significant.

	Controls (n=100)	**TBD patients (n=6)**	**p-value**
Age	44.5±14.3	55.3±16.1	0.147
TA	0.56±0.34	0.11±0.07	<0.001
RTL	1.69±0.25	0.99±0.21	<0.001
S-phase %	35.7±7.4	30.1±4.7	0.060

Telomerase activity, telomere length, and proliferative capacity in TBD and controls (III)

The TBD patients had significantly decreased TA ($p<0.001$) and RTL ($p<0.001$) compared to controls but no significant difference in S-phase % ($p=0.060$) or age ($p=0.147$) (Table 6). There have been conflicting results regarding the association between TA and age in activated lymphocytes [39,54,237]. Some studies have reported a negative correlation between TA and age in activated T-cells [39,237] while others have not found any association [54]. There was no significant correlation between TA and age in our control cohort (adjusted $R2=0.014$, $p=0.123$) (Figure 12). However, increasing age was associated with reduced proliferative capacity (adjusted $R2=0.473$, $p<0.001$) and there was a weak positive correlation between TA and S-phase % (adjusted $R2=0.149$, $p<0.001$) (Figure 12). These results were in accordance with a previous study that identified a significant correlation between TA in activated T-cells and proliferative capacity [35].

We performed a multiple linear regression analysis with TA as the dependent variable and S-phase %, age, and RTL as the independent variables. RTL was not significantly associated with TA ($p=0.068$) and was excluded from the model. After the exclusion of RTL, age was no longer significantly associated with TA ($p=0.075$). The results suggest that proliferative capacity is a confounding variable that influences the association between age and TA. Hence, we adjusted the TA-values for S-phase % by fitting a linear regression model to the control data: $Y = \beta_0 + \beta_1 x_1 + \epsilon$. Y was the ΔCt-values (since the TA-values had a logarithmic scale, see section titled Telomerase activity measurements, page 30), β_0 the intercept, β_1 the slope coefficient of S-phase %, and ε the model's prediction error. Negative ΔCt-values meant that the sample had a lower Ct than the positive control and correspondingly a higher TA. The adjusted ΔCt (ΔCt_{adj}) was calculated for the controls and TBDs using the intercept and slope coefficients.

Pathogenicity evaluation of genetic variants (III)

The controls had a ΔCt_{adj} range of ± 3 SD. Three of the TBD patients had ΔCt_{adj}-values 1-3 SD above the mean of the controls (Figure 13). These patients had likely-pathogenic variants (*TERT*p.Leu1017_Leu1019del, *TERT* p.Asp684Gly, and *TERC* n.64G>A*)* according to the ACMG criteria [75] and the TA analysis confirmed pathogenicity at a lower level. The other three TBD patients had ΔCt_{adj}-values 3 SD or more above the mean of controls which clearly separated them from controls. These patients had a pathogenic (*TERT* p.Arg865His) or likely-pathogenic variant (*TERT* p.Tyr1002Cys, n=2) according to the ACMG criteria. Reduced TA in the *TERT* p.Arg865His variant have been published previously by

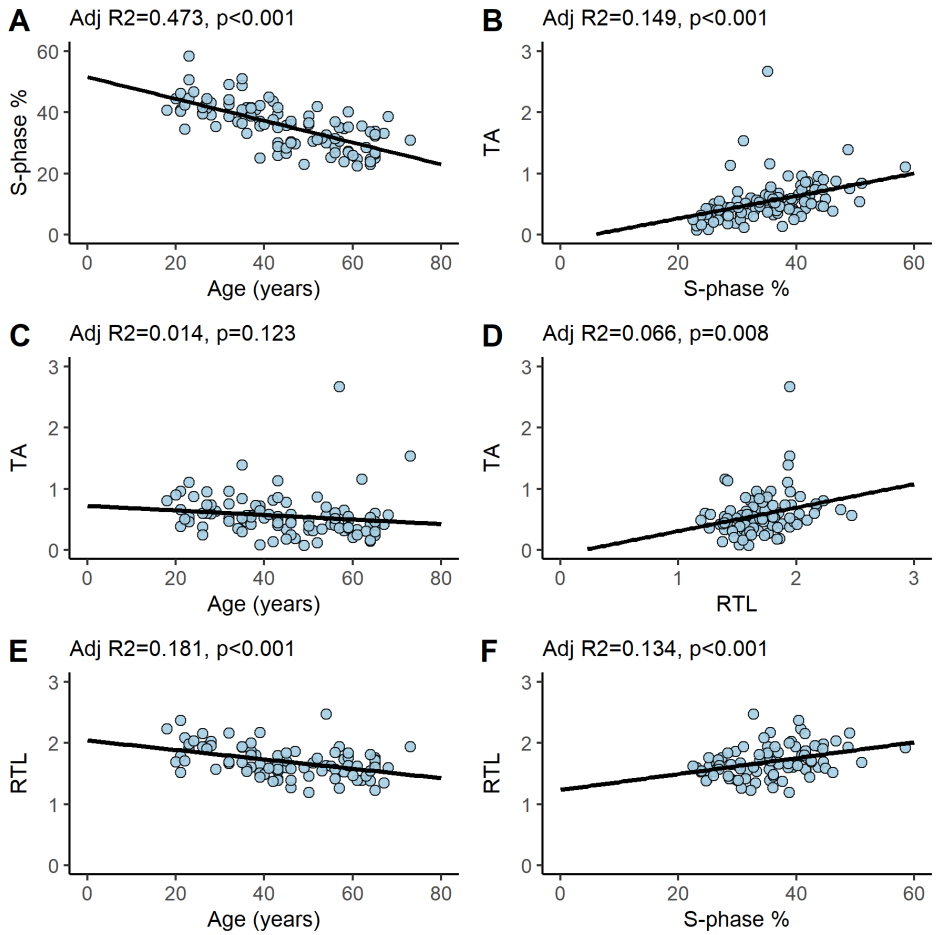


Figure 12. Linear regression analysis of the relationship between proliferative capacity (S-phase %), telomerase activity (TA), age, and relative telomere length (RTL). The S-phase and TA was measured in activated T-cells (n=100). RTL was measured in leukocytes (n=90). A) S-phase % and age. B) TA and S-phase %. C) TA and age. D) TA and RTL. E) RTL and age. F) RTL and S-phase %.

others [238-240]. The *TERT* p.Tyr1002Cys variant is novel and identified during clinical evaluation at the Department of Clinical Genetics, Umeå University Hospital, Sweden. The significant reduction in TA and short RTL supports that the *TERT* p.Tyr1002Cys variant is pathogenic.

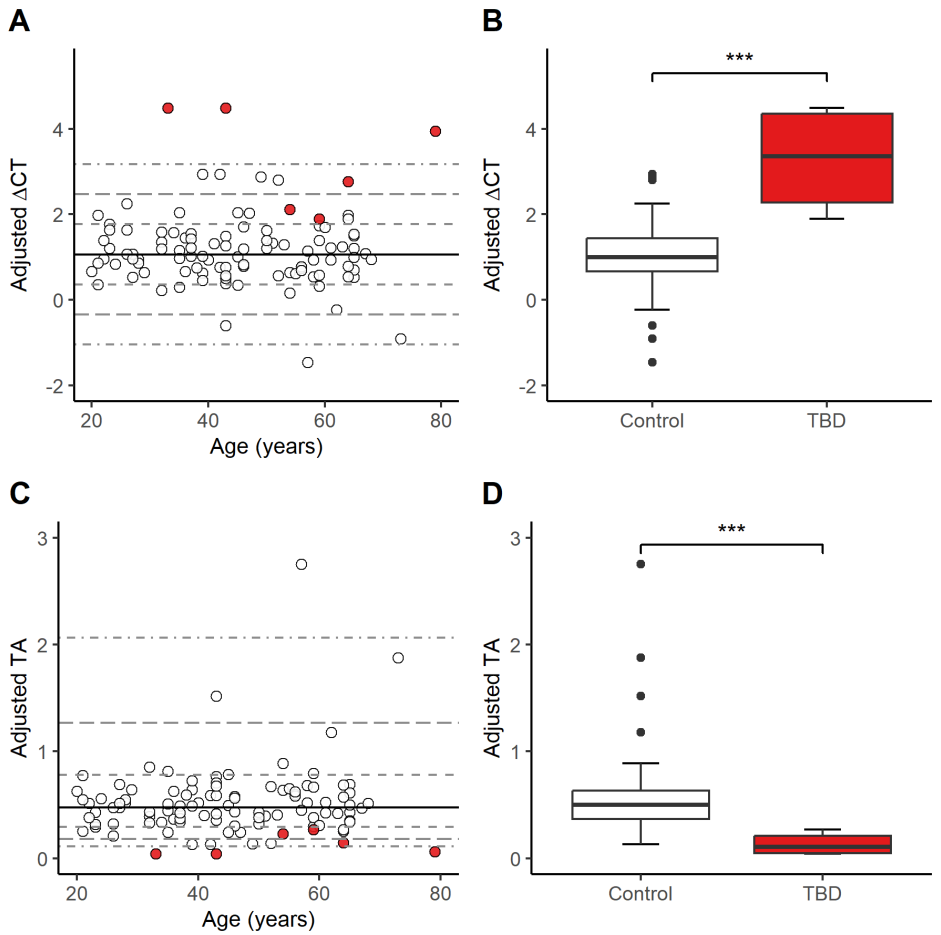


Figure 13. ΔCT and TA adjusted for S-phase % and plotted against age. Controls are represented as white circles and TBDs as red circles. *** corresponds to p-values <0.001. A) Adjusted ΔCT in controls and TBD plotted against age. The black solid line represents the mean of the controls and the black dashed lines standard deviation ±1, ±2, and ±3. B) Boxplots of adjusted ΔCT in controls (white) and TBD (red). The black solid line represents the median. C) Adjusted TA in controls and TBD plotted against age. The black solid line represents the mean adjusted TA (converted from the mean ΔCT) of the normal controls and the black dashed lines standard deviation ±1, ±2, and ±3 from the mean (converted from the ΔCT standard deviations). D) Boxplots of adjusted TA in controls and TBD. The black solid line represents the median.

Assay reproducibility and preanalytical effect on telomerase activity (III)

The TA measurement by qPCR was run on two separate occasions for sixteen samples (the six TBDs and ten randomly selected controls) to monitor the TA variation between PCR plates. The inter-assay coefficient of variation was considerate acceptable (median=9.6%, range: 0-24%). We further analyzed if time between blood withdrawal and mononuclear cell (MNC) isolation had any preanalytical effect on TA. Eight additional normal peripheral venous blood samples were collected. The blood samples were aliquoted and the MNC were isolated at the same day as blood withdrawal (n=8), after one day (n=6), and after two days (n=3). There was no significant difference in TA_{adj} between day 0 and day 1 (p=0.06) or between day 0 and day 2 (p=0.025). The TBD samples in our cohort were referred to the Umeå University Hospital from other medical centers in Sweden and MNC was isolated within 32 hours (range: 7-32 hours) of blood withdrawal. Based on our small-scale evaluation, we consider a transportation time of up to 48 hours acceptable for TA analysis. However, there were few measuring points included in this analysis and the preanalytical effect on TA should be evaluated further.

Conclusion (III)

Activated T-cells from TBD patients had reduced TA compared to normal activated T-cells. Pathogenicity was strongly supported for variants with a TA of more than 3 SD below the mean TA of controls. Thus, functional analysis of TA in patient-derived cells could support pathogenicity evaluation in clinical diagnostics and reduce the number of reported VUS in TBD.

Strengths and limitations of this thesis

Paper I

The TBD cohort was well-characterized and large considering the rarity of the disease. To our knowledge, this was the first study of DNAm in a cohort including not only dyskeratosis congenita but also other forms of TBD. The limitations of the study were that we did not have gene expression data, and thus, we could not investigate whether the DNAm alterations affected gene expression. Furthermore, we did not have enough cases to examine DNAm alterations in asymptomatic TBD with ES-RTL or non-hematological TBD with S-RTL. Such analysis could have provided valuable insights of whether DNAm alterations resulted from severe telomere attrition or were involved in disease progression and phenotypic presentation.

Paper II

Our cohort included well-characterized cases of DLBCL, HGBL, PCNSL, and t-DLBCL. The large sample size of DLBCL allowed for the evaluation of telomere and DNAm alterations between the GC and nonGC subtypes. As far as we know, no previous study of DNAm alterations in HGBL has been conducted in humans. A limitation with the study was the retrospective design and the small sample sizes of HGBL, PCNSL, t-DLBCL, and the normal GC B-cells. Also, we did not have a validation cohort to confirm the prognostic relevance of global hypomethylation and short telomere length in DLBCL.

Paper III

The control cohort was large and evenly distributed in terms of gender and age across several decades. Although the TA assay was evaluated in a few TBD patients, we demonstrated its relevance for assessing genetic variants in telomere-associated genes. The study's limitation were the few TBD cases, restricted to variants in *TERT* and *TERC*.

General conclusions and future perspectives

The overall aim of this thesis was to investigate alterations in telomere biology and epigenetics to improve the understanding of underlying factors contributing to telomere biology disorders and large B-cell lymphoma.

We demonstrated that a functional assay of TA in activated T-cells from TBD patients supported pathogenicity evaluation in clinical diagnostics. This analysis would particularly benefit patients with a variant of unknown significance (VUS), since VUS is not actionable in clinical practice. Five *TERT* and one *TERC* variants have been analyzed so far and pathogenicity was strongly supported for *TERT* p.Arg865His and *TERT* p.Tyr1002Cys. The other variants were classified as likely-pathogenic according to the ACMG criteria. Further evaluation of variants in telomere-related genes is necessary before the assay can be implemented as a diagnostic tool. Blood samples from patients with suspected TBD may be transported between medical centres for evaluation. Therefore, the time from blood sampling to T-cell stimulation should be tested on a larger scale to ensure that the quality of the samples is not compromised.

The telomere shortening resulting from pathogenic genetic variants do not fully explain the heterogenous clinical presentation in TBD. The absence of a direct correlation suggests that other molecular mechanisms are involved in the disease processes. We identified DNAm alterations in symptomatic TBD patients, indicating that DNAm could be involved in disease progression. However, similarities in the DNAm profiles of asymptomatic and symptomatic patients with extremely short telomeres was also identified. These results could indicate that the DNAm alterations were markers of extensive telomere shortening or that the asymptomatic patients were in transition towards symptom development. However, the study included a limited number of asymptomatic patients with extremely short telomeres and these results should be confirmed in a larger cohort. A longitudinal study of DNAm profiles in asymptomatic TBD could be valuable to clarify whether DNAm in TBD results from telomere attrition, symptom development or both. Furthermore, we identified seven genes with two or more differentially methylated CpG sites. Functional analysis of these genes in terms of gene expression, telomere maintenance, and disease progression would provide further insights of their potential role in TBD.

We identified that LBCL was characterized by extensive semimethylation that was absent in other B-cell neoplasms and in normal cells. DLBCL-GC and DLBCL-nonGC subtypes did not have entity-specific DNAm profiles, which was in accordance with several other studies. Interestingly, a subset of DLBCL-GC patients had extremely long telomeres. This phenotype is associated with the ALT mechanism rather than upregulated telomerase activity. Although, ALT has not

been described in hematological malignancies, our results indicate that it might be present in DLBCL. We will continue investigating telomere maintenance mechanisms in an extended DLBCL cohort. Additionally, global hypomethylation and short telomere length were significant prognostic markers associated with worse prognosis. Tumor samples with a high percentage of hypomethylated CpGs also had a high percentage of hypermethylated CGIs, a pattern that is common in aging and malignant cells. Hypermethylation of promoter-associated CGIs could inactivate tumor suppressor genes and thereby promote malignant progression. Gene expression profiling of the hypomethylated/CGI hypermethylated tumor samples would provide valuable insights into affected pathways. If tumor suppressor genes are inactivated in these tumors, hypomethylating drugs targeting the CGI hypermethylation could be relevant. So far, we have not investigated if genetic alterations are associated with DNAm or telomere length in LBCL. The LymphGen algorithm, a new classification tool that subgroups DLBCL based on genetic alterations, could be used in future studies to investigate the association between genetic, DNAm, and telomere alterations in DLBCL.

Acknowledgments

I am deeply grateful to everyone who has contributed to and supported me throughout my doctoral studies and the writing of this thesis. I want to express my gratitude towards all the participants included in my research. Without your sample donations, there would not have been any thesis to write. I would also like to thank the financial support from the Medical Faculty of Umea University, the Kempe Foundation, the Swedish Cancer Society, the Cancer Research Foundation in Northern Sweden, the Lion's Cancer Research Foundation in Northern Sweden, and the regional agreement between Umeå University and Västerbottens County Council on cooperation in the fields of Medicine, Odontology, and Health. Furthermore, I want to acknowledge the Hultdin/Degerman research group, my colleagues at TRC, friends, and family.

To my supervisor **Magnus Hultdin**. Working together has been a joy, and I have learned so much from you. We have had so many fun and interesting discussions and a few valuable disagreements. You must also be the least stressed person I have met, and I am saying that because I used to be the least stressed person I had ever met. Although you have a ton of work to do, you have always taken time for me and my questions whenever I pop into your office. Some of those questions could definitely have waited, but you never made me feel that way.

My co-supervisor **Sofie Degerman**. My thoughts and ideas have been flying around in my head so many times and you have been the best at settling them down. When I've been completely stuck in a perspective (my specialty) you have provided calmness, structure, alternative perspectives, and pep talks. If I could get half as organized as you are, I will do good.

My co-supervisor **Anna Norberg**. I have enjoyed running down to your office with all my thoughts and questions about telomeres and telomere biology disorders. You are a great storyteller and your knowledge have been of great value to me.

To my colleagues in the Hultdin/Degerman group. **Mattias**, unfortunately for you, I became interested in bioinformatics at the beginning of my PhD and I am so thankful for all time and effort you have put into explaining the field to me. You are a great teacher with a huge amount of patience. **Pia**, I have enjoyed our travels and all the great conversations throughout the years. You have taught me that research is one of the most interesting things you can do and also that it is okay not to love every moment of it. **Fernanda**, working with you and becoming your friend has been a joy. You have shown me what hard work and dedication can do and also that it is important to dress up and drink beer (I have been better

at one of them). **Nina**, my first own PhD colleague and exactly the energy vibe we all needed. You are one of the most kind-hearted and caring people I have met. I have really enjoyed all the dog activities we've had outside of work and although they are on pause now, I hope they will continue in the future. **Elina** and **Sigrid**, one sad part about finishing my PhD is that I didn't get to spend more time with you, but our short time was lovely. I look forward to reading your papers and follow your research. Thank you for all your hard work on our paper, Elina, it turned out great!

To the dedicated lunch/social corner group: **Isabelle, Laura, Nina, Fernanda, Tommy, Erik, Britta, Marga, Parniyan, Jessica, Sai, Helena, Arvin, Joshua, Nick, Vicky, Justin, Alva, Tova, Alicia,** and I'm so sorry If I forgot to add someone's name, it's been a couple of rough weeks. I've had a blast with you all for the last five years. Thank you, **Isabelle, Laura, Matthew, Arvin**, and **Niklas** for adopting me to your ALS community.

To **all my friends**. I am so lucky to have you all in my life.

To **Johanna** and **Karolina**, who shared the PhD journey with me. Thank you for all the pep talks and discussions, and thank you Karro for letting me use your beautiful painting as my thesis cover. To **Dennis** for all the scientific discussions, weird fun facts, and for taking such good care of Sivan. To my furry friend **Sivan** that swept into my life a year ago. You have no idea what I do at work and you could not care less and that is all I need sometimes.

To my **family**, and especially to **mormor** and **morfar**. Tack för att ni alltid finns vid min sida.

To **Anton**. My greatest supporter both in and out of science. Thank you for supporting my PhD journey with everything that came with it. Your passion for the biotechnology and medicine field is admirable, and whenever I am sad and tired, you remind me why we chose this path. You also remind me that work and achievements are just one part and that there is so much more to life. I love you.

References

1 Aunan, J. R., Watson, M. M., Hagland, H. R. & Søreide, K. Molecular and biological hallmarks of ageing. *Br J Surg* **103**, e29-46, doi:10.1002/bjs.10053 (2016).

2 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115, doi:10.1186/gb-2013-14-10-r115 (2013).

3 Marioni, R. E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**, 25, doi:10.1186/s13059-015-0584-6 (2015).

4 López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243-278, doi:10.1016/j.cell.2022.11.001 (2023).

5 Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46, doi:10.1158/2159-8290.Cd-21-1059 (2022).

6 López-Otín, C., Pietrocola, F., Roiz-Valle, D., Galluzzi, L. & Kroemer, G. Meta-hallmarks of aging and cancer. *Cell Metab* **35**, 12-35, doi:10.1016/j.cmet.2022.11.001 (2023).

7 Revy, P., Kannengiesser, C. & Bertuch, A. A. Genetics of human telomere biology disorders. *Nat Rev Genet* **24**, 86-108, doi:10.1038/s41576-022-00527-z (2023).

8 Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514, doi:10.1016/s0092-8674(00)80760-6 (1999).

9 de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**, 2100-2110, doi:10.1101/gad.1346005 (2005).

10 Shay, J. W. & Wright, W. E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867-874, doi:10.1093/carcin/bgh296 (2005).

11 O'Sullivan, R. J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol* **11**, 171-181, doi:10.1038/nrm2848 (2010).

12 de Lange, T. Shelterin-Mediated Telomere Protection. *Annu Rev Genet* **52**, 223-247, doi:10.1146/annurev-genet-032918-021921 (2018).

13 Watson, J. D. Origin of concatemeric T7 DNA. *Nat New Biol* **239**, 197-201, doi:10.1038/newbio239197a0 (1972).

14 Grill, S. & Nandakumar, J. Molecular mechanisms of telomere biology disorders. *J Biol Chem* **296**, 100064, doi:10.1074/jbc.REV120.014017 (2020).

15 Turner, K. J., Vasu, V. & Griffin, D. K. Telomere Biology and Human Phenotype. *Cells* **8**, doi:10.3390/cells8010073 (2019).
16 Shay, J. W. Role of Telomeres and Telomerase in Aging and Cancer. *Cancer Discov* **6**, 584-593, doi:10.1158/2159-8290.Cd-16-0062 (2016).
17 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
18 Alberts., B. *et al. Essential Cell Biology*. Vol. Fourth Edition 726 (Garland Science (Taylor & Francis Group), 2014).
19 Pfeiffer, V. & Lingner, J. Replication of telomeres and the regulation of telomerase. *Cold Spring Harb Perspect Biol* **5**, a010405, doi:10.1101/cshperspect.a010405 (2013).
20 Zhu, X. *et al.* The association between telomere length and cancer risk in population studies. *Sci Rep* **6**, 22243, doi:10.1038/srep22243 (2016).
21 Ma, H. *et al.* Shortened telomere length is associated with increased risk of cancer: a meta-analysis. *PLoS One* **6**, e20466, doi:10.1371/journal.pone.0020466 (2011).
22 Wang, Q., Zhan, Y., Pedersen, N. L., Fang, F. & Hägg, S. Telomere Length and All-Cause Mortality: A Meta-analysis. *Ageing Res Rev* **48**, 11-20, doi:10.1016/j.arr.2018.09.002 (2018).
23 Newton, C. A. *et al.* Telomere length in patients with pulmonary fibrosis associated with chronic lung allograft dysfunction and post-lung transplantation survival. *J Heart Lung Transplant* **36**, 845-853, doi:10.1016/j.healun.2017.02.005 (2017).
24 Demanelis, K. *et al.* Determinants of telomere length across human tissues. *Science* **369**, doi:10.1126/science.aaz6876 (2020).
25 Njajou, O. T. *et al.* Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci U S A* **104**, 12135-12139, doi:10.1073/pnas.0702703104 (2007).
26 Diaz de Leon, A. *et al.* Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. *PLoS One* **5**, e10680, doi:10.1371/journal.pone.0010680 (2010).
27 Nordfjäll, K., Svenson, U., Norrback, K. F., Adolfsson, R. & Roos, G. Large-scale parent-child comparison confirms a strong paternal influence on telomere length. *Eur J Hum Genet* **18**, 385-389, doi:10.1038/ejhg.2009.178 (2010).
28 Broer, L. *et al.* Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *Eur J Hum Genet* **21**, 1163-1168, doi:10.1038/ejhg.2012.303 (2013).

29 Cowell, W. *et al.* Telomere dynamics across the early life course: Findings from a longitudinal study in children. *Psychoneuroendocrinology* **129**, 105270, doi:10.1016/j.psyneuen.2021.105270 (2021).
30 Hunt, S. C. *et al.* Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. *Aging Cell* **7**, 451-458, doi:10.1111/j.1474-9726.2008.00397.x (2008).
31 Steenstrup, T. *et al.* Leukocyte telomere dynamics in the elderly. *Eur J Epidemiol* **28**, 181-187, doi:10.1007/s10654-013-9780-4 (2013).
32 Aubert, G., Baerlocher, G. M., Vulto, I., Poon, S. S. & Lansdorp, P. M. Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. *PLoS Genet* **8**, e1002696, doi:10.1371/journal.pgen.1002696 (2012).
33 Nordfjäll, K. *et al.* The individual blood cell telomere attrition rate is telomere length dependent. *PLoS Genet* **5**, e1000375, doi:10.1371/journal.pgen.1000375 (2009).
34 Benetos, A. *et al.* Telomere length tracking in children and their parents: implications for adult onset diseases. *Faseb j* **33**, 14248-14253, doi:10.1096/fj.201901275R (2019).
35 Tedone, E. *et al.* Telomere length and telomerase activity in T cells are biomarkers of high-performing centenarians. *Aging Cell* **18**, e12859, doi:10.1111/acel.12859 (2019).
36 Gorenjak, V., Petrelis, A. M., Stathopoulou, M. G. & Visvikis-Siest, S. Telomere length determinants in childhood. *Clin Chem Lab Med* **58**, 162-177, doi:10.1515/cclm-2019-0235 (2020).
37 Henckel, E. *et al.* Hematopoietic cellular aging is not accelerated during the first 2 years of life in children born preterm. *Pediatr Res* **88**, 903-909, doi:10.1038/s41390-020-0833-6 (2020).
38 Lin, J. *et al.* Analyses and comparisons of telomerase activity and telomere length in human T and B cells: insights for epidemiology of telomere maintenance. *J Immunol Methods* **352**, 71-80, doi:10.1016/j.jim.2009.09.012 (2010).
39 Lin, Y. *et al.* Age-associated telomere attrition of lymphocytes in vivo is co-ordinated with changes in telomerase activity, composition of lymphocyte subsets and health conditions. *Clin Sci (Lond)* **128**, 367-377, doi:10.1042/cs20140481 (2015).
40 Andreu-Sánchez, S. *et al.* Genetic, parental and lifestyle factors influence telomere length. *Commun Biol* **5**, 565, doi:10.1038/s42003-022-03521-7 (2022).
41 Zhao, B., Vo, H. Q., Johnston, F. H. & Negishi, K. Air pollution and telomere length: a systematic review of 12,058 subjects.

Cardiovasc Diagn Ther **8**, 480-492, doi:10.21037/cdt.2018.06.05 (2018).
42 Blackburn, E. H., Epel, E. S. & Lin, J. Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **350**, 1193-1198, doi:10.1126/science.aab3389 (2015).
43 Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413, doi:10.1016/0092-8674(85)90170-9 (1985).
44 Greider, C. W. & Blackburn, E. H. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331-337, doi:10.1038/337331a0 (1989).
45 Srinivas, N., Rachakonda, S. & Kumar, R. Telomeres and Telomere Length: A General Overview. *Cancers (Basel)* **12**, doi:10.3390/cancers12030558 (2020).
46 Blackburn, E. H. & Collins, K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a003558 (2011).
47 Kiss, T., Fayet-Lebaron, E. & Jády, B. E. Box H/ACA small ribonucleoproteins. *Mol Cell* **37**, 597-606, doi:10.1016/j.molcel.2010.01.032 (2010).
48 Collins, K. Physiological assembly and activity of human telomerase complexes. *Mech Ageing Dev* **129**, 91-98, doi:10.1016/j.mad.2007.10.008 (2008).
49 Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644-648, doi:10.1126/science.1165357 (2009).
50 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* **18**, 173-179, doi:10.1002/(sici)1520-6408(1996)18:2<173::Aid-dvg10>3.0.Co;2-3 (1996).
51 Yamada, O., Motoji, T. & Mizoguchi, H. Up-regulation of telomerase activity in human lymphocytes. *Biochim Biophys Acta* **1314**, 260-266, doi:10.1016/s0167-4889(96)00104-8 (1996).
52 Shay, J. W. & Wright, W. E. Telomeres and telomerase: three decades of progress. *Nat Rev Genet* **20**, 299-309, doi:10.1038/s41576-019-0099-1 (2019).
53 Bodnar, A. G., Kim, N. W., Effros, R. B. & Chiu, C. P. Mechanism of telomerase induction during T cell activation. *Exp Cell Res* **228**, 58-64, doi:10.1006/excr.1996.0299 (1996).
54 Son, N. H., Murray, S., Yanovski, J., Hodes, R. J. & Weng, N. Lineage-specific telomere shortening and unaltered capacity for telomerase expression in human T and B lymphocytes with age. *J*

Immunol **165**, 1191-1196, doi:10.4049/jimmunol.165.3.1191 (2000).
55 Lupatov, A. Y. & Yarygin, K. N. Telomeres and Telomerase in the Control of Stem Cells. *Biomedicines* **10**, doi:10.3390/biomedicines10102335 (2022).
56 Vulliamy, T., Marrone, A., Dokal, I. & Mason, P. J. Association between aplastic anaemia and mutations in telomerase RNA. *Lancet* **359**, 2168-2170, doi:10.1016/s0140-6736(02)09087-6 (2002).
57 Fogarty, P. F. *et al.* Late presentation of dyskeratosis congenita as apparently acquired aplastic anaemia due to mutations in telomerase RNA. *Lancet* **362**, 1628-1630, doi:10.1016/s0140-6736(03)14797-6 (2003).
58 Vulliamy, T. *et al.* Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nat Genet* **36**, 447-449, doi:10.1038/ng1346 (2004).
59 Armanios, M. Y. *et al.* Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* **356**, 1317-1326, doi:10.1056/NEJMoa066157 (2007).
60 Savage, S. A. *et al.* TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. *Am J Hum Genet* **82**, 501-509, doi:10.1016/j.ajhg.2007.10.004 (2008).
61 Du, H. Y. *et al.* Complex inheritance pattern of dyskeratosis congenita in two families with 2 different mutations in the telomerase reverse transcriptase gene. *Blood* **111**, 1128-1130, doi:10.1182/blood-2007-10-120907 (2008).
62 Norberg, A. *et al.* Novel variants in Nordic patients referred for genetic testing of telomere-related disorders. *Eur J Hum Genet* **26**, 858-867, doi:10.1038/s41431-018-0112-8 (2018).
63 Calado, R. T. *et al.* Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. *Proc Natl Acad Sci U S A* **106**, 1187-1192, doi:10.1073/pnas.0807057106 (2009).
64 Schratz, K. E. & Armanios, M. Cancer and myeloid clonal evolution in the short telomere syndromes. *Curr Opin Genet Dev* **60**, 112-118, doi:10.1016/j.gde.2020.02.019 (2020).
65 Alder, J. K. *et al.* Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* **115**, E2358-e2365, doi:10.1073/pnas.1720427115 (2018).
66 Holohan, B., Wright, W. E. & Shay, J. W. Cell biology of disease: Telomeropathies: an emerging spectrum disorder. *J Cell Biol* **205**, 289-299, doi:10.1083/jcb.201401012 (2014).

67 Zhang, K., Xu, L. & Cong, Y. S. Telomere Dysfunction in Idiopathic Pulmonary Fibrosis. *Front Med (Lausanne)* **8**, 739810, doi:10.3389/fmed.2021.739810 (2021).
68 Savage, S. A. Beginning at the ends: telomeres and human disease. *F1000Res* **7**, doi:10.12688/f1000research.14068.1 (2018).
69 Kam, M. L. W., Nguyen, T. T. T. & Ngeow, J. Y. Y. Telomere biology disorders. *NPJ Genom Med* **6**, 36, doi:10.1038/s41525-021-00198-5 (2021).
70 Terada, K. *et al.* TERT and TERC mutations detected in cryptic dyskeratosis congenita suppress telomerase activity. *Int J Lab Hematol* **42**, 316-321, doi:10.1111/ijlh.13176 (2020).
71 Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat Rev Genet* **13**, 693-704, doi:10.1038/nrg3246 (2012).
72 Armanios, M. The Role of Telomeres in Human Disease. *Annu Rev Genomics Hum Genet* **23**, 363-381, doi:10.1146/annurev-genom-010422-091101 (2022).
73 Heiss, N. S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19**, 32-38, doi:10.1038/ng0598-32 (1998).
74 Vulliamy, T. J. *et al.* Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations. *PLoS One* **6**, e24383, doi:10.1371/journal.pone.0024383 (2011).
75 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424, doi:10.1038/gim.2015.30 (2015).
76 Du, H. Y. *et al.* TERC and TERT gene mutations in patients with bone marrow failure and the significance of telomere length measurements. *Blood* **113**, 309-316, doi:10.1182/blood-2008-07-166421 (2009).
77 Bhoopalan, S. V., Wlodarski, M., Reiss, U., Triplett, B. & Sharma, A. Reduced-intensity conditioning-based hematopoietic cell transplantation for dyskeratosis congenita: Single-center experience and literature review. *Pediatr Blood Cancer* **68**, e29177, doi:10.1002/pbc.29177 (2021).
78 Barbaro, P. & Vedi, A. Survival after Hematopoietic Stem Cell Transplant in Patients with Dyskeratosis Congenita: Systematic Review of the Literature. *Biol Blood Marrow Transplant* **22**, 1152-1158, doi:10.1016/j.bbmt.2016.03.001 (2016).

79 Hou, K. *et al.* Alternative Lengthening of Telomeres and Mediated Telomere Synthesis. *Cancers (Basel)* **14**, doi:10.3390/cancers14092194 (2022).
80 Recagni, M., Bidzinska, J., Zaffaroni, N. & Folini, M. The Role of Alternative Lengthening of Telomeres Mechanism in Cancer: Translational and Therapeutic Implications. *Cancers (Basel)* **12**, doi:10.3390/cancers12040949 (2020).
81 Gao, J. & Pickett, H. A. Targeting telomeres: advances in telomere maintenance mechanism-specific cancer therapies. *Nat Rev Cancer* **22**, 515-532, doi:10.1038/s41568-022-00490-1 (2022).
82 De Vitis, M., Berardinelli, F. & Sgura, A. Telomere Length Maintenance in Cancer: At the Crossroad between Telomerase and Alternative Lengthening of Telomeres (ALT). *Int J Mol Sci* **19**, doi:10.3390/ijms19020606 (2018).
83 Hu, Y. *et al.* Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. *Sci Rep* **6**, 32280, doi:10.1038/srep32280 (2016).
84 Mascarenhas, J. *et al.* Randomized, Single-Blind, Multicenter Phase II Study of Two Doses of Imetelstat in Relapsed or Refractory Myelofibrosis. *J Clin Oncol* **39**, 2881-2892, doi:10.1200/jco.20.02864 (2021).
85 Walsh, S. H. *et al.* Telomere length and correlation with histopathogenesis in B-cell leukemias/lymphomas. *Eur J Haematol* **78**, 283-289, doi:10.1111/j.1600-0609.2007.00817.x (2007).
86 Barthel, F. P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet* **49**, 349-357, doi:10.1038/ng.3781 (2017).
87 Grabowski, P. *et al.* Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. *Blood* **105**, 4807-4812, doi:10.1182/blood-2004-11-4394 (2005).
88 Kroupa, M. *et al.* Relationship of telomere length in colorectal cancer patients with cancer phenotype and patient prognosis. *Br J Cancer* **121**, 344-350, doi:10.1038/s41416-019-0525-3 (2019).
89 Hosnijeh, F. S. *et al.* Prediagnostic telomere length and risk of B-cell lymphoma-Results from the EPIC cohort study. *Int J Cancer* **135**, 2910-2917, doi:10.1002/ijc.28934 (2014).
90 Kumar, V., Abbas, A. K. & Aster, J. C. Vol. 10th Edition 1379 (Elsevier, Philadelphia, 2021).
91 Greer, J. P. *et al. Wintrobe's Clinical Hematology*. Vol. 14 (Wolters Kluwer, 2018).

92 Dunn-Walters, D., Townsend, C., Sinclair, E. & Stewart, A. Immunoglobulin gene analysis as a tool for investigating human immune responses. *Immunol Rev* **284**, 132-147, doi:10.1111/imr.12659 (2018).
93 Martin-Subero, J. I. & Oakes, C. C. Charting the dynamic epigenome during B-cell development. *Semin Cancer Biol* **51**, 139-148, doi:10.1016/j.semcancer.2017.08.008 (2018).
94 Young, C. & Brink, R. The unique biology of germinal center B cells. *Immunity* **54**, 1652-1664, doi:10.1016/j.immuni.2021.07.015 (2021).
95 Swerdlow, S. *et al.* 586 (IARC, 2017).
96 Organization, W. H. *International Agency for Research on Cancer, Global Cancer Observatory*, <https://gco.iarc.who.int/today/en/dataviz/bars?mode=population. > (2022).
97 Li, S., Young, K. H. & Medeiros, L. J. Diffuse large B-cell lymphoma. *Pathology* **50**, 74-87, doi:10.1016/j.pathol.2017.09.006 (2018).
98 Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511, doi:10.1038/35000501 (2000).
99 Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* **346**, 1937-1947, doi:10.1056/NEJMoa012914 (2002).
100 Pasqualucci, L. & Dalla-Favera, R. Genetics of diffuse large B-cell lymphoma. *Blood* **131**, 2307-2319, doi:10.1182/blood-2017-11-764332 (2018).
101 Sehn, L. H. & Salles, G. Diffuse Large B-Cell Lymphoma. *N Engl J Med* **384**, 842-858, doi:10.1056/NEJMra2027612 (2021).
102 Gutiérrez-García, G. *et al.* Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood* **117**, 4836-4843, doi:10.1182/blood-2010-12-322362 (2011).
103 Scott, D. W. *et al.* Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. *J Clin Oncol* **33**, 2848-2856, doi:10.1200/jco.2014.60.2383 (2015).
104 Hans, C. P. *et al.* Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103**, 275-282, doi:10.1182/blood-2003-05-1545 (2004).

105 Rosenthal, A. & Younes, A. High grade B-cell lymphoma with rearrangements of MYC and BCL2 and/or BCL6: Double hit and triple hit lymphomas and double expressing lymphoma. *Blood Rev* **31**, 37-42, doi:10.1016/j.blre.2016.09.004 (2017).
106 Horn, H. *et al.* MYC status in concert with BCL2 and BCL6 expression predicts outcome in diffuse large B-cell lymphoma. *Blood* **121**, 2253-2263, doi:10.1182/blood-2012-06-435842 (2013).
107 Savage, K. J. *et al.* MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood* **114**, 3533-3537, doi:10.1182/blood-2009-05-220095 (2009).
108 Grommes, C. & DeAngelis, L. M. Primary CNS Lymphoma. *J Clin Oncol* **35**, 2410-2418, doi:10.1200/jco.2017.72.7602 (2017).
109 Holdhoff, M. *et al.* Challenges in the Treatment of Newly Diagnosed and Recurrent Primary Central Nervous System Lymphoma. *J Natl Compr Canc Netw* **18**, 1571-1578, doi:10.6004/jnccn.2020.7667 (2020).
110 Wright, G. W. *et al.* A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell* **37**, 551-568.e514, doi:10.1016/j.ccell.2020.03.015 (2020).
111 Jiang, Y. & Melnick, A. The epigenetic basis of diffuse large B-cell lymphoma. *Semin Hematol* **52**, 86-96, doi:10.1053/j.seminhematol.2015.01.003 (2015).
112 Rosenthal, A. C., Munoz, J. L. & Villasboas, J. C. Clinical advances in epigenetic therapies for lymphoma. *Clin Epigenetics* **15**, 39, doi:10.1186/s13148-023-01452-6 (2023).
113 Bakhshi, T. J. & Georgel, P. T. Genetic and epigenetic determinants of diffuse large B-cell lymphoma. *Blood Cancer J* **10**, 123, doi:10.1038/s41408-020-00389-w (2020).
114 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500, doi:10.1038/nrg.2016.59 (2016).
115 Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep* **16**, 1439-1453, doi:10.15252/embr.201540749 (2015).
116 Morrison, O. & Thakur, J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. *Int J Mol Sci* **22**, doi:10.3390/ijms22136922 (2021).
117 Michalak, E. M., Burr, M. L., Bannister, A. J. & Dawson, M. A. The roles of DNA, RNA and histone methylation in ageing and

cancer. *Nat Rev Mol Cell Biol* **20**, 573-589, doi:10.1038/s41580-019-0143-1 (2019).
118 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661-678, doi:10.1038/nrg.2016.112 (2016).
119 Liu, H. *et al.* Three-dimensional genome structure and function. *MedComm (2020)* **4**, e326, doi:10.1002/mco2.326 (2023).
120 Skvortsova, K., Stirzaker, C. & Taberlay, P. The DNA methylation landscape in cancer. *Essays Biochem* **63**, 797-811, doi:10.1042/ebc20190037 (2019).
121 Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20**, 590-607, doi:10.1038/s41580-019-0159-6 (2019).
122 Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**, 393-396, doi:10.1126/science.6164095 (1981).
123 Mattei, A. L., Bailly, N. & Meissner, A. DNA methylation: a historical perspective. *Trends Genet* **38**, 676-707, doi:10.1016/j.tig.2022.03.010 (2022).
124 Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365, doi:10.1038/366362a0 (1993).
125 Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926, doi:10.1016/0092-8674(92)90611-f (1992).
126 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257, doi:10.1016/s0092-8674(00)81656-6 (1999).
127 Schermelleh, L. *et al.* Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* **35**, 4301-4312, doi:10.1093/nar/gkm432 (2007).
128 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, doi:10.1126/science.1169786 (2009).
129 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).

130 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).
131 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303, doi:10.1126/science.1210597 (2011).
132 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).
133 Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086, doi:10.1038/nature08924 (2010).
134 Blackledge, N. P. *et al.* CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* **38**, 179-190, doi:10.1016/j.molcel.2010.04.009 (2010).
135 Du, Q., Luu, P. L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7**, 1051-1073, doi:10.2217/epi.15.39 (2015).
136 Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480-492, doi:10.1016/j.cell.2013.03.011 (2013).
137 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-492, doi:10.1038/nrg3230 (2012).
138 Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313, doi:10.1016/j.cell.2006.02.043 (2006).
139 Teschendorff, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* **20**, 440-446, doi:10.1101/gr.103606.109 (2010).
140 Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet* **39**, 232-236, doi:10.1038/ng1950 (2007).
141 Cabezas-Wallscheid, N. *et al.* Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507-522, doi:10.1016/j.stem.2014.07.005 (2014).
142 Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. *Cell Stem Cell* **5**, 442-449, doi:10.1016/j.stem.2009.08.016 (2009).
143 Accomando, W. P., Wiencke, J. K., Houseman, E. A., Nelson, H. H. & Kelsey, K. T. Quantitative reconstruction of leukocyte

subsets using DNA methylation. *Genome Biol* **15**, R50, doi:10.1186/gb-2014-15-3-r50 (2014).
144 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31, doi:10.1186/gb-2014-15-2-r31 (2014).
145 Oakes, C. C. *et al.* DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet* **48**, 253-264, doi:10.1038/ng.3488 (2016).
146 Kulis, M. *et al.* Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet* **47**, 746-756, doi:10.1038/ng.3291 (2015).
147 Dominguez, P. M. *et al.* DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell Rep* **12**, 2086-2098, doi:10.1016/j.celrep.2015.08.036 (2015).
148 Teater, M. *et al.* AICDA drives epigenetic heterogeneity and accelerates germinal center-derived lymphomagenesis. *Nat Commun* **9**, 222, doi:10.1038/s41467-017-02595-w (2018).
149 Ehrlich, M. DNA hypermethylation in disease: mechanisms and clinical relevance. *Epigenetics* **14**, 1141-1163, doi:10.1080/15592294.2019.1638701 (2019).
150 Goyama, S. & Kitamura, T. Epigenetics in normal and malignant hematopoiesis: An overview and update 2017. *Cancer Sci* **108**, 553-562, doi:10.1111/cas.13168 (2017).
151 Gore, A. V. & Weinstein, B. M. DNA methylation in hematopoietic development and disease. *Exp Hematol* **44**, 783-790, doi:10.1016/j.exphem.2016.04.013 (2016).
152 Gadalla, S. M. *et al.* The relationship between DNA methylation and telomere length in dyskeratosis congenita. *Aging Cell* **11**, 24-28, doi:10.1111/j.1474-9726.2011.00755.x (2012).
153 Weidner, C. I. *et al.* DNA methylation in PRDM8 is indicative for dyskeratosis congenita. *Oncotarget* **7**, 10765-10772, doi:10.18632/oncotarget.7458 (2016).
154 Bejaoui, Y. *et al.* DNA methylation signatures in Blood DNA of Hutchinson-Gilford Progeria syndrome. *Aging Cell* **21**, e13555, doi:10.1111/acel.13555 (2022).
155 Guastafierro, T. *et al.* Genome-wide DNA methylation analysis in blood cells from patients with Werner syndrome. *Clin Epigenetics* **9**, 92, doi:10.1186/s13148-017-0389-4 (2017).
156 Maierhofer, A. *et al.* Epigenetic signatures of Werner syndrome occur early in life and are distinct from normal epigenetic aging processes. *Aging Cell* **18**, e12995, doi:10.1111/acel.12995 (2019).
157 Horvath, S. *et al.* Accelerated epigenetic aging in Down syndrome. *Aging Cell* **14**, 491-495, doi:10.1111/acel.12325 (2015).

158 Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494.e415, doi:10.1016/j.cell.2017.09.027 (2017).
159 Martín-Subero, J. I. *et al.* New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood* **113**, 2488-2497, doi:10.1182/blood-2008-04-152900 (2009).
160 Richter, J. *et al.* Array-based DNA methylation profiling of primary lymphomas of the central nervous system. *BMC Cancer* **9**, 455, doi:10.1186/1471-2407-9-455 (2009).
161 Shaknovich, R. *et al.* DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood* **116**, e81-89, doi:10.1182/blood-2010-05-285320 (2010).
162 De, S. *et al.* Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet* **9**, e1003137, doi:10.1371/journal.pgen.1003137 (2013).
163 Pan, H. *et al.* Epigenomic evolution in diffuse large B-cell lymphomas. *Nat Commun* **6**, 6921, doi:10.1038/ncomms7921 (2015).
164 Duran-Ferrer, M. *et al.* The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome. *Nat Cancer* **1**, 1066-1081, doi:10.1038/s43018-020-00131-2 (2020).
165 Asmar, F. *et al.* Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. *Haematologica* **98**, 1912-1920, doi:10.3324/haematol.2013.088740 (2013).
166 Pike, B. L. *et al.* DNA methylation profiles in diffuse large B-cell lymphoma and their relationship to gene expression status. *Leukemia* **22**, 1035-1043, doi:10.1038/leu.2008.18 (2008).
167 Chambwe, N. *et al.* Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood* **123**, 1699-1708, doi:10.1182/blood-2013-07-509885 (2014).
168 Wedge, E. *et al.* Global hypomethylation is an independent prognostic factor in diffuse large B cell lymphoma. *Am J Hematol* **92**, 689-694, doi:10.1002/ajh.24751 (2017).
169 Nakamura, T. *et al.* Genome-wide DNA methylation profiling identifies primary central nervous system lymphoma as a distinct entity different from systemic diffuse large B-cell lymphoma. *Acta Neuropathol* **133**, 321-324, doi:10.1007/s00401-016-1664-8 (2017).
170 O'Riain, C. *et al.* Array-based DNA methylation profiling in follicular lymphoma. *Leukemia* **23**, 1858-1866, doi:10.1038/leu.2009.114 (2009).

171 Youn, A. & Wang, S. The MiAge Calculator: a DNA methylation-based mitotic age calculator of human tissue types. *Epigenetics* **13**, 192-206, doi:10.1080/15592294.2017.1389361 (2018).
172 Yang, Z. *et al.* Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol* **17**, 205, doi:10.1186/s13059-016-1064-3 (2016).
173 Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573-591, doi:10.18632/aging.101414 (2018).
174 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367, doi:10.1016/j.molcel.2012.10.016 (2013).
175 Chen, B. H. *et al.* Leukocyte telomere length, T cell composition and DNA methylation age. *Aging (Albany NY)* **9**, 1983-1995, doi:10.18632/aging.101293 (2017).
176 Degerman, S. *et al.* Maintained memory in aging is associated with young epigenetic age. *Neurobiol Aging* **55**, 167-171, doi:10.1016/j.neurobiolaging.2017.02.009 (2017).
177 Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)* **10**, 1758-1775, doi:10.18632/aging.101508 (2018).
178 Kabacik, S., Horvath, S., Cohen, H. & Raj, K. Epigenetic ageing is distinct from senescence-mediated ageing and is not prevented by telomerase expression. *Aging (Albany NY)* **10**, 2800-2815, doi:10.18632/aging.101588 (2018).
179 Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686, doi:10.1073/pnas.96.15.8681 (1999).
180 Kelly, A. D. *et al.* A CpG island methylator phenotype in acute myeloid leukemia independent of IDH mutations and associated with a favorable outcome. *Leukemia* **31**, 2011-2019, doi:10.1038/leu.2017.12 (2017).
181 Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-522, doi:10.1016/j.ccr.2010.03.017 (2010).
182 Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med* **3**, 75ra25, doi:10.1126/scitranslmed.3001875 (2011).
183 Borssén, M. *et al.* Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* **8**, e65373, doi:10.1371/journal.pone.0065373 (2013).

184 Borssén, M. *et al.* DNA Methylation Adds Prognostic Value to Minimal Residual Disease Status in Pediatric T-Cell Acute Lymphoblastic Leukemia. *Pediatr Blood Cancer* **63**, 1185-1192, doi:10.1002/pbc.25958 (2016).
185 Borssén, M. *et al.* DNA methylation holds prognostic information in relapsed precursor B-cell acute lymphoblastic leukemia. *Clin Epigenetics* **10**, 31, doi:10.1186/s13148-018-0466-3 (2018).
186 Haider, Z. *et al.* DNA methylation and copy number variation profiling of T-cell lymphoblastic leukemia and lymphoma. *Blood Cancer J* **10**, 45, doi:10.1038/s41408-020-0310-9 (2020).
187 World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama* **310**, 2191-2194, doi:10.1001/jama.2013.281053 (2013).
188 Armitage, J. O. Staging non-Hodgkin lymphoma. *CA Cancer J Clin* **55**, 368-376, doi:10.3322/canjclin.55.6.368 (2005).
189 Lister, T. A. *et al.* Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J Clin Oncol* **7**, 1630-1636, doi:10.1200/jco.1989.7.11.1630 (1989).
190 Oken, M. M. *et al.* Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* **5**, 649-655 (1982).
191 A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* **329**, 987-994, doi:10.1056/nejm199309303291402 (1993).
192 Dernstedt, A. *et al.* Regulation of Decay Accelerating Factor Primes Human Germinal Center B Cells for Phagocytosis. *Front Immunol* **11**, 599647, doi:10.3389/fimmu.2020.599647 (2020).
193 Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res* **30**, e47, doi:10.1093/nar/30.10.e47 (2002).
194 Nordfjäll, K., Osterman, P., Melander, O., Nilsson, P. & Roos, G. hTERT (-1327)T/C polymorphism is not associated with age-related telomere attrition in peripheral blood. *Biochem Biophys Res Commun* **358**, 215-218, doi:10.1016/j.bbrc.2007.04.099 (2007).
195 Weisenberger, D. J. *et al.* Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* **33**, 6823-6836, doi:10.1093/nar/gki987 (2005).
196 Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189-196, doi:10.1093/bioinformatics/bts680 (2013).
197 Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA

methylation microarrays. *Bioinformatics* **30**, 1363-1369, doi:10.1093/bioinformatics/btu049 (2014).
198 Triche, T. J., Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res* **41**, e90, doi:10.1093/nar/gkt090 (2013).
199 Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45**, e22, doi:10.1093/nar/gkw967 (2017).
200 Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* **17**, 61, doi:10.1186/s13059-016-0926-z (2016).
201 McClay, J. L. *et al.* High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome Biol* **16**, 291, doi:10.1186/s13059-015-0842-7 (2015).
202 FlowSorted.Blood.EPIC: Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells (2020).
203 Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* **19**, 64, doi:10.1186/s13059-018-1448-7 (2018).
204 de Punder, K., Heim, C., Przesdzing, I., Wadhwa, P. D. & Entringer, S. Characterization in humans of in vitro leucocyte maximal telomerase activity capacity and association with stress. *Philos Trans R Soc Lond B Biol Sci* **373**, doi:10.1098/rstb.2016.0441 (2018).
205 Vindeløv, L. L., Christensen, I. J. & Nissen, N. I. A detergent-trypsin method for the preparation of nuclei for flow cytometric DNA analysis. *Cytometry* **3**, 323-327, doi:10.1002/cyto.990030503 (1983).
206 Patthey, A. *et al.* Combination of aneuploidy and high S-phase fraction indicates increased risk of relapse in stage I endometrioid endometrial carcinoma. *Acta Oncol* **60**, 1218-1224, doi:10.1080/0284186x.2021.1939146 (2021).
207 Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015, doi:10.1126/science.7605428 (1994).
208 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849, doi:10.1093/bioinformatics/btw313 (2016).

209 Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software* **25**, 1 - 18, doi:10.18637/jss.v025.i01 (2008).
210 Database, R. P. *Epigenetic regulation of gene expression (Homo sapiens)*, <https://reactome.org/> (2023-01-20).
211 Lu, A. T. *et al.* DNA methylation-based estimator of telomere length. *Aging (Albany NY)* **11**, 5895-5923, doi:10.18632/aging.102173 (2019).
212 Cypris, O. *et al.* PRDM8 reveals aberrant DNA methylation in aging syndromes and is relevant for hematopoietic and neuronal differentiation. *Clin Epigenetics* **12**, 125, doi:10.1186/s13148-020-00914-5 (2020).
213 Armando, R. G., Mengual Gomez, D. L., Maggio, J., Sanmartin, M. C. & Gomez, D. E. Telomeropathies: Etiology, diagnosis, treatment and follow-up. Ethical and legal considerations. *Clin Genet* **96**, 3-16, doi:10.1111/cge.13526 (2019).
214 Hohenauer, T. & Moore, A. W. The Prdm family: expanding roles in stem cells and development. *Development* **139**, 2267-2282, doi:10.1242/dev.070110 (2012).
215 Rubio Gomez, M. A. & Ibba, M. Aminoacyl-tRNA synthetases. *Rna* **26**, 910-936, doi:10.1261/rna.071720.119 (2020).
216 Lee, Y. *et al.* Epigenome-wide association study of leukocyte telomere length. *Aging (Albany NY)* **11**, 5876-5894, doi:10.18632/aging.102230 (2019).
217 Zhao, Z., Wang, X., Ding, Y., Cao, X. & Zhang, X. SMC4, a novel tumor prognostic marker and potential tumor therapeutic target. *Front Oncol* **13**, 1117642, doi:10.3389/fonc.2023.1117642 (2023).
218 Codd, V. *et al.* Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet* **53**, 1425-1433, doi:10.1038/s41588-021-00944-6 (2021).
219 Weyburne, E. & Bosco, G. Cancer-associated mutations in the condensin II subunit CAPH2 cause genomic instability through telomere dysfunction and anaphase chromosome bridges. *J Cell Physiol* **236**, 3579-3598, doi:10.1002/jcp.30113 (2021).
220 Fernandez, R. J., 3rd & Johnson, F. B. A regulatory loop connecting WNT signaling and telomere capping: possible therapeutic implications for dyskeratosis congenita. *Ann N Y Acad Sci* **1418**, 56-68, doi:10.1111/nyas.13692 (2018).
221 Rim, E. Y., Clevers, H. & Nusse, R. The Wnt Pathway: From Signaling Mechanisms to Synthetic Modulators. *Annu Rev Biochem* **91**, 571-598, doi:10.1146/annurev-biochem-040320-103615 (2022).

222 Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999, doi:10.1016/j.cell.2017.05.016 (2017).
223 Yang, T. B. *et al.* Mutual reinforcement between telomere capping and canonical Wnt signalling in the intestinal stem cell niche. *Nat Commun* **8**, 14766, doi:10.1038/ncomms14766 (2017).
224 Lee, D. D., Komosa, M., Nunes, N. M. & Tabori, U. DNA methylation of the TERT promoter and its impact on human cancer. *Curr Opin Genet Dev* **60**, 17-24, doi:10.1016/j.gde.2020.02.003 (2020).
225 Heng, J. *et al.* Integrated analysis of promoter methylation and expression of telomere related genes in breast cancer. *Oncotarget* **8**, 25442-25454, doi:10.18632/oncotarget.16036 (2017).
226 Peters, T. J. *et al.* Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res* **49**, e109, doi:10.1093/nar/gkab637 (2021).
227 Maierhofer, A. *et al.* Accelerated epigenetic aging in Werner syndrome. *Aging (Albany NY)* **9**, 1143-1152, doi:10.18632/aging.101217 (2017).
228 Weidner, C. I. *et al.* Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol* **15**, R24, doi:10.1186/gb-2014-15-2-r24 (2014).
229 Liu, Y. & Barta, S. K. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am J Hematol* **94**, 604-616, doi:10.1002/ajh.25460 (2019).
230 Kristensen, L. S. *et al.* Aberrant methylation of cell-free circulating DNA in plasma predicts poor outcome in diffuse large B cell lymphoma. *Clin Epigenetics* **8**, 95, doi:10.1186/s13148-016-0261-y (2016).
231 Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390, doi:10.1182/blood-2016-01-643569 (2016).
232 Loeffler, M. *et al.* Genomic and epigenomic co-evolution in follicular lymphomas. *Leukemia* **29**, 456-463, doi:10.1038/leu.2014.209 (2015).
233 Oakes, C. C. & Martin-Subero, J. I. Insight into origins, mechanisms, and utility of DNA methylation in B-cell malignancies. *Blood* **132**, 999-1006, doi:10.1182/blood-2018-02-692970 (2018).
234 Henson, J. D., Neumann, A. A., Yeager, T. R. & Reddel, R. R. Alternative lengthening of telomeres in mammalian cells. *Oncogene* **21**, 598-610, doi:10.1038/sj.onc.1205058 (2002).

235 Henson, J. D. *et al.* The C-Circle Assay for alternative-lengthening-of-telomeres activity. *Methods* **114**, 74-84, doi:10.1016/j.ymeth.2016.08.016 (2017).
236 Zhang, C. *et al.* The Association between Telomere Length and Cancer Prognosis: Evidence from a Meta-Analysis. *PLoS One* **10**, e0133174, doi:10.1371/journal.pone.0133174 (2015).
237 Hiyama, K. *et al.* Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J Immunol* **155**, 3711-3715 (1995).
238 Tsakiri, K. D. *et al.* Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A* **104**, 7552-7557, doi:10.1073/pnas.0701009104 (2007).
239 Zaug, A. J., Crary, S. M., Jesse Fioravanti, M., Campbell, K. & Cech, T. R. Many disease-associated variants of hTERT retain high telomerase enzymatic activity. *Nucleic Acids Res* **41**, 8969-8978, doi:10.1093/nar/gkt653 (2013).
240 Tsang, A. R., Wyatt, H. D., Ting, N. S. & Beattie, T. L. hTERT mutations associated with idiopathic pulmonary fibrosis affect telomerase activity, telomere length, and cell growth by distinct mechanisms. *Aging Cell* **11**, 482-490, doi:10.1111/j.1474-9726.2012.00810.x (2012).