GENETIC MODIFIERS OF TELOMERE MAINTENANCE AND THEIR CONTRIBUTIONS TO PHENOTYPIC VARIATIONS IN TELOMERE BIOLOGY DISORDERS

by

Jialin Xu

B.Sc., Fudan University, 2009

M.Sc., Fudan University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pharmaceutical Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Genetic modifiers of telomere maintenance and their contributions to phenotypic variations in telomere biology disorders.

submitted by Jialin Xu in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

In Pharmaceutical Sciences

Examining Committee:

Dr. Judy Wong
Supervisor

Dr. Andrew Sandford
Supervisory Committee Member

Supervisory Committee Member

Dr. Suzanne Vercauteran
University Examiner

Dr. Louis Lefebvre
University Examiner

Additional Supervisory Committee Members:

Dr. Bruce Carleton
Supervisory Committee Member

Supervisory Committee Member

Abstract

Telomeres, the protective caps at the end of human chromosomes, are shortened during cellular proliferation in normal aging. Telomere biology disorders (TBDs) refer to a spectrum of tissue degenerative disorders caused by accelerated shortening of telomeres secondary to genetic defects in telomere biology genes. Defects in eleven genes involved in telomere length maintenance have been found to cause TBDs. Accelerated telomere shortening leads to premature aging at the cellular level and regenerative defects at the tissue level. TBD-related genetic defects, in concert with intrinsic tissue turn-over rate and various environmental insults that precipitate telomere shortening, determine the aging process of specific tissues and the clinical presentations of TBDs.

The main **objective** of this dissertation is to investigate the genetic factors that contribute to phenotypic variations of TBDs. The thesis is divided into two sections based on the presentations of TBDs in fast- and slow-turnover tissues.

In Chapter 2, using patient-derived cell models and DNA samples, I comprehensively assessed the molecular and cellular phenotypes of X-linked dyskeratosis congenita (X-DC) in female *DKC1* mutation carriers. I demonstrated that successful X chromosome inactivation (XCI) in their blood cells led to normal dyskerin expression and function and thus normal telomere length maintenance. These populations should be free of hematopoietic disease manifestations. In contrast, protection from XCI in tissue compartments other than the hematopoietic system may not be complete, and DC manifestations could be observed in a patient-specific manner depending on the sum total of the environmental and inherited telomere lengths as confounding factors.

Extending from the observation with phenotypic variations in female *DKC1* mutation carriers, I further investigated how incomplete genetic perturbations of telomerase activity may impact clinical presentations of a common disease. In Chapter 3, using a combination of cell and clinical disease models, I showed that functional defects in telomerase catalytic activity, caused by selected genetic polymorphisms in *TERT*, led to suboptimal telomere length maintenance. Rapid progression of chronic obstructive pulmonary diseases is associated with patients' carrying status of the minor allele of rs61748181.

Collectively, my study revealed two genetic modifiers for potential causes of phenotypic variations in TBDs.

Lay Summary

X-linked dyskeratosis congenita (X-DC) is a rare disorder that affects 1 in a million people, whereas chronic obstructive pulmonary disease (COPD) affects more than 1.6 million people in Canada. Despite the different prevalence, patients with both diseases have accelerated telomere shortening. Telomeres are the protective caps found at the end of human chromosomes.

In this dissertation, I comprehensively assessed female family members of X-DC patients carrying rare genetic changes. My research suggested that these female subjects were free of bone marrow failure, the severe presentation of X-DC, due to their normal telomere lengths in blood. I also evaluated the effect of common genetic changes in telomere biology genes on telomere lengths in a cell model and in COPD patients. Rapid disease progression in COPD patients is associated with genetic changes in telomere biology genes. This study will help to build individualized models for the prediction of short telomere-associated disease presentations.

Preface

Chapter 1 is an introduction of telomere biology and the clinical manifestations of telomere biology disorders. A version of this Chapter will be submitted for an invited review article in 2019.

Chapter 2: a version of this material has been previously published in **Xu J**, Khincha PP, Giri N, Alter BP, Savage SA, Wong JMY. Investigation of chromosome X inactivation and clinical phenotypes in female carriers of *DKC1* mutations. *Am J Hematol* 2016; 91(12): 1215-20. All laboratory work was conducted in the Faculty of Pharmaceutical Sciences, UBC. I collected all the *in vitro* data and contributed to most figures in this manuscript, except for Figure 2.2, Figure 2.3, and Table 2.3: I designed the study, performed the experiments, analyzed data and wrote the first draft of the manuscript. The co-first author Dr. Payal Khincha and her supervisor Dr. Sharon Savage designed the study, evaluated patients, and contributed to Figure 2.2 and Table 2.3. They sent clinical samples to RepeatDx laboratory for telomere length measurement to generate Figure 2.3. They also edited the patient recruitment and clinical evaluation content of the manuscript. Dr. Neelam Giri and Dr. Blanch Alter evaluated patients and edited the manuscript. My supervisor Dr. Judy Wong helped me with the study design, data analysis, and manuscript revision. In this study, I used patient materials collected from an Institutional Review Board-approved longitudinal cohort study at the National Cancer Institute (NCI) entitled "Etiologic Investigation of Cancer Susceptibility in Inherited Bone Marrow Failure Syndromes" (www.marrowfailure.cancer.gov, NCI 02–C–0052, ClinicalTrials.gov identifier: NCT0027274). Copyright permission has been obtained from the journal.

The Appendix to Chapter 2 is based upon a previously published research article: **Xu J**, Gu AY, Thumati NR, Wong JMY. Quantification of pseudouridine levels in cellular RNA pools

with a modified HPLC-UV assay. *Genes* 2017; 8(9): 219. I performed 90% of the work in this manuscript. Dr. Naresh Thumati initiated the work by optimizing the HPLC-UV method with the help of Dr. Judy Wong. I optimized the assay, validated the method, conceived the biological application for the method and performed related experiments, analyzed the data and wrote the manuscript. Alice Gu assisted in method validation and manuscript revisions. Dr. Judy Wong assisted in the study design and manuscript revision. Copyright permission has been obtained from the journal.

Chapter 3: part of this material is based upon two unpublished research articles: (1) **Xu, J,** Trudeau, MA, Lee, J, Sin, DD, Sandford, AJ, Wong, JMY. Effects of single nucleotide polymorphisms in telomerase reverse transcriptase gene on progression of chronic obstructive pulmonary disease. (2) **Xu, J.,** Trudeau, MA, Sandford, AJ, Wong, JMY. Evaluations of short telomere risk-associated single nucleotide polymorphisms in telomerase reverse transcriptase gene on telomere length maintenance. Both articles have been submitted for publication and are still under review. The laboratory work was conducted in the Faculty of Pharmaceutical Sciences at UBC and the Centre for Heart Lung Innovation at St Paul's Hospital. I performed 90% of the work in this manuscript. A previous graduate student in the Wong Lab, Matthew Trudeau, initiated the study by running site-directed mutagenesis for two of the eight selected *TERT* SNPs. I designed the rest of the study, performed the experiments, analyzed the data and wrote the manuscript. A previous graduate student in the Sandford Lab, Jee Lee, contributed to the telomere length measurement by qPCR in Figure 3.11. Dr. Sandford and Dr. Judy Wong contributed to the study design, data analysis, and manuscript revisions. Dr. Sin contributed to manuscript revisions. In this study, I used patient materials collected from an Institutional Review Board-approved cohort study at the University of British Columbia (Providence Health

Care Research Ethics Board certificate number H09-02042, ClinicalTrials.gov Identifier: NCT00000568).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aa	amino acid
AD	autosomal dominant
AEC2	type 2 alveolar epithelial cells
ALT	alternative lengthening of telomeres
ANOVA	analysis of variance
AR	autosomal recessive
AS	alternative splicing
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
CB	Cajal body
CBCA	cap binding complex A
CI	confidence interval
CMCT	N-cyclohexyl-N'-β-(4-methylmorpholinium)ethylcarbodiimide p-tosylate
COPD	chronic obstructive pulmonary disease
CR	conserved region (in telomerase RNA)
CST	a complex made up of CTC1, STN1 and TEN1
CTC1	conserved telomere maintenance component 1
CTE	C-terminal extension domain (in telomerase reverse transcriptase)
D-loop	displacement loop
DC	dyskeratosis congenita
DMEM	Dulbecco's Modified Eagle's Medium

DMSO	Dimethyl sulfoxide
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
ETS	E-twenty-six transcription factors
FBS	fetal bovine serum
FEV1	forced expiratory volume in one second
flow FISH	flow cytometry with fluorescence *in situ* hybridization
GABPA	GA-binding protein alpha subunit
GWAS	genome-wide association study
H/ACA	hairpin-hinge (H-box)-hairpin-ACA structure
HPLC	high performance liquid chromatography
HUMARA	human androgen receptor assay
IFD	insertion in fingers domain
iPSC	induced pluripotent stem cells
kb	kilobase
LARP7	La Autoantigen Related Protein 7
LHS	Lung Health Study
LOD	limit of detection
LOQ	limit of quantitation
MAF	minor allele frequency
MDS	myelodysplastic syndromes
mRNA	messenger RNA
MS	mass spectrometry

NAF1	nuclear association factor 1
NAP	nucleotide addition processivity
ncRNA	non-coding RNA
NEXT	nuclear exosome targeting complex
NOP10	nucleolar protein 10
nt	nucleotide(s)
OB	oligonucleotide/oligosaccharide-binding domain
OR	odds ratio
PAGE	polyacrylamide gel electrophoresis
PARN	poly(A)-specific ribonuclease
PAVM	pulmonary arteriovenous malformation
PBMC	peripheral blood mononuclear cell
PDL	population doubling level
PCR	polymerase chain reaction
POT1	protection of telomeres 1
PUS	pseudouridine synthase
RAP	repeat addition processivity
RAP1	repressor/activator protein I
Rb	retinoblastoma protein
RNP	ribonucleoprotein
RPA	replication protein A
rRNA	ribosomal RNA
RE	relative errors

RSD	relative standard deviations
RT	reverse transcriptase domain (in telomerase reverse transcriptase)
RT-PCR	reverse transcription polymerase chain reaction
RTEL1	regulator of telomere elongation helicase 1
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
scaRNA	small Cajal body RNA
SEM	standard error of the mean
SHQ1	snoRNA of the box H/ACA family quantitative accumulation protein
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
ssDNA	single-stranded DNA
STE	stem terminus element domain (in telomerase RNA)
T-loop	telomere loop
TBDs	telomere biology disorders
TBE	template boundary element (in telomerase RNA)
TCAB1	telomerase Cajal body protein I
TEL patch	TPP1 glutamate (E) and leucine (L)-rich patch
TEN	telomerase essential N-terminus domain (in telomerase reverse transcriptase)
TER	telomerase RNA
TERT	telomerase reverse transcriptase
TIN2	TRF2- and TRF1-interacting nuclear protein 2
TPP1	TIN2 and POT1-Interacting Protein

TRAMP	the human TRF4-2/AIR2/MTR4 polyadenylation complex
TRAP	telomeric repeat amplification protocol
TRBD	telomerase RNA binding domain (in telomerase reverse transcriptase)
TRF	telomere restriction fragment analysis
TRF1	telomere repeat binding factor 1
TRF2	telomere repeat binding factor 2
tRNA	transfer RNA
UV	ultraviolet
WT	wild-type
WCL	whole cell lysate
X-DC	X-linked dyskeratosis congenita
XCI	X chromosome inactivation
7 met-G	7-methylguanosine
Ψ	pseudouridine

Acknowledgments

It has been a long journey towards a Ph.D. The research described here is made possible through the help and support from a lot of faculty, staff, and friends at the University of British Columbia.

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Judy Wong for supporting my graduate study and research work at UBC. I am inspired by her enthusiasm for science. I benefited greatly from her extensive knowledge, rigorous scientific training, and her diligence towards work.

I would also like to express my gratitude to my supervisory committee members, Drs. Mary Ensom, Ronald Reid, Bruce Carleton, Andrew Sandford, and Larry Lynd. They guided me throughout the progress and provided insights and advice on my research. Specifically, Dr. Ensom is always supportive and she is an excellent role model for how a female scientist could be both professional and elegant. Dr. Reid introduced me into the field of epigenetics, and I finally applied the knowledge in epigenetics in my dissertation. Dr. Carleton has a great knowledge of pharmacogenomics and epidemiology. I am always inspired by his enthusiasm for applying cutting-edge science to clinical practice. My collaborator Dr. Sandford, with his expertise in the genetics of lung diseases, taught me the laboratory techniques in person and guided me for the statistical analysis parts of my study. I also benefited greatly from the discussion with him and during the writing of the manuscript. I am so grateful to him for joining my committee after Dr. Ensom and Dr. Reid retired. How fortunate I am to have all these great professors as advisors during my Ph.D.!

I am honored to collaborate with Dr. Payal Khincha and Dr. Sharon Savage from NIH. These physician-scientists enrolled into the study the clinical cases and provided insights into the

clinical aspects of my thesis. My research was made possible through various grant supports from NSERC and NIH.

I also want to thank other members of the Wong Lab. Dr. Naresh Thumati and Kyle Hukezalie taught me the techniques and guided me into the field of telomere biology. Fellow graduate students and lab technicians Connor Thompson, Alice Gu, and Sunny Yang provided valuable suggestions and assistance in the lab. I also appreciate administrative support from previous and current program managers Rachel Wu and Shirley Wong. I enjoyed the annual fireside chat with Dr. Thomas Chang for checking on my program progress. The studies in this dissertation could not have been finished without financial supports by Westat, Inc., Natural Sciences and Engineering Research Council of Canada, and Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

I would like to thank all my friends for their help along the way, including those at UBC and in China. It was their mental support and friendship that helped me through the ups and downs of my graduate study.

Last but not least, I would like to express my deepest gratitude to my parents, Linquan Xu and Suping Feng. Ever since I was a little kid, they have encouraged me to pursue my dreams – any of them – no matter if it was to be a musician, a teacher, or a scientist. My growing up years as their only daughter and the youngest child in the extended family cannot be that delightful without the nature and nurture they provided to me. Their unconditional love and support have made my life so beautiful and extraordinary.

To my dear parents

For your love and care through the decades

Chapter 1: Introduction

1.1 Telomeres

1.1.1 Telomere structure

Telomeres are the regions of repetitive nucleotide sequences found at each end of a linear chromosome. Made up of a tandem hexanucleotide repeat $(TTAGGG)_n$, human telomeres measure 5~15 kb in length (1). They consist of a TTAGGG lagging G-strand sequence and its complementary CCCTAA leading C-strand sequence. Telomeres end with a single-stranded (ss) 3' G-strand that is 50 – 400 nucleotides (nt) long (2). Telomeres are further arranged into tightly-packed nucleosomes that have shorter and slightly altered spacing than bulk nuclear chromatins (3). The G-rich overhang invades the duplex telomeric repeats and forms a lariat loop structure (Telomere Loop, T-loop) (4) under electron microscopy (5) and Stochastic Optical Reconstruction Microscopy (6). Thus, a single-stranded displacement loop (D-loop) is created, where the overhang is base-paired to the C-strand (Figure 1.1A).

Shelterin, a six-member protein complex, is bound to telomere sequence specifically (Figure 1.1B) (7). Among the six shelterin subunits, telomere repeat binding factors I and II (TRF1 and TRF2) directly recognize and bind to double-stranded (ds) telomeric region, whereas Protection of Telomeres I (POT1) binds to the 3' G-rich overhang of telomeres to prevent the single-stranded part being recognized by the DNA repair machinery (8). They are interconnected by the remaining three subunits: Repressor/Activator Protein I (RAP1), TIN2 and POT1-Interacting Protein (TPP1), and TRF1- and TRF2-Interacting Nuclear Protein 2 (TIN2). RAP1 and TPP1 are bound to TRF2 and POT1, respectively. Finally, TIN2 stabilizes and binds to TRF1, TRF2, and the TPP1-POT1 complex. It serves as a bridge between the subunits bound to double-stranded and single-stranded DNA (7).

In humans, the $(TTAGGG)_n$ telomeric sequences can also form the stacked Hoogsteen-base-paired G-quadruplex structure *in vitro* (9) and *in vivo* (10). However, the formation of such secondary structure requires the telomeric DNA to be single-stranded, which could mostly happen during DNA replication. Generally, the binding of shelterin masks the G-rich sequences and maintains the integrity and proper folding of the telomeric region (10).

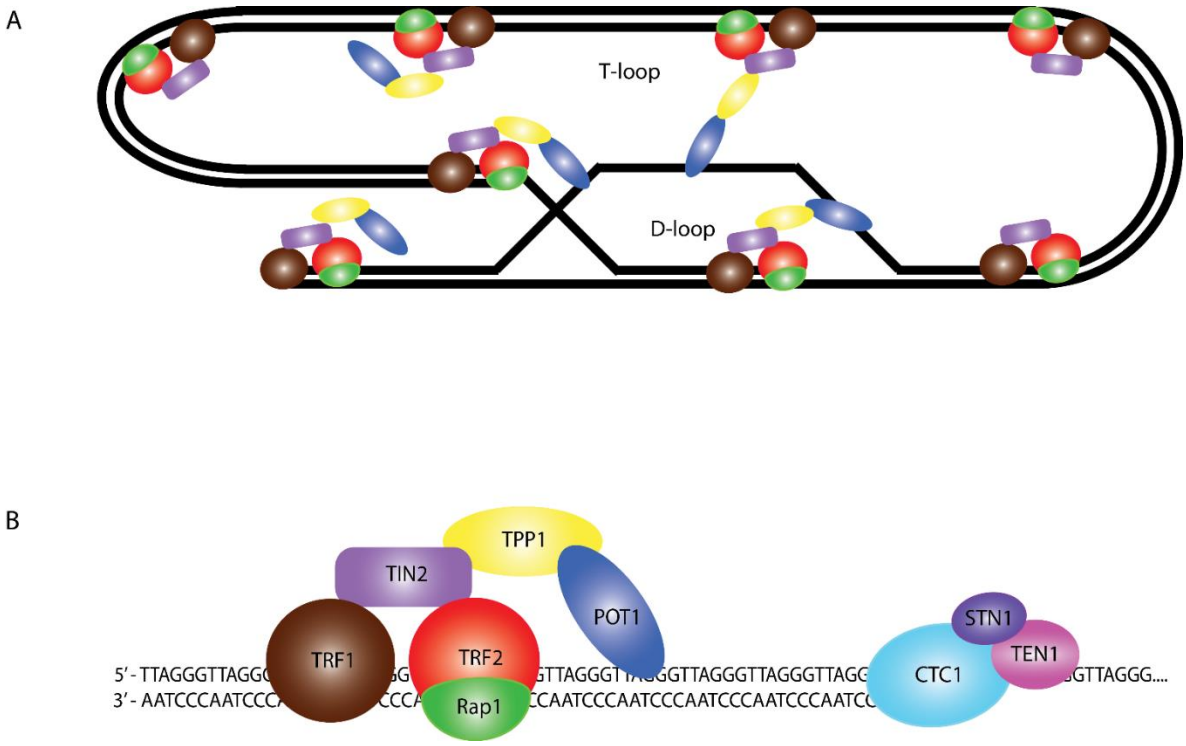

Figure 1.1 Telomere cap structure in humans.

(A) Telomeric loops (T-loops) are formed through the invasion of the G-rich overhang into the duplex telomeric repeats. A single-stranded displacement loop (D-loop) is created as the overhang is base-paired to the C-strand. (B) The six-member protein complex shelterin binds to telomeric DNA sequences specifically. The CST complex at the very end coordinates with shelterin to inhibit telomerase access to telomere and to recruit DNA polymerase α for the fill-in synthesis of the 3' overhang.

1.1.2 Chromosome end problems

Chromosome end problems refer to the end-protection problem and the end-replication problem (11).

When linear DNAs are introduced into eukaryotic cells, they are unstable because the ends of linear DNAs can get recombined with the genome (12). In fact, the introduced linear DNAs are recognized as broken chromosomes and are subject to two types of DNA repair: homologous recombination and non-homologous end joining (13). Shelterin binding and the higher-order chromatin T-loop structure distinguish normal chromosome ends from broken DNA ends and protect them from degradation, erroneous recombination repair, and end-to-end fusions (14).

The replication at the very end of chromosomes involves a few specific problems. First, the T-loops at the telomeres need to be disassembled by the RecQ helicase family members WRN, BLM, and the DNA helicase regulator of telomere elongation helicase 1 (RTEL1) (15-19). RTEL1 is an essential helicase that acts at both telomeric and non-telomeric regions. Through the unwinding of the D-loop structure within the T-loop (Figure 1.1A), RTEL1 facilitates proper replication at the end and prevents the 3’ ssDNA D-loop from being recognized as a strand-invasion intermediate in homologous repair (20, 21). Second, during DNA replication, DNA-dependent DNA polymerases can only synthesize DNA in the 5’ to 3’ direction and are not able to replicate linear chromosomes completely to the very end in the lagging strand, where there is a gap left behind by the 5’ most RNA primer (22). Third, to create the 3’ overhang of the telomere, the leading strand needs to be 5’-resected by the exonucleases Apollo and ExoI before being filled-in by the CST complex (CTC1-STN1-TEN1) (led by CTC1) recruitment of DNA polymerase α (23-26) (Figure 1.1B) (to be introduced in detail in Section1.3.2). Together, these processes contribute to a loss of 50 - 100 nt of the 3' terminal telomeric DNA with each

mammalian cell division (27-29). Telomere length serves as a mitotic clock, counting down the number of cell divisions in each cell lineage before they become senescent.

1.1.3 Short telomeres and cancer susceptibility

Progressive loss of telomeric DNA with repeated cell divisions leads to the activation of a p53 and/or retinoblastoma (Rb)-dependent short telomere checkpoint where cells are driven into senescence (30) or apoptosis (31).

With the inactivation of genome surveillance mechanisms, some cells can evade the short telomere checkpoint. Cells lacking proper tumor suppressor functions keep on dividing until they reach a stage of "crisis", where telomeres are no longer able to form T-loops and are recognized as double-strand DNA breaks. At this point, the formation of dicentric chromosomes through end-to-end fusions can lead to mis-segregation, genomic instability, and chromosomal instability (induced by the chromosomal breakage-fusion-bridge cycle), fueling the transformation process necessary for cancer development (32, 33). Very short telomeres were universally observed in cells undergoing crisis (34) and in epithelial cancer precursor lesions (35). Cells in crisis stage undergo frequent cellular death due to chromosome breakage and mis-segregation (33).

Occasionally, a cell can escape death induced by crisis through the activation of cellular mechanisms to maintain its telomere length. The resultant cell lineage can become immortalized through genomic mutations and rearrangement, especially through gains and amplifications of chromosome 5p15 to increase the copy number of the *TERT* gene region (36) and through point mutations in the *TERT* promoter region that regulates *TERT* expression and hence telomerase activity (37). In 85 – 90% of human tumors, telomerase is reactivated to counterbalance telomere loss (37). Two models have been proposed to explain the observed reactivation of telomerase activity and the following cellular immortalization process. In the classical crisis model, a point

mutation in a single cell is acquired during the crisis stage and leads to reactivation of telomerase for telomere length maintenance and immortalization of the cell lineage (38). In the other two-step model, a point mutation can be acquired at any time independent of telomere length. The upregulated telomerase activity can extend the shortest telomeres but cannot prevent the majority of cells from telomere shortening. Genomic instability caused by critically short telomeres helps to further upregulate telomerase activity and leads to tumorigenesis and immortalization (39). For the 10 – 15% of human tumors that do not express telomerase, alternative lengthening of telomeres (to be introduced in Section 1.3.4) are used to maintain telomere length for further cellular proliferation.

1.2 Telomerase

Telomerase is a ribonucleoprotein (RNP) complex that is able to counterbalance the proliferation-dependent telomere shortening by *de novo* synthesis of telomeric repeats. It is comprised of two core catalytic subunits, telomerase RNA (TER, also known as hTR or TERC) and telomerase reverse transcriptase (TERT, also known as hTERT for human), as well as the box H/ACA protein complex (Figure 1.2A). Expression of both TERT and TER is the minimal requirement for reconstituting telomerase activity *in vitro* (40, 41), whereas the H/ACA and other protein factors are required for regulation of telomerase biogenesis (42), subcellular localization (43), maintenance of TER stability and holoenzyme function (44) *in vivo*. Excluding the subpopulations of both TERT and TER that are not used for the assembly of the telomerase holoenzymes (45), it is estimated that the number of catalytically active telomerase molecules vary between approximately 20 to 250 in an average telomerase-positive cell (i.e., cancer cell or immortalized cell) (45-47). Measurements of TER copy number have substantially different

estimates between around 120 to 70,000 molecules per cell (47, 48), and the most recent measurement yielded an estimate of about 750 to 1,150 molecules per cell (45). While the copy number of TER is in great excess compared to that of TERT, concomitant overexpression of both TERT and TER showed an elevated telomerase activity compared to either TERT or TER overexpression (49). Furthermore, overexpression of either one of them in human cells can promote the assembly of active telomerase RNP assembly and result in increased telomerase activity *in vivo* (45). Therefore, both TERT and TER serve as limiting factors for telomere elongation.

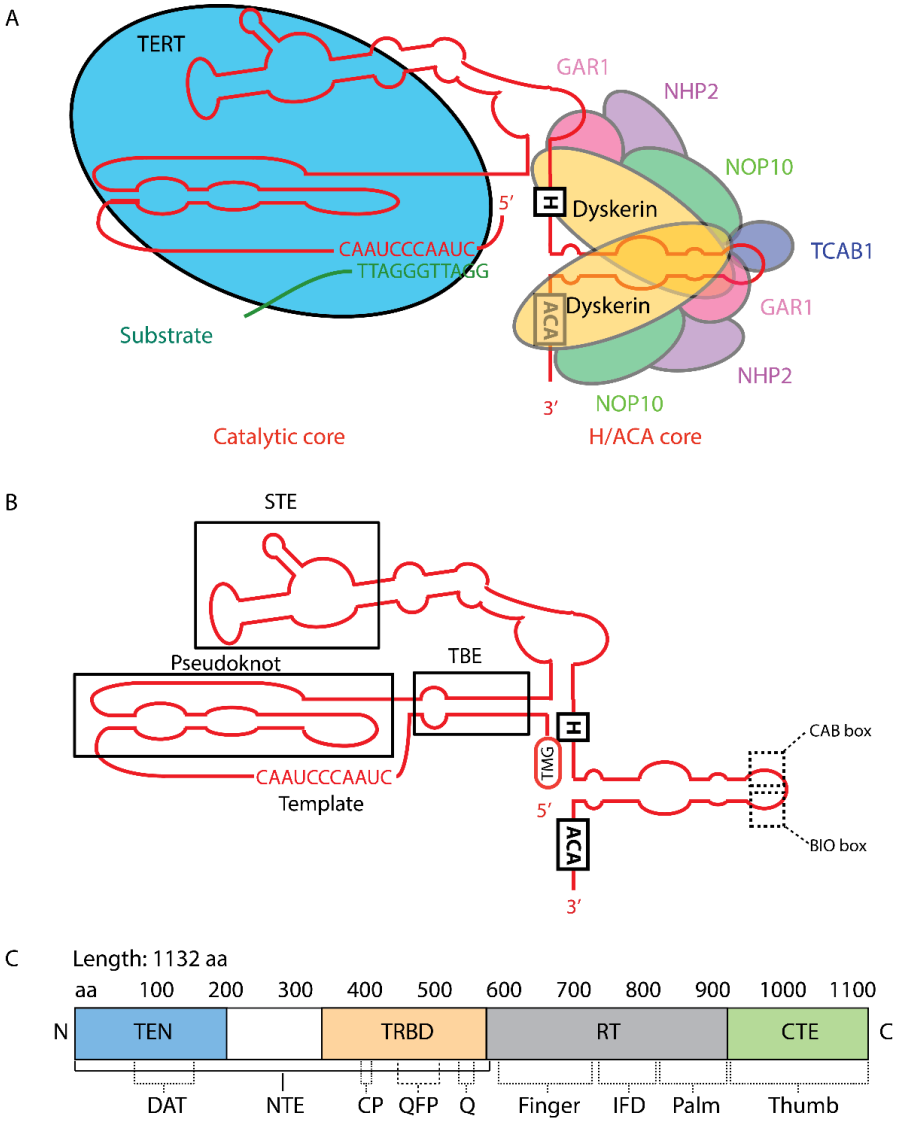

Figure 1.2 Telomerase

(A) The telomerase RNP is made up of telomerase RNA (TER), telomerase reverse transcriptase (TERT), and two copies of H/ACA RNP (dyskerin, NHP2, GAR1, NOP10). TCAB1 brings the holoenzyme complex into Cajal bodies, where telomerase meets the substrate. (B) Secondary structure of telomerase RNA. TMG, the 2,2,7-trimethylguanosine cap structure; TBE, template boundary element; STE, the stem terminus element. (C) Telomerase reverse transcriptase has four evolutionarily conservative functional domains: the long N-terminal extension (NTE) of TERT containing the telomerase essential N-terminal domain (TEN) and the telomerase RNA-binding domain (TRBD), the catalytic reverse transcriptase (RT) domain, and the C-terminal extension (CTE), with the functional motifs illustrated.

1.2.1 Telomerase RNA

The 451-nucleotide (nt) mature TER belongs to the evolutionarily ancient family of box H/ACA small nucleolar RNAs (snoRNAs) / small Cajal body RNAs (scaRNAs) (42). It contains the following structural conserved domains: 1) the TER core containing pseudoknot and template domain; 2) the stem terminus element (STE); and 3) the box H/ACA scaRNA domain (Figure 1.2B) (50).

The 5' half of TER (nt 1 – 210) is made up of the template boundary element (TBE), the template, and a helical region that can fold at one end to form a pseudoknot (51). The TBE domain upstream of the template at the 5' end of TER functions as a physical template boundary and pauses nucleotide addition by TERT at specific residues to maintain the precision of telomeric repeat addition (52, 53). Since the first 17 nt sequence at the 5' end of TER is rich in guanine, it can form the G-quadruplex structure (54). These special structures are unwound by the RNA helicase DHX36 for the proper folding and function of the TBE domain (55-57). The 11 nt-template (5'-CUAACCCUAAC-3') (nt 46 – 56) is complementary to human telomere sequence. It aligns to the telomere and serves as a template for further extension of telomeric repeats by TERT (see Section 1.3.2 for telomerase-dependent telomere extension) (58). The pseudoknot is evolutionarily conserved (59). It facilitates the repeat synthesis (repeat addition processivity, RAP) of telomerase through the interaction with 3' terminus of TERT to improve the stability of TERT (see Section 1.3.1 for telomerase RAP) (51, 58, 60).

The second domain in TER that interacts with TERT is the STE domain (nt 243 – 328). It is also known as conserved region 4/5 domain (CR4-CR5 domain). It is a three-way junction element that interacts with both the template domain of TER and TERT for optimal telomerase activity *in vitro* and *in vivo* (61, 62). The most conserved residues in the STE domain are

localized in the P6.1 hairpin (nt 302 – 314), which is critical for telomerase activity, but not RNA-protein binding (62-64). At the 3' half of TER, there is a conserved H/ACA structure (nt 211 – 237 and nt 334 – 451, see Section 1.2.2 for details) for the maintenance of its stability (42, 65). Two additional motifs, the CAB box (nt 411 - 414), and the 3' end processing signal (nt 409 – 411 and nt 417 – 420) within the TER H/ACA 3' hairpin loop are found to facilitate Cajal body localization and to promote the biogenesis of TER, respectively (43, 50, 66).

It was postulated that the 5' and 3' of TER evolved in distinct pathways: while the 5' half of TER resembles ciliate TER for interaction with telomerase, the 3' H/ACA motif was adopted from other snoRNA/scaRNAs (67). The deduction was further supported by a recent structural biology study showing that the catalytic core and the H/ACA lobe were spatially separated (68). Therefore, little interaction between H/ACA proteins binding to the H/ACA motif and TERT binding to the catalytic core was observed (Figure 1.2A).

1.2.2 The H/ACA proteins

The box H/ACA RNP complex is comprised of one copy of box H/ACA RNA and two copies of a protein complex consisting dyskerin, GAR1, NHP2, and NOP10 (Figure 1.2A). The term box H/ACA RNA refers to a group of non-coding RNAs (snoRNA/scaRNA) with conserved 5'-hairpin-hinge (H box, ANANNA)-hairpin-ACA-3' structure. Each hairpin within the box H/ACA structure contains a pocket for direct interaction with dyskerin (69).

The H/ACA proteins have dual functions. In the RNP formation with TER, they contribute to the stabilization, cellular localization, and biogenesis of the telomerase RNP. A detailed introduction of the biogenesis process can be found in Section 1.2.4. Genetic mutations in either dyskerin, NOP10 or NHP2 result in low levels of TER and lead to premature aging symptoms (see Section 1.4 for details). In addition to TER, box H/ACA proteins can also form RNP with

other box H/ACA snoRNAs or scaRNAs, where the RNP complex serves as a pseudouridine synthase (PUS) to catalyze the isomerization of uridine to pseudouridine (Ψ) in a guide RNA-dependent manner (70). The hairpins in the H/ACA RNAs serve as sequence-specific guides for base-pairing with target RNAs for pseudouridylation, and these targets include rRNA, spliceosomal small nuclear RNA (snRNA), mRNA, and scaRNA (70). The catalytic activity of pseudouridylation is carried out by dyskerin, which shares structural similarities with the bacterial TruB family of PUS (71). A detailed introduction for the function of the H/ACA RNP complex as a PUS can be found in the Appendix (Section A.2).

1.2.3 Telomerase reverse transcriptase

1.2.3.1 Structure of telomerase reverse transcriptase

The human *TERT* gene is located on the short (p) arm of chromosome 5 at position 15.33 spanning around 42 kb. It consists of 16 exons and 15 introns. The encoded TERT protein has 1,132 amino acids (aa) and a molecular weight of approximately 127 kDa. To date, four functional domains have been defined (Figure 1.2C): the long N-terminal extension of TERT containing the telomerase essential N-terminal domain (TEN) and the telomerase RNA-binding domain (TRBD), the catalytic reverse transcriptase (RT) domain, and the C-terminal extension (CTE). As an evolutionarily conserved protein, the identification of highly-conserved residues across species facilitates the study of TERT domain functions.

The TEN domain (1 – 195 aa) forms an anchor site for the binding of telomeric DNA substrate upstream of the telomere 3’ end and the binding of TER (72-74). The conserved residue L14 in ciliate TERT is implicated in the processive telomeric repeat addition of TERT (75). Two residues Y18 and Q169 are located on the surface of the DNA binding groove that was found to be associated with telomeric DNA through crosslinking studies (72, 76-79). A putative primer

grip sequence spanning 5 aa ($_{137}$WGLLL$_{141}$) was also found to be indispensable for telomeric DNA specific substrate binding (74). A "dissociates activities of telomerase" (DAT) motif (68 – 128 aa), as the name suggests, was found to be dispensable for telomerase catalytic activity *in vitro* but not *in vivo* (74, 80). It was later discovered to interact with TPP1 of the shelterin complex to recruit telomerase to telomeres (see Section 1.3.1 for details) (81-83).

A flexible linker region (231 – 324 aa) is located between the TEN and TRBD domains. It is not as conserved as the other regions in TERT. Cells expressing a *TERT* mutant in the linker region did not show growth defect and telomere length maintenance defect at early passages (80).

In the TRBD domain (322 – 594 aa), there exist the CP motif (405 – 415 aa) and the T motif (561 – 581 aa) that can recognize the STE domain in TER, directly bind to TER, maintain the conformation of STE and therefore modulate the catalysis by controlling the rate of template copying during telomere synthesis (74, 76, 84-89). A third motif, the QFP motif (460 – 520 aa) in the TRBD domain, comprises several hydrophobic amino acids and is implicated in the assembly of telomerase holoenzyme (87-89).

The RT domain (595 – 935 aa) is the catalytic center of telomerase (90, 91). It has seven evolutionarily conserved motifs that are common to all reverse transcriptases (1, 2, A, B', C, D, E) (91, 92). The seven motifs can be further subdivided into two groups based on the classical "right-hand" structure of nucleic acid 5'-polymerase: the "finger" (motifs 1, 2 & A) and the "palm" (motifs B', C, D &E) subdomains (90-93). A catalytic triad of conserved aspartates (D712 in motif A, both D868 and D869 in motif C) coordinates the magnesium ions for nucleotide addition by the two metal-ion mechanism (90, 94). A primer-grip sequence ($_{930}$WCGLL$_{934}$) in motif E, as the name suggests, stabilizes telomerase binding to substrate

telomeric DNA and is, therefore, a processivity determinant (74, 95, 96). An insertion-in-fingers domain (IFD) (734 – 802 aa) that is unique for telomerase but not for other RTs interconnects the "finger" and the "palm" subdomains (97). The IFD domain was found to stabilize the RNA-DNA hybrids within TERT and mediate enzyme processivity as well as telomerase recruitment to telomeres (97-99). An additional telomerase-specific motif (motif 3) between motifs 2 and A in the "finger" subdomain coordinates the translocation of TERT such that the template in TER separates and realigns with the substrate telomeric DNA in each repeat synthesis (100).

Compared to the TRBD and RT domains, the CTE domain (936 – 1132 aa) is less conserved across species, indicating a species-specific function. It resembles the "thumb" subdomain in the "right-hand" structure of TERT (101). The CTE domain in TERT has been implicated in enzyme processivity (102), catalytic activity (nucleotide addition processivity, NAP, see Section 1.3.1 for details) (102), nuclear localization (103), and telomeric DNA binding (104).

1.2.3.2 Expression of telomerase reverse transcriptase

In contrast to the ubiquitously transcribed TER, the expression of TERT is tightly regulated at multiple levels, including transcription and mRNA splicing. Consequently, after early fetal development, low levels of telomerase activity are only found in certain types of epithelial cells, progenitor cells, stem cells (especially hematopoietic stem cells) and activated lymphocytes, but is otherwise absent in most human somatic cells (105).

The *TERT* promoter is located within a core region including 330 bp upstream of the translation start site (ATG), exon 1 (219 nt), intron 1 (104 nt) and the first 37 bp of exon 2 in *TERT* (106). Numerous transcription activators and repressors that interact with this core region have been identified (107). Recurrent somatic mutations in the *TERT* promoter (chr5 1,295,228 C>T and 1,295,250 C>T, corresponding to the positions -124 and -146 bp upstream of the TERT

translation start site) have been reported in different types of human cancers, including melanoma (108, 109), glioma (110, 111), thyroid cancer (112, 113), urothelial carcinoma (114, 115), and bladder cancer (116) (Bell *et al.* provided a comprehensive review on the prevalence of *TERT* promoter mutations in different types of cancers in (117)). It remains unknown why different types of tumors contain exclusively these two specific mutations in the *TERT* promoter. However, the mechanism behind the mutations has been studied. Both mutations generate an 11-bp consensus sequence (5'-CCCCTTCCGGG-3') within the *TERT* promoter region for E-twenty-six (ETS) transcription factors, which share a common recognition motif GGA(A/T) (108, 109). Specifically, the newly formed consensus sequence can recruit the multimeric GA-binding protein alpha subunit (GABPA) (118, 119). Circularized chromosome conformation capture analyses demonstrated that GABPA can initiate the intra-chromosomal long-range interaction between the *TERT* promoter and the chr5: 1,556,087 – 1,558,758 region through its heterodimerization, thereby enhance the enrichment of histone marks for active chromatin and recruit RNA polymerase II to drive the transcription of TERT (120). Luciferase reporter gene assays were carried out in various types of cancer cell lines and confirmed that cells with the *TERT* promoter mutations had two to four-fold increased transcriptional activity (108, 116).

Alternative splicing is one of the major mechanisms for post-transcriptional regulation of *TERT* gene expression. To date, twenty-three alternatively spliced variants (ASVs) of TERT have been identified (121, 122). Except for the full-length TERT transcript that expresses all 16 exons, none of the identified ASVs harbor telomerase catalytic activity for telomere elongation (121). Instead, emerging evidence has revealed that some TERT ASVs have functions beyond telomere length maintenance (see Section 1.2.5). For instance, Killedar *et al.* report that the minor (A) allele of a TERT single nucleotide polymorphism (SNP) rs10069690, located in intron

4 of *TERT*, can create an additional splice donor site and lead to the expression of an ASV INS1b transcript (122). The retention of an additional 480 bp from intron 4 introduces a premature stop codon. Therefore, the INS1b transcript encodes a truncated form of TERT, with the TEN and TRBD domains intact, but no RT or CTE domains (122). The functional consequence for the expression of this SNP (and this AS transcript) will be discussed in Chapter 3.

1.2.4 Biogenesis of telomerase ribonucleoproteins

The biogenesis of telomerase RNP can be separated into two parts: maturation of TER and formation of H/ACA RNP.

The snoRNA of the box H/ACA family quantitative accumulation protein (SHQ1) is an essential assembly factor for the H/ACA RNP. In the cytoplasm, the chaperone SHQ1 binds to the pseudouridine synthase, dyskerin. SHQ1 functions as an RNA mimic to prevent dyskerin from nonspecific RNA binding before the association with other H/ACA RNP core proteins and a snoRNA (123, 124). Next, dyskerin binds to two other proteins NOP10 and NHP2 sequentially to form the core heterotrimer in the cytoplasm. Both reptin and pontin belong to the ATPases Associated with various cellular Activities (AAA+) helicase family. They dissociate the assembly factor SHQ1 from dyskerin and promote the recruitment of an assembly chaperone nuclear association factor I (NAF1) to the heterotrimer (125). NAF1 stabilizes the newly-formed heterotrimer and brings the complex into the nucleus, where H/ACA RNA precursors (including the TER precursor) are transcribed by RNA polymerase II (65).

The nascent 5' terminus of TER is modified with a 7-methyl guanosine (m^7G) cap in a para-transcription manner. After 5' capping of TER, the cap binding complex A (CBCA) binds to the m^7G cap and recruits the nuclear exosome targeting (NEXT) complex to terminate transcription (126). The human TRF4-2/AIR2/MTR4 polyadenylation (TRAMP) complex is also recruited to

the 3’ of TER upon the transcription termination signal to add short oligo(A) tails that can promote the recognition by exonuclease and degradation by exosome (126, 127). At the same time, H/ACA RNP binds to the newly transcribed TER that is essential for the accumulation and 3’ processing of TER (42). The binding of dyskerin prevents TER from further cleavage (128). Two copies of the dyskerin complex protein tetrad are bound to the H/ACA domain of TER before the assembly factor NAF1 being exchanged by the mature RNP component GAR1 for RNP remodeling. Finally, mature mammalian H/ACA snoRNA (including TER) are deadenylated by poly(A) specific ribonuclease (PARN) (129). The oligo(A) tails of nascent TER precursor are trimmed by PARN to promote their maturation (126, 127, 130). Inhibition of TRAMP components such as TRF4-2 can rescue TER level in cell lines derived from PARN-mutant patients (131). The equilibrium between CBCA-NEXT-TRAMP initiated degradation and PARN-promoted maturation determine the final TER level *in vivo* (126, 127).

The mature H/ACA RNP is recruited to enter Cajal bodies (CBs) through the binding of TCAB1 (telomerase Cajal body protein 1) to the CAB box in the mature TER. TCAB1 also facilitates correct folding of the STE domain in TER for its binding to TRBD domain in active TERT (132). Originally described as nucleolus accessory bodies, CBs are the spherical sub-organelles found near the nucleolus of neurons and some proliferative cells, such as embryonic cells and tumor cells. They are considered a processing center for RNPs involved in splicing, ribosome biogenesis, and telomere maintenance. Within the CB, the m^7G cap at the 5’ terminus of TER is hypermethylated into 2,2,7-trimethylguanosine (43).

Following TERT peptide synthesis by the ribosome, the chaperones p23 and heat shock protein 90 stabilizes TERT and maintains it in a conformation that is compatible with the assembly (133, 134). The telomerase complex is enzymatically active once TERT is loaded onto

the H/ACA RNP (135). However, to be functionally active in telomere maintenance, a few processes are needed for the correct subcellular localization of telomerase. These processes include: 1) post-translational phosphorylation of TERT by the protein kinases Akt (136) and Src (137) that facilitates nuclear import and export of TERT and was demonstrated in non-transformed human embryonic kidney 293 (HEK293) cell line and human breast cancer cell line MCF, respectively; 2) the binding of interaction protein 14-3-3 to unmask a nuclear export signal motif in TERT (970 – 981 aa) in HEK293 and human high grade serous ovarian carcinoma cell line A2780 (103); as well as 3) the binding of serine/arginine-protein phosphatase PP2A to TERT for the trafficking between nucleus and cytoplasm and for the regulation of TERT activity in human cervical carcinoma cell line HeLa (138).

Early evidence in *Tetrahymena* showed that the protein p65 in La Autoantigen Related Protein 7 (LARP7) family recognizes TER sequence and stabilizes the RNA to promote TERT assembly (139). Although it remains elusive how LARP7 affects telomerase assembly in human, functional studies in fission yeast have revealed that Lar7 (ortholog of human LARP7) also utilizes its evolutionarily conserved RNA-recognition domain in TER binding and stabilization (140, 141). The discovery of short telomeres and impaired telomerase activity in patients with LARP7 deficiency also indicates the potentially similar role of LARP7 in telomerase assembly in humans (142).

1.2.5 Non-canonical functions of telomerase

In addition to telomere maintenance, telomerase, especially TERT, is also found to have telomere-independent non-canonical functions. These extra-telomeric functions, identified mostly in tumor cells, include regulation of DNA damage responses (143-147), promotion of tumor formation in mouse models (148), regulation of gene expression (149-152), presentation

of RNA-dependent polymerase activity (153), facilitation of rDNA transcription and cell replication (154), survival or cell cycle progression under stress and mitogen-depleted conditions (146, 148, 155-157), protection of mitochondrial integrity under oxidative stress (158, 159), as well as promotion of stem cell self-renewal, survival and replication (149, 160, 161). These activities are all connected to promote cellular fitness during proliferation. As such, the restoration of stable telomerase activity is not only essential for tumor cells to maintain telomere length, but also advantageous against cellular stresses arising from the increasing metabolic demands due to accelerated proliferation. Based on the cells' needs for telomere maintenance and the benefits they gain from the constitutive expression of telomerase, it is not surprising that telomerase is reactivated in 85~90% of all human tumors through transcriptional activation of TERT, despite the existence of an alternate pathway for telomere length maintenance (162, 163).

1.3 Telomere Length Maintenance

1.3.1 Canonical telomerase activity and processivity

As an RNA-dependent DNA polymerase, telomerase has the nucleotide addition (type I) form of processivity (NAP) that is shared by all polymerases. NAP allows telomerase to add each nucleotide within the 6-nt repeats continuously. In addition to NAP, telomerase also has the unique repeat addition (type II) form of processivity (RAP) that allows it to add multiple rounds of DNA repeats after a single primer-binding step. Repeat addition by telomerase includes the following processes: 1) alignment of the 3’ ends of telomeric DNA and the five nucleotides (nt 52 – 56, 3’-CAAUC-5’) in the template recognition region of TER; 2) nucleotide addition to the telomere ends by copying the RNA template to its boundary (nt 45 – 51, 3’-CCAAUC-5’); 3) translocation of the RNA template and/or DNA template within the active site to facilitate the

realignment of the 3’ end of newly synthesized telomeric DNA and 3’ boundary of the RNA template (11). A strong pause is generated after every 6-nt addition by the translocation process, which serves as a rate-limiting step for telomere synthesis. As a result, the assays that measure telomerase activity present a ladder pattern of primer extension products (to be introduced in Section 3.3.5). Cells expressing telomerase with reduced RAP activity but intact NAP activity have been found to have deficient telomere maintenance (96). Patients carrying mutations with defective RAP activity were also shown to suffer from telomere biology disorders (164).

1.3.2 Telomere elongation by telomerase

Telomeres are synthesized during S phase of the cell cycle. The active telomerases are recruited to telomeres by TPP1. Specifically, a 7 aa conservative cluster (termed the TEL patch, TPP1 glutamate (E) and leucine (L)-rich patch) on the surface of OB (oligonucleotide/oligosaccharide-binding) domain in TPP1 can bind to the TEN domain of TERT (specifically, residues K78 and R132) (81-83, 165). The recruitment is tethered by TIN2, the shelterin complex that interconnects TRF1 and TRF2 (166, 167). TPP1 also binds tightly to the telomeric ssDNA overhang-binding protein POT1. The 3’ end of chromosomal DNA is buried into one of the OB domains in POT1 (168). In corroboration of its important regulatory role, depletion of the OB folds in POT1 resulted in significant telomere elongation (169). Together, the POT1-TPP1 heterodimer functions to: 1) repress an ataxia telangiectasia and Rad3-related protein (ATR)-dependent DNA damage response at telomeres (ATR is a type of serine/threonine-protein kinases that responds to single-strand breaks) (8); 2) negatively regulate the accessibility of telomere to telomerase (168, 170); and 3) stimulate the translocation efficiency of repeat addition by telomerase at telomere ends (170, 171).

In addition to the POT1-TPP1 heterodimer, the aforementioned trimeric "CST" complex, consisting of CTC1 (conserved telomere maintenance component 1), STN1 (suppressor of cdc thirteen homolog) and TEN1 (telomeric pathways with Stn1), is also implicated in telomere elongation. First, the complex promotes DNA replication restart after replication fork stalling, which happens frequently in telomeric DNA. In order to prevent G-strand overextension, after the successive promotion of telomere elongation by telomerase via the POT1-TPP1 complex during S phase, the CST complex is recruited by TPP1 to suppress telomerase access to telomeres and inhibit both telomerase and the telomerase-stimulator TPP1 (172). Following telomere synthesis at the G-strand, resection of the leading strand is mediated by the nucleases Apollo and ExoI to form the 3' overhang (24). In late S/G2 phase, the trimeric complex promotes the C-strand fill-in synthesis by DNA polymerase α primase so as to coordinate the G- and C-strand length in the newly synthesized telomere (173). By limiting the telomerase-mediated elongation to a single binding and extension event through the CST complex, approximately 60 nucleotides are added to a single chromosome end during each cycle of replication (45, 174).

1.3.3 Telomere length regulation

Telomere length homeostasis is tightly regulated. Telomerase activity is the major positive regulation mechanism of telomere length. In an individual cell, it is the few shortest telomeres, instead of the average telomere length, that determine the cellular fate in terms of senescence, viability and chromosome stability (175-177). During early S phase, there should be 92 chromosomes ends from 23 pairs of chromosomes present in a normal diploid human somatic cell. However, telomere elongation is stochastic and only a small subset of telomeres are co-localized with telomerase during S phase (178). To prevent cellular senescence, telomerase needs

to find the chromosome ends with the shortest telomeres. A few models have been proposed to explain the maintenance of telomere length homeostasis.

In the protein-counting model, the telomere-bound proteins, especially the shelterin complex, exhibit additive negative effects such that longer telomeres have a stronger repressive effect upon telomerase binding to the 3' end of telomeres and *vice versa*. As a result, the short telomeres are preferentially elongated due to their weak inhibitory effect (179). However, the major deficit of the model is the failure to explain how the additive inhibitory effect of telomere-bound proteins cumulates and integrates over the many kilobases of telomere sequences.

A few modifications to the protein counting model have been proposed. For example, in a POT1-TRF1 transduction model, the amount of TRF1 is positively correlated with telomere length. It was inferred that telomere length information is transduced through the interaction of TRF1 with POT1, which can block the access of telomerase to telomeric ends (4, 180). Similarly, the transduced cumulative negative signals from RIF1 and RIF2, both of which interact with another shelterin component RAP1, were considered to drive conformational changes at telomeric ends and therefore a higher density of these proteins at the telomere correlates with a higher chance of switching them from an "extendible" to a "non-extendible" status (181). In these protein counting models, conformational changes of telomeric ends are driven by the accumulation or loss of telomere binding proteins to control the accessibility of telomerase. These models predict that telomere binding proteins are evolved to coordinate with telomerase for telomere lengthening, and their regulation of telomere length should be dependent on telomerase. However, genetic modulations on telomere lengthening were found to be effective in telomerase-depleted yeast cells, suggesting that telomerase regulation cannot explain telomere length homeostasis completely (182).

In 2016, a novel replication fork model was proposed by Dr. Carol Greider (183). In this model, the telomerase complex moves with the replication fork. The repressive effects exerted by the telomere-bound proteins increase the probability of telomerase falling off the fork before they reach chromosome ends. Consequently, short telomeres are preferentially elongated and long telomeres exert negative regulation effects on telomere lengths (183). The RAP1-interacting and telomere-binding protein RIF1 is indeed an evolutionarily conserved protein for negative regulation of telomeric replication origins (184). The removal of RIF1 initiates origin firing during late-S phase through promoting protein phosphatase 1-dependent dephosphorylation of replication initiation factors (184, 185). The replication fork model explains the experimental observation that the regulation of origin firing could affect telomere length. Furthermore, the essential components of lagging strand synthesis – DNA primase and DNA polymerases α and δ – are also required for telomerase-mediated telomere addition (186). The strength of this new model is that it explains the role of origin firing and lagging strand synthesis well. However, more mechanistic and experimental work is needed to support and polish this model.

1.3.4 Alternative lengthening of telomeres

As discussed in Section 1.1.2, telomere length maintenance is essential for cellular immortalization. To achieve immortalization, telomerase reactivation is found in 85 – 90% of human tumors (37). In the remaining cases, telomere elongation was achieved by alternative lengthening of telomeres (ALT) (187). ALT-positive cells are characterized by copious extrachromosomal telomeric DNA, heterogeneous telomeric sequences and lengths, and ALT-associated promyelocytic leukemia bodies (APBs) that contain telomeric DNA and shelterin proteins (188). Based on the abundance of homologous recombination proteins in APBs and the aforementioned characteristics of ALT cell lines, it has been proposed that the ALT mechanism

is mediated by homology-directed strand exchange and telomeric DNA synthesis (188). The reason why most human tumors utilize telomerase reactivation rather than ALT for telomere length maintenance is not fully understood but could be partially explained by the non-canonical functions of telomerase.

1.4 Telomere biology disorders

1.4.1 Definition

At the cellular level, short telomeres trigger cells to enter replicative senescence. At the tissue level, the premature induction of senescence, caused by short telomeres, would induce tissue failure and accelerated aging phenotypes. Short telomeres serve as the molecular marker for telomere biology disorders (TBDs). TBDs are comprised of a spectrum of tissue degenerative disorders caused by genetic defects in telomere biology genes (189).

1.4.2 Symptoms and clinical manifestations

Telomere lengths vary substantially in the human population and show an overall trend of telomere shortening as a function of normal aging. Therefore, telomere lengths are usually discussed with respect to an individual's age to determine whether or not they are in the normal range (190). At the severe end of the spectrum, TBDs can present as rare, premature bone marrow failure syndrome, dyskeratosis congenita (DC), where the patients exhibit exceedingly short telomeres (less than the first percentile in the population). Since telomere shortening is a function of cell division, the highly proliferative tissues, such as the skin, bone marrow, and epithelial cells, are highly sensitive to accelerated shortening of telomeres. Therefore, early disease manifestations of severe TBDs are usually most evident in these tissues. At the other end of the spectrum, the slow turnover tissues, including the lung, liver, and bone, are also affected

by short telomeres. Diseases in these tissues may have a relatively late onset compared to the symptoms in fast turnover tissues. Several studies have investigated whether common genetic variations are also associated with short telomeres and whether they predispose to chronic diseases in slow turnover tissues. There is no definite boundary between the two ends of the spectrum (191). In most cases, the severity of symptoms is determined by telomere length, as a result of inheritance, telomere synthesis and telomere attrition under normal aging and pathological conditions (191). For example, among 28 patients with DC that have the same mutation in *DKC1* (c.1058 C>T, p.Ala353Val), 8 patients had disease onset at 15 years of age or later and their disease presentations include the mild forms of DC such as skin pigmentation, greying hair, but no aplastic anemia. However, another 4 patients from the same cohort had the most severe form of DC – Hoyeraal-Hreidersson syndrome (to be introduced in Section 1.4.4.2) – and these four patients presented during infancy and early childhood with bone marrow failure, immunodeficiency and delayed development (192).

As discussed in depth in Section 1.1.2, telomere shortening is essential for cell transformation and very short telomeres are found in precancerous lesions (35). Therefore, TBD patients are at risk of developing cancers of hematopoietic and epithelial origin – i.e., cancers in high turnover tissues (193). In TBDs, telomerase deficiency serves as one of the pathological pathways that trigger the development of cancer (194).

The spectrum of TBD symptoms is summarized in Table 1.1 (189, 195).

Table 1.1 Clinical manifestations of TBDs in different tissues

Adapted and expanded from Armanios M & Blackburn EH (195).

Turnover Rate	Tissue type	Disease manifestations in people with TBD
Fast	Hair	Premature hair greying
	Skin	Reticular skin pigmentation Oral leukoplakia
	Nail	Nail dystrophy
	Bone marrow	Bone marrow failure (aplastic anemia)
	Immune system	Immunodeficiency Opportunistic infections
	Intestinal epithelium	Enterocolitis
Slow	Lung	Idiopathic pulmonary fibrosis Emphysema Pulmonary arteriovenous malformations
	Liver	Liver cirrhosis
	Bone	Osteopenia and poor bone healing Osteoporosis
High	Cancers of epithelial or hematopoietic origins	

1.4.3 Genetic mechanisms of TBDs

By definition, TBD is a collection of genetic disorders with molecular defects in telomere maintenance and where the disease severities are determined by the attrition rate of telomere lengths. To date, eleven genes involved in telomere biology have been found to cause TBD. Table 1.2 is a summary of these genes and their molecular epidemiology profiles in DC, the most well-characterized TBD (196). These 11 genes could be divided by inheritance patterns into three forms: X-linked form (*DKC1* gene), autosomal dominant forms (*TERC, TERT, TINF2* (encoding TIN2), *RTEL1*, and *ACD* (encoding TPP1)) and autosomal recessive forms (*NOP10, NHP2, WRAP53* (encoding TCAB1), *RTEL1*, *TERT*, *ACD*, *CTC1*, and *PARN*) (197). Not only germline inheritance but also *de novo* mutations have been reported. Mutations in telomere biology genes *POT1* (198), *DCLRE1B* (encoding the exonuclease Apollo) (199), *SHQ1* (200), *NAF1* (201) and *STN1* (202) may also cause other TBD such as Coats plus syndromes, Hoyeraal-

Hreidersson syndrome or idiopathic pulmonary fibrosis (IPF), that share similar phenotype with DC. However, the genetic cause in around 40% of known DC probands remains unidentified (197).

In addition to genetic mutations and variations in telomere biology genes, telomere length itself is a unique inheritable trait. Although there is no consensus among researchers regarding the paternal/maternal inheritance pattern for telomeres, based on large-population studies in twins, siblings and multi-generational families, the heritability of telomere length is estimated to be 0.32 – 0.84 (203). In families affected by TBDs, progressive telomere shortening in each successive generation causes disease anticipation, i.e., more severe disease and earlier disease onset in the later generations (164, 204-206). Genotypically normal offspring of parents with mutations in telomere biology genes were also found to have stable but short telomeres (164, 206, 207). It remains unclear whether short telomeres alone without any mutations in telomere biology genes would induce the disease manifestation. However, individuals from TBD families with short telomeres but without known mutations have been shown to exhibit pulmonary malfunctions, including emphysema (164).

Table 1.2 Genes involved in TBDs with their molecular epidemiology profile in DC

Adapted from Wegman-Ostrosky T & Savage SA (196), Dokal J, *et al.* (208) & Bertuch AA (197).

Gene	Gene function	Cytogenetic band	Approx. % of patients in DC	Inheritance pattern	Association with other TBDs
ACD (TPP1)	Shelterin	16q22.1	<1%	AR	BMF, DC, HHS
TINF2 (TIN2)		1q11.2	11-20%	AD	DC, HHS, IPF, RS
POT1[*]		7q31.33	Unknown	Unknown	CP
DKC1	Telomerase structure	Xq28	>40%	XLR	DC, HHS, IPF
NHP2		5q35.3	<1%	AR	DC
NOP10		15q14	<1%	AR	DC
NAF1[$]		4q32.2	Unknown	Unknown	IPF
TERT	Telomerase	5q15.33	5%	AD, AR	BMF, DC, IPF, HHS
TERC		3q26.2	5%	AD	BMF, DC, IPF
CTC1	Telomere capping	17p13.1	<1%	AR	CP, DC
STN1[*]		10q24.33	Unknown	Unknown	CP
DCLRE1B		1p13.2	Unknown	Unknown	HHS
SHQ1[*]	Biogenesis	3p13	Unknown	Unknown	CP
PARN		16p13.12	<1%	AR, AD	BMF, DC, HHS, IPF
WRAP53 (TCAB1)		17p13.1	<1%	AR	DC
RTEL1	Telomere stability and helicase	20q13.33	5-10%	AD, AR	HHS

AR, autosomal recessive; AD, autosomal dominant; XLR, X-linked recessive.
BMF, bone marrow failure (including aplastic anemia and myelodysplastic syndrome); CP, Coats plus syndrome; DC, dyskeratosis congenita; HHS, Hoyeraal-Hreidarsson syndrome; IPF, idiopathic pulmonary fibrosis; RS, Revesz syndrome.
[$] *NAF1* AD pathogenic variants have been reported in pulmonary fibrosis and associated with telomere length and levels of telomerase RNA which suggests *NAF1* is also a DC gene.
[*] There have been emerging clinical cases where patients with mutations in *POT1*, *STN1*, and *SHQ1* are suffering from Coats plus syndromes.

1.4.4 TBDs in fast-turnover tissues – dyskeratosis congenita

1.4.4.1 Diagnosis

The first discovered TBD – DC – presents as an inherited bone marrow failure syndrome. The classical mucocutaneous triad of DC includes nail dystrophy, reticular skin pigmentation, and oral leukoplakia (209). Leukocyte telomere lengths less than the first percentile for age, are considered as a diagnostic criterion of DC in the presence of other clinical manifestations (191, 209). As discussed in Section 1.4.2, DC also presents as emphysema, IPF, liver fibrosis and cirrhosis at a lower penetrance (195). In some DC cases, even if the patients survived life-threatening bone marrow failure and immunodeficiency in childhood, they are at a higher risk of developing cancers. A combination of clinical assessment and candidate gene/whole exome sequencing is applied for the diagnosis of DC.

1.4.4.2 Differential diagnosis

Hoyeraal-Hreidersson syndrome and Revesz syndrome are both considered severe variants of DC (189). Hoyeraal-Hreidersson syndrome is characterized by cerebellar hypoplasia and immunodeficiency, whereas Revesz syndrome is characterized by bilateral exudative retinopathy, in addition to their shared symptoms of DC.

Patients with Coats plus syndrome have characteristic bilateral retinopathy and intracranial calcification (189). Most patients with Coats plus syndrome have mutations in *CTC1*, yet not every one of them has telomere length less than the first percentile for age (210). Due to the shared disease etiology in telomere maintenance defects, Coats plus patients may present symptoms of DC (210, 211).

1.4.4.3 Phenotypic variations in X-DC

The X-linked recessive form of dyskeratosis congenita (X-DC), caused by mutations in *DKC1*, accounts for over 40% of all DC cases (Table 1.2).

X chromosome inactivation (XCI) is the gene dosage compensation process by which one of the two copies of the X chromosome in females is inactivated. Heterozygous female carriers of *DKC1* mutant alleles are typically protected from X-DC clinical manifestations through skewed XCI. After gastrulation, cells expressing the *DKC1* wild-type (WT) allele have a survival advantage over their counterparts expressing the mutant allele (212-214). This growth selection likely results in the depletion of cells expressing the mutant dyskerin allele and thus may explain the lack of DC phenotypes in female carriers. However, DC-like premature aging symptoms, such as early hair greying and delayed wound healing have been reported in a small subset of female *DKC1* mutation carriers, the biological mechanism of which is not clearly understood (213, 215).

1.4.5 TBDs in the slow-turnover tissues – pulmonary diseases

1.4.5.1 Idiopathic pulmonary fibrosis

Pulmonary fibrosis is featured by the scarring and stiffness of air sacs in the lung (alveoli), leading to reduced oxygen absorption into the bloodstream. It occurs most frequently in the elderly. When the cause of pulmonary fibrosis is unknown – which happens in most cases – the disease is termed idiopathic pulmonary fibrosis (IPF). Despite being a rare, chronic, progressive, and irreversible life-threatening disease, IPF is the most common form of interstitial lung disease. It is estimated that in every 100,000 Canadians, 41.8 will develop IPF in their lifetime (216).

Among all slow turnover tissues, the lung is the most sensitive to short telomeres. Pulmonary diseases are present in up to 20% of DC patients (209). Mutations in telomere

biology genes are found in 20 – 25% of patients with family history of IPF and in 5 – 10% of patients with sporadic IPF (217-222). Patients with IPF may suffer from systematic stem cell failure secondary to telomerase deficiency. As a result, patients may present comorbidities such as bone marrow failure, enteropathy, and liver diseases (223). For the same reason, they may also suffer from post-transplant complications such as cytopenia and immunosuppression after lung transplantation (224, 225).

1.4.5.2 Chronic obstructive pulmonary disease

Chronic obstructive pulmonary disease (COPD) is a multi-factorial trait where inflammation of the peripheral airways and/or loss of elastic recoil due to the destruction of alveoli (emphysema) cause irreversible progressive airflow limitation. In 2015, 174.5 million individuals worldwide had COPD and 3.2 million people died from it (226). The morbidity and progression of COPD are considered to result from a complex interplay among genetic factors, environmental factors, and lifestyle issues (227). While factors such as smoking and exposure to ambient particulate matter could explain 73.3% of disability-adjusted life years due to COPD (226), to date, alpha-1 antitrypsin (AAT) deficiency remains the only well-characterized genetic cause of emphysema (228).

Patients with COPD are known to have short telomere length in their peripheral blood leukocytes (229-231). Notably, the discovery of rare telomerase mutations in emphysema patients was previously reported (201, 232). In 2015, Stanley *et al.* sequenced the telomerase genes (*TERT* and *TER*) in COPD patients across two large cohorts (COPDgene and the Lung Health Study (LHS)) and discovered that 1% of patients with early-onset severe emphysema had deleterious TERT mutations (233). The prevalence of *TERT* mutations in COPD is comparable to AAT deficiency.

1.4.5.3 Pulmonary vascular diseases

Pulmonary arteriovenous malformation (PAVM) refers to the abnormal vascular connection between arteries and veins in the lung. The bypass of the normal pulmonary capillary bed for the capillary – alveolar barrier leads to decreased amount of gas exchange. Therefore, patients with PAVM suffer from reduced level of oxygen in blood (hypoxemia) and shortness of breath (dyspnea) (234). Initially, hepatopulmonary syndrome, the manifestation of which is PAVM, was found to be the cause of dyspnea in DC patients (235). One of the proposed mechanisms underlying PAVM is that short telomeres could induce senescence in hepatic endothelial cells, which further trigger a vascular response in the lung. The conditions in the lung are exacerbated with dysfunctional pulmonary endothelial cells (again, due to short telomeres) and finally result in PAVM (235). It was not until 2017 that PAVM was recognized as a phenotype of TBDs that was present in up to 3% of DC patients (236). However, 69% of patients with PAVM did not have evidence of any liver disease (236). Therefore, the precise mechanism of how telomere dysfunction leads to pulmonary vascular malformation remains to be discovered (222, 236).

1.4.5.4 Pathogenesis of TBDs in the lung

Despite the distinct clinical symptoms and physiological effects between IPF and emphysema, the two diseases share a few similarities: they are both chronic and progressive disorders that are mostly seen in the elderly; they are both heavily influenced by bronchial irritants such as occupational exposure or cigarette smoking (237, 238); most importantly, they are both caused by alveolar senescence and premature aging of the lung (239-241). In both diseases, the presentations in the lung are driven by injury or senescence in type 2 alveolar epithelial cells (AEC2), which serve as stem cells in the lung (242-244).

In such a slow turnover tissue as the lung, multiple factors such as smoking, aging and lung pathologies that increase the need to replenish the loss of pulmonary cells play important roles in disease manifestation. The substantial contributions of such factors predict that effects of genetic dysfunctions in telomere biology genes could be partially masked. Hence, IPF was identified in only 20% of DC patients, even though those patients were carrying mutations with predicted high penetrance in telomere biology genes (209). Correspondingly, only female smokers with mutations in *TERT* showed severe, early onset emphysema while their carrier male relatives were free of disease (233).

1.5 Rationale and hypothesis

1.5.1 Research questions

As discussed in Sections 1.4.4.3 and 1.4.5.4, phenotypic variations are observed extensively in people with mutations in telomere biology genes. For example, female carriers of rare *DKC1* mutations have been observed to display disease symptoms, even though the penetrance of *DKC1* mutations was thought to be restricted through XCI. Furthermore, the penetrance and expressivity of disease-associated autosomal *TERT* mutations were influenced by gender and environmental factors. However, the degree to which mild genetic perturbations, especially common variations in telomere biology genes, are implicated in the manifestations of pulmonary diseases remains unclear.

1.5.2 Overall hypothesis

I **HYPOTHESIZED** that the absence of skewed X chromosome inactivationcontributed to the disease onset in female carriers of rare *DKC1* mutations, whereas genetic polymorphisms in telomere biology genes, while not sufficient to drive the disease onset, acted in concert with

additional pathological and environmental factors to modify the manifestations of pulmonary diseases.

Chapter 2: Investigation of X Chromosome Inactivation on Phenotypic Variations in X-linked Dyskeratosis Congenita

2.1 Introduction

2.1.1 Dyskerin

The human *DKC1* gene is located at the tip of the X chromosome (Xq28). It consists of 15 exons and 14 introns. The encoded dyskerin protein is an evolutionarily conservative PUS of 57 kDa with 514 aa (245, 246). Studies from orthologs in other species, and from biochemical studies of human dyskerin in cell models, have mapped the following functional features in the dyskerin protein (Figure 2.1). It has a TruB domain (aa107 – aa247), a PUA domain (aa297 – aa371) and two nuclear localization signal regions (11 –20 aa,a446 – aa458). The TruB domain (named after an *Escherichia coli* pseudouridine synthase) serves as the catalytic domain of PUS. The PUA domain (PUS and archaeosine-specific transglycosylase domain) is the putative RNA-binding domain (247).

Dyskerin is a highly conserved and essential protein that also binds to the H/ACA box in TER and maintains the *in vivo* stability of TER (248). After assembly with TERT, the TERT-TER-dyskerin complex travels to chromosome ends and adds telomeric TTAGGG repeats during the replication phase of the cell cycle. *DKC1* mutations result in a dysfunctional dyskerin protein (248, 249). In comparison with WT cells, human cells with different dyskerin mutations have a two to five-fold reduction in TER levels, resulting in deficient telomerase activation and defective telomere maintenance (250). Co-expression of recombinant TERT and TER can restore telomerase activity and telomere maintenance in *DKC1* mutant patient fibroblasts without correcting for dyskerin sequence and function (250, 251). Co-expression of TERT and TER also

restored the growth capacity of X-DC cells, suggesting the phenotypic effects of X-DC are primarily mediated through telomerase insufficiency (250).

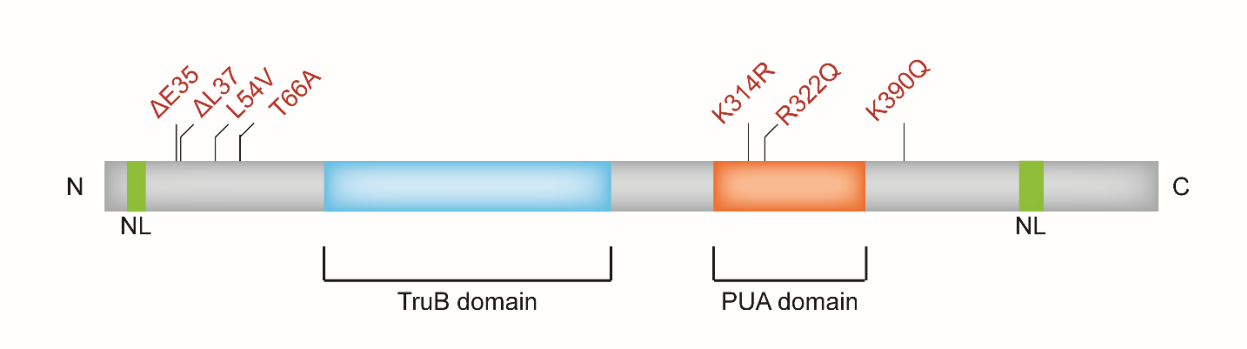

Figure 2.1 Schematic of dyskerin functional domains.

Non-synonymous amino acid changes in this study are illustrated in red. Studies from orthologs in other species, and from biochemical studies of human dyskerin in cell models, have mapped the following functional features in the dyskerin protein. It has a TruB domain (aa 107 – 247), a PUA domain (aa 297 – 371) and two nuclear localization signal regions (aa 11 – 20, aa 446 – 458). The TruB domain (named after an *Escherichia coli* pseudouridine synthase) serves as the catalytic domain of pseudouridine synthases. The PUA domain (pseudouridine synthase and archaeosine-specific transglycosylase domain) is the putative RNA-binding domain.

2.1.2 X chromosome inactivation

As discussed briefly in Section 1.4.4.3, XCI is the gene dosage compensation process by which one of the two copies of the X chromosome in females is inactivated. In an individual cell, the inactivation of either the maternal or paternal X is a random process. The inactivation happens during early blastocyst stage and the status can be stably inherited. Non-random inactivation, termed skewing, is defined by the inactivation of the same allele in over 75% of cells. Primary non-random inactivation happens by chance. In clinical settings, secondary (or acquired) skewing may result from the proliferative disadvantage of cells expressing mutant alleles on one of the X chromosomes. As such, females are protected from developing some X-linked recessive diseases and remain heterozygotes for X-linked mutations that may influence

development and cell viability. It is worthy to mention that the onset or severity of clinical symptoms in heterozygous female carriers of X-linked mutations can be influenced by the extent of skewing (252).

In heterozygous female carriers of *DKC1* mutant alleles, after gastrulation, their cells expressing the *DKC1* WT allele have a survival advantage over their counterparts expressing the mutant allele (212-214). This growth selection likely results in the depletion of cells expressing the mutant dyskerin allele and thus may explain the lack of DC phenotypes in female carriers.

2.1.3 Disease manifestations of X-DC in female *DKC1* mutation carriers

However, DC-like premature aging symptoms and phenotypic variations have been observed in a subset of female *DKC1* mutation carriers (213, 215). Since epigenetic regulation should have inhibited the expression of the mutant *DKC1* allele through XCI skewing in female carriers and protected them from disease manifestations, the biological mechanism of disease presentation in female *DKC1* mutation carriers has not been clearly addressed.

Recent findings suggest that variable XCI skewing patterns, such as germline mosaicism, could be the cause of clinical manifestations and disease variability in carrier females of X-linked neurological diseases (253). Similarly, an alternative explanation for the observed DC-like symptoms in *DKC1* mutation carriers is that this mosaic XCI skewing pattern could be specific to tissues or organ compartments where clinical features were observed (212, 215, 254). Additionally, XCI status in blood cells may not be representative of XCI status in tissues where cell lineage representation is usually invariable, such as in the epithelial compartments. However, neither of these hypotheses has yet been fully supported by direct evidence.

2.2 Hypothesis and specific aims

The objective of this chapter is to investigate phenotypic variations of X-DC symptoms in female *DKC1* mutation carriers. I **HYPOTHESIZED** that incomplete XCI skewing caused the expression of the *DKC1* mutant allele and hence the presentation of X-DC clinical symptoms in female *DKC1* mutation carriers.

The **SPECIFIC AIMS** for this chapter were as follows:

Aim 1: To determine telomere lengths in affected and asymptomatic female *DKC1* mutation carriers. We used multiple telomere length measurement assays to test the telomere lengths in the blood and disease-affected skin patches from affected and asymptomatic female *DKC1* mutation carriers. Telomere length measurements allowed us to understand whether these subjects had clinical diagnosis of DC, which was less than the first percentile in the age-adjusted population.

Aim 2: To evaluate XCI status in female *DKC1* mutation carriers. We used the human androgen receptor assay to investigate XCI status on the X chromosomes in DNA samples collected from patients. We also used an RT-PCR method followed by direct sequencing to check the monoallelic expression at the *DKC1* mutant loci in female carrier-derived cell lines. XCI status confirmed whether or not WT dyskerin is expressed, therefore providing direct evidence to support/disprove our hypothesis that the female carriers showed disease presentation due to their expression of mutant *DKC1* allele.

Aim 3: To test the expression level and function of dyskerin in patient-derived cell lines. We determined dyskerin expression level by western blot and tested the function of dyskerin in female *DKC1* mutation carriers for total RNA ψ level by HPLC and the maintenance of telomerase RNA stability by a competitive RT-PCR method. The functional changes in dyskerin protein could have resulted in reduced turnover potential of the affected cells/tissues.

To answer these questions, I performed the stated comprehensive assessments of female *DKC1* mutation carriers with clinical phenotypes suggestive of DC, and compared their molecular phenotypes with an extended panel of clinically unaffected female *DKC1* mutation carriers as well as healthy mutation-negative individuals.

2.3 Materials and methods

2.3.1 Study participants

Participants in this study are enrolled in the Institutional Review Board-approved longitudinal cohort study at the National Cancer Institute (NCI) entitled "Etiologic Investigation of Cancer Susceptibility in Inherited Bone Marrow Failure Syndromes" (www.marrowfailure.cancer.gov, NCI 02-C-0052, ClinicalTrials.gov Identifier: NCT00027274) (255). Affected individuals and their family members completed comprehensive family history and medical history questionnaires. My collaborators at NIH conducted detailed medical record reviews, collected biospecimens, and performed thorough clinical evaluations of affected individuals and their relatives at the NIH Clinical Center (255). Whole exome sequencing was performed as previously described (256). Clinical genetic testing was performed to confirm research-based genetic testing results. The dyskerin mutations involved in this study are illustrated in the schematic of dyskerin functional domains in Figure 2.1.

2.3.2 Cell models

EBV-transformed lymphoblastoid cell lines were derived from female *DKC1* mutation carriers and *DKC1* mutation-negative (wild-type, WT) control subjects. Another EBV-transformed lymphoblastoid cell line (GM03650, heterozygous for *DKC1* T66A mutation), was obtained from the Coriell Cell Repository. GM03650 was immortalized by retroviral vector-

mediated expression of TERT. Cell culture media, antibiotics, and other cell culture reagents were commercially available from Invitrogen/Life Technologies. Cells were maintained under standard culture conditions of 37˚C with 5% CO_2. All cell lines were cultured in RPMI 1640 medium (Gibco/BRL) with 10% fetal bovine serum (FBS).

Cells were harvested at around 80% confluency (about once per week). As these EBV-transformed lymphoblastoid cells were suspension cells, they were transferred to a fresh 15 mL conical tube, centrifuged at 1,500 rpm for 5 min, washed once with phosphate-buffered saline, aliquoted in 1.5 mL microcentrifuge tubes, and then stored at -80 ˚C for further analysis.

2.3.3 DNA and RNA isolation

Genomic DNA from whole blood, buccal cells, and fibroblasts was extracted by QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Total RNA from EBV-transformed cells was extracted with Trizol (Invitrogen, Carlsbad, CA). All purified nucleic acids were quantified by Nanodrop Spectrophotometer (Nanodrop Technologies, Oxfordshire, UK).

2.3.4 Telomere length measurement

Telomere lengths in patients' cells or patients' DNA were measured by two different methods: 1) flow FISH (flow cytometry with fluorescence *in situ* hybridization) in 6-panel leukocyte subsets using previously published protocols (257) and 2) Southern blot-based terminal restriction fragment analysis (TRF) to determine average telomere length (251). While samples were sent out by my collaborators to an external laboratory (RepeatDx) for the measurement by the flow FISH method, data from the TRF assay were generated by myself.

2.3.4.1 Flow FISH

A well-established automated flow-FISH assay was applied for telomere length measurement in patients' nucleated blood cells under a contract between NIH and RepeatDx

(257). Briefly, blood cell samples were added into a tube containing 100 μl of suspension solution (glucose 5%, 20 mM Hepes, 0.1% (w/v) bovine serum albumin (BSA)) and 2×10^5 control cells (cow thymocytes). After centrifugation, the supernatant was removed. The cell pellet, recovered in 10 μl of suspension solution, was resuspended with 170 μl of hybridization solution (75% deionized formamide, 20mM Tris, pH 7.1, 1% (w/v) BSA and 20 mM NaCl). 0.3 μg/ml FITC-conjugated PNA probe (5'-CCC TAA CCC TAA CCC TAA-3') was added to the hybridization solution. A no probe control was run in parallel. After 10 min incubation under room temperature, the cell suspension was denatured at 87 ˚C for 15 min in a circulating water bath. Hybridization was performed for 90 min at room temperature in the dark. Four wash cycles were performed with 1 ml of wash solution (75% formamide, 10 mM Tris, 0.1% BSA, 0.1% Tween 20) each before the cells were mixed and counterstained with FACSFlow containing 0.1% BSA, 10 μg/ml RNase A and 0.01 μg/ml LDS 751 for 20 min.

Telomere fluorescence for all the samples were acquired by FACSCalibur. The median telomere fluorescence in each cell type per sample was calculated by the specific fluorescence in cells hybridized with the FITC labeled telomere PNA probe minus the median fluorescence of unstained controls (background). Mean of duplicates were used for the final report. The fluorescence signals were further converted to average telomere length in kilobases based on the internal standard of the control cells (cow thymocytes).

To separate different types of cells in the blood samples and analyze median telomere lengths in each type of cells, after the washes, cells pellets were collected and incubated with 30 μl antibody solution containing anti-human CD45RA-Cy5 and anti-human CD20-PE, or with anti-human CD45RA-Cy5 and anti-human CD57-biotin, and then streptavidin-PE. Despite the

acquisition of telomere lengths from granulocytes, naïve T cells, memory T cells, B cells, natural killer cells, only results from the lymphocytes were used in the present study.

2.3.4.2 Preparation of terminal restriction fragment DNA

20 – 30 µg of genomic DNA was digested overnight at 37 ˚C with 20 units each of *Hinf*I and *Rsa*I (New England Biolabs, Ipswich, MA, USA). The total volume of the digestion reaction adjusted by H_2O to make 200 µl. These two restriction enzymes recognized and digested most regions of genomic DNA, but left the telomeric and subtelomeric regions intact. Digested DNA was mixed with 20 µl (1/10X volume) 3M sodium acetate (pH 5.2), 50 µg (2 µl of 25 mg/ml) linear polyacrylamide, and 600 µl (3X volume) of anhydrous ethanol and stored for >3 hours at -20 ˚C. Digested DNA was recovered by centrifugation at 13,200 rpm for 20 min. DNA pellets were resuspended in 30 µl of TE buffer (pH 8) and measured by the fluorescent QuantIT PicoGreen assay (Section 2.3.5).

2.3.4.3 Southern blot-based TRF assay

Digested DNA (2 µg) was resolved on 0.5% agarose-TBE gel overnight (approximately 16 hours). Two DNA marker sets, 1 kb DNA ladder and GeneRuler High Range DNA ladder (Fermentas, Burlington, ON, USA), were end-labeled with [γ-^{32}P]-ATP (3000 Ci/mmol, 5 mCi/mL, Perkin Elmer, Boston, MA, USA) and resolved in parallel with digested DNA. Agarose gels were dried for 2.5 hours at 60°C, denatured in 0.5 M NaOH/1.5 M NaCl for 20 min, and neutralized in 0.5 M Tris·HCl, pH 7.4/1.5 M NaCl for 20 min. Gels were prehybridized in Church's Hybridization Buffer (1 mM EDTA, 500 mM Na_2HPO_4, 7% SDS, 10 mg/mL BSA) for 1 h at 42 ˚C followed by overnight hybridization with a telomeric probe (5'-TTA GGG TTA GGG TTA GGG-3') end labelled with [γ-^{32}P]-ATP. Blots were 3X washed (30 min per wash) in 0.5X SSC (sodium chloride-sodium citrate buffer)/0.1% SDS), and exposed to a Phosphor screen

(GE Healthcare Lifesciences). Images were obtained at 100 µm resolution using the Typhoon Imager (GE Healthcare, Buckinghamshire, UK).

2.3.4.4 Quantification of TRF assay

Data from the TRF assay were analyzed by ImageQuant Software (v5.2, GE Healthcare, Buckinghamshire, UK). TRF length was determined as a weighted average. The gel image was partitioned into horizontal segments according to the size markers. For each TRF smear, the following equation was used to calculate mean telomere length: $TRF = \Sigma(OD_i * L_i)/\Sigma OD_i$, where OD_i is optical density at interval i, and L_i is the average length at interval i. TRF lengths were reported as kb (258).

2.3.5 QuantIT PicoGreen assay

The fluorescent QuantIT PicoGreen assay (Sigma-Aldrich, Oakville, ON) was used to determine the concentration of dsDNA after TRF digestion. The PicoGreen reagent was diluted 1:200 using TE buffer (pH 8) prior to the assay and stored in -20 ˚C until use. The 1 kb DNA ladder (50 µg/µl) New England Biolabs, Ipswich, MA, USA) was diluted with TE buffer (pH 8), and a two-fold dilution series was performed to provide a dsDNA standard in the range of 0.625 to 10 µg/mL. TRF samples were diluted 1:80 with TE buffer (pH 8). The diluted PicoGreen reagent was added to each well (57 µL per well) and 3 µL of either sample or standard were added to the diluted PicoGreen reagent. The plate was wrapped with foil to protect from light, and reagent was given 5 min incubation time with the samples and standards. The samples were excited at 480 nm and the fluorescence emission intensity was measured at 520 nm using PHERAstar FS microplate reader (BMG LABTECH, Ortenberg, Germany). A linear regression curve generated from the sample standards was used to determine the digested DNA concentration. The mean from triplicate measurements was adopted for further application.

2.3.6 XCI status measurement by the human androgen receptor (HUMARA) assay

XCI status at the DNA level was measured with slight modification to the standard protocol (259). During embryogenesis, the randomly inactivated X chromosome is methylated. Since XCI is irreversible, the methylation status of the inactive X chromosome is kept throughout the lifetime of the individual. Experimentally, upon digestion of a methylation-sensitive restriction endonuclease, the inactivated X chromosome remains intact whereas the active X chromosome is digested. In the HUMARA assay, we amplified the region containing the highly polymorphic CAG trinucleotide repeat in the promoter area of the human androgen receptor gene to study the XCI pattern. For females with random XCI (50% of the cells expressing one X and the other 50% of the cells expressing the other X), the PCR products of the amplified region would show two distinct bands on the gel (Figure 2.6). In contrast, only one band could be observed for females with completely skewed XCI (Figure 2.5A). The assay was performed for both digested and undigested DNA samples to exclude the possibility that the amplified DNA from an individual with random XCI showing only one band on the gel due to the inheritance of the same number of CAG trinucleotide repeats in the promoter area of the gene in two X chromosomes.

2 µg of genomic DNA from patients' cells was digested with 20 units of methylation-sensitive restriction endonuclease *Hpa*II (New England Biolabs) at 37 ^{o}C overnight in a total reaction volume of 20 µl. At the same time, another 2 µg of genomic DNA was incubated with the enzyme digestion buffer but no enzyme. The samples were heated at 70 ^{o}C to terminate the reaction. The CAG repeat regions from digested and undigested DNA samples were PCR amplified with the previously described primers (259): 2 µl of DNA template was added to 50 µl total volume PCR reaction containing 20 units of *Pfu* polymerase (homemade), 5 µl of 10x *Pfu* buffer (200 mM Tris·HCl pH 8.0, 15 mM $MgCl_2$, 680 mM KCl, 0.5% Tween 20, and 10 mM

EGTA), 1 µl of 10 mM dNTP mix, 2.5 µl (5%) of DMSO, and 1 µl of 25 µM forward and reverse primer each. The sequences of the primers were as follows: forward primer 5'-GCT GTG AAG GTT GCT GTT CCT CAT-3'; and reverse primer 5'-TCC AGA ATC TGT TCC AGA GCG TGC-3'. The forward primer was end-labeled with [γ-^{32}P]-ATP. 10 µl of the PCR products were resuspended in equal volume of 2X formamide loading dye (95% formamide, 10 mM EDTA, 0.01% SDS, 1% bromophenol blue and xylene), heated at 95 ^{o}C for 3 min and kept on ice. The products were resolved on a 17.5% denaturing polyacrylamide gel electrophoresis (PAGE) (17.5% acrylamide/bis (19:1), 0.5x Tris-boric acid-EDTA, 7M urea) for 4 hours at 650V. Gels were dried for 5 hours (3.5 hours at 80 ^{o}C and 1.5 hours cooling), then exposed to a storage phosphor screen. Images (100 µm resolution) were obtained using the Typhoon Imager.

2.3.7 Direct sequencing of expressed dyskerin allele by RT-PCR

Allelic expression of the *DKC1* gene in patient-derived cultured cells was studied using reverse transcription-PCR (RT-PCR) consisting of 1 µl of total RNA, 1 µl of SuperScript III Reverse Transcriptase (Life Technologies, Gaithersburg, MD, USA), 4 µl of 5x first-strand buffer, 2 µl of 100 mM DTT, 2 µl of 10 mM dNTP mix, and 1 µl of 25 µM specific reverse primer in a total reaction volume of 20 µl. The reaction was incubated at 42 ˚C for an hour before the inactivation of the enzyme at 70 ˚C for 15 minutes. Half of the first-strand cDNA was used for PCR amplification with specific primers. Genomic DNA was PCR amplified in parallel. The PCR products from both cDNA and DNA were resolved by agarose gel electrophoresis, purified by QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA), and analyzed by Sanger sequencing (Genewiz, South Plainfield, NJ, USA). The primer pairs for each mutation are listed in Table 2.1. Sequences for the primers are listed in Table 2.2.

Table 2.1 The primer pairs for genomic DNA-PCR and RT-PCR

Genomic DNA PCR				
Family		Forward primer	Reverse primer	Sequencing primer
#	Location			
ΔE35	Exon 3	DKC-I3-F	DKC-E3-R	DKC-I3-F
T66A	Exon 4	DKC-I4-F	DKC-E5-R	DKC-I4-F
K314R	Exon 10	DKC-I10-F	DKC-E10-R	DKC-E10-R
R322Q	Exon 10	DKC-I10-F	DKC-E10-R	DKC-E10-R
Complementary DNA RT-PCR				
Family		Forward primer	Reverse primer	Sequencing primer
#	Location			
ΔE35	Exon 3	DKC-E2-F	DKC-E6-R	DKC-E5-F
T66A	Exon 4	DKC-E2-F	DKC-E6-R	DKC-E5-F
K314R	Exon 10	DKC-E8-F	DKC-E10-R	DKC-E10-R
R322Q	Exon 10	DKC-E8-F	DKC-E10-R	DKC-E8-F

Table 2.2 Sequences for the primers used in the study

Primer	Location	Sequence
DKC-E2-F	Exon 2	AAG GAG CGG AAG TCA TTG CC
DKC-I3-F	Intron 3	CTT ATG CCA CCC ACA GAC CC
DKC-E3-R	Exon 3	AAA AGG GGC CAC TGA GAC G
DKC-I4-F	Intron 4	TGT ACT TAG TCC ATC TTG CAG ACC
DKC-E5-R	Exon 5	TGG GAA GAG GGG TTA GAG GG
DKC-E6-R	Exon 5 & 6	TCC CCA CAT ACT CTT TGC CTG C
DKC-E8-F	Exon 8	GGG AGT TGG TGG TCA GAT GC
DKC-I10-F	Intron 10	GAG TCG TGT GTA ATG TGC GG
DKC-E10-R	Exon 10	CAG ATT GCT TCT CCT TTG GTG G

2.3.8 Protein extraction

Patient-derived cultured cells were resuspended in hypotonic lysis buffer (20 mM HEPES pH 8.0, 2 mM $MgCl_2$, 0.2 mM EGTA, 10% glycerol, 1 mM DTT, and 0.1 mM PMSF). Cell suspensions were subjected to four consecutive freeze-thaw cycles from liquid nitrogen to a 37 ^{o}C water bath. NaCl was then added to the lysate in two parts to a final concentration of 400 mM. Cell suspensions were incubated on ice for 15 min before clearing the lysate by

centrifugation at 13,200 rpm for 15 min. The supernatant (whole cell lysate, WCL) was transferred to a fresh tube, aliquoted, and stored in -80 ºC for further application.

2.3.9 Bradford protein assay

Protein concentrations of WCLs were measured using the Bradford protein assay (Bio-Rad Laboratories Inc., Hercules, CA, USA). Dye reagent concentrate was diluted using distilled H_2O and a 1:5 dilution was prepared as the dye solution. BSA was diluted using hypotonic lysis buffer and a two-fold dilution series was performed to provide a protein standard series in the range of 62.5 to 500 µg/µL. They were stored in -20 ºC prior to use. Protein samples were prepared by diluting 2 µl of each sample with 48 µl of hypotonic lysis buffer. 10 µl of each standard and diluted sample were then pipetted into separate wells in a 96-well plate. 200 µl of the dye solution was added into each well, mixed with the protein standards and samples, and incubated at room temperature for 3 min. Absorbance was measured at 595 nm. The linear regression generated from the sample standards was used to determine protein concentration. The mean from triplicate measurements was adopted for further application.

2.3.10 Protein expression measurement by western blot

30 µg whole cell extracts were resolved in non-continuous Tris-glycine sodium dodecyl sulphate gels. The resolved protein samples were transferred to polyvinylidene fluoride membranes (GE HealthCare Life Sciences, Piscataway, NJ). The blots were incubated with anti dyskerin polyclonal antibodies (1:1,000, Santa Cruz Biotech) and anti-β-actin monoclonal antibody (1:40,000, Sigma). Protein signals were labeled with Alexa Fluor 680 Dye (Thermo Fisher Scientific, Rockford, IL, USA), detected by Licor Odyssey CLx Infrared Imaging System and quantified by ImageJ software.

2.3.11 TER copy number measurement by competitive RT-PCR

I used a previously well-established method to quantify the amount of telomerase RNA at steady state in patient-derived cultured cells (251, 260).

2.3.11.1 Preparation of TER competitors

A competitor RNA transcription template was amplified in a total PCR reaction volume of 50 µl consisting of 90 ng of plasmid BalI-Tag-TER-pRcCMV (gift from Dr. Kathleen Collins lab), 5 µl of 10X PCR buffer, 1 µl of 10 mM dNTP, 1 µl of 25µM forward primer (5'-GAA ATT AAT ACG ACT CAC TAT AGG TTC-3'), 1 µl of 25µM TER371-Rev (5'- TTC CTG CGG CCT GAA AGG CCT GAA CC-3'), 2.5 µl (5%) of DMSO) and 20 units of *Pfu* (homemade). PCR products were resolved on a 10% PAGE gel (10% 19:1 Acrylamide/Bis), excised and extracted from the gel with QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The plasmid contains entire TER sequence with a 20-nt tag (5'-GCT GAT ATA TAA CCT TCA GGG G-3') such that tagged recombinant TER could compete against endogenous TER for primer binding.

Competitor RNA was created through *in vitro* transcription with Ambion MAXIscript T7 transcription kit (Thermo Fisher Scientific, Rockford, IL, USA). Competitor RNA was resolved on 5% denaturing PAGE gel (5% Acrylamide/Bis (19:1), 8M urea, 0.5x TBE), extracted with phenol/chloroform, and purified with isopropanol. DNA concentration was determined by Nanodrop Spectrophotometer and converted to copy number (1 mole = 6×10^{23} copies). Competitor RNA was aliquoted and stored in -20 ^{o}C prior to use.

2.3.11.2 RT-PCR

1 µg of total RNA from each sample was reverse-transcribed along with increasing amounts of TER competitor RNA in a reaction of 20 µl consisting of 2 µl of 10X M-MuLV buffer, 2 µl of

10 mM DTT, 1 µl of 25 µM hTR-162-Rev (5'-GGC AGG CCG AGG CTT TTC-3'), 10 units of RNase Inhibitor (Applied Biosystems, Foster City, CA, USA), and 10 units of M-MuLV (New England Biolabs). Following reverse transcription at 42 ºC for 1 hour, PCR was performed with the hTR-162-Rev primer and the CY5-fluorophor labeled hTR-17-For primer (5'-GCC TGG GAG GGG TGG TG-3'). Following RT-PCR, amplification products (126 bp for endogenous TER and 146 bp for competitor RNA) were resolved on a 10% PAGE gel (10% Acrylamide/Bis (19:1), 1x TBE). The fluorescent image was scanned by a Typhoon imager, and the densitometry readings were quantified with ImageQuant software v5.2. The initial titration used a 10-fold dilution series of competitor RNA from 1×10^5 to 1×10^9 RNA copies. A finer titration was performed in a 2-fold dilution series over a range that matched the intensity of PCR products from recombinant and endogenous transcripts.

2.3.12 Pseudouridine level measurement by HPLC

The analytical method for the determination of pseudouridine in RNA (from patient-derived cultured cells) was modified from previous reports (261, 262). Quantitative analysis was performed by HPLC-UV system consisting of a Waters 2695 HPLC system and a Waters 2996 UV detector (Waters, Milford, MA). Data acquisition and processing was performed by Empower software. The chromatography was performed on a Waters Nova-Pak C18 column 3.9 mm x 300 mm, 4 µm. A detailed description of the method can be found in Appendix A , in which I developed a modified HPLC method to determine ψ levels in different types of RNA.

2.3.13 Statistical analysis

All data were analyzed by GraphPad Prism software (GraphPad Software, San Diego, CA). Error bars denote standard error of the mean (SEM). The student's t-test was used for two group

comparisons, and the one-way ANOVA with *post-hoc* Bonferroni correction was used to adjust for multiple comparisons. Differences were considered significant at $p < 0.05$.

2.4 Results

2.4.1 Some female *DKC1* mutation carriers showed DC symptoms

We evaluated two female *DKC1* mutation carriers who had phenotypic features overlapping with DC (Table 2.3).

Subject NCI-288-4, a female carrier of the *DKC1* L54V mutation, had alopecia and ridged nails. She came from family NCI-288, which consists of a male proband with classic features of DC (Figure 2.2A). Genetic testing for known telomere biology genes revealed a single mutation in *DKC1* (p.L54V). The proband's mother (NCI-288-3), who was confirmed to be a carrier of the same *DKC1* mutation (p.L54V), had multiple phenotypic features consistent with DC including aplastic anemia, non-alcoholic liver cirrhosis, abnormal nails, and early tooth loss. She died secondary to end-stage liver disease and hepatic encephalopathy at 65 years of age. Her sister (NCI-288-4) was also evaluated as part of this study.

Proband NCI-219-2 was a teenage girl whose brother and maternal uncle were both affected with DC. Her brother (NCI-219-1) died of post-hematopoietic stem cell transplant complications at age 3 years. She had two unaffected sisters (Figure 2.2B). The proband had nail dystrophy since birth, and skin hyperpigmentation and early graying of hair since the age of 10 years, all of which are phenotypes associated with DC (Figure 2.2C). Whole exome sequencing performed on the proband and her unaffected mother revealed a heterozygous mutation causing a single amino acid deletion in the coding sequence of *DKC1* (p.ΔE35) that was clinically validated to be present in both females. Genetic testing of the unaffected sisters revealed that they did not harbor

the mutation. Whole exome sequencing did not reveal mutations in any of the other known DC genes.

Table 2.3 Characteristics of female heterozygous *DKC1* mutation carriers and healthy controls.

Clinical status	Participant ID	Age at sample collection (years)	*DKC1* mutation
Affected carrier	NCI 219-2	15	ΔE35
	NCI 288-4	65	L54V
Unaffected carrier	NCI 96-2	55	K390Q
	NCI 103-2	59	ΔL37
	NCI 103-3	30	ΔL37
	NCI 106-4	36	K314R
	NCI 167-4	51	R322Q
	NCI 219-5	40	ΔE35
	GM03650	47	T66A
Healthy wild-type control	NCI 140	49	None
	NCI 156	31	None
	NCI 160	31	None
	NCI 204	50	None
	NCI 228	33	None

In addition to these two female *DKC1* mutation carriers with phenotypic features, we also identified six clinically unaffected female *DKC1* carriers, including the mother of NCI-219-2, all of whom were immediate family of males with X-linked DC. They served as a comparison group in the clinical and functional evaluations in this study. We also evaluated five healthy age-matched female WT controls who were mutation negative for *DKC1* and all other DC-associated genes (Table 2.3).

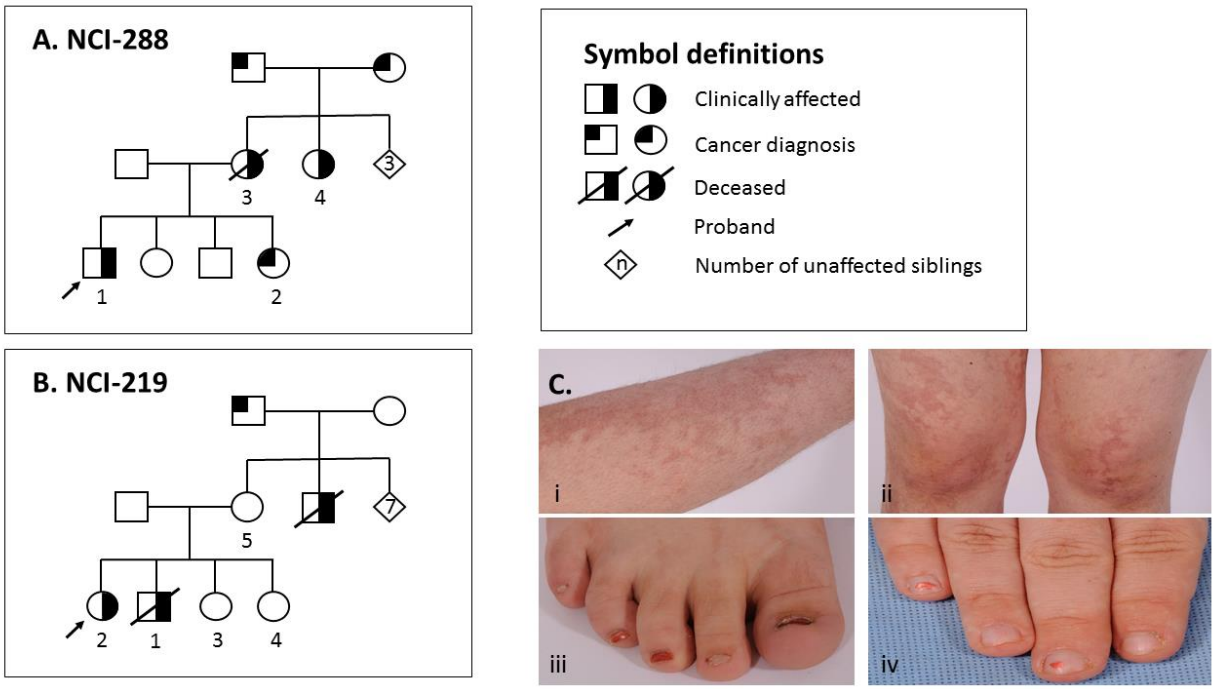

Figure 2.2 Family pedigrees of clinically affected female *DKC1* mutation carriers.

(A) Family NCI 288 with affected proband NCI 288-1, his affected carrier mother NCI 288-3, and affected carrier maternal aunt NCI 288-4. (B) Family NCI 219 with affected proband NCI 219-2 and her unaffected carrier mother NCI 219-5C. (C) Clinical manifestations of DC-associated phenotypes in female *DKC1* mutation carrier NCI 219-2. (i & ii) Characteristic skin pigmentation (iii &iv) Characteristic nail dysplasia.

2.4.2 Female *DKC1* mutation carriers had normal leukocyte telomere length

To explore the hypothesis that clinical manifestations of DC-associated symptoms in female *DKC1* mutation carriers were caused by short telomeres, we measured 6-panel leukocyte subset telomere lengths (granulocytes, total lymphocytes, and lymphocyte subsets) of all affected and unaffected female *DKC1* mutation carriers and compared them with healthy female controls. Female *DKC1* mutation carriers had normal telomere lengths in granulocytes and all lymphocyte subsets, regardless of DC phenotypes, and were comparable to age-matched healthy controls (Figure 2.3). In families where data were available, flow FISH telomere lengths of the male probands with *DKC1* mutations were very short (less than the first percentile for age) in all

leukocyte and lymphocyte subsets (Figure 2.3). The average leukocyte telomere length measured by TRF analysis of Southern blots correlated with and confirmed the flow FISH findings in the female *DKC1* carriers and healthy female controls (Figure 2.4).

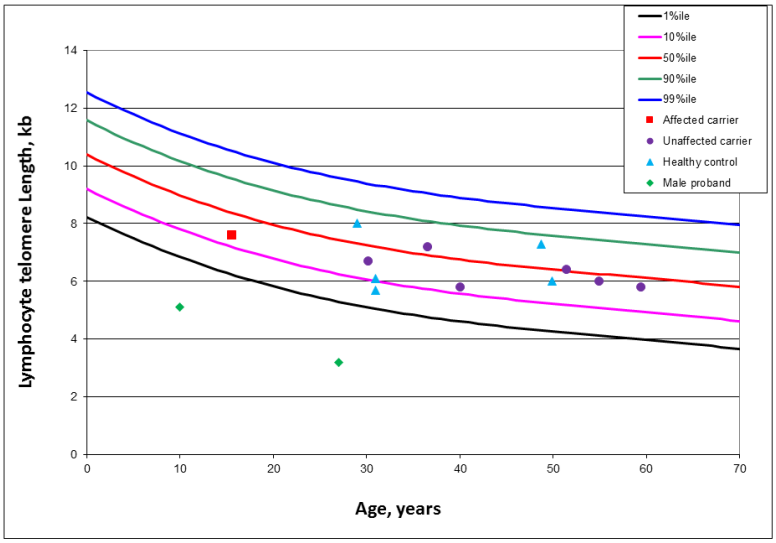

Figure 2.3 Mean lymphocyte telomere length in study participants

Total lymphocyte telomere length was measured by flow-FISH. The vertical axis represents telomere length in kilobases. Lines in the figures indicate the first, tenth, 50th, 90th, and 99th percentiles of results from 400 normal control subjects. Data include the telomere length measurements from two DC male probands for comparison.

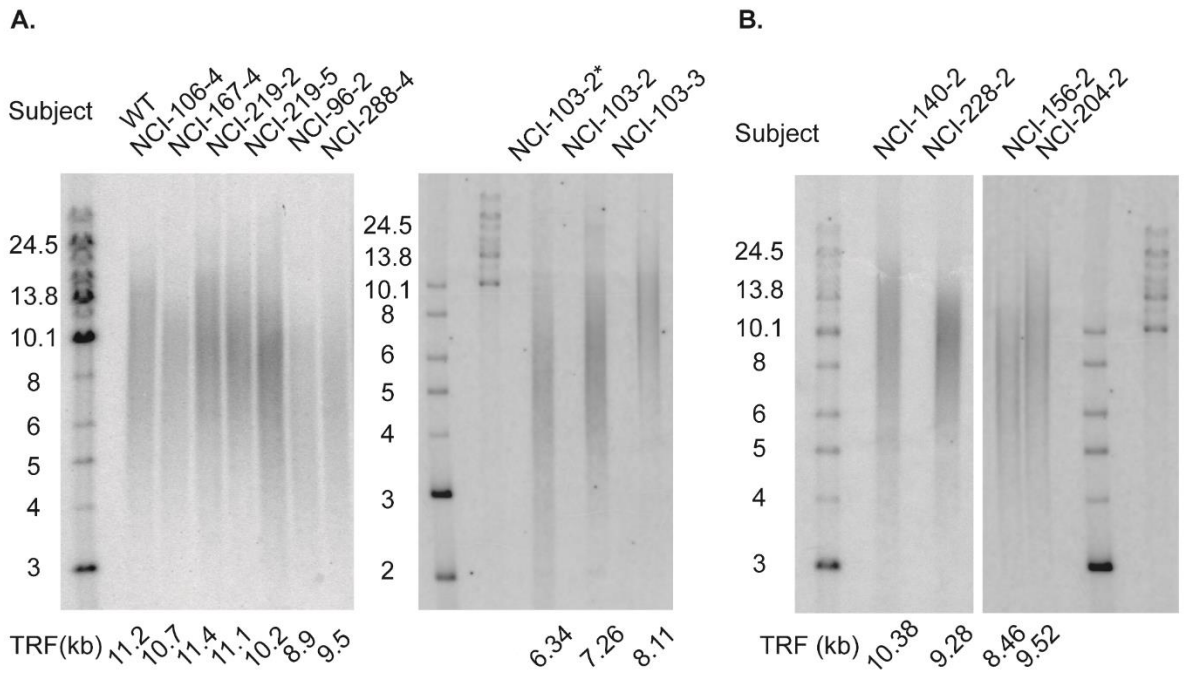

Figure 2.4 Mean leukocyte telomere lengths in female *DKC1* mutation carriers and mutation-negative controls.
Terminal restriction fragment analysis was used to quantify telomere lengths. The measured signal represents the sum of telomeric and some sub-telomeric region. Estimated telomere length was calculated by the weighted average of the lane and is shown at the bottom of the gel. All samples tested here were collected from leukocyte DNA without cell culture. The only exception (buccal cell DNA sample from NCI-103-2) is labeled with the asterisk. The figure shows that telomere length in parenchymal tissues is similar to that in blood cells. (A) Female *DKC1* mutation carriers. (B) Female mutation-negative healthy controls

2.4.3 Female *DKC1* mutation carriers showed completely skewed X chromosome inactivation

The modified HUMARA assay for XCI was performed on samples from seven subjects in five X-DC families: families NCI-219 (with *DKC1* mutation p.ΔE35), NCI-103 (p.ΔL37), NCI-288 (p.L54V), NCI-96 (p.K390Q) and the p.T66A sample (Table 2.3 & Figure 2.5A). XCI status in DNA samples from both blood and buccal mucosa were tested simultaneously. For the female proband NCI-219-2, the HUMARA assay was also performed on DNA from fibroblasts obtained from skin areas with and without pigmentation changes (Figure 2.5A). We observed two PCR

products of different sizes on the gel for all the undigested DNA samples, suggesting different numbers of CAG repeats in the promoter area of the human androgen receptor gene on the two copies of X chromosome in these subjects. Upon digestion with *HpaII* for the DNA samples, only one PCR product remained, indicating the expression of one specific X chromosome in all cells. Therefore, we concluded that skewed XCI existed in all samples, regardless of their *DKC1* mutations, clinical phenotype, or tissue of origin. In contrast, using the same assay, we observed random XCI patterns in all leukocyte DNA samples from healthy controls (Figure 2.6).

Talebizadeh *et al.* reported that factors other than the XCI process could affect the *in vivo* allelic expression of X-linked genes and that evaluation of XCI status at the gene expression level was essential to understand the effects of XCI for each genetic loci (263). To confirm that the mutant dyskerin allele did not escape XCI, we evaluated dyskerin mRNA expression by sequencing the dyskerin locus from genomic DNA to confirm their *DKC1* mutant carrier status, and compared the genomic sequencing data with RT-PCR-sequencing analysis of dyskerin mRNA expression (Figure 2.5B). Only wild-type (WT) dyskerin mRNA expression was detected in the EBV-transformed lymphoblastoid cells from one affected and four unaffected female heterozygous *DKC1* mutation carriers evaluated.

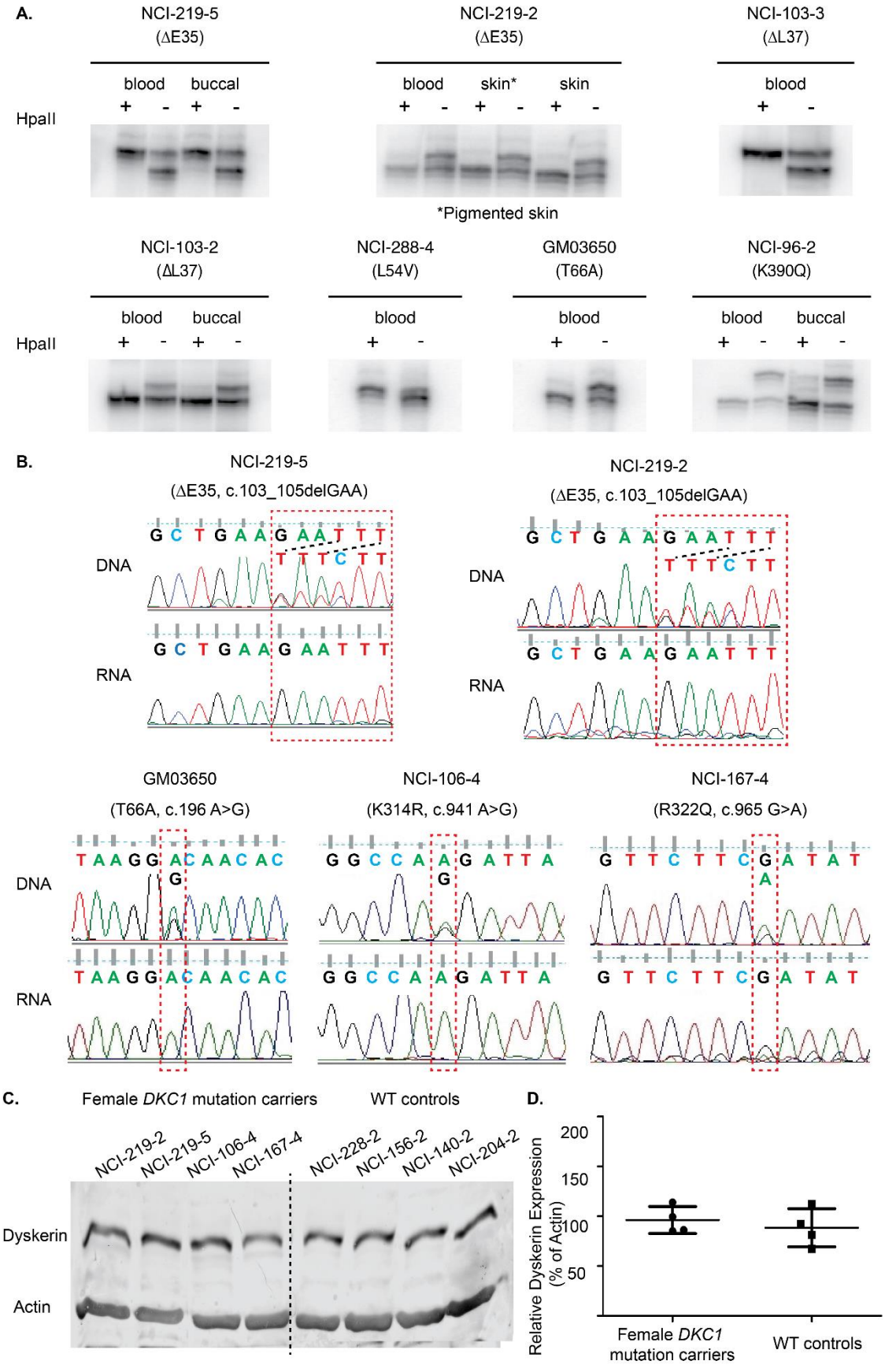
A.
NCI-219-5
(ΔE35)
NCI-219-2
(ΔE35)
NCI-103-3
(ΔL37)
blood
buccal
blood
skin*
skin
blood
HpaII
+ - + -
+ - + - + -
+ -
*Pigmented skin
NCI-103-2
(ΔL37)
NCI-288-4
(L54V)
GM03650
(T66A)
NCI-96-2
(K390Q)
blood
buccal
blood
blood
blood
buccal
HpaII
+ - + -
+ -
+ -
+ - + -
B.
NCI-219-5
(ΔE35, c.103_105delGAA)
NCI-219-2
(ΔE35, c.103_105delGAA)
DNA
RNA
G C T G A A G A A T T T
T T T C T T
GM03650
(T66A, c.196 A>G)
NCI-106-4
(K314R, c.941 A>G)
NCI-167-4
(R322Q, c.965 G>A)
T A A G G A C A A C A C
G
G G C C A A G A T T A
G T T C T T C G A T A T
A
C.
Female DKC1 mutation carriers
WT controls
NCI-219-2
NCI-219-5
NCI-106-4
NCI-167-4
NCI-228-2
NCI-156-2
NCI-140-2
NCI-204-2
Dyskerin
Actin
D.
Relative Dyskerin Expression
(% of Actin)
200
150
100
50
Female DKC1
mutation carriers
WT controls

Figure 2.5 Skewed X-chromosome inactivation and normal dyskerin protein expression in female *DKC1* mutation carriers, regardless of tissue origins or *DKC1* mutations.

(A) Skewed XCI patterns were detected in leukocytes and buccal cells for female mutation carriers in *DKC1* ΔE35, ΔL37, L54V, T66A, and K390Q families by HUMARA assay. In addition to PBMC, differently pigmented skin samples were collected from the female proband in family NCI-219, *DKC1* ΔE35. We tested the XCI status in her PBMC and fibroblasts derived from her pigmented skin (denoted with an asterisk) as well as from normal skin areas. All samples from this individual showed complete skewed X-inactivation. (B) Only wild-type (WT) dyskerin was observed in EBV-transformed lymphoblastoid cells derived from female carriers. The nucleotide sequences of both genomic DNA and complementary DNA of female carriers are illustrated on top of the chromatogram. *DKC1* genotypes at the variant nucleotide(s) for each family are illustrated in parentheses on top of the DNA sequences. (C) A representative western blot for dyskerin; β-actin was used to control the loading. (D) The dyskerin expression in EBV-transformed lymphoblastoid cells was quantified and normalized to their β-actin expression. The bold line and whiskers represent mean dyskerin expression level (normalized to β-actin) ± 1 standard deviation. NS, not significant.

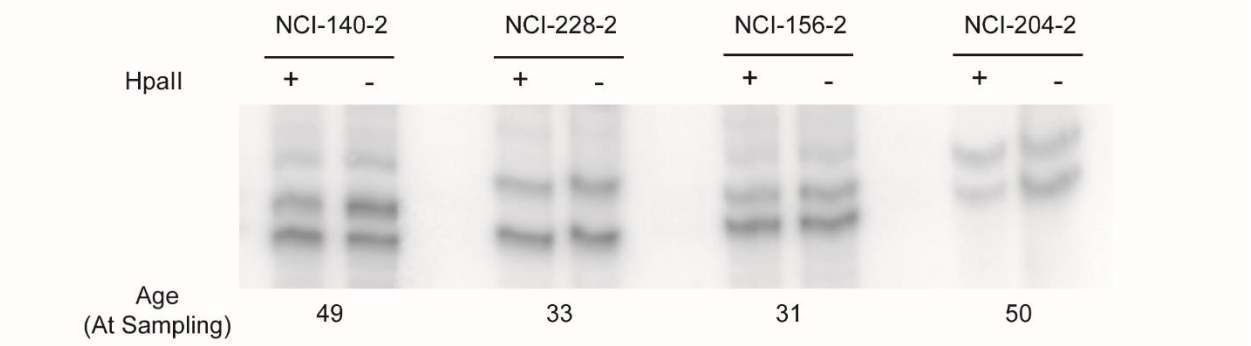

Figure 2.6 Mosaic XCI patterns observed in DNA from leukocytes of healthy *DKC1* mutation negative women by HUMARA assay.

As expected, random XCI patterns were detected in leukocytes from healthy females free of X-linked mutations. Therefore, the methylation-sensitive restriction enzyme *Hpa*II successfully digested the highly methylated (inactive) X chromosome in female carriers of *DKC1* mutation (Figure 2.5A). Our assay was unbiased.

2.4.4 Female *DKC1* mutation carriers showed normal dyskerin function and protein expression.

Lower steady-state level of TER is a molecular marker of X-linked DC. We evaluated and compared the level of stable TER in female mutation carriers with that of healthy controls in order to assess the ability of dyskerin to maintain TER accumulation (Figure 2.7A). Our data show that TER levels were significantly higher in both female *DKC1* mutation carriers and female *DKC1*-WT controls than in males with X-linked DC (data from (251)) (one-way ANOVA with Bonferroni post hoc test, $p < 0.0001$, Figure 2.7B). However, there were no significant differences in TER expression levels among healthy WT controls and female mutation carriers (Student's t-test, $p = 0.33$) (Figure 2.7B).

In addition to maintaining TER stability, dyskerin also has an essential function as a pseudouridine synthase in the site-specific modification of ribosomal RNAs (rRNAs) (264, 265), small nuclear RNAs (snRNAs) (247), non-coding RNA (ncRNA) and a subset of mRNA (266). Previous studies have shown that at least two disease-associated *DKC1* mutations are associated with a reduction in total rRNA pseudouridine levels (261). As several *DKC1* mutant alleles in our current study had not been previously evaluated, we quantified pooled pseudouridine levels in total RNA from our panel of female carriers of *DKC1* mutations. There were comparable levels between *DKC1* mutation carriers and healthy WT controls (Student's t-test, $p = 0.52$), suggesting that dyskerin's pseudouridinylase function is not compromised in female *DKC1* mutations carriers (Figure 2.8).

To evaluate whether there were differences in dyskerin protein expression between female *DKC1* mutation carriers and their age- and gender-matched healthy WT controls, we measured dyskerin expression levels by western blot analysis, and normalized dyskerin signals to the

expression of β-actin, as a loading control. No statistically significant difference was observed in normalized dyskerin protein expression level between the two groups (Student's t-test, p = 0.53) (Figure 2.5C & Figure 2.5D).

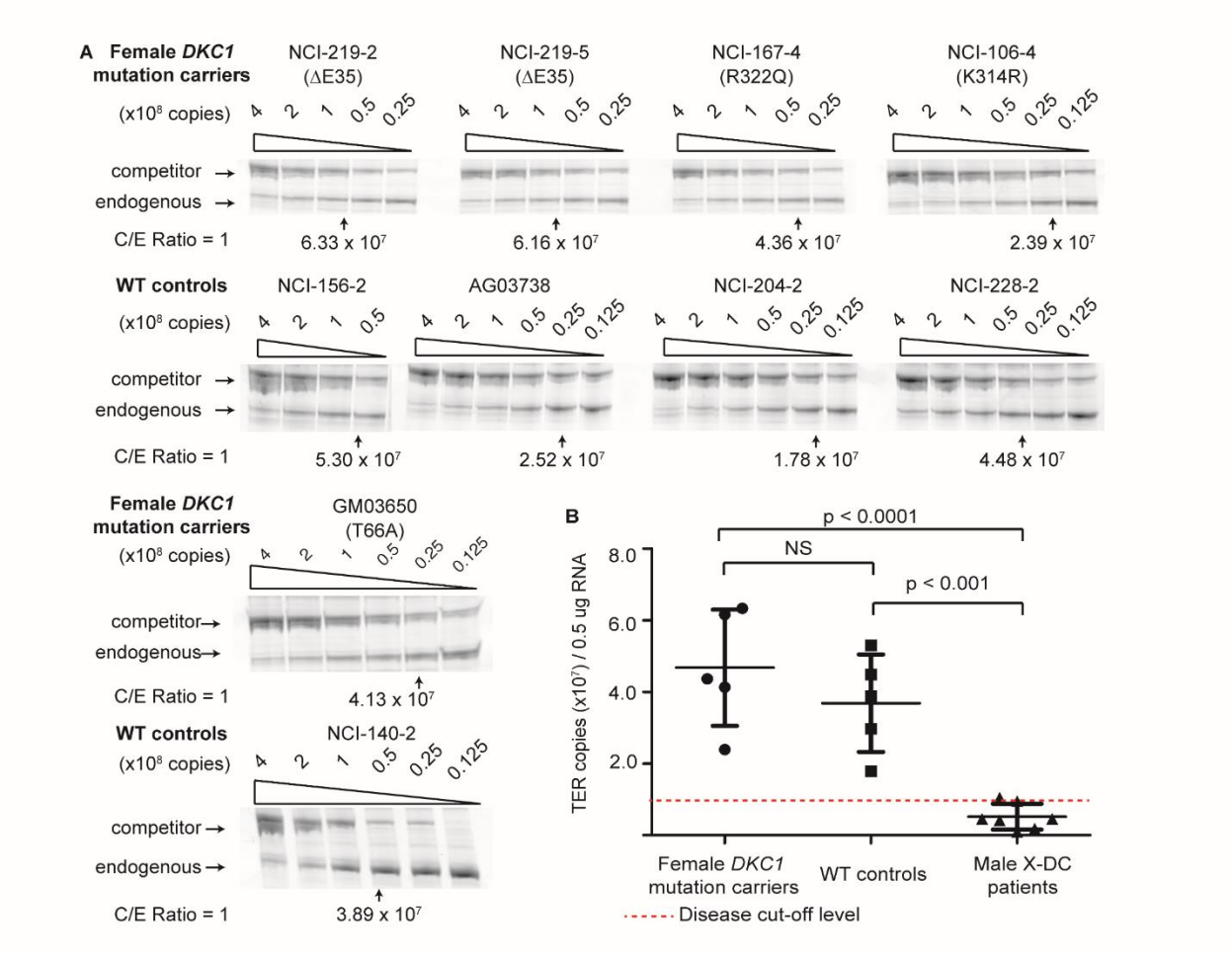

Figure 2.7 Female *DKC1* mutation carriers have normal TER levels at steady state.

(A) Steady-state TER levels were measured in EBV-transformed lymphoblastoid cells using a previously published competitive RT-PCR assay (see Section 2.3.11 for details). Input-tagged competitor RNA competitor RNA copy numbers are indicated for each series. The arrows indicate the approximate competitor inputs where 1:1 ratio of competitor to endogenous TER RT-PCR signals were obtained. For the quantification experiment, each RNA sample was measured first with a 10-fold serial dilution range of input competitor, then repeated with an

adequate 2-fold serial dilution range for final copy number calculation. (B) The densitometry quantification data are shown in the dot plot. The bold line and whiskers represent mean TER level (expressed as copy numbers of the competitor RNA) ± 1 standard deviation. NS, not significant. Though TER levels in both *DKC1* mutant carriers and WT control samples are highly variable, all of them are higher than that in TERT-expressing fibroblasts from X-DC male patient (data from (251)) ($p < 0.0001$, one-way ANOVA with *post-hoc* Bonferroni correction).

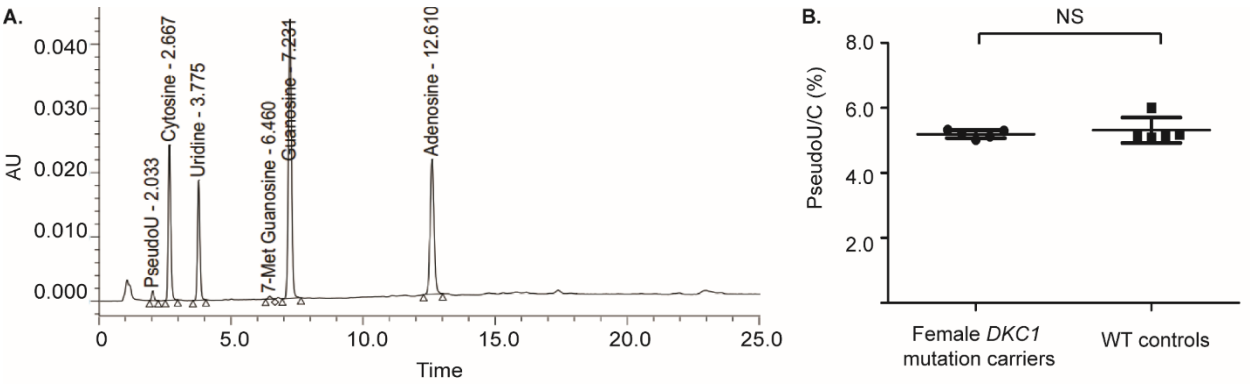

Figure 2.8 Female *DKC1* mutation carriers have normal pseudouridine modification levels.

(A) A representative HPLC of digested total RNA profile: the retention time for each nucleoside is illustrated in the chromatogram, and 7-methylguanosine (7-met G) is used as an internal control. (B) Pseudouridine level in total RNA. Steady-state pseudouridine levels in total RNA from female *DKC1* mutation carriers are comparable to those from WT controls. The bold line and whiskers represent mean pseudouridine level in total RNA (normalized to cytosine) ± 1 standard deviation. NS, not significant.

2.5 Discussion

This study revealed that there appears to be normal telomere length maintenance in the examined tissues from all female *DKC1* mutation carriers, including those affected with features of DC and carriers from families of male X-DC probands. We did not observe random XCI patterns or aberrant dyskerin gene expression levels in female *DKC1* mutation carriers. Our data also showed no functional changes in dyskerin protein expression, TER stability, and total RNA pseudouridine levels when comparing female *DKC1* mutation carriers (affected or unaffected)

with healthy WT controls. This is consistent with previous reports of normal telomere biology in female *DKC1* mutation carriers (215).

Functional analysis of the mutation p.Leu54Val (in family NCI-288) was previously performed in induced pluripotent stem cells (iPSCs) derived from male probands (267). In agreement with classical DC, p.L54V-derived fibroblasts and iPSCs were found to harbor reduced TER level, impaired telomerase activity, and defective telomere maintenance. Blood abnormality (i.e., aplastic anemia) was found in a carrier mother (NCI-288-3). However, since she passed away before the initiation of my study, no postmortem specimen was available to us for further investigation as to whether this case represents any cellular and molecular abnormality. The mutation p.ΔE35 in family NCI-219 has never been reported before our study. In this family, two male members displayed classical DC manifestations of immunodeficiency and bone marrow failure early in life, and they both died at ages 2 to 3 years. Since no other mutations in coding sequences of known telomere biology genes were found in this family, we infer that p.ΔE35 is responsible for the clinical phenotypes. Notably, both p.ΔE35 and p.L54V are located in a mutation hot spot in exon 3 at the N-terminus of dyskerin coding sequence (Figure 2.1). Additionally, a closely related residue (p.ΔL37) had been reported in numerous biochemical studies to cause classic DC, from our laboratory and others (44, 251).

It remains unclear why certain female *DKC1* mutation carriers develop DC-like phenotypes, in the presence of functionally normal dyskerin and normal telomere maintenance. One hypothesis is that secondary skewing of XCI may not apply to parenchymal tissues because their compositions of cell lineages are not as dynamically edited. As a consequence, dysfunctional dyskerin expressing cells may still exist in a subset of parenchymal cells and contribute to the mosaic epithelial symptoms observed (215). Our evaluation of XCI status in DNA samples

collected directly from unprocessed buccal cells (from all *DKC1* mutation carriers) and from normal and abnormally pigmented skin from NCI-219-2, an affected carrier, showed completely-skewed XCI (Figure 2.5A). These data suggest that the X chromosome encoding the mutant *DKC1* allele was not expressed *in vivo*.

We also observed WT dyskerin mRNA expression (Figure 2.5B), normal dyskerin expression (Figure 2.5C & D) and function (Figure 2.7 and Figure 2.8) with cultured cells derived from primary patient samples. While these data were in complete agreement with the skewed-XCI status and normal telomere length observed with uncultured clinical materials, it is important to note the assays' limitation: culturing of dermal tissue as fibroblasts, and blood cells as EBV-transformed lymphoblasts may have allowed for growth selection against cells expressing mutant dyskerin; thus, the negative results from our study using fibroblasts and lymphoblasts as sample sources could be explained by the selection bias favoring WT dyskerin-expressing cells in *ex vivo* culture. With the recent developments in *in situ* PCR with single cell resolution, targeted collection of clinical samples such as skin biopsies from affected and unaffected areas and the determination of allelic-specific expression at single cell resolution, may help to shed light on the potential mosaic dyskerin expression in parenchymal tissues at *DKC1* inactivation stages (268) or at steady-state allelic expression level (268, 269).

There were no concurrent mutations found in other known DC genes in the affected female carrier that may explain the sporadic appearance of DC-like phenotypes in some but not other *DKC1* mutation carriers. However, the lack of overt changes in dyskerin function in affected female carriers could still be explained by the compound presence of genetic and/or environmental modifiers that have not yet been identified to influence telomere biology. Telomere length maintenance is a dynamic process that involves both synthesis and attrition.

Weak perturbations of telomere pathway components, such as non-uniform expression of mutant dyskerin in female carriers, could predispose individuals to DC-like symptoms. But other genetic predispositions and/or environmental cues may be needed to induce clinical manifestations (270).

The dyskerin complex has recently been found to potentiate *in vitro* OCT4/SOX2-mediated transcription in embryonic stem cells and depletion of dyskerin in human fibroblasts resulted in decreased efficiency of iPSC conversion (271). As pluripotent embryonic stem cells are produced before the gastrulation stage during embryogenesis when XCI is established, female heterozygous *DKC1* mutation carriers expressing both *DKC1* alleles (WT and mutant) may have differential expression of pluripotency genes in their embryonic stem cells. These changes in early developmental stages may be followed by normal XCI and selection for WT-*DKC1* expressing cells. It is possible that mutant dyskerin may affect cellular differentiation and maturation in the female heterozygous mutant carriers.

In summary, we systematically studied dyskerin biology in a panel of female heterozygous carriers of disease-associated *DKC1* alleles. We observed extensively skewed XCI patterns in all female carriers, in samples from blood and buccal cells, as well as in patient-derived primary fibroblast cultures and lymphoblastoid cell lines. Exclusive expression of WT-dyskerin in these female subjects may protect them from severe manifestations of DC-associated molecular phenotypes. Female *DKC1* mutation carriers had normal telomere lengths, as well as normal dyskerin expression and function. These findings are surprising given the extent of the clinical phenotypes observed in patients such as NCI-219-2, NCI-288-3, and NCI-288-4. Clinical manifestations of DC phenotypes in female carriers may be a result of multiple factors, including the interactions between mosaic (temporal and positional) expression of mutant *DKC1*,

environmental and behavioural factors (such as smoking or recurrent infections) that influence the rate of telomere attrition as well as telomere length inheritance from affected family members. Further studies are warranted to elucidate the disease mechanism in female carriers.

Chapter 3: Investigation of the Effect of Single Nucleotide Polymorphisms on Telomere Length Maintenance and Phenotypic Variations in Chronic Obstructive Pulmonary Disease

3.1 Introduction

As discussed in Sections 1.4.5.2 and 1.4.5.4, telomerase dysfunction is involved in the pathogenesis of COPD. Telomere maintenance defects limit the proliferative lifespan of stem cells, the senescence and exhaustion of which is considered an important pathogenic origin of COPD (272). This provides a theoretical foundation to explain the observation that up to 1% of severe, early-onset patients with emphysema were carriers of deleterious mutations in *TERT* (233). Single nucleotide polymorphisms (SNPs) in both *TERT* and *TER* have been linked to telomere biology disorders, including, but not limited to, idiopathic pulmonary fibrosis (273) and bone marrow failure (274, 275). Rare, disease-causing *TERT* mutations precipitate the development of tissue regenerative deficiency such as emphysema (242). In this study, we were interested in understanding whether common *TERT* and/or *TER* SNPs associated with mild telomere maintenance defects could modify the risk and clinical presentation of COPD.

The activities of some telomerase variants have been assessed in many genetic studies. In these studies, telomerase activity is usually reconstituted through TERT and TER expression in rabbit reticulocyte lysates or through transfection of recombinant *TERT* and/or *TER* coding sequence harboring selective mutations into the WT-38 VA-13 cell line, which lacks both endogenous TERT/TER expression. The expression of telomerase in both systems is subject to inherent variations in the transient expression assay. Telomerase activity reconstituted through recombinant protein/RNA expression is measured by a PCR-based telomerase repeat

amplification protocol (TRAP) assay or by direct extension of a telomeric DNA primer (primer extension assay). A detailed comparison of these two assays was reviewed by Podlevsky *et al.* (276) and Aubert *et al.* (277). While the TRAP assay and its derivative methods provide rapid and sensitive detection, they are still limited by PCR-mediated amplification that they can only be used for the assessment of telomerase catalytic activity (nucleotide addition processivity) (278). On the other hand, the primer extension assay provides more accurate quantification of both nucleotide addition processivity and repeat addition processivity. However, it has low sensitivity and usually requires activity level from a highly concentrated TERT protein source, mostly through immunoprecipitation of cell extracts, which could introduce systematic errors for quantitative purposes. Since primers are provided to telomerase for the addition of telomeric repeats, neither of the assays can be used to detect any defects in localization of telomerase to telomere ends.

3.1.1 Single nucleotide polymorphisms in *TERT*

We selected eight nonsynonymous SNPs that are located in different functional domains of TERT to screen for potential SNPs that harbor deficient telomerase activity and telomere length maintenance. Four of these SNPs were identified in genome-wide association studies (GWAS) or in targeted association studies of telomere biology pathway genes with different types of TBDs. Due to the rarity of classic TBDs (for instance, DC is estimated to occur in approximately 1 in 1 million people), in most targeted association studies, there is a lack of sufficient patient numbers to establish the association between the SNPs and TBDs conclusively with enough statistical power. The drivers of TBD development are expected to have high penetrance. Therefore, common genetic variations in the *TERT* coding sequence should not induce the onset of TBD or serve as the primary driver of TBD progression. Nonetheless, the frequent occurrence of these

SNPs in different types of TBDs suggests that they may carry potential functional defects in telomerase activity and telomere length maintenance. The *TERT* nonsynonymous variants involved in this study are illustrated in the schematic of TERT functional domains in Figure 3.1 and listed in Table 3.1.

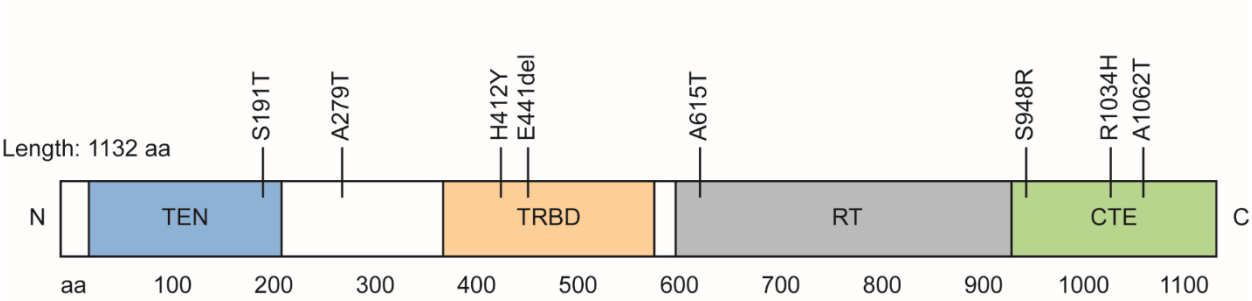

Figure 3.1 Schematic of TERT functional domains, showing the variants in this study

Nonsynonymous amino acid changes in this study are illustrated in the schematic of TERT functional domains (see Section 1.2.3.1 for details). Studies from orthologs in other species, and from biochemical studies of human telomerase reverse transcriptase have mapped the following functional features in the TERT protein. It has a telomerase essential N-terminal domain (TEN) (1 – 195 aa), the telomerase RNA-binding domain (TRBD) (322 – 594 aa), the reverse transcriptase (RT) domain (595 – 935 aa), and the C-terminal extension (CTE) (936 – 1132 aa).

The nonsynonymous *TERT* variant rs61748181 features a C>T nucleotide substitution, resulting in an amino acid change from alanine to threonine (TERT p.A279T). It is located at a linker domain of the N-terminus of TERT. Targeted deep sequencing has linked this SNP to bone marrow failure (DC and aplastic anemia) (274), aplastic anemia (275), myelodysplastic syndromes (MDS, a type of hematopoietic stem cell disorder) (279), lung cancer (280), and esophageal cancer (281). Compared to other disease-associated *TERT* variants involved in this study, it has the highest minor allele frequency (MAF) of 3.58% in Caucasians (282). Particularly, in a recent search for candidate genetic factors of COPD susceptibility, there were seven unrelated homozygotes for the rs61748181 minor allele in the case group (n = 1769) but

none in the control group (n=723), although the association between the variant and COPD was not statistically significant (*P*-value = 0.167) (283). Telomerase catalytic activity of p.A279T was reported to be around 80 – 100% of WT level (220, 284, 285).

The nonsynonymous *TERT* variant rs35719940 features a G>A nucleotide substitution, resulting in an amino acid change from alanine to threonine (TERT p.A1062T) in the C-terminus of TERT. It was one of the first *TERT* SNPs identified in association studies for TBDs. To date, it has been found in patients with aplastic anemia (275), acute myeloid anemia (286, 287), MDS (279), combined pulmonary fibrosis and emphysema (288), liver cirrhosis (289), and hepatocellular carcinoma secondary to cirrhosis (290). Telomerase catalytic activity of p.A1062T was found to be normal (220, 284) or around 60% of WT telomerase (286).

The MAFs for both rs61748181 and rs36719940 (Table 3.1) are high enough (>1%) that these two variants are considered SNPs. It is important to understand their genetic effects in the general population for susceptibility to potential TBDs. Meanwhile, *TERT* variations of lower MAFs were also found to exist in both patients and healthy subjects. Some rare *TERT* variants were prevalent in different types of TBDs.

The *TERT* SNP rs34094720 features an amino acid change from histidine to tyrosine (TERT p.H412Y). Similar to the conversion from arginine to threonine in the cases of rs61748181 and rs35719940, the switch from a positively charged (basic) amino acid in rs34094720 (histidine) to an uncharged one (tyrosine) at neutral pH in the TRBD domain of TERT could potentially have profound effect on protein function. Indeed, it was first identified in two unrelated patients with aplastic anemia (275). Later, it was also found in patients with autosomal recessive DC (291), IPF (220, 292), bone marrow failure (285), systemic sclerosis (an autoimmune connective tissue disease featured by fibrosis) (293) and in healthy volunteers (220, 285). Despite the consistent

identification of this SNP in various association studies to TBDs, there were contrasting reports about its telomerase catalytic activity: while p.H412Y was reported to have drastic reduction in telomerase catalytic activity by TRAP assay (275, 285), studies in other groups revealed a mild reduction or no change in activity by the primer extension assay (220, 284).

The *TERT* SNP rs377639087 features the deletion of codon 441 (TERT p. ΔE441). It has been discovered in patients with typical TBDs, including aplastic anemia (275), acute myeloid leukemia (286), liver cirrhosis (289), and cancer of epithelial origin such as renal cell carcinoma (294). As this SNP is known to have reduced telomerase catalytic activity by TRAP and to confer short telomeres (286), it will be used as a known telomerase activity defective control for our human cell model study.

Across the entire genome, nonsynonymous SNPs in coding regions only make up 0.60% of all known SNPs, whereas SNPs in introns account for 35.82% of the total variants (295). Intronic SNPs can have functional consequences through: 1) generation of new splice sites and/or modulation of alternative splicing in post-transcriptional regulation; and 2) modulation of protein expression level through the binding of transcriptional factors or the generation of alternative promoters (296).

An intronic SNP in *TERT* rs10069690 has been linked by association studies to multiple cancers, including estrogen receptor-negative breast cancer (297), *BRCA1* mutation carrier breast and ovarian cancers (298), colorectal cancer (299) and lung cancer (300). It was reported that the minor allele at rs10069690 (chr5: g.1,279,675 C>T) in intron 4 of *TERT* introduced an additional splice site donor and induced the expression of an alternatively spliced variant transcript INS1b (122). The MAF for rs10069690 is 27.6% in Caucasians (282). The INS1b transcript encodes a truncated form of TERT, with the telomerase TEN and TRBD domains remaining intact, but

without the RT or CTE domains. It is possible that the lack of reverse transcription activity of TERT INS1b transcript caused deficiency in telomere length maintenance itself or by competing with the expressed WT-TERT for TER binding. Therefore, an increase in expression of the INS1b transcript could be implicated in TBDs.

Table 3.1 List of *TERT* SNPs: nucleotide and amino acid changes and their frequencies

Amino Acid Change	Nucleotide Change	dbSNP Number	Overall		Caucasians		MAF		Functional domain	Data source[#]
							Overall	Caucasians		
S191T	572G>C	rs11952056	2	5006	0	1006	0.040%	0.000%	TEN	(282)
A279T	835G>A	rs61748181	48	4960	36	970	0.958%	3.579%	between TEN and TRBD	(282)
H412Y	1234C>T	rs34094720	3	5005	3	1003	0.060%	0.298%	TRBD	(282)
E441del	1323_1325 delGGA	rs377639087	4	5004	4	1002	0.080%	0.398%	TRBD	(282)
A615T	1843G>A	rs112614087	3	246159	1	111629	0.001%[*]	0.001%	RT	(301)
S948R	2844C>A	rs34062885	1	246099	0	111664	0.000%	0.000%	CTE	(301)
R1034H	3101G>A	rs62331332	1	240679	1	111603	0.000%	0.001%	CTE	(301)
A1062T	3184G>A	rs35719940	26	4982	22	984	0.519%	2.187%	CTE	(282)
N/A	1951-205G>A	rs10069690	1741	3267	278	728	34.764%	27.634%	intron	(282)

[#] Data retrieval date: August 31, 2018.

[*] The heterozygosity of rs112614087 was listed as 0.500 in the NCBI dbSNP database when we initiated the study (302). Therefore, during the initiation of my thesis study, this variant was deemed unlikely to change TERT function and was selected as a known telomerase activity intact control for our human cell model study.

3.1.2 Single nucleotide polymorphisms in *TER*

The *TER* SNP rs113487931 (r.58 G>A) features the transition of guanine to adenine and is 3-nt downstream of the highly conserved template region within the pseudoknot domain (Figure 3.2). It has been found frequently in patients with aplastic anemia (303-307) and MDS (304). The *TER* G58A variant has a MAF of 6.35% in the African population (282). Previous studies showed that telomerase holoenzyme reconstituted with this TER variant had normal function (128, 308, 309). In a long-term human cell-based study of up to 16 population doubling levels (PDLs, refers to the total number of times the G58A-expressing primary dermal fibroblast have doubled since the initial isolation of the strain), it was shown to have telomere length maintenance capacity comparable to WT telomerase (310).

The *TER* SNP rs141686314 (r.228 G>A) is located in the hypervariable paired region that links the STE domain to the H/ACA box (Figure 3.2). It has been identified in patients with aplastic anemia (304, 306). G228A is rarer than G58A and has a MAF of 2.42% in the African population (282). Telomerase activity reconstituted with G228A version of TER was shown to have normal activity (128, 308, 309). However, long-term, human cell-based telomere length maintenance was never tested.

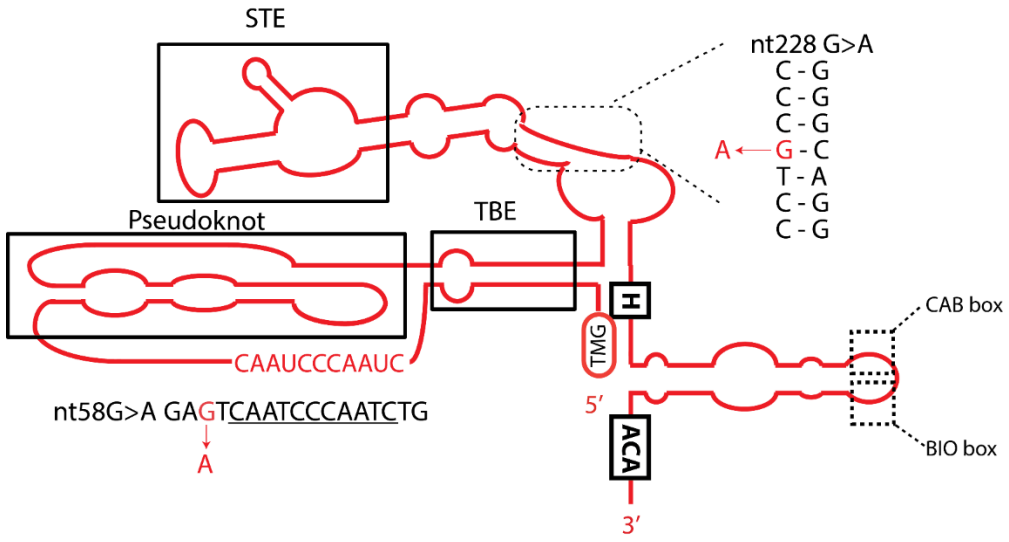

Figure 3.2 The secondary structure of TER, showing the variants studied in this thesis.

TER variants G58A and G228A were illustrated in secondary structure of telomerase RNA. TMG, the 2,2,7-trimethylguanosine cap structure; TBE, template boundary element; STE, the stem terminus element.

3.2 Hypothesis and specific aims

The objective of this chapter was to investigate the genetic effects of common variants in *TERT* and *TER*. I **HYPOTHESIZED** that selected genetic polymorphisms in *TERT* and *TER* genes caused functional defects in telomerase and compromised telomere maintenance *in vivo*. These dysfunctional *TERT* and *TER* variants, in turn, modified the progression of chronic degenerative disease, such as COPD, in patients carrying these variants.

The **SPECIFIC AIMS** for this chapter were as follows:

Aim 1: To determine telomere length maintenance dynamics and growth kinetics mediated by *TERT/TER* variants in a human cell model. We created a laboratory human cell model to test the telomere maintenance phenotype using isogenic cell lines expressing eight nonsynonymous variants in *TERT* and two SNPs in *TER*. These human cell models were carried continuously in culture for an extensive period of time to mimic the proliferative lifespan of human cells. In

addition to the examination of cellular growth kinetics, we also monitored telomere length maintenance dynamics in these human cell models by the Southern blot-based terminal restriction fragment analysis over the collection time course.

Aim 2: To characterize the mechanism of potential telomerase defects for the variants that show telomere length maintenance dysfunction in cell culture from Aim 1. We applied the primer extension assay to measure telomere catalysis using extracts from primary cells expressing TERT and/or TER isoforms. Using the primer extension assay, the mechanisms of telomerase defects were revealed by measuring telomerase nucleotide addition processivity and repeat addition processivity separately.

Aim 3: To evaluate the impact of telomerase defects in COPD disease progression. We investigated the association between *TERT* and/or *TER* SNPs and disease progression in subjects from the Lung Health Study (LHS) cohort. We tested the genotypes of the selected *TERT* and/or *TER* SNPs that showed functional defects in telomere length maintenance in LHS participants. Previously collected clinical data were used to establish a statistical correlation between patients' genotypes and disease progression.

3.3 Materials and methods

3.3.1 Cell culture

The primary human foreskin fibroblast BJ cell line was obtained from ATCC. The dermal fibroblast cell line TA1061 was collected from an identity-blinded, male X-DC participant in the National Cancer Institute's IRB approved study (the same source as previously described in Section 2.3.1).

All genetically-modified primary cell lines were cultured in DMEM high-glucose medium (Gibco, Waltham, MA, USA) with 15% FBS (Gibco) and were maintained at 37 °C with 5% CO_2.

Cells were harvested at 70 – 80% confluence (around once every 7 – 10 days) by treating cells with 0.5% trypsin-EDTA. Detached cells were transferred to a 15 ml conical tube. At this stage, cell numbers were counted and calculated according to the manufacturer's instructions using a Coulter counter (Beckman Coulter, Brea, CA, USA). Cells were centrifuged for 5 min at 1,500 rpm, washed once with phosphate-buffered saline, transferred to 1.5 ml microcentrifuge tubes, and stored at -80 ^{o}C for further analysis. Population doubling levels (PDLs) were calculated based on the following equation:

$$PDL = (\log Xe - \log Xb)/0.3 + S$$

where Xe was the cell harvest number, Xb was the inoculum number, and S was the starting PDL.

293T is a human embryonic kidney cell transformed with adenovirus and trasnduced with the SV40 T antigens. It provides high efficiency of transient transfection that is required for producing retrovirus packaging materials. GM00847 is an SV40-immortalized fibroblast cell line from a patient with Lesch-Nyhan syndrome. The two cell lines were obtained from ATCC and the Coriell Institute, respectively. They were cultured in DMEM high-glucose medium with 5% HyClone® FBS (GE Healthcare, Piscataway, NJ, USA) and were maintained at 37 °C with 5% CO_2.

3.3.2 Retroviral gene delivery

The retroviral protein expression vector pBabe-hygro-Flag3B-WT-hTERT (a gift from Dr. Kathleen Collins) was used as a template for *Dpn*I-mediated site-directed mutagenesis. Primers

for the mutagenesis were synthesized by IDT DNA (Coralville, IA, USA) and are listed in Table 3.2. Silent mutations were introduced in each pair of primers to create a restriction site for easy identification of positive clones through restriction digestion with a selected endonuclease. The restriction enzymes of choice are also listed in Table 3.2. Fidelity of positive clones was confirmed with Sanger sequencing of the entire *TERT* coding sequence (Genewiz, South Plainfield, NJ, USA). Retroviral infection was performed by the 293T cell-based system (311) to integrate *TERT* (both WT and variant) coding sequences into the BJ genome.

Briefly, the 293T cell line was triple-transfected through calcium phosphate method with the transfer plasmid (pBabe-hygro), the packaging plasmid pUMVC and the envelop plasmid pMD2.G in a 10:9:1 ratio. The medium was replaced after 24 hours. 48 hours after the transfection, polybrene was mixed with the medium to a final concentration of 5 μg/ml. The medium was filtered through a 0.22 μm syringe-tip filter (Millipore, Darmstadt, Germany) and applied to BJ fibroblasts. The filtered medium was placed again to BJ fibroblasts after 24 hours. Hygromycin selection at 50 μg/ml was applied to select for positive clones for at least 14 days. Positive clones were pooled for continuous culture. Biological repeats of cell lines were performed for a selected set of *TERT* SNPs. All the retrovirus-infected cell lines were polyclonal. The first passage following antibiotic selection was set as PDL = 1.

The reconstitution of telomerase activity mediated by SNP-TER and WT-TERT in the TA1061-TERT cell line was performed according to a previously published protocol (250). The dermal fibroblast TA1061 was from the skin biopsy of an X-DC patient with a dyskerin mutation A353V. It was previously immortalized by the aforementioned retroviral vector-mediated expression of TERT. The retroviral protein expression vector pBabe-puro-U3-WT-hTR (a gift from Dr. Kathleen Collins) was used as a template for *Dpn*I-mediated site-directed mutagenesis.

Similarly, retroviral infection was performed by the 293T cell-based system (311) to integrate *TER* (both WT and variant) sequences into the TA1061-TERT immortalized cell line. Puromycin selection at 5 μg/ml was applied to select for positive clones for 3 days.

Table 3.2 Primer design for site-directed mutagenesis

Primer #	dbSNP ID		Sequence (5' – 3')
TERT S191T with *Age*I-F	rs11952056	Forward	CCC CGC CAC ACG CTA CCG GTC CCC GAA GGC GT
TERT S191T with *Age*I -R		Reverse	ACG CCT TCG GGG ACC GGT AGC GTG TGG CGG GG
TERT A279T with *Eag*I-F	rs61748181	Forward	GTG GTG TCA CCT GCC CGG CCG ACC GAA GAA GCC ACC TCT
TERT A279T with *Eag*I-R		Reverse	AGA GGT GGC TTC TTC GGT CGG CCG GGC AGG TGA CAC CAC
TERT H412Y with *Bbs*I-F	rs34094720	Forward	CCC TAC GGG GTG CTC TTG AAG ACG TAC TGC CCG CTG CGA
TERT H412Y with *Bbs*I-R		Reverse	TCG CAG CGG GCA GTA CGT CTT CAA GAG CAC CCC GTA GGG
TERT E441del with *Bbs*I-F	rs377639087	Forward	GGC GGC CCC CGA GGA AGA CAC AGA CCC CCG T
TERT E441del with *Bbs*I-R		Reverse	ACG GGG GTC TGT GTC TTC CTC GGG GGC CGC C
TERT A615T-*Hinc*II-F	rs112614087	Forward	GAA GCC AGG CCC ACC CTG TTG ACG TCC AGA C
TERT A615T-*Hinc*II-R		Reverse	GTC TGG ACG TCA ACA GGG TGG GCC TGG CTT C
TERT S948R with *Nru*I-F	rs34062885	Forward	GCA GAG CGA CTA CTC GCG ATA TGC CCG GAC CTC C
TERT S948R with *Nru*I-R		Reverse	GGA GGT CCG GGC ATA TCG CGA GTA GTC GCT CTG C
TERT R1034H with *EcoR*V-F	rs62331332	Forward	CCA CAT TTT TCC TGC ACG TGA TAT CTG ACA CGG CCT CC
TERT R1034H with *EcoR*V-R		Reverse	GGA GGC CGT GTC AGA TAT CAC GTG CAG GAA AAA TGT GG
TERT A1062T with *Age*I-F	rs35719940	Forward	GGG GCC AAG GGC GCC ACC GGT CCT CTG CCC TCC GAG
TERT A1062T with *Age*I-R		Reverse	CTC GGA GGG CAG AGG ACC GGT GGC GCC CTT GGC CCC
TER G58A -F	rs113487931	Forward	CTA ACC CTA ACT AAG AAG GGC GTA G
TER G58A-R		Reverse	CTA CGC CCT TCT TAG TTA GGG TTA G
TER G228A-F	rs141684314	Forward	GGC GGG TCG CCT ACC CAG CCC CCG AAC
TER G228A-R		Reverse	GTT CGG GGG CTG GGT AGG CGA CCC GCC

3.3.3 Transfection

Based on the open reading frame of the known TERT-INS1b splicing variant (122), a truncated version of the *TERT* coding sequence that includes *TERT* exons 1-4 was cloned into the protein expression vector pBabe-hygro. The GM00847 cell line expresses endogenous TER, but not TERT. Cells were co-transfected with jetPRIME reagent (Polyplus transfection, Illkirch, France) with equal amounts of WT + INS1b-TERT, WT-TERT + vector, and INS1b-TERT + vector according to manufacturer's instruction. The medium was refreshed after 4 hours of transfection and the cells were harvested after 24 hours of the initial transfection.

3.3.4 Telomere length measurement

The Southern blot-based terminal restriction fragment analysis (TRF) was applied to determine average telomere lengths in cells (see Section 2.3.4 for a detailed description of the assay).

Telomeres in peripheral blood leukocytes in the LHS cohort were measured by qPCR as previously described by Jee Lee (the co-author of the manuscript in preparation). The data were previously published for the association between telomere length and mortality in COPD (230).

They used a modified qPCR protocol described by Cawthon for telomere length measurement and used 36B4 as the reference single copy gene (312). Two pairs of primers were used in the assay:

270 nM tel 1: 5'-GGT TTT TGA GGG TGA GGG TGA GGG TGA GGG TGA GGG T-3'

900 nM tel 2: 5'-TCC CGA CTA TCC CTA TCC CTA TCC CTA TCC CTA TCC CTA-3'

300 nM 36B4u: 5'-CAG CAA GTG GGA AGG TGT AAT CC-3'

500 nM 36B4d: 5'-CCC ATT CTA TCA TCA ACG GGT ACA A-3'

In addition to the primers, each reaction of 20 μL contained 10 μL of SYBR Green PCR Master Mix and 5 ng of DNA. All the samples were run in triplicates. The reactions were performed in an ABI PRISM 7900HT Sequence Detection System. For the telomere PCR, the reactions were run at 50 °C for 2 min then 95 °C for 2 min, followed by 30 cycles of 95 °C for 15s and 54 °C for 2 min. For 36B4 amplification, the reactions were run at 50 °C for min then 95 °C for 2 min, followed by 35 cycles of 95 °C for 15s and 58 °C for 1 min. telomere lengths were quantified as a relative T/S ratio (T = telomere, S = single copy gene, i.e., 36B4), calculated based on Cawthon's formula (230, 312):

$$2^{-(\Delta Ct1-\Delta Ct2)} = 2^{-\Delta\Delta Ct}$$

3.3.5 Telomerase activity

Telomerase activity was measured by two different methods: (1) telomerase repeat amplification protocol (TRAP) as previously described (250) and (2) primer extension assay (313).

3.3.5.1 Telomerase repeat amplification protocol (TRAP)

The PCR-based TRAP assay was used to measure telomerase activity in the transfected GM00847 cell lines. Whole cell lysates (WCL) from GM00847 transfected with different versions of TERT transcripts (see Section 3.3.3) were used as the source of telomerase. TRAP assay was performed using 500, 250, 125, and 62.5 ng (2-fold dilution series) of WCL. The 6-nt telomeric repeat addition by telomerase was carried out at 30 °C for 1 hr, in the presence of 100 ng M2 primer that served as a template, 1x PCR buffer (20 mM Tris·HCl pH 8.0, 1.5 mM $MgCl_2$, 68 mM KCl, 0.05% Tween 20, and 1 mM EGTA), and 2.5 nmol of dNTP. Reactions were heat-inactivated at 95 °C for 3 minutes. The primer extension products were then PCR-amplified using 2.5 units of *Pfu* DNA polymerase (homemade) in 1x PCR buffer with 100 ng of

a CY5-fluorophor-labeled reverse primer CX3, 90 ng of NS primer, and 5.0 x 10^3 copies NSCX3 internal control primer. The products were resolved on a 10% PAGE gel (10% acrylamide/bis (19:1), 0.5x TBE, 0.1% APS and 0.1% TEMED) and visualized using the Typhoon imager with the Cy5 670BP30 filter (wavelength 655 – 685 nm). The sequences of primers used in TRAP assay are listed in Table 3.3.

Table 3.3 List of primers in TRAP assay

Primer	Sequence
M2	5'-AAT CCG TCG AGC AGA GTT-3'
CY5-CX3	CY5-5'-CCG CGC CCT AAC CCT AAC CT-3'
NS	5'-AAA AGG CCG AGA AGC GAT-3'
NSCX3 (internal control)	5'-AAA AGG CCG AGA AGC GAT TAG GGT TAG GGT TAG GG-3'

3.3.5.2 Primer extension assay (conventional assay)

The classical radioactive primer extension assay was applied to a selected panel of primary BJ cell lines with recombinant TERT (WT and variant) expression (313).

A typical reaction of 40 µl consisted of telomerase (100 µg of WCL, up to 20 µl), 40 nmol of dATP, 40 nmol of dTTP, 4 nmol of dGTP, 250 nCi of [α-^{32}P]-labelled dGTP, 0.1 nmol of an 18 nt primer (5'-GGG TTA GGG TTA GGG TTA-3'), and 10x assay buffer (50 mM Tris-acetate pH 8.3, 1 mM $MgCl_2$, 50 mM potassium acetate, 1 mM spermidine, and 5 mM β-mercaptoethanol). The sodium concentration in WCL was diluted down from 400 mM to 200 mM to prevent the inhibition of telomerase activity under high-salt conditions.

Reactions were incubated at 30 ^oC for 2 hours and stopped by 15 min incubation with 50 µl Stop Solution A (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.1 mg/mL RNase A), followed by another 15 min incubation with 50 µl Stop Solution B (10 mM Tris-HCl pH 7.5, 0.4% sodium dodecyl sulfate, 0.2 mg/mL Proteinase K). After the addition of a 15 nt [γ-^{32}P]-ATP (3000

Ci/mmol, 5 mCi/ml, Perkin Elmer, Boston, MA, USA) end-labelled recovery control, DNA products were purified by adding an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0), then precipitated in 1/10X volume of 3M sodium acetate (pH 5.2) and 3X volume of anhydrous ethanol, with 50 µg linear polyacrylamide as a carrier. Following overnight precipitation at -20 ^{o}C, products were recovered by centrifugation at 14,000 rpm for 40 mins at room temperature, resuspended in equal volumes of Tris-EDTA buffer (pH 8.0) and 2x formamide loading dye, and boiled for 3 min before snap-chilling on ice. Products were resolved on a 17.5% denaturing PAGE gel for 4.5 hours at 450V. The gel was dried for 5 hours (3.5 hours at 80 ^{o}C and 1.5 hours cooling), then exposed to a phosphor screen for six days. Images (50 µm resolution) were obtained using the Typhoon Imager.

Quantitation of telomerase catalysis and repeat addition processivity were previously described using ImageQuant softweare v5.2 (314). Total enzyme activity (nucleotide addition processivity, NAP) for each reaction was determined by normalizing the total product ladder intensity from an entire gel lane (starting from the primer +1 nt product) to the intensity of the RC. The NAP activity-normalized values were then normalized to the reaction of BJ cells expressing WT-TERT.

For the calculation of repeat addition processivity (RAP), the counts for each telomerase band was obtained individually, normalized for the number of radiolabeled guanosines added. For each single band, the intensity of signal was corrected for the number of ^{32}P-G's incorporated to the product. The fraction of total remaining signal (FLB, fraction left behind) for a band ***n*** was calculated by taking the sum of counts for repeats (total – ***n***) divided by the total signal. Ln(1-FLB) was plotted versus repeat number. Each lane was fit with a linear regression equation with the slope ***m***. The processivity value equals -0.693/***m***. Due to the abortive initiation

by telomerase in the first few repeats (314, 315) and the weak signals for the repeats at the high molecular weight range, only the 3rd to 12th repeats were included for the linear regression. RAP activity was then normalized across samples.

3.3.6 Immunoblot

100 μg WCL were resolved in 5% SDS-PAGE gel. A standard western blotting protocol was applied for the detection of TERT expression in the cell lysates. A sheep anti-human TERT polyclonal antibody (0.5 μg/ml, Abbexa, Cambridge, UK) and a Donkey anti-sheep horseradish peroxidase-conjugated antibody (0.2 μg/ml, Jackson ImmunoResearch, West Grove, PA, USA) were sequentially applied before detection by ECL (PerkinElmer, Waltham, MA, USA) on Kodak X-ray film. The cytoskeletal protein vinculin was used as a normalization control. It has a molecular weight of 124 kDa, which is close to TERT (127 kDa). Rabbit anti-human vinculin monoclonal antibody (0.1 μg/ml, Cell Signaling, Danvers, MA, USA) and Alexa Fluor 680 Dye (ThermoFisher Scientific, Waltham, MA, USA) were applied before detection by Licor Odyssey CLx Infrared Imaging System (LI-COR Biotechnology, Lincoln, NE, USA) and quantification by ImageJ software (NIH, Bethesda, MD, USA).

3.3.7 COPD patients

This study used clinical data and biological materials from the LHS. The LHS cohort initially recruited 5,887 smokers, aged 35 – 60, with mild airway obstruction across 10 clinical centers in North America and collected their written informed consents (316). Spirometry was performed annually over a period of 5 years for all study participants and in a subset after 11 years of study initiation, as previously described (317, 318). Lung function was assessed as % predicted forced expiratory volume in one second (FEV1) and was adjusted for age, height and sex (319). A subset of the study participants that had the fastest disease progression (average decrease in %

predicted FEV1 of 3.59 per year) and slowest disease progression (average increase in % predicted FEV1 of 0.45% per year) were selected for the preliminary study.

3.3.8 Genotyping

Two polymorphisms from *TERT* (rs61748181 and rs10069690) were genotyped by Taqman® SNP Genotyping Assay (Applied Biosystems, Carlsbad, CA, USA) on an ABI GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). Template-free controls and known genotype controls were included in each experiment. Data were analyzed by ViiA^{TM} 7 Software (Applied Biosystems). For assay quality control, 5% of the samples were re-genotyped to check data reproducibility of the assay.

3.3.9 Statistical analysis

Genotype distributions were assessed for Hardy-Weinberg equilibrium. Linkage disequilibrium between the rs61748181 and rs10069690 SNPs was assessed using R packages 'genetics' v1.3.8.1 for χ^2 test. Pearson's χ^2 test and student's t-test were used to assessing the differences in dichotomous and continuous demographic parameters between patients with extremely rapid decline of lung function and those with no decline (rapid decliners and non-decliners), as well as patients with and without the minor allele of rs61748181. The frequencies of the alleles and genotypes between groups were analyzed by Pearson's χ^2 analysis for 2 x 2 contingency tables for crude odds ratios (OR) and by multiple logistic regression for adjusted ORs. The ORs and their 95% confidence intervals (CI) were calculated. Confounding factors included age, gender, pack-years of smoking (one pack year is defined by smoking 20 cigarettes per day for one year) and smoking status. Statistical analyses were done by the JMP Statistics software package version 13 (SAS Institute Inc.) and R version 3.5.1. Figures were generated by GraphPad Prism version 7.0 (La Jolla, CA, US).

3.4 Results

3.4.1 Five variants in *TERT* lead to compromised telomere length maintenance

We applied an *ex vivo* human cell model to check whether the minor alleles of eight *TERT* variants, when translated into variant versions of TERT, were functionally different from WT-TERT with long-term telomere length maintenance as a readout. We used site-directed mutagenesis to introduce the sequences encoded by the minor alleles of the SNPs into *TERT* open reading frame. Like other human fibroblasts, the human foreskin fibroblast BJ cell line silenced its endogenous TERT expression, whereas its TER expression is intact. Stable genomic integration of the coding sequences of *TERT* (WT and the SNPs) through infection with retrovirus vectors (pBabe-Hygro-flag3B-TERT) in BJ cells restored telomerase activity (Figure 3.3A). An examination of protein expression revealed that TERT protein levels in these reconstituted cell lines were comparable to each other (Figure 3.3B). While we observed a higher expression of TERT in the A1062T-TERT BJ cell line, this may lead to the cell line having potentially increased telomere length maintenance capacity than BJ cell line expressing WT-TERT (leading to a conservative bias against the association of A1062T-TERT with decreased telomere length maintenance). The population growth kinetics of the TERT-expressing BJ cell lines followed typical logarithmic curves. While the BJ vector control cell line (BJ-vector), which does not express telomerase activity, entered replicative senescence following 10 PDLs post-antibiotic selection, all BJ fibroblasts expressing recombinant telomerase (WT/SNP variants) were proliferating continuously in our *ex vivo* cell model system. Although the variants were carried in cell culture during different time period, the WT-TERT and the BJ vector control cell line were always carried at the same time. Notably, WT-TERT and TERT variant-expressing cell lines did not show any visible difference in growth kinetics up to 50 – 60 PDLs (Figure

3.3C). There were a few variants showed fluctuations in telomere length maintenance (detailed in Section 3.4.3). Yet none of them showed differential growth kinetics from WT-TERT-expressing cell line up to 100 PDLs.

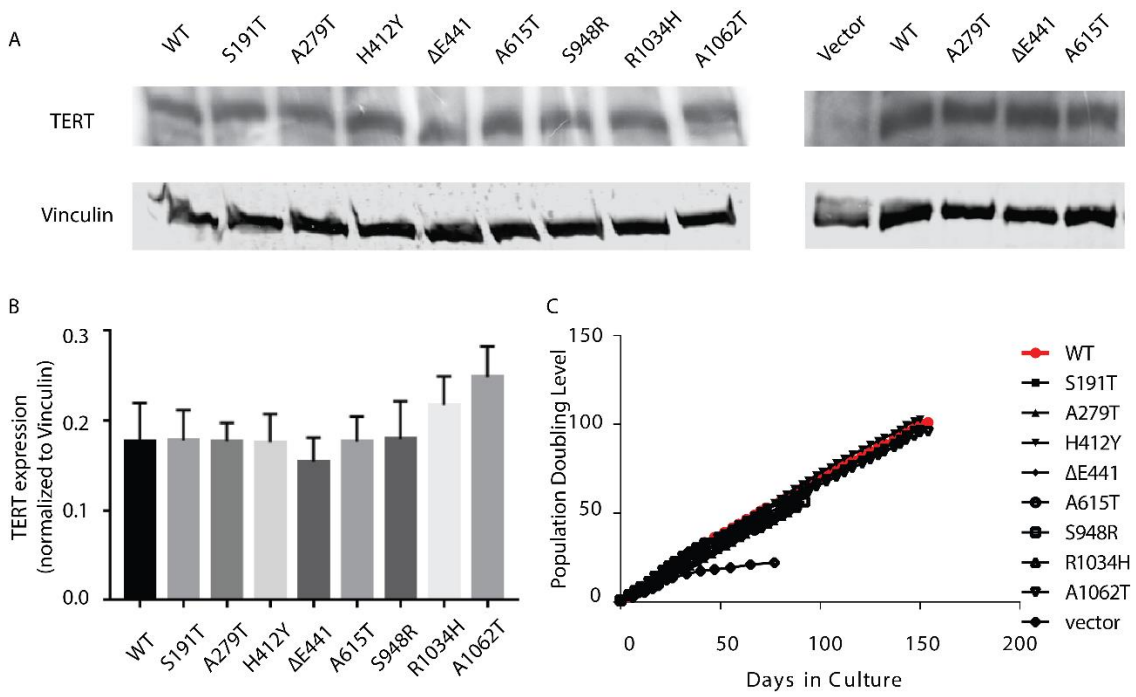

Figure 3.3 Restoration of recombinant telomerase mediated by the TERT variants in BJ cell lines.

(A) Recombinant expression of TERT in BJ fibroblasts. TERT expression levels (127 kDa) are normalized to those of vinculin (124 kDa). The specificity of the TERT antibody is illustrated in the right panel. (B) TERT expression in different cell lines is comparable and the data are presented as mean ± SEM (n = 6). (C) Cellular growth kinetics of BJ cell lines expressing recombinant SNP-version of telomerase.

Next, we checked the dynamics of mean telomere length in these cell populations over time. The measurement of telomere length confirmed that senescence in the BJ-vector cell line was caused by progressive loss of telomeric repeats (Figure 3.4). In contrast to their comparable cellular growth kinetics, WT-BJ and variant-BJ cell lines presented differential telomere length

dynamics. While BJ cell lines expressing WT-TERT and S191T, H412Y and A615T-TERT variants showed moderate telomere lengthening as a function of proliferation (Figure 3.4F), BJ cell lines expressing A279T, ΔE441, S948R, R1034H, and A1062T-TERT variants had proliferation-dependent decreases in TRF within the same period (Figure 3.4E). In short, while the restoration of telomerase activity mediated by some *TERT* SNPs (i.e., A279T, ΔE441, S948R, R1034H, and A1062T) prevented BJ cells from entering replicative senescence, these cell lines showed compromised telomere maintenance capacity within 40 PDLs, compared to WT-TERT expressing cells.

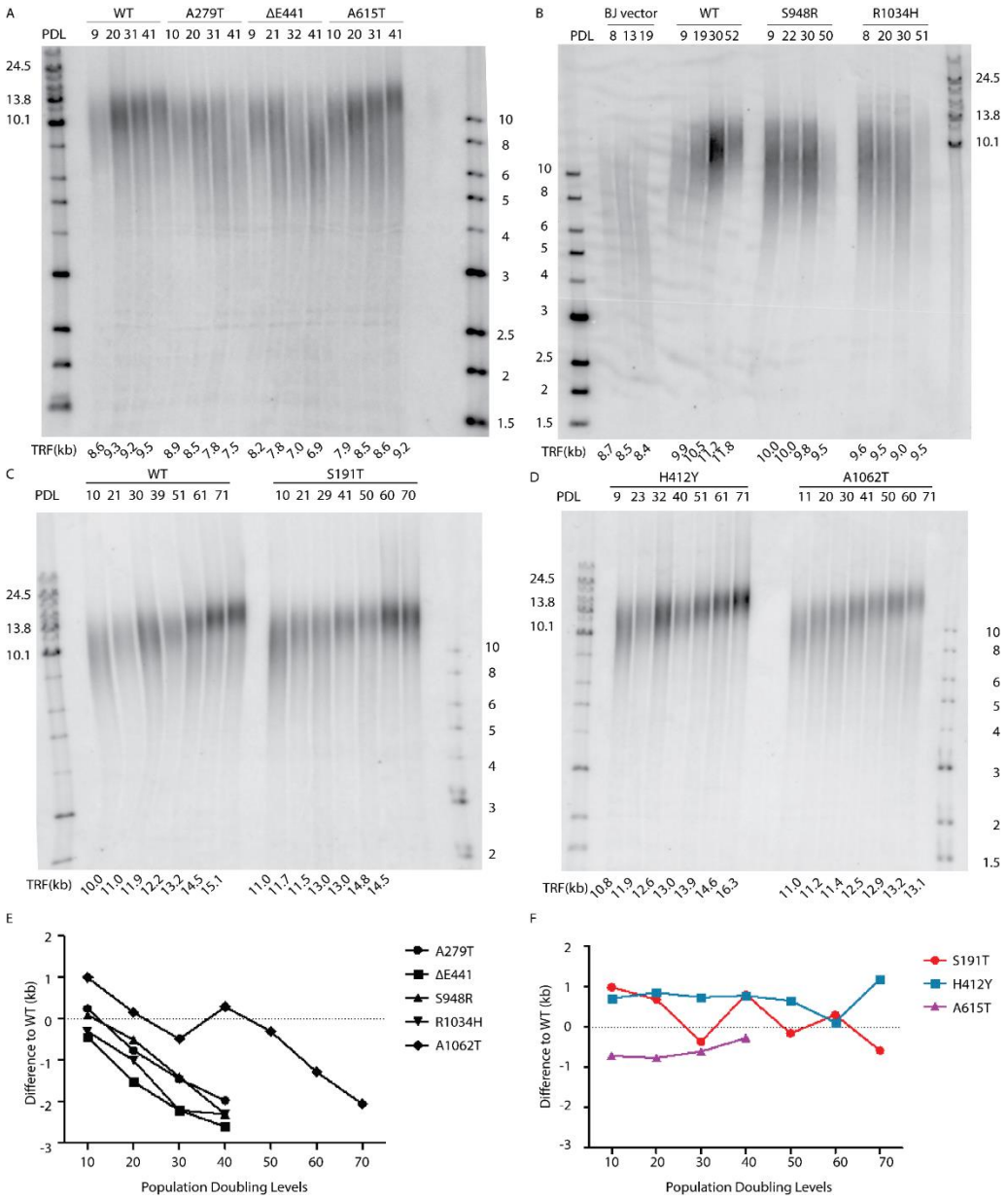

Figure 3.4 Defective telomere length maintenance was found in cells expressing 5 of 8 *TERT* variants.

(A-D) Terminal restriction fragment analysis was used to quantify telomere lengths. The measured signal represents the sum of telomeric and sub-telomeric region. Estimated telomere length was calculated by the weighted average of the lane and is shown at the bottom of the gel. All samples tested here were collected from cell populations with different doubling levels. The PDL for each lane is also shown on the top of the gel. (E) Five out of eight *TERT* SNPs showed telomere length maintenance defects compared to WT-TERT. (F) The other three TERT SNPs showed comparable telomere length increment to WT.

3.4.2 The two *TER* SNPs had normal telomere length maintenance capacity

Previous studies in our lab had confirmed that recombinant expression of both WT-TERT and WT-TER could restore normal telomerase activity and telomere lengthening in fibroblasts carrying *DKC1*-mutations (250), whereas the expression of WT-TERT and TER constructs harboring disease-associated TER mutations led to reduced telomere length maintenance (310). We used site-directed mutagenesis to introduce the two SNPs G58A and G228A to *TER* and expressed them in the TA1061 TERT-immortalized cell line. The TA1061 fibroblast cell line features a *DKC1* mutation A353V and therefore has reduced capacity to stabilize TER level at steady state (251). Stable genomic integration of TER variants through infection with retrovirus vectors (pBabe-puro-U3-TER) in TA1061-TERT immortalized cells increased the basal telomerase activity (Figure 3.5). An examination of recombinant TER expression revealed that there were comparable TER levels at steady state in these cell lines, whereas endogenous TER with mutant dyskerin in the parent TA1061-TERT cell line showed evident defect of reduced TER levels (Figure 3.6A). The population growth kinetics of the TA1061 cell lines expressing different TER variants and the vector control followed typical logarithmic curves (Figure 3.6B).

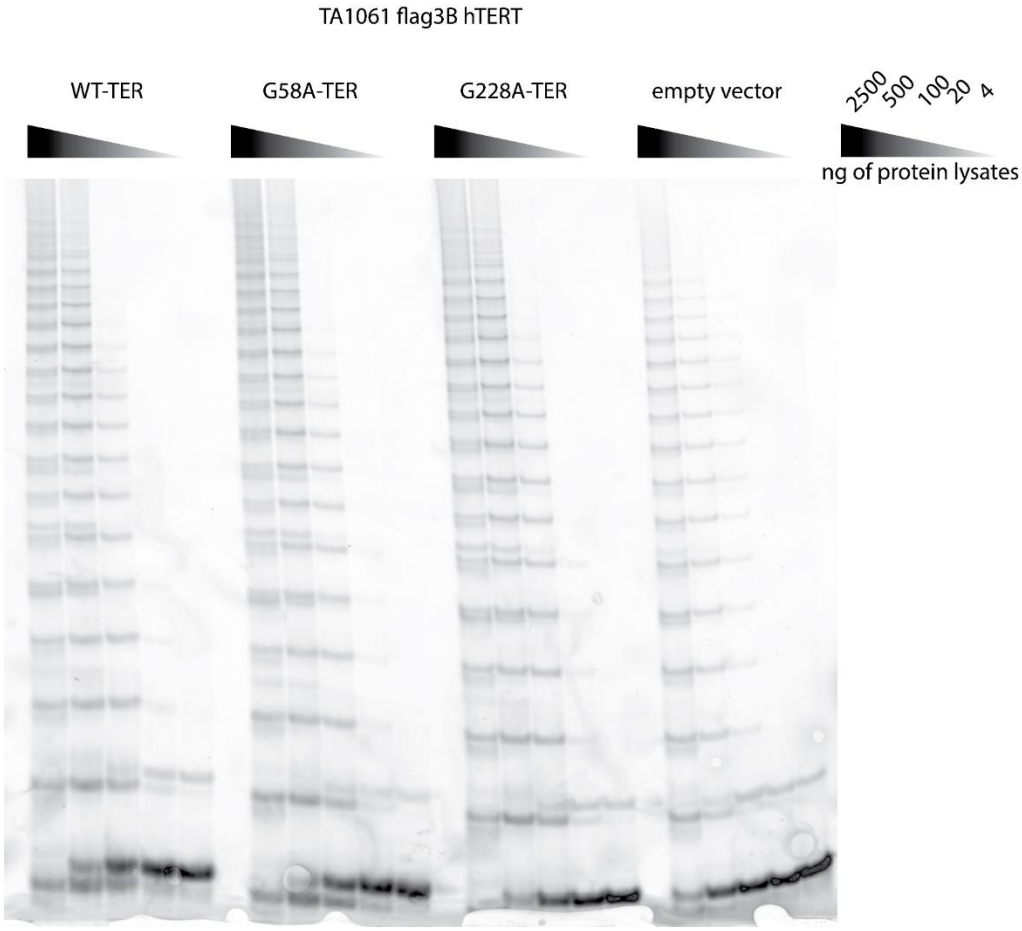

Figure 3.5 Stable integration of WT/variant-TER in TA1061-TERT immortalized cell line increased the basal telomerase activity level.

The gel-based telomeric repeat amplification protocol (TRAP) was applied to measure and compare telomerase activity in the TA1061-TERT immortalized cell line with recombinant expression of WT/variant TER. There were comparable signals in cell lines expressing WT, G58A and G228. All these three lines showed a 2.5 – 5 fold increase in telomerase activity compared to the TA1061-TERT immortalized cells.

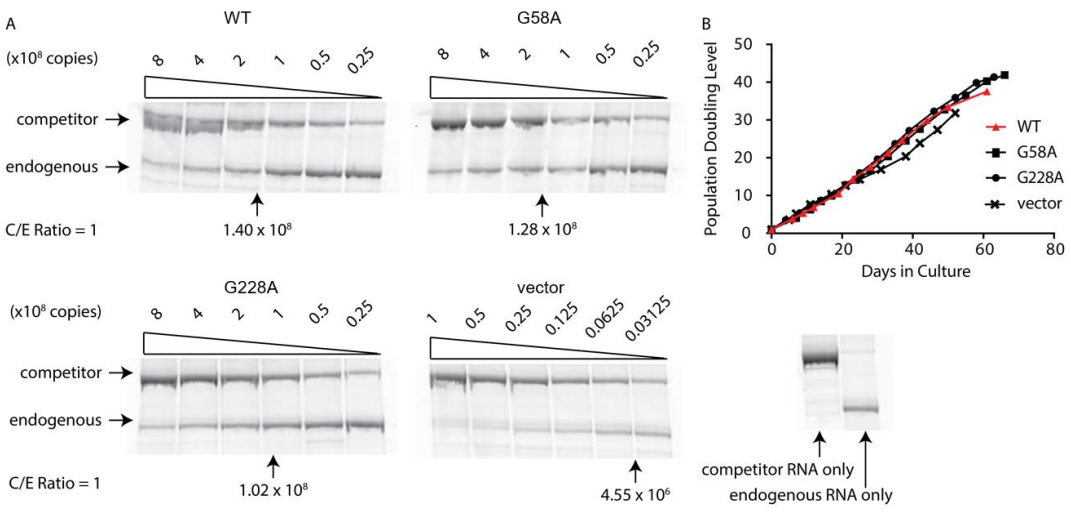

Figure 3.6 Restoration of recombinant telomerase mediated by the TER variants in TA1061-TERT cell lines.

(A) Steady-state TER levels were measured in TA1061-TERT fibroblasts expressing recombinant TER using a previously published competitive RT-PCR assay (see Section 2.3.11 for details). Input-tagged competitor RNA copy numbers are indicated for each series. The arrows indicate the approximate competitor inputs where 1:1 ratio of competitor to endogenous TER RT-PCR signals were obtained. For the quantification experiment, each RNA sample was measured first with a 10-fold serial dilution range of input competitor, then repeated with an adequate 2-fold serial dilution range for final copy number calculation. (B) Cellular growth kinetics of the TA1061-TERT cell lines expressing recombinant SNP-version of TER.

Despite the increase in the basal suboptimal telomerase activity by recombinant expression of WT-TERT in the TA1061 cell line, there was still progressive loss of telomeric repeats in the absence of overexpression of TER (refer to vector control in Figure 3.7). TA1061-TERT cell lines expressing G58A and G228A TER variants showed moderate telomere lengthening as a function of proliferation that is comparable to WT-TER (Figure 3.7B). Therefore, the two *TER* SNPs did not show any functional defects in our cellular model. It is worthwhile to note that the structural changes in TER may affect telomere catalysis in a much more subtle context, which

may not have been revealed with our cell model system as the overexpression of recombinant TER could compensate for the catalysis defects (250).

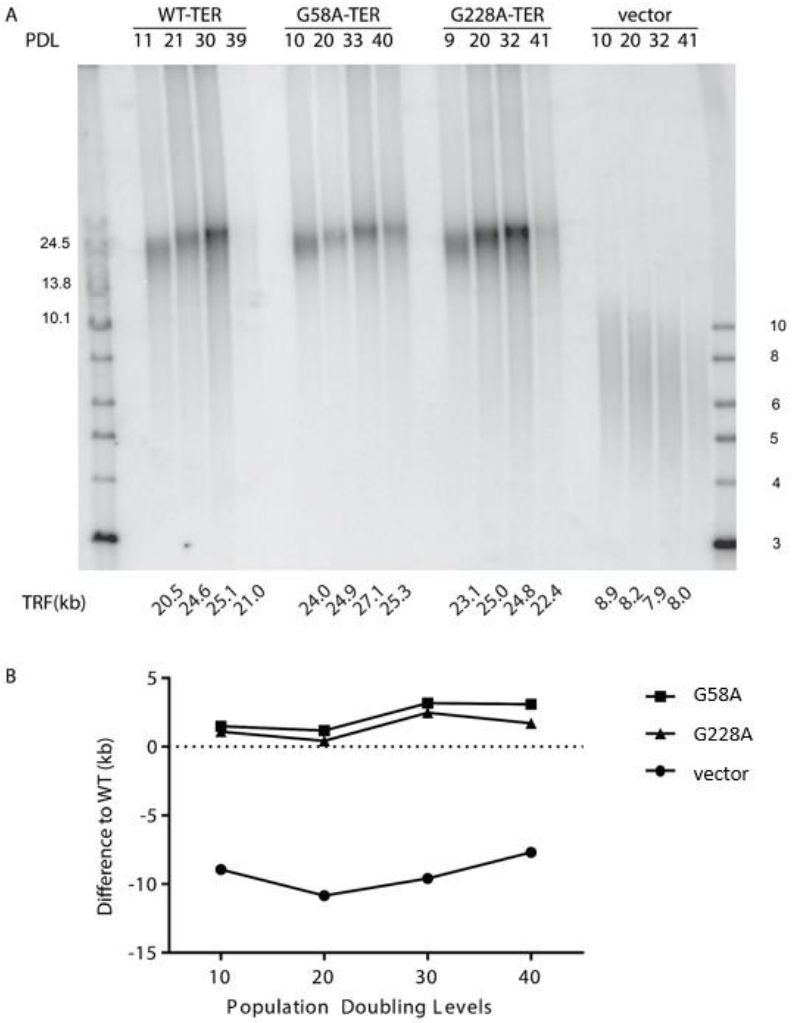

Figure 3.7 No telomere length maintenance defects were found in TA1061-TERT cell lines expressing *TER* SNPs.

(A) Terminal restriction fragment analysis was used to quantify telomere lengths. The measured signal represents the sum of telomeric and sub-telomeric region. Estimated telomere length was calculated by the weighted average of the lane and is shown at the bottom of the gel. All samples tested here were collected from cell populations with different doubling levels. The PDL for each lane is also shown on the top of the gel. The lane in WT-TER with PDL 39 is underloaded. (B) The two *TER* SNPs showed normal telomere length maintenance capacity compared to WT-TERT.

3.4.3 Defective telomere length maintenance of *TERT* SNPs is caused by reduced telomerase nucleotide addition processivity

To understand the mechanism behind defective telomere length maintenance by TERT variants, we measured telomerase activity *in vitro* for the cell lines expressing A279T, ΔE441, S948R, R1034H, and A1062T-TERT.

We applied the primer extension assay to check single nucleotide incorporation (nucleotide addition processivity, NAP) and repeat synthesis (repeat addition processivity, RAP) using cell extracts from WT-TERT and TERT variant cell lines. In this assay, an 18 nt telomeric oligonucleotide primer ending with TTA is extended in the presence of [α-^{32}P]-dGTP. With this specific primer, telomerase incorporates a single radiolabeled guanidine, stalls and translocates as it reaches the end of the TER template region. As a result, the characteristic 6 nt telomerase repeat pattern is observed, starting at TTA-primer +1, +7, +13, and so on (Figure 3.8A). Within each repeat, the guanosines reside in TTA-primer +5 and +6 positions are also radiolabeled at the α phosphate position, yet these guanosines are in the middle of the templated repeat and the chance of telomerase stalling at these locations is low. As such, they do not show signals as strong as guanosines at the TTA-primer +1 do. The activities are normalized to a radioactive end-labeled 15 nt oligo recovery control so that the DNA extraction/purification efficiency in different samples are comparable.

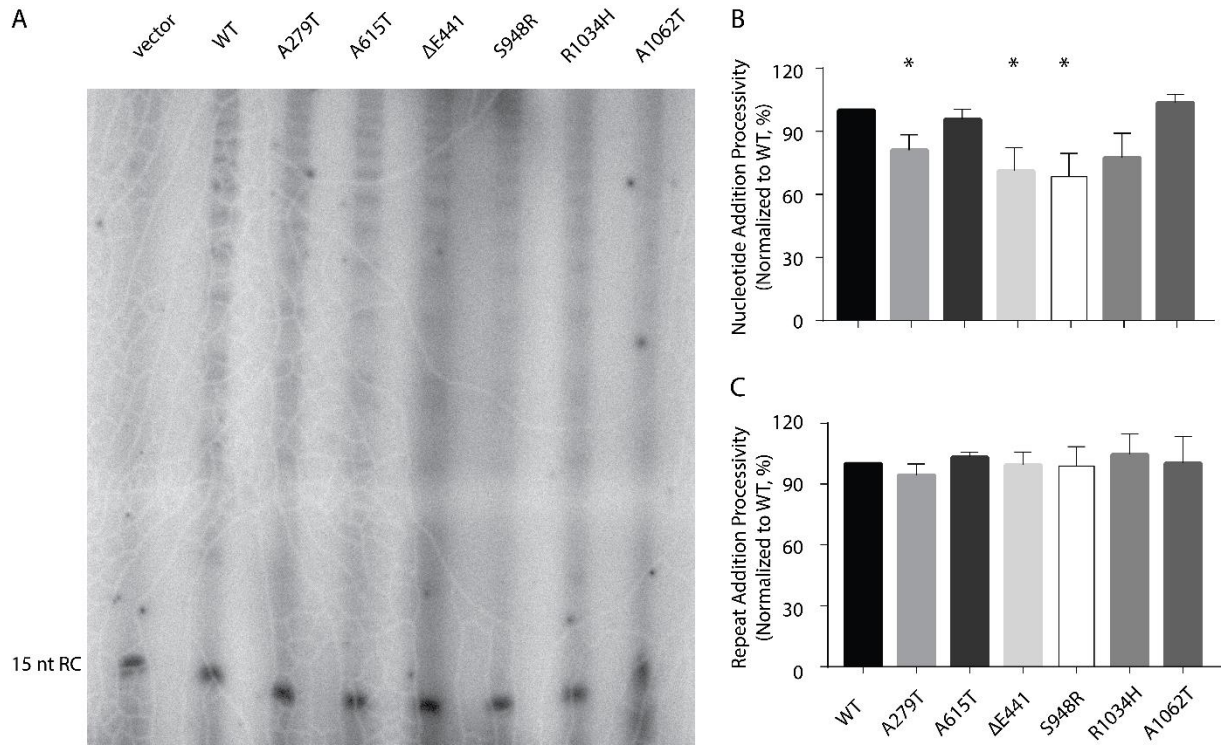

Figure 3.8 *TERT* SNPs affect telomerase activity and processivity in a SNP-specific manner.

(A) Expression of variant TERT proteins in BJ cell lines in the primer extension assay. RC, recovery control. (B) Nucleotide addition processivity in BJ cell lines expressing TERT variants. (C) Repeat addition processivity in BJ cell lines expressing TERT variants. Data are presented as mean ± SEM (n=6).

The NAP and RAP activities mediated by the TERT variants were normalized to WT and illustrated in Figure 3.8B & Figure 3.8C and summarized in Table 3.4. The telomerase from all the genetically-modified BJ cell lines under investigation showed comparable RAP. However, telomerase from three of the five cell lines (BJ cell lines expressing A279T, ΔE441, and S948R) showed slight but statistically significant decreases in their NAP activities. For example, the RAP activity of the A279T-TERT was 94.4 ± 5.5% compared to the WT-TERT ($n = 6$, $p = 0.3313$). Its NAP activity was 81.1 ± 7.3% of WT ($n = 6$, $p = 0.0323$). The overall trend showed

that the TERT protein encoded by A279T had similar repeat addition capacity as WT-TERT, but reduced nucleotide addition processivity. For the BJ cell line expressing R1034H-TERT and A1062T-TERT, we did not find any functional defects in its telomerase activity.

Table 3.4 Summary of NAP and RAP activity for the selected *TERT* SNPs

Variant	Nucleotide Addition Processivity[$]			Repeat Addition Processivity[$]		
	Mean	SEM	*P* Value	Mean	SEM	*P* Value
A279T	81.05%	7.33%	0.032[*]	94.41%	5.52%	0.331
ΔE441	71.16%	11.22%	0.022[*]	99.41%	6.48%	0.929
A615T	95.78%	4.77%	0.347	103.2%	2.50%	0.110
S948R	68.49%	11.04%	0.014[*]	98.97%	9.45%	0.915
R1034H	77.43%	11.77%	0.065	104.50%	10.28%	0.667
A1062T	103.80%	3.82%	0.224	100.20%	13.26%	0.985

[*] Both NAP and RAP are normalized to WT-TERT. These data are calculated based on six biological repeats.
[$] $p < 0.05$.

3.4.4 rs10069690 induced expression of a splicing variant acted as a dominant negative inhibitor of TERT activity

As introduced in Section 3.1.1, the *TERT* intronic SNP rs10069690 is potentially implicated in TBDs through the increased expression of TERT ASV INS1b. We cloned the corresponding coding sequence for the predicted INS1b open reading frame and replaced the full-length TERT sequences with it in WT-TERT expressing vector. To confirm the biological effect of INS1b as a dominant negative competitor for TERT activity, we expressed equal amounts of WT-TERT + empty vector, INS1b + WT-TERT, and INS1b + empty vector in TERT-negative GM00847 fibroblast to test the *in vivo* effects of INS1b expression on WT-telomerase activity. As expected, expression of INS1b protein itself did not restore any telomerase activity. In contrast, WT-TERT expression in the presence of empty vector (which does not contain any TERT open reading frame) restored telomerase activity as shown by the signature 6 bp ladders in the TRAP assay

(Figure 3.9). When both INS1b and WT-TERT polypeptides were present, they could compete against each other for the access to limited copies of endogenous telomerase RNA expressed in the GM00847 cell line, resulting in reduced telomerase activity by TRAP assay when compared with the WT-TERT+ empty vector sample (Figure 3.9). Our data agreed with the previous report that the spliced variant INS1b-TERT acted as a dominant negative inhibitor of telomerase activity (122).

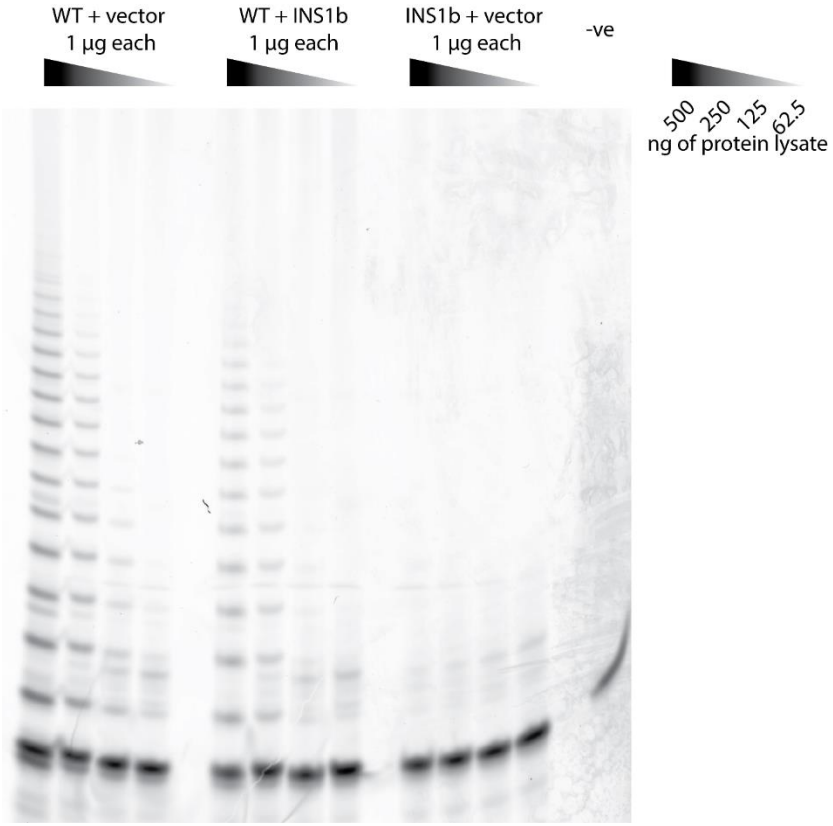

Figure 3.9 The INS1b transcript generated by rs10069690 is a dominant negative inhibitor of telomerase activity.

A dilution series of whole cell lysate from each transfected cell line was used to check the amount of protein lysate required for the restoration of telomerase activity in the *in vitro* TRAP assay. There was similar telomerase activity restored with 250 ng of WT/vector-expressing cell line and with 500 ng of WT/INS1b-expressing cell line. The INS1b competed against WT-TERT in the GM00847 cell line for endogenous TER binding. Therefore, more protein lysates were needed to restore the same level of telomerase activity.

3.4.5 Carriers of TERT variant A279T were more prevalent in patients with rapid COPD progression

Next, we asked whether deficient telomere repeat synthesis and length maintenance mediated by common TERT variants in our *ex vivo* study could affect disease progression in COPD patients. We selected a catalytically defective *TERT*-SNP with the highest MAF (rs61748181, p.A279T) in the clinical association study.

It was known that emphysema patients had short telomeres and accelerated senescence in their AEC2 and endothelial cells (242). Specifically, there was a positive correlation between senescence in AEC2 cells and airflow limitation, i.e., senescence was related to disease severity (242). In a mouse model, it was also shown that short telomeres could trigger alveolar stem cell failure, leading to the regenerative defects and inflammatory responses which are frequently observed in emphysema (320). Since disease severity, measured by FEV1% of predicted, and leukocyte telomere length measurement were both subject to variations during the period of measurement, we decided to use disease progression (measured by the rate of decline of FEV1 over 5 years) as an alternative to disease severity. In fact, in the LHS cohort, disease progression was positively correlated to disease severity in the 5^{th} year of study ($r = 0.57$, $p < 0.0001$), but was weakly correlated to that in the 1^{st} year of study ($r = 0.12$, $p < 0.0001$) (Figure 3.10). We hypothesized that patients with rapid disease progression had stem cell exhaustion and that the maximal contribution to telomere uncapping by telomerase dysfunction (in the alveolar cells) could be revealed in that population.

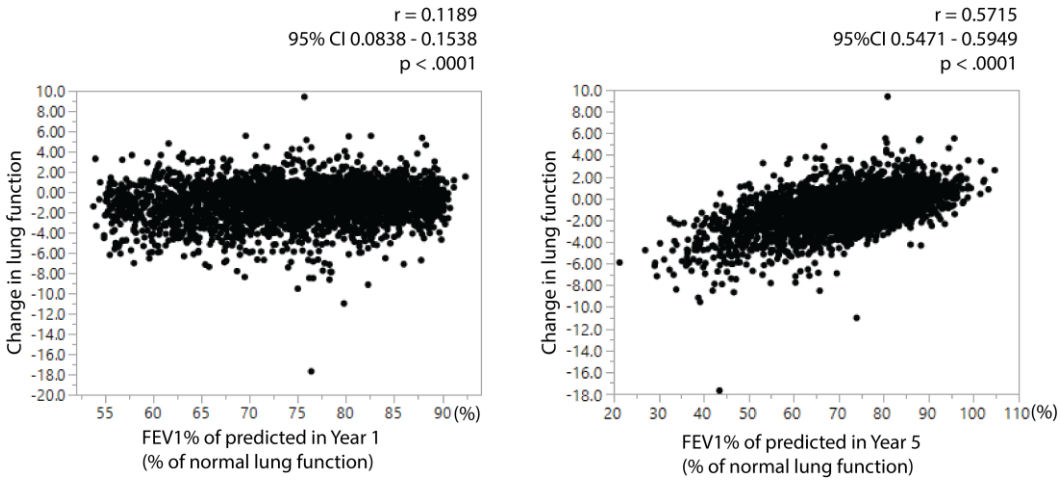

Figure 3.10 The relationship between COPD progression and disease severity.

FEV1% of predicted value is calculated by the FEV1% of the individual divided by the average FEV1% in the population for people of similar age, sex, and body composition. In other words, this refer to an individual's normalized lung function by age, sex and body size. According to the COPD GOLD guidelines (2018 edition), mild, moderate, severe, and very severe COPD are defined by FEV1% of predicted value over 80%, between 50 – 79%, 30 – 49%, and less than 30%, respectively. In year 1 of LHS, all participants had mild to moderate COPD, whereas in year 5 of LHS, more than a half of the participants showed moderate to severe COPD. The progression of COPD (measured by average change in lung function per year) is more strongly associated with disease severity (measured by FEV1% of predicted value) in the 5th year of study, than in the 1st year of study.

In our *ex vivo* study, we did not observe any growth defects up to 60 PDLs (Figure 3.3C). However, from the earliest time point following polyclonal culture establishment, BJ cells expressing TERT variants had shorter average telomere length compared to those expressing WT-TERT (Figure 3.4E). The potential clinical implications of these data were that in a subpopulation of COPD patients who are carriers of the defective allele of SNP rs61748181 would have a lower capacity to replenish telomere loss when compared with the homozygous WT-TERT patient population. Hence, these SNP rs61748181 carrier patients were predicted to progress faster to stem cell exhaustion than patients carrying only WT-TERT sequence, when controlled for all other known risk factors in COPD disease progression. To test this hypothesis,

we set out to explore the effect of TERT-SNP in patients with extremely rapid disease progression. We sorted all the LHS study participants by their rate of decline of FEV1 over 5 years and selected equal numbers of rapid- and non-decliners from each end of the spectrum. The characteristics of these selected study participants are shown in Table 3.5. In a subset of 464 individuals (15% of the LHS population), those who were rs61748181 heterozygotes or minor allele homozygotes were more prevalent among rapid decliners (18/214 = 7.76%) than non-decliners (7/225 = 3.02%) (adjusted OR = 2.49, 95% CI 1.01 – 6.14, p = 0.038) (Table 3.6).

Table 3.5 Characteristics of the study participants in the rapid and non-decliner groups.

	Rapid Decliners		Non-decliners		
N	232		232		
Male, %	136	58.62%	154	66.38%	
	Mean	SD	Mean	SD	*P* value Rapid vs Slow
Age	49.58	6.53	47.45	6.81	0.0006
Pack years	42.80	18.67	38.43	17.83	< 0.0001
FEV1 (% of predicted value) at baseline	74.95	9.53	79.72	7.84	< 0.0001
FEV1 (% of predicted value) after 5 years	58.68	13.20	82.77	9.30	< 0.0001
ΔFEV1 (% predicted/year)	-3.59	1.41	0.45	1.02	-

Table 3.6 The odds ratio of *TERT* SNPs rs61748181 and rs10069690 among rapid versus non-decliners in the study group.

SNP		rs61748181	rs10069690
Rapid Decliners	SNP[$]	18	119
	WT	214	113
	%SNP	7.76%	51.29%
Non-decliners	SNP[$]	7	97
	WT	225	135
	%SNP	3.02%	41.81%
Crude OR	OR	2.70	1.47
	95% CI	1.11 – 6.60	1.02 – 2.11

	P value	0.029*	0.041*
Adjusted OR#	OR	2.49	1.39
	95% CI	1.01 – 6.14	0.95 – 2.03
	P value	0.038*	0.088

$ SNP carriers, including homozygous and heterozygous carriers of the minor allele of the SNP.
# The odds ratio adjusted for age and smoking history.
* $p < 0.05$.

Next, we tested the association between rs61748181 and disease progression in 5 years in the entire LHS population by a multivariate linear regression model. The demographic profiles of the study group are shown in Table 3.7. We could not determine the genotypes for 516 out of 3,741 samples (13.79%). When we assessed the demographic profile of these individuals, we noted that there were slight differences in age and cigarette consumption, but otherwise unremarkable changes in baseline characteristics between the typed and undetermined participants (Table 3.8). The MAF of rs61748181 in the Caucasians from the LHS population was 3.29%, which is in agreement with the known MAF of 3.6% in Caucasians for this SNP (282). The locus showed a slight deviation from Hardy-Weinberg equilibrium ($p = 0.040$), with an increase in homozygotes for the minor allele than expected. The deviation agrees to a recent report on the excess of homozygotes for the rs61748181 minor allele in COPD patients than in control subjects (283). In addition to the template-free and known-genotype control samples in each experiment, to ensure that the deviation was not driven by genotyping errors, we typed 5% of the samples again to check the repeatability of the genotyping assay. Genotypes were 100% reproducible in these samples.

Table 3.7 The demographic profile of the LHS population (Caucasians).

	WT		rs61748181		
Number of subjects	2899		197		
Age range, min – max	34 – 62		35 – 59		
Gender, male N (%)	1957 (67.5)		125 (63.5)		
Continuous smokers, N (%)	1587 (54.7)		119 (60.4)		
Intermittent quitters, N (%)	861 (29.7)		45 (22.8)		
Sustained quitters, N (%)	451 (15.6)		33 (16.8)		
	Mean	SD	Mean	SD	*P* value*
Age (years)	48.43	6.71	48.49	6.69	0.92
Pack years	40.10	17.97	41.95	18.69	0.18
FEV1 (% of predicted value) at baseline	78.26	9.09	78.94	9.00	0.28
FEV1 (% of predicted value) after 5 years	74.90	12.35	75.32	12.65	0.62

* Unpaired student's t test between WT and carriers of minor allele of rs61748181.

Table 3.8 Demographic profile of the entire cohort, with comparisons between genotyped and undetermined samples.

	Entire cohort		Genotyped		Undetermined		
Number of subjects (%)	3741		3225 (86.2)		516 (13.8)		
Age range, min - max	34 – 67		34 – 62		35 – 67		
Gender, male N (%)	2367 (63.3)		2026 (62.8)		341 (66.1)		
Race, Caucasian N (%)	3596 (96.1)		3096 (96.0)		500 (96.9)		
Continuous smokers, N (%)	2067 (55.2)		1787 (55.4)		280 (54.3)		
Intermittent quitters, N (%)	1075 (28.7)		935 (29.0)		140 (27.1)		
Sustained quitters, N (%)	598 (16.0)		502 (15.6)		96 (18.6)		
	Mean	SD	Mean	SD	Mean	SD	*P* value*
Age (years)	48.53	6.69	48.44	6.71	49.23	6.23	0.013
Pack years	40.5	18.3	40.22	18.02	41.83	19.92	0.016
FEV1 (% of predicted value) at baseline	78.32	9.05	78.25	9.07	78.75	8.91	0.23
FEV1 (% of predicted value) after 5 years	74.97	12.34	74.83	12.36	75.91	12.19	0.060
ΔFEV1 (% of predicted value) over 5 years	-0.99	1.89	-1.00	1.86	-0.90	1.77	0.25

* Unpaired student's t test between participants with and without genotype information.

While age and smoking (both smoking history in pack-years and smoking status during the 5 years of the LHS study) affected disease progression, rs61748181 status did not significantly affect the rate of decline of FEV1 in the entire cohort. There was no significant difference in the rate of decline of FEV1 between the SNP carriers (n = 197) and those with the WT genotype (n = 2,899) (Table 3.9).

Table 3.9 Statistical analysis for changes in FEV1 over 5 years (rate of decline) in rs61748181 and WT carriers.

	WT	rs61748181
Median	-0.900	-0.836
Mean (95% CI)	-0.999	-0.956
Std. Deviation	1.981	2.774
Std. Error of Mean	0.037	0.198
95% CI	-1.072 – -0.927	-1.347– -0.565
P value*	0.89	

* Unpaired student's t test between WT and carriers of minor allele of rs61748181.

Our data suggest that the functional effect of the minor allele of rs61748181 was only evident in those individuals with the most extreme phenotype. Considering the telomere maintenance defect observed in the A279T-BJ cell line over long term in our cell model, we investigated the association between disease progression (stratified by SNP carrying status) and short telomeres. Caucasian patients from the entire cohort with measurable genotypes (n = 3096) were divided into four groups based on their disease progression (Figure 3.11A). In each group, the rate of decline in lung function for the SNP carriers were compared to their WT counterparts. SNP carriers (n = 52) in the 4^{th} quartile group (where the patients had the fastest disease progression) tended to have shorter leukocyte telomere lengths than WT patients (n = 722), although statistical significance was not reached (p = 0.1784) due to high intra-group variance and imbalanced sample size between the two genotype groups. The difference in telomere length between SNP and WT carriers diminished in the upper 2^{nd} and 1^{st} quartile groups (Figure 3.11B). The blood samples for leukocyte telomere length measurement were collected in the fifth year of the LHS.

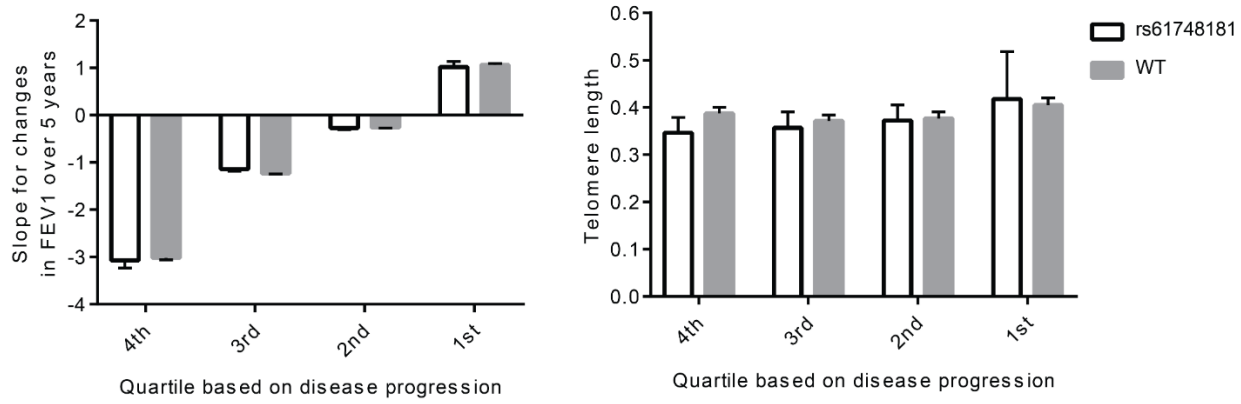

Figure 3.11 Rapid decliners carrying rs61748181 had short telomeres before they showed disease progression.

(A) All patients in the LHS were divided into four subgroups. The slope for changes in FEV1 over 5 years (rate of decline) for each group is shown as mean ± SEM. (B) Leukocyte telomere lengths for SNP and WT carriers in the four subgroups are shown as mean ± SEM.

3.4.6 rs10069690 status does not affect COPD progression

We examined the genotype distribution of rs10069690 in rapid- and non-decliners. The MAF of rs10069690 in Caucasians from the studied population was 26.37%. It conforms to Hardy-Weinberg equilibrium ($p = 0.505$). Given the high MAF, almost half of the studied population (45.8%) were heterozygotes or homozygotes for the minor allele of rs10069690.

Study participants with one or two copies of the rs10069690 minor allele did not show increased prevalence among rapid decliners (119/232 = 51.29%) than non-decliners (97/232 = 41.81%) (adjusted OR = 1.39, 95% CI 0.95 – 2.03, $p = 0.088$) (Table 3.6).

3.4.7 The combined clinical effects of rs61748181 and rs10069690

The rs61748181 and rs10069690 SNPs exhibited minimal linkage disequilibrium ($r^2 = 0.04$, $p < 0.0001$) in the LHS cohort. Knowing that there were more carriers of the minor allele of rs61748181 in the rapid-decliners group, we further asked whether carrying two minor alleles would show an additive risk of rapid COPD progression. Compared to people that did not have

either minor allele, those with both minor alleles were not significantly increased among rapid decliners (14/130 = 10.77%) than non-decliners (6/136 = 4.41%) (adjusted OR = 2.39, 95% CI 0.88 – 6.49, p = 0.077) (Scenario 1 in Table 3.10). Similarly, compared to patients who had zero or one minor alleles, those with two minor alleles were not significantly increased among rapid decliners (14/232 = 6.03%) compared with non-decliners (6/232 = 2.59%) (adjusted OR = 2.26, 95% CI 0.84 – 6.07, p = 0.092) (Scenario 2 in Table 3.10).

Table 3.10 Combined analysis for the odds ratio of *TERT* SNPs rs61748181 and rs10069690 among rapid versus non-decliners in the study group.

	Scenario 1		Scenario 2	
Rapid Decliners	Double SNP carriers*	14	Double SNP carriers*	14
	Noncarriers	116	Carriers of one SNP and noncarriers	218
	% Double SNP carriers	10.77%	% Double SNP carriers	6.03%
Non-decliners	Double SNP carriers*	6	Double SNP carriers*	6
	Noncarriers	130	Carriers of one SNP and noncarriers	226
	% Double SNP carriers	4.41%	% Double SNP carriers	2.59%
Crude OR	OR	2.62	OR	2.42
	95% CI	0.97 – 7.03	95% CI	0.91 – 6.41
	P value	0.057	*P* value	0.076
Adjusted OR #	OR	2.39	OR	2.26
	95% CI	0.88 – 6.49	95% CI	0.84 – 6.06
	P value	0.077	*P* value	0.092

* SNP carriers, including homozygous and heterozygous carriers of the minor allele of both SNPs rs61748181 and rs10069690.
# The odds ratio adjusted for age and smoking history.

3.5 Discussion

We investigated the biological and clinical effects of single nucleotide polymorphisms in the *TERT* gene which have previously been linked to telomere biology disorders. Using a

combination of molecular and cell biology tools in the laboratory, we have demonstrated that five out of eight nonsynonymous *TERT* variants and the intronic *TERT* SNP rs10069690 could lead to deficient telomere length maintenance. With clinical samples collected in the LHS, we also demonstrated that carrier status of the SNP rs61748181 aggravated existing disease conditions in COPD patients. These results implicated *TERT* as a biologically plausible candidate gene for lung function decline and COPD pathogenesis. Not only did our results agree with the previous observation that deleterious mutations in *TERT* contribute to COPD pathogenesis (233), we also extended this observation to demonstrate that even common variants in *TERT* that are likely to have low penetrance may nonetheless modify disease presentations.

Our *ex vivo* results demonstrated in a pure laboratory setting that, when other confounding factors for telomere attrition were fixed, fibroblasts with recombinant expressions of the disease-associating TERT variants were predetermined to suboptimal telomere maintenance. However, deficient telomere length maintenance in this context did not translate to a loss of proliferative capacity, as BJ cells expressing the minor allele of rs61748181 (A279T) (as well as the other four variants) were kept in culture for up to 50 – 60 PDLs, the limit of fibroblast divisions during a lifetime (321), and demonstrated the same growth rate as WT-TERT expressing cells. To our knowledge, this was the first study to investigate the long-term effects of common TERT variants on telomeres in primary cell lines. Our results confirmed that the short telomere risk-associated SNPs encoded TERT variants A279T, ΔE441 and A1062T were deficient in telomere length maintenance. We provided experimental evidence to support the previous clinical observations that in subjects with bone marrow failure, carriers (heterozygous and homozygous) of these variants had shorter telomeres than their age-matched peers (274). Our investigation of their biochemical properties revealed that while the variants' RAP activities were intact, both

A279T and ΔE441 displayed statistically significant lower NAP activity compared to WT-TERT. These biochemical results are also supported by reports in different cellular and experimental systems (281, 284). Despite the observed telomere length shortening deficiency in long term cell culture, we did not find any defects for S948R, R1034H and A1062T in their RAP or NAP activity through the *in vitro* primer extension assay. All three variants are located in the CTE domain of TERT, which was known to be implicated in TERT nuclear localization (103) and telomeric DNA engagement of the holoenzyme (104). It is thus conceivable that these three variants may affect TERT trafficking between nucleus and cytoplasm and/or reduced telomerase holoenzyme binding to telomeres *in vivo.*

Interestingly, in the cell model over long-term culture (up to 70 PDLs), we also revealed that short telomere risk-associated TERT variant H412Y had similar telomere length maintenance as WT-TERT. Thus, the frequent identification of H412Y-TERT in association studies with various TBDs was not supported by our *ex vivo* investigation. While H412Y-TERT itself does not cause short telomeres, it remains possible that this rare variant is in linkage disequilibrium with another yet unidentified variant that negatively impacts telomere length maintenance. This unidentified linkage disequilibrium may then contribute to the identification of H412Y-TERT in disease association studies.

This was also the first study to show the association of rs61748181 minor allele with increased rate of decline of lung function in a selected patient group from a large cohort study of COPD (adjusted OR 2.49, 95% CI 1.01 – 6.14). Regression analysis showed that the association was independent of known risk factors such as age and smoking history. Telomeres shorten at similar rates in blood and somatic tissues (322). In a previous study, telomere lengths in leukocytes were correlated to those in lung tissues (323). Therefore, our measurement of

leukocyte telomere lengths in COPD patients should be representative of the lengths in their lung tissues. Despite the lack of significant, our observation of relatively shorter telomeres in homozygous or heterozygous carriers of rs61748181 minor allele than in WT patients was consistent with the rs61748181 minor allele contributing to shorter telomeres in the *ex vivo* human cell model. The observation further supported the cellular senescence-based pathogenic model of COPD (272) that the carriers' stem cell compartments would remain proliferative until their telomeres were too short and trigger stem cell exhaustion.

The SNP rs10069690 was not correlated to COPD progression. The genetic effects were not statistically significant and were less than the effects observed with rs61748181. When considering a biological explanation, the INS1b transcript, the alternatively spliced TERT product whose expression is proposed to be induced by the presence of rs10069690, is known to have cell- and tissue-specific expression through a complex regulatory mechanism (122). The carrier status of rs10069690 does not necessarily mean there will be equal or even any expression of the INS1b isoform in the carriers' AEC2 cells. Thus, the dominant negative effects of INS1b are not expected to follow a Mendelian inheritance pattern, which may explain the lack of effect on COPD progression.

An individual's telomere length is determined by multiple factors, including genetics, inherited telomere length, and environmental influences. While the effects of environmental insults, such as smoking or infection, are likely subject to huge inter-individual and temporal variations, the effects represented by inherited telomere length and from the genotype of telomere-related genes are fixed. The comprehension of these fixed effects could facilitate understanding of the pathogenesis and progression in complex disorders.

In well-characterized telomere biology disorders such as DC, highly penetrant disease-causing mutations in telomere biology genes mostly present molecularly as rapid telomere shortening followed by replicative senescence in fast turnover tissues, leading to clinical phenotypes including bone marrow failure, immunodeficiency, and skin hyperpigmentation. While suboptimal telomere synthesis occurs in parallel in slow turnover tissues from these patients, the rate of telomere attrition will not exceed that observed in fast turnover tissues such as bone marrow. Consequently, patients with exceedingly short telomeres succumb to diseases of high turnover tissues (such as aplastic anemia and epithelium presentations), before the defects in their slow turnover tissues (i.e. pulmonary fibrosis in the lung and cirrhosis in the liver) have a chance to develop and accumulate to a clinically relevant extent. It is noteworthy that due to the variable expressivity of telomere biology disorders, once patients carrying highly penetrant mutations survived bone marrow failure and immunodeficiency at early ages, they very often will present with lung and liver dysfunctions in middle adulthood (195). Hence, the fixed variable – short telomeres, caused by rare mutations – remains the major driver of these various clinical presentations from telomere biology disorders.

In contrast, for carriers with genetic variants of low penetrance in telomere biology genes, as demonstrated in our *ex vivo* cell model using *TERT* genetic variants, there is still accelerated telomere shortening due to the genetic changes. In this case, the moderately accelerated telomere attrition rate is tolerated, as the functional telomerase enzyme of reduced activity mediated by the nonsynonymous variant is still able to protect against the premature onset of proliferative senescence, within a human lifespan. Carriers of these low penetrance variants remain asymptomatic when the telomere length maintenance defects keep accumulating. Such individuals are not expected to show any profound clinical effects in the absence of other fixed

and/or random factors that may accelerate cellular turnover, such as exposure to cell death stimulants including smoking or recurrent infections. In these situations, when carriers of *TERT* SNPs are in constant need of telomere replenishment to maintain tissue integrity, the defects in telomerase catalysis would be exacerbated. Over time, carriers of *TERT* SNPs may experience accelerated worsening of existing pathological conditions, as a function of the sum total of all etiologic factors that contributed to the loss of tissue integrity, leading to earlier onset or progression of multiple aging-related diseases, such as cancer (37), idiopathic pulmonary fibrosis (272), or COPD (272).

This proof-of-concept study showed in a pure laboratory setting that common genetic variants in *TERT* conferred telomere length maintenance defects. In clinical settings, variants associated with mild penetrance in telomerase deficiency did not lead to cellular senescence in the absence of other confounding factors, and will not, on their own, drive the onset of chronic degenerative disorders. However, in the presence of accelerated cellular turnover, resulting from exposure to chronic environmental irritants and/or other genetic perturbations that affect tissue-specific functions, the presence of these genetic variants in the telomere biology pathway would be combined, resulting in faster telomere attrition. The effect of mildly deficient genetic polymorphisms in telomere biology genes can be seen as modifiers in accelerated COPD disease progression, when the pathogenetic process leads to excessive demand in lung tissue regeneration. Although the understanding of genetic profile in telomere biology genes does not predict one's disease susceptibility, we contend that a thorough understanding of all the determinants that affect cellular regenerative capacity is important in the prognosis and management of all existing chronic tissue degenerative disorders.

In summary, we investigated telomere length maintenance effect for a panel of variants in TERT and TER in *ex vivo* cellular models. While the two *TER* SNPs (G58A and G228A) and three TERT variants (S191T, H412Y, and A615T) showed comparable telomere lengthening to WT-TERT, five variants in TERT (A279T, ΔE441, S948R, R1034H, A1062T) showed defective telomere maintenance. For three of them (A279T, ΔE441, S948R), the defects were secondary to reduced telomerase nucleotide addition processivity. Additionally, we selected the nonsynonymous *TERT* SNP rs61748181 (encoding variant A279T) and an intronic *TERT* SNP rs10069690 for study in a clinical cohort (LHS). The intronic SNP rs10069690 was verified to induce expression of a splicing variant that inhibited WT-telomerase activity in a dominant negative manner. COPD patients carrying *TERT* SNP rs61748181 minor allele had increased risk of rapid decline of lung function (adjusted OR = 2.49, 95% CI 1.01 – 6.14) independent of known COPD pathogenesis factors such as age and smoking history.

Chapter 4: Concluding Remarks

4.1 Summary and conclusion

In this dissertation, I focused my studies on three core telomere biology genes: *DKC1*, *TERT,* and *TER*. I studied the effects of specific genetic mutations/variations in these genes at the molecular and cellular levels and revealed their implications on the phenotypic variations observed in the clinics from carriers of these mutations/variations.

In the X-DC project in Chapter 2, I revealed extensively skewed XCI patterns in a panel of female heterozygous carriers of disease-associated *DKC1* alleles, in samples from blood and buccal cells, as well as in patient-derived primary fibroblast cultures and lymphoblastoid cell lines. My comprehensive assessment of molecular and cellular phenotypes in female *DKC1* mutation carriers showed normal telomere lengths, as well as normal dyskerin expression and function in this population. My data suggested that these female *DKC1* mutation carriers were protected from severe manifestations of DC, namely, bone marrow failure, owing to epigenetic regulation through XCI and the resultant exclusive expression of WT-dyskerin in their blood cells. I postulated that the epithelial manifestations of DC phenotypes in a subset of female carriers may be a result of multiple factors, including the interactions between mosaic (temporal and positional) expression of mutant *DKC1,* environmental and behavioral factors that influence the rate of telomere attrition as well as telomere length inheritance from affected family members.

In the COPD project in Chapter 3, I investigated telomere length maintenance capacity of telomerase variants encoded by eight *TERT* nonsynonymous SNPs and two *TER* SNPs in *ex vivo* cellular models. Defective telomere length maintenance was observed in TERT variants A279T, ΔE441, S948R, R1034H, and A1062T. None of the other three TERT variants (p.S191T,

p.H412Y, and p.A615T) nor the TER variants (G58A and G228A substitutions) showed telomere length maintenance defects in our *ex vivo* models. I further showed reduced telomerase NAP is responsible for the telomere synthesis deficiency in three out of five defective TERT variations (p.A279T, p.ΔE441, and p.S948R), while telomerase RAP for these variants remained in the normal range. For the other two TERT variants (p.R1034H and p.A1062T), their defects in telomere length maintenance cannot be explained by a reduction in catalytic activity (neither NAP nor RAP). Further investigations on substrate engagement of these variant telomerase to telomeres should be considered. In parallel, I confirmed data from a previous report on a TERT splicing variant INS1b that inhibited WT-telomerase activity in a dominant negative manner. Finally, I studied the genetic effects of rs61748181 (encoding p.A279T) and rs10069690 (potentially inducing the expression of the alternatively spliced INS1b TERT variant) in a clinical COPD cohort (LHS). COPD patients carrying rs61748181 in the LHS cohort were predisposed to rapid rate of decline of lung function.

In conclusion, I showed that genetic modifiers – both epigenetic regulation through XCI and genetic polymorphisms in telomere biology genes – are implicated in telomere length maintenance at the cellular level. Phenotypic variations of telomere biology disorders in human tissues are a result of multiple factors, including the inheritance of these genetic modifiers and other disease-associating genetic factors, the altered rate of telomere attrition due to environmental and/or behavioral factors, and inherited telomere lengths which were set during embryogenesis. Understanding the contributions of the fixed variables will benefit modeling of telomere maintenance dynamics, and inform personalized decisions on risk minimizing behaviors.

4.2 Significance and future directions

One of the primary motivations for the development of personalized medicine is to accurately estimate an individual's disease risk, computation of which necessarily includes a combination of genetic, clinical, and environmental factors (324). Since clinical (pathological/physiological) and environmental factors are very often stochastic and subject to changes over time, it is of particular interest for researchers to determine contributions of genetic factors – to build a plausible nature-nurture model for a single disease (325). While the genetic risks for thousands of Mendelian (monogenic) diseases have been assessed, the quantification of genetic risks for most complex disorders remains unsolved (324). The study of genetic modifiers on telomere length maintenance in this dissertation provides clinical evidence and biological validations for potential genetic risks conferred by a single enzyme, telomerase, involved in tissue degenerative disorders. My data reveal that variants in telomere biology genes, rare and common, contribute to phenotypic variations and they should be incorporated in the assessment of genetic risks for short telomere-related disorders.

4.2.1 The many facets of telomere homeostasis in TBDs

Telomere biology disorders, as the name suggests, derive from telomere maintenance defects. In the narrow sense, TBDs are historically considered monogenic as DC, the most well-characterized TBD that is caused by single gene mutations in one of the telomere biology genes. In the broad sense, the concept of TBDs can be extended to many common and heterogeneous chronic diseases, wherever short telomere length serves as a contributing factor in the etiology and pathogenesis of the disorder. In this dissertation, I used a cohort study of COPD as an example to demonstrate the effects of common variants in *TERT* with suboptimal telomerase catalytic capacity on disease progression. Theoretically, our understanding of the tested *TERT*

SNPs should be applicable to other functionally defective SNPs in *TERT* as well as additional telomere biology genes. The impact of *TERT* SNPs associated with suboptimal telomere length maintenance in COPD could also be applied to other chronical degenerative diseases, where there is an increase in tissue renewal demand, including, but not limited to, liver cirrhosis and osteoporosis (270).

Short telomere length – the shared etiology of TBDs in all tissues – is a sum total of the inherited starting telomere length following embryogenesis, the negative effects from cell turnover and/or telomeric DNA damages, and the positive effect from optimal telomere synthesis by telomerase when the enzyme is expressed. Following early embryogenesis (326), both intrinsic and extrinsic factors contribute to the negative and positive effects in telomere maintenance dynamics. Intrinsically, variations and/or mutations of telomere maintenance pathway genes combine with the normal functional demands for tissues of different turnover rates to determine the rate of telomere attrition in a cell-autonomous manner. On the other hand, extrinsic factors can promote rapid depletion of telomeric repeats with an increase in cellular proliferation. Some examples of these extrinsic factors include the induction of clonal expansion following active immune response to infections (327), as well as the increased repair demands from damaged/destructed bronchial epithelium induced by cigarette smoke (328) and oxidative stress (329). Consequently, inflammation induced by cellular senescence exacerbates COPD (330) and other age-related diseases (331).

In patients with highly-penetrant mutations in telomere biology genes, the deleterious mutations resulted in severe defects in telomere synthesis and rapid accumulations of short telomeres, even in the presence of minimal environmental factors that induce telomere attrition. Consequently, disease presentations in mutation carriers become a competition of telomere

shortening between fast- and slow-turnover tissues: the tissue types that have developed and accumulated exceedingly short telomeres at a clinically relevant level would show disease manifestations first. Since the bone marrow and the hematopoietic stem cells belong to this category, it can explain why severe X-DC patients often present bone marrow failure as their first symptom in the clinics. Development of clinical phenotypes in slow-turnover tissues may take decades longer, due to the nature of slow turnover rate. In these cases, clinical manifestations of slow-turnover tissue defects in severe DC patients are understandably rare, as these patients will have long succumbed to diseases of high-turnover tissue defects before they show any clinical symptoms associated with the slow-turnover tissues.

On the other hand, in patients with genetic variations of mild/moderate functional perturbations, the genetic defects in telomere length maintenance dynamics are not profound enough to induce growth disadvantage at the cell/tissue level or to trigger an early onset of disease in the body. This explains why these variations, though suboptimal telomere synthesis, are carried over generations in human population during evolution. In chronic aging-related diseases such as COPD, the onset and manifestations of the disease are largely attributed to environmental insults (226). However, such aging-associated diseases are, by nature, consequences from the loss of proliferation capacity induced by an accumulation of the senescent cell population (332). Therefore, the existing short telomeres and pathological conditions in patients with aging-associated diseases can be exacerbated by these intrinsic genetic factors, due to their contributions to accelerated telomere shortening.

4.2.2 Potential models for heritability in TBD

If short telomeres and cellular senescence are so important for the pathogenesis of TBDs or tissue degenerative disorders in general, why are the SNPs in telomere biology genes not usually

picked up in genome-wide association studies (GWAS)? In addition to the limitations of statistical power from sample sizes (including the rarity of some variants), two mathematical models for GWAS could be used to explain the discrepancy between our expectation and clinical observations.

The first is an infinitesimal model based on a common disease-common variant hypothesis: a large number of small-effect common variants have cumulative effects on disease association (333). To support this model, genome-wide polygenic scores have been developed to assess genetic risks in five common complex diseases, including coronary artery disease, atrial fibrillation, type 2 diabetes mellitus, inflammatory bowel disease, and breast cancer (334). With the newly developed SNP-based heritability algorithm, the population with high genetic risks of common diseases were identified and they had significantly higher risks of developing the diseases than patients with rare, monogenic mutations (334). Based on our functional validation of *TERT* SNPs in *ex vivo* cell model and the observed minor, but significant odds ratio, it is probable that SNPs in telomere biology pathway need to be integrated with a combination of other genetic factors to predict one's genetic risk for COPD, other tissue degenerative disorders, or TBDs.

Second, an omnigenic model could also be applied to explain the "missing heritability" for SNPs of telomere biology pathway in GWAS of complex disorders: genetic networks have such strong implications on the regulation of core disease-related genes that the direct effects from the nonsynonymous SNPs in telomere biology genes are potentially masked under gene regulation (335). Under this hypothesis, cell- or tissue-specific regulatory networks need to be identified before we could fully dissect the contributions of regulatory genes and disease-causing genes. TERT expression is subject to strict regulation, such as in the case of INS1b splicing variant that

can be induced by the intronic SNP rs10069690 in a cell- and tissue-specific manner (Section 3.4.6) (122). To understand whether this SNP is contributing to COPD progression, a better comprehension of the expression of telomere biology genes in AEC2 is essential. The concept for the functional regulation of splicing variants can be generalized to the expression of all the telomere biology pathway components: there should be different algorithms for the model of genetic risks of TBDs in fast- and slow-turnover tissues, due to tissue-specific regulation of telomerase expression and telomere length maintenance.

Most *TERT* SNPs picked up in GWAS studies are in association with different types of cancers. This phenomenon is understandable considering the strong selection factor of short telomere-induced genomic instability, a late initiation step in carcinogenesis. The critical role of short telomeres in late stage of carcinogenesis acts as a strong selection filter to amplify the effects of small contributing factors in determining telomere attrition rate. This knowledge could be further extended to targeted association studies of tissue degenerative disorders where short telomeres serve as important road markers for disease pathogenesis.

4.2.3 Prediction of X-DC manifestations in female *DKC1* mutation carriers

The concept of the prevalence of genetic variations in telomere biology genes and their differential penetrance in slow- vs. fast-turnover tissues could also be applied to female *DKC1* mutation carriers discussed in Chapter 2: of this dissertation. In normal females free of any X-linked mutations, skewed XCI presents in an age- and tissue-specific pattern, and that severe skewing was observed in the blood cells from older females, potentially resulting from secondary selection in the context of rapid turnover rate in the blood (336). Theoretically, the protection provided by XCI skewing in female *DKC1* mutation carriers should only be effective so long as the growth disadvantage of cells expressing the *DKC1* mutant allele is present. Thus, cells

expressing the *DKC1* WT allele were selected for in the fastest turnover tissue – blood (Figure 2.5A & B). Contrarily, in slower turnover tissue, WT-*DKC1* expressing cells may need a much longer time to show their competitive growth advantage. The fact that the female carriers with disease symptoms shows patchy skin hyperpigmentation also suggests that the parenchymal tissues lack the ability to tease out and eliminate defective tissues effectively like the hematopoietic system does. Thus, there might be undetected mosaic *DKC1* expression in these tissues (see Section 2.5 for details). With exposure to environmental/disease factors later in life, these extrinsic factor-induced accelerations of cellular turnover may trigger disease manifestations in slow turnover tissues, such as the epithelial manifestations (hair greying, skin and nail lesions) observed in my study (337) and other reports (213, 215). Nonalcoholic liver cirrhosis (337) and early onset of emphysema (215) were also reported, supporting our theory of disease manifestation predominantly in slow turnover tissues of female *DKC1* mutation carriers.

It is worth mentioning that during the preparation of this dissertation, a new study was published reporting on the differential XCI skewing patterns in blood, buccal mucosa, and tongue in a DC family of three female *DKC1* mutation carriers with a novel *DKC1* mutation (c.1218_1219insCAG, p.D406_S407insQ) (338). As predicted, there was minimal (<5%) expression of *DKC1* mutant allele in whole blood in two of the three female carriers, which is consistent to my report (337) and others (213, 215). Interestingly, using a highly sensitive droplet digital RT-PCR method targeting the site of *DKC1* mutation, the authors showed the expression of the *DKC1* mutant allele in 7 – 45% of naïve cells from buccal mucosa and tongue of all the three carriers and in 45% of naïve whole blood cells from one female carrier. It is probable that this particular female carrier with high level of mutant *DKC1* expression may have defects in her XCI machinery or has a second X-linked genetic change on the other X chromosome that

prevented complete XCI skewing from DC protection, which was expected and observed in all the other female *DKC1* mutation carriers. Two of the three female carriers had anemia/cytopenia in addition to presentations of the mucocutaneous triad (characteristic of DC) (338). While the underlying mechanism of the escape from XCI skewing remained unknown, the relatively high expression levels of *DKC1* mutant allele in epithelial cells (buccal mucosa and tongue) than in blood again supported our contention that *DKC1* mutant expressions were selected against through XCI in female carriers and that the extent of skewness was a function of tissue turnover rate.

Findings in this new study also confirmed that unmanipulated biopsy samples from areas of disease manifestations were superior over EBV-transformed blood cells or any patient-derived, laboratory-cultured samples in the investigation of X-DC disease etiology. Blood is usually the only tissue collected for the measurement of telomere length, which serves as a diagnostic criterion for DC. While this is a completely acceptable surrogate measurement for the prediction of bone marrow failure, as discussed extensively in my study and others, DC-like symptoms could show up in slow turn-over tissues in the absence of short telomeres in blood. Furthermore, in the aforementioned study, two of the three female carriers had anemia/cytopenia in addition to the mucocutaneous triad, though their telomere lengths in blood cells were within normal range (338). Therefore, telomere length in blood is not necessarily an ideal surrogate for the prediction and prognosis of bone-marrow related X-DC disease manifestations. I would suggest collecting cells from all the pertinent tissues, if possible, for better diagnosis of DC and DC-like syndromes.

4.3 Closing remarks

As genome sequencing becomes cheaper for research and more accessible to the general public, personalized genomic prediction tools are in high demand. However, our knowledge of data utilization and interpretation are still not comprehensive. A big challenge for the era of genomics is to deduce complex disease models incorporating functional genetic network knowledge and gene-environment interactions. My study of genetic modifiers of telomere biology disorders serves as a proof of principle for the roles of mild genetic changes in building such models, contributing to the understanding of biological mechanism and the validation of putative genetic risks.

Bibliography

1. Allsopp RC, Harley CB. Evidence for a critical telomere length in senescent human fibroblasts. Experimental cell research. 1995;219(1):130-6.

2. McElligott R, Wellinger RJ. The terminal DNA structure of mammalian chromosomes. EMBO J. 1997;16(12):3705-14.

3. Tommerup H, Dousmanis A, de Lange T. Unusual chromatin in human telomeres. Mol Cell Biol. 1994;14(9):5777-85.

4. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, et al. Mammalian telomeres end in a large duplex loop. Cell. 1999;97(4):503-14.

5. Nikitina T, Woodcock CL. Closed chromatin loops at the ends of chromosomes. J Cell Biol. 2004;166(2):161-5.

6. Doksani Y, Wu JY, de Lange T, Zhuang X. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. Cell. 2013;155(2):345-56.

7. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005;19(18):2100-10.

8. Denchi EL, de Lange T. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. Nature. 2007;448(7157):1068-71.

9. Parkinson GN, Lee MP, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. Nature. 2002;417(6891):876-80.

10. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013;5(3):182-6.

11. Nandakumar J, Cech TR. Finding the end: recruitment of telomerase to telomeres. Nat Rev Mol Cell Biol. 2013;14(2):69-82.

12. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A. 1981;78(10):6354-8.

13. Liang F, Han M, Romanienko PJ, Jasin M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A. 1998;95(9):5172-7.

14. de Lange T. Shelterin-Mediated Telomere Protection. Annu Rev Genet. 2018.

15. Opresko PL, von Kobbe C, Laine JP, Harrigan J, Hickson ID, Bohr VA. Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. J Biol Chem. 2002;277(43):41110-9.

16. Crabbe L, Verdun RE, Haggblom CI, Karlseder J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science. 2004;306(5703):1951-3.

17. Opresko PL, Otterlei M, Graakjaer J, Bruheim P, Dawut L, Kølvraa S, et al. The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. Mol Cell. 2004;14(6):763-74.

18. Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MI, et al. RTEL1 maintains genomic stability by suppressing homologous recombination. Cell. 2008;135(2):261-71.

19. Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. Cell. 2012;149(4):795-806.

20. Walne AJ, Vulliamy T, Kirwan M, Plagnol V, Dokal I. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. Am J Hum Genet. 2013;92(3):448-53.

21. Uringa EJ, Youds JL, Lisaingo K, Lansdorp PM, Boulton SJ. RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. Nucleic Acids Res. 2011;39(5):1647-55.

22. Olovnikov AM. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. J Theor Biol. 1973;41(1):181-90.

23. Soudet J, Jolivet P, Teixeira MT. Elucidation of the DNA end-replication problem in Saccharomyces cerevisiae. Mol Cell. 2014;53(6):954-64.

24. Wu P, Takai H, de Lange T. Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. Cell. 2012;150(1):39-52.

25. Stewart JA, Wang F, Chaiken MF, Kasbek C, Chastain PD, Wright WE, et al. Human CST promotes telomere duplex replication and general replication restart after fork stalling. EMBO J. 2012;31(17):3537-49.

26. Feng X, Hsu SJ, Kasbek C, Chaiken M, Price CM. CTC1-mediated C-strand fill-in is an essential step in telomere length maintenance. Nucleic Acids Res. 2017;45(8):4281-93.

27. Hayflick L. The Limited in Vitro Lifetime of Human Diploid Cell Strains. Experimental cell research. 1965;37:614-36.

28. Huffman KE, Levene SD, Tesmer VM, Shay JW, Wright WE. Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. J Biol Chem. 2000;275(26):19719-22.

29. Levy MZ, Allsopp RC, Futcher AB, Greider CW, Harley CB. Telomere end-replication problem and cell aging. J Mol Biol. 1992;225(4):951-60.

30. Campisi J. The biology of replicative senescence. European journal of cancer. 1997;33(5):703-9.

31. Vaziri H, West MD, Allsopp RC, Davison TS, Wu YS, Arrowsmith CH, et al. ATM-dependent telomere loss in aging human diploid fibroblasts and DNA damage lead to the post-translational activation of p53 protein involving poly(ADP-ribose) polymerase. The EMBO journal. 1997;16(19):6018-33.

32. Hackett JA, Feldser DM, Greider CW. Telomere dysfunction increases mutation rate and genomic instability. Cell. 2001;106(3):275-86.

33. Maciejowski J, de Lange T. Telomeres in cancer: tumour suppression and genome instability. Nat Rev Mol Cell Biol. 2017;18(3):175-86.

34. Counter CM, Avilion AA, LeFeuvre CE, Stewart NG, Greider CW, Harley CB, et al. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 1992;11(5):1921-9.

35. Meeker AK, Hicks JL, Iacobuzio-Donahue CA, Montgomery EA, Westra WH, Chan TY, et al. Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. Clin Cancer Res. 2004;10(10):3317-26.

36. Bryce LA, Morrison N, Hoare SF, Muir S, Keith WN. Mapping of the gene for the human telomerase reverse transcriptase, hTERT, to chromosome 5p15.33 by fluorescence in situ hybridization. Neoplasia. 2000;2(3):197-201.

37. Shay JW. Role of Telomeres and Telomerase in Aging and Cancer. Cancer Discov. 2016;6(6):584-93.

38. Wright WE, Shay JW. The two-stage mechanism controlling cellular senescence and immortalization. Exp Gerontol. 1992;27(4):383-9.

39. Chiba K, Lorbeer FK, Shain AH, McSwiggen DT, Schruf E, Oh A, et al. Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. Science. 2017;357(6358):1416-20.

40. Weinrich SL, Pruzan R, Ma L, Ouellette M, Tesmer VM, Holt SE, et al. Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. Nat Genet. 1997;17(4):498-502.

41. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB, et al. Extension of life-span by introduction of telomerase into normal human cells. Science. 1998;279(5349):349-52.

42. Mitchell JR, Cheng J, Collins K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. Mol Cell Biol. 1999;19(1):567-76.

43. Jády BE, Bertrand E, Kiss T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. J Cell Biol. 2004;164(5):647-52.

44. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. Nature. 1999;402(6761):551-5.

45. Xi L, Cech TR. Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. Nucleic Acids Res. 2014;42(13):8565-77.

46. Cohen SB, Graham ME, Lovrecz GO, Bache N, Robinson PJ, Reddel RR. Protein composition of catalytically active human telomerase from immortal cells. Science. 2007;315(5820):1850-3.

47. Yi X, Shay JW, Wright WE. Quantitation of telomerase components and hTERT mRNA splicing patterns in immortal human cells. Nucleic Acids Res. 2001;29(23):4818-25.

48. Cao Y, Huschtscha LI, Nouwens AS, Pickett HA, Neumann AA, Chang AC, et al. Amplification of telomerase reverse transcriptase gene in human mammary epithelial cells with limiting telomerase RNA expression levels. Cancer Res. 2008;68(9):3115-23.

49. Cristofari G, Lingner J. Telomere length homeostasis requires that telomerase levels are limiting. EMBO J. 2006;25(3):565-74.

50. Zhang Q, Kim NK, Feigon J. Architecture of human telomerase RNA. Proc Natl Acad Sci U S A. 2011;108(51):20325-32.

51. Theimer CA, Blois CA, Feigon J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol Cell. 2005;17(5):671-82.

52. Chen JL, Greider CW. Template boundary definition in mammalian telomerase. Genes Dev. 2003;17(22):2747-52.

53. Brown AF, Podlevsky JD, Qi X, Chen Y, Xie M, Chen JJ. A self-regulating template in human telomerase. Proc Natl Acad Sci U S A. 2014;111(31):11311-6.

54. Martadinata H, Phan AT. Formation of a stacked dimeric G-quadruplex containing bulges by the 5'-terminal region of human telomerase RNA (hTERC). Biochemistry. 2014;53(10):1595-600.

55. Lattmann S, Stadler MB, Vaughn JP, Akman SA, Nagamine Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. Nucleic Acids Res. 2011;39(21):9390-404.

56. Booy EP, Meier M, Okun N, Novakowski SK, Xiong S, Stetefeld J, et al. The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. Nucleic Acids Res. 2012;40(9):4110-24.

57. Sexton AN, Collins K. The 5' guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. Mol Cell Biol. 2011;31(4):736-43.

58. Blackburn EH, Collins K. Telomerase: an RNP enzyme synthesizes DNA. Cold Spring Harb Perspect Biol. 2011;3(5).

59. Lin J, Ly H, Hussain A, Abraham M, Pearl S, Tzfati Y, et al. A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. Proc Natl Acad Sci U S A. 2004;101(41):14713-8.

60. Wu RA, Tam J, Collins K. DNA-binding determinants and cellular thresholds for human telomerase repeat addition processivity. EMBO J. 2017;36(13):1908-27.

61. Ueda CT, Roberts RW. Analysis of a long-range interaction between conserved domains of human telomerase RNA. RNA. 2004;10(1):139-47.

62. Mitchell JR, Collins K. Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. Mol Cell. 2000;6(2):361-71.

63. Robart AR, Collins K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. J Biol Chem. 2010;285(7):4375-86.

64. Chen JL, Opperman KK, Greider CW. A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res. 2002;30(2):592-7.

65. Feng J, Funk WD, Wang SS, Weinrich SL, Avilion AA, Chiu CP, et al. The RNA component of human telomerase. Science. 1995;269(5228):1236-41.

66. Egan ED, Collins K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. Mol Cell Biol. 2012;32(13):2428-39.

67. Podlevsky JD, Chen JJ. Evolutionary perspectives of telomerase RNA structure and function. RNA Biol. 2016;13(8):720-32.

68. Nguyen THD, Tam J, Wu RA, Greber BJ, Toso D, Nogales E, et al. Cryo-EM structure of substrate-bound human telomerase holoenzyme. Nature. 2018;557(7704):190-5.

69. Kiss T, Fayet-Lebaron E, Jády BE. Box H/ACA small ribonucleoproteins. Mol Cell. 2010;37(5):597-606.

70. Li X, Ma S, Yi C. Pseudouridine: the fifth RNA nucleotide with renewed interests. Curr Opin Chem Biol. 2016;33:108-16.

71. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. IUBMB Life. 2000;49(5):341-51.

72. Jacobs SA, Podell ER, Cech TR. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat Struct Mol Biol. 2006;13(3):218-25.

73. Lue NF. A physical and functional constituent of telomerase anchor site. J Biol Chem. 2005;280(28):26586-91.

74. Wyatt HD, Lobb DA, Beattie TL. Characterization of physical and functional anchor site interactions in human telomerase. Mol Cell Biol. 2007;27(8):3226-40.

75. Zaug AJ, Podell ER, Cech TR. Mutation in TERT separates processivity from anchor-site function. Nat Struct Mol Biol. 2008;15(8):870-2.

76. Wyatt HD, West SC, Beattie TL. InTERTpreting telomerase structure and function. Nucleic Acids Res. 2010;38(17):5609-22.

77. Robart AR, Collins K. Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation. Mol Cell. 2011;42(3):308-18.

78. Wyatt HD, Tsang AR, Lobb DA, Beattie TL. Human telomerase reverse transcriptase (hTERT) Q169 is essential for telomerase function in vitro and in vivo. PLoS One. 2009;4(9):e7176.

79. Sealey DC, Zheng L, Taboski MA, Cruickshank J, Ikura M, Harrington LA. The N-terminus of hTERT contains a DNA-binding domain and is required for telomerase activity and cellular immortalization. Nucleic Acids Res. 2010;38(6):2019-35.

80. Armbruster BN, Banik SS, Guo C, Smith AC, Counter CM. N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo. Mol Cell Biol. 2001;21(22):7775-86.

81. Schmidt JC, Dalby AB, Cech TR. Identification of human TERT elements necessary for telomerase recruitment to telomeres. Elife. 2014;3.

82. Sexton AN, Regalado SG, Lai CS, Cost GJ, O'Neil CM, Urnov FD, et al. Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. Genes Dev. 2014;28(17):1885-99.

83. Zhong FL, Batista LF, Freund A, Pech MF, Venteicher AS, Artandi SE. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. Cell. 2012;150(3):481-94.

84. Huang J, Brown AF, Wu J, Xue J, Bley CJ, Rand DP, et al. Structural basis for protein-RNA recognition in telomerase. Nat Struct Mol Biol. 2014;21(6):507-12.

85. Bley CJ, Qi X, Rand DP, Borges CR, Nelson RW, Chen JJ. RNA-protein binding interface in the telomerase ribonucleoprotein. Proc Natl Acad Sci U S A. 2011;108(51):20333-8.

86. Bryan TM, Goodrich KJ, Cech TR. Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase. Mol Cell. 2000;6(2):493-9.

87. Bosoy D, Peng Y, Mian IS, Lue NF. Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. J Biol Chem. 2003;278(6):3882-90.

88. Friedman KL, Cech TR. Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. Genes Dev. 1999;13(21):2863-74.

89. Rouda S, Skordalakes E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure. 2007;15(11):1403-12.

90. Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. Reverse transcriptase motifs in the catalytic subunit of telomerase. Science. 1997;276(5312):561-7.

91. Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, et al. Telomerase catalytic subunit homologs from fission yeast and human. Science. 1997;277(5328):955-9.

92. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 1990;9(10):3353-62.

93. Autexier C, Lue NF. The structure and function of telomerase reverse transcriptase. Annu Rev Biochem. 2006;75:493-517.

94. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A. 1993;90(14):6498-502.

95. Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. Nat Struct Mol Biol. 2010;17(4):513-8.

96. Peng Y, Mian IS, Lue NF. Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance. Mol Cell. 2001;7(6):1201-11.

97. Lue NF, Lin YC, Mian IS. A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. Mol Cell Biol. 2003;23(23):8440-9.

98. Chu TW, D'Souza Y, Autexier C. The Insertion in Fingers Domain in Human Telomerase Can Mediate Enzyme Processivity and Telomerase Recruitment to Telomeres in a TPP1-Dependent Manner. Mol Cell Biol. 2016;36(1):210-22.

99. Bosoy D, Lue NF. Functional analysis of conserved residues in the putative "finger" domain of telomerase reverse transcriptase. J Biol Chem. 2001;276(49):46305-12.

100. Xie M, Podlevsky JD, Qi X, Bley CJ, Chen JJ. A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. Nucleic Acids Res. 2010;38(6):1982-96.

101. Hossain S, Singh S, Lue NF. Functional analysis of the C-terminal extension of telomerase reverse transcriptase. A putative "thumb" domain. J Biol Chem. 2002;277(39):36174-80.

102. Huard S, Moriarty TJ, Autexier C. The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity. Nucleic Acids Res. 2003;31(14):4059-70.

103. Seimiya H, Sawada H, Muramatsu Y, Shimizu M, Ohko K, Yamane K, et al. Involvement of 14-3-3 proteins in nuclear localization of telomerase. EMBO J. 2000;19(11):2652-61.

104. Tomlinson CG, Holien JK, Mathias JA, Parker MW, Bryan TM. The C-terminal extension of human telomerase reverse transcriptase is necessary for high affinity binding to telomeric DNA. Biochimie. 2016;128-129:114-21.

105. Wright WE, Piatyszek MA, Rainey WE, Byrd W, Shay JW. Telomerase activity in human germline and embryonic tissues and cells. Dev Genet. 1996;18(2):173-9.

106. Cong YS, Wen J, Bacchetti S. The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter. Hum Mol Genet. 1999;8(1):137-42.

107. Ramlee MK, Wang J, Toh WX, Li S. Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene. Genes (Basel). 2016;7(8).

108. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013;339(6122):957-9.

109. Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, et al. TERT promoter mutations in familial and sporadic melanoma. Science. 2013;339(6122):959-61.

110. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc Natl Acad Sci U S A. 2013;110(15):6021-6.

111. Nonoguchi N, Ohta T, Oh JE, Kim YH, Kleihues P, Ohgaki H. TERT promoter mutations in primary and secondary glioblastomas. Acta Neuropathol. 2013;126(6):931-7.

112. Landa I, Ganly I, Chan TA, Mitsutake N, Matsuse M, Ibrahimpasic T, et al. Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. J Clin Endocrinol Metab. 2013;98(9):E1562-6.

113. Liu X, Bishop J, Shan Y, Pai S, Liu D, Murugan AK, et al. Highly prevalent TERT promoter mutations in aggressive thyroid cancers. Endocr Relat Cancer. 2013;20(4):603-10.

114. Borah S, Xi L, Zaug AJ, Powell NM, Dancik GM, Cohen SB, et al. Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. Science. 2015;347(6225):1006-10.

115. Kinde I, Munari E, Faraj SF, Hruban RH, Schoenberg M, Bivalacqua T, et al. TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine. Cancer Res. 2013;73(24):7162-7.

116. Rachakonda PS, Hosen I, de Verdier PJ, Fallah M, Heidenreich B, Ryk C, et al. TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. Proc Natl Acad Sci U S A. 2013;110(43):17426-31.

117. Bell RJ, Rube HT, Xavier-Magalhães A, Costa BM, Mancini A, Song JS, et al. Understanding TERT Promoter Mutations: A Common Path to Immortality. Mol Cancer Res. 2016;14(4):315-23.

118. Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SD, Nagarajan RP, et al. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. Science. 2015;348(6238):1036-9.

119. Stern JL, Theodorescu D, Vogelstein B, Papadopoulos N, Cech TR. Mutation of the TERT promoter, switch to active chromatin, and monoallelic TERT expression in multiple cancers. Genes Dev. 2015;29(21):2219-24.

120. Akıncılar SC, Khattar E, Boon PL, Unal B, Fullwood MJ, Tergaonkar V. Long-Range Chromatin Interactions Drive Mutant TERT Promoter Activation. Cancer Discov. 2016;6(11):1276-91.

121. Wong MS, Wright WE, Shay JW. Alternative splicing regulation of telomerase: a new paradigm? Trends Genet. 2014;30(10):430-8.

122. Killedar A, Stutz MD, Sobinoff AP, Tomlinson CG, Bryan TM, Beesley J, et al. A Common Cancer Risk-Associated Allele in the hTERT Locus Encodes a Dominant Negative Inhibitor of Telomerase. PLoS Genet. 2015;11(6):e1005286.

123. Walbott H, Machado-Pinilla R, Liger D, Blaud M, Réty S, Grozdanov PN, et al. The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. Genes Dev. 2011;25(22):2398-408.

124. Grozdanov PN, Roy S, Kittur N, Meier UT. SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. RNA. 2009;15(6):1188-97.

125. Venteicher AS, Meng Z, Mason PJ, Veenstra TD, Artandi SE. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. Cell. 2008;132(6):945-57.

126. Tseng CK, Wang HF, Burns AM, Schroeder MR, Gaspari M, Baumann P. Human Telomerase RNA Processing and Quality Control. Cell Rep. 2015;13(10):2232-43.

127. Nguyen D, Grenier St-Sauveur V, Bergeron D, Dupuis-Sandoval F, Scott MS, Bachand F. A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA. Cell Rep. 2015;13(10):2244-57.

128. Fu D, Collins K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. Mol Cell. 2003;11(5):1361-72.

129. Berndt H, Harnisch C, Rammelt C, Stöhr N, Zirkel A, Dohm JC, et al. Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. RNA. 2012;18(5):958-72.

130. Moon DH, Segal M, Boyraz B, Guinan E, Hofmann I, Cahan P, et al. Poly(A)-specific ribonuclease (PARN) mediates 3'-end maturation of the telomerase RNA component. Nat Genet. 2015;47(12):1482-8.

131. Boyraz B, Moon DH, Segal M, Muosieyiri MZ, Aykanat A, Tai AK, et al. Posttranscriptional manipulation of TERC reverses molecular hallmarks of telomere disease. J Clin Invest. 2016;126(9):3377-82.

132. Chen L, Roake CM, Freund A, Batista PJ, Tian S, Yin YA, et al. An Activity Switch in Human Telomerase Based on RNA Conformation and Shaped by TCAB1. Cell. 2018;174(1):218-30.e13.

133. Holt SE, Aisner DL, Baur J, Tesmer VM, Dy M, Ouellette M, et al. Functional requirement of p23 and Hsp90 in telomerase complexes. Genes Dev. 1999;13(7):817-26.

134. Forsythe HL, Jarvis JL, Turner JW, Elmore LW, Holt SE. Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. J Biol Chem. 2001;276(19):15571-4.

135. Egan ED, Collins K. Biogenesis of telomerase ribonucleoproteins. RNA. 2012;18(10):1747-59.

136. Jeong SA, Kim K, Lee JH, Cha JS, Khadka P, Cho HS, et al. Akt-mediated phosphorylation increases the binding affinity of hTERT for importin α to promote nuclear translocation. J Cell Sci. 2015;128(12):2287-301.

137. Haendeler J, Hoffmann J, Brandes RP, Zeiher AM, Dimmeler S. Hydrogen peroxide triggers nuclear export of telomerase reverse transcriptase via Src kinase family-dependent phosphorylation of tyrosine 707. Mol Cell Biol. 2003;23(13):4598-610.

138. Xi P, Zhou L, Wang M, Liu JP, Cong YS. Serine/threonine-protein phosphatase 2A physically interacts with human telomerase reverse transcriptase hTERT and regulates its subcellular distribution. J Cell Biochem. 2013;114(2):409-17.

139. Singh M, Wang Z, Koo BK, Patel A, Cascio D, Collins K, et al. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. Mol Cell. 2012;47(1):16-26.

140. Collopy LC, Ware TL, Goncalves T, Í Kongsstovu S, Yang Q, Amelina H, et al. LARP7 family proteins have conserved function in telomerase assembly. Nat Commun. 2018;9(1):557.

141. Mennie AK, Moser BA, Nakamura TM. LARP7-like protein Pof8 regulates telomerase assembly and poly(A)+TERRA expression in fission yeast. Nat Commun. 2018;9(1):586.

142. Holohan B, Kim W, Lai TP, Hoshiyama H, Zhang N, Alazami AM, et al. Impaired telomere maintenance in Alazami syndrome patients with LARP7 deficiency. BMC Genomics. 2016;17(Suppl 9):749.

143. Masutomi K, Possemato R, Wong JM, Currier JL, Tothova Z, Manola JB, et al. The telomerase reverse transcriptase regulates chromatin state and DNA damage responses. Proc Natl Acad Sci U S A. 2005;102(23):8222-7.

144. Shin KH, Kang MK, Dicterow E, Kameta A, Baluda MA, Park NH. Introduction of human telomerase reverse transcriptase to normal human fibroblasts enhances DNA repair capacity. Clin Cancer Res. 2004;10(7):2551-60.

145. Sharma GG, Gupta A, Wang H, Scherthan H, Dhar S, Gandhi V, et al. hTERT associates with human telomeres and enhances genomic stability and DNA repair. Oncogene. 2003;22(1):131-46.

146. Thompson CAH, Gu A, Yang SY, Mathew V, Fleisig HB, Wong JMY. Transient Telomerase Inhibition with Imetelstat Impacts DNA Damage Signals and Cell-Cycle Kinetics. Mol Cancer Res. 2018;16(8):1215-25.

147. Fleisig HB, Hukezalie KR, Thompson CA, Au-Yeung TT, Ludlow AT, Zhao CR, et al. Telomerase reverse transcriptase expression protects transformed human cells against DNA-damaging agents, and increases tolerance to chromosomal instability. Oncogene. 2016;35(2):218-27.

148. Stewart SA, Hahn WC, O'Connor BF, Banner EN, Lundberg AS, Modha P, et al. Telomerase contributes to tumorigenesis by a telomere length-independent mechanism. Proc Natl Acad Sci U S A. 2002;99(20):12606-11.

149. Choi J, Southworth LK, Sarin KY, Venteicher AS, Ma W, Chang W, et al. TERT promotes epithelial proliferation through transcriptional control of a Myc- and Wnt-related developmental program. PLoS Genet. 2008;4(1):e10.

150. Ghosh A, Saginc G, Leow SC, Khattar E, Shin EM, Yan TD, et al. Telomerase directly regulates NF-κB-dependent transcription. Nat Cell Biol. 2012;14(12):1270-81.

151. Park JI, Venteicher AS, Hong JY, Choi J, Jun S, Shkreli M, et al. Telomerase modulates Wnt signalling by association with target gene chromatin. Nature. 2009;460(7251):66-72.

152. Koh CM, Khattar E, Leow SC, Liu CY, Muller J, Ang WX, et al. Telomerase regulates MYC-driven oncogenesis independent of its reverse transcriptase activity. J Clin Invest. 2015;125(5):2109-22.

153. Maida Y, Yasukawa M, Furuuchi M, Lassmann T, Possemato R, Okamoto N, et al. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. Nature. 2009;461(7261):230-5.

154. Gonzalez OG, Assfalg R, Koch S, Schelling A, Meena JK, Kraus J, et al. Telomerase stimulates ribosomal DNA transcription under hyperproliferative conditions. Nat Commun. 2014;5:4599.

155. Fleisig HB, Wong JM. Telomerase promotes efficient cell cycle kinetics and confers growth advantage to telomerase-negative transformed human cells. Oncogene. 2012;31(8):954-65.

156. Mukherjee S, Firpo EJ, Wang Y, Roberts JM. Separation of telomerase functions by reverse genetics. Proc Natl Acad Sci U S A. 2011;108(50):E1363-71.

157. Zhu J, Wang H, Bishop JM, Blackburn EH. Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. Proc Natl Acad Sci U S A. 1999;96(7):3723-8.

158. Ahmed S, Passos JF, Birket MJ, Beckmann T, Brings S, Peters H, et al. Telomerase does not counteract telomere shortening but protects mitochondrial function under oxidative stress. J Cell Sci. 2008;121(Pt 7):1046-53.

159. Haendeler J, Dröse S, Büchner N, Jakob S, Altschmied J, Goy C, et al. Mitochondrial telomerase reverse transcriptase binds to and protects mitochondrial DNA and function from damage. Arterioscler Thromb Vasc Biol. 2009;29(6):929-35.

160. Nitta E, Yamashita M, Hosokawa K, Xian M, Takubo K, Arai F, et al. Telomerase reverse transcriptase protects ATM-deficient hematopoietic stem cells from ROS-induced apoptosis through a telomere-independent mechanism. Blood. 2011;117(16):4169-80.

161. Sarin KY, Cheung P, Gilison D, Lee E, Tennen RI, Wang E, et al. Conditional telomerase induction causes proliferation of hair follicle stem cells. Nature. 2005;436(7053):1048-52.

162. Hukezalie KR, Wong JM. Structure-function relationship and biogenesis regulation of the human telomerase holoenzyme. The FEBS journal. 2013;280(14):3194-204.

163. Shay JW, Wright WE. Role of telomeres and telomerase in cancer. Seminars in cancer biology. 2011;21(6):349-53.

164. Alder JK, Cogan JD, Brown AF, Anderson CJ, Lawson WE, Lansdorp PM, et al. Ancestral mutation in telomerase causes defects in repeat addition processivity and manifests as familial pulmonary fibrosis. PLoS Genet. 2011;7(3):e1001352.

165. Nandakumar J, Bell CF, Weidenfeld I, Zaug AJ, Leinwand LA, Cech TR. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. Nature. 2012;492(7428):285-9.

166. Abreu E, Aritonovska E, Reichenbach P, Cristofari G, Culp B, Terns RM, et al. TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. Mol Cell Biol. 2010;30(12):2971-82.

167. Takai KK, Kibe T, Donigian JR, Frescas D, de Lange T. Telomere protection by TPP1/POT1 requires tethering to TIN2. Mol Cell. 2011;44(4):647-59.

168. Lei M, Podell ER, Cech TR. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat Struct Mol Biol. 2004;11(12):1223-9.

169. Kelleher C, Kurth I, Lingner J. Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro. Mol Cell Biol. 2005;25(2):808-18.

170. Rice C, Shastrula PK, Kossenkov AV, Hills R, Baird DM, Showe LC, et al. Structural and functional analysis of the human POT1-TPP1 telomeric complex. Nat Commun. 2017;8:14928.

171. Wang F, Podell ER, Zaug AJ, Yang Y, Baciu P, Cech TR, et al. The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature. 2007;445(7127):506-10.

172. Chen LY, Redon S, Lingner J. The human CST complex is a terminator of telomerase activity. Nature. 2012;488(7412):540-4.

173. Wang F, Stewart JA, Kasbek C, Zhao Y, Wright WE, Price CM. Human CST has independent functions during telomere duplex replication and C-strand fill-in. Cell Rep. 2012;2(5):1096-103.

174. Zhao Y, Abreu E, Kim J, Stadler G, Eskiocak U, Terns MP, et al. Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions. Mol Cell. 2011;42(3):297-307.

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

176. Xu Z, Duc KD, Holcman D, Teixeira MT. The length of the shortest telomere as the major determinant of the onset of replicative senescence. Genetics. 2013;194(4):847-57.

177. Martens UM, Chavez EA, Poon SS, Schmoor C, Lansdorp PM. Accumulation of short telomeres in human fibroblasts prior to replicative senescence. Exp Cell Res. 2000;256(1):291-9.

178. Xi L, Schmidt JC, Zaug AJ, Ascarrunz DR, Cech TR. A novel two-step genome editing strategy with CRISPR-Cas9 provides new insights into telomerase action and TERT gene expression. Genome Biol. 2015;16:231.

179. Marcand S, Gilson E, Shore D. A protein-counting mechanism for telomere length regulation in yeast. Science. 1997;275(5302):986-90.

180. Loayza D, De Lange T. POT1 as a terminal transducer of TRF1 telomere length control. Nature. 2003;423(6943):1013-8.

181. Teixeira MT, Arneric M, Sperisen P, Lingner J. Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states. Cell. 2004;117(3):323-35.

182. Ballew BJ, Lundblad V. Multiple genetic pathways regulate replicative senescence in telomerase-deficient yeast. Aging Cell. 2013;12(4):719-27.

183. Greider CW. Regulating telomere length from the inside out: the replication fork model. Genes Dev. 2016;30(13):1483-91.

184. Hafner L, Lezaja A, Zhang X, Lemmens L, Shyian M, Albert B, et al. Rif1 Binding and Control of Chromosome-Internal DNA Replication Origins Is Limited by Telomere Sequestration. Cell Rep. 2018;23(4):983-92.

185. Davé A, Cooley C, Garg M, Bianchi A. Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. Cell Rep. 2014;7(1):53-61.

186. Diede SJ, Gottschling DE. Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta. Cell. 1999;99(7):723-33.

187. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet. 2010;11(5):319-30.

188. Pickett HA, Reddel RR. Molecular mechanisms of activity and derepression of alternative lengthening of telomeres. Nat Struct Mol Biol. 2015;22(11):875-80.

189. Savage SA. Beginning at the ends: telomeres and human disease. F1000Res. 2018;7.

190. Dodson D, Bertuch AA. Dyskeratosis Congenita and the Telomere Biology Disorders. In: Kupfer G. RG, Smith F., editor. Bone Marrow Failure Pediatric Oncology: Springer, Cham; 2018.

191. Alter BP, Rosenberg PS, Giri N, Baerlocher GM, Lansdorp PM, Savage SA. Telomere length is associated with disease severity and declines with age in dyskeratosis congenita. Haematologica. 2012;97(3):353-9.

192. Vulliamy TJ, Marrone A, Knight SW, Walne A, Mason PJ, Dokal I. Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. Blood. 2006;107(7):2680-5.

193. Bessler M, Wilson DB, Mason PJ. Dyskeratosis congenita and telomerase. Current opinion in pediatrics. 2004;16(1):23-8.

194. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.

195. Armanios M, Blackburn EH. The telomere syndromes. Nat Rev Genet. 2012;13(10):693-704.

196. Wegman-Ostrosky T, Savage SA. The genomics of inherited bone marrow failure: from mechanism to the clinic. Br J Haematol. 2017;177(4):526-42.

197. Bertuch AA. The molecular genetics of the telomere biology disorders. RNA Biol. 2016;13(8):696-706.

198. Takai H, Jenkinson E, Kabir S, Babul-Hirji R, Najm-Tehrani N, Chitayat DA, et al. A POT1 mutation implicates defective telomere end fill-in and telomere truncations in Coats plus. Genes Dev. 2016;30(7):812-26.

199. Touzot F, Callebaut I, Soulier J, Gaillard L, Azerrad C, Durandy A, et al. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. Proc Natl Acad Sci U S A. 2010;107(22):10097-102.

200. Bizarro J, Meier UT. Inherited SHQ1 mutations impair interaction with NAP57/dyskerin, a major target in dyskeratosis congenita. Mol Genet Genomic Med. 2017;5(6):805-8.

201. Stanley SE, Gable DL, Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. Sci Transl Med. 2016;8(351):351ra107.

202. Simon AJ, Lev A, Zhang Y, Weiss B, Rylova A, Eyal E, et al. Mutations in STN1 cause Coats plus syndrome and are associated with genomic and telomere defects. J Exp Med. 2016;213(8):1429-40.

203. Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, 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. 2013;21(10):1163-8.

204. Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, Griffin CA, et al. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proc Natl Acad Sci U S A. 2005;102(44):15960-4.

205. Vulliamy T, Marrone A, Szydlo R, Walne A, Mason PJ, Dokal I. Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. Nat Genet. 2004;36(5):447-9.

206. Diaz de Leon A, Cronkhite JT, Katzenstein AL, Godwin JD, Raghu G, Glazer CS, et al. Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. PLoS One. 2010;5(5):e10680.

207. Chiang YJ, Calado RT, Hathcock KS, Lansdorp PM, Young NS, Hodes RJ. Telomere length is inherited with resetting of the telomere set-point. Proc Natl Acad Sci U S A. 2010;107(22):10148-53.

208. Dokal I, Vulliamy T, Mason P, Bessler M. Clinical utility gene card for: Dyskeratosis congenita - update 2015. Eur J Hum Genet. 2015;23(4).

209. Ballew BJ, Savage SA. Updates on the biology and management of dyskeratosis congenita and related telomere biology disorders. Expert Rev Hematol. 2013;6(3):327-37.

210. Anderson BH, Kasher PR, Mayer J, Szynkiewicz M, Jenkinson EM, Bhaskar SS, et al. Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. Nat Genet. 2012;44(3):338-42.

211. Keller RB, Gagne KE, Usmani GN, Asdourian GK, Williams DA, Hofmann I, et al. CTC1 Mutations in a patient with dyskeratosis congenita. Pediatr Blood Cancer. 2012;59(2):311-4.

212. Vulliamy TJ, Knight SW, Dokal I, Mason PJ. Skewed X-inactivation in carriers of X-linked dyskeratosis congenita. Blood. 1997;90(6):2213-6.

213. Devriendt K, Matthijs G, Legius E, Schollen E, Blockmans D, van Geet C, et al. Skewed X-chromosome inactivation in female carriers of dyskeratosis congenita. Am J Hum Genet. 1997;60(3):581-7.

214. Ferraris AM, Forni GL, Mangerini R, Gaetani GF. Nonrandom X-chromosome inactivation in hemopoietic cells from carriers of dyskeratosis congenita. Am J Hum Genet. 1997;61(2):458-61.

215. Alder JK, Parry EM, Yegnasubramanian S, Wagner CL, Lieblich LM, Auerbach R, et al. Telomere phenotypes in females with heterozygous mutations in the dyskeratosis congenita 1 (DKC1) gene. Hum Mutat. 2013;34(11):1481-5.

216. Hopkins RB, Burke N, Fell C, Dion G, Kolb M. Epidemiology and survival of idiopathic pulmonary fibrosis from national data in Canada. Eur Respir J. 2016;48(1):187-95.

217. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007;356(13):1317-26.

218. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A. 2007;104(18):7552-7.

219. Diaz de Leon A, Cronkhite JT, Yilmaz C, Brewington C, Wang R, Xing C, et al. Subclinical lung disease, macrocytosis, and premature graying in kindreds with telomerase (TERT) mutations. Chest. 2011;140(3):753-63.

220. Alder JK, Chen JJ, Lancaster L, Danoff S, Su SC, Cogan JD, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proc Natl Acad Sci U S A. 2008;105(35):13051-6.

221. Armanios M. Telomerase and idiopathic pulmonary fibrosis. Mutat Res. 2012;730(1-2):52-8.

222. Hoffman TW, van Moorsel CHM, Borie R, Crestani B. Pulmonary phenotypes associated with genetic variation in telomere-related genes. Curr Opin Pulm Med. 2018;24(3):269-80.

223. Armanios M. Telomeres and age-related disease: how telomere biology informs clinical paradigms. J Clin Invest. 2013;123(3):996-1002.

224. Stanley SE, Noth I, Armanios M. What the genetics "RTEL"ing us about telomeres and pulmonary fibrosis. Am J Respir Crit Care Med. 2015;191(6):608-10.

225. Silhan LL, Shah PD, Chambers DC, Snyder LD, Riise GC, Wagner CL, et al. Lung transplantation in telomerase mutation carriers with pulmonary fibrosis. Eur Respir J. 2014;44(1):178-87.

226. Collaborators GCRD. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and

asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med. 2017;5(9):691-706.

227. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. Lancet. 2015;385(9971):899-909.

228. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. Am J Respir Crit Care Med. 2013;188(8):941-7.

229. Savale L, Chaouat A, Bastuji-Garin S, Marcos E, Boyer L, Maitre B, et al. Shortened telomeres in circulating leukocytes of patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2009;179(7):566-71.

230. Lee J, Sandford AJ, Connett JE, Yan J, Mui T, Li Y, et al. The relationship between telomere length and mortality in chronic obstructive pulmonary disease (COPD). PLoS One. 2012;7(4):e35567.

231. Rode L, Bojesen SE, Weischer M, Vestbo J, Nordestgaard BG. Short telomere length, lung function and chronic obstructive pulmonary disease in 46,396 individuals. Thorax. 2013;68(5):429-35.

232. Nunes H, Monnet I, Kannengiesser C, Uzunhan Y, Valeyre D, Kambouchner M, et al. Is telomeropathy the explanation for combined pulmonary fibrosis and emphysema syndrome?: report of a family with TERT mutation. Am J Respir Crit Care Med. 2014;189(6):753-4.

233. Stanley SE, Chen JJ, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, et al. Telomerase mutations in smokers with severe emphysema. J Clin Invest. 2015;125(2):563-70.

234. Cartin-Ceba R, Swanson KL, Krowka MJ. Pulmonary arteriovenous malformations. Chest. 2013;144(3):1033-44.

235. Gorgy AI, Jonassaint NL, Stanley SE, Koteish A, DeZern AE, Walter JE, et al. Hepatopulmonary syndrome is a frequent cause of dyspnea in the short telomere disorders. Chest. 2015;148(4):1019-26.

236. Khincha PP, Bertuch AA, Agarwal S, Townsley DM, Young NS, Keel S, et al. Pulmonary arteriovenous malformations: an uncharacterised phenotype of dyskeratosis congenita and related telomere biology disorders. Eur Respir J. 2017;49(1).

237. Sgalla G, Iovene B, Calvello M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: pathogenesis and management. Respir Res. 2018;19(1):32.

238. Scanlon PD, Connett JE, Waller LA, Altose MD, Bailey WC, Buist AS, et al. Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study. Am J Respir Crit Care Med. 2000;161(2 Pt 1):381-90.

239. Aoshiba K, Nagai A. Senescence hypothesis for the pathogenetic mechanism of chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2009;6(7):596-601.

240. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AN, et al. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. Proc Natl Acad Sci U S A. 2006;103(35):13180-5.

241. Chilosi M, Poletti V, Rossi A. The pathogenesis of COPD and IPF: distinct horns of the same devil? Respir Res. 2012;13:3.

242. Tsuji T, Aoshiba K, Nagai A. Alveolar cell senescence in patients with pulmonary emphysema. Am J Respir Crit Care Med. 2006;174(8):886-93.

243. Sisson TH, Mendez M, Choi K, Subbotina N, Courey A, Cunningham A, et al. Targeted injury of type II alveolar epithelial cells induces pulmonary fibrosis. Am J Respir Crit Care Med. 2010;181(3):254-63.

244. Barkauskas CE, Cronce MJ, Rackley CR, Bowie EJ, Keene DR, Stripp BR, et al. Type 2 alveolar cells are stem cells in adult lung. J Clin Invest. 2013;123(7):3025-36.

245. Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet. 1998;19(1):32-8.

246. Hassock S, Vetrie D, Giannelli F. Mapping and characterization of the X-linked dyskeratosis congenita (DKC) gene. Genomics. 1999;55(1):21-7.

247. Angrisani A, Vicidomini R, Turano M, Furia M. Human dyskerin: beyond telomeres. Biol Chem. 2014;395(6):593-610.

248. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. Nature. 1999;402(6761):551-5.

249. Knight SW, Heiss NS, Vulliamy TJ, Greschner S, Stavrides G, Pai GS, et al. X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. Am J Hum Genet. 1999;65(1):50-8.

250. Wong JM, Collins K. Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. Genes Dev. 2006;20(20):2848-58.

251. Zeng XL, Thumati NR, Fleisig HB, Hukezalie KR, Savage SA, Giri N, et al. The accumulation and not the specific activity of telomerase ribonucleoprotein determines telomere maintenance deficiency in X-linked dyskeratosis congenita. Hum Mol Genet. 2012;21(4):721-9.

252. Migeon BR. The role of X inactivation and cellular mosaicism in women's health and sex-specific diseases. JAMA. 2006;295(12):1428-33.

253. Wu H, Luo J, Yu H, Rattner A, Mo A, Wang Y, et al. Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. Neuron. 2014;81(1):103-19.

254. Vulliamy TJ, Knight SW, Heiss NS, Smith OP, Poustka A, Dokal I, et al. Dyskeratosis congenita caused by a 3' deletion: germline and somatic mosaicism in a female carrier. Blood. 1999;94(4):1254-60.

255. Alter BP, Giri N, Savage SA, Peters JA, Loud JT, Leathwood L, et al. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. Br J Haematol. 2010;150(2):179-88.

256. Ballew BJ, Yeager M, Jacobs K, Giri N, Boland J, Burdett L, et al. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. Human genetics. 2013;132(4):473-80.

257. Baerlocher GM, Lansdorp PM. Telomere length measurements in leukocyte subsets by automated multicolor flow-FISH. Cytometry A. 2003;55(1):1-6.

258. Vaziri H, Schächter F, Uchida I, Wei L, Zhu X, Effros R, et al. Loss of telomeric DNA during aging of normal and trisomy 21 human lymphocytes. Am J Hum Genet. 1993;52(4):661-7.

259. Thouin MM, Giron JM, Hoffman EP. Detection of nonrandom X chromosome inactivation. Curr Protoc Hum Genet. 2003;Chapter 9:Unit9.7.

260. Wong JM, Kyasa MJ, Hutchins L, Collins K. Telomerase RNA deficiency in peripheral blood mononuclear cells in X-linked dyskeratosis congenita. Hum Genet. 2004;115(5):448-55.

261. Thumati NR, Zeng XL, Au HH, Jang CJ, Jan E, Wong JM. Severity of X-linked dyskeratosis congenita (DKCX) cellular defects is not directly related to dyskerin (DKC1) activity in ribosomal RNA biogenesis or mRNA translation. Hum Mutat. 2013;34(12):1698-707.

262. Russo T, Salvatore F, Cimino F. Determination of pseudouridine in tRNA and in acid-soluble tissue extracts by high-performance liquid chromatography. J Chromatogr. 1984;296:387-93.

263. Talebizadeh Z, Simon SD, Butler MG. X chromosome gene expression in human tissues: male and female comparisons. Genomics. 2006;88(6):675-81.

264. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev. 1998;12(4):527-37.

265. Ni J, Tien AL, Fournier MJ. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell. 1997;89(4):565-73.

266. Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell. 2014;159(1):148-62.

267. Batista LF, Pech MF, Zhong FL, Nguyen HN, Xie KT, Zaug AJ, et al. Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells. Nature. 2011;474(7351):399-402.

268. Janiszewska M, Liu L, Almendro V, Kuang Y, Paweletz C, Sakr RA, et al. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. Nat Genet. 2015;47(10):1212-9.

269. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing in situ. Science. 2014;343(6177):1360-3.

270. Trudeau MA, Wong JM. Genetic Variations in Telomere Maintenance, with Implications on Tissue Renewal Capacity and Chronic Disease Pathologies. Curr Pharmacogenomics Person Med. 2010;8(1):7-24.

271. Fong YW, Ho JJ, Inouye C, Tjian R. The dyskerin ribonucleoprotein complex as an OCT4/SOX2 coactivator in embryonic stem cells. Elife. 2014;3.

272. Chilosi M, Carloni A, Rossi A, Poletti V. Premature lung aging and cellular senescence in the pathogenesis of idiopathic pulmonary fibrosis and COPD/emphysema. Transl Res. 2013;162(3):156-73.

273. Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, et al. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. J Med Genet. 2008;45(10):654-6.

274. Vulliamy TJ, Walne A, Baskaradas A, Mason PJ, Marrone A, Dokal I. Mutations in the reverse transcriptase component of telomerase (TERT) in patients with bone marrow failure. Blood Cells Mol Dis. 2005;34(3):257-63.

275. Yamaguchi H, Calado RT, Ly H, Kajigaya S, Baerlocher GM, Chanock SJ, et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. N Engl J Med. 2005;352(14):1413-24.

276. Podlevsky JD, Chen JJ. It all comes together at the ends: telomerase structure, function, and biogenesis. Mutat Res. 2012;730(1-2):3-11.

277. Aubert G, Hills M, Lansdorp PM. Telomere length measurement-caveats and a critical assessment of the available technologies and tools. Mutat Res. 2012;730(1-2):59-67.

278. Fajkus J. Detection of telomerase activity by the TRAP assay and its variants and alternatives. Clin Chim Acta. 2006;371(1-2):25-31.

279. Rollison DE, Epling-Burnette PK, Park JY, Lee JH, Park H, Jonathan K, et al. Telomere length in myelodysplastic syndromes. Leuk Lymphoma. 2011;52(8):1528-36.

280. Kachuri L, Amos CI, McKay JD, Johansson M, Vineis P, Bueno-de-Mesquita HB, et al. Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. Carcinogenesis. 2016;37(1):96-105.

281. Zhang Y, Calado R, Rao M, Hong JA, Meeker AK, Dumitriu B, et al. Telomerase variant A279T induces telomere dysfunction and inhibits non-canonical telomerase activity in esophageal carcinomas. PLoS One. 2014;9(7):e101010.

282. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

283. Qiao D, Ameli A, Prokopenko D, Chen H, Kho AT, Parker MM, et al. Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. Hum Mol Genet. 2018.

284. Zaug AJ, Crary SM, Jesse Fioravanti M, Campbell K, Cech TR. Many disease-associated variants of hTERT retain high telomerase enzymatic activity. Nucleic Acids Res. 2013;41(19):8969-78.

285. Du HY, Pumbo E, Ivanovich J, An P, Maziarz RT, Reiss UM, et al. TERC and TERT gene mutations in patients with bone marrow failure and the significance of telomere length measurements. Blood. 2009;113(2):309-16.

286. Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, et al. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc Natl Acad Sci U S A. 2009;106(4):1187-92.

287. Aref S, El-Ghonemy MS, Abouzeid TE, El-Sabbagh AM, El-Baiomy MA. Telomerase reverse transcriptase (TERT) A1062T mutation as a prognostic factor in Egyptian patients with acute myeloid leukemia (AML). Med Oncol. 2014;31(9):158.

288. George G, Rosas IO, Cui Y, McKane C, Hunninghake GM, Camp PC, et al. Short telomeres, telomeropathy, and subclinical extrapulmonary organ damage in patients with interstitial lung disease. Chest. 2015;147(6):1549-57.

289. Calado RT, Brudno J, Mehta P, Kovacs JJ, Wu C, Zago MA, et al. Constitutional telomerase mutations are genetic risk factors for cirrhosis. Hepatology. 2011;53(5):1600-7.

290. Donaires FS, Scatena NF, Alves-Paiva RM, Podlevsky JD, Logeswaran D, Santana BA, et al. Telomere biology and telomerase mutations in cirrhotic patients with hepatocellular carcinoma. PLoS One. 2017;12(8):e0183287.

291. Du HY, Pumbo E, Manley P, Field JJ, Bayliss SJ, Wilson DB, et al. Complex inheritance pattern of dyskeratosis congenita in two families with 2 different mutations in the telomerase reverse transcriptase gene. Blood. 2008;111(3):1128-30.

292. Sugino S, Bortz BJ, Vaida S, Karamchandani K, Janicki PK. Peripartum Anesthetic Management and Genomic Analysis of Rare Variants in a Patient with Familial Pulmonary Fibrosis. A A Case Rep. 2015;5(10):169-72.

293. Mak AC, Tang PL, Cleveland C, Smith MH, Kari Connolly M, Katsumoto TR, et al. Brief Report: Whole-Exome Sequencing for Identification of Potential Causal Variants for Diffuse Cutaneous Systemic Sclerosis. Arthritis Rheumatol. 2016;68(9):2257-62.

294. Casuscelli J, Becerra MF, Manley BJ, Zabor EC, Reznik E, Redzematovic A, et al. Characterization and Impact of TERT Promoter Region Mutations on Clinical Outcome in Renal Cell Carcinoma. Eur Urol Focus. 2017.

295. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C. Silent (synonymous) SNPs: should we care about them? Methods Mol Biol. 2009;578:23-39.

296. Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet. 2011;12(10):683-91.

297. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011;43(12):1210-4.

298. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet. 2013;45(4):371-84, 84e1-2.

299. Pellatt AJ, Wolff RK, Herrick J, Lundgreen A, Slattery ML. TERT's role in colorectal carcinogenesis. Mol Carcinog. 2013;52(7):507-13.

300. Gao L, Thakur A, Liang Y, Zhang S, Wang T, Chen T, et al. Polymorphisms in the TERT gene are associated with lung cancer risk in the Chinese Han population. Eur J Cancer Prev. 2014;23(6):497-501.

301. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

302. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308-11.

303. Xin ZT, Beauchamp AD, Calado RT, Bradford JW, Regal JA, Shenoy A, et al. Functional characterization of natural telomerase mutations found in patients with hematologic disorders. Blood. 2007;109(2):524-32.

304. Yamaguchi H, Baerlocher GM, Lansdorp PM, Chanock SJ, Nunez O, Sloand E, et al. Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. Blood. 2003;102(3):916-8.

305. Vulliamy T, Marrone A, Dokal I, Mason PJ. Association between aplastic anaemia and mutations in telomerase RNA. Lancet. 2002;359(9324):2168-70.

306. Wilson DB, Ivanovich J, Whelan A, Goodfellow PJ, Bessler M. Human telomerase RNA mutations and bone marrow failure. Lancet. 2003;361(9373):1993-4.

307. Field JJ, Mason PJ, An P, Kasai Y, McLellan M, Jaeger S, et al. Low frequency of telomerase RNA mutations among children with aplastic anemia or myelodysplastic syndrome. J Pediatr Hematol Oncol. 2006;28(7):450-3.

308. Ly H, Calado RT, Allard P, Baerlocher GM, Lansdorp PM, Young NS, et al. Functional characterization of telomerase RNA variants found in patients with hematologic disorders. Blood. 2005;105(6):2332-9.

309. Marrone A, Stevens D, Vulliamy T, Dokal I, Mason PJ. Heterozygous telomerase RNA mutations found in dyskeratosis congenita and aplastic anemia reduce telomerase activity via haploinsufficiency. Blood. 2004;104(13):3936-42.

310. Errington TM, Fu D, Wong JM, Collins K. Disease-associated human telomerase RNA variants show loss of function for telomere synthesis without dominant-negative interference. Mol Cell Biol. 2008;28(20):6510-20.

311. Swift S, Lorens J, Achacoso P, Nolan GP. Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. Curr Protoc Immunol. 2001;Chapter 10:Unit 10.7C.

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

313. Hukezalie KR, Thumati NR, Côté HC, Wong JM. In vitro and ex vivo inhibition of human telomerase by anti-HIV nucleoside reverse transcriptase inhibitors (NRTIs) but not by non-NRTIs. PLoS One. 2012;7(11):e47505.

314. Latrick CM, Cech TR. POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. EMBO J. 2010;29(5):924-33.

315. Lee MS, Blackburn EH. Sequence-specific DNA primer effects on telomerase polymerization activity. Mol Cell Biol. 1993;13(10):6586-99.

316. Connett JE, Kusek JW, Bailey WC, O'Hara P, Wu M. Design of the Lung Health Study: a randomized clinical trial of early intervention for chronic obstructive pulmonary disease. Control Clin Trials. 1993;14(2 Suppl):3S-19S.

317. Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of Lung Health Study participants after 11 years. Am J Respir Crit Care Med. 2002;166(5):675-9.

318. Anthonisen NR, Connett JE, Kiley JP, Altose MD, Bailey WC, Buist AS, et al. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. JAMA. 1994;272(19):1497-505.

319. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meet ATS recommendations. Am Rev Respir Dis. 1981;123(6):659-64.

320. Alder JK, Barkauskas CE, Limjunyawong N, Stanley SE, Kembou F, Tuder RM, et al. Telomere dysfunction causes alveolar stem cell failure. Proc Natl Acad Sci U S A. 2015;112(16):5099-104.

321. Shay JW, Wright WE. Hayflick, his limit, and cellular ageing. Nat Rev Mol Cell Biol. 2000;1(1):72-6.

322. Daniali L, Benetos A, Susser E, Kark JD, Labat C, Kimura M, et al. Telomeres shorten at equivalent rates in somatic tissues of adults. Nat Commun. 2013;4:1597.

323. Saferali A, Lee J, Sin DD, Rouhani FN, Brantly ML, Sandford AJ. Longer telomere length in COPD patients with α1-antitrypsin deficiency independent of lung function. PLoS One. 2014;9(4):e95600.

324. Green ED, Guyer MS, Institute NHGR. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011;470(7333):204-13.

325. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet. 2008;9(4):255-66.

326. Kalmbach K, Robinson LG, Wang F, Liu L, Keefe D. Telomere length reprogramming in embryos and stem cells. Biomed Res Int. 2014;2014:925121.

327. Cohen S, Janicki-Deverts D, Turner RB, Casselbrant ML, Li-Korotky HS, Epel ES, et al. Association between telomere length and experimentally induced upper respiratory viral infection in healthy adults. JAMA. 2013;309(7):699-705.

328. Tsuji T, Aoshiba K, Nagai A. Cigarette smoke induces senescence in alveolar epithelial cells. Am J Respir Cell Mol Biol. 2004;31(6):643-9.

329. von Zglinicki T. Oxidative stress shortens telomeres. Trends Biochem Sci. 2002;27(7):339-44.

330. Amsellem V, Gary-Bobo G, Marcos E, Maitre B, Chaar V, Validire P, et al. Telomere dysfunction causes sustained inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2011;184(12):1358-66.

331. Franceschi C, Campisi J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. J Gerontol A Biol Sci Med Sci. 2014;69 Suppl 1:S4-9.

332. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, et al. Geroscience: linking aging to chronic disease. Cell. 2014;159(4):709-13.

333. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13(2):135-45.

334. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-24.

335. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017;169(7):1177-86.

336. Sharp A, Robinson D, Jacobs P. Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. Hum Genet. 2000;107(4):343-9.

337. Xu J, Khincha PP, Giri N, Alter BP, Savage SA, Wong JM. Investigation of chromosome X inactivation and clinical phenotypes in female carriers of DKC1 mutations. Am J Hematol. 2016;91(12):1215-20.

338. Hirvonen EAM, Peuhkuri S, Norberg A, Degerman S, Hannula-Jouppi K, Välimaa H, et al. Characterization of an X-chromosome-linked telomere biology disorder in females with DKC1 mutation. Leukemia. 2018.

339. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. 2013;41(Database issue):D262-7.

340. COHN WE. Some results of the applications of ion-exchange chromatography to nucleic acid chemistry. J Cell Physiol Suppl. 1951;38(Suppl. 1):21-40.

341. Jack K, Bellodi C, Landry DM, Niederer RO, Meskauskas A, Musalgaonkar S, et al. rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells. Mol Cell. 2011;44(4):660-6.

342. Ofengand J, Bakin A. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. J Mol Biol. 1997;266(2):246-68.

343. Karijolich J, Yi C, Yu YT. Transcriptome-wide dynamics of RNA pseudouridylation. Nat Rev Mol Cell Biol. 2015;16(10):581-5.

344. Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature. 2014;515(7525):143-6.

345. Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One. 2014;9(10):e110799.

346. Li X, Zhu P, Ma S, Song J, Bai J, Sun F, et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat Chem Biol. 2015;11(8):592-7.

347. Gilbert WV, Bell TA, Schaening C. Messenger RNA modifications: Form, distribution, and function. Science. 2016;352(6292):1408-12.

348. Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G. The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. J Biol Chem. 2002;277(47):45235-42.

349. Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT. Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J Cell Biol. 2006;173(2):207-18.

350. Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell. 2002;109(2):145-8.

351. McMahon M, Contreras A, Ruggero D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. Wiley Interdiscip Rev RNA. 2015;6(2):173-89.

352. Ho NW, Gilham PT. Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. Biochemistry. 1971;10(20):3651-7.

353. Ofengand J, Del Campo M, Kaya Y. Mapping pseudouridines in RNA molecules. Methods. 2001;25(3):365-73.

354. Tomikawa C, Yokogawa T, Kanai T, Hori H. N7-Methylguanine at position 46 (m7G46) in tRNA from Thermus thermophilus is required for cell viability at high temperatures through a tRNA modification network. Nucleic Acids Res. 2010;38(3):942-57.

355. Dudley E, Bond L. Mass spectrometry analysis of nucleosides and nucleotides. Mass Spectrom Rev. 2014;33(4):302-31.

356. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet. 2010;6(12):e1001247.

357. Su D, Chan CT, Gu C, Lim KS, Chionh YH, McBee ME, et al. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. Nat Protoc. 2014;9(4):828-41.

358. Fu D, Collins K. Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation. Mol Cell. 2007;28(5):773-85.

359. Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. 2006;34(Database issue):D145-9.

360. Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, et al. MODOMICS: a database of RNA modification pathways. 2008 update. Nucleic Acids Res. 2009;37(Database issue):D118-21.

361. Gillery P, Georges N, Wegrowski J, Randoux A, Borel JP. Protein synthesis in collagen lattice-cultured fibroblasts is controlled at the ribosomal level. FEBS Lett. 1995;357(3):287-9.

362. Halle JP, Müller S, Simm A, Adam G. Copy number, epigenetic state and expression of the rRNA genes in young and senescent rat embryo fibroblasts. Eur J Cell Biol. 1997;74(3):281-8.

363. Yi X, Tesmer VM, Savre-Train I, Shay JW, Wright WE. Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. Mol Cell Biol. 1999;19(6):3989-97.

364. He J, Navarrete S, Jasinski M, Vulliamy T, Dokal I, Bessler M, et al. Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. Oncogene. 2002;21(50):7740-4.

365. Lodish H, Berk A, Zipursky S. Section 11.6 Processing of rRNA and tRNA. Molecular Cell Biology 4th Edition: W. H. Freeman & Co.; 2000.

366. King TH, Liu B, McCully RR, Fournier MJ. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. Mol Cell. 2003;11(2):425-35.

Appendix

Appendix A Quantification of pseudouridine levels in cellular RNA pools with a modified HPLC-UV assay

The following material has previously been published as:

Xu J, Gu AY, Thumati NR, Wong JMY. Quantification of pseudouridine levels in cellular RNA pools with a modified HPLC-UV assay. *Genes* 2017; 8(9): 219.

It describes an HPLC-UV assay used for the quantification of ψ level in Chapter 2. Specifically, part of Section 2.4.4 (results presented in Figure 2.8) was measured by the following method.

A.1 Abstract

Pseudouridine (Ψ) is the most abundant post-transcriptionally modified ribonucleoside. Different Ψ modifications correlate with stress responses and are postulated to coordinate the distinct biological responses to a diverse panel of cellular stresses. With the help of different guide RNAs, the dyskerin complex pseudouridylates ribosomal RNA, small nuclear RNA and selective messenger RNAs. To monitor Ψ levels quantitatively, a previously reported high-performance liquid chromatography method coupled with ultraviolet detection (HPLC-UV) was modified to determine total Ψ levels in different cellular RNA fractions. Our method was validated to be accurate and precise within the linear range of 0.06 – 15.36 pmol/μL and to have absolute Ψ quantification levels as low as 3.07 pmol. Using our optimized HPLC assay, we found that 1.20% and 1.94% of all ribonucleosides in nuclear-enriched RNA and small non-coding RNA pools from the HEK293 cell line, and 1.77% and 0.98% of ribonucleosides in 18S and 28S rRNA isolated from the HeLa cell line, were pseudouridylated. Upon knockdown of

dyskerin expression, a consistent and significant reduction in total Ψ levels in nuclear-enriched RNA pools was observed. Our assay provides a fast and accurate quantification method to measure changes in Ψ levels of different RNA pools without sample derivatization.

Keywords: pseudouridine; dyskerin; HPLC; RNA modification

A.2 Introduction

Of the 166 known RNA nucleoside modifications, pseudouridine (Ψ) is the most abundant post-transcriptional modification and was the first to be discovered (339, 340). The isomerization of 1-ribosyluracil (uridine) to 5-ribosyluracil provides an additional hydrogen bond donor on the uracil base that contributes to the stabilization of base-stacking and RNA structures (71).

Known as the fifth ribonucleoside, Ψ is widely distributed in almost all RNA-containing species and constitutes over 1% of total nucleosides and over 7% of all uridines in ribosomal RNA (rRNA) fractions. Ψ has been shown to enhance transfer RNA (tRNA) binding to ribosomes and translational fidelity of rRNAs (341, 342). It is also involved in pre-mRNA splicing regulation in spliceosomal small nuclear RNAs (snRNA) and in decoding and stabilizing tRNA (reviewed in (343)).

Recently, with the advent of massive parallel sequencing of the transcriptome, hundreds of naturally modified Ψ sites in mRNA were discovered in yeast and human cells (266, 344, 345). Specifically, around 0.2-0.6% of uridines in mammalian mRNA are pseudouridylated (346). Pseudouridylation profile in mRNA was dynamically regulated in a stress-specific manner (266, 344, 346). Although the biological consequences of mRNA pseudouridylation remain unknown, it is postulated that the irreversible changes in the pseudouridylation profile of the transcriptome could result in functional changes including protein recoding, reduced translational efficiency, and altered transcript structure (347).

In humans, there are 13 known Ψ synthases (PUS) that can catalyze the isomerization of uridine to Ψ (70). Based on their respective mechanisms of action, these enzymes can be classified into RNA-independent and RNA-dependent PUS families. RNA-independent PUS enzymes promote Ψ formation with classical enzymatic catalysis, through protein-RNA interactions between the enzyme's catalytic domain and the structural context of the modified target uridine. In contrast, Ψ modifications of specific rRNA and snRNA residues are carried out by the box H/ACA ribonucleoprotein (RNP) complex in a guide RNA-dependent manner.

The H/ACA RNP complex is comprised of one box H/ACA small nucleolar/Cajal body RNA (snoRNA or scaRNA) and two copies of a protein catalytic subunit comprised of dyskerin, GAR1, NHP2 and NOP10. The catalytic core, formed by the dyskerin/NHP2/NOP10 trimers complexed with a single guide RNA, was assembled through sequential binding with biogenesis factors including SHQ1 and NAF1 (348). GAR1 brings the mature H/ACA RNP assembly to Cajal bodies and nucleoli, where nascent target RNA are modified (349). The specificity of pseudouridylation in this system comes from base-pair interactions between the antisense elements in the box H/ACA snoRNA and the sequence of the target RNA. Guided by the snoRNA, the dyskerin/NHP2/NOP10 complex carries out pseudouridylation of specific uridine residues in non-coding, as well as coding RNA (71).

While the dyskerin/NHP2/NOP10 complex is the only known RNA-dependent PUS, it is responsible for the modification of various targets, by virtue of its ability to switch RNA guides. All box H/ACA guide RNAs are non-coding small RNAs with a hairpin-hinge-hairpin structure and the signature ACA residues at the 3' terminus at maturation (350). The most abundant H/ACA-RNAs are found localized at the nucleolus (H/ACA-snoRNAs). These guide RNAs are responsible for directing the majority of rRNA Ψ modifications. In comparison, H/ACA-

scaRNAs have much greater structural diversity than H/ACA-snoRNAs, with a subset of scaRNAs containing both guiding Ψ (H/ACA) and guiding 2'O-methyl (structural C/D box motif) modifications. H/ACA-scaRNAs are found localized to the Cajal body and are involved in the nucleoside modifications of selected small nuclear RNAs involved in the spliceosome complex (351).

Since Ψ's discovery, multiple methods have been developed to map the location and quantify the levels of Ψ. The most popular method for studying Ψ in RNA is based on chemical derivatization of RNA by N-cyclohexyl-N'-β-(4-methylmorpholinium)ethylcarbodiimide p-tosylate (CMCT) (352). CMCT reacts with all guanosine, uridine, guanosine-like and uridine-like residues to form adducts. A subsequent cleavage step hydrolyzes all adducts except those formed at the Ψ bases (342, 353). As the bulky CMC group at the Ψ site sterically stops reverse transcription, truncated products from the primer extension assay with reverse transcriptase will be able to sequence-specifically determine the location of each Ψ modification site. Although this method is useful for the determination of site-specific pseudouridylation, the derivatization, cleavage and reverse transcription steps introduce variables that may influence accurate quantification.

In contrast, the traditional physicochemical properties-based chromatography method provides an alternative way to quantify relative Ψ levels by simple enzymatic digestion of cellular RNA into single nucleosides. Upon separation of different nucleosides by high-performance liquid chromatography (HPLC), a few detectors have been applied to determine Ψ levels in cellular RNA pools. Absorbance detectors, such as ultraviolet (UV) detectors which is based on light absorbing capacity of the analytes, are the most commonly used ones for HPLC analysis. Tomikawa *et al*. established an HPLC-UV method for the detection and determination

of modified RNA (354). The method allows the detection of ten different modified RNA nucleosides from the same digested RNA sample within a single injection. However, it was used as a semi-quantitative method for the comparison of relative RNA modification levels among different bacterial strains. In addition, the chromatographic conditions were not ideal (it required 85 minutes for the analysis of each sample).

Mass spectrometry (MS) is also a popular choice of detectors for the determination of Ψ levels in recent years (355). The fragmentation of molecules by electric fields provides individual molecules different base-to-charge ratios. Triple quadrupole MS was applied to the detection of modified nucleosides in a 40-min HPLC-MS run (356). The high sensitivity and specificity of MS facilitated the quantification of over 20 different modified nucleosides simultaneously. With the improvement of separation on this method, a modified UPLC-MS method was also reported where a single sample could be analyzed within 15 minutes (357). Modifications of low prevalence could be detected by LC/MS. However, the analysis of comprehensive RNA modification profile may not be necessary for a lot of researchers, considering the high cost of MS analysis and the availability of facilities.

In the current study, a previously reported HPLC-UV method (262) was modified to determine the Ψ levels in various cellular RNA pools. The modified method was successful in detecting changes in Ψ levels upon knockdown of dyskerin expression. Our measurement of Ψ levels remained precise and accurate at picomole levels. We contend that the modified HPLC-UV assay provides an economic option for rapid screening of small changes in Ψ level.

A.3 Methods

A.3.1 Chemicals and expression vector

The chemical reference compounds Ψ and 7-methylguanosine (7-metG) were obtained from Santa Cruz Biotechnology and Sigma, respectively. Nucleoside test mix from Sigma was used as a reference to identify major nucleosides (A, U, G, C) in RNA pools. The aqueous mobile phase was prepared with ammonium dihydrophosphate from Sigma. Methanol (HPLC grade) was obtained from Fisher Scientific. Ultrapure water was obtained from a MilliQ water purification system (Millipore). A previously described small hairpin (sh) RNA construct was applied to reduce dyskerin expression level by RNA interference (358). It was a generous gift from Dr. Kathleen Collins' laboratory.

A.3.2 Cell culture and transfection

The human embryonic kidney cell line HEK 293 was obtained from American Type Culture Collection. The human cervical adenocarcinoma cell line Hela was a generous gift from Dr. Eric Jan's laboratory. Cells were cultured in 1x DMEM high glucose media (Gibco) supplemented with 5% FBS and maintained at 37^{o}C with 5% CO2. Cells were trypsinized, quantified using a Coulter counter (Beckman Coulter), and seeded at a density of 3 million cells per 100 mm plate. Cells were allowed to recover overnight until they were 40 – 60% confluent at the time of transfection.

Standard calcium phosphate mediated transfection was performed. The cells were maintained under puromycin selection to inhibit the growth of cells that did not express the shRNA plasmid. Cells were harvested 96, 120, and 144 hours after the transfection.

A.3.3 RNA extraction and digestion

For the collection of nuclear-enriched RNA pool, cells were resuspended in hypotonic lysis buffer (20 mM HEPES pH 8.0, 2 mM $MgCl_2$, 0.2 mM EGTA, 10% glycerol, 1mM DTT, 0.1 mM PMSF) and subjected to four consecutive freeze-thaw cycles in liquid nitrogen and a 37oC water bath. Following centrifugation at 1,000 g for 15 minutes, the supernatant containing hypotonic lysis buffer and cytoplasmic components was removed. The remaining pellet was washed once with hypotonic lysis buffer. The nuclear RNA pool was then extracted with Trizol (Invitrogen).

Total RNA was extracted with Trizol (Invitrogen). Total RNA was separated by size on denaturing urea-polyacrylamide (PAGE) gel electrophoresis and the gel fraction between the region with xylene cyanol dye migration (around 75 nt) and 5S rRNA (121 nt) was harvested. The excised band was sliced into small pieces with a clean razor and moved to an 15 ml conical tube. 3x vol. of 0.3M NaOAc, pH 5.2 and 0.1x vol. of phenol chloroform were added to the sliced gel pieces. The tube was kept shaking at 37^{o}C overnight. Nucleic acid in the supernatant was extracted and precipitated with ethanol. Similarly, 18S and 28S rRNA were separated by size using a denaturing formaldehyde agarose gel, gel purified by the Ultrafree-DA centrifugal filter unit (Millipore), and precipitated with ethanol.

RNA levels were quantified using Nanodrop (Nanodrop Technologies) and the concentrations of selected RNA samples were verified with the Ribogreen assay (Thermo Fisher). 5 μg of each RNA sample was sequentially hydrolyzed by 5 units of RNase T2 (Worthington Biochem) and dephosphorylated by 5 units of Shrimp Alkaline Phosphatase (Invitrogen) following two overnight incubations at 37^{o}C. The final volume of each digested

RNA sample was adjusted to 200 μl and 210 pmol of 7-metG was added to each sample as an internal control.

A.3.4 Chromatographic conditions

Nucleoside separation and quantification were performed with a Waters 2695 HPLC system and Waters 2996 UV detector (Waters). Chromatographic separation was performed with a 4 μm Waters Nova-Pak C18 column 3.9 mm x 150 mm (Waters). The column was kept at room temperature and the detection wavelength was set at 254 nm. The two mobile phase components were adapted from a previous study (262). Mobile phase A (0.01M ammonium dihydrophosphate adjusted with phosphoric acid to pH 5.1) was filtered through a 0.22 μm membrane filter (Millipore) before use. Mobile phase B was composed of methanol/water (1:1 v/v). A linear gradient was programmed from 0% to 40% mobile phase B over the first 30 minutes, followed by 40% to 0% in the next minute, and then re-equilibration with 100% mobile phase A for 4 minutes (262). The flow rate was set at 1 mL/min and the injection volume was 50 μl. Data acquired were processed with Empower software.

A.3.5 Method validation

Stock solutions of Ψ was made by dissolving the appropriate amount of pure Ψ in water to yield a final concentration of 200 ng/μl. Working solutions with concentrations of 0.10, 0.26, 0.64, 1.6, 4, 10 and 25 ng/μl was prepared every month by a 2.5-fold serial dilution from the stock. Calibration curves for Ψ levels were created on each analysis day by diluting 30 μl of working solution with water and digestion enzyme buffers into 200 μl (the final concentration of Ψ is equivalent to 0.06, 0.16, 0.39, 0.98, 2.46, 6.14, and 15.36 pmol/μl). The analyte to internal standard (I.S.) peak area ratios were plotted against matched amounts of Ψ added in the injected blank sample. The calibration curves were calculated by the least squares method. Linearity was

assessed by determining the coefficient of correlation (r2) of data points on the plotted curves. The absolute Ψ amount was expressed as picomoles per microliter (pmol/μL) and converted to picomoles later.

For method validation, spiked samples were prepared at three different concentrations that cover the low, middle and high ranges of the standard curve. Relative error (RE) between the nominal and measured concentrations was expressed as accuracy of the method, whereas relative standard deviation (RSD) of repeated measurements was expressed as precision.

Sensitivity of the method was estimated using the calibration curve method. The limit of detection (LOD) and limit of quantitation (LOQ) of the present method were calculated by the following formula: $A = k\sigma/S$, where A was LOD or LOQ, k was the coefficient for the two parameters (k = 3.3 for LOD and k = 10 for LOQ), σ was the standard deviation of the response (i.e. the intercept of the calibration curve), and S was the slope of the curve.

Specificity of the nucleosides in RNA pools was tested by injecting nucleoside test mix and referring to the relative retention time in the standard.

A.3.6 Protein expression measurement by western blot

The western blot protocol was the same as described in (337). Whole cell extracts were quantified by Bradford protein assay and 30 μg of protein was resolved in 10% SDS-PAGE gel. After transfer of the resolved protein samples to polyvinylidene fluoride membranes, the blots were incubated with anti-dyskerin rabbit polyclonal antibodies (200 ng/ml, 1:1,000 dilution, Santa Cruz Biotech) and anti-β-actin mouse monoclonal antibody (50 ng/ml, 1,40,000 dilution, Sigma). Alexa Fluor 680 goat anti-rabbit IgG and Alexa Fluor 790 donkey anti-mouse IgG were used as secondary antibodies (both from Thermo Fisher, 1:10,000 dilution). The signals were detected by Licor Odyssey CLx Infrared Imaging System and analyzed by ImageJ software.

A.3.7 Statistical Analysis

Data were analyzed using GraphPad Prism software (GraphPad Software, San Diego, CA). Error bars denote SEM. The one-way ANOVA with *post-hoc* Bonferroni correction was applied to adjust for multiple comparisons. Differences were considered significant at $p < 0.05$.

A.4 Results

A.4.1 Method validation

Representative chromatograms of a calibration standard sample and a digested RNA sample are shown in (Figure 4.1A & Figure 4.1B), respectively. The retention times for Ψ and 7-metG were 2.0 and 6.5 minutes. Although other RNA nucleosides were not involved in our quantification method, their retention times were also stable and listed in Table 4.1.

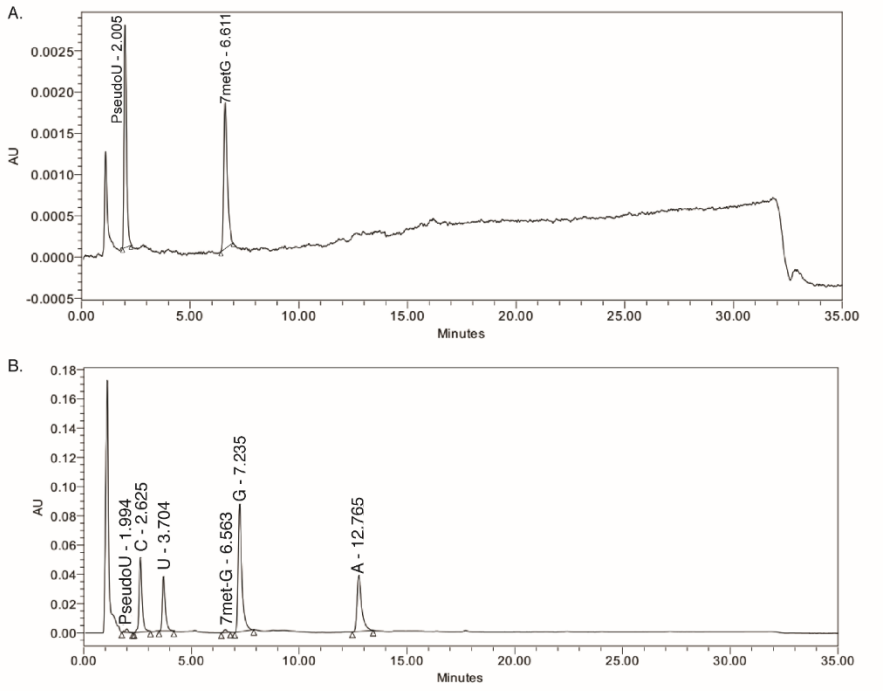

Figure 4.1 Representative HPLC chromatograms

(A) a Ψ quantification standard sample at 2.46 pmol/μL and (B) a digested nuclear-enriched RNA sample from the HEK293 cell line.

Table 4.1 Retention times for major nucleosides using the HPLC-UV method.

Analyte	Retention Time (min)
Ψ	1.97
Cytosine	2.67
Uridine	3.67
7-methyl Guanosine (I.S.)	6.49
Guanosine	7.15
Adenosine	12.66

The peak area ratios of the calibration standard were proportional to the concentration of Ψ input over the range of 0.06 – 15.36 pmol/μL. The means ± SDs of four standard curve slopes and their intercepts for Ψ were 0.768 ± 0.059 and 0.002 ± 0.001, respectively. The regression coefficients (r2) of all calibration curves were greater than 0.999. Therefore, the LOD for this method was estimated to be 0.01 pmol/μL and LOQ estimated to be 0.02 pmol/μL.

Assay accuracy and precision were shown in Table 4.2. Since the above data showed that our method was valid for measuring a wide range of Ψ levels, we first determined the average Ψ levels in different cellular RNA pools and then applied this validated method to RNAi-mediated knockdown of dyskerin expression.

Table 4.2 Accuracy and precision of the HPLC-UV method.

Spiked conc. (pmol/μl)	Intra-day (n = 3)			Inter-day (n = 3)		
	Observed conc. (pmol/μl)	Precision (RSD, %)	Accuracy (RE, %)	Observed conc. (pmol/μl)	Precision (RSD, %)	Accuracy (RE, %)
0.19	0.20 ± 0.01	5.70	104.19	0.21 ± 0.02	8.34	108.12
1.54	1.66 ± 0.01	0.50	108.07	1.66 ± 0.04	2.92	108.70
12.28	12.90 ± 0.25	1.86	107.02	12.90 ± 0.26	2.00	106.37

Table 4.3 Average ψ levels in different RNA pools.

Cell Line	Fraction	RNA Size (nt)	Conc. (pmol/μl)	Ψ%	RSD (%)
HEK293	Nuclear-enriched RNA	N/A	0.88	1.20%	8.84%
HEK293	Small RNA	75 – 121	1.43	1.94%	14.77%
Hela	18S rRNA	1868	1.30	1.77%	0.87%
Hela	28S rRNA	5025	0.72	0.98%	0.76%

A.4.2 Average Ψ levels in different cellular RNA pools

From the standard curve, the mean concentration of Ψ in nuclear-enriched RNA samples was found to be 0.88 pmol/μL. The total amount of Ψ involved in the injected sample (50 μL of RNA sample containing 1.25 μg of nuclear RNA) was calculated to be 44.16 pmol. Since 1.25 μg of the starting material, i.e. intact nuclear RNA is equal to 3.68 nmol, we concluded that approximately 1.20% of nucleosides in nuclear-enriched RNA are Ψs. This is in agreement with previous reports (71).

Using the same method, the amount of Ψ in 18S and 28S rRNA from HeLa cells in log-phase growth were calculated to be 1.77% and 0.98%, respectively (Table 4.3). It is known that in 28S rRNA (found in the 60S large ribosomal subunits), 57 out of 5025 nucleosides (1.13%) could be pseudouridylated (339). Our result indicated that 86.7% of the maximum calculated level of Ψs in 28S rRNA in HeLa cells were indeed pseudouridylated. For 18S rRNA, 1.87% (35/1868) of total nucleosides are reported to be pseudouridylated (339), and our result of 1.77% was comparable to this value (94.7% of the maximum level). Thus, our experimental data agreed with published data on the extent of Ψ modifications within rRNAs.

It was also found in our study that 1.94% of gel-purified small RNA populations are pseudouridylated. The average Ψ concentration measured by our assay, and the equivalent amount and levels in different cellular RNA pools are listed in Table 4.3. While it is beyond the scope of our current report, we have previously shown that our Ψ quantification method can expand to include the measurement of Ψ levels in RNAs isolated through purification of specific RNP complexes, such as the 40S and 60S ribosomal subunits (261), thereby restricting Ψ analysis in functionally distinct populations of mature rRNAs. Similarly, we envisioned that Ψ

quantifications of RNA components in specific RNP complexes could be analyzed following immunoprecipitation-purification.

A.4.3 Ψ reduction is not proportional to dyskerin knock-down

Next, we applied our assay to quantify changes in Ψ levels upon the reduction of dyskerin expression. Only three out of thirteen known PUS can catalyze pseudouridylation in rRNAs (70). Out of the three enzymes, the dyskerin complex is the only mechanism that uses guide RNAs to target the positions of rRNA pseudouridylation. At the time of our work, the dyskerin complex is known to modify 35 sites on 18S rRNA, 57 sites on 28S rRNA and 2 sites on 5.8S rRNA comprising the major rRNA Ψ modification mechanism (70, 339, 359, 360).

We reasoned that as rRNA are known to have long half-lives (vary from 3 to 8 days in mammalian cells (361-363)), stable rRNA pools in cytoplasm may not reflect the changes in Ψ modification rates upon dyskerin knockdown. To enrich for nascent rRNA, we isolated the nuclear-enriched cell fractions, where newly transcribed rRNAs are modified. We expected to see a proportional relationship between Ψ levels and dyskerin protein expression in the nuclear fractions.

The HEK293 cells were directed to express an shRNA duplex against dyskerin and kept under antibiotic (puromycin) selection for at least 48 hours. Antibiotic selection inhibited the growth of cells that were negative for the expression of the shRNA construct. With this approach, we confirmed that dyskerin expression in HEK293 cells was reduced to about 45% of regular levels (Figure 4.2A and Figure 4.2B, $p = 0.0188$). As expected, there was also a consistent and significant reduction in Ψ levels in nuclear RNA fractions 96, 120 and 144 hours post-shRNA transfection (Figure 4.2C and Figure 4.2D, $p = 0.0291$). The reduction was not directly proportional to the reduction in dyskerin expression. Although HEK293 cells showed

widespread cell death (data not shown), suggesting that they were under severe stress with the reduction in dyskerin expression, the average Ψ levels in these shRNA-treated cells were only around 10% lower than the control group (no dyskerin shRNA transfection).

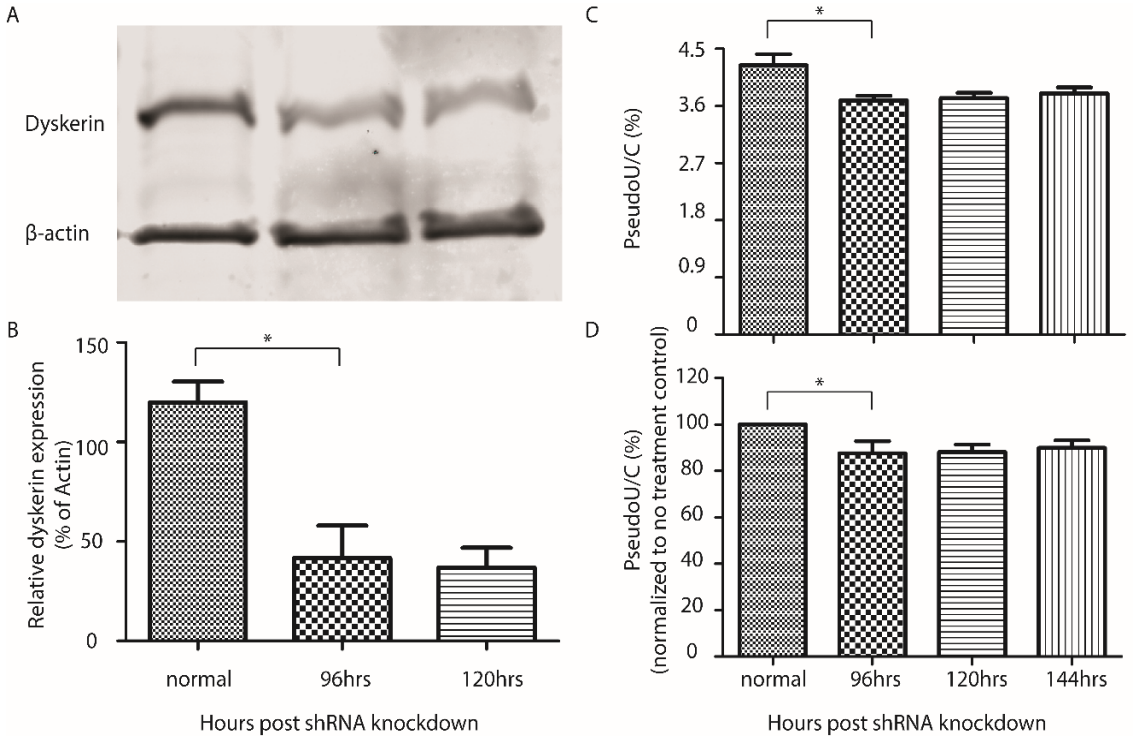

Figure 4.2 A reduction of ψ was observed after dyskerin knock-down.

(A) & (B) Representative immunoblot and quantification of dyskerin protein expression at 96 and 120 hours after dyskerin knockdown with shRNA. (C) & (D) Change in Ψ expression after dyskerin knockdown with shRNA. Ψ level was expressed as a function of cytosine (Ψ/C). It is worthy to note that the Ψ/C (%) is merely the ratio of the integrated absorbance, i.e., peak areas, for Ψ and C. To facilitate the comparison of Ψ levels at different time points after the knockdown of dyskerin expression, the Ψ/C ratio was further normalized to the baseline level and illustrated in Figure 4.2D.

A.5 Discussion

We developed a quantitative HPLC-UV-based assay to determine Ψ, the most abundant post-transcriptionally modified RNA base. We found that Ψ made up of 1.20% and 1.94% of all nucleosides in nuclear RNA and small RNA pools from the HEK293 cell line and 0.98% and

1.77% of nucleosides in 28S and 18S rRNA from the HeLa cervical cancer cell line. These results are comparable to previous reports (71, 339).

Our linear range for quantification was 0.06 – 15.36 pmol/μL, which is equivalent to 3.07 – 767.81 pmol of Ψ in this assay. We calculated the mean amount of Ψ in 1.25 μg of nuclear RNA pools to be 44.16 pmol. With the assay being able to determine Ψ levels as low as at the picomole level, theoretically, the minimal amount of RNA input prior to the digestion step could be as low as about 100 ng (assuming the average Ψ content of the RNA sample being 1.20%, the same as the level in nuclear RNA pools in our assay). This is an easily collected sample amount from biochemical experiments, as well as from clinical/patient sources.

The LOD and LOQ for our assay were estimated to be 0.01 pmol/μL and 0.02 pmol/μL, respectively. These values equaled 0.31 and 0.92 pmol of Ψ or 10 and 31 ng of RNA input (assuming the percentage of pseudouridylation in the RNA sample for detection is around 1%). As well, the intra-day and inter-day coefficients of variation for precision and accuracy were below 10% for concentrations tested within the quantification range. In other words, minimal changes in Ψ levels as low as 0.31 pmol in RNA samples may be accurately and precisely analyzed by the current assay. Therefore, our assay provides a sufficient window for measuring changes in pseudouridylation under different conditions.

Based on the observed retention times, the major nucleosides were well separated. Importantly, Ψ had a stable retention time of 2.0 minutes that allowed it to be easily identified. Our method has already been applied to quantify the steady-state Ψ levels in rRNA samples from X-DC patients with mutations in the dyskerin-encoding *DKC1* gene (261, 337). We found that specific mutations of the *DKC1* gene, but not all disease-associating mutations, led to reproducible but modest loss of steady-state Ψ levels in rRNA. We contended that while these

moderate changes in rRNA Ψ levels are not enough to drive the disease, they could nonetheless modulate the rate of disease onset and/or severity by a mechanism that involves reduced stress tolerance conferred by optimal rRNA Ψ modifications.

The observed non-proportional relationship between Ψ levels and dyskerin expression may be explained by the existence of different mechanisms for pseudouridylation. There are at least 12 additional PUS known in humans, on top of the H/ACA RNP complex. Specifically, pseudouridylations in rRNA by PUS7 and PUS7L are carried out in a guide-RNA independent manner and are unaffected upon the knockdown of dyskerin expression (70). On the other hand, *DKC1* is an essential gene in metazoan, and knockout of *DKC1* leads to embryonic lethality in mouse models (364). It is thus conceivable that reduction of dyskerin expression beyond a threshold level is cytotoxic. In agreement, we observed that despite the use of puromycin selection over the course of 144 hours, dyskerin protein expression did not reduce beyond 55%. This level of dyskerin expression may represent the minimal level compatible with life and thus, reduction of dyskerin expression to these levels may cause minimal disruptions to rRNA modification and the structure of the protein synthesis machinery.

It is worthy to note that the nuclear-enriched RNA pool we collected from our shRNA experiments is comprised of almost all organelles, including the nucleus, ribosomes, endoplasmic reticulum, Golgi body, cell membrane, and mitochondria. This RNA collection contains not only nascent rRNA, but also includes mRNA, tRNA, and other small RNAs. As dyskerin is not known to be substantially involved in the modifications of these other RNA populations, with the notable exceptions of snRNA and a small percentage of mRNAs, Ψ levels in these RNA fractions will not be affected by dyskerin knockdown. Even though rRNAs constitute over 80% of total cellular RNA (365), the nascent rRNA population residing in the

nucleus is expected to be substantially less. In this case, our modest reduction in Ψ levels could also be explained by the relative abundance of co-purifying tRNA and mRNA in our analyte.

Previously, functional studies of Ψ focused on the different pseudouridylation levels at specific Ψ residues in rRNA, tRNA and mRNA (341, 344, 366). While this information is invaluable to study the mechanism of Ψ regulation, mapping the changes in individual Ψ targets is time-consuming and may not reflect comprehensive Ψ regulation. With our assay, the global Ψ levels in a specific RNA pool can be quantified to serve as a quick readout of changes in Ψ levels (either in total RNA or in an isolated RNA pool) in response to different types of cellular stress including, but not limited to, heat shock, hypoxia, and modified growth conditions. The changes in global Ψ levels, either in a dose-response or in a temporal manner, can then be correlated with viability, toxicity and/or other functional endpoints. Once the consequences associated with pseudouridylation defects are observed, further studies could then be performed to determine the mechanisms, i.e., screening for the target residues of pseudouridylation that are responsible for the outcome.

With only one digestion step prior to analysis, our method provides accurate quantification of Ψ levels without any additional sample derivatization or manipulation. It has been used to measure Ψ levels *in vivo* and to measure changes in pseudouridylation over time. Our fast and economical HPLC-UV method can also be applied to detect and determine low levels of Ψ in different RNA fractions with small amounts of RNA input.