Different Faces of the Human Telomeric Protein TPP1

by

Sherilyn A. Grill

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
in the University of Michigan
2019

Doctoral Committee:

Assistant Professor Jayakrishnan Nandakumar, Chair
Professor Ursula Jakob
Professor John V. Moran
Associate Professor Lyle A. Simmons

Sherilyn A. Grill

sagrill@umich.edu

ORCID iD: 0000-0003-3201-0214

“Friends are the family you get to choose for yourself”

To Dad, for teaching me to be rational in an irrational world
To Mom, for teaching me to be just crazy enough to believe that I can do anything
And to Mandy, for being the sister I was given, but the best friend I chose. I love you.

Acknowledgements

First and foremost, I would like to thank JK. It seems almost silly to try to put into words all of the support, inspiration, and encouragement JK has given me for the past six years. JK's enthusiasm for science is contagious, and from the very beginning (literally the first day of graduate school) I knew that I wanted to work in a lab where that enthusiasm radiated through every student and post-doc that had the pleasure of working there. Throughout my entire stay in the Nandakumar lab, I can say with confidence that this was the case. JK cultivated an environment where we could learn, make mistakes, test the limits of any scientific question, and most importantly have fun while doing it. He has mastered the art of bringing together a group of people that not only complement each other, but who also push each other to be better scientists, mentors, and critical thinkers. JK himself embodies all of these traits, as his unwavering love of science, constant drive, and unparalleled scientific acumen was an inspiration to watch. In the Nandakumar lab no question was ever too silly to ask, and no idea was every too crazy to propose. This freedom to think critically was the most valuable thing anyone could have given me during my time in graduate school. To this day I cannot thank JK enough for the hundreds of hours that we spent thinking, talking, and brainstorming about science, because those hours truly made me in to the scientist I am today.

To the current and former Nandakumar lab members, thank you for creating the best lab environment I could have asked for. To Eric, Valerie, Kamlesh, Devon, Oana, Alam, Shilpa, Ritvija, Jackie, Jonathan, John, and Cassie thank your friendship, support, and shared obsessions with coffee, bubble tea, tequila, and all the other vices we indulged in over the past six years. I want to specifically thank Kamlesh for keeping us sane during those early years in the Nandakumar lab by never taking anything too seriously. I want to thank Valerie for being the most incredible friend and mentor I have ever met. I am so lucky that for the last six years I have gotten to witness, benefit, and learn from your scientific brilliance. Without you, I would not be here. And finally, I would like to thank Eric, for your unwavering support, constant encouragement, and your never-ending ability to amaze me with your determination. You inspire me to be a better friend and scientist, and for that I am eternally grateful.

I would also like to take this opportunity to thank all of the scientists who helped mentor me along the way. To Dr. Steve Clark, thank you for taking a chance on an overly exuberant undergrad. You and Dr. Lindsey Gish introduced me to lab work in the most inclusive, exciting, and illuminating way possible. I am in graduate school because of the excitement you two passed down to me. I would also like to thank my committee members, Dr. Ursula Jakob, Dr. John Moran, and Dr. Lyle Simmons for their support, feedback, and suggestions over the past five years. To Lyle, thank you for being a second mentor to me, for talking to me whenever I needed, and for always helping me as if I were your own student. I would like to thank Dr. Sue Hammoud and Adrienne Shami, for your help throughout the last year, and for inspiring my love of germ cell biology. I would like to thank the members of the Shelterin Group, including

Dr. Ivan Maillard, Dr. Catherine Keegan, Eric Perkey, and Ann Friedman, for their continued support throughout my time in the Nandakumar lab. Finally, I would like to thank the Genetics Training Program (T32GM007544) for not only providing me funding, but also providing me a platform to learn and grow as a scientist. I recognize my years in the GTP as some of the most enriching I had at Michigan, and I appreciate all of the opportunities that I was provided through the GTP to expand my knowledge both on and off the bench.

Finally, I would like to thank my friends and family for being there through all the ups and downs during this whole process. To mom and dad, thank you for helping me live out my dreams each and every day, your love and support mean the world to me, but your example of good work ethic is what got me where I am today. To Adam, Nirav, Shikha, Lindsey, Binu, Kelly, Jason, Stephanie, Alex, the whole Women's team, and everyone I have played with past and present, thank you for the years of great volleyball and even better friendship. To the members of the Simmons lab, thank you for the laughs and the ability to sympathize with how much science sucks sometimes. Specifically I would like to thank Lindsay Matthews, for all those late night talks when nothing was working, for the multiple projects of mine that you saved because of your comp cells, and for the positivity that you always seemed to have even in the worst of times. Additionally, I would like to thank the whole crew over at Zingermans, let's not undersell how essential you were in this whole process. Thank you for being the cheer my mornings so desperately needed. Even the smallest acts of kindness that you showed me every day made the biggest difference. To Jackie Tobey, more than anyone else my success is a result of your support and the tools you taught me. You

have consistently changed my life for the better, and I can't thank you enough. Finally, to my sister, thank you for being my best friend, my cheering squad, my biggest supporter.

And last but not least, I would like to thank the University of Michigan. I was lucky enough to attend the University of Michigan for both undergrad and graduate school, and both experiences were better than I could have ever imagined. To say that Michigan has changed my life is too small of a statement. I am who I am today because of Michigan, and the years I have spent in Ann Arbor have been the happiest of my life. I will be forever grateful for my time here, and my blood will always run maize and blue.

For those who leave Michigan, but for whom Michigan never leaves. Go blue.

Table of Contents

Dedication ..... **ii**

Acknowledgements ..... **iii**

List of Figures ..... **xiv**

List of Appendices ..... **xvii**

ABSTRACT ..... **xx**

CHAPTER I Introduction **1**

1.1 Introduction *1*

1.2 Linear chromosomes and the problems they pose *2*

1.3 Telomeric DNA *3*

1.3.1 Telomere length and senescence 4

1.4 Telomerase *5*

1.4.1 Telomerase expression in somatic and germ cells 8

1.4.2 Telomerase and cancer 9

1.5 Shelterin *10*

1.5.1 Shelterin protects telomeric DNA 11

1.5.1 TRF1 and TRF2 12

1.5.2 Rap1 15

1.5.3 TIN2 15

1.5.4 POT1 16

1.5.5 TPP1 18

1.6 Telomerase recruitment to telomeres 21

1.7 Hierarchical assembly of shelterin at telomeres 22

1.8 Telomeropathies 24

1.9 Figures 27

CHAPTER II The N-terminus of the OB Domain of Telomere Protein TPP1 is Critical for Telomerase Action. 30

2.1 Abstract 30

2.2 Introduction 31

2.3 Results 34

2.3.1 NOB region of TPP1 is critical for stimulating processivity of telomerase 34

2.3.2 Effects of the NOB region on telomerase processivity are independent of TPP1 binding to POT1 or TIN2 36

2.3.3 TPP1ΔNOB fails to recruit telomerase to telomeres 37

2.3.4 TPP1ΔNOB overexpression fails to stimulate telomere lengthening in human cells . 38

2.3.5 The human TPP1 NOB region improves stimulation of human telomerase by mouse TPP1 40

2.4 Discussion 41

2.5 Materials and Methods 44

2.5.1 Primer extension assay 45

2.5.2 Immunofluorescence (IF) and fluorescence in situ hybridization (FISH) microscopy 45

2.5.3 Telomere restriction fragment length analysis 46

2.5.4 Plasmid constructs and mutagenesis 47
2.5.5 Purification of protein expressed in insect cells 49
2.5.6 HeLa culture and stable cell line generation 50
2.5.7 Co-immunoprecipitation 51
2.5.8 Immunoblotting 51
2.5.9 Telomere localization of TPP1 constructs 52
2.6 Figures and Tables 54

CHAPTER III Two separation-of-function isoforms of human TPP1 and a novel intragenic RNA dictate telomerase regulation in somatic and germ cells 66

3.1 Abstract 66
3.2 Introduction 67
3.3 Results 70
3.3.1 TPP1-S, but not TPP1-L, overexpression causes hyperelongation of telomeres 70
3.3.2 Both TPP1-S and TPP1-L recruit telomerase to telomeres 71
3.3.3 Both TPP1-S and TPP1-L protect chromosome ends 71
3.3.4 TPP1-L blocks telomere extension by telomerase 72
3.3.5 TPP1-L and TPP1-S isoforms are both primed for transcription but only TPP1-S accumulates in most cells 74
3.3.6 An intragenic noncoding RNA in the 3'-UTR shuts down TPP1-L, but not TPP1-S 76
3.3.7 TPP1-L expression switches on and TERT expression continues throughout spermatogenesis 80
3.4 Discussion 82
3.5 Materials and Methods 86
3.5.1 Molecular Cloning and site-directed mutagenesis 86

3.5.2 Cell culture 87
3.5.3 Generation of stable cell lines expressing TPP1-L and TPP1-L M87A 88
3.5.4 Telomere restriction fragment length analysis 88
3.5.5 Immunofluorescence-fluorescence *in situ* hybridization 89
3.5.6 RNAscope 93
3.5.7 Co-immunoprecipitation assay for protein-protein interaction 93
3.5.8 Immunoblotting 94
3.5.9 Testes tissue processing and immunoblotting 95
3.5.10 Enrichment of endogenous TPP1 using POT1-DNA pulldown. 95
3.5.11 CRISPR-Cas9 mediated insertion of a C-terminal FLAG tag at the endogenous locus of human TPP1 96
3.5.12 CRISPR-Cas9 mediated TPP1-L KO generation 96
3.5.13 5' RACE to clone sunRNA 97
3.5.14 Cloning, expression, and silencing experiments with sunRNA 98
3.5.15 Northern blotting 99
3.5.16 siRNA knockdown experiments 100
3.5.17 Protein purification 100
3.5.18 DNA-binding experiments 100
3.5.19 Direct telomerase activity assays 100
3.5.20 Visualization of data on UCSC genome browser 101
3.5.21 GTEx data 102
3.5.22 Testes-cell specific RNA-Seq analysis 102
3.6 Figures 103
CHAPTER IV Defining the Complete TPP1-TIN2 Interface 127
4.1 Abstract 127

4.2 Introduction *128*

4.3 Results *130*

4.3.1 Homology-assisted site-directed mutagenesis of TPP1 reveals extended TPP1-TIN2 interaction interface 130

4.3.2 TPP1 enhances the TIN2-TRF2 interaction 132

4.3.3 Extending the definition of the TPP1-binding region of $TIN2_{TRFH}$ 132

4.4 Discussion *133*

4.5 Materials and methods *135*

4.5.1 Molecular cloning and site-directed mutagenesis 135

4.5.2 Co-immunoprecipitation 135

4.5.3 Immunoblotting 136

4.5.4 HeLa cell culture 137

4.6 Figures and Tables *138*

CHAPTER V Concluding Remarks and Future Directions 142

5.1 Overview *142*

5.2 The N-terminus of the OB domain of TPP1 and its role in telomerase action *143*

5.3 The separation-of-function of TPP1-L and TPP1-S and its effect on germline biology .. *145*

5.4 The end-protection surfaces of TPP1 ..... *151*

5.5 Figures and Tables *156*

Appendix A A Lentivirus-Free Inducible CRISPR-Cas9 System for Efficient Targeting of Human Genes 157

A.1 Abstract *157*

A.2 Introduction *158*

A.3 Results *162*

A.3.1 Generation of a Flp recombinase-targetable vector for stable expression of two guide RNAs and a dox-inducible FLAG-Cas9 protein in HeLa cells 162

A.3.2 Cas9 in pG1G2-FLAG-Cas9-F3 can be induced to efficiently edit the HeLa genome 163

A.3.3 Virus-free generation of a cell line stably encoding a guide RNA targeting *AGO2* and an inducible Cas9 gene 166

A.3.4 Induction of Cas9 results in cleavage of the genomic target, reduction of target protein levels, and loss of target protein function 167

A.4 Discussion *169*

A.5 Materials and Methods *172*

A.5.1 Reagents and kits for molecular biology 172

A.5.2 Sanger sequencing 173

A.5.3 Parental plasmids 173

A.5.4 Cloning of the g1/g2-inducible Cas9 F3 vector 174

A.5.5 Cell Culture 175

A.5.6 Immunoblotting 175

A.5.7 Transient transfections 176

A.5.8 Guide RNA design 176

A.5.9 Surveyor nuclease assay 176

A.5.10 Lentivirus-free production of clonal stable cell lines encoding dox-inducible Cas9 177

A.6 Figures and Tables *179*

Appendix B Novel Interaction Between Proteins Responsible for Chromosome End Protection 184

B.1 Introduction 184

B.2 Results 186

B.2.1 Describing shelterin protein-protein interactions *in vitro* 186

B.2.2 TRF1 directly binds shelterin protein TRF2 187

B.2.3 Binding of either Rap1 or TIN2 inhibits the TRF1-TRF2 interaction. 188

B.2.4 Determining the region on TRF2 which is responsible for interacting with TRF1 189

B.2.5 TRF1-TRF2 weakly interact *in vivo* 191

B.3 Discussion 191

B.4 Materials and Methods 194

B.4.1 Molecular cloning 194

B.4.2 Purification of protein expressed in insect cells 195

B.4.3 Purification of protein expressed in bacteria 196

B.4.4 Flow-cytometry based assay for quantifying protein-protein interactions 196

B.4.5 Direct binding assays 197

B.4.6 Competition experiments 197

B.4.7 Co-immunoprecipitation assay for protein-protein interaction 198

B.4.8 GST pulldown assay 198

B.5 Figures and Tables 199

Bibliography 206

List of Figures

Figure 1.1 Domain diagrams and interaction map of shelterin. 27

Figure 1.2 Shelterin and telomerase solve the chromosome end protection and end replication problems. 28

Figure 1.3 Structure of the TPP1 OB domain. 29

Figure 2.1 NOB region of TPP1 is critical for facilitating repeat addition processivity of telomerase. 55

Figure 2.2 Structure of TPP1 OB, and domain diagram and binding properties of human TERT. 56

Figure 2.3 Effects of the NOB region on telomerase processivity are independent of TPP1 binding to POT1 and TIN2. 57

Figure 2.4 Deletion of the NOB of TPP1 disrupts both telomerase recruitment to telomeres and telomere lengthening. 59

Figure 2.5 Cell lines stably expressing TPP1 variants for telomere lengthening and telomerase recruitment studies. 61

Figure 2.6 Inability of mouse TPP1 to stimulate human telomerase is partially rescued by including the human NOB sequence. 63

Figure 2.7 Mouse OB-DBD stimulates mouse telomerase but not human telomerase. 64

Figure 2.8 Relative locations of the TEL patch, NOB, and L104 regions on the surface of the TPP1 OB domain. 65

Figure 3.1 TPP1-S but not TPP1-L overexpression causes robust telomere elongation. 103

Figure 3.2 Telomere restriction fragment length analysis of TPP1-L and TPP1-S clonal cell lines. 104
Figure 3.3 TPP1-S and TPP1-L recruit telomerase and protect chromosome ends. 106
Figure 3.4 Both TPP1-L and TPP1-S are proficient in chromosome end protection and telomerase recruitment. 108
Figure 3.5 TPP1-L is deficient in activating telomere synthesis. 110
Figure 3.6 Direct telomerase activity assays with TPP1-L and TPP1-S proteins. 111
Figure 3.7 TPP1-S but not TPP1-L accumulates in human cell lines. 113
Figure 3.8 Transcriptome and protein analysis suggest separate transcription cassettes for TPP1-L and TPP1-S, although TPP1-S protein is the predominant form in cells. 115
Figure 3.9 An intragenic RNA in the 3'-UTR of TPP1 completely shuts down TPP1-L. 117
Figure 3.10 Silencing of the sunRNA is specific for TPP1-L. 119
Figure 3.11 sunRNA silences TPP1-L but not TPP1-S. 121
Figure 3.12 sunRNA mediated regulation of TPP1 is specific to human TPP1 sequence but not dependent on any unique RNA surveillance or silencing machinery. 123
Figure 3.13 TPP1-L is upregulated and TERT persists throughout spermatogenesis. 125
Figure 3.14 RNAscope data in human testes. 126
Figure 4.1 Mutagenesis screen to identify the TPP1-POT1 and TPP1-TIN2 interaction surfaces. 138
Figure 4.2 Mutations in the $TPP1_{PBD}$ and the TPP1 C-terminus disrupt POT1 or TIN2 binding. 139
Figure 4.3 Mutations in the $TPP1_{TBM}$ and $TPP1_{TBM\text{-}ext}$ disrupt cooperativity of the TPP1-TIN2-TRF2 interaction. 140
Figure 4.4 Mutations in TIN2 alpha helix 7 result in loss of TPP1 and TRF2 binding. 141
Figure 5.1 Structure of *Tetrahymena thermophila* telomerase. 156

Figure A.1 Cloning strategy for generation of a vector allowing for site-specific virus-free integration into the genome, and stable co-expression of doxycycline-inducible Cas9 and two guide RNAs. 179
Figure A.2 pG1G2-FLAG-Cas9-F3 can be adapted to efficiently cleave the endogenous human *AGO2* and *DICER* loci. 180
Figure A.3 Virus-free generation of a cell line stably expressing dox-inducible Cas9 and guide RNAs against human *AGO2*. 181
Figure A.4 Induction of Cas9 in the presence of an *AGO2*-targeting guide RNA results in robust cleavage of the human *AGO2* locus, rapid depletion of AGO2 protein, and loss of RNA interference capacity. 182
Figure A.5 Timeline for generating stable cell lines expressing dox-inducible Cas9 and guide RNAs. 183

Figure B.1 Describing shelterin protein-protein interactions *in vitro* using a flow cytometry protein interaction assay. 200
Figure B.2 TRF1 directly binds shelterin protein TRF2. 202
Figure B.3 Rap1 and TIN2-mediated inhibition of the TRF1-TRF2 interaction. 203
Figure B.4 Determining the TRF1 binding region of TRF2. 204
Figure B.5 TRF2 binds TRF1 *in vivo*. 205

List of Appendices

Appendix A A Lentivirus-Free Inducible CRISPR-Cas9 System for Efficient Targeting of Human Genes 157

A.1 Abstract 157

A.2 Introduction 158

A.3 Results 162

A.3.1 Generation of a Flp recombinase-targetable vector for stable expression of two guide RNAs and a dox-inducible FLAG-Cas9 protein in HeLa cells 162

A.3.2 Cas9 in pG1G2-FLAG-Cas9-F3 can be induced to efficiently edit the HeLa genome 163

A.3.3 Virus-free generation of a cell line stably encoding a guide RNA targeting *AGO2* and an inducible Cas9 gene 166

A.3.4 Induction of Cas9 results in cleavage of the genomic target, reduction of target protein levels, and loss of target protein function 167

A.4 Discussion 169

A.5 Materials and Methods 172

A.5.1 Reagents and kits for molecular biology 172

A.5.2 Sanger sequencing 173

A.5.3 Parental plasmids 173

A.5.4 Cloning of the g1/g2-inducible Cas9 F3 vector 174

A.5.5 Cell Culture 175
A.5.6 Immunoblotting 175
A.5.7 Transient transfections 176
A.5.8 Guide RNA design 176
A.5.9 Surveyor nuclease assay 176
A.5.10 Lentivirus-free production of clonal stable cell lines encoding dox-inducible Cas9 177
A.6 Figures and Tables 179
Appendix B Novel Interaction Between Proteins Responsible for Chromosome End Protection 184
B.1 Introduction 184
B.2 Results 186
B.2.1 Describing shelterin protein-protein interactions *in vitro* 186
B.2.2 TRF1 directly binds shelterin protein TRF2 187
B.2.3 Binding of either Rap1 or TIN2 inhibits the TRF1-TRF2 interaction. 188
B.2.4 Determining the region on TRF2 which is responsible for interacting with TRF1 189
B.2.5 TRF1-TRF2 weakly interact *in vivo* 191
B.3 Discussion 191
B.4 Materials and Methods 194
B.4.1 Molecular cloning 194
B.4.2 Purification of protein expressed in insect cells 195
B.4.3 Purification of protein expressed in bacteria 196
B.4.4 Flow-cytometry based assay for quantifying protein-protein interactions 196
B.4.5 Direct binding assays 197
B.4.6 Competition experiments 197

B.4.7 Co-immunoprecipitation assay for protein-protein interaction 198

B.4.8 GST pulldown assay..... 198

B.5 Figures and Tables..... 199

ABSTRACT

Eukaryotic cells must overcome two major biological problems associated with linear chromosomes: the end protection and the end replication problems. The end protection problem occurs when the natural ends of linear chromosomes are misrecognized as DNA breaks within the genome, eliciting a DNA damage response. To counteract this, chromosome ends are capped with of a repetitive DNA sequence called telomeric DNA. Telomeric DNA is coated by the protein complex shelterin, thereby sequestering it from DNA damage response and repair machinery. The end replication problem occurs because replicative DNA polymerases are unable to catalyze the synthesis of DNA at the extreme ends of chromosomes, leading to the gradual loss of DNA during every replication cycle. The ribonucleoprotein complex telomerase helps compensate for this end attrition by catalyzing the addition of telomeric repeats to chromosome ends, enabling somatic and germ stem cells to replicate throughout the lifetime of a species. Together, shelterin and telomerase solve both problems associated with linear chromosomes.

TPP1 is the only shelterin protein known to play a dual role at chromosome ends. Not only does it help protect telomeres from erroneous DNA damage response and repair machinery, but it also recruits and facilitates processivity of telomerase. My doctoral studies have focused on how TPP1 interacts with telomerase and other

shelterin components to mediate its vital functions in somatic and germ cells. I demonstrate the importance of five residues at the N-terminus of the oligosaccharide/oligonucleotide binding (OB) domain (NOB) of TPP1 in telomerase recruitment to the telomere, telomerase stimulation by TPP1, and telomere lengthening in cells. Together with the previously discovered TEL patch, we have now defined an elaborate surface on TPP1 that we predict binds telomerase. Additionally, I describe the expression and separation-of-function of two isoforms of human TPP1. One, described here as TPP1-L, is annotated as a 544 amino acid protein initiating at methionine one. The other, referred to here as TPP1-S, is related to the rodent TPP1 and initiates at methionine 87. While both TPP1-S and TPP1-L recruit telomerase to the telomere, only TPP1-S activates telomerase for efficient telomere elongation. TPP1-S is the predominant isoform in somatic cells, but TPP1-L is the major isoform in differentiated male germ cells. I found that telomerase expression persists in germ cells, suggesting that TPP1-L may function to restrain telomerase in this context.

TPP1 is not only important in end replication, but it also plays an essential role in end protection by interacting with its shelterin partners POT1 and TIN2. To understand how each of these interactions contributes to end protection, I conducted a homology assisted site-directed mutagenesis screen to map the POT1 and TIN2 interaction surfaces on TPP1. I find that the TPP1-TIN2 surface is larger than previously appreciated and is important for TIN2 mediated cooperativity of TPP1 and TRF2. Together these studies serve to better define how TPP1 functions in human cells to protect the integrity of chromosome ends.

CHAPTER I

Introduction

1.1 Introduction

In 1911 Thomas Hunt Morgan hypothesized that genes were arranged in a particular order, with a beginning and an end, like 'beads on a string' (Morgan, 1911). Years later, Hermann Muller, one of Morgan's students, recognized that these natural ends behaved differently than the X-ray induced broken ends of linear chromosomes (H.J., 1938). Not to be confused with the broken ends, Muller coined these natural ends 'telomeres' using the Greek words for 'end' (telos) and 'part' (meros). Although Muller did not know it at the time, he had started a century long field of study aimed at understanding the 'end-part' of linear chromosomes.

Since the time of Thomas Hunt Morgan and Hermann Muller, hundreds of researchers have advanced the telomere field from an abstract theory to a detailed understanding of the complex biological system that is in place at every human chromosome end. They have found that the majority of eukaryotic chromosomes end in a repetitive DNA sequence (GGTTAG in humans) composed of both double stranded repeats and a single stranded G-rich 3' overhang. This telomeric DNA is coated with a

six-protein complex, called shelterin, that specifically binds to both the single stranded and double stranded telomeric DNA repeats. Together, these proteins and the telomeric DNA they bind to, make up the telomere.

1.2 Linear chromosomes and the problems they pose

By 1941, Barbara McClintock, who was best known for her work on transposable elements, had concluded that although the radiation induced broken ends of a chromosome were distinctly different from the natural ends, these induced ends could be altered to look and act much like their natural telomeric counterparts (McClintock, 1939). This observation was the first glimpse into the future of the telomere field and its focus on the two main problems posed by linear chromosomes: the end protection problem and the end replication problem. The end protection problem arises when the natural ends of linear chromosomes are misrecognized as induced double stranded DNA breaks within the chromosome. The shelterin protein complex solves this end protection problem by sequestering away the telomeric DNA and preventing the DNA damage repair and response machinery from recognizing the telomere as a DNA break in need of repair. Just as Barbara McClintock had observed, the natural chromosome ends are distinctly different from induced chromosome breaks because they are protected by shelterin. In contrast, the end replication problem occurs because DNA polymerase is unable to add nucleotides to the extreme ends of linear chromosomes. As a result, a small amount of DNA is lost at the ends of each chromosome during every replication cycle. Left unsolved, the end replication problem would result in a catastrophic loss of important DNA after many replication cycles. However, the ribonucleoprotein telomerase solves the end replication problem by adding telomeric

repeats to chromosome ends, thus preventing the loss of important DNA in continually dividing cells. Together, telomeric DNA, the shelterin protein complex, and telomerase work in concert to solve both problems posed by linear chromosomes.

1.3 Telomeric DNA

The first telomeric DNA was characterized in 1978 from the ciliated protozoa *Tetrahymena thermophila* (Blackburn and Gall, 1978), however it would be another ten years before the characterization of human telomeric DNA in 1988 (Moyzis et al., 1988). While the length and sequence of telomeres varies widely across organisms, the ends of human chromosomes are made up of a 10-15 kb stretch of hexad GGTTAG repeats (de Lange et al., 1990). Telomeric DNA is double stranded throughout, barring a short single stranded 3' overhang that exists at the ends of all chromosomes. This mammalian telomeric 3' overhang is rich in guanosine and devoid of cytosine, prompting the naming of the two strands of a telomeric duplex as the G strand or the C strand. The human single stranded G overhang varies between 50 and 500 nucleotides in length and is not only essential for proper end protection by the shelterin protein complex (discussed below), but is also an integral component of the t-loops that occur at chromosome ends (Palm and de Lange, 2008). In a t-loop, the single stranded telomeric overhang loops back to invade the double stranded telomeric repeats, hiding any exposed free ends that may trigger a DNA damage response (Griffith et al., 1999). t-loops were initially characterized in protein-free DNA after introduction of interstrand cross-links (Griffith et al., 1999), however they were later observed in native chromatin as well (Nikitina and Woodcock, 2004).

1.3.1 Telomere length and senescence

Since 1961 with the emergence of what is often referred to as the Hayflick limit, cell aging has been a topic of continued controversy. Leonard Hayflick suggested that cells grown in culture were not immortal, a widely accepted theory of the time, but instead undergo a phenomenon known as cellular senescence (HAYFLICK, 1965; HAYFLICK and MOORHEAD, 1961). Hayflick hypothesized that there was a counting mechanism for replicative senescence, however it wasn't until 1989 that Lundblad and Szostak gave Hayflicks's theory a molecular basis. They definitively demonstrated that an "ever shortening telomere" (EST) phenotype in budding yeast led to senescence, suggesting that telomeres acted as the counting mechanism Hayflick had referred to (Lundblad and Szostak, 1989). Just a year later, three seminal papers demonstrated that telomere shortening occurs during cellular aging, both in cultured human fibroblasts and in normal human tissues (de Lange et al., 1990; Harley et al., 1990; Lindsey et al., 1991). While the link between telomere shortening and cellular senescence is well studied, the implications of telomere shortening on longevity are an active area of research today. It is well established that telomeres progressively shorten as humans age, and that short telomeres are associated with an increased risk for age related diseases (Opresko and Shay, 2017). Telomere attrition is also classified as a hallmark of aging, yet the degree to which it is a principal cause of aging is still controversial (Lopez-Otin et al., 2013). Although numerous studies have worked to understand the mechanisms behind telomere shortening and longevity, the exact underpinnings of this process are still an active area of research today.

1.4 Telomerase

After shortening telomeres were proposed to lead to cellular senescence, it was suggested that cancer cells and germline cells, which require continued replication, must be able to solve the end replication problem and bypass the senescence that occurs in normal somatic cells. In 1985 the answer came from Carol Greider and Elizabeth Blackburn, who discovered enzymatic activity in *Tetrahymena thermophila* extracts that specifically elongated DNA of telomeric sequence (Greider and Blackburn, 1985). Although initially termed 'terminal transferase activity', they later described it as telomerase, a unique ribonucleoprotein-based reverse transcriptase that harbors an RNA template for telomeric DNA synthesis.

Human telomerase consists of a reverse transcriptase protein subunit TERT (Telomerase Reverse Transcriptase) and an RNA subunit TR (Telomerase RNA) (Blackburn and Collins, 2011; Greider and Blackburn, 1987, 1989; Lingner et al., 1997). The active telomerase enzyme uses a template sequence in its RNA subunit to specifically add telomeric repeats (GGTTAG in mammals) to the 3' ends of chromosomes, combatting the telomere loss that occurs during every replication cycle. In addition to providing a template for DNA synthesis, TR contains a template boundary element which prevents the incorporation of non-telomeric nucleotides, a pseudoknot and triple helix which contribute to the catalytic activity of telomerase, and a CR4-CR5 domain that is partially responsible for binding TERT (Chen and Greider, 2003b; Chen et al., 2002; Greider and Blackburn, 1989; Mitchell and Collins, 2000; Qiao and Cech, 2008; Shefer et al., 2007; Theimer et al., 2005; Theimer and Feigon, 2006). In addition, TR has both structural and sequence elements that are responsible for binding

accessory proteins (Nandakumar and Cech, 2013). Specifically, the accessory proteins dyskerin, NOP10, NHP2, and GAR1 bind the 3' end of TR known as the H/ACA motif, where they confer both TR stability and telomerase RNP biogenesis (Nandakumar and Cech, 2013; Schmidt and Cech, 2015).

TERT on the other hand, can be divided into four critical domains; the telomerase essential N-terminal domain (TEN) which is unique to telomerase, the TERT RNA-binding domain (TRBD), the reverse transcriptase (RT) domain, and the C-terminal extension (CTE) (Podlevsky and Chen, 2012). The reverse transcriptase domain of TERT contains the active site for telomere synthesis, sharing significant homology with the reverse transcriptase domains of both retrotransposon and retroviral reverse transcriptases (Lingner et al., 1997; Nakamura et al., 1997). Together, the TRBD, RT, and CTE domains are responsible for forming the TERT ring, which creates the catalytic core of the TERT protein where the canonical hand-like structure is formed between the RT domain (the fingers and the palm) and the CTE domain (the thumb) (Gillis et al., 2008; Jiang et al., 2018; Mitchell et al., 2010). Within this catalytic core is a positively charged cavity where TR interacts with telomeric DNA to form an RNA-DNA hairpin (Gillis et al., 2008; Jiang et al., 2018). Together these elements form an active site within telomerase which facilitates the addition of new telomeric repeats at chromosome ends. The final domain of TERT, the TEN domain, serves a function specific to telomerase and separate from the TERT ring. The TEN domain is essential for telomerase recruitment to the telomeres, and specific contacts between the TEN domain and the shelterin protein TPP1 show a direct link between telomerase and the shelterin protein complex at chromosome ends (Schmidt et al., 2014).

Although TPP1 resides at the telomere during all phases of the cell cycle, hTERT is only recruited to chromosome ends in S phase (Jady 2006, Tomlinson 2006), when hTERT transcription and activity peaks (Schmidt and Cech, 2015). Outside of S phase, TERT and TR are targeted to Cajal bodies, subnuclear sites where snRNAs and scaRNAs are synthesized and processed (Jády et al., 2004; Kiss et al., 2010; Zhu et al., 2004). TR is trafficked to Cajal bodies through its interaction with the protein TCAB1, which binds to the Cajal box (CAB) motif found at the 3' end of TR (Tycowski et al., 2009; Venteicher et al., 2009). Once in Cajal bodies, TR is assembled with TERT to form the biologically active telomerase RNP. Finally, telomerase is recruited from the Cajal bodies to the telomeres (Nandakumar and Cech, 2013), in a process that is described in depth in section 1.6.

Once recruited to telomeres, the active telomerase RNP works by binding to the single stranded 3' end of a chromosome and facilitating the addition of new telomeric repeats. For this, the alignment region in TR must anneal to the very end of the single stranded telomeric DNA, thus allowing TERT to add a new telomeric repeat in a TR template-dependent manner. Once the first repeat is completely synthesized, the template-primer base pairing is loosened, allowing telomerase to translocate and re-bind its template/alignment region to the newly formed 3' end of the DNA primer such that additional repeats can be added. The process of telomerase synthesizing multiple hexad repeats upon a single primer-binding event is called RAP, or repeat addition processivity (Greider, 1991). While telomerase alone can add multiple repeats to a telomeric primer *in vitro*, its processivity is stimulated by the presence of the shelterin

subcomplex POT1-TPP1, which resides on the 3' tail of telomeric DNA (Wang et al., 2007). These proteins are discussed in detail in section 1.5.

1.4.1 Telomerase expression in somatic and germ cells

Telomerase solves the end replication problem by adding telomeric repeats to the ends of chromosomes during every replication cycle. However the chromosome shortening that telomerase combats is what sets a limit on the number of times most somatic cells can divide, providing a natural anti-tumorigenic mechanism in long lived animals such as humans (Gomes et al., 2011). Because of this, telomerase is specifically shut down in most human somatic cells (see telomerase and cancer section for more detail on this process) (Bodnar et al., 1998). However, germline and somatic stem cells must preserve their ability to self-renew over the course of a species' lifetime. Consequently, these cells express telomerase, which counters the telomere shortening, and thus senescence, that would normally occur upon continued cell division (Shay and Wright, 2010). Therefore it is not surprising that germline mutations which affect telomerase activity result in diseases referred to as telomeropathies, discussed in detail in section 1.8 (Dokal, 2011; Savage, 2014). For these reasons, telomerase expression is tightly regulated in all human cells. Interestingly, telomerase is constitutively expressed in all cells of mice and other short-lived rodents (Gomes et al., 2011; Prowse and Greider, 1995). This is likely because they have fewer cell divisions over the course of their lifetime, nullifying their risk for oncogenesis and making strict regulation of telomerase unnecessary. These concepts will be revisited in detail in chapter three.

1.4.2 Telomerase and cancer

Since Leonard Hayflick first suggested that normal human cells have a limited capacity to divide in 1961, researchers have worked to understand the mechanisms and regulators that determine a cell's replicative lifespan. Once replicative capacity has been reached, cells undergo replicative senescence, a process that limits a cell's proliferative potential (Campisi, 1997). However in 1989 a groundbreaking paper from the Shay lab suggested that cells which bypassed senescence eventually entered a second stage of growth arrest called crisis (Wright et al., 1989). Cells in crisis contain fused chromosomes, almost universally leading to apoptosis. These two mechanisms were hypothesized to be potent anti-cancer protection, in place to prevent unwanted cell division (Shay and Wright, 2011). Despite this, a rare cell that acquires the ability to maintain or lengthen telomeres can escape crisis and continue dividing (Hanahan and Weinberg, 2000; Shay and Wright, 2011). This observation was confirmed when researchers found that introduction of TERT was enough to activate telomerase activity in telomerase negative cells and was sufficient to bypass both cellular senescence and crisis to induce cell immortalization (Bodnar et al., 1998). This limitless replicative potential is a hallmark of cancer, suggesting that TERT may have immense oncogenic potential. Indeed, telomerase is upregulated in more than 90% of all human cancers (Kim et al., 1994).

Mutations in the TERT promoter have been found at high frequency in both early and late stage human tumors, suggesting that they may be important for tumorigenesis (Horn et al., 2013; Huang et al., 2013). The first mechanistic insights into telomerase upregulation in cancer came in 2015 when it was discovered that the multimeric GA-

binding protein (GABP) transcription factor is recruited to the TERT mutant promoter, resulting in TERT activation (Bell et al., 2015). Soon after, it was discovered that these mutations extended the life of cells by allowing telomerase to lengthen the shortest telomeres and escape crisis, however further up-regulation of telomerase was needed to fully immortalize them (Chiba et al., 2017).

Because of the importance and prevalence of telomerase upregulation in cancer, strategies that block telomerase activity are prime candidates for anti-cancer drug development. This is especially appealing because normal human somatic cells do not express telomerase, therefore providing a natural protection mechanism against anti-telomerase therapeutics. Additionally, stem cells retain telomerase expression yet maintain longer telomere lengths than cancer cells, providing them with a significant survival advantage during the treatment window (Jafri et al., 2016). Unfortunately there is no telomerase inhibitor approved for the treatment of cancer available today, however multiple promising treatments are currently awaiting clinical trials (Bryan et al., 2015; Jafri et al., 2016).

1.5 Shelterin

As more studies elucidated the importance of telomeres and telomerase in end replication, discovering the proteins responsible for protecting telomeric DNA became the new focus of the telomere field. With telomere binding proteins already discovered in yeast and ciliates (Conrad et al., 1990; Gottschling and Zakian, 1986; Klobutcher et al., 1981; Lustig et al., 1990), the search for human telomere binding proteins began. In 1997, the first human telomere binding protein, telomeric repeat binding factor 1 or TRF1, was discovered by the de Lange lab (Chong et al., 1995; van Steensel and de

Lange, 1997). By 2004, all six members of the human shelterin protein complex had been discovered (Palm and de Lange, 2008). These proteins are discussed in detail below.

1.5.1 Shelterin protects telomeric DNA

Shelterin is a six-membered protein complex that protects telomeric DNA from being erroneously recognized by the DNA damage response and repair machinery (Figure 1.2). Shelterin is composed of three DNA binding proteins: TRF1 and TRF2 (telomeric repeat binding factor 1 and 2) which specifically bind to double stranded telomeric DNA, and POT1 (protection of telomeres 1) which binds to single stranded G-rich telomeric DNA. TPP1 (adrenocortical dysplasia protein homolog) binds to POT1 at the single stranded telomeric DNA and is responsible for recruiting telomerase to the telomere. TIN2 (TRF1-interacting nuclear factor 2) bridges TRF1 and TRF2 on the double stranded telomeric DNA to TPP1 at the single stranded telomeric DNA, while Rap1 (telomeric repeat-binding factor 2-interacting protein 1) binds specifically to TRF2.

Deletion of any shelterin protein, with the exception of Rap1, is embryonic lethal in mice (Palm and de Lange, 2008). These phenotypes are a result of end-deprotection, or the loss of shelterin proteins from the telomere and the activation of a DNA damage response at the natural ends of our chromosomes. In mammalian cells, two protein kinases are responsible for alerting DNA damage repair machinery to lesions in the genome. The first is the ataxia talengiectasia mutated kinase, or ATM kinase, which is thought to respond primarily to double stranded breaks within the genome (Palm and de Lange, 2008). The second is the ataxia talengiectasia and Rad3 related kinase, or ATR kinase, which is thought to primarily respond to single stranded

breaks within the genome (Palm and de Lange, 2008). In accordance with this, knockdown of the single stranded telomeric binding protein POT1 leads to an ATR mediated DNA damage response at chromosome ends, while knockdown of the double stranded telomeric binding protein TRF2 leads to an ATM mediated DNA damage response at chromosome ends (Celli and de Lange, 2005; Denchi and de Lange, 2007; Karlseder et al., 1999).

Shelterin forms a stable six-membered complex in the absence of telomeric DNA (Liu et al., 2004a; Ye et al., 2004b), however sub-complexes of shelterin have also been observed *in vitro* (Hu et al., 2017; Lim et al., 2017; Wang et al., 2007). *In vivo* the shelterin proteins have vastly different stoichiometries at telomeres (Takai et al., 2010). TRF1, TRF2, TIN2, and Rap1 are significantly more abundant than TPP1 or POT1, potentially reflecting the length disparity between the single stranded tail and the double stranded telomeric repeats. These differences necessitate the presence of shelterin sub-complexes *in vivo*, although current imaging techniques have been unable to distinguish these sub-complexes in cells. While the exact composition of shelterin at human telomeres is unknown, many studies have suggested that these shelterin sub-complexes are an important and functional component of the telomere (Lim et al., 2017).

1.5.1 TRF1 and TRF2

TRF1 and TRF2 are the shelterin double stranded DNA binding proteins. They share a common domain structure consisting of the TRF1 homology (TRFH) domain and a C-terminal myb DNA-binding domain that specifically interacts with the double stranded telomeric repeats (Figure 1.1) (Palm and de Lange, 2008). Their myb

domains are largely identical and confer TRF1 and TRF2's specificity for double stranded telomeric DNA through interactions with both the DNA backbone and the individual bases (Bianchi et al., 1999; Court et al., 2005). Because of a long flexible hinge region that separates the TRFH and myb domains, the two myb domains of a TRF1 homodimer can engage two separate sites on the telomeric DNA (Palm and de Lange, 2008). This flexibility allows TRF1 to interact with two different regions of telomeric DNA at a large distance, on two different molecules, and even in different relative orientations (Bianchi et al., 1999; Griffith et al., 1998). TRF2 similarly binds double stranded telomeric DNA, however it also plays an important function in the structure of the telomere by facilitating the formation of t-loop-like structures when given a telomeric substrate (Stansel et al., 2001). This is dependent on TRF2 binding the telomeric DNA sequence, as non-telomeric sequences at the single stranded/double stranded junction greatly reduces t-loop formation (Griffith et al., 1999).

Both TRF1 and TRF2 act as homodimers, tethered together through interactions in their TRFH domains (Bianchi et al., 1997; Broccoli et al., 1997). Despite their overall structural similarity, these TRFH domains cannot heterodimerize, but instead only recognize their own TRFH dimerization surface (Broccoli et al., 1997; Fairall et al., 2001). These domains function as versatile docking sites for telomere accessory proteins (Chen et al., 2008). The TRFH domain of TRF2 preferentially recognizes the Y-X-L-X-P motif of its accessory proteins, while the TRF1 TRFH preferentially interacts with the F-X-L-X-P motif, where X is any amino acid (Chen et al., 2008). While TIN2 is the primary binding partner of TRF1 utilizing the F-X-L-X-P binding motif (258 FNLAP 262), it is not the only protein that exploits this site. This motif is also used by proteins

that require temporary access to telomeres, such a TERB1 (telomere repeat-binding bouquet formation protein 1) which binds TRF1 to help tether the telomeres to the inner nuclear membrane during meiosis (Shibuya et al., 2014). While TIN2 binds both the TRF1 and TRF2 TRFH domains, it interacts with TRF2 with significantly lower affinity than TRF1, as its F-X-L-X-P motif is not optimal for a TRF2 interaction (Chen et al., 2008). Instead, TIN2 exploits a different region on TRF2 for optimal binding (Hu et al., 2017), explained in detail below. TRF2 on the other hand, utilizes the Y-X-L-X-P motif for optimal binding by the exonuclease Apollo and the double stranded break sensing complex MRN (MRE11-RAD50-NBS1) (Fairall et al., 2001; Rai et al., 2017; Wu et al., 2012).

Given the importance of TRF1 and TRF2 for protecting the double stranded telomeric DNA, it is not surprising that knock out of either TRF1 or TRF2 results in mouse embryonic lethality (Celli and de Lange, 2005; Karlseder et al., 2003). Although both TRF1 and TRF2 bind double stranded telomeric repeats, they elicit distinct telomeric dysfunction phenotypes when knocked out. TRF1 is essential for telomere replication, aiding in the polymerization of the G-rich repetitive telomeric DNA sequence. When knocked out, TRF1 results in the characteristic “fragile telomere” phenotype where long stretches of telomeric DNA are incompletely replicated (Sfeir et al., 2009). In contrast, TRF2 protects double stranded telomeric DNA from activating the ATM kinase DNA damage response as already mentioned above (Karlseder et al., 1999; Karlseder et al., 2004; Takai et al., 2003).

1.5.2 Rap1

Rap1 is a constitutive TRF2 binding partner, forming a 1:1 complex with high affinity (Palm and de Lange, 2008; Zhu et al., 2000). It has three distinct domains: a myb domain, an N-terminal BRCT motif, and a C-terminal domain (Figure 1.1). While the Rap1 myb domain is similar to that of TRF1 and TRF2, it is unable to bind telomeric DNA and is dependent on TRF2 for telomeric localization (Li and de Lange, 2003; Li et al., 2000). Instead, the Rap1 myb domain has been hypothesized to confer a protein-protein interaction with an unknown partner (Hanaoka et al., 2001). The C-terminal domain of Rap1 is responsible for interacting with the Rap1 binding motif on TRF2, and a solved crystal structure details this interaction (Chen et al., 2011). Unlike all other shelterin proteins, Rap1 is not embryonic lethal in mice, does not elicit any DNA damage response when knocked down in human cells, and has no known effect on telomere length maintenance (Kabir et al., 2014; Sfeir et al., 2010). However a recent study found that the TRF2-Rap1 complex plays an important role in preventing unwanted homology directed repair (HDR) at telomeres (Rai et al., 2016).

1.5.3 TIN2

TIN2 bridges the single stranded and double stranded shelterin proteins by binding to TRF1, TRF2, and TPP1 (Figure 1.1) (Houghtaling et al., 2004; Kim et al., 2004; Kim et al., 1999). The interaction between TRF1 and TIN2 is mediated through the canonical FxLxP motif in the C-terminus of TIN2 and the TRFH domain of TRF1 (Chen et al., 2008). While TIN2 is also able to interact with the TRFH domain of TRF2 using this same site, it does so with significantly lower affinity than that for TRF1 (Chen et al., 2008). Instead, the optimal binding site between TIN2 and TRF2 resides in the N-

terminus of TIN2, which interacts with the TIN2-binding motif (TBM) on TRF2. In fact, TPP1 and TRF2 both interact with TIN2 through its N-terminus, and a recently published paper gives the first structural insights into the TIN2-TPP1-TRF2 complex (Hu et al., 2017). Through this study, Hu et al. found that the N-terminus of TIN2 structurally resembles the TRFH domains of TRF1 and TRF2, even though sequence alignments show less than nine percent identity between the domains. Both the TIN2 TRFH domain and the TRFH domains of TRF1 and TRF2 are used as platforms for protein-protein interactions, utilizing similar surfaces to interact with their respective binding partners (Hu et al., 2017). These interactions will be discussed in detail in chapter 4.

The ability of TIN2 to bind three separate shelterin proteins, and its relative abundance compared to POT1/TPP1 (Takai et al., 2010), suggests a possible role in forming shelterin sub-complexes at the telomere. Indeed, multiple studies have shown that TPP1 binding promotes the interaction between TIN2 and TRF2 (O'Connor et al., 2006). Additionally TIN2 can simultaneously interact with both TRF1 and TRF2, suggesting that TIN2 may bridge these two double stranded DNA binding proteins (Ye et al., 2004a). Consistent with TIN2's central role, depletion of TIN2 has a significantly destabilizing effect on shelterin and overall telomere end protection, eliciting both an ATM- and ATR- mediated DNA damage response (Kim et al., 2004; Ye et al., 2004a).

1.5.4 POT1

POT1 is the only shelterin protein known to bind to the single stranded G-rich tail of human telomeres (Palm and de Lange, 2008). It was originally identified through sequence similarity to the alpha subunit of the telomere end-binding protein (TEBP) of the ciliate *Oxytricha nova* (Baumann and Cech, 2001). Like TEBP-α, human POT1

contains two oligosaccharide/oligonucleotide binding (OB) domains that form its N-terminal DNA binding domain (DBD) which can recognize and bind to the single stranded G-strand telomeric sequence (Figure 1.1) (Baumann and Cech, 2001; Kelleher et al., 2005; Lei et al., 2004; Loayza et al., 2004). A solved crystal structure of the POT1 DBD in complex with single stranded telomeric DNA details how the first OB domain (OB1) interacts with the first six bases of the telomeric sequence using both hydrogen bonds and pi stacking interactions (Lei et al., 2003). This high specificity and affinity for telomeric DNA provides POT1 with an obvious advantage for protecting chromosome ends from the ATR kinase mediated DNA damage response (Denchi and de Lange, 2007). By binding the single stranded telomeric DNA, POT1 prevents replication protein A (RPA) from gaining access to chromosome ends and triggering subsequent homologous recombination (HR) at telomeres (Takai et al., 2017).

POT1 not only interacts with single stranded telomeric DNA with high specificity and affinity, but it also binds to its shelterin partner protein TPP1 (Palm and de Lange, 2008). Two solved crystal structures of the C-terminus of POT1 in complex with the POT1 binding domain (PBD) of TPP1 show an extensive POT1-TPP1 binding interface (Chen et al., 2017; Rice et al., 2017), supporting the observation that these two proteins interact with high affinity. Indeed, binding of TPP1 to POT1 actually increases POT1's affinity for single stranded telomeric DNA, although the structural basis for this increased affinity remains unknown (Wang et al., 2007). Because of this, it is not surprising that mutations in the TPP1-binding domain of POT1 are associated with several different leukemias and melanomas (Chen et al., 2017; Lei et al., 2004; Rice et

al., 2017), highlighting the POT1-TPP1 heterodimer's essential role in preventing chromosome instability.

1.5.5 TPP1

In 2004 researchers from three separate labs conducted immunoprecipitation-mass spectrometry (IP-MS) experiments using TIN2 protein as bait to discover the shelterin protein TPP1 (Houghtaling et al., 2004; Liu et al., 2004b; Ye et al., 2004b). TPP1 plays a central role in shelterin by binding both POT1 and TIN2 (Palm and de Lange, 2008). While all other shelterin proteins were named for their telomeric function, "TPP1" is actually the result of combining the three different names given by the groups that initially characterized the protein. TINT1 (Houghtaling et al., 2004), PTOP (Liu et al., 2004b), and PIP1 (Ye et al., 2004b) were all described concurrently, leading to the renaming of the described protein as TPP1. To add to this confusion, the accepted gene name for TPP1 is *ACD*, after the adrenocortical dysplasia phenotype that mice exhibited upon deletion of this gene (Keegan et al., 2005). Like other shelterin components, TPP1 plays an essential role in end protection. Knockout of TPP1 in mouse cells results in a DNA damage response, chromosome instability, and an increase in chromosome fusions (Else et al., 2007; Guo et al., 2007). It is thought that TPP1 participates in its end-protection functions through its interaction with POT1, rather than through a direct interaction with telomeric DNA, as TPP1 increases POT1's affinity for telomeric DNA by ten-fold (Lei et al., 2004; Loayza et al., 2004; Wang et al., 2007; Xin et al., 2007).

Like other shelterin members, TPP1 is a multi-domain protein where different domains perform distinct functions (Figure 1.1). The N-terminus of TPP1 (aa 1-86) is

the only region within the 544 amino acid TPP1 protein that is not conserved among mouse, rat, and most other species, and has no known or predicted structural motifs. Based on this, and the fact that it is seemingly dispensable for all of TPP1's known functions (Hwang et al., 2012; Nandakumar et al., 2012; Schmidt et al., 2014), it has been argued that the physiological TPP1 protein starts at methionine 87. I will explore this idea further in chapter three.

The TPP1 OB domain begins at methionine 87 and ends at aa 250. While TPP1 knockdown had previously been implicated in telomere over-elongation and a DNA damage response at telomeres (Ye et al., 2004b), deletion of the OB domain of TPP1 (TPP1ΔOB) protected chromosome ends as well wild-type (WT) TPP1 and led to telomere shortening (Xin et al., 2007). From this observation, the TPP1 OB domain was suggested to play a role in telomere length maintenance by interacting with telomerase, as well as through its interaction with POT1. This was supported by pulldown experiments showing co-immunoprecipitating of TERT with TPP1, but not with TPP1ΔOB (Xin et al., 2007). Further studies confirmed this association when TPP1 was shown to recruit telomerase to telomeres in both human and mouse cells (Abreu et al., 2010; Tejera et al., 2010). Together these data suggested that TPP1, and specifically the TPP1 OB domain, was an essential component for solving the end replication problem. This was the first evidence of a shelterin protein operating in both end protection and end replication, functioning as the link between the telomere and telomerase. These data are explored in greater detail below.

C-terminal to the TPP1 OB domain is the POT1-binding domain (PBD; aa 250-334), which is responsible for the interaction between POT1 and TPP1. Two crystal

structures detail an extensive interaction surface between the TPP1 PBD and the POT1 C-terminus (Chen et al., 2017; Rice et al., 2017). This is consistent with TPP1's reduced stability when purified individually *in vitro*, and the suggestion that POT1-TPP1 function explicitly as a heterodimer *in vivo*. Independent of telomerase recruitment to the telomere *in vivo*, the POT1-TPP1 heterodimer also increases telomerase processivity *in vitro* (Wang et al., 2007). While addition of POT1 alone inhibited telomerase extension of a telomeric primer in primer extension assays, addition of the POT1-TPP1 complex increased telomerase activity and processivity 2-3 fold compared to telomerase alone controls (Wang et al., 2007). However while this processivity is dependent on a POT1-TPP1 interaction, mutations in the TPP1 OB domain that do not affect POT1 binding abrogate this effect, suggesting that the OB domain of TPP1, as well as the interaction with POT1, is essential for telomerase stimulation (Nandakumar et al., 2012).

At the far C-terminus of the TPP1 protein lies the TIN2 binding domain (aa 480-544) (Houghtaling et al., 2004; O'Connor et al., 2006). While previous studies have confirmed the importance of this region for interacting with TIN2 and preserving end protection (Frescas and de Lange, 2014a), it wasn't until recently that a solved crystal structure gave the first structural insights into this interaction. Using yeast-two hybrid analysis, a small peptide of TPP1 containing amino acids 510-544 was found to be necessary and sufficient for TIN2 binding (Hu et al., 2017). This region was termed the TIN2 binding motif of TPP1, or TBM. The TPP1 TBM is not only essential for TIN2 binding, but it is also important for stabilizing larger shelterin subcomplexes. While TIN2 has been shown to interact with TPP1 as a stable heterodimer, the TPP1-TIN2 complex

is important for stabilizing the TIN2-TRF2 interaction (O'Connor et al., 2006). This suggests TIN2 confers cooperativity between TPP1 and TRF2 to form a stable three membered shelterin sub-complex. However, despite the solved crystal structure of the TIN2 TRFH domain in complex with peptides of both TPP1 and TRF2, the mechanism for this cooperativity remains unclear. In chapter 4 I discuss this TPP1-TIN2 interface in detail. I find that the discovered interface is not complete, and I describe the full TPP1-TIN2 interface and its effect on TIN2 mediated cooperativity of TPP1 and TRF2.

1.6 Telomerase recruitment to telomeres

Telomerase recruitment to the telomere relies on the direct link between TPP1 and TERT. Because of the extraordinarily low abundance of telomerase in cells (250 telomerase molecules per human cell), it is not surprising that telomerase must be recruited to chromosome ends rather than having to find them through simple diffusion (Schmidt and Cech, 2015). In 2010 the first *in vivo* assays showed that TPP1 was essential for this telomerase recruitment. Through immunofluorescence/fluorescence *in situ* hybridization and chromatin immunoprecipitation researchers discovered that TPP1, but not POT1, was necessary for telomerase recruitment to the telomere. Upon siRNA knockdown of TPP1, telomerase accumulated in Cajal bodies. This recruitment defect could be rescued through overexpression of an siRNA resistant TPP1, giving the first definitive evidence of TPP1's essential role in telomerase recruitment in human cells (Abreu et al., 2010). Additional studies confirmed that the TPP1 OB domain was sufficient for this interaction, as an OB domain tethered in an artificially created region in the genome was sufficient to recruit telomerase to this non-telomeric sequence (Zhong et al., 2012).

These findings suggested that TPP1 was an essential component of both end protection, by interacting with the shelterin protein POT1, and end replication, by binding telomerase. To understand these distinct roles, a search for separation-of-function mutants of TPP1 that would affect telomerase recruitment but not telomere end protection was conducted by researchers in the Cech lab. Using a site-directed mutagenesis screen of conserved and surface exposed residues in the OB domain of TPP1, Nandakumar et al. discovered a patch of amino acids, known as the TEL patch, that is essential for telomerase recruitment to the telomere *in vivo* and telomerase processivity *in vitro* (Figure 1.3) (Nandakumar et al., 2012). TEL patch residue E215 was later found to specifically contact the TEN domain of TERT, giving the first evidence of a direct interaction between TPP1 and telomerase (Schmidt et al., 2014).

Although the TPP1 TEL patch is necessary for telomerase regulation, it cannot describe the entire TPP1-telomerase interface. All seven TEL patch residues identified in the human TPP1 screen are conserved in mouse TPP1, yet mouse TPP1 is unable to stimulate human telomerase (Zaug et al., 2010). This suggests additional species-specific contacts within TPP1 may mediate the TPP1-telomerase interaction, a hypothesis that I tested and describe in chapter two.

1.7 Hierarchical assembly of shelterin at telomeres

Although each shelterin member has a unique and essential function at telomeres, these proteins do not act autonomously. When TPP1 is knocked down in human cells POT1 signal is also lost, triggering an ATR mediated DNA damage response at chromosome ends (Hockemeyer et al., 2007). This suggests TPP1 may be required for POT1 localization to the telomere. In confirmation of this theory, expression

of a POT1 mutant that cannot bind TPP1 results in loss of POT1 signal at the telomere and an ATR-mediated DNA damage response. In contrast, expression of a mutant POT1 that does not contain the DNA binding domain still localizes to telomeres (Chen et al., 2007; Denchi and de Lange, 2007; Hockemeyer et al., 2007; Liu et al., 2004b). While it is clear that TPP1 is important for POT1 localization to the telomere, TPP1 is also reliant on a shelterin protein for its localization. The TPP1-POT1 heterodimer appears to rely solely on TIN2 for both its nuclear import and single stranded telomere localization (Abreu et al., 2010). When TIN2 is knocked down in mouse cells, there is a loss of POT1 and TPP1 signal at the telomere and an accumulation of RPA, leading to an ATR dependent DNA damage response (Takai et al., 2011). Because of the loss of TPP1 at telomeres, knockdown of TIN2 also results in a lack of telomerase recruitment to the telomere (Abreu et al., 2010). To complicate things further, TIN2 localization also depends on an interaction with TRF1 (Frescas and de Lange, 2014b). While TRF1 localization is unaffected by knockout of TIN2, knockout of TRF1 results in loss of TIN2 signal at the telomere. However when TRF2 is knocked out, TIN2 signal at telomeres is apparently unaffected. This suggests that the TIN2-TRF1 interaction is essential for proper TIN2 localization to telomeric DNA, and thus proper POT1-TPP1 localization as well (Frescas and de Lange, 2014b). This hierarchy of localization creates a complicated system for understanding the physiological functions of individual shelterin proteins, as knockout of TIN2 results in a similar phenotype as knockout of TIN2, TPP1, and POT1. Because of this, knockout of TIN2, TPP1, and POT1 all elicit an ATR mediated DNA damage response at telomeres, while knockout of TRF2 results in an ATM mediated DNA damage response (Palm and de Lange, 2008; Takai et al., 2011).

Interestingly knockout of TIN2 signals both ATM and ATR mediated DNA damage, suggesting TIN2 may play a role in stabilizing and/or loading TRF2 on to telomeres as well (Takai et al., 2011).

1.8 Telomeropathies

While expression of telomerase in somatic cells is constitutively shut down to prevent unwanted cell division, telomerase expression in stem cells is essential for continued cell renewal over the course of a species' lifetime (Shay and Wright, 2010). Genetic mutations in telomerase or telomerase-associated factors can affect the ability of stem cells to self-renew, leading to a wide variety of telomere biology disorders known as telomeropathies (Savage, 2018). The most severe of these disorders, known as dyskeratosis congenita or DC, presents as a symptomatic triad of abnormal skin pigmentation, nail dystrophy, and oral leukoplakia (Savage, 2018). However patients with DC are also at a much higher risk for bone marrow failure, cancer, and pulmonary fibrosis. Since the first reported case of dyskeratosis congenita in 1906, eleven different genes have been found to be mutated in patients with DC, including the genes that encode for TERT, TR, TIN2, and TPP1 (Guo et al., 2014; Kocak et al., 2014). Mutations in these genes cause the blood leukocyte telomere lengths of patients to be in the less than first percentile for their age. These critically short telomere lengths often result in reduced hematopoietic stem cell pools and eventually lead to bone marrow failure, the leading cause of death for DC patients (Dokal, 2011; Jones et al., 2016; Savage and Bertuch, 2010).

Patients suffering from DC have mutations in telomerase or telomerase associated factors and are characterized as having critically short telomeres. This led

researchers to ask how specific disease mutations were preventing telomerase from extending telomeric repeats at chromosome ends. Given the importance of the TPP1 TEL patch in both the recruitment and processivity of telomerase, mutations in the TEL patch would be expected to result in telomere shortening and telomeropathies. Indeed, two studies have identified germline mutations in *ACD* in two unrelated patients suffering from severe telomeropathies (Guo et al., 2014; Kocak et al., 2014). This *ACD* mutation causes an in-frame deletion of lysine 170 of TPP1. Researchers in the Nandakumar lab discovered that deletion of lysine 170 distorts the structure of the TPP1 TEL patch loop, inhibiting TPP1's ability to recruit telomerase to telomeres in cells (Bisht et al., 2016). This was the first study to report a mechanistic understanding of how a DC mutation directly leads to telomere shortening in cells (Bisht et al., 2016).

Unfortunately, the underlying molecular basis for many patients suffering from DC remains poorly understood. The first shelterin gene that was implicated in DC was *TINF2*, which encodes the TIN2 protein (Savage et al., 2008; Walne et al., 2008). Patients with *TINF2* mutations have extremely short telomeres and exhibit a severe form of DC at an early age (Sasa et al., 2012; Savage et al., 2008; Walne et al., 2008). More than twenty different mutations have been identified in the *TINF2* gene of DC patients, all clustering in a "DC" hotspot (Figure 1.1) (Glousker et al., 2015). These mutations act in an autosomal dominant manner, however unlike most forms of DC where genetic anticipation plays a large role in the pathogenesis, *TINF2* mutations arise *de novo* (Walne et al., 2008). While two separate studies have suggested that these *TINF2* mutants reduce telomerase recruitment to the telomere (Frank et al., 2015; Yang et al., 2011), no mechanistic link between TIN2 and telomerase has been found.

Additionally, these studies fail to explain the extraordinarily short telomeres or the early onset of the disease in these individuals compared to patients suffering from other DC mutations. More research is necessary to understand the molecular basis of DC in these patients and many others.

Since the first characterized case, significant progress has been made to understand the genetic causes that underlie DC and other related telomeropathies. Yet even today, 30% of all patients with DC still have no known genetic deficiencies (Savage, 2018). For this reason, continued collaborative research is crucial for understanding and treating DC in affected individuals. However this research cannot be conducted in a vacuum; clinicians and basic scientists must work together to achieve a better understanding of the underlying causes of DC and related telomeropathies. Together, basic scientists and clinicians can pioneer new treatments for patients who suffer from these debilitating diseases, allowing for advances that are seldom ever achieved independently.

1.9 Figures

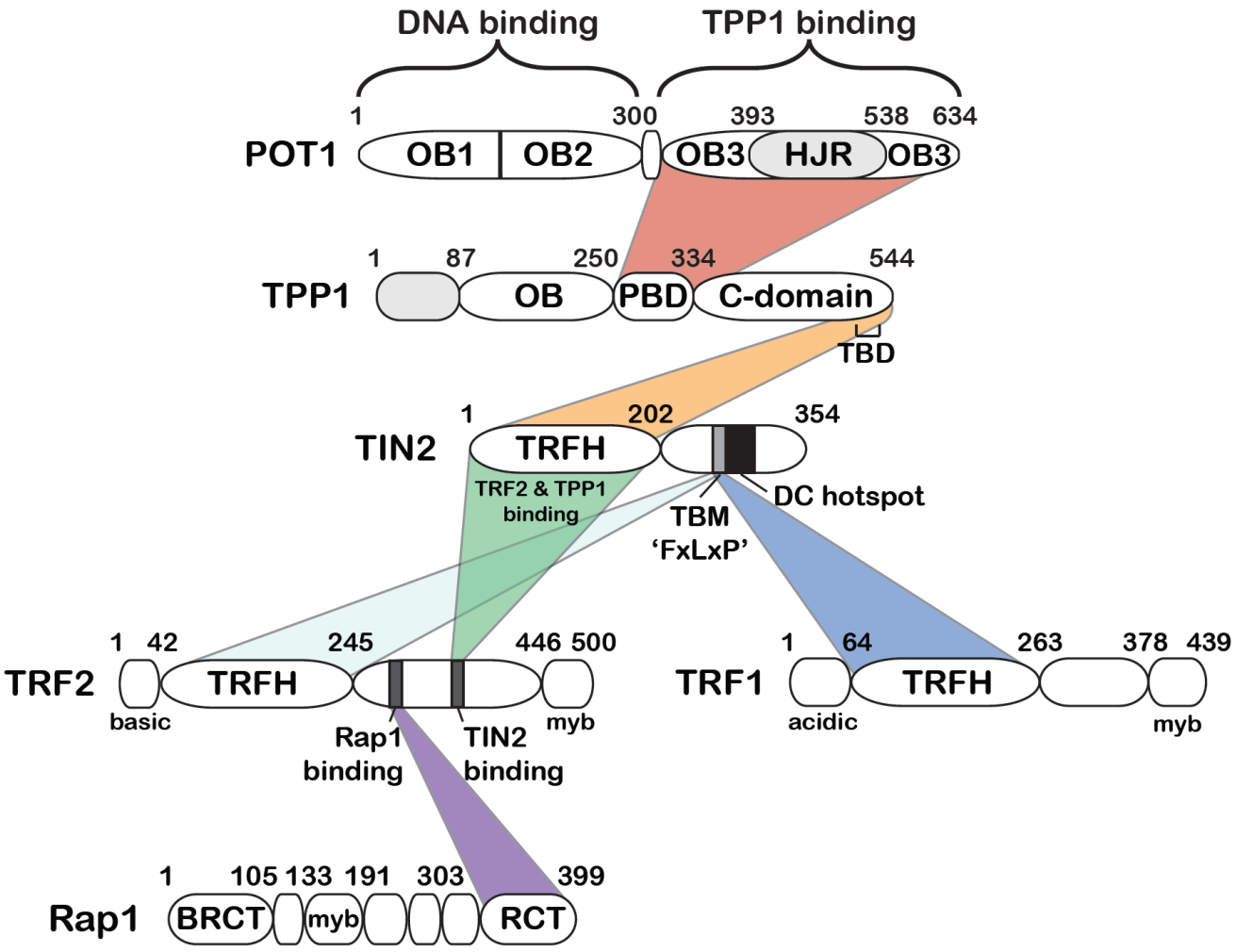


Figure 1.1 Domain diagrams and interaction map of shelterin.
Domain diagrams of the six shelterin proteins. OB: oligonucleotide/oligosaccharide-binding domain. HJR: Holliday junction resolvase-like domain. PDB: POT1 binding domain. CTD: C-terminal domain. TBD: TIN2 binding region of TPP1. TRFH: TRF homology domain TBM: TRF1-binding F-X-L-X-P motif. RCT: Rap1 C-terminal domain.

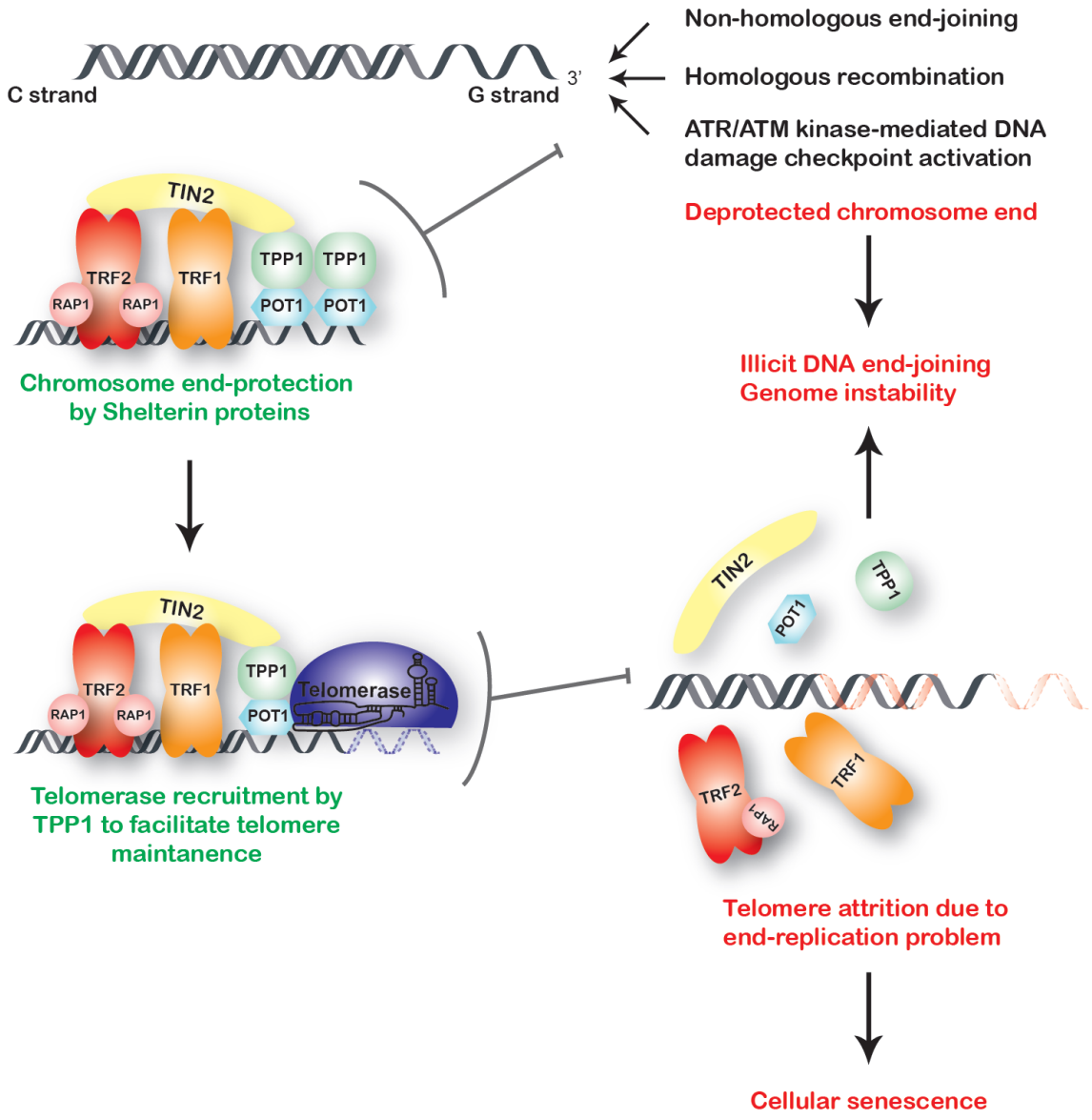


Figure 1.2 Shelterin and telomerase solve the chromosome end protection and end replication problems.
Top: Schematic of naked telomeric DNA eliciting a DNA damage response which results in genome instability. Shelterin protects telomeric DNA by blocking the ATM/ATR mediated DNA damage responses at chromosome ends. *Bottom*: Schematic of telomerase bound to the single stranded telomeric DNA tail and the shelterin protein complex through its interaction with TPP1. Loss of telomerase results in telomere attrition and cellular senescence. (Adapted from (Nandakumar and Cech, 2013)

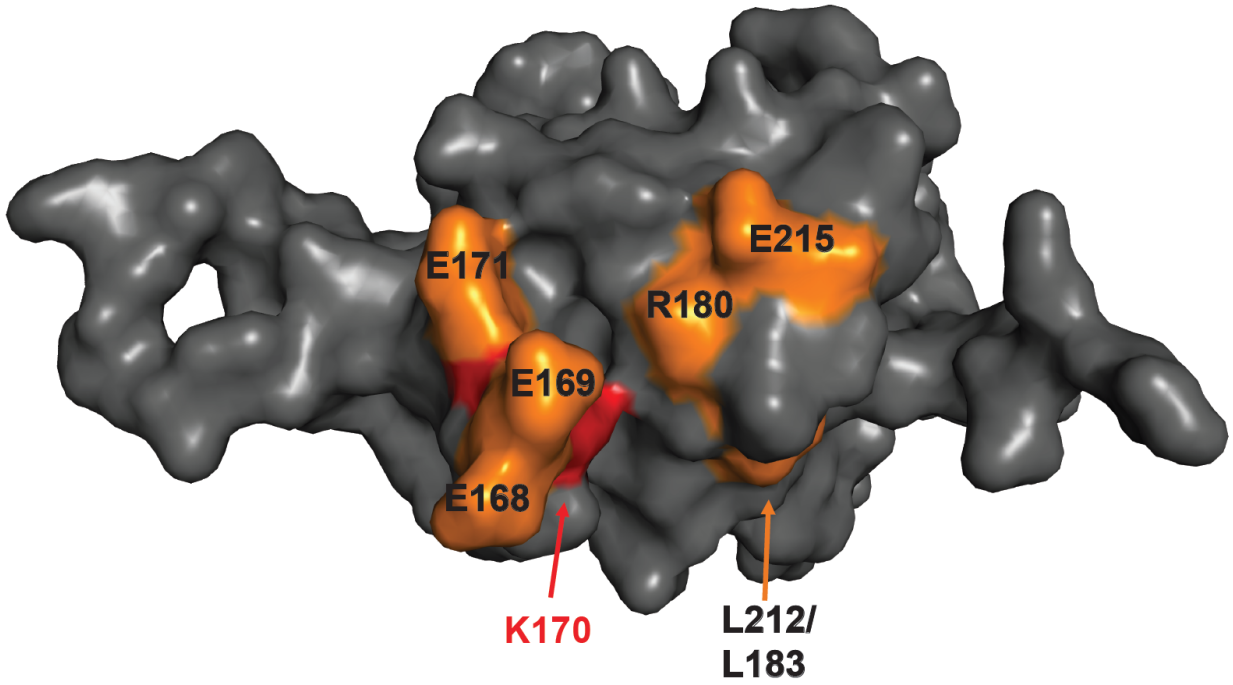


Figure 1.3 Structure of the TPP1 OB domain.
Structure of the TPP1 OB domain (gray; PDB ID: 2146) with TEL patch residues shown in orange and residue K170 in red.

CHAPTER II

The N-terminus of the OB Domain of Telomere Protein TPP1 is Critical for Telomerase Action1.

2.1 Abstract

Telomerase recruitment to telomeres and enzymatic processivity are mediated by TPP1, an essential component of telomere integrity and telomerase function. A surface on the OB domain of TPP1 called the TEL patch is critical for TPP1's telomerase-associated functions. Here we identify a separate region in the N-terminus of the OB domain (termed NOB) of TPP1 that, like the TEL patch, is essential for telomerase repeat addition processivity in vitro as well as telomerase recruitment to telomeres and telomere lengthening in cells. Although well-conserved among most

[1] A modified version of this chapter is published in Cell Reports (DOI: 10.1016/j.celrep.2018.01.012). I performed all cell biological experiments, including stable cell line generation, co-immunoprecipitation, IF-FISH, and telomere length analysis. Valerie Tesmer performed protein purifications, prepared telomerase extracts, and conducted direct telomerase activity assays. Both Valerie Tesmer and I contributed to experimental design and molecular cloning. I wrote the manuscript with detailed input from Valerie Tesmer and Jayakrishnan Nandakumar.

mammalian TPP1 homologs, the NOB region in mouse is distinct. Swapping the sequence of human NOB into mouse TPP1 allows it to stimulate human telomerase, qualifying NOB as an important determinant of species specificity for the TPP1-telomerase interaction. Our studies show that TPP1 NOB is critical for telomerase function and demonstrate that the telomerase interaction surface on TPP1 is more elaborate than previously appreciated.

2.2 Introduction

In most eukaryotes, continued cell division is made possible through activation of the ribonucleoprotein (RNP) enzyme telomerase (Hahn et al., 1999). Telomerase compensates for the shortening of chromosomes that occurs during DNA replication by synthesizing the telomeric repeat sequence (GGTTAG in humans) at chromosome ends (Greider and Blackburn, 1985). This process is essential for stem cell renewal (Shay and Wright, 2010) and mutations that negatively impact telomerase function result in inheritable telomeropathies such as dyskeratosis congenita (Dokal, 2011; Savage, 2014). Telomerase is also responsible for conferring replicative immortality in ~90% of all known cancers, qualifying it as a prime target for anti-cancer drug development (Kim et al., 1994).

Telomerase biogenesis and catalysis involve several steps. Formation and stabilization of the RNP complex, which includes the protein subunit TERT (telomerase reverse transcriptase) (Lingner et al., 1997) and the RNA subunit TR (telomerase RNA), occurs in the nucleus (Blackburn et al., 1989; Egan and Collins, 2012; Greider and Blackburn, 1989). The assembled telomerase RNP is then trafficked through sub-nuclear structures called Cajal bodies (Venteicher and Artandi, 2009) before it is finally

recruited to telomeres for chromosome end extension (Nandakumar and Cech, 2013). Telomeres are coated by shelterin, a six-protein complex [consisting of POT1, TPP1, TIN2, TRF1, TRF2, and Rap1] that protects chromosome ends from erroneously eliciting a DNA damage response (Kibe et al., 2016; Palm and de Lange, 2008). Shelterin component TPP1 (Houghtaling et al., 2004; Liu et al., 2004b; Ye et al., 2004b) is uniquely involved in facilitating telomerase action as well. TPP1 not only binds shelterin proteins POT1 (protection of telomeres 1) and TIN2 (TERF1-interacting nuclear factor 2) (Figure 2.1A), but it also recruits telomerase to telomeres (Nandakumar and Cech, 2013; Palm and de Lange, 2008). The primary binding partner of TPP1 within telomerase is the TEN (telomerase essential N-terminal) domain of TERT (Schmidt et al., 2014; Sexton et al., 2012; Zaug et al., 2010; Zhong et al., 2012) (Figure 2.2A), although other regions of TERT including the CTE (carboxy-terminal extension, or thumb) domain (Zhong et al., 2012) and the IFD (insertions in fingers domain) (Chu et al., 2016) have also been implicated in this function (Figure 2.2A). Once recruited, TPP1 associates with telomerase to stimulate telomerase repeat addition processivity (or RAP; hereby referred to as processivity) *in vitro* and promote telomere lengthening in telomerase-positive cells (Nandakumar et al., 2012; Wang et al., 2007).

The region of TPP1 responsible for telomerase-related functions has been mapped to an OB domain (Abreu et al., 2010; Wang et al., 2007; Xin et al., 2007) [corresponding to human TPP1 amino acids (aa) 87-250 or mouse TPP1 aa 1-162; Figure 2.1A]. In human TPP1, seven conserved and surface-exposed OB residues (E168, E169, E171, R180, L183, L212, and E215) were found to be critical for

telomerase association and are collectively referred to as the TEL [TPP1 glutamate (E) and leucine (L)-rich] patch (Nandakumar et al., 2012; Sexton et al., 2012; Zhong et al., 2012) (Figure 2.1A; Figure 2.2B,C). The importance of the TEL patch is further highlighted by the discovery of a TPP1 K170Δ mutation in two unrelated individuals presenting with telomeropathies (Guo et al., 2014; Kocak et al., 2014). TPP1 K170Δ structurally distorts the peptide backbone of the loop harboring TEL patch glutamate residues E168, E169, and E171 resulting in telomere shortening (Bisht et al., 2016). Additionally, TPP1 knock-in experiments in human embryonic stem cells (hESCs) confirm that deletion of this acidic loop leads to severely shortened telomeres and reduced long-term cell viability (Sexton et al., 2014). Other TPP1 residues implicated in mediating telomerase function include W98, which likely serves a structural role (Rao et al., 2014; Zhang et al., 2013); phosphorylation site S111, whose mutation altered telomere length regulation in cells (Zhang et al., 2013) through a mechanism not revealed in direct telomerase extension assays (Nandakumar et al., 2012); and L104, whose mutation reduces telomerase processivity but not association (Nandakumar et al., 2012) and which has been proposed to be involved in regulating telomere length homeostasis (Sexton et al., 2014).

The TPP1 TEL patch is necessary for telomerase regulation, however it does not describe the entire telomerase interaction surface on TPP1. All seven TEL patch residues identified in human TPP1 are conserved in mouse TPP1 (Figure 2.2C), yet mouse TPP1 is unable to stimulate processivity of human telomerase (Zaug et al., 2010). N-terminal to the first alpha helix in the TPP1 OB domain resides a largely hydrophobic sequence that is well conserved among many mammals but is less

conserved between human and mouse (Figure 2.1B, top). In the crystal lattice formed by the human TPP1 OB domain these residues are buried to form a homodimeric interface [Figure 2.2B; (Wang et al., 2007)]. Given that the TPP1 OB domain is monomeric in solution [data not shown; (Wang et al., 2007)], it is likely that the hydrophobic N-terminus of the OB domain (NOB) is buried in a different interface to shield it from the aqueous environment in the cell. We asked whether the NOB region of TPP1 contributes to telomerase binding and dictates human versus mouse TPP1 specificity for stimulating telomerase. Here we show that human TPP1 NOB mutants reduce the stimulation of telomerase processivity by TPP1, abrogate telomerase recruitment to telomeres, and shorten telomeres of human cells in culture. Simply replacing mouse TPP1 NOB with the analogous residues of the human homolog improves the ability of mouse TPP1 to stimulate human telomerase. Our study reveals that the NOB region, together with the TEL patch, provides a more complete model for the surface of TPP1 that is critical for telomerase function.

2.3 Results

2.3.1 NOB region of TPP1 is critical for stimulating processivity of telomerase

To determine if the N-terminus of the TPP1 OB domain is important for telomerase processivity we purified wild-type TPP1-N (TPP1 aa 90-334; TPP1-N WT) and mutant TPP1-N proteins harboring mutations or deletions within the NOB region (Figure 2.1A; Figure 2.1B, bottom). All TPP1 mutant and WT proteins were recombinantly expressed in *Escherichia coli* and purified as previously described (Bisht

et al., 2016) (Figure 2.1C; see EXPERIMENTAL PROCEDURES). We then performed direct telomerase primer extension assays using super-telomerase extract prepared in HEK 293T cells (Cristofari and Lingner, 2006) and POT1 purified from baculovirus-infected insect cells (Figure 2.1D,E; see EXPERIMENTAL PROCEDURES). TPP1-N WT displayed the characteristic increase in telomerase processivity compared to telomerase extract alone, as indicated by a larger fraction of long DNA products (compare lanes 1 and 2, Figure 2.1D). Deletion of the NOB region of TPP1 (TPP1-N ΔNOB) resulted in a striking decrease in telomerase processivity compared to TPP1-N WT (38% decrease; $p < 0.05$) (compare lanes 2 and 3 in Figure 2.1D; Figure 2.1E). To pinpoint which residues within TPP1 NOB contribute to telomerase stimulation, we engineered a series of alanine mutations between TPP1 aa 92-104. Mutating residues 92-95 to alanine (92-95A) resulted in a significant decrease in telomerase processivity relative to TPP1-N WT, and comparable to that of TPP1-N ΔNOB (compare lanes 2-4 in Figure 2.1D; Figure 2.1E). TPP1-N with alanine substitutions at aa 92-93 (92-93A) or 94-95 (94-95A) similarly resulted in a decrease in processivity (compare lanes 2 with 5 and 6 in Figure 2.1D; Figure 2.1E). In contrast, downstream alanine substitutions of aa 96 (96A) or aa 100-101 (100-101A) had no effect on telomerase processivity compared to WT (compare lane 2 with lanes 7 and 8 in Figure 2.1D; Figure 2.1E). Mutation of TPP1 aa L104 has previously been shown to moderately reduce telomerase processivity (Nandakumar et al., 2012). Because L104 and the NOB region lie in close proximity, we directly compared the contributions of these regions to telomerase function. In contrast to the NOB mutants that showed a considerable decrease in telomerase processivity relative to WT, the effects of the L104A mutation were more

modest (13% decrease; compare lane 9 to lanes 2-5 in Figure 2.1D; Figure 2.1E). In summary, our data defines the NOB region as a new functional element within TPP1 that is critical for stimulating telomerase processivity.

2.3.2 Effects of the NOB region on telomerase processivity are independent of TPP1 binding to POT1 or TIN2

The OB domain of TPP1 has not only been implicated in telomerase binding but also in stabilizing the POT1-DNA complex (Rajavel et al., 2016). Telomerase activity and processivity measured in our primer extension assay depends not just on the TPP1-telomerase interaction but also on tethering TPP1 to POT1-bound DNA. To confirm that the defects of NOB mutants result solely from telomerase-binding deficiencies, we performed primer extension assays with a series of chimeric proteins that completely lack the POT1 and TIN2 binding regions of TPP1. In these fusions, the OB domain of human TPP1 is covalently linked to the DNA binding domain of human POT1 (hOB-hDBD), enabling the recruitment of the TPP1 OB domain directly to the single stranded telomeric DNA primer (Figure 2.1A). hOB-hDBD chimeric proteins were recombinantly expressed in *Escherichia coli* and purified (Figure 2.3A; see EXPERIMENTAL PROCEDURES). WT hOB-hDBD stimulated telomerase processivity to levels comparable to that of the POT1 + TPP1-N heterodimeric complex (compare lanes 1-3 in Figure 2.3B; Figure 2.3C). To further confirm that our hOB-hDBD fusion system recapitulates the phenotypes of the covalently unlinked protein complex, we examined the E169A-E171A (EE-AA) TEL patch double mutation previously shown to be severely defective in telomerase processivity stimulation (Nandakumar et al., 2012). EE-AA in the hOB-hDBD backbone was defective in stimulating telomerase processivity

(58% decrease; Figure 2.3B,C). The most instructive finding was that deletion or mutation of NOB (constructs hOB-hDBD ΔNOB and hOB-hDBD 92-95A, respectively) caused defects in telomerase processivity (43% and 61% decrease, respectively) that are comparable to that of the EE-AA mutant (Figure 2.3B,C).

To further exclude the possibility that TPP1 NOB affects TIN2 or POT1 binding we performed co-immunoprecipitation experiments. Both transiently expressed FLAG-tagged TPP1 WT and TPP1ΔNOB bound Myc-POT1 or Myc-TIN2 on anti-FLAG conjugated beads (Figure 2.3D). These data confirm that the TPP1 NOB region stimulates telomerase processivity independent of POT1 or TIN2 binding.

2.3.3 TPP1ΔNOB fails to recruit telomerase to telomeres

Because TEL patch mutants of TPP1 fail to stimulate processivity of telomerase due to poor telomerase association (Nandakumar et al., 2012; Sexton et al., 2012; Zhong et al., 2012), we asked how deleting the NOB would affect TPP1-telomerase association in cells. To address this, we engineered stable cell lines that utilize a tetracycline-inducible promoter to control the expression of a single copy gene cassette encoding FLAG-tagged TPP1 ΔNOB (ΔNOB; aa 96-544) or TPP1 L104A (L104A; aa 87-544), using a published method (Bisht et al., 2016; Nandakumar et al., 2012) (Figure 2.5A; Figure 2.4A). These lines were directly compared with a previously characterized TPP1 WT (WT; aa 87-544) stable cell line (Bisht et al., 2016; Nandakumar et al., 2012). Individual clones were selected to equalize FLAG-TPP1 expression levels, although we note ΔNOB consistently expressed at a higher level than WT and L104A (Figure 2.4A; Figure 2.5B). All TPP1 constructs successfully localized to telomeres (Figure 2.5C), in agreement with their ability to bind TIN2 and POT1. To evaluate telomerase

recruitment to telomeres we performed immunofluorescence (IF) to visualize FLAG-TPP1 and compared its localization with that of TR, which was visualized using fluorescence *in situ* hybridization (FISH). More than 90% of TPP1 WT foci colocalized with telomerase foci, indicative of robust telomerase recruitment to telomeres with TPP1 WT (Figure 2.4B,C). In contrast, cells expressing TPP1ΔNOB exhibited drastically reduced levels of telomerase recruitment (Figure 2.4B), with less than 20% of ΔNOB foci colocalizing with telomerase (Figure 2.4C; $p < 0.01$ when compared to WT). These results are consistent with what has previously been observed for TPP1 TEL patch mutations (Nandakumar et al., 2012). Although L104 is adjacent to the NOB region, previous studies suggest that L104A exhibits a slight, if any, reduction in telomerase association (Nandakumar et al., 2012; Sexton et al., 2014). Indeed, L104A successfully recruited telomerase to telomeres, although a slight reduction in the percent of telomerase recruitment (relative to WT) in each L104A clone was observed ($p = 0.18$ and 0.11, respectively) (Figure 2.4B,C). The severe recruitment defects observed for the ΔNOB clones but not the L104A clones are even more striking when considering the fact that the steady-state protein levels of ΔNOB are consistently higher than that of L104A (Figure 2.4A; Figure 2.5B). In summary, despite its proximity to L104, the critical role of the NOB region in recruiting telomerase to telomeres resembles that of the TEL patch.

2.3.4 TPP1ΔNOB overexpression fails to stimulate telomere lengthening in human cells

Although TPP1 WT overexpression induces significant telomere elongation, overexpression of TPP1 TEL patch mutants results in telomere shortening due to

reduced telomerase recruitment (Nandakumar et al., 2012). Given that TPP1ΔNOB is unable to efficiently recruit telomerase to telomeres, we asked how deleting the TPP1 NOB region would affect telomere length maintenance. For this we conducted telomere restriction fragment length (TRF) analysis of our engineered stable cell lines overexpressing TPP1 WT, ΔNOB, and L104A proteins. A stable cell line containing the tetracycline inducible promoter but an empty cassette (vector) was engineered and propagated in parallel to serve as a negative control. As expected, the telomeres of the vector cell line remained relatively stable over 45 days in culture, similar to those of untransfected HeLa cells (Nandakumar et al., 2012) (Figure 2.4D,E). Overexpression of TPP1 WT led to a robust increase in telomere length over time (~70 bp per day) (Figure 2.4D,E), consistent with previous reports (Bisht et al., 2016; Nakashima et al., 2013; Nandakumar et al., 2012). In sharp contrast, overexpression of TPP1ΔNOB resulted in a decrease in telomere length (Figure 2.4D; Figure 2.5D) at a rate of 22 and 17 bp per day for clones 1 and 2, respectively (Figure 2.4E). This rate of telomere shortening is reminiscent of that observed for the TPP1 EE-AA TEL patch mutant expressing cells (~18 bp/day; (Bisht et al., 2016; Nakashima et al., 2013; Nandakumar et al., 2012). The eventual rescue of telomere length observed in prolonged cultures of TPP1ΔNOB stable cell lines (Figure 2.4D; Figure 2.5D) coincides with a loss of TPP1ΔNOB expression (Figure 2.5E), an effect previously documented for cell lines expressing TEL patch mutants (Nakashima et al., 2013). In contrast to this trend, L104A overexpression resulted in a constant increase in telomere length (Figure 2.4D; Figure 2.5D) at a rate of 20 bp per day for each clone. The rate of telomere elongation for L104A (Figure 2.4E) was less dramatic than that of WT, which is in accordance with our biochemical (Figure

2.1D,E) and cytological data (Figure 2.4B,C). From these data we conclude that the critical role of TPP1 NOB in telomere length maintenance closely resembles that of the previously identified TEL patch.

2.3.5 The human TPP1 NOB region improves stimulation of human telomerase by mouse TPP1

Given that some NOB residues are different between mouse and human TPP1 (Figure 2.1B; top), we asked whether the mouse NOB sequence (mNOB) accounts for the inability of mouse TPP1 to stimulate human telomerase (Zaug et al., 2010). To address this, we exploited our OB-hDBD fusion system so that species-specific differences in POT1-TPP1 binding would not influence our results. We purified a chimera in which the mouse TPP1 OB domain was engineered upstream of hDBD (mOB-hDBD) (Figure 2.6A; Figure 2.7A). The chimeric mOB-hDBD protein successfully stimulated telomerase reconstituted with mTERT (compare lanes 8-10 with 4-6; Figure 2.7B). In sharp contrast, mOB-hDBD was defective in stimulating human telomerase activity or processivity compared to hOB-hDBD (compare lanes 1-3 in Figure 2.6B; Figure 2.6C,D; for replicate compare lanes 3 and 7 in Figure 2.7B). We next asked whether replacing hNOB with the mNOB sequence in human TPP1 (Figure 2.1B; bottom) would compromise human telomerase processivity. Indeed hTPP1-N mNOB reduced human telomerase processivity relative to hTPP1-N WT (compare lanes 1 and 2 in Figure 2.7B). Finally, to test whether human NOB can revive mOB-hDBD function with human telomerase, we designed a variant in which three of the mouse NOB residues (L6R, A8V, Q10R) were replaced with their human equivalents in mOB-hDBD (hNOB/mOB-hDBD) (Figure 2.6A; Figure 2.7A; see EXPERIMENTAL PROCEDURES).

Compared to mOB-hDBD, the chimeric hNOB/mOB-hDBD protein was able to partially stimulate both processivity and activity of human telomerase (compare lanes 3 and 4 in Figure 2.6B; Figure 2.6C,D). We conclude that the NOB region is an important determinant of mouse versus human species-specificity of the TPP1-telomerase interaction. However, hNOB/mOB-hDBD did not completely rescue the activity and processivity of telomerase to the levels observed with WT hOB-hDBD (compare lanes 2 and 4 in Figure 2.6B, Figure 2.6C,D), indicating that the complete telomerase-binding surface of TPP1 exceeds the list of identified TEL patch and NOB residues.

2.4 Discussion

TPP1 is the telomere-associated factor that is primarily responsible for recruiting telomerase to chromosome ends. Mutagenesis screens by three independent groups converged on a strictly conserved surface of TPP1 OB called the TEL patch as being a critical determinant of telomerase action (Nandakumar et al., 2012; Sexton et al., 2012; Zhong et al., 2012). In this study we uncover NOB as another distinct element in TPP1 that is critical for telomerase function, but which likely escaped discovery because it is less conserved between mouse and human. We show that deletion or mutation of human TPP1 NOB residues results in a reduced ability to stimulate telomerase processivity, diminished telomerase recruitment, and shortening of telomeres in cells. Furthermore, the extreme phenotype of deleting the NOB region matches that of mutations within the TEL patch, suggesting that both of these TPP1 elements are critical for telomerase function.

Inspection of the crystal structure of TPP1 OB reveals that although distant from the acidic residues of the TEL patch (E168, E169, and E171), the hydrophobic NOB

region is proximal to TEL patch residues L183 and L212 (Figure 2.6E). Labeling both TEL patch and NOB residues on the surface of TPP1, two clefts for binding telomerase emerge, one laced with primarily acidic residues (E168, E169, and E171, R180, and E215), and another with largely hydrophobic residues (R92, L93, V94, L95, L183, and L212) (Figure 2.6E). Analysis of TERT has revealed the highly basic TEN domain as the binding partner of the TEL patch (Sexton et al., 2012; Zhong et al., 2012) (Figure 2.2A). However only one binary interaction between TPP1 and telomerase (TERT K78 and TPP1 E215) has been firmly established (Schmidt et al., 2014) (Figure 2.2A). The IFD of TERT has also been suggested as a potential binding partner of TPP1 based on observations that TERT IFD mutant V791Y is neither recruited efficiently to telomeres, nor stimulated by TPP1 in primer extension analysis (Chu et al., 2016) (Figure 2.2A). Finally, the CTE domain of TERT has also been proposed to contact TPP1 (Zhong et al., 2012) (Figure 2.2A). Based on the presence of two distinct telomerase-binding clefts in TPP1 OB, it is tempting to speculate that the highly basic TEN domain docks into the acidic cleft, while the hydrophobic pocket is occupied by residues residing elsewhere in TERT.

Given the importance of the TPP1 NOB region in stimulating telomerase processivity, it is intriguing that the N-terminal regions of the distant yet well-characterized TPP1 homologs Tpz1 of the fission yeast *S. pombe* (Miyoshi et al., 2008), TEBPβ of ciliate *S. nova*, and Est3 of budding yeast *S. cerevisiae* all show structural similarities to mammalian TPP1 (Horvath et al., 1998; Rao et al., 2014; Wang et al., 2007). In each, the N-terminus contains an integral alpha helix that precedes the canonical beta barrel core (prediction, in case of Tpz1). In both mammals and budding

yeast, the NOB region borders this helix, which is capped at its N-terminus by a “PWI” sequence (Figure 2.1B; top). In the structures of Est3 and human TPP1 the conserved tryptophan is stabilized by an essential aspartate located in a loop between the second and third beta strands of the OB domain (Lubin et al., 2013; Rao et al., 2014) that has also been identified as important for DNA binding with POT1 (Rajavel et al., 2016). Both the tryptophan and the aspartate residues are conserved in Tpz1, and an analogous interaction occurs in ciliate TEBPβ with a phenylalanine taking the place of the tryptophan (Rao et al., 2014). Given that the structural elements at the N-terminus of the OB domain of distant TPP1 homologs are so well conserved, it is possible that the NOB region similarly facilitates telomerase recruitment in at least some of these diverse eukaryotic organisms.

In human TPP1, the first alpha helix of the N-terminal OB domain terminates at L104 (Wang et al., 2007). Mutation of L104 does not influence telomerase binding to TPP1 (Nandakumar et al., 2012), but when introduced at the endogenous gene locus in hESCs it causes resetting of telomere length to a lower length without affecting long-term viability (Sexton et al., 2014). Yet only subtle telomerase recruitment and telomere length phenotypes of L104A are observed in the overexpression system used in the current study. These results open the possibility that L104 allosterically affects telomerase function through the NOB region that resides at the opposite end (i.e., N-terminus) of the helix harboring L104. Alternatively, phenotypes of L104 mutants may result from protein folding/stability defects in TPP1, as the low B-factor of L104 in the crystal structure suggests a structural role for this residue. This possibility is consistent

with the fact that in multiple clones we observed that steady state levels of L104A were lower than that of both WT and ΔNOB (Figure 2.5B).

Sequence conservation is a powerful tool to probe protein structure and function, including protein-protein interactions. It was through mutagenesis of strictly conserved surface residues that the TEL patch of TPP1 was mapped (Figure 2.2C). However the telomere-telomerase interaction is species-specific (Zaug et al., 2010), which suggests there are key determinants of the TPP1-telomerase binding surface that are not strictly conserved. In fact, the NOB region embodies one such species-specific element in TPP1 that is important for telomerase function. However, hNOB/mOB-hDBD, despite harboring all seven TEL patch residues and the human TPP1 NOB sequence, is not able to stimulate telomerase processivity to hOB-hDBD levels (Figure 2.6B-D). Further mutagenesis exploiting differences in mouse versus human TPP1 sequences will provide an efficient avenue to completely define the telomerase-binding surface of TPP1. The sheer size combined with the macromolecular complexity of the telomerase RNP has been a major impediment to comprehensively determining the TPP1-binding surface of telomerase. We propose that probing species-specific sequence differences and involving cross-species hybrid proteins of TERT will greatly streamline the search for the more elusive half of the TPP1-telomerase interface.

2.5 Materials and Methods

2.5.1 Primer extension assay

Extension assays included super telomerase extract, 1 μM primer a5 (5'-TTAGGGTTAGCGTTAGGG-3'), dNTPs, α^{32}P-dGTP, and 500 nM of POT1/TPP1-N variant or OB-hDBD chimeras, and were performed for one hour at ~30°C. Extension products were resolved on a 10% polyacrylamide/ 7 M urea/ 1x TBE sequencing gel and the data were imaged using a Phosphorimager (Storm; GE Healthcare Life Sciences). Activity was scored by quantitating the signal for the entire lane. Processivity values were obtained by dividing the total intensity from bands representing addition of 9 or more hexad repeats by the total intensity in the lane (omitting repeats 1 and 2) using the rolling ball method to define background.

2.5.2 Immunofluorescence (IF) and fluorescence in situ hybridization (FISH) microscopy

Stable cell lines expressing either TPP1 WT, ΔNOB, or L104A were induced with doxycycline for 3 days and then ~100,000 cells were seeded on coverslips in a 12 well plate with growth medium containing doxycycline. Twenty-four hours post-seeding the medium was removed and the cells were washed with PBS. All subsequent steps were performed at room temperature. Cells were fixed with 4% formaldehyde in PBS for 10 min and washed three times for 5 min in PBS before permeabilization in PBS containing 0.5% Triton-X 100 for 10 min. IF-FISH experiments for telomerase recruitment were performed as described previously (Nandakumar et al., 2012). Briefly, IF was first performed to visualize FLAG-TPP1 proteins using mouse monoclonal anti-FLAG M2 (Sigma; F1804; 1:500) in combination with Alexa Fluor 568-conjugated anti-mouse IgG (Life Technologies). Subsequently a mixture of Cy5-conjugated probes complementary

to TR was used at a concentration of 30 ng per probe per coverslip to detect TR by FISH (Abreu et al., 2010). The cells were washed three times in PBS and mounted on microscope slides using ProLong Gold mounting medium with DAPI (Life Technologies). Coverslips were sealed with transparent nail polish and stored at -20°C until the time of imaging. A laser scanning confocal microscope (SP5; Leica, Germany) equipped with a 100x oil objective was used to image IF-FISH experiments. The images were processed with ImageJ and Adobe Photoshop, and colocalizations were quantified manually by two separate individuals.

2.5.3 Telomere restriction fragment length analysis

Genomic DNA was purified by the GenElute kit (Sigma, G1N350-1KT) from stable cell lines overexpressing either FLAG-TPP1 constructs or a vector control. DNA (2 µg) was digested with HinfI and RsaI and incubated overnight at 37°C. The digested DNA was run on a 25 cm long 0.8% agarose-1X TBE gel along with a lambda DNA-HindIII digest ladder (NEB) at a constant 50 V for 21-23 h. The gel was imaged with a fluorescent ruler and then dried at 55°C for one hour prior to denaturation in 0.5 M NaOH for 30 min. The gel was then rinsed with water, neutralized with 1.5 M NaCl and 0.5 M Tris-Cl (pH 7.5) for 30 min, and prehybridized in Church buffer [0.5 M sodium phosphate buffer (pH 7.2), 1% bovine serum albumin, 1 mM EDTA, and 7% SDS] for 30 min at 65°C in a rotating hybridization oven. T4 polynucleotide kinase was used to 5' ^{32}P-label a $(TTAGGG)_4$ oligonucleotide, which was added at 20 million cpm to the gel. Hybridization was continued overnight at 55°C. The gel was then washed three times with 2X SSC for 10 min at 55°C and exposed to a phophorimager screen. The gel was analyzed using the Imagequant TL software and calibrated using the molecular weights

of the lambda DNA-HindIII digest ladder. The mean telomere length for each lane was plotted as a function of days in culture for each cell line.

2.5.4 Plasmid constructs and mutagenesis

The following parental constructs used in this study have been described elsewhere: pFBHTb-Smt3star-hPOT1 for purification of insect cell-expressed human POT1 (Kocak et al., 2014); pET-Smt3-TPP1-N WT and L104A for bacterial expression of TPP1-N (aa 90-334) (Kocak et al., 2014; Nandakumar et al., 2012; Wang et al., 2007); pET-Smt3 vector for bacterial expression of constructs newly described in this study (MTA with Cornell University) (Mossessova and Lima, 2000); pTERT-cDNA6/Myc-HisC and phTR-BluescriptIISK(+) constructs for overexpression of telomerase in HEK 293T cells (Nandakumar et al., 2012); pTet-IRES-eGFP-BI4 vector, p3x-FLAG-TPP1-BI4 plasmid for Tet-inducible expression of FLAG-tagged human TPP1 (aa 87-544) in HeLa-EM2-11ht cells, and pd1gfpPtetmiR vector for Flp recombinase-mediated stable clone generation (MTA with Tet System Holdings GmbH & Co KG) (Nandakumar et al., 2012).

For expression of N-terminal 3X-FLAG-tagged mTERT in human cell lines, we substituted the mTERT cDNA sequence (pCiteE-mTERT plasmid was a kind gift from Julian Chen, Arizona State University) for the human TERT cDNA sequence in pTERT-cDNA6/Myc-HisC using Gibson Assembly (New England Biolabs; NEB).

3X-FLAG-tagged TPP1 ΔNOB and L104A were cloned into the pTet-IRES-eGFP-BI4 and p3X-FLAG-TPP1-F3 vectors as described previously (Nandakumar et al., 2012). Myc-tagged POT1 and TIN2 were similarly amplified from plasmids containing the respective cDNA sequences (Kocak et al., 2014; Nandakumar et al.,

2012) and cloned into the pTet-IRES-eGFP-BI4 vector to furnish p6X-Myc-POT1-BI4 and p6X-Myc-TIN2-BI4, respectively.

For bacterial expression of TPP1-N variants ΔNOB, 92-95A, 92-93A, 94-95A, and mNOB, appropriate TPP1-N sequences were cloned into the in pET-Smt3 vector using standard restriction endonuclease-mediated cloning. TPP1-N variants 96A and 100-101A were generated by QuikChange site-directed mutagenesis (Agilent technologies) using pET-Smt3-TPP1-N WT as template.

The construct pET-Smt3-hOB-hDBD was created for expression of hOB-hDBD WT, which consists of the OB domain of TPP1 (aa 88-250) and the DNA binding domain (DBD) of human POT1 (aa 2-303), linked by a 29 aa glycine and serine-rich linker NH_2-VDGSGGSSGSGGSGSSGSSGGSKLAAALD-COOH. This was established by first splitting the multiple cloning site of the pET-Smt3 vector by inserting the cDNA sequence coding for the glycine-serine stretch (IDT) between the SalI and HindIII restriction sites. The human OB domain was cloned between the upstream BamHI and SalI sites, while the DBD was cloned between the downstream HindIII and XhoI sites. A similar strategy was used to create the mOB-hDBD and the hNOB/mOB-hDBD constructs (the N-termini of these proteins are specified in Figure 2.7A). hOB-hDBD mutants (ΔNOB, 92-95A, and EE-AA) were generated by QuikChange site-directed mutagenesis (Agilent technologies) using hOB-hDBD WT as template.

For insect cell expression of TPP1-N we engineered a pFBHTb-sumostar-TPP1-N plasmid using the strategy described for engineering $(His)_6$-SUMOstar human POT1 (Kocak et al., 2014). All oligonucleotides used in this study were purchased from IDT.

2.5.5 Purification of protein expressed in insect cells

Full-length human POT1 was expressed as a SUMOstar-$(His)_6$-POT1 fusion protein in baculovirus-infected High Five cells (Thermo Fisher Scientific), as described previously (Kocak et al., 2014). Briefly, POT1 was purified from the soluble cellular lysate with Ni-NTA agarose resin (Qiagen). After tag cleavage with SUMOstar protease (LifeSensors), untagged protein further purified by Superdex 200 size-exclusion chromatography (GE Healthcare Life Sciences) in 25 mM Tris-Cl (pH 8), 150 mM NaCl, and 2 mM DTT. The POT1-TPP1-N heterodimeric complex was similarly prepared from insect cells co-infected with viruses separately expressing $(His)_6$-SUMOstar-POT1 and $(His)_6$-SUMOstar-TPP1-N fusion proteins. Typical final yields were 2 and 6 mg per liter of culture for POT1 and the POT1-TPP1-N heterodimer, respectively.

Purification of protein expressed in bacteria: WT and variants of TPP1-N and OB-hDBD were expressed in *E. coli* BL21(DE3) strain as Sumo-$(His)_{10}$-fusion proteins. For protein purification, cells were harvested, sonicated, and centrifuged to clarify the soluble cellular lysates. Fusion proteins were purified with Ni-NTA agarose resin (Qiagen) and eluted with buffer containing 300 mM imidazole. Following tag cleavage with Ulp1 protease (MTA with Cornell University) (Mossessova and Lima, 2000) TPP1-N or OB-hDBD was further purified by Superdex 75 size-exclusion chromatography (GE Healthcare Life Sciences) in 25 mM Tris-Cl (pH 8), 100 mM NaCl, and 2 mM DTT. Typical final yields were 2 mg and 1-2 mg per liter of culture for TPP1-N and the OB-DBD constructs, respectively.

2.5.6 HeLa culture and stable cell line generation

Tetracycline-inducible protein expression was performed in HeLa-EM2-11ht cells that constitutively express a tetracycline-controlled transcriptional activator and contain a thymidine kinase gene that can be excised with Flp recombinase for facilitating cassette exchange (Nandakumar et al., 2012; Weidenfeld et al., 2009). HeLa-EM2-11ht cells were cultured at 37°C in the presence of 5% CO_2 and propagated in modified DMEM (Dulbecco's Modified Eagle Medium; Gibco 11995-065) medium containing 100 U/mL penicillin, 100 µg/mL streptomycin, and 10% fetal bovine serum. Stable clones for doxycycline-inducible expression of TPP1 (ΔNOB and L104A) were engineered by the same procedure used to create the TPP1 WT line (Nandakumar et al., 2012). Briefly, cells were co-transfected with Lipofectamine 2000 (Thermo Fisher Scientific) and 1 ug each of the p3X-FLAG-TPP1-F3 (TPP1ΔNOB or TPP1-L104A) and a Flp recombinase-expressing plasmid also encoding puromycin resistance. One day of positive selection was performed with puromycin (5 ug/mL; Sigma-Aldrich), followed by 10 days of negative selection in fresh medium adjusted to 50 µM ganciclovir (Sigma-Aldrich). Individual clones were picked and expanded, and two positive clones of each TPP1 variant were selected based on GFP fluorescence (From IRES-GFP locus downstream of TPP1 construct) and Western blot analysis of FLAG-TPP1 signal after overnight induction with doxycycline (200 ng/mL). Protein expression of POT1, TIN2, and TPP1 from transiently transfected p6X-Myc-POT1-BI4 and p6X-Myc-TIN2-BI4 plasmids and from the stable cell lines expressing FLAG-TPP1 (WT, ΔNOB, and L104A) were induced with doxycycline (200 ng/mL).

2.5.7 Co-immunoprecipitation

HeLa-EM2-11ht cells were transfected with 1 ug each of plasmids containing FLAG-TPP1 WT, FLAG-TPP1 ΔNOB, Myc-POT1, or Myc-TIN2 plasmid. 48 h after transfection/induction with doxycycline, cells were washed with PBS, trypsinized, and dislodged with medium containing 50% fetal bovine serum to inactivate the trypsin. Cell pellets were resuspended in 400 µL of lysis buffer [50 mM Tris-Cl (pH 7.4), 20% glycerol, 1 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.02% SDS, 1 mM dithiothreitol, 2 mM phenylmethylsulfonyl fluoride, and complete protease inhibitor cocktail (Roche)] and kept on ice. Next, 33 µL of 4 M NaCl was added followed by 433 µL of water before centrifuging at 16,000 x g for 10 min. Supernatants were used directly in immunoprecipitation. A portion of the lysate (50 µL) was kept aside to serve as an "input" sample. 30 µL of pre-washed anti-FLAG M2 affinity beads slurry (Sigma; A2220) was added to the remaining supernatant and the samples were rocked overnight at 4°C. The beads were washed three times with 0.5X lysis buffer and the bound proteins were eluted in 60 µL of 2X SDS gel loading buffer. All samples were heated for 10 min at 95°C and analyzed by SDS-PAGE followed by immunoblotting with HRP-conjugated anti-FLAG or anti-Myc antibodies.

2.5.8 Immunoblotting

Proteins from cellular lysates or immunoprecipitation were resolved by SDS-PAGE, transferred to nitrocellulose (BioRad), and blocked with StartingBlock (TBS) blocking buffer (Thermo Scientific). Immunoblotting was performed following standard procedures and 1:10,000 dilution of the following antibodies: mouse monoclonal anti-FLAG M2-HRP conjugate (Sigma; A8592), mouse monoclonal anti-c-Myc

(Developmental Studies Hybridoma Bank; 9E10) HRP conjugate (Santa Cruz; sc-40 HRP), and mouse monoclonal anti-β-actin antibody (Sigma; A5441) in conjunction with secondary horseradish peroxidase-conjugated goat antibody against mouse IgG (Santa Cruz Biotechnology). Antibodies were detected by chemiluminescence with ECL plus reagents (Pierce ECL Western Blotting Substrate; Thermo Scientific). The data were visualized using a gel-documentation system (ChemiDoc™ MP System; BioRad). Quantitation of actin or FLAG band intensity was performed using ImageJ software.

2.5.9 Telomere localization of TPP1 constructs

Telomeres were visualized by fluorescence in situ hybridization (FISH) using a fluorescent telomeric PNA probe. Briefly, cells were fixed for 10 min with 4% formaldehyde in PBS, washed, then permeablized for 10 min in PBS containing 0.5% Triton X-100. Cells were then washed twice in PBS and soaked in 2X SSC, 50% formamide for 5 min. 40 µl of hybridization solution containing 0.3 µg/ml Cy3-conjugated PNA-$(CCCTAA)_3$ probe was placed on a microscope slide, and a coverslip containing cells from each engineered stable cell line (WT, ΔNOB, L104A) was inverted on the hybridization solution. DNA was denatured by heat for 6 min at 80°C followed by hybridization for two hours at room temperature. After hybridization the slides were washed twice at room temperature with 50% formamide in 2X SSC for 30 min. Slides were then washed with PBS and used for subsequent immunofluorescence (IF). IF was performed to visualize FLAG-TPP1 constructs to determine telomere localization. For IF, cells were blocked for 30 min with 1 mg/ml BSA, 3% goat serum, 0.1% Triton X-100, and 1 mM EDTA (pH 8.0) in PBS. FLAG-TPP1 constructs were visualized using mouse monoclonal anti-FLAG M2 (Sigma; F1804; 1:500) in combination with Alexa Fluor 633-

conjugated anti-mouse IgG (Life Technologies; A21052). Cells were washed three times with PBS and mounted on microscope slides using ProLong Gold mounting medium with DAPI (Life Technologies). Imaging and analysis were performed as described above.

2.6 Figures and Tables

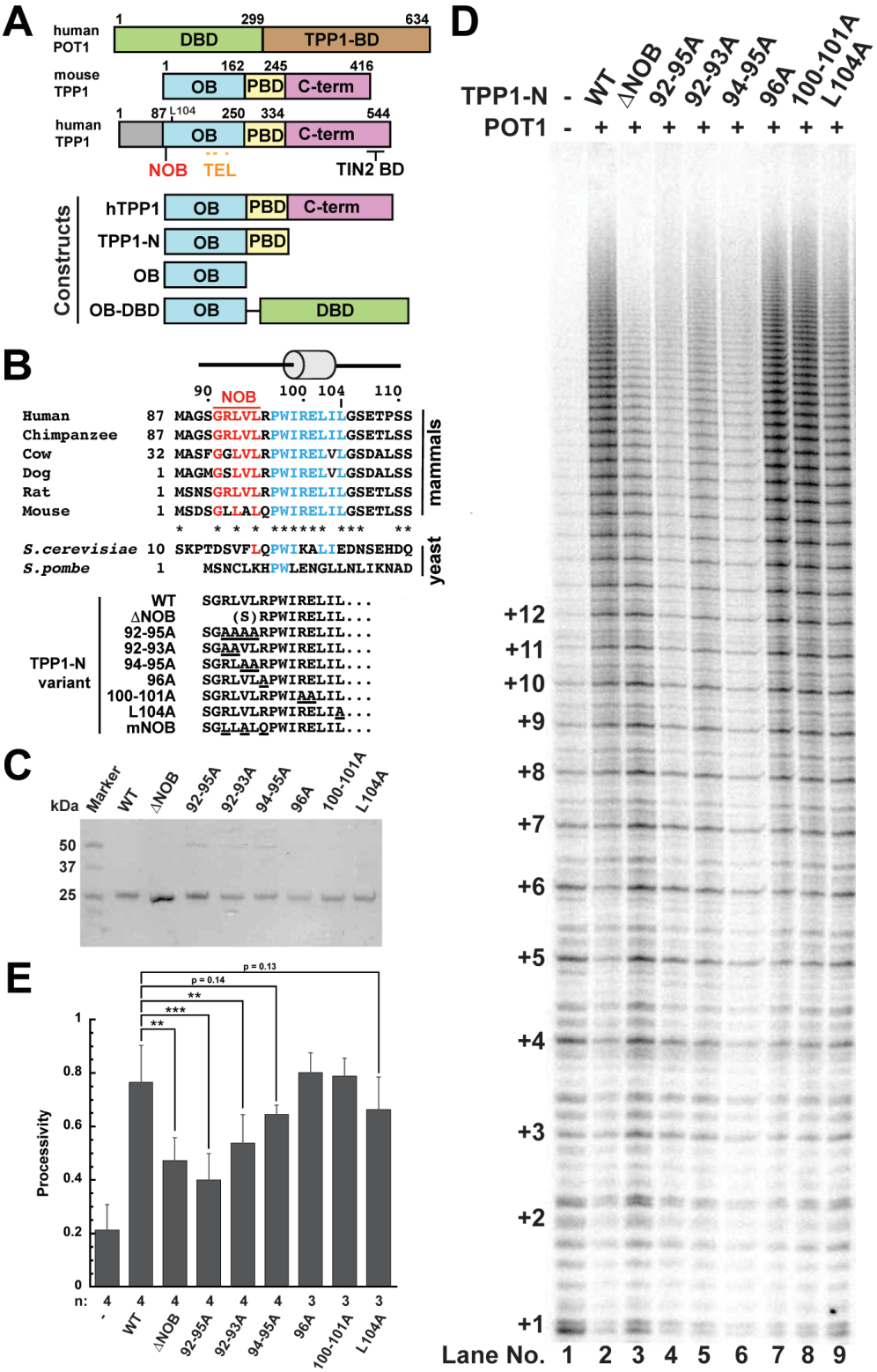


Figure 2.1 NOB region of TPP1 is critical for facilitating repeat addition processivity of telomerase.
(A) Schematic of POT1 and TPP1 proteins and constructs used in this study. The domain organization of human POT1 [DNA-binding domain (hDBD) and TPP1-binding domain (TPP1-BD)], and mouse or human TPP1 [OB domain (OB), POT1-binding domain (PBD), and C-terminal region (C-term) that includes the TIN2-binding domain (TIN2 BD)] are shown. Amino acid numbering above the schematics indicates domain boundaries. NOB (red) and TEL patch (mustard) residues are indicated. Eighty-six aa N-terminal to Met87 in human TPP1 encode a region of unknown function somewhat conserved in primates but not other mammals. (B) *Top*: Primary sequence for the N-terminus of the TPP1 OB domain in representative mammals and yeast homologs Est3 (*S. cerevisiae*) and Tpz1 (*S. pombe*). Est3 protein is homologous to TPP1-OB, and lacks C-terminal sequences that correspond to POT1- and TIN2-binding regions. The conserved PWI residues and first alpha helix are part of the structural core of the OB domain (shown in blue). N-terminal to this region is a largely hydrophobic sequence, the N-terminus of OB domain or NOB, shown in red. Asterisks in the mammalian line-ups indicate sequence identity. *Bottom:* The distinctive N-termini of the TPP1-N variants analyzed in telomerase assays are shown with site-directed mutations underlined. The N-terminus of the ΔNOB variant begins with a serine followed by aa R96. (C) Coomassie blue-stained SDS-PAGE gel of the purified TPP1-N variant proteins. (D) Direct telomerase activity assay with purified POT1 and TPP1-N variants. Lane 1: telomerase extract alone; lanes 2-9 include POT1 and indicated TPP1-N proteins. (E) Quantitative comparison of telomerase processivity for replicate experiments of which panel D is representative. 'n' indicates the number of replicates performed for each construct. Processivity is defined as the ratio of the total intensity from bands representing addition of 9 or more hexad repeats over the total intensity in the lane (omitting repeats 1 and 2) using the rolling ball method to define background. Statistical significance was scored with a two-tailed student's *t*-test.

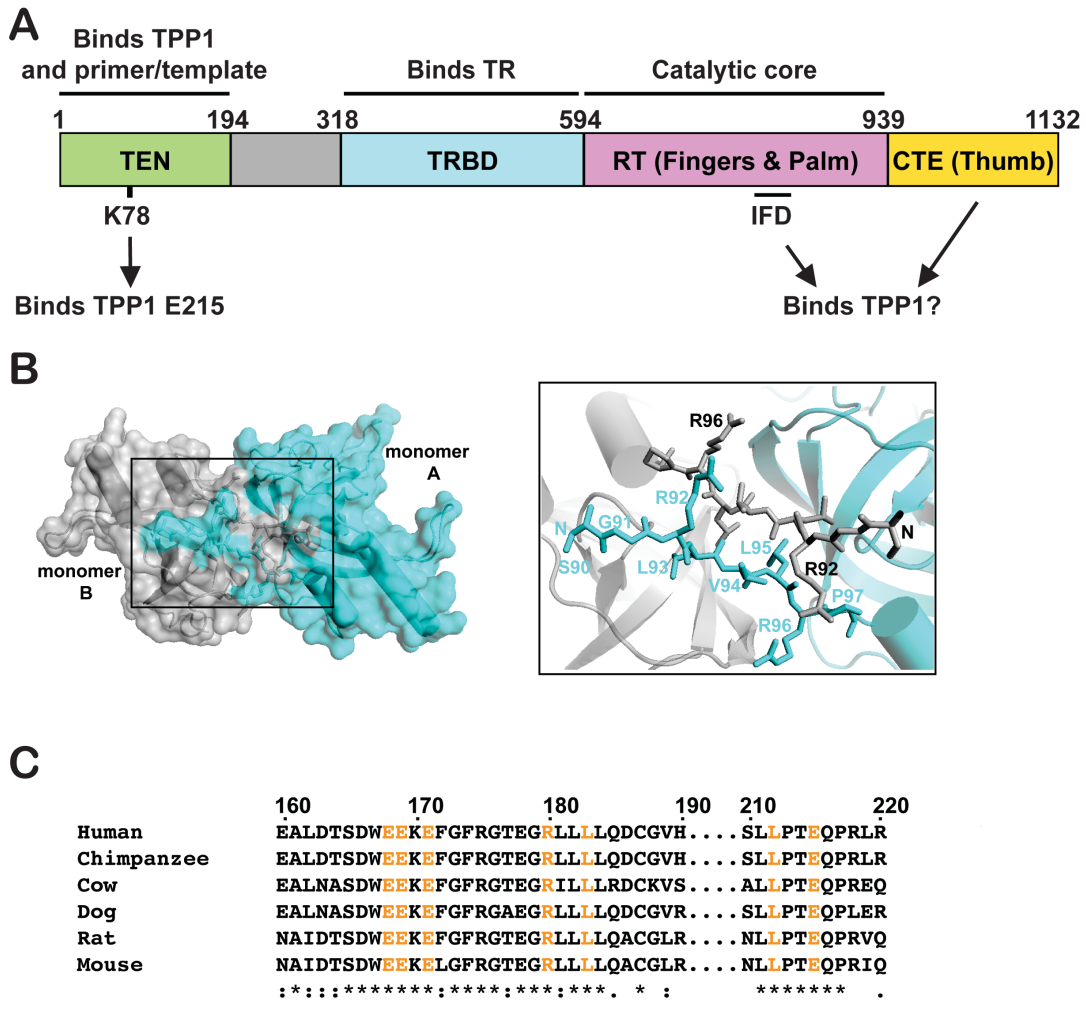


Figure 2.2 Structure of TPP1 OB, and domain diagram and binding properties of human TERT. A) The arrangement of the TEN, TRBD (telomerase RNA binding domain), RT (reverse transcriptase), and CTE (C-terminal extension) domains in the primary structure of human TERT is shown along with annotations for various (known and putative) binding regions (Gillis et al., 2008; Wang and Feigon, 2017). (B) The TPP1 N-terminal residues 90-96 (NH_2-SGRLVLR-COOH), which include the NOB region, of two OB monomers (labeled A and B) form extended conformations and align roughly anti-parallel to each other to mediate homodimerization in the human TPP1 OB domain crystal structure (PDB accession: 2I46) (Wang et al., 2007). The overall structure is shown as a surface model and the site of dimerization (boxed area) is also shown on the right in stick representation for aa 90-96. (C) The TEL patch is highly conserved across mammalian species. Sequence alignment of selected regions of indicated TPP1 homologs with TEL patch residues shown in orange. Amino acid numbers represent the human sequence. " * " beneath sequence lineups indicate identical residues, " : " represent strongly conserved residues, and " . " represent weakly conserved residues as described by the MUSCLE algorithm.

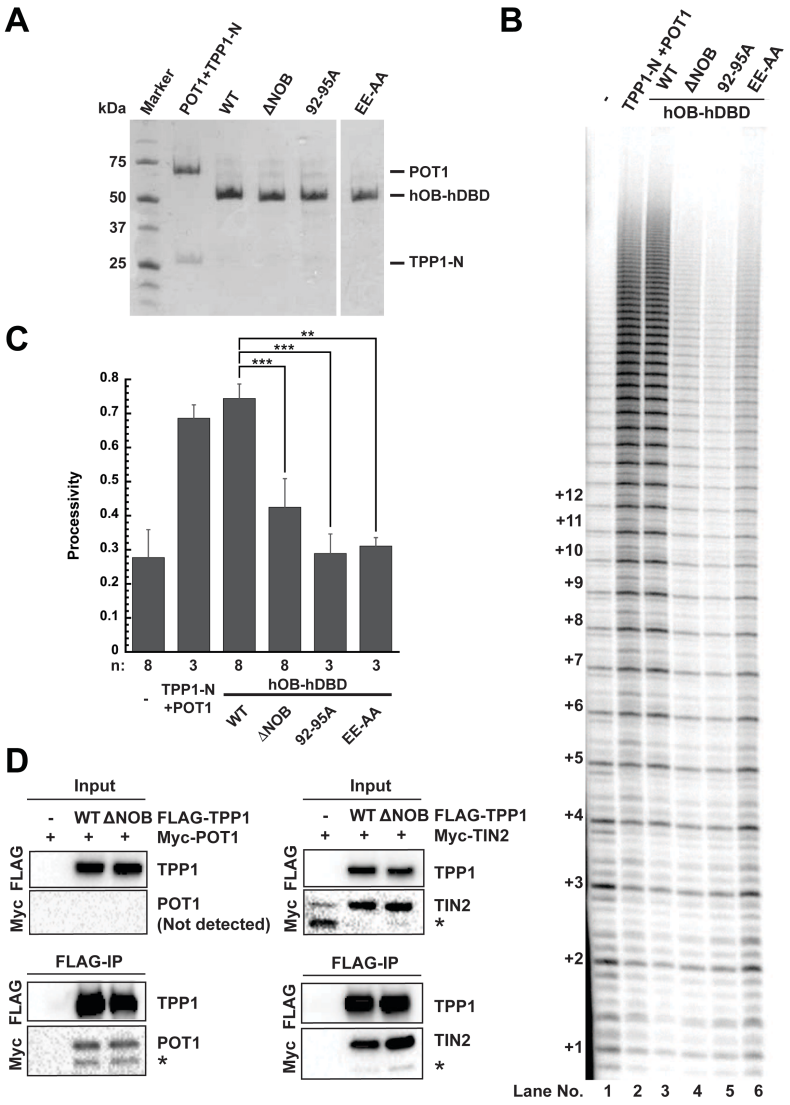


Figure 2.3 Effects of the NOB region on telomerase processivity are independent of TPP1 binding to POT1 and TIN2.
(A) Coomassie blue-stained SDS-PAGE gel of purified POT1/TPP1-N complex, and hOB-hDBD fusion proteins used in panel B. (B) Direct telomerase activity assay. Lane 1: telomerase extract alone; lane 2: includes human POT1 and TPP1-N; and lanes 3-6: include the indicated hOB-hDBD variant. (C) Quantitation of telomerase processivity from indicated number of replicates 'n' of which panel B is representative. Statistical significance was scored with a two-tailed student's *t*-test. (D) Pulldown of FLAG-TPP1 (WT or ΔNOB) and Myc-POT1 or Myc-TIN2 transiently expressed in HeLa cells on anti-FLAG conjugated beads. "Input" and "FLAG-IP" immunoblots refer to soluble lysates prior to incubation with anti-FLAG beads and anti-FLAG beads after overnight immunoprecipitation, respectively. POT1 was not detected in the input sample but was detected in FLAG-IP samples that contained TPP1, consistent with POT1 stabilization by TPP1. Asterisks indicate POT1- and TIN2-derived degradation products.

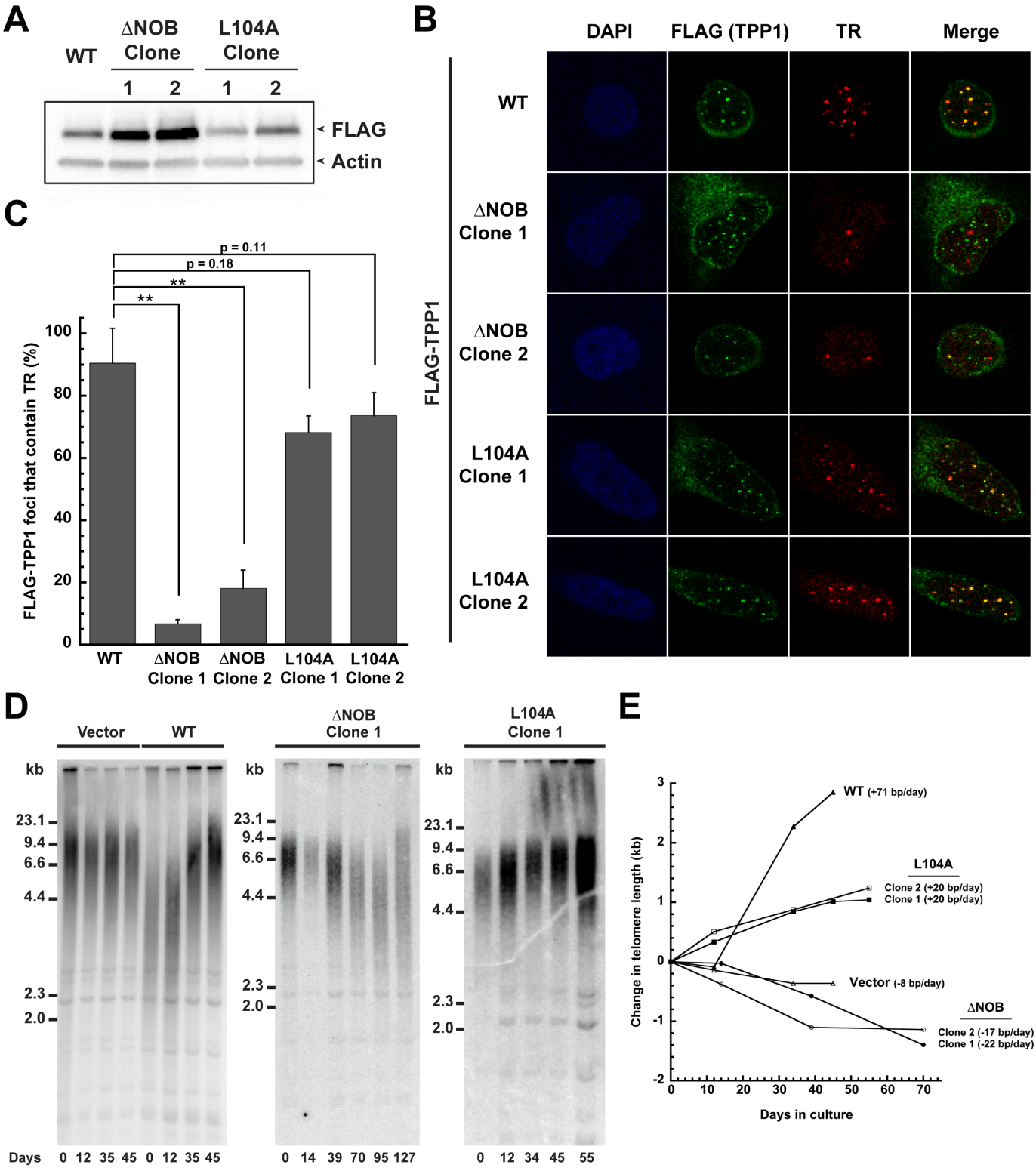

A
WT
ΔNOB Clone
L104A Clone
1
2
1
2
FLAG
Actin
B
DAPI
FLAG (TPP1)
TR
Merge
FLAG-TPP1
WT
ΔNOB Clone 1
ΔNOB Clone 2
L104A Clone 1
L104A Clone 2
C
p = 0.11
p = 0.18
**
**
FLAG-TPP1 foci that contain TR (%)
100
80
60
40
20
0
WT
ΔNOB Clone 1
ΔNOB Clone 2
L104A Clone 1
L104A Clone 2
D
Vector
WT
ΔNOB Clone 1
L104A Clone 1
kb
23.1
9.4
6.6
4.4
2.3
2.0
Days
0 12 35 45 0 12 35 45
0 14 39 70 95 127
0 12 34 45 55
E
Change in telomere length (kb)
3
2
1
0
-1
-2
0 10 20 30 40 50 60 70
Days in culture
WT (+71 bp/day)
L104A
Clone 2 (+20 bp/day)
Clone 1 (+20 bp/day)
Vector (-8 bp/day)
ΔNOB
Clone 2 (-17 bp/day)
Clone 1 (-22 bp/day)

Figure 2.4 Deletion of the NOB of TPP1 disrupts both telomerase recruitment to telomeres and telomere lengthening.
(A) Immunoblot showing expression levels of FLAG-TPP1 WT or mutant protein in the indicated stable cell lines. (B) Stable HeLa-EM2-11ht clones overexpressing the indicated FLAG-TPP1 constructs were analyzed for telomerase recruitment to telomeres using immunofluorescence-fluorescence *in situ* hybridization. "FLAG (TPP1)" shows the FLAG immunofluorescence signal indicative of telomeres (green). Fluorescence *in situ* hybridization was used to detect telomerase RNA (TR; red). Merge panels reveal the extent of recruitment of telomerase to telomeres (yellow). (C) Quantitation of telomerase recruitment data of which panel B is representative. The mean percentage of FLAG-TPP1 foci containing TR and standard deviation for triplicate measurements (>100 telomere foci scored per clone) were plotted. Statistical significance was scored with a two-tailed student's *t*-test. (D) Telomeric restriction fragment (TRF) Southern blot analysis of genomic DNA from HeLa-EM2-11ht cell lines stably expressing the indicated TPP1 constructs (Vector, WT, ΔNOB, and L104A) for the indicated number of days in culture. (E) The change in mean telomere length for data shown in panel D was plotted against the number of days in culture.

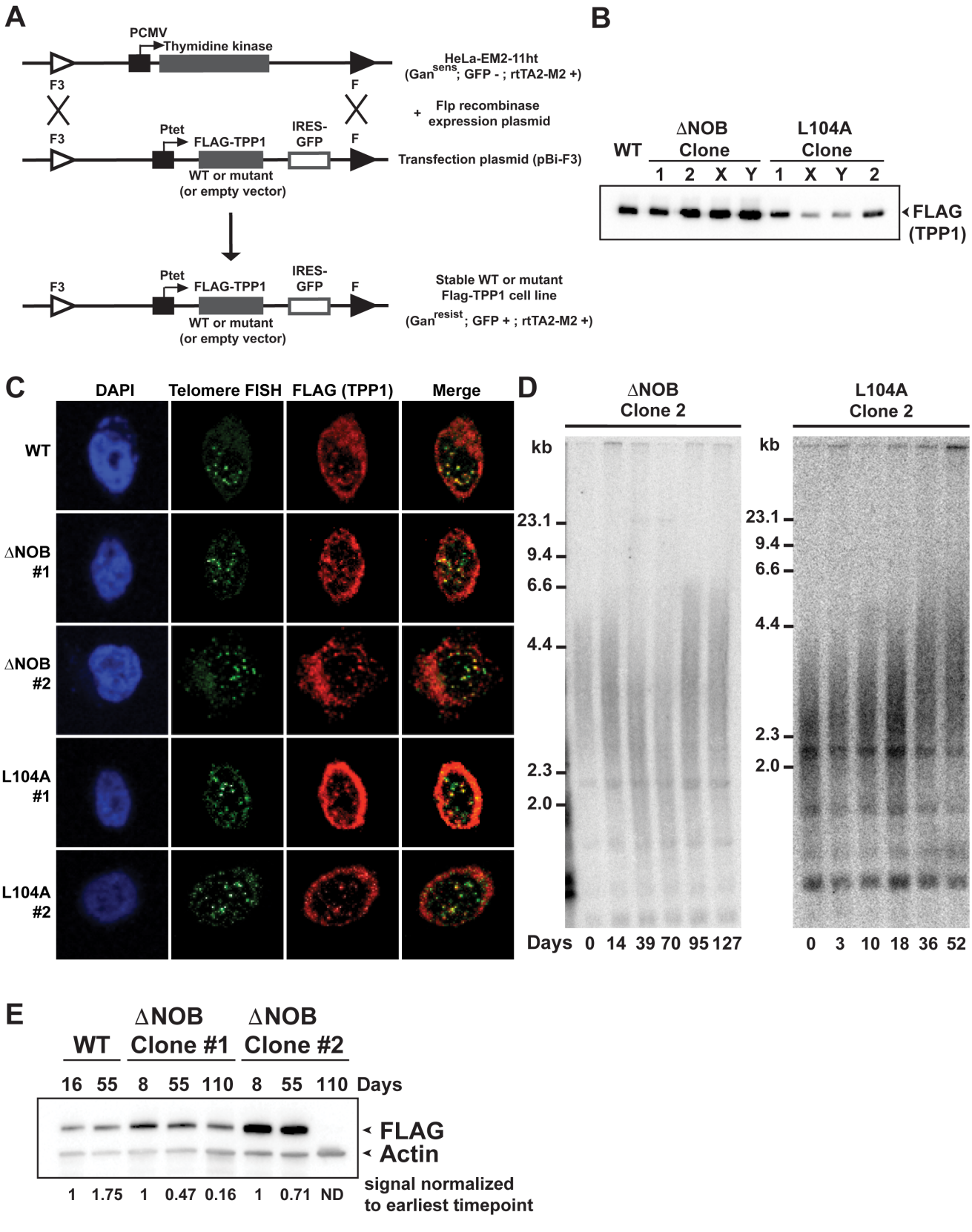

A
PCMV
Thymidine kinase
F3
F
HeLa-EM2-11ht
(Gan^sens; GFP - ; rtTA2-M2 +)
+ Flp recombinase expression plasmid
Ptet
FLAG-TPP1
IRES-GFP
WT or mutant
(or empty vector)
Transfection plasmid (pBi-F3)
Stable WT or mutant Flag-TPP1 cell line
(Gan^resist; GFP + ; rtTA2-M2 +)
B
WT
ΔNOB Clone
L104A Clone
1 2 X Y 1 X Y 2
FLAG (TPP1)
C
DAPI
Telomere FISH
FLAG (TPP1)
Merge
WT
ΔNOB #1
ΔNOB #2
L104A #1
L104A #2
D
ΔNOB Clone 2
L104A Clone 2
kb
23.1
9.4
6.6
4.4
2.3
2.0
Days 0 14 39 70 95 127
0 3 10 18 36 52
E
WT
ΔNOB Clone #1
ΔNOB Clone #2
16 55 8 55 110 8 55 110 Days
FLAG
Actin
1 1.75 1 0.47 0.16 1 0.71 ND
signal normalized to earliest timepoint

Figure 2.5 Cell lines stably expressing TPP1 variants for telomere lengthening and telomerase recruitment studies.
(A) Scheme for generating cell lines stably integrated with TPP1 WT or mutant, or empty vector sequence. (B) Western blot for four clones each of ΔNOB and L104A alongside a WT clone shows high expression for ΔNOB and low expression for L104A. Clones 1 and 2 of each mutant were selected for further analysis of telomerase recruitment and telomere length. (C) IF-FISH to show colocalization of indicated TPP1 constructs with telomeres. “FLAG (TPP1)” shows the FLAG immunofluorescence signal indicative of TPP1 (red), and “Telomere FISH” indicates telomeric DNA probed with a C-rich, Cy3-labeled PNA probe (green). Merge panels reveal colocalization of TPP1 variants with telomeric DNA (yellow). (D) Telomeric restriction fragment (TRF) analysis of clone 2 for HeLa-EM2-11ht cell lines stably expressing the indicated TPP1 constructs (ΔNOB and L104A) for the indicated number of days in culture. (E) FLAG immunoblots from indicated cell lines at indicated days in culture. The FLAG signal was divided by the actin signal in each lane to account for loading differences. The numbers below the blot indicate actin-normalized FLAG signal in that lane divided by the corresponding signal in the earliest time-point for that clone.

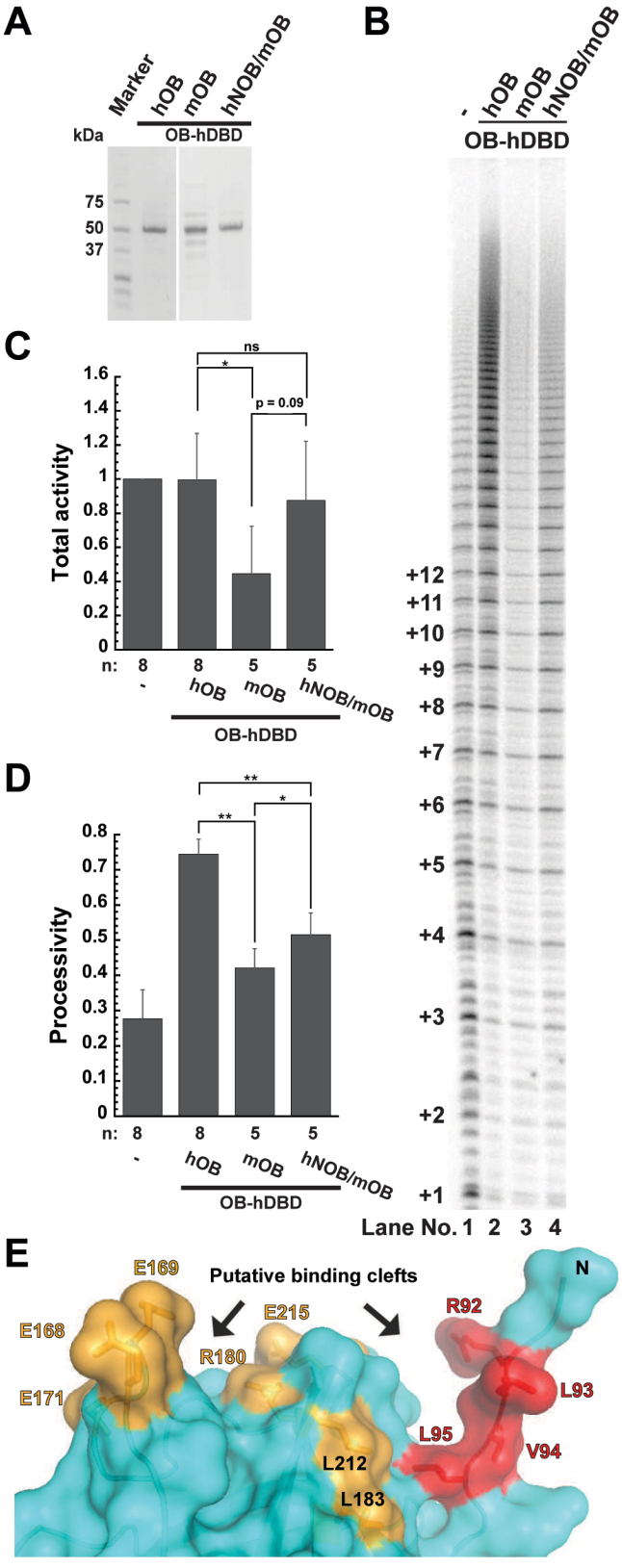

A
Marker
hOB
mOB
hNOB/mOB
kDa
OB-hDBD
75
50
37
B
-
hOB
mOB
hNOB/mOB
OB-hDBD
+12
+11
+10
+9
+8
+7
+6
+5
+4
+3
+2
+1
Lane No. 1 2 3 4
C
Total activity
ns
*
p = 0.09
n: 8 8 5 5
hOB
mOB
hNOB/mOB
OB-hDBD
D
Processivity
**
**
*
n: 8 8 5 5
hOB
mOB
hNOB/mOB
OB-hDBD
E
Putative binding clefts
E169
E168
E171
E215
R180
R92
N
L93
L95
V94
L212
L183

Figure 2.6 Inability of mouse TPP1 to stimulate human telomerase is partially rescued by including the human NOB sequence.
(A) Coomassie blue-stained SDS-PAGE gel of purified OB-hDBD variants containing either hTPP1 OB sequence (hOB), mTPP1 OB sequence (mOB), or mTPP1 OB sequence with site-directed mutations to reconstitute the human TPP1 NOB sequence (hNOB/mOB). (B) Direct telomerase activity assay. Lane 1: telomerase extract alone; lanes 2-4: include indicated OB-hDBD variant. Quantitative comparison of telomerase activity (C) and processivity (D) from the indicated number of replicates 'n' for which panel B is representative. Activity was determined using the entire signal in the lane with rolling ball background correction, normalized against the signal observed for telomerase extract alone. Statistical significance was scored with a two-tailed student's *t*-test. (E) The NOB (red) and TEL patch (mustard) residues in the structure of the human TPP1 OB domain (PDB accession: 2I46) (Wang et al., 2007) line two prominent clefts (indicated with arrows) that could bind telomerase.

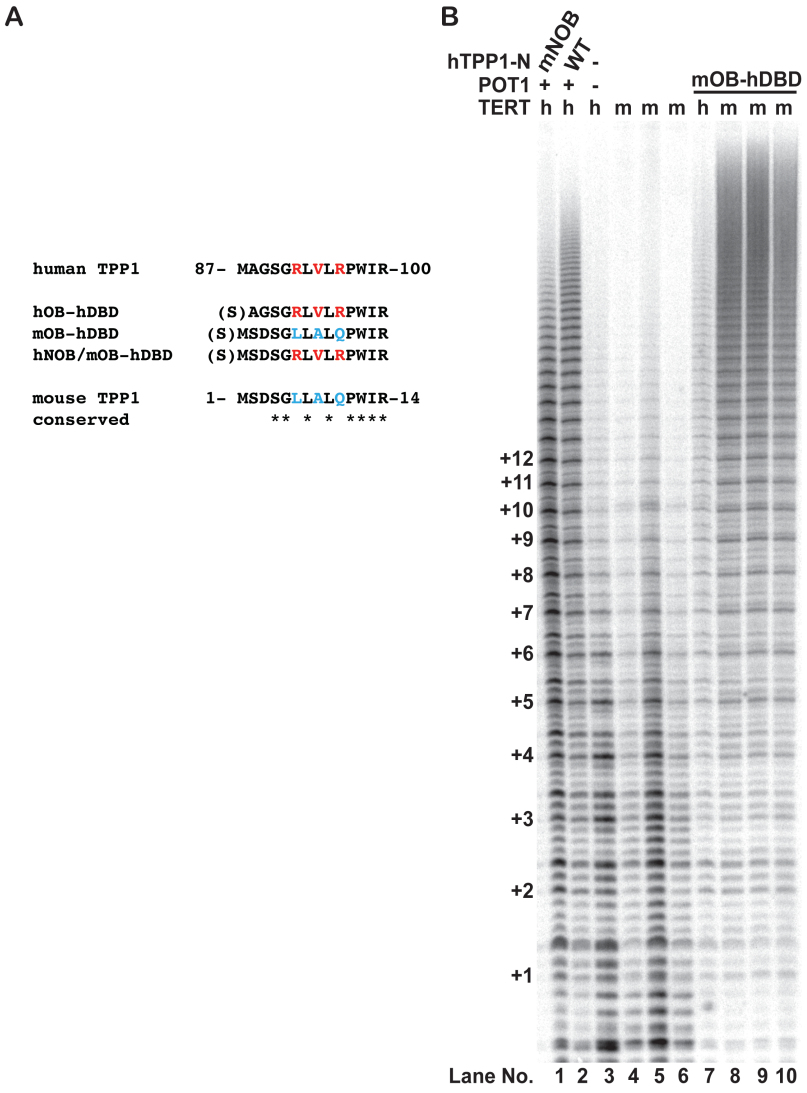


Figure 2.7 Mouse OB-DBD stimulates mouse telomerase but not human telomerase.
(A) Schematic of N-termini for mouse and human OB-DBD variants compared with wild-type sequence for human TPP1 (above) and mouse TPP1 (below). Conserved amino acids between these N-termini are marked with asterisks. (B) Mouse NOB stimulates reconstituted mouse telomerase but not human telomerase. Direct activity assay using extracts with telomerase reconstituted in HEK 293T cells from hTR and either hTERT (human telomerase) or mTERT (mouse telomerase). Lane 1: human telomerase plus human POT1 and TPP1-N mNOB; lane 2: human telomerase plus human POT1/ TPP1-N WT; lane 3: human telomerase extract alone; lanes 4-6: mouse telomerase reconstituted with differing ratios of transfected mTERT/TR plasmids (mTERT/TR is 1:3 in lane 4, 1:1 in lane 5, and 3:1 in lane 6); lane 7: human telomerase (hTERT/TR = 1:3) plus mOB-hDBD; lanes 8-10: same as lanes 4-6 except that mOB-hDBD was included in the reaction. hTR was used to reconstitute activity with mTERT, because telomerase reconstituted with mTR exhibits very low telomerase processivity, as previously described (Chen and Greider, 2003a; Zaug et al., 2010).

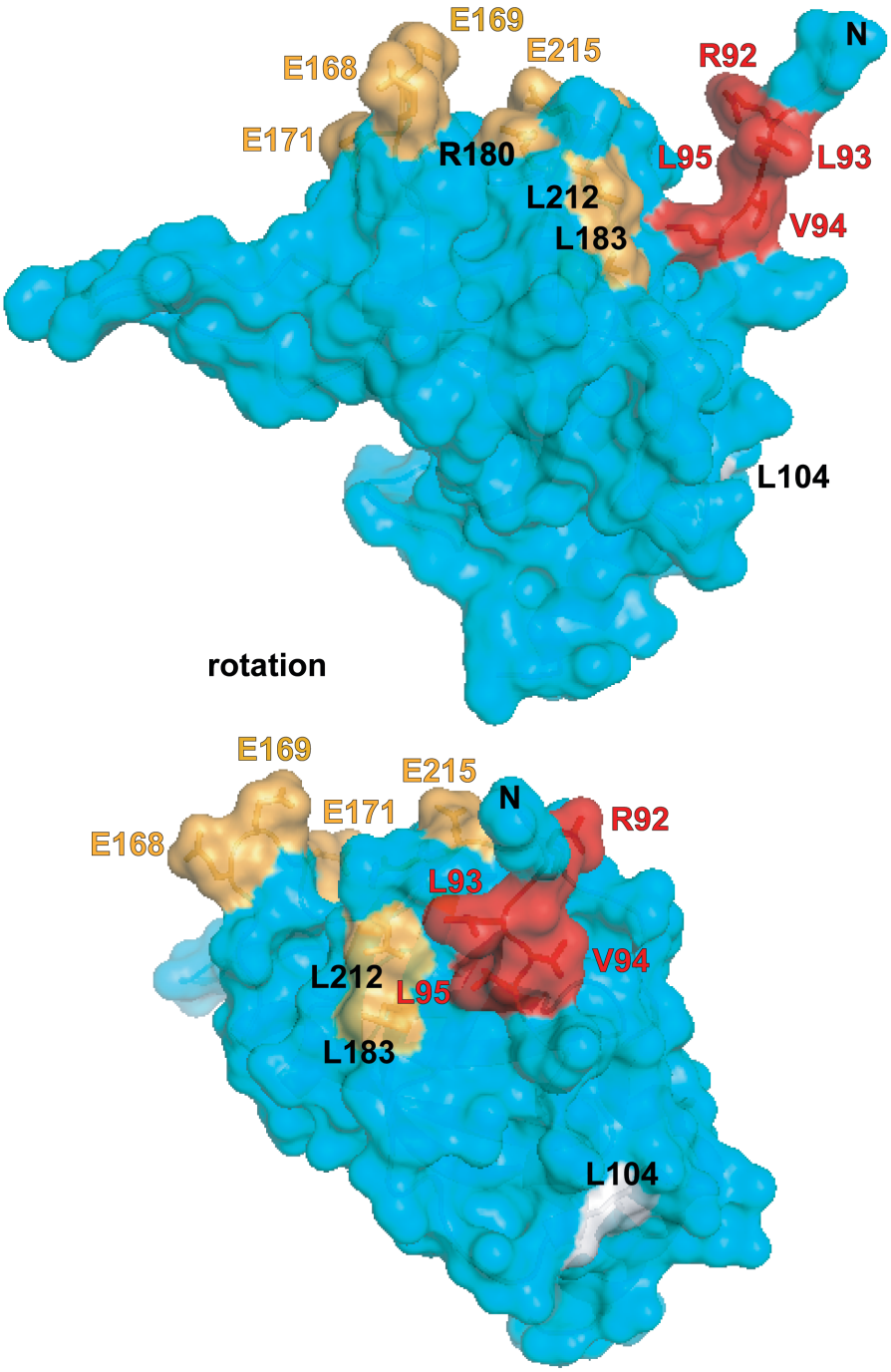


Figure 2.8 Relative locations of the TEL patch, NOB, and L104 regions on the surface of the TPP1 OB domain.
Two separate surface views of TPP1 OB with TEL patch residues shown in mustard, NOB residues in red, and L104 in grey.

CHAPTER III

Two separation-of-function isoforms of human TPP1 and a novel intragenic RNA dictate telomerase regulation in somatic and germ cells[2]

3.1 Abstract

Telomerase replicates chromosome ends in germ and somatic stem cells to facilitate continued proliferation. Telomerase action depends on the telomeric protein TPP1, which recruits telomerase to telomeres and facilitates processive DNA synthesis. Here we identify separation-of-function long (TPP1-L) and short (TPP1-S) isoforms of

[2] A modified version of this chapter is published in Cell Reports (DOI: 10.1016/j.celrep.2019.05.073). I performed experiments involving TPP1-L and TPP1-S expression, separation-of-function, CRISPR KO, localization, and RNAscope. I performed primer extension experiments and protein purification with the help of Valerie Tesmer. Kamlesh Bisht performed sunRNA-related experiments except for sunRNA rescue experiments and sunRNA 4, 5, and 6 expression experiments which I performed. Sue Hammoud and Adrienne Shami provided expertise on spermatogenesis aspects, imaged and interpreted RNAscope testes sections with me, and conducted bioinformatics analysis of archived RNA-Seq data. I wrote the manuscript with significant input from Jayakrishnan Nandakumar.

TPP1 differing only in 86 amino acids at their N-terminus. While both isoforms retain the ability to recruit telomerase, only TPP1-S facilitates efficient telomere synthesis. We identify an intragenic RNA in the 3'-UTR of the TPP1-encoding gene that specifically shuts down TPP1-L to establish TPP1-S as the predominant isoform in somatic cells. Strikingly, TPP1-L is the major isoform in differentiated germ cells where we observed that TERT expression persists, suggesting that TPP1-L restrains telomerase in this context. We show how expression and regulation of TPP1 isoforms determine telomerase function, and demonstrate how alternative transcription start sites allow one gene to perform distinct functions in different biological contexts.

3.2 Introduction

The end replication problem arises due to incomplete chromosome end synthesis by DNA polymerases. This leads to the gradual loss of DNA at the ends of chromosomes during every replication cycle. This chromosome shortening sets a limit on the number of times most somatic cells can divide, thereby providing a natural anti-tumorigenic mechanism in large, long-lived mammals such as humans (Gomes et al., 2011). However somatic and germline stem cells must preserve their ability to self-renew over long periods of time. Telomerase, a unique ribonucleoprotein complex, is a reverse transcriptase that synthesizes DNA at the 3' ends of chromosomes (Greider and Blackburn, 1985). Using a template sequence in its RNA subunit (TR) and a reverse transcriptase protein subunit (TERT), telomerase synthesizes multiple telomeric repeats (GGTTAG in mammals) at chromosome ends compensating for incomplete DNA replication (Greider and Blackburn, 1989; Lingner et al., 1997; Meyerson et al., 1997). Not surprisingly, germline mutations in telomerase or in genes important for

telomerase function result in diseases referred to as telomeropathies (Dokal, 2011; Savage, 2014).

While reduced telomerase function in stem cells can result in telomeropathies, aberrant reactivation of telomerase in somatic cells is a hallmark of a majority of cancers (Kim et al., 1994). Thus, telomerase must be tightly regulated; requiring sustained expression in stem cells, but complete shutdown upon differentiation. Interestingly, telomerase is constitutively expressed in most somatic cells of smaller, short-lived mammals such as rodents (Gomes et al., 2011; Prowse and Greider, 1995). Fewer total cell divisions in their lifetime likely enable these organisms to sustain the benefits of telomerase expression without exacerbating the risk of oncogenesis.

Human telomeres are composed of telomeric DNA repeats bound to a six-protein complex called shelterin (Figure 3.1A) (Palm and de Lange, 2008). Shelterin protects natural chromosome ends from being recognized as double stranded DNA breaks requiring repair. While TRF1 and TRF2 bind the double stranded telomeric DNA (Broccoli et al., 1997), POT1 protects the single stranded 3' overhang (Baumann and Cech, 2001; Lei et al., 2004). POT1 binds TPP1 to form a heterodimer with greater affinity for single stranded telomeric DNA than POT1 alone (Wang et al., 2007). The TIN2 protein connects TPP1 to TRF1 and TRF2 (Frescas and de Lange, 2014b; Kim et al., 1999), while the protein Rap1 constitutively binds TRF2 (Li et al., 2000).

Shelterin also provides a mechanism for recruiting telomerase to chromosome ends (Figure 3.1A) (Nandakumar and Cech, 2013). This is facilitated by an OB (oligonucleotide/oligosaccharide-binding) fold domain in the shelterin protein TPP1 (encoded by the *ACD* gene), which recruits telomerase to telomeres (Figure 3.1A)

(Abreu et al., 2010; Xin et al., 2007). Once recruited, telomerase synthesizes telomeric DNA with high processivity in a POT1-TPP1 dependent manner (Wang et al., 2007). Two regions in the OB domain, the TEL (TPP1's glutamate (E) and leucine (L) rich) patch and the NOB (N-terminus of OB domain), are critical for all of TPP1's telomerase-associated functions, (Grill et al., 2018; Nakashima et al., 2013; Nandakumar et al., 2012; Sexton et al., 2012; Zhong et al., 2012).

A distinct difference at the N-terminus between the human and mouse TPP1 orthologs was noted during their discovery. Human TPP1 is annotated to encompass 544 amino acids (aa) initiating at Met1 (Figure 3.1A) (Houghtaling et al., 2004; Liu et al., 2004b; Ye et al., 2004b); we refer to this isoform as TPP1-L. As such a majority of studies on human TPP1 involved TPP1-L overexpression (Abreu et al., 2010; Sexton et al., 2014; Ye et al., 2004b; Zhong et al., 2012). Yet, the existence of a shorter isoform that initiates at Met87, and referred to here as TPP1-S, was suggested by the realization that rodent TPP1 unequivocally initiates at a Met equivalent to human TPP1 Met87 (Figure 3.1A,B) (Hockemeyer et al., 2007; Houghtaling et al., 2004; Nandakumar et al., 2012; Xin et al., 2007). Thus the TPP1-OB crystal structure (Wang et al., 2007) and other functional studies of TPP1 have used TPP1-S protein (Hwang et al., 2012; Nandakumar et al., 2012; Schmidt et al., 2014). Although TPP1-L is also able to recruit telomerase to the telomere (Abreu et al., 2010; Zhong et al., 2012), telomere hyperelongation that is characteristic of TPP1-S overexpression has not been noted for TPP1-L (Houghtaling et al., 2004; Nandakumar et al., 2012; Xin et al., 2007; Ye et al., 2004b). Here we unveil the existence and separation-of-function of the two isoforms of

human TPP1, discover an intragenic noncoding RNA (ncRNA) that differentially regulates these isoforms, and suggest distinct roles for them in germ and somatic cells.

3.3 Results

3.3.1 TPP1-S, but not TPP1-L, overexpression causes hyperelongation of telomeres

To directly compare the effects of TPP1-L and TPP1-S in telomere length regulation, we used a single-site integration, doxycycline (dox)-inducible stable cell line strategy that we have used previously (Nandakumar et al., 2012). We engineered a stable cell line that expresses C-terminally FLAG-tagged TPP1-L (aa 1-544) and compared it to a previously characterized C-terminally FLAG-tagged TPP1-S (aa 87-544) stable cell line (Nandakumar et al., 2012). Individual clones were selected to equalize TPP1 protein level. In all TPP1-L clones we also observed a modest amount of shorter FLAG-tagged protein that was similar in size to that of FLAG-tagged TPP1-S (Figure 3.1C). We reasoned that the M87 start codon in TPP1-L cDNA was being utilized to produce FLAG-tagged TPP1-S protein. To circumvent this alternative translation initiation site, we engineered another HeLa cell line that stably expresses FLAG-tagged TPP1-L protein harboring an M87A mutation. As expected, the M87A mutation blocked the production of the shorter TPP1 protein (Figure 3.1D). Consistent with previous results, overexpression of TPP1-S resulted in the characteristic lengthening of telomeres over time (+42 bp/day), while telomeres of the “Vector” cell line remained relatively stable (-6 bp/day) (Figure 3.1E,F; Figure 3.2) (Nandakumar et al.,

2012). In contrast, overexpression of TPP1-L resulted in either no lengthening (0 bp/day) or a moderate lengthening (17 bp/day) of telomeres (Figure 3.1E,F; Figure 3.2A,B). The low level of telomere lengthening in TPP1-L overexpressing cells is likely attributable to co-expression of TPP1-S protein, as the TPP1-L M87A cell lines had short telomeres (~3 kb) that did not elongate over time (Figure 3.1D,E,G; Figure 3.2A,C). Thus, TPP1-S and TPP1-L play contrasting roles in telomere length maintenance.

3.3.2 Both TPP1-S and TPP1-L recruit telomerase to telomeres

We asked if TPP1-L M87A, like TEL patch or NOB mutants, was impaired in its ability to recruit telomerase to telomeres. As expected, telomerase (TR) co-localized with over 90 percent of TPP1-S foci, indicative of robust telomerase recruitment to telomeres (Figure 3.3A,B). Surprisingly we also observed robust recruitment of telomerase to telomeres in both TPP1-L M87A clones (91% and 84% for clone 1 and clone 2, respectively; Figure 3.3A,B; Figure 3.4A) and both TPP1-L WT clones (Figure 3.4B). These data distinguish TPP1-L from the previously characterized TEL patch and NOB mutants of TPP1-S.

3.3.3 Both TPP1-S and TPP1-L protect chromosome ends

TPP1 is unique as it not only recruits telomerase to telomeres for end replication, but it also helps protect chromosome ends. Co-immunoprecipitation experiments showed that both FLAG-tagged TPP1-S and TPP1-L M87A efficiently pulled down transiently co-expressed POT1 or TIN2 on anti-FLAG conjugated beads (Figure 3.4C,D). TPP1-L and TPP1-S were also indistinguishable in their abilities to form larger

shelterin subcomplexes with POT1, TIN2, and TRF2 (Figure 3.3C; Figure 3.4E). Although neither protein bound DNA directly, both TPP1-L and TPP1-S further stabilized the POT1-DNA complex *in vitro* (Figure 3.3D) and co-localized with telomeric DNA in cells (Figure 3.4F). Finally, neither protein gave rise to a substantial number of telomere-dysfunction induced foci when overexpressed (>70% of cells had no TIFs) (Figure 3.4G) and no differences in TERRA accumulation were observed at TPP1-L versus TPP1-S telomeres (Figure 3.4H). Taken together these data suggest that both TPP1-L and TPP1-S are fully proficient at end protection.

3.3.4 TPP1-L blocks telomere extension by telomerase

We asked if the lack of telomere elongation in TPP1-L M87A overexpressing cells was indicative of a telomerase activation defect. *In vitro* analysis of TPP1-L in telomerase primer extension assays suggested a modest defect in stimulating telomerase processivity compared to TPP1-S, but the propensity for N-terminal degradation of TPP1-L complicated this analysis (Figure 3.6A-D). To circumvent this problem, we utilized an in-cell telomere extension assay to visualize newly synthesized telomeric repeats (Diolaiti et al., 2013). We transiently transfected HeLa-EM2-11ht cells and stable cell lines with WT TERT and a mutant telomerase RNA (TR) that contains a non-telomeric GCCAAC (WT: CCAAUC) template sequence. We then simultaneously determined the localization of FLAG-TPP1 protein (green; Figure 3.5B) by IF, as well as telomerase (telomerase^{mutTR}; cyan; Figure 3.5B) and newly added mutant telomeres (red; Figure 3.5B) by FISH (see Figure 3.5A for two contrasting outcomes of the experiment). As expected, TPP1-S overexpression resulted in telomerase^{mutTR}

recruitment and synthesis of mutant telomeres, with more than 70% of TR foci containing mutant telomeres (Figure 3.5B,D). In contrast, overexpression of TPP1-L M87A resulted in significantly fewer foci with mutant repeats, with only 35% of TR foci containing detectible mutant telomere signal (Figure 3.5B,D). The mutant telomere signal in TPP1-L M87A cells was also significantly less intense than that in TPP1-S cells (Figure 3.5E). In HeLa cells (that lack TPP1 overexpression) over 75% of telomerase^{mutTR} contained mutant telomeres (Figure 3.5B,D), suggesting that the striking TPP1-L phenotype cannot be attributed to differences in TPP1-L and TPP1-S overexpression levels. Instead these data suggest that TPP1-L blocks telomerase action at telomeres.

To examine how TPP1-L restricts telomerase action, we inspected the 86 amino acids unique to TPP1-L. This region is highly basic (theoretical pI: 12.3) and predicted to be unstructured, consistent with its glycine/proline/arginine-rich composition (Figure 3.1B) and our circular dichroism results (Figure 3.6E). To test the importance of the basic region, we mutated three arginine residues to glutamate residues (R43E/R46E/R48E; TPP1-L M87A R_3E_3; red "R" residues in Figure 3.1B; Figure 3.5C). Transient transfection of TPP1-L M87A R_3E_3 rescued telomerase activity with 63% of telomerase^{mutTR} foci containing mutant telomere sequence, compared to only 37% with transient transfection of TPP1-L M87A (Figure 3.5B,D). TPP1-L M87A R_3E_3 mutant telomeres were also more intense than TPP1-L M87A telomeres (Figure 3.5E). Finally, cells stably overexpressing TPP1-L M87A R_3E_3 maintained longer telomeres over time than cells expressing TPP1-L M87A (Figure 3.5F,G). These data confirm that the arginine-rich region unique to TPP1-L is important for its inhibition of telomerase

function, providing a molecular basis for the separation-of-function between TPP1-S and TPP1-L.

3.3.5 TPP1-L and TPP1-S isoforms are both primed for transcription but only TPP1-S accumulates in most cells

The GENCODE transcript set describes two distinct transcripts for the human *ACD* locus that differ only at their 5' ends (Figure 3.7A) (Harrow et al., 2012). The shorter transcript (TPP1-S mRNA) initiates downstream of the codon for TPP1 Met1 and can be translated into TPP1-S protein (starting at Met87) but not TPP1-L protein. The longer transcript (TPP1-L mRNA) can code for both TPP1-L and TPP1-S proteins. Genome-wide ChIP studies in 91 cell lines, including cancer cell lines, fibroblasts, and embryonic stem cells, show a bimodal enrichment of transcription factors at the 5' ends of TPP1-L and TPP1-S transcripts indicative of two distinct promoters (Figure 3.7A; Figure 3.8A) (Gerstein et al., 2012; Wang et al., 2012; Wang et al., 2013). A similar accumulation of active chromatin marks is observed at both transcription start sites (Figure 3.8A). GRO-cap and GRO-seq data, which capture nascent RNA, similarly show that both TPP1-L and TPP1-S promoters fire in all cell lines tested (Figure 3.7A; Figure 3.8A) (Core et al., 2014; Core et al., 2008). In contrast, RNA-seq data suggest that the steady state level of TPP1-S mRNA is significantly higher than that of TPP1-L mRNA (10-100 fold more reads common to both isoforms than reads specific for TPP1-L; Figure 3.7A; Figure 3.8A) (Consortium, 2012) (data deposited by Wold Lab at Caltech). Cap analysis gene expression (CAGE) data, which capture the 5' ends of capped RNA in cells (Carninci et al., 1996), reveal only a small number of reads for the 5' end of TPP1-L relative to TPP1-S mRNA (note that y-axis is in log scale; Figure 3.7A),

suggesting that although both transcripts are primed for transcription, only TPP1-S transcript accumulates at steady state. The underrepresentation of TPP1-L by RNA-seq and CAGE was observed in all the cell lines tested, including several cancer cell lines, and embryonic stem cells (Figure 3.7A; Figure 3.8A). The high ratio of TPP1-S:TPP1-L mRNA correlates with a high ratio for the rates of protein translation, as ribosome profiling reads for the two isoforms mirror the RNA-seq data in all of the cell lines tested (Figure 3.7A; Figure 3.8A) (Michel et al., 2014).

We independently explored TPP1 isoform abundance using immunoblotting. A polyclonal rabbit antibody detects a discrete band for TPP1-S in HeLa cells (close to 50 kDa) that is diminished in the presence of an shRNA that targets a region common to both isoforms (Figure 3.7B) (Bisht et al., 2016). Even when we enriched for endogenous TPP1 in BJ fibroblasts, HeLa, and HEK 293T cell lysates using POT1-biotinylated DNA immobilized on streptavidin beads we were only able to detect the TPP1-S isoform (Figure 3.7C). To increase detection sensitivity we engineered a 3X-FLAG tag at the C-terminus of the endogenous *ACD* locus in HEK 293T cells using CRISPR-Cas9 (Figure 3.8B), but still only observed a band for TPP1-S (Figure 3.7D; Figure 3.8C). We then used CRISPR-Cas9 technology to engineer TPP1-L knockout (KO) clonal HeLa cell lines. We targeted the endogenous *ACD* locus to generate a stop codon in the place of Met11 (the only Met between Met1 and Met87), preventing translation upstream of aa 86 (Figure 3.7E-G; Figure 3.8D). TPP1-L KO clones (Figure 3.8D) did not exhibit any telomere length phenotype, suggesting that HeLa cell telomere length is unaffected by KO of TPP1-L (Figure 3.7H). In summary, although both TPP1-

L and TPP1-S transcripts are similarly primed for transcription, TPP1-S is the major functional isoform in the cell lines examined.

3.3.6 An intragenic noncoding RNA in the 3'-UTR shuts down TPP1-L, but not TPP1-S

Given that both TPP1-L and TPP1-S promoters displayed comparable potential for mRNA production, we wished to determine how TPP1-L mRNA is selectively down regulated in most human cells. We noted that the short (93 nt) 3'-UTR of the gene (*ACD*) coding for human TPP1 contains a stretch of 26 nucleotides that are almost fully complementary to regions in the TPP1-L open reading frame (ORF) (Fig 5A). Furthermore, the presence of CAGE tags close to the 3' end of the *ACD/TPP1* gene were suggestive of a ncRNA that encompasses the 3'-UTR sequence of TPP1 (see CAGE sub-panel in Figure 3.7A). The existence of such an RNA was also suggested by Northern blot analysis of HEK 293T cells (Figure 3.10A). To further verify that such a 3'-UTR derived ncRNA exists and to examine what influence it might bear on TPP1-L expression, we performed 5' RACE in HEK 293T cells. We identified a set of ncRNAs with the complete 3'-UTR sequence but differing sequences at their 5' ends, in agreement with the broad peaks in the CAGE data (Figure 3.9B; Figure 3.7A). We call these ncRNAs silencing 3'-UTR-derived noncoding RNAs or sunRNAs, based on their function described later.

We studied how the sunRNA affects TPP1-L levels using a bidirectional expression plasmid to co-express a protein-coding cDNA and a ncRNA gene as described previously (Bisht et al., 2017; Nandakumar et al., 2012). We cloned FLAG-TPP1-L cDNA ("start codon to stop codon" cDNA) in one multiple cloning site, and the

cloned intragenic sunRNA candidates in the other (Figure 3.9C). A plasmid encoding FLAG-TPP1-L, but lacking the sunRNA (Empty) was engineered as a negative control. Strikingly, while expression of FLAG-TPP1-L/Empty resulted in the expected accumulation of FLAG-TPP1-L protein (lane 1; Figure 3.9D), co-expression of FLAG-TPP1-L and the sunRNA completely abrogated detection of FLAG-TPP1-L protein (Figure 3.9D; lanes 4, 6, & 8). sunRNA-1 (121 nt; all sunRNA lengths exclude polyA and assume shared 3' end with reported TPP1 mRNA), sunRNA-2 (139 nt), sunRNA-3 (196 nt), and sunRNA-4 (247 nt) were all able to shutdown TPP1-L expression, suggesting that the 3'-UTR sequence common to them is important for the silencing (Figure 3.9D; Figure 3.10B). Indeed, the 93 nt TPP1 3'-UTR RNA common to all sunRNAs was sufficient to silence TPP1-L (Figure 3.9D; lane 2). Interestingly, the two longest sunRNAs, sunRNA-5 (495 nt) and sunRNA-6 (701 nt), failed to silence TPP1-L, implying an upper limit on sunRNA length for effective silencing (Figure 3.10B). To demonstrate that silencing of TPP1-L is specific to the sunRNA, we designed plasmids encoding FLAG-TPP1-L and the reverse complement (RC) of each sunRNA, so as to conserve the RNA length, GC content, and secondary structure, but not sequence. These control constructs resulted in an accumulation of TPP1-L protein similar to that of FLAG-TPP1-L/Empty (Figure 3.9D; lanes 5, 7, & 9), suggesting TPP1-L silencing by the sunRNA is sequence specific. CAGE tags at the 3' end of the *ACD* gene point to the sunRNA transcript being separate from the mRNAs encoding the TPP1 isoforms. Accordingly, TPP1-L was silenced when its 3'UTR was co-expressed in *trans* (#3 in Figure 3.9E) but not when the 3'UTR was part of the TPP1-L mRNA (#2; Figure 3.9E).

Taken together, these data suggest that TPP1-L is silenced by a ncRNA that is derived from its own 3'-UTR.

We verified that TPP1-L protein expression levels correlated with mRNA abundance by Northern blot analysis, which depicted a clear loss of TPP1-L mRNA in the presence of the sunRNA, but not RC (Figure 3.9F). Robust silencing was also recapitulated when TPP1-L and the sunRNA were expressed from separate dox-inducible plasmids (Figure 3.10C). Finally, the silencing is TPP1 mRNA-specific, as the 3'-UTR of TPP1 is unable to silence other shelterin components such as TIN2 or POT1 (Figure 3.10D).

Intriguingly, the sunRNA contains a stretch of 11 nt that are fully complementary to a region at the 5' end of the TPP1-L mRNA, but absent in the TPP1-S mRNA (Figure 3.9A). To test the role of this selective complementarity in the silencing of TPP1-L versus TPP1-S, we co-expressed TPP1-S with the sunRNA. In sharp contrast to the complete silencing of TPP1-L, the sunRNA failed to diminish the levels of TPP1-S mRNA or protein (Figure 3.11A,B). To better understand if the complementarity between the sunRNA and the TPP1-L mRNA was responsible for this TPP1-L specific silencing, we mutated the 11 complementary nt in the sunRNA (MutG; Figure 3.9A). Indeed these mutations completely abrogated the sunRNA's ability to silence TPP1-L (Figure 3.11C), confirming the importance of this region for sunRNA silencing. Notably, compensatory mutations in TPP1-L could not reinstate silencing by MutG sunRNA, suggesting that mere complementarity is not sufficient for silencing (Figure 3.10E). To further demonstrate specificity of sunRNA for TPP1-L versus TPP1-S, we engineered 5'

deletions in TPP1-L. All deletions tested, including deletion of the 5' terminal 30 nt of the TPP1-L ORF, decreased the sensitivity of silencing by sunRNA (Figure 3.10F).

Because mouse TPP1 begins at a Met equivalent to human Met87 (Figure 3.1B), it lacks the first target site for the sunRNA (red in Figure 3.12A). However mouse TPP1 also lacks 29 aa that are present in both human TPP1-S and TPP1-L (aa 334-362) but which lie outside TPP1's TERT, POT1, and TIN2 binding regions. Most remarkably, the mRNA that codes for these residues harbors a putative second target site of complementarity for the sunRNA (TPP1-L mRNA nt 1386-1403 shown in black in Figure 3.9A and red in Figure 3.11D, coded aa are shown in red in Figure 3.12A). To test the importance of this potential target site, we created a series of deletion mutations within the sunRNA. Notably, only sunRNA constructs with deletions that significantly disrupted complementarity (i.e. 3 bp) lost silencing capacity and rescued TPP1-L expression (deletions marked in red in Figure 3.9A and Figure 3.11E). In contrast, sunRNA constructs with deletions disrupting less than 3 bp of complementarity continued to silence TPP1-L expression (deletions marked in green in Figure 3.9A; Figure 3.11E). Furthermore, neither the human TPP1 sunRNA nor the 3'-UTR of mouse TPP1 is able to knockdown expression of mouse TPP1, confirming that this phenomenon is absent in mice (bottom image of Figure 3.11F: compare lanes 2 and 3; compare lanes 5 and 6).

CAGE data imply that the TPP1 sunRNA is 5'-capped and 3'-polyadenylated, suggesting that it is an RNA polymerase II transcript (Figure 3.7A; Figure 3.12B). Interestingly it is enriched in the nucleus, which is therefore the likely location of TPP1-L silencing. To further characterize the pathway through which the sunRNA inhibits TPP1-L mRNA, we knocked down key proteins in the RNA interference pathway.

Downregulation of *AGO2* and *DICER1* did not disrupt the silencing function, suggesting that the TPP1 sunRNA is not a cryptic microRNA (Figure 3.11G,H). Similarly, knockdown of essential components of the nuclear or cytosolic RNA exosomes (involved in 3' RNA degradation), decapping enzyme DCP2 (involved in 5' mRNA degradation), XRN1 or XRN2 ribonucleases, or components of Lsm complexes (involved in splicing and RNA turnover) were unable to perturb TPP1-L silencing by the sunRNA (Figure 3.12C-H). The lack of CAGE reads for sunRNA in the polysome enriched fraction precludes sunRNA-derived peptides as a driving force of TPP1-L silencing (Figure 3.12B; K562 data). Moreover, our observation that the short (93 nt) 3'-UTR of TPP1 is sufficient to silence TPP1-L (Figure 3.9D) suggests that peptides, if any, encoded from the TPP1 ORF regions present in longer sunRNAs are not necessary for silencing. Together our data reveal a phenomenon whereby a specific protein-coding mRNA isoform is silenced by an intragenic ncRNA that we accordingly named silencing 3'-UTR-derived noncoding RNA or sunRNA. However the precise molecular mechanism by which sunRNA acts remains to be determined.

3.3.7 TPP1-L expression switches on and TERT expression continues throughout spermatogenesis

TPP1-L is the minor TPP1 isoform in all somatic cell lines, embryonic stem cell lines, and cancer cell lines analyzed here. Yet the presence of an actively firing promoter for TPP1-L and the unique ability for TPP1-L to recruit but not efficiently activate telomerase, suggests that it is unlikely to be a vestigial protein. Inspection of archived RNA-seq data revealed that although TPP1-S expression dominated in most human tissues/organs, several reads specific to the unique sequence of TPP1-L mRNA

were detectible in testes (Figure 3.13A) (Burge lab data visualized on UCSC genome browser; (Wang et al., 2008)). Genotype-Tissue Expression (GTEx) data for TPP1 confirmed that TPP1-L mRNA is the major isoform in testes, expressing at ~10-fold over TPP1-S (Figure 3.14A; see Methods). The production of sperm during spermatogenesis begins with the spermatagonial stem cells that reside along the basement membrane of the seminiferous tubules of the testes. These specialized stem cells undergo differentiation into spermatocytes, then complete two rounds of meiosis to form haploid spermatids which ultimately mature into spermatozoa after exiting via the lumen (Fayomi and Orwig, 2018) (Figure 3.13B). Analysis of RNA-Seq data for human testis-specific cells revealed very few TPP1-L specific reads in spermatocytes (Lesch et al., 2016) (Figure 3.13C). Remarkably, the TPP1-L specific reads were greatly enriched in round spermatids, with peak heights comparable to other exons common to both TPP1-L and TPP1-S mRNA. These data suggest that the boost in TPP1-L expression in testes occurs after differentiation and completion of meiosis. Additionally, the only detectible TPP1 protein band in the chromatin enriched fraction of testes lysate corresponds to the length of TPP1-L, providing evidence for TPP1-L protein in testes (Figure 3.13D). Together our analysis suggests that TPP1-S is the predominant isoform in the early stages of spermatogenesis, while TPP1-L dominates as these germ cells mature.

To understand how TPP1-L may affect telomerase function in testes, we investigated TERT status in this organ. Because spermatogonial stem cells self-renew, it is not surprising that they are telomerase-positive (Izadyar et al., 2011). However spermatogonial stem cells differentiate and undergo extensive maturation as well

(Figure 3.13B). Unlike differentiated somatic cells, differentiated germ cells must not only block excessive telomerase action but they must also be able to swiftly re-activate telomerase following fertilization, during embryogenesis (Wright et al., 1996). To ask if differentiated human germ cells retain telomerase expression, we performed RNAscope *in situ* hybridization to detect TERT expression in human testes tissue using probes that were previously validated for TERT detection in human cells (for controls, see Figure 3.14B) (Ravindranathan et al., 2018). As expected, robust TERT signal was detected in cells lining the basement membrane of the seminiferous tubules (indicative of spermatogonial stem cells), but was absent in non-germ cells of the testes such as myoid cells (Figure 3.13E&F; Figure 3.14C). Intriguingly, TERT signal was also consistently detected within the adluminal compartment, in the differentiating germ cells located between the spermatagonial stem cell layer and the lumen. On average fifteen spermatogonia were positive for TERT signal per tubule, while an additional eight differentiating germ cells (either meiotic or post-meiotic) were positive for TERT. In contrast, the average number of myoid cells that were positive for TERT signal per tubule was less than one (Figure 3.13F). Together these data reveal continued TERT expression in differentiating germ cells and invoke a plausible function for TPP1-L over TPP1-S expression in testes: TPP1-L could block telomerase action in these cells until TPP1-S expression resumes after fertilization.

3.4 Discussion

Our curiosity for the identities and functions of human TPP1 isoforms led to the discovery of two isoforms distinguished by their ability to activate telomere synthesis by telomerase and their distinct representation in somatic cells versus differentiated cells of

the germline. We uncovered a regulatory RNA, which we call sunRNA, that dictates TPP1 isoform expression. Finally, our observation of an isoform expression switch in TPP1 unveils a new function for TPP1 specific to spermatogenesis, prompting the investigation of alternative isoforms of other telomeric and non-telomeric proteins in this unique stage of reproductive biology.

Our telomerase recruitment analysis, along with previously published reports, demonstrates clearly that TPP1-L is proficient at recruiting telomerase to telomeres. This suggests that both the TEL patch and NOB in TPP1-L are able to fully engage their target sites on TERT. However, telomerase recruited to the telomere by TPP1-L cannot efficiently extend chromosome ends. We therefore characterize TPP1-L as a telomerase activation-incompatible isoform of TPP1 that is distinct from all previously characterized TEL patch and NOB mutants of TPP1-S.

The separation-of-function for TPP1-S and TPP1-L isoforms implies that they have unique roles in distinct biological contexts. We observed that TPP1-S is the main isoform in all somatic cells and in embryonic stem cells. The presence of TPP1-S in stem cells is fully consistent with their need to support telomerase activity for continued self-renewal. After differentiation, cells use a robust mechanism to switch off telomerase expression (Aubert, 2014; Günes and Rudolph, 2013), rendering downregulation of TPP1-S (or upregulation of TPP1-L) unnecessary. Thus it is not surprising that TPP1-S is the major isoform in somatic cells, both before and after differentiation. We observed that TPP1-L is specifically upregulated in maturing germ cells in the testes, which intriguingly retain TERT expression. We propose that TPP1-L in these cells will block telomerase action keeping telomere length in check throughout

the remainder of spermatogenesis. Upon fertilization, TPP1-S expression would resume, ensuring end replication (and protection) during embryogenesis and throughout life thereafter. In contrast to humans, mice express telomerase in most cells, making a check on telomere elongation in differentiated germ cells unnecessary (Prowse and Greider, 1995). Thus it is not surprising that TPP1-L is absent in mice, however other functions of TPP1-L in the human germline cannot be discounted. The arginine-glycine richness of TPP1-L aa 1-86 is reminiscent of RGG/RG-domain proteins involved in RNA binding and phase separation (Chong et al., 2018). Future studies will determine whether TPP1-L exhibits such attributes and how these attributes might impact TPP1-L function.

TPP1-L mRNA is selectively diminished by an intragenic RNA that we call sunRNA. The sunRNA appears to be transcribed by RNA polymerase II and accumulates in the nucleus. It is known that RNA polymerase II transcript length is used to sort RNAs for export into the cytoplasm such that only transcripts shorter than 200-300 nt (e.g., snRNAs) are retained in the nucleus (McCloskey et al., 2012). Our observation that only sunRNAs shorter than ~250 nt are effective in silencing TPP1-L suggests that longer sunRNAs are exported into the cytoplasm. Finally, based on its origin from the 3' end of the TPP1 gene, the sunRNA can be placed in two known ncRNA categories: "termini-associated sRNA" or TASR (Affymetrix and Cold Spring Harbor Laboratory, 2009; Kapranov et al., 2007), and 3' UTR-associated RNA or uaRNA (Mercer et al., 2011).

To our knowledge, this is the first report of differential silencing of protein isoforms by an intragenic ncRNA. We show that two essential sequence elements in

the TPP1 sunRNA are complementarity to two separate target regions in the TPP1-L mRNA sequence. The upstream target region is present only in TPP1-L, and absent in TPP1-S, and appears to be the basis for the isoform specificity. The downstream target region, shared by both human TPP1-S and TPP1-L, codes for a stretch of seemingly unstructured amino acids with no known function. Both of these sunRNA-recognition regions are conserved among primates but absent in the mouse TPP1 mRNA, suggesting that they play an important regulatory function at the mRNA level. However the failure of compensatory mutations in TPP1-L to rescue silencing by mutations in sunRNA and the dispensability of several known RNA silencing/processing machineries for this silencing phenomenon suggest that sunRNA acts through a mechanism that awaits further characterization. We propose a model whereby TPP1-L mRNA is primed in all human cells, but it is shut down by the TPP1 sunRNA in somatic cells, allowing TPP1-S to emerge as the major isoform. The switch that regulates sunRNA levels during spermatogenesis remains to be established, pending further experimentation including CAGE analysis in cells of the testes.

There are several conceivable advantages of sunRNA-like intragenic regulatory mechanisms in complex genomes. (1) The *cis* nature of the transcription start sites of the silencer RNA and the target mRNA could provide for an easier “search” for the complementary sequence due to enhanced local concentrations of the two RNAs. This is particularly relevant in complex mammalian transcriptomes where RNA-RNA recognition in *trans* is entropically disfavored without involvement of specialized machineries such as in RNAi. It is interesting to note that in less complex genomes such as those in bacteria, ncRNAs called sRNAs expressed from the 3’-UTR of protein-

coding genes, can silence mRNA targets in *trans* using sequence-complementarity driven mechanisms (Kunne et al., 2014; Miyakoshi et al., 2015; Waters and Storz, 2009). It is therefore tempting to speculate that sRNA-mediated mechanisms were largely lost (and replaced by RNAi) when more complex transcriptomes evolved, and only cis-sRNA (like sunRNA) mediated mechanisms were conserved due to proximity between the silencer and the target. (2) sunRNAs can utilize the mRNA polyadenylation signal for their own 3' end maturation thus obviating the need for further 3' maturation. (3) The selection pressure to conserve essential mRNA sequences will also help preserve the corresponding sunRNA sequence. Therefore, we hypothesize that the overlapping nature of mRNA and sunRNA has prevented identification of other sunRNAs, and many such ncRNAs and their associated biological functions await discovery.

3.5 Materials and Methods

3.5.1 Molecular Cloning and site-directed mutagenesis

3X-FLAG-tagged constructs and 6x-Myc-tagged constructs for human cell expression were cloned into the pTET-IRES-eGFP-Bi4 vector; and 3X-FLAG-TPP1-L and 3X-FLAG-TPP1-L-M87A constructs were cloned into p3X-FLAG-TPP1-F3 vectors as previously described (Grill et al., 2018; Nandakumar et al., 2012). Additionally, 3X-FLAG-TPP1-S and 3X-FLAG-TPP1-L M87A constructs were cloned into a pcDNA3 derived vector and used to generate the M87A R_3E_3 construct for transient transfection in HeLa cells. All TPP1-S constructs used for human cell expression contained cDNA

sequences corresponding to aa 87-544 of the TPP1 protein, while all TPP1-L constructs contained cDNA sequences corresponding to aa 1-544. The M87A mutation in the TPP1-L expression plasmid was introduced using the QuikChange® Site-Directed Mutagenesis Kit (Agilent Technologies) and complementary mutagenic primers (Integrated DNA Technologies). The resulting TPP1-M87A plasmid was sequenced to confirm both the presence of the intended mutation and the absence of unwanted errors introduced during PCR amplification/cloning. All additional site-directed mutagenesis, including mutation of TPP1-L M87A to M87A R_3E_3, and mutation of TR to mutant TR was performed using the same protocol as above. For bacterial expression of $TPP1^{90-334}$ and $TPP1^{1-334}$, the appropriate TPP1 sequence was cloned into the pET-Smt3 vector using standard restriction endonuclease cloning as described previously (Grill et al., 2018). For insect cell expression and purification of human POT1 the pFBHTb-Smt3star-hPOT1 was used as described previously (Kocak et al., 2014).

3.5.2 Cell culture

Cell culture was performed as described previously (Bisht et al., 2016; Grill et al., 2018). HeLa-EM2-11ht cells were used in all transient transfection, while both HeLa-EM2-11ht and HEK 293T cells were used in CRISPR-Cas9 experiments. All cells were cultured at 37°C in the presence of 5% CO2 and propagated in growth medium containing modified DMEM (Gibco; Dulbecco's Modified Eagle Medium; 11995-065), 100 U/mL penicillin, 100 µg/mL streptomycin, and 10% FBS. For all experiments requiring induction, doxycycline was added to a final concentration of 200 ng/mL to drive a tetracycline-inducible promoter within the p6X-Myc-BI4 or p6X-FLAG-BI4 plasmids (FLAG-TPP1-S, FLAG -TPP1-L, FLAG -TPP1-L M87A, ΔNOB).

3.5.3 Generation of stable cell lines expressing TPP1-L and TPP1-L M87A

Stable cell line generation was performed exactly as described previously (Grill et al., 2018). Briefly, HeLa-EM2-11ht cells were co-transfected with 1 µg each of the p3X-FLAG-TPP1-F3 (TPP1-L or TPP1-M87A) and Flp recombinase-expressing plasmid (that also codes for puromycin resistance). Cells were selected for one day using puromycin (5 µg/ml; Sigma-Aldrich), and then fresh medium with ganciclovir (50 µM; Sigma-Aldrich) was added for 10 days of negative selection. Individual clones were picked and expanded, and positive clones were selected based on Western blot analysis of FLAG-TPP1 signal after overnight induction with doxycycline (200 ng/ml). Generation of the stable cell lines expressing vector, TPP1-S, and TPP1ΔNOB constructs used in this study have been described previously (Grill et al., 2018; Nandakumar et al., 2012)

3.5.4 Telomere restriction fragment length analysis

Telomere length analysis was performed as described previously (Grill et al., 2018; Nandakumar et al., 2012). Briefly, genomic DNA from stable cell lines expressing FLAG-TPP1 constructs (TPP1-S/TPP1-L/TPP1-M87A) and a vector control was isolated from confluent 6 cm dishes using the GenElute kit (Sigma). 2 µg of DNA was digested with *Hinf*1 and *Rsa*1 and incubated overnight at 37°C. The digested DNA was run on a 0.8% 25 cm long Agarose-1X TBE gel along with a Lambda DNA-*Hind*III digest ladder (NEB) at a constant 50 V for 20-23 h. After imaging with a florescent ruler, the gel was transferred to a sheet of dry Whatman filter paper and dried at 55°C for one h. After drying, the filter paper was removed and the gel was soaked in buffer containing 0.5 M NaOH for 30 min, rinsed with water, and shaken in a solution of 0.5 M Tris-Cl and 1.5 M NaCl (pH 7.5) for 30 min. The gel was then prehybridized in Church buffer (0.5 M

sodium phosphate buffer [pH 7.2], 1% bovine serum albumin [BSA], 1 mM EDTA, 7% SDS) for 30 min at 65°C in a rotating hybridization oven. A telomeric probe of sequence $(TTAGGG)_4$ was 5' ^{32}P-labeled with T4 PNK (NEB) and added at 20 million cpm to the gel. Hybridization continued overnight at 55°C. After overnight hybridization the gel was washed three times with 2X SSC for 10 min and exposed to a phophorimager screen for 24-72 h. The gel was analyzed using the Imagequant TL software and calibrated using the molecular weights of the lambda DNA ladder. The mean telomere length for each lane was plotted as a function of days in culture for each cell line. A linear regression (MS Excel) was used to calculate the rate of telomere elongation/shortening.

3.5.5 Immunofluorescence-fluorescence *in situ* hybridization

3.5.5.1 TIF analysis (co-IF)

Co-IF experiments for TIF analysis were adapted from protocols described previously (Nandakumar et al., 2012). Briefly, co-IF was performed using stable cell lines expressing either FLAG-TPP1-S or FLAG-TPP1-L M87A. Cells were induced with doxycycline for 3 days and ~100,000 cells were seeded on coverslips in growth medium containing doxycycline. 24 hours after seeding, medium was removed, cells were washed with PBS, and all subsequent steps were performed at room temperature. Cells were fixed with 4% formaldehyde in PBS for 10 min and then washed in PBS before permeabilization with 0.5% Triton-X 100. Cells were blocked (PBS-T containing 3% FBS) for 30-60 minutes and incubated with mouse monoclonal anti-FLAG M2 antibody (Sigma; F1804; 1:500) and rabbit polyclonal anti-53BP1 antibody (Novus Biologicals; NB100-304; 1:1,000) diluted in blocking buffer. After 60 min of incubation, cells were washed with PBS and incubated with Alexa Fluor 568-conjugated anti-mouse

IgG (Life Technologies) and Alexa Fluor 647-conjugated anti-rabbit IgG (Life Technologies) diluted 1:500 in blocking buffer for 30 minutes. Cells were mounted on microscope slides using ProLong Gold mounting medium with DAPI (Life Technologies). Coverslips were sealed with transparent nail polish and stored at 4°C until the time of imaging. All imaging was performed using a laser scanning confocal microscope (SP5; Leica, Germany) equipped with a 100X oil objective. ImageJ and Adobe Photoshop were used to process all images. Colocalizations of FLAG-TPP1 foci and 53BP1 foci were quantified manually by two separate individuals.

3.5.5.2 Telomerase recruitment to telomeres (IF-FISH)

IF-FISH experiments for telomerase recruitment were performed exactly as described previously (Grill et al., 2018). Briefly, IF was performed prior to FISH using mouse monoclonal anti-FLAG M2 (Sigma; F1804; 1:500) in combination with Alexa Fluor 568-conjugated anti-mouse IgG (Life Technologies) to visualize FLAG TPP1 proteins (FLAG -TPP1-S, FLAG -TPP1-L, FLAG -TPP1-M87A) by IF. A mixture of Cy5-conjugated probes complementary to TR was used to detect TR by FISH. Imaging was performed as described above. Colocalizations were quantified manually by two separate individuals.

3.5.5.3 Telomere localization of TPP1-S and TPP1-L M87A (IF-FISH)

Telomere localization of all TPP1 constructs was performed exactly as previously described (Grill et al., 2018). Briefly, a Cy3-conjugated PNA-$(CCCTAA)_3$ probe was used to visualize telomeres by FISH. Cells were fixed, permeabilized, and soaked in 2X SSC, 50% formamide before being heated for 6 min at 80°C in the presence of the

probe. After 2 h of hybridization cells were washed and subjected to immunofluorescence for FLAG-TPP1. Imaging was performed as described above.

3.5.5.4 TERRA localization at telomeres

This protocol was adapted from TERRA-FISH protocols previously published in the literature (Azzalin et al., 2007). 48 h after doxycycline induction, FLAG-TPP1-S and FLAG-TPP1-L M87A stable cell lines were washed with PBS and permeabilized in CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM MgCl2, 10 mM HEPES (pH 7), 0.5% Triton X-100, and 10mM VRC) for 7 min. Cells were washed in PBS and fixed in 4% formaldehyde for 10 min followed by consecutive dehydration with 70%, 85%, and 100% ethanol. Cells were then rehydrated in 2X SSC 50% formamide before hybridization with a Cy3-conjugated PNA-$(CCCTAA)_3$ probe for 16 h. After hybridization, cells were washed with 2X SSC 50% formamide and directly used for immunofluorescence to visualize FLAG-TPP1 constructs. Imaging was performed as described above.

3.5.5.5 Detection of nascent telomeric synthesis using mutant TR (IF-coFISH)

For detection of newly synthesized mutant telomeric DNA in HeLa, FLAG-TPP1-S, and FLAG-TPP1-L M87A stable cell lines, cells were first induced with doxycycline for 48 h prior to transfection with mutant telomerase. After 48 h induction, 1 µg of pTERT-cDNA6/Myc-His C and 3 µg of phTRmut-Bluescript II SK (+) plasmids were transfected into approximately 1 million cells with lipofectamine LTX (Fisher; 15338100) using the manufacturer recommended protocol. 20-24 h post transfection, immunofluorescence was performed exactly as described above to visualize FLAG-TPP1 constructs. Subsequently cells were analyzed by FISH to detect mutant TR and

newly synthesized mutant telomeres. Cells were fixed with 4% formaldehyde for 10 min at room temperature, washed with PBS, and consecutively dehydrated with 70%, 95%, and 100% ethanol. Cells were then rehydrated with 2X SSC 50% formamide and pre-hybridized for 1 hour using hybridization solution containing 100 mg/ml dextran sulfate, 0.125 mg/ml yeast tRNA, 1 mg/ml BSA, 0.5 mg/ml salmon sperm DNA, 1 mM vanadyl ribonucleoside complexes (VRC), and 50% formamide in 2X SSC. After pre-hybridization cells were transferred to hybridization solution containing a mixture of three Cy5-conjugated probes against TR (described previously) and a Cy3-conjugated PNA-$(CCGCAA)_3$ probe. Cells were heat denatured at 80°C for 5 min then incubated at 37°C for ~16 h. After incubation cells were washed with 2X SSC 50% formamide and mounted on microscope slides using ProLong Gold mounting medium with DAPI (Life Technologies). Detection of newly synthesized mutant telomeres in HeLa cells transiently transfected with FLAG-TPP1-L M87A or FLAG-TPP1-L M87A R_3E_3 was performed similarly 40 h after transfection with 1 μg FLAG-TPP1 construct, 1 μg of pTERT-cDNA6/Myc-His C plasmid, and 3 μg of phTRmut-Bluescript II SK (+) plasmid. HeLa cells with no exogenous TPP1 expression were analyzed both 20 h and 40 h after transfection with the TERT and mutant TR plasmids. All imaging was performed using a laser scanning confocal microscope (SP5; Leica, Germany) equipped with a 100X oil objective. ImageJ and Adobe Photoshop were used to process all images. Colocalization between TR and mutant telomere signal was quantified manually by two separate individuals. Mutant telomere signal was only scored when it was detectible above the background signal (lower pixel intensity threshold set at 30, upper pixel

intensity threshold set at 255). Statistical significance was scored using a two-tailed student's t test.

3.5.6 RNAscope

Normal testes tissue was cut in to 5 μm thick sections and put in 10% buffered formalin for 1-7 d before embedding. After embedding following standard procedures, FFPE testes sections were processed and analyzed for RNAscope according to manufacturer's protocol. Briefly, target retrieval was performed for 15 min at 95-99°C in a steamer. Slides were then dehydrated and dried before a 15 min protease treatment. Probe hybridization was performed using DapB negative control probes (ACDBio; 310043), PPIB positive control probes (ACDBio; 313901), or catalog probes for TERT (ACDBBio; 481961), followed by staining with 50% hematoxylin, drying, clearing with xylene, and mounting exactly as described in the manufacturer's protocol. Slides were dried overnight and imaged on a light microscope at 40X or 63X magnification.

3.5.7 Co-immunoprecipitation assay for protein-protein interaction

Co-immunoprecipitation experiments were performed as described previously (Grill et al., 2018). Briefly, HeLa-EM2-11ht cells were transfected with 1 μg each of plasmids containing Flag-TPP1-S, Flag-TPP1-M87A, Myc-POT1, FLAG-POT1, or Myc-TIN2. 48 h after transfection/ induction with doxycycline, cells were washed with PBS, trypsinized, and dislodged with medium containing 50% FBS. Cells were then pelleted, resuspended in 400 μL of lysis buffer (50 mM Tris-Cl (pH 7.6), 20% glycerol, 1 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.02% SDS, 1 mM dithiothreitol, 2 mM phenylmethylsulfonyl fluoride, complete protease inhibitor cocktail [Roche]) and kept on ice. 33 μL of 4 M NaCl, and 433 μL of water was added and lysates were centrifuged at

13,600 rpm for 10 min. After centrifugation, 40 µL of lysate was added to SDS gel loading buffer and kept aside for analysis of input samples. Remaining supernatant was used directly for immunoprecipitation with 30 µL of pre-washed anti-FLAG M2 affinity gel (Sigma; A2220). After overnight incubation, beads were washed and protein was eluted from the beads by adding 60 µL of 2X SDS gel loading buffer. All samples were analyzed by SDS-PAGE followed by immunoblotting with HRP conjugated FLAG or Myc antibodies.

3.5.8 Immunoblotting

Immunoblotting was performed using standard procedures with the following antibodies: mouse monoclonal anti-FLAG M2-HRP conjugate (Sigma; A8592; 1:10,000), mouse monoclonal anti-c-Myc (9E10) HRP conjugate (Santa Cruz; sc-40 HRP; 1:10,000), mouse monoclonal anti-β-actin antibody (Sigma; A5441; 1:10,000), and rabbit polyclonal TPP1 antibody (Bethyl Laboratories; A303-069A; 1:2,500). Secondary horseradish peroxidase-conjugated goat antibodies against rabbit or mouse IgG (Santa Cruz Biotechnology; 1:10,000) were used for detection with chemiluminescence by ECL plus reagents (Pierce ECL Western Blotting Substrate; Thermo Scientific). All antibodies used to assess knockdown of components in RNA decay and processing events were a kind gift from Dr. Peter Baumann. The rat monoclonal antibody against Ago2 was a kind gift from Dr. Gunter Meister. The rabbit polyclonal anti-Dicer antibody was a kind gift from Dr. Nils Walter. The data were visualized and quantified using a gel-documentation system (ChemiDoc™ MP System; BioRad).

3.5.9 Testes tissue processing and immunoblotting

Normal testes tissue was obtained from the Tissue Core of the University of Michigan Rogel Cancer Center. Testes tissue used for immunoblotting was homogenized in cell extraction buffer (Invitrogen; FNN011) supplemented with PMSF and protease inhibitor cocktail (Sigma 11836170001 Roche). Homogenized tissue was incubated on ice for 1 h followed by centrifugation at 4°C and 13,000 rpm for 15 min. Soluble lysate was removed and the resulting pellet was resuspended in buffer containing 5 M urea, 2 mM MgCl2, 4% SDS, 140 mM Tris-Cl (pH 8), and 50 mM DTT. The resulting chromatin fraction was boiled with 5X SDS loading dye and analyzed by SDS-PAGE and immunoblotting with rabbit polyclonal TPP1 antibody following standard procedures.

3.5.10 Enrichment of endogenous TPP1 using POT1-DNA pulldown.

Immunoprecipitation experiments to detect endogenous TPP1 protein were performed by first incubating 10 µg of purified POT1 protein with a biotinylated telomeric DNA primer (5'-biotin-ACGTA$(GGTTAG)_3$) bound to 20 µl of streptavidin beads (Thermo; Pierce; 20357) for 2 h on ice. During this incubation, approximately 4 million HeLa, HEK 293T, and BJ fibroblast cells (kind gift from Dr. Ursula Jakob) were lysed in 50 mM Tris-Cl (pH 7.6), 20% glycerol, 1 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.02% SDS, and 1 mM DTT and centrifuged at 4°C and 13,000 rpm for 10 min. After 2 h, the soluble lysates were added to the POT1-DNA-bead slurry and incubated for ~16 h at 4°C. Beads were then washed and proteins were eluted with 2X SDS loading buffer. Samples were analyzed by SDS-PAGE and immunoblotting with TPP1 antibody.

3.5.11 CRISPR-Cas9 mediated insertion of a C-terminal FLAG tag at the endogenous locus of human TPP1

Insertion of the FLAG tag into the HEK 293T genome was performed using a strategy reported previously for introducing the K170Δ into the HEK 293T genome (Bisht et al., 2016). The design of the ssODN is shown in Figure 3.8B and involves insertion of a 3X-FLAG coding DNA sequence in-frame and immediately upstream of the natural stop codon of TPP1-L and TPP1-S. Single clones were analyzed using FLAG-immunoblotting.

3.5.12 CRISPR-Cas9 mediated TPP1-L KO generation

Generation of TPP1 knockout clones was performed using a strategy previously reported (Bisht et al., 2016). A guide RNA was designed with the Zhang Lab open access CRISPR design tool. The ssODN (described in Figure 3.7E) was designed to encompass the desired mutations as well as a silent *Aat*II restriction endonuclease cleavage site. A surveyor nuclease assay (Surveyor® Mutation Detection Kit; catalog #706025; Transgenomic) was used to determine cleavage efficiency 48 h after transfection of 1 μg of guide RNA and Cas9 encoding plasmids into HeLa-EM2-11ht cells following the manufacturer's protocol. Upon completion of the reaction, products were visualized using ethidium bromide-stained agarose gels and robust cleavage was observed. HeLa-EM2-11ht cells were then transfected with guide RNA and Cas9 encoding plasmids along with indicated ssODN as previously described. 3 d after transfection 2,000 cells were plated for colony formation while the rest were used to test the efficiency of repair. To check repair efficiency genomic DNA was prepared using GenElute Mammalian Genomic DNA Purification Kit (#G1N70; Sigma) and the locus

encompassing our desired mutations was PCR amplified. Amplified DNA was then digested in the presence or absence of *Aat*II and reaction products were visualized on an ethidium bromide stained agarose gel. ~2 weeks after seeding, colonies were picked and transferred to a 96 well plate where they remained for 1 week. They were then trypsinized and half the cells were transferred to a replicate plate for genomic DNA preparation. Isolated genomic DNA from each clone within the 96 well plate was PCR amplified and subjected to digestion by *Aat*II. Positive clones were expanded and genotyped using Sanger DNA sequencing. Selected clones were then propagated for telomere length analysis.

3.5.13 5’ RACE to clone sunRNA

5' RACE was performed using the 5’ RACE kit from Life Technologies (#18374058). First strand cDNA synthesis was performed with SuperScript™ II RT using 2 µg total RNA from HEK 293T cells, a gene-specific antisense oligonucleotide GSP1 (TCCCTGATCCTCTCCTCTCC) and abridged anchor primer, AAP (GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG) which contains a complementary homopolymeric tail annealing sequence and allows amplification from a homopolymeric tail. Following cDNA synthesis, the first strand product was treated with RNAse and purified from unincorporated dNTPs and GSP1 using S.N.A.P. columns provided with the kit. 10 µl of S.N.A.P. column purified cDNA was tailed at the 3’ end with CTP using TdT (Terminal deoxynucleotidyl transferase) enzyme. After heat inactivation of TdT, 5 µl of tailed cDNA was amplified by PCR for 30 to 35 cycles using a nested gene-specific primer GSP2 (TCCTCTCCTCTCCTGCCGC) which anneals 3' to GSP1 and a complementary homopolymer-containing abridged anchor

primer (AAP) which permits amplification from the homopolymeric tail. 0.1% of the PCR reaction from this step was subjected to re-amplification using a nested gene specific primer GSP3 (GGAAGCAGAGTGTGGAGCGG) and abridged universal amplification primer AUAP (GGCCACGCGTCGACTAGTAC). The PCR products were resolved on a 2% GTG agarose gel and discrete bands were gel eluted and subjected to TA cloning and sequencing. Sequences showing a homopolymer C tail after sequencing were considered positive. All sequenced cDNA were positive for the 3'-UTR sequence of TPP1, but contained different extents of upstream exonic sequences.

3.5.14 Cloning, expression, and silencing experiments with sunRNA

Protein-coding inserts contained cDNA for only the ORF cloned into multiple cloning site 1 of the bi-directional dox-inducible pTet-BI4 vector using *Not*1/*Xho*1 restriction based cloning. FLAG-TPP1-L was cloned into pTet-Bi4 vector with a 3X-FLAG tag on the C-terminus and a 1X-FLAG tag on N-terminus of the TPP1-L protein and used in all sunRNA experiments requiring TPP1-L expression. The sunRNA sequence was similarly cloned into multiple cloning site 2 using *Spe*1/*Nhe*1 sites. Because *Spe*1 and *Nhe*1 generate compatible sticky ends, the same inserts, ligation reactions, and transformations were used to generate plasmid constructs with sunRNAs and cognate reverse complements (RC). SV40 polyA sequences downstream of the multiple cloning sites were used for proper RNA maturation of the cloned protein and sunRNA/RC cDNA constructs. Plasmids encoding TPP1-L (or other protein-coding genes used in the study) and Vector/sunRNA/RC cDNA were transfected into HeLa-EM2-11ht as already described. 24 h post-transfection, cells were lysed in SDS-gel

loading buffer and analyzed by immunoblotting (for detecting proteins) as already described, or by Northern blot analysis as described below.

3.5.15 Northern blotting

10 µg of total RNA was dissolved in buffer containing 5.7 µl DEPC-treated water, 1 µl 10 X MOPS, 3.3 µl formaldehyde, and 10 µl formamide and mixed well. The mixture was heated at 72°C for 5 min, followed by the addition of 5 µl of 6X DNA loading dye and transfer to ice. The agarose gel for running the RNA samples was prepared as follows. 126.75 ml water was added to 1.5 g agarose heated in a microwave to get a homogenous solution that was kept aside until the temperature reduced to approximately 55 °C. 10 ml 10X MOPS, 8.25 ml 37% formaldehyde, and 3 µl ethidium bromide were heated in a separate tube in a 65°C water bath. The two solutions were mixed together and cooled to form the gel matrix for electrophoresis. The RNA samples were loaded on the gel and run in 1X MOPS buffer at 50 V until the bromophenol blue front traversed approximately half the length of the gel. The gel was visualized under UV light to detect and capture an image of the bands for the 18S and 28S ribosomal RNA subunits. The gel was then transferred to Hybond nylon membrane using 10X SSC overnight. The protocol described for Southern blotting was applied for detecting the indicated RNA bands in the figures. A mixture of five probes, TPP1 ORF: 1R, 2R, 3R, 4R, and 5R, were used for detecting the TPP1-L or TPP1-S mRNA. All these probes were complementary to regions common to both TPP1-L and TPP1-S mRNA. Probes containing the TPP1 3'-UTR sequence and its reverse complement were used to detect RC and sunRNA, respectively.

3.5.16 siRNA knockdown experiments

HeLa-EM2-11ht cells were transfected with 80 nmol of the indicated siRNA using Oligofectamine (Life Technologies) in a 24-well culture plate. All siRNAs were a kind gift from Dr. Peter Baumann. 24 h after transfection, cells were trypsinized and split into three wells of a 12-well plate. 48 h post transfection, 500 ng plasmids encoding TPP1-L and vector/sunRNA/RC constructs were transfected using Lipofectamine 2000 (Life Technologies). 10 h after the second round of transfection, the media was changed to include doxycycline (200 ng/ml). Lysates were prepared 24 h post-induction and immunoblotting was performed as already described.

3.5.17 Protein purification

$TPP1^{90-334}$ and POT1 proteins were purified exactly as described previously (Grill et al., 2018; Nandakumar et al., 2012). $TPP1^{1-334}$ was purified in parallel with $TPP1^{90-334}$ using the same protocol.

3.5.18 DNA-binding experiments

Filter-binding assays with POT1-DNA, $TPP1^{90-334}$-POT1-DNA, and $TPP1^{1-334}$-POT1-DNA complexes were performed and analyzed exactly as previously described (Nandakumar et al., 2012; Nandakumar et al., 2010).

3.5.19 Direct telomerase activity assays

Telomerase assays were performed as described previously (Grill et al., 2018). Briefly, HEK 293T cells were transfected with 1 µg of pTERT-cDNA6/Myc-His C and 3 µg of phTR-Bluescript II SK (+) plasmids and soluble cell extracts were prepared as previously described(Kocak et al., 2014). This extract was used as the source of telomerase enzyme in direct telomerase primer extension assays based on previously

published protocols (Kocak et al., 2014). In each reaction, 2 µl of cell extract containing telomerase enzyme was incubated at 30°C for 10, 20, 30, or 60 min. Full reactions contained 50 mM Tris-Cl (pH 8.0), 1 mM MgCl2, 1 mM spermidine, 30 mM KCl, 5 mM β-mercaptoethanol, 1 µM of primer a5 (TTAGGGTTAGCGTTAGGG), 500 µM dATP, 500 µM dTTP, 2.92 µM unlabelled dGTP, 0.17 µM radiolabeled dGTP (3000 Ci/mmol). Purified POT1, TPP1$^{90-334}$, or TPP1$^{1-334}$ proteins were added to the reactions at indicated concentrations. 100 µl of buffer containing 3.6 M ammonium acetate and 20 µg of glycogen was used to quench each reaction, and the DNA products were precipitated with 70% ethanol. The pellets were resuspended in 7 µl water, then mixed with 7 µl loading buffer containing 95% formamide, and heated at 95°C for 5 min. Samples were resolved on a 10% acrylamide, 7M urea, 1X TBE sequencing-size gel and gels were dried then imaged on a phosphorimager (Storm; GE). All assays were analyzed using Imagequant TL (GE Life Sciences) software. Processivity calculations were performed as described previously (Grill et al., 2018).

3.5.20 Visualization of data on UCSC genome browser

Indicated tracks were loaded on to the UCSC genome browser (Casper et al., 2018) and saved as .ps files. The sessions were processed in Adobe Illustrator to furnish final figures. All UCSC genome browser associated data were previously deposited in databases by other groups and have been cited appropriately. Description of data displayed in Figure 3.7A: GENCODE reveals two major variations of transcripts differing only in their 5' definition. GRO-cap data for indicated cell lines show similar nascent transcription activity for TPP1-L and TPP1-S mRNA. Cumulative transcription factor ChIP data from a large number of human cell lines reveals two distinct peaks,

suggesting independent transcription of TPP1-L and TPP1-S mRNA from two dedicated promoters. Cumulative RNA seq from nine cell lines (GM12878, H1-human embryonic stem cell line, HeLa-S3, HepG2, HSMM, HUVEC, K562, NHEK, and NHLF) shows very few reads specific to TPP1-L compared to TPP1-S mRNA at steady-state. Note that the cumulative plot is fully representative of data from each individual cell line (see Figure 3.8A). CAGE data plotted in natural logarithm scale capturing capped RNA 5' ends is shown for the polyA enriched nuclear fraction of the indicated cell lines. Cumulative ribosome profiling (ribo-seq) data obtained for a large number of human cell lines spanning twelve independent studies is shown. The cumulative plot is fully representative of data from each individual cell line (see Figure 3.8A for data from an embryonic stem cell line).

3.5.21 GTEx data

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 05/10/18.

3.5.22 Testes-cell specific RNA-Seq analysis

Human pachytene spermatocyte and round spermatid RNA-seq data were obtained from NCBI Sequence Read Archive: SRP057141 (Lesch et al., 2016). RNA-seq quality was analyzed using fastQC. STAR version 2.52 was used to perform alignment of RNA-seq data to hg38 genome assembly, available from Ensembl87. Merged BAM files were visualized using Integrative Genomics Viewer version 2.3.

3.6 Figures

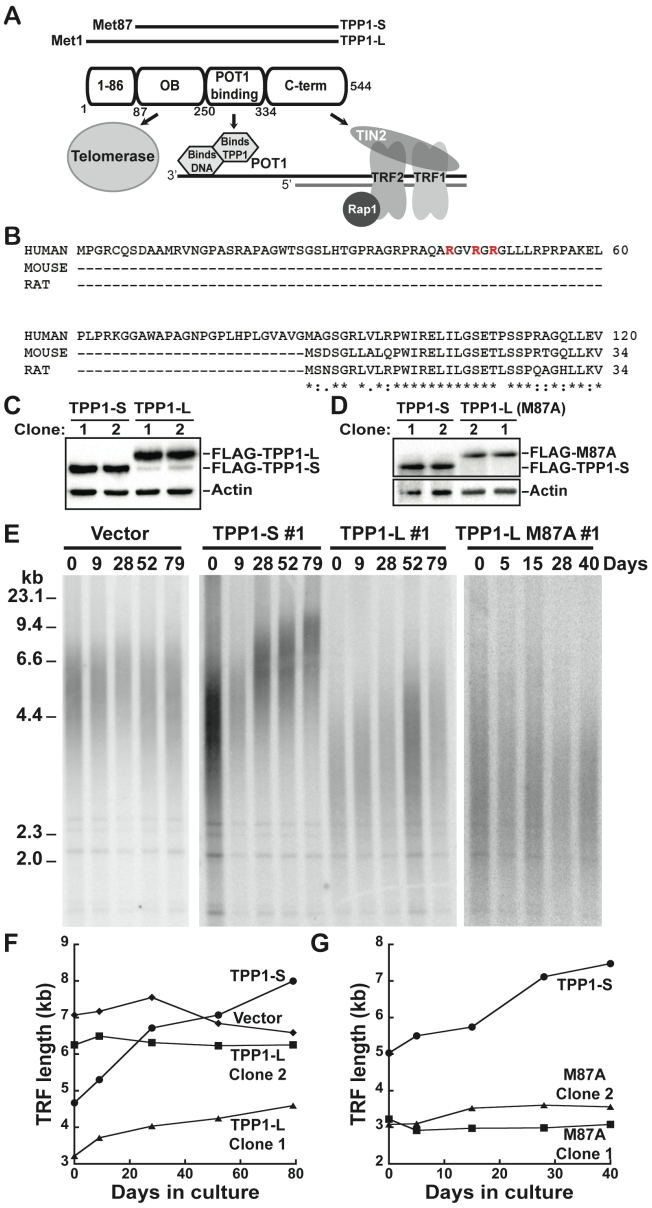


Figure 3.1 TPP1-S but not TPP1-L overexpression causes robust telomere elongation.
(A) Schematic showing a TPP1-centric view of the shelterin complex in humans. (B) Sequence for the N-terminus of human, mouse, and rat TPP1 proteins. Arginine residues in red indicate those mutated in the R_3E_3 mutant. (C and D) FLAG and actin immunoblots of lysates from indicated stable cell lines. (E) Telomere restriction fragment (TRF) Southern blot analysis of indicated HeLa-EM2-11ht clonal stable cell lines. (F and G) Plot of the mean telomere restriction fragment (TRF) length of indicated HeLa-EM2-11ht clonal stable cell lines.

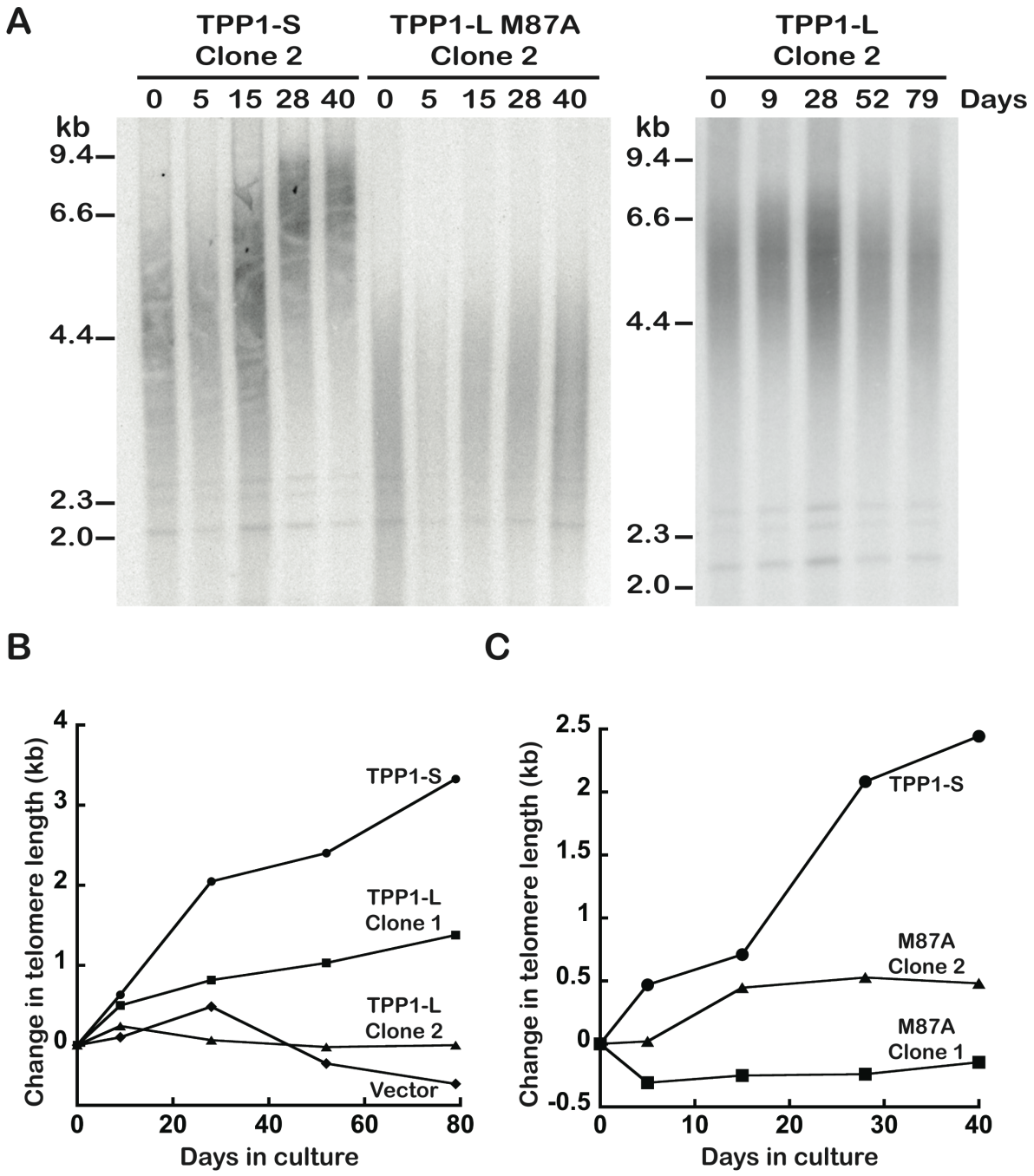


Figure 3.2 Telomere restriction fragment length analysis of TPP1-L and TPP1-S clonal cell lines. (A) Telomere restriction fragment (TRF) analysis of a second clone for HeLa-EM2-11ht cells expressing TPP1-S, TPP1-L, or TPP1-L M87A constructs for the indicated number of days in culture. (B and C) Plot of the change in mean telomere length for HeLa-EM2-11ht cells stably expressing the indicated constructs for the indicated number of days in culture

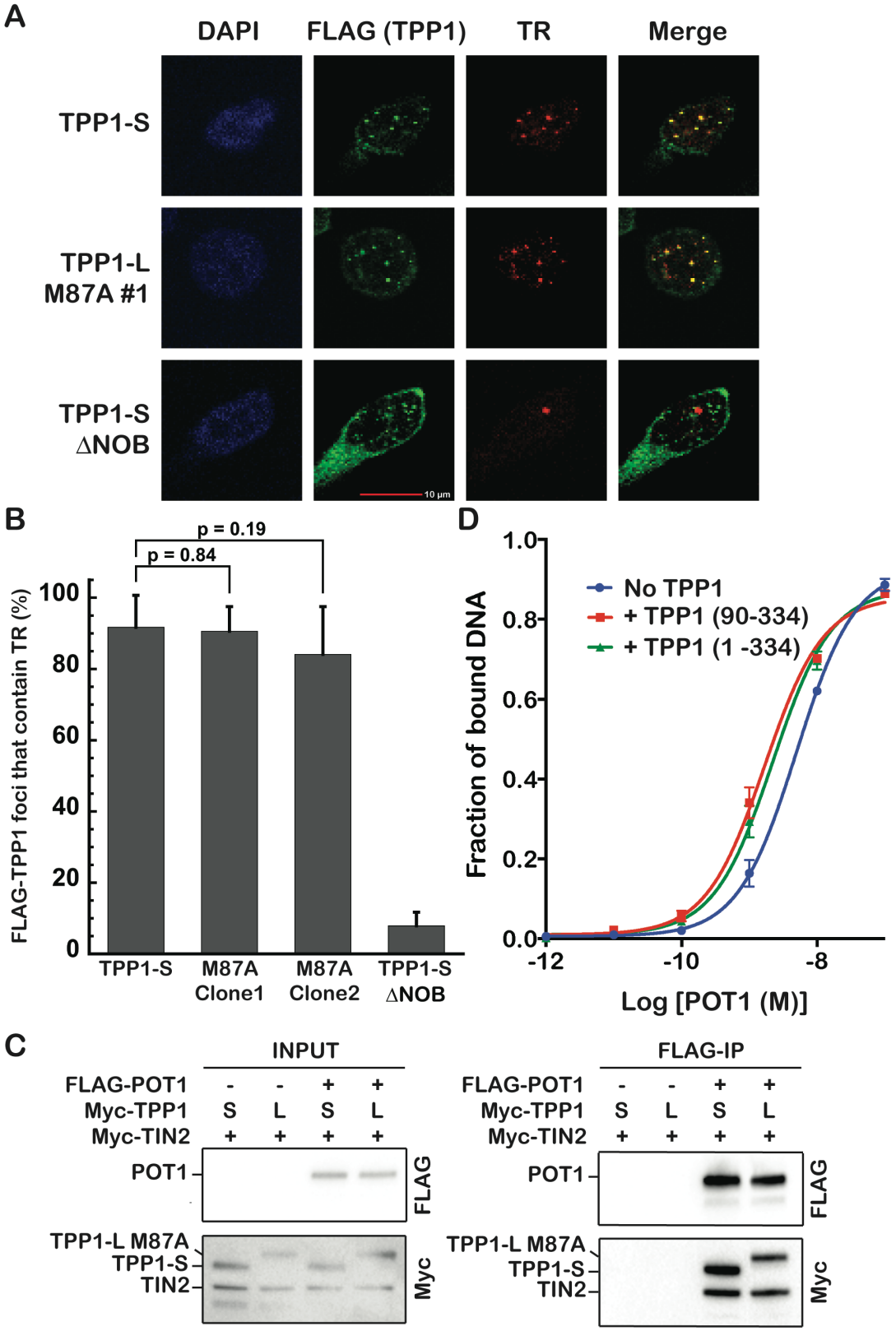

A
DAPI
FLAG (TPP1)
TR
Merge
TPP1-S
TPP1-L M87A #1
TPP1-S ΔNOB
10 μm
B
p = 0.19
p = 0.84
FLAG-TPP1 foci that contain TR (%)
100
80
60
40
20
0
TPP1-S
M87A Clone1
M87A Clone2
TPP1-S ΔNOB
D
Fraction of bound DNA
1.0
0.8
0.6
0.4
0.2
0.0
No TPP1
+ TPP1 (90-334)
+ TPP1 (1 -334)
-12
-10
-8
Log [POT1 (M)]
C
INPUT
FLAG-IP
FLAG-POT1 - - + +
Myc-TPP1 S L S L
Myc-TIN2 + + + +
POT1
FLAG
TPP1-L M87A
TPP1-S
TIN2
Myc

Figure 3.3 TPP1-S and TPP1-L recruit telomerase and protect chromosome ends.
(A) HeLa-EM2-11ht cells expressing TPP1-S, TPP1-L M87A, or TPP1ΔNOB were analyzed for telomerase recruitment to telomeres by immunofluorescence (IF) and fluorescence *in situ* hybridization (FISH). “FLAG (TPP1)” indicates IF signal of the indicated TPP1 construct at telomeres (green). Telomerase RNA (“TR”) was detected by FISH with a fluorescently tagged DNA probe (red). The “Merge” panel depicts the extent of telomerase recruitment to telomeres (yellow). (B) Quantitation of telomerase recruitment data of which A is representative. For each clone >100 telomere foci were scored and the mean percentage (bar) and standard deviation (error bar) of FLAG-TPP1 foci containing TR was plotted for triplicate measurements. Statistical significance was scored with a two-tailed student’s t test. (C) Pulldown of transiently expressed FLAG-POT1 on anti-FLAG conjugated beads with Myc-TPP1 and Myc-TIN2 constructs. (D) DNA binding curves from filter binding analysis of increasing concentrations of purified POT1 protein in the absence of any TPP1 protein, or in the presence of 200 nM of either TPP1-S^{90-334} or TPP1-L^{1-334}. 10 pM ^{32}P-end labeled GGTTAGGGTTAG DNA oligonucleotide was used in all binding experiments

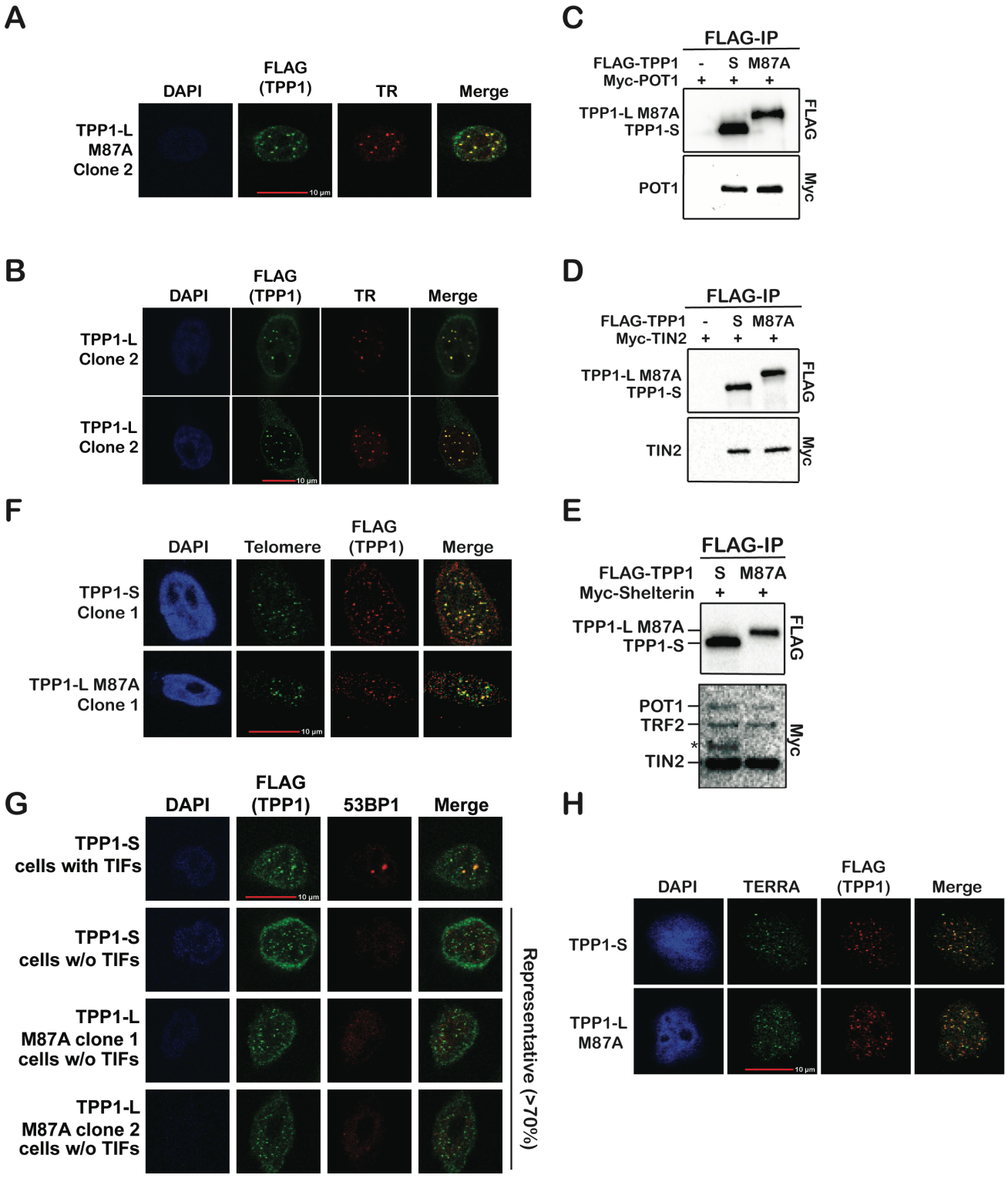

A
DAPI
FLAG (TPP1)
TR
Merge
TPP1-L M87A Clone 2
10 μm
B
DAPI
FLAG (TPP1)
TR
Merge
TPP1-L Clone 2
TPP1-L Clone 2
10 μm
C
FLAG-IP
FLAG-TPP1 - S M87A
Myc-POT1 + + +
TPP1-L M87A
TPP1-S
FLAG
POT1
Myc
D
FLAG-IP
FLAG-TPP1 - S M87A
Myc-TIN2 + + +
TPP1-L M87A
TPP1-S
FLAG
TIN2
Myc
E
FLAG-IP
FLAG-TPP1 S M87A
Myc-Shelterin + +
TPP1-L M87A
TPP1-S
FLAG
POT1
TRF2
*
TIN2
Myc
F
DAPI
Telomere
FLAG (TPP1)
Merge
TPP1-S Clone 1
TPP1-L M87A Clone 1
10 μm
G
DAPI
FLAG (TPP1)
53BP1
Merge
TPP1-S cells with TIFs
TPP1-S cells w/o TIFs
TPP1-L M87A clone 1 cells w/o TIFs
TPP1-L M87A clone 2 cells w/o TIFs
10 μm
Representative (>70%)
H
DAPI
TERRA
FLAG (TPP1)
Merge
TPP1-S
TPP1-L M87A
10 μm

Figure 3.4 Both TPP1-L and TPP1-S are proficient in chromosome end protection and telomerase recruitment.
(A and B) IF/FISH analysis of telomerase recruitment for TPP1-L M87A clone #2 (A) and two clones of TPP1-L WT (B). “FLAG (TPP1)” refers to immunofluorescence signal of FLAG-TPP1-L or FLAG-TPP1-L M87A at telomeres (green). “TR” represents FISH signal for telomerase RNA (red). “Merge” demonstrates successful telomerase recruitment to telomeres (yellow). See also Figure 3.3 for quantitation. (C) Pulldown of transiently expressed FLAG-TPP1-S or FLAG-TPP1-L M87A and Myc-POT1 on anti-FLAG conjugated beads. (D) Pulldown of transiently expressed FLAG-TPP1-S or FLAG-TPP1-L M87A and Myc-TIN2 on anti-FLAG conjugated beads. (E) Pulldown of transiently expressed Myc-TIN2, Myc-POT1, Myc-TRF2, Myc-Rap1, and FLAG-TPP1-S or FLAG-TPP1-L M87A, on anti-FLAG conjugated beads. Note that co-migration of TRF2 and Rap1 precluded distinction between these proteins in the blots. The asterisk indicates bleed through of the FLAG-TPP1-S signal on the anti-Myc blot. (F) IF-FISH depicting colocalization of indicated TPP1 construct with telomeres. “Telomere” refers to signal from a C-rich, Cy3-labled PNA probe against telomeric DNA (green). “FLAG (TPP1)” refers to FLAG immunofluorescence signal of indicated TPP1 construct (red). “Merge” panel reveals extent of TPP1 colocalization to the telomere (yellow). (G) Cells stably expressing TPP1-S or TPP1-L M87A were analyzed for TIFs using co-immunofluorescence (co-IF) against “FLAG (TPP1)” as a telomere marker and “53BP1” as a DNA damage marker. Co-localization of 53BP1 foci and FLAG-TPP1 foci within “Merge” panels indicates DNA damage at the telomere. An example of a cell with TIFs is shown along with images of cells that are more representative of the majority of cells. (H) IF/RNA-FISH analysis of TERRA colocalization with the indicated TPP1 construct. “TERRA” depicts signal from a C-rich, Cy3-labled telomeric PNA probe bound to RNA. “FLAG (TPP1)” depicts immunofluorescence signal from indicated TPP1 construct. “Merge” represents extent of TERRA co-localization with the telomere (yellow).

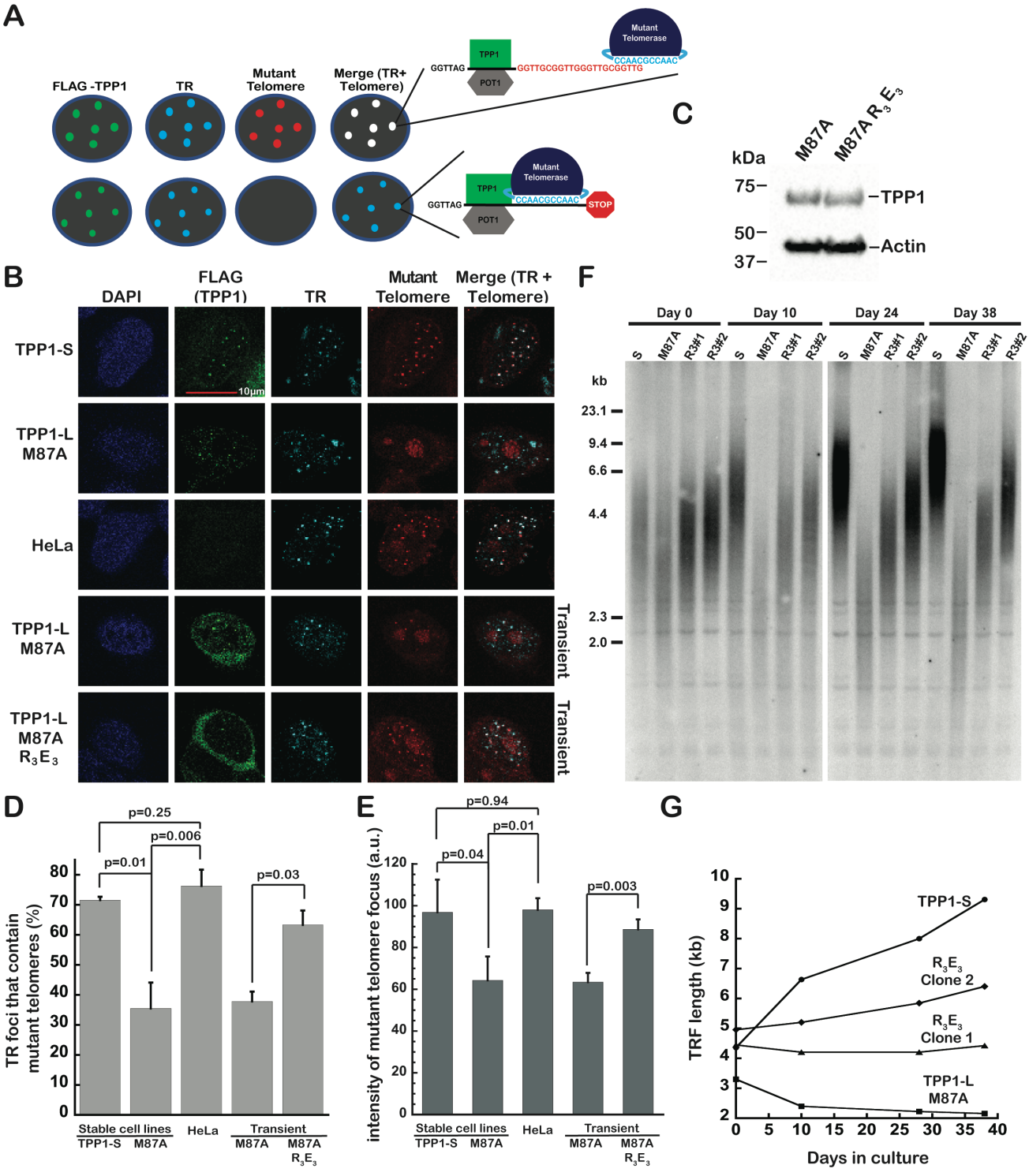

A
FLAG -TPP1
TR
Mutant Telomere
Merge (TR+ Telomere)
Mutant Telomerase
CCAACGCCAAC
TPP1
POT1
GGTTAG
GGTTGCGGTTGGGTTGCGGTTG
STOP
B
DAPI
FLAG (TPP1)
TR
Mutant Telomere
Merge (TR + Telomere)
TPP1-S
TPP1-L M87A
HeLa
TPP1-L M87A R3E3
Transient
10μm
C
kDa
M87A
M87A R3E3
75–
50–
37–
–TPP1
–Actin
F
Day 0
Day 10
Day 24
Day 38
S
M87A
R3#1
R3#2
kb
23.1
9.4
6.6
4.4
2.3
2.0
D
TR foci that contain mutant telomeres (%)
p=0.25
p=0.006
p=0.01
p=0.03
Stable cell lines
TPP1-S
M87A
HeLa
Transient
M87A
M87A R3E3
E
intensity of mutant telomere focus (a.u.)
p=0.94
p=0.01
p=0.04
p=0.003
G
TRF length (kb)
TPP1-S
R3E3 Clone 2
R3E3 Clone 1
TPP1-L M87A
Days in culture

Figure 3.5 TPP1-L is deficient in activating telomere synthesis.
(A) Cartoon representation of contrasting outcomes of the *in vivo* IF-coFISH telomere extension assay involving “FLAG-TPP1” (green), “TR” (cyan), and newly synthesized mutant telomeric repeats “Mutant Telomere” (red). White foci in “Merge” of TR and mutant telomere signals depicts colocalization. (B) HeLa-EM2-11ht cells transiently transfected with WT TERT and mutant telomerase RNA (TR) that contains a non-telomeric GCCAAC (WT: CCAAUC) template sequence. IF-coFISH was used to visualize telomere extension by telomerase in parental HeLa-EM2-11ht cells, indicated stable cell lines, and transiently transfected cells. “FLAG-TPP1” and “TR” were visualized as in Figure 3.3. “Mutant telomere” refers to signal from a Cy3-labled PNA probe against the mutant telomere sequence (red) that was synthesized by telomerase containing a mutant RNA template. Non-specific staining of the nucleolus by the mutant telomere probe occurred in all cells including HeLa cells that were not transfected with mutant telomerase (data not shown). (C) Immunoblot of lysates from HeLa-EM2-11ht cells transiently transfected with FLAG-TPP1-L M87A or FLAG-TPP1-L M87A-R_3E_3. (D and E) Quantitation of *in vivo* telomere extension data for which (B) is representative. >300 TR foci were scored and the mean percent of TR foci that contain mutant telomeres (bar; D) or mean spot intensity (bar; E), and standard deviation (error bar) were plotted for triplicate measurements. Statistical significance was scored using a two-tailed student’s t test. (F) TRF analysis of indicated clonal stable cell lines. (G) Plot of the mean telomere restriction fragment (TRF) length of indicated clonal stable cell lines.

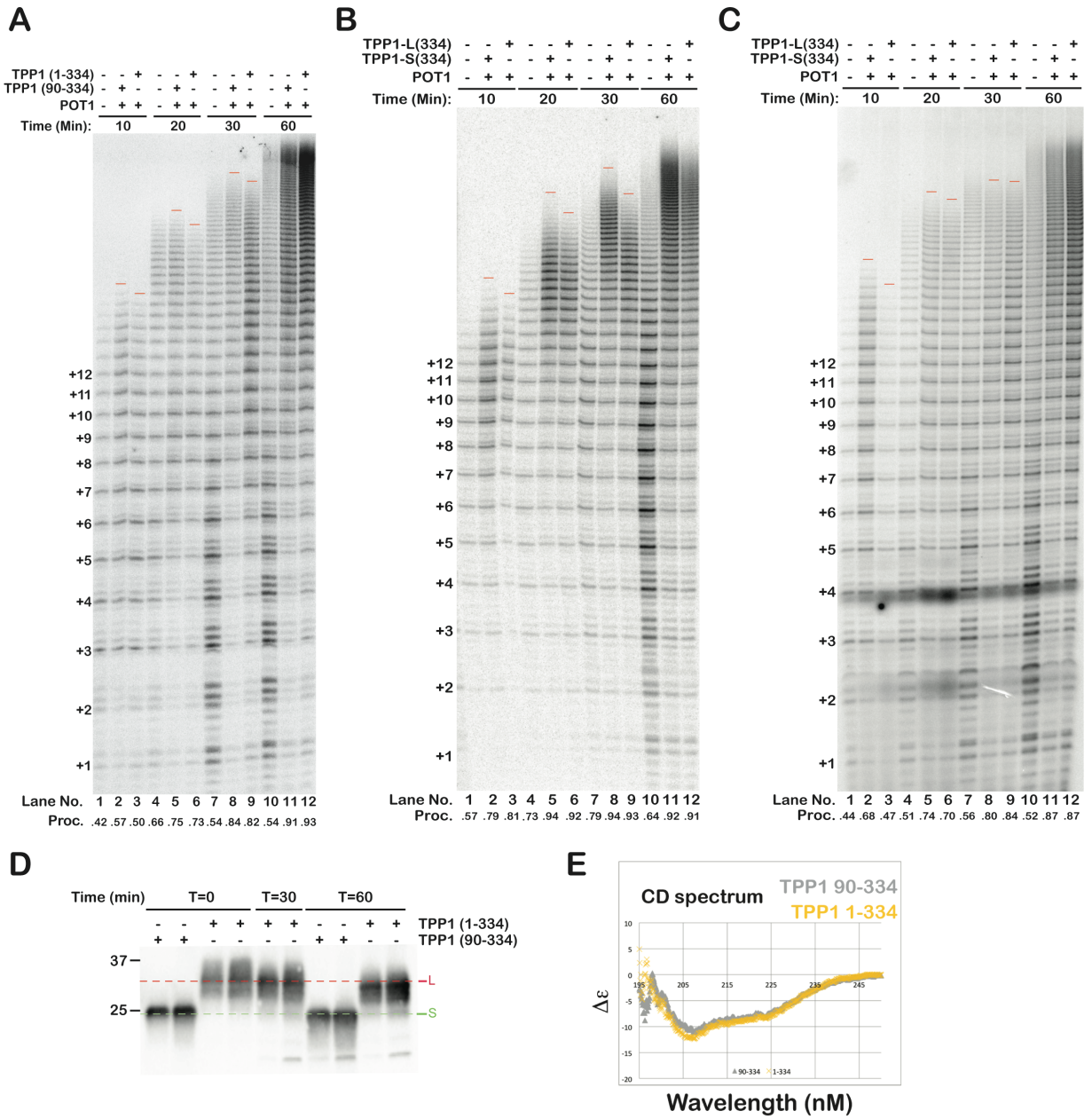


Figure 3.6 Direct telomerase activity assays with TPP1-L and TPP1-S proteins.
(A, B, and C) Three independent replicates of the direct telomerase activity assay with purified POT1, TPP1-S$^{90\text{-}334}$, and TPP1-L$^{1\text{-}334}$ proteins at 30°C for the indicated incubation times. Red horizontal bars approximately mark the longest detectable product. The ratio of the total intensity from bands representing the addition of 9 or more hexad repeats over the total intensity of the lane is shown below as processivity (proc.). (D) Anti-TPP1 immunoblot of telomerase activity reactions (minus radiolabeled dGTP) at the indicated time-points. Green and red dashed lines indicate the center of the TPP1-S and TPP1-L proteins, respectively, at time = 0. (E) Circular Dichroism (CD) spectra of TPP1$^{1\text{-}334}$ and TPP1$^{90\text{-}334}$ proteins showing almost superimposable profiles and suggesting the absence of any significant secondary structure in TPP1-L aa 1-86.

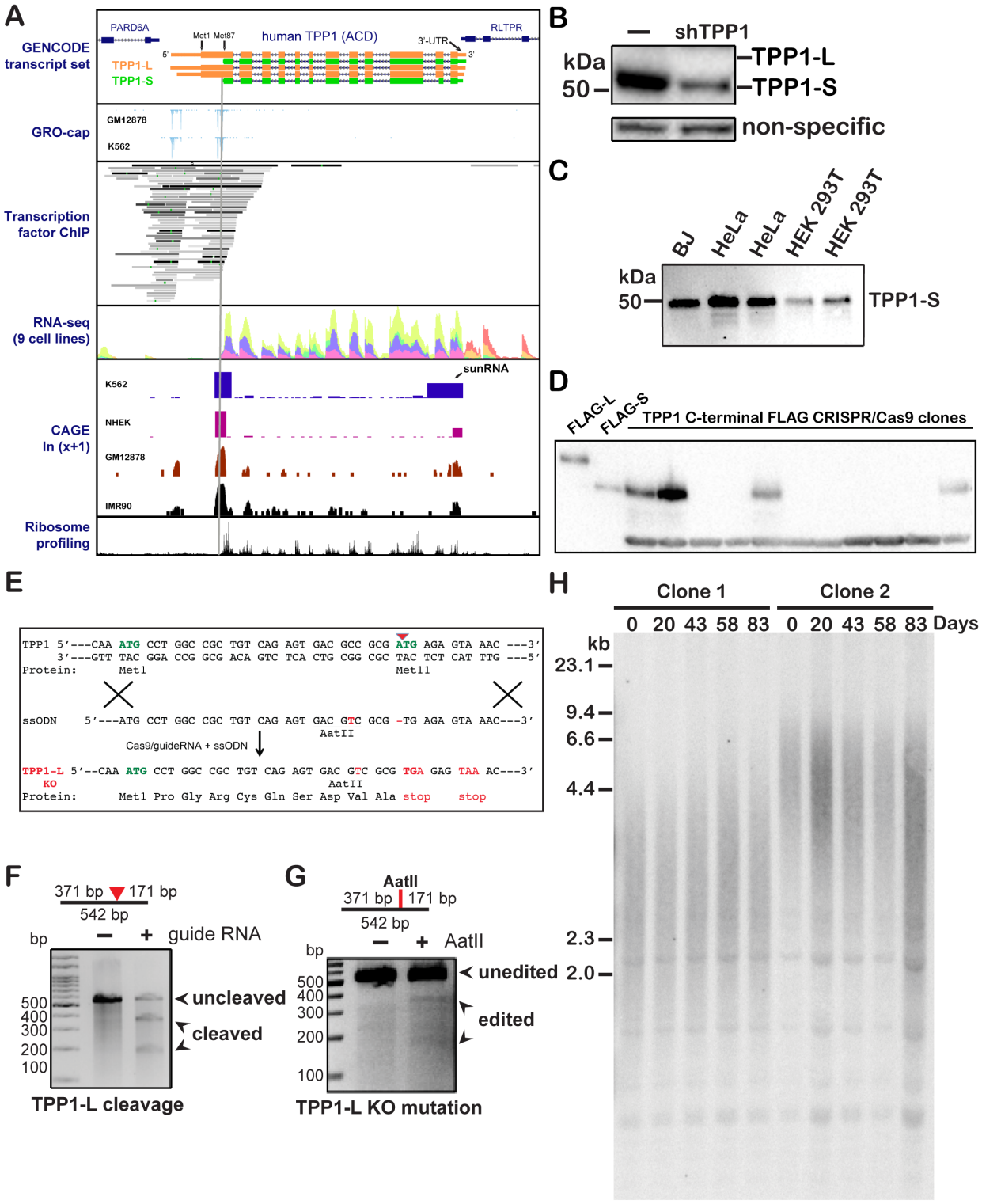

A
GENCODE transcript set
PARD6A
Met1 Met87
human TPP1 (ACD)
3'-UTR
RLTPR
TPP1-L
TPP1-S
GRO-cap
GM12878
K562
Transcription factor ChIP
RNA-seq (9 cell lines)
sunRNA
K562
NHEK
GM12878
IMR90
CAGE ln (x+1)
Ribosome profiling
B
shTPP1
kDa
50
TPP1-L
TPP1-S
non-specific
C
BJ
HeLa
HeLa
HEK 293T
HEK 293T
kDa
50
TPP1-S
D
FLAG-L
FLAG-S
TPP1 C-terminal FLAG CRISPR/Cas9 clones
E
TPP1 5'---CAA ATG CCT GGC CGC TGT CAG AGT GAC GCC GCG ATG AGA GTA AAC ---3'
3'---GTT TAC GGA CCG GCG ACA GTC TCA CTG CGG CGC TAC TCT CAT TTG ---5'
Protein: Met1 Met11
ssODN 5'---ATG CCT GGC CGC TGT CAG AGT GAC GTC GCG -TG AGA GTA AAC---3'
AatII
Cas9/guideRNA + ssODN
TPP1-L KO 5'--CAA ATG CCT GGC CGC TGT CAG AGT GAC GTC GCG TGA GAG TAA AC---3'
AatII
Protein: Met1 Pro Gly Arg Cys Gln Ser Asp Val Ala stop stop
F
371 bp
171 bp
542 bp
bp
guide RNA
500
400
300
200
100
uncleaved
cleaved
TPP1-L cleavage
G
AatII
371 bp
171 bp
542 bp
AatII
bp
500
400
300
200
100
unedited
edited
TPP1-L KO mutation
H
Clone 1
Clone 2
0 20 43 58 83 0 20 43 58 83 Days
kb
23.1
9.4
6.6
4.4
2.3
2.0

Figure 3.7 TPP1-S but not TPP1-L accumulates in human cell lines.
(A) Existing high-throughput sequencing information for the *ACD* locus was visualized using the UCSC genome browser. Vertical grey bar denotes the 5' end of TPP1-S mRNA. CAGE tags at the 3' end of TPP1 mRNA indicative of the 5' end of the sunRNA are indicated with an arrow. See also Figure 3.8A and Methods. (B) Immunoblot of HeLa-EM2-11ht cells using a rabbit polyclonal antibody against human TPP1 protein (epitope in OB domain; Bethyl Labs) in the presence or absence of shTPP1. (C) TPP1 immunoblot from indicated cell types after enrichment using a biotinylated telomeric DNA oligonucleotide bound to POT1 immobilized on streptavidin beads. (D) FLAG immunoblot of lysates from CRISPR-edited HEK 293T clones containing endogenous 3X-FLAG tagged TPP1. "FLAG-L" and "FLAG-S" proteins from transiently transfected HeLa-EM2-11ht cells served as size markers. (E) Strategy to engineer TPP1-L KO cells. A guide RNA was designed to cleave the ATG codon for M11. A repair ssODN was designed to disrupt the M11 ATG codon, create an in-frame stop codon (TGA), and introduce an *Aat*II site using silent mutations. (F) A PCR-based Surveyor assay to assess the efficiency of cleavage by Cas9. (G) An *Aat*II digest performed two days post-transfection indicated a detectible level of edited cells in the population. (H) TRF Southern blot analysis of CRISPR-Cas9 derived HeLa-EM2-11ht TPP1-L KO clones.

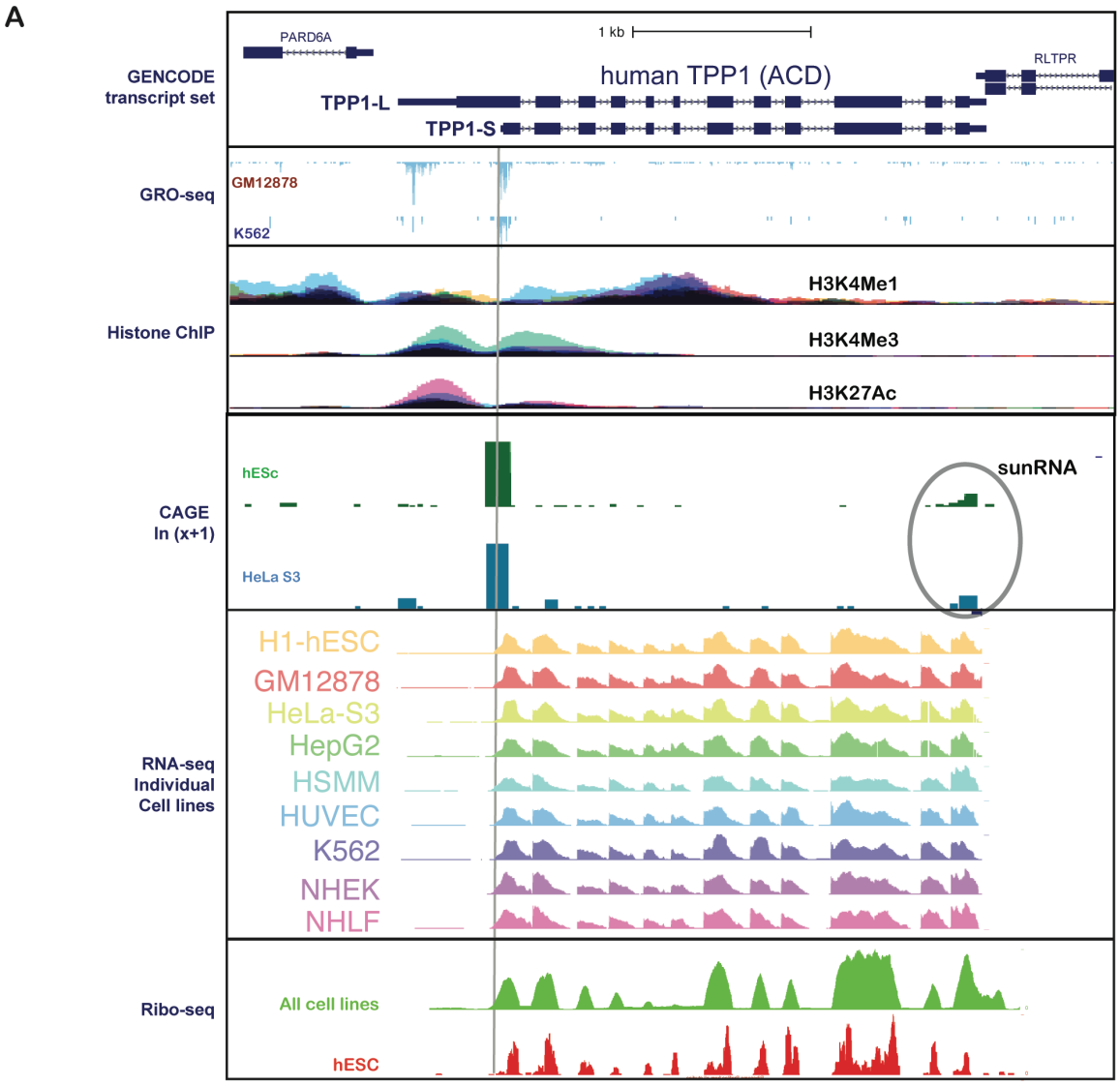

A
PARD6A
1 kb
RLTPR
GENCODE transcript set
human TPP1 (ACD)
TPP1-L
TPP1-S
GRO-seq
GM12878
K562
Histone ChIP
H3K4Me1
H3K4Me3
H3K27Ac
CAGE ln (x+1)
hESc
HeLa S3
sunRNA
RNA-seq Individual Cell lines
H1-hESC
GM12878
HeLa-S3
HepG2
HSMM
HUVEC
K562
NHEK
NHLF
Ribo-seq
All cell lines
hESC
B
TPP1-L
TPP1-S
STOP
WT
--GAGGAGAAGGAGTTCGGCTTCCGCGGGACAGAGGG--
guide RNA + Cas9
+
FLAG
ssODN
C
INPUT
FLAG-IP
unedited
1
2
Edited clone
FLAG
TPP1-S
Actin
D
WT: CAA ATG CCT GGC CGC TGT CAG AGT GAC GCC GCG ATG AGA GTA AAC
TPP1 KO clone 1:
Allele 1: CAA ATG CCT GGC CGC TGT CAG AGT GAC GCC GCG ATT GAG AGT AAA (2)
Allele 2: CAA ATG CCT GGC CGC TGT CAG AGT GAC GCC GCG _TG AGA GTA AAC (2)
TPP1 KO clone 2 (mixed clone):
Allele 1: CAA ATG CCT GGC CGC TGT CAG AGT GAC GTC GCG _TG AGA GTA AAC (7)
Allele 2: CAA ATG CCT GGC CGC TGT CAG AGT GAC GCC GCG ATT GAG AGT AAA (2)
Allele 3: CAA ATG CCT GGC CGC _ _ _C_G _ _T _ _ _ _ _CAG _TG AGA GTA AAC (1)

Figure 3.8 Transcriptome and protein analysis suggest separate transcription cassettes for TPP1-L and TPP1-S, although TPP1-S protein is the predominant form in cells.
(A) High-throughput sequencing data visualized on the UCSC genome browser at the *ACD* locus. GRO-seq data for assessing nascent mRNA from the indicated cell lines, cumulative ChiP data from a collection of cell lines for indicated histone modifications for assessing gene expression potential, CAGE data for human embryonic stem cells and HeLa S3 cells, RNA-seq tracks from individual cell lines, and Ribo-seq data (cumulative versus embryonic stem cell line) are shown along with the TPP1-L and TPP1-S transcript annotations from GENCODE. The vertical grey bar denotes the 5' end of TPP1-S mRNA. (B) CRISPR-Cas9 strategy to introduce the 3X FLAG tag at the C-terminus of endogenous TPP1 in HEK 293T cells is shown. (C) Immunoprecipitation of lysates from clones isolated from the FLAG-TPP1-S CRISPR experiment using anti-FLAG M2 affinity gel followed by blotting using the anti-actin and anti-FLAG-HRP conjugate. An unedited clone was included as negative control. (D) Allele status of *ACD* (TPP1) in the CRISPR-mediated KO of TPP1-L.

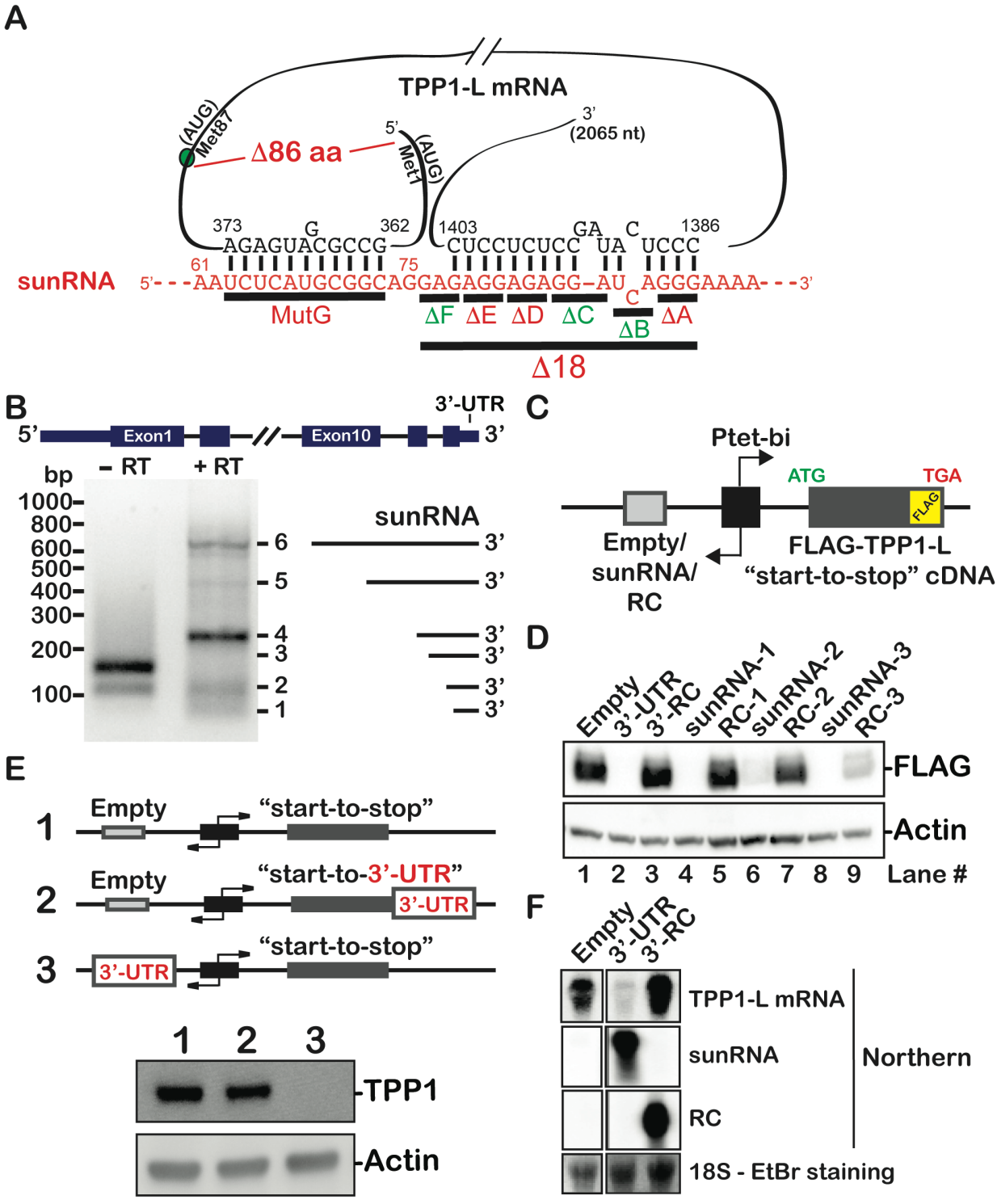

A
TPP1-L mRNA
(AUG) Met87
Δ86 aa
5'
(AUG) Met1
3'
(2065 nt)
373 G 362
AGAGUACGCCG
1403 GA C 1386
CUCCUCUCC UA UCCC
sunRNA
61 75
5'- - -AAUCUCAUGCGGCAGGAGAGGAGAGG–AU C AGGGAAAA- - -3'
MutG
ΔF ΔE ΔD ΔC ΔB ΔA
Δ18
B
3'-UTR
5'
Exon1
Exon10
3'
bp
- RT
+ RT
1000
800
600
500
400
300
200
100
sunRNA
6
5
4
3
2
1
3'
C
Ptet-bi
ATG
TGA
FLAG
Empty/
sunRNA/
RC
FLAG-TPP1-L
"start-to-stop" cDNA
D
Empty
3'-UTR
3'-RC
sunRNA-1
RC-1
sunRNA-2
RC-2
sunRNA-3
RC-3
FLAG
Actin
1 2 3 4 5 6 7 8 9 Lane #
E
1 Empty "start-to-stop"
2 Empty "start-to-3'-UTR" 3'-UTR
3 3'-UTR "start-to-stop"
1 2 3
TPP1
Actin
F
Empty
3'-UTR
3'-RC
TPP1-L mRNA
sunRNA
RC
Northern
18S - EtBr staining

Figure 3.9 An intragenic RNA in the 3'-UTR of TPP1 completely shuts down TPP1-L.
(A) Extensive complementarity was detected between two regions in the TPP1-L ORF and 3'-UTR (sunRNA). TPP1-S mRNA lacks the sequence that is complementary to the 63-73 nt region of 3'-UTR; MutG corresponds to a sunRNA mutant encompassing all complementary nucleotides in this region. Three nucleotide deletions were engineered in the sunRNA. Deletions in red are in regions of perfect predicted complementarity between the sunRNA and TPP1 mRNA, while deletions in green are in regions where the predicted complementarity is imperfect. (B) TPP1 sunRNAs were cloned from HEK 293T cells using 5' RACE. (C) Schematic of the vector used to study the effect of sunRNA expression on TPP1 protein. Expression of both the protein and RNA constructs is driven by a bi-directional promoter using doxycycline ("Tet-on" system). "RC" is a reverse complement of the sunRNA. (D) Western blot showing the silencing of TPP1-L by sunRNA constructs. "3'-UTR" indicates a minimal sunRNA-like construct that contains only the 93 bp between the stop codon and the start of the polyA tail of TPP1 mRNA. (E) The 3'-UTR sequence of TPP1 silences TPP1-L when expressed as a ncRNA (in *trans*), but not when it is inherent to TPP1-L mRNA (in *cis*). (F) Northern blot analysis of TPP1-L mRNA, sunRNA, and RC in the indicated transfections. 18S signal from ethidium bromide staining was used as a loading control.

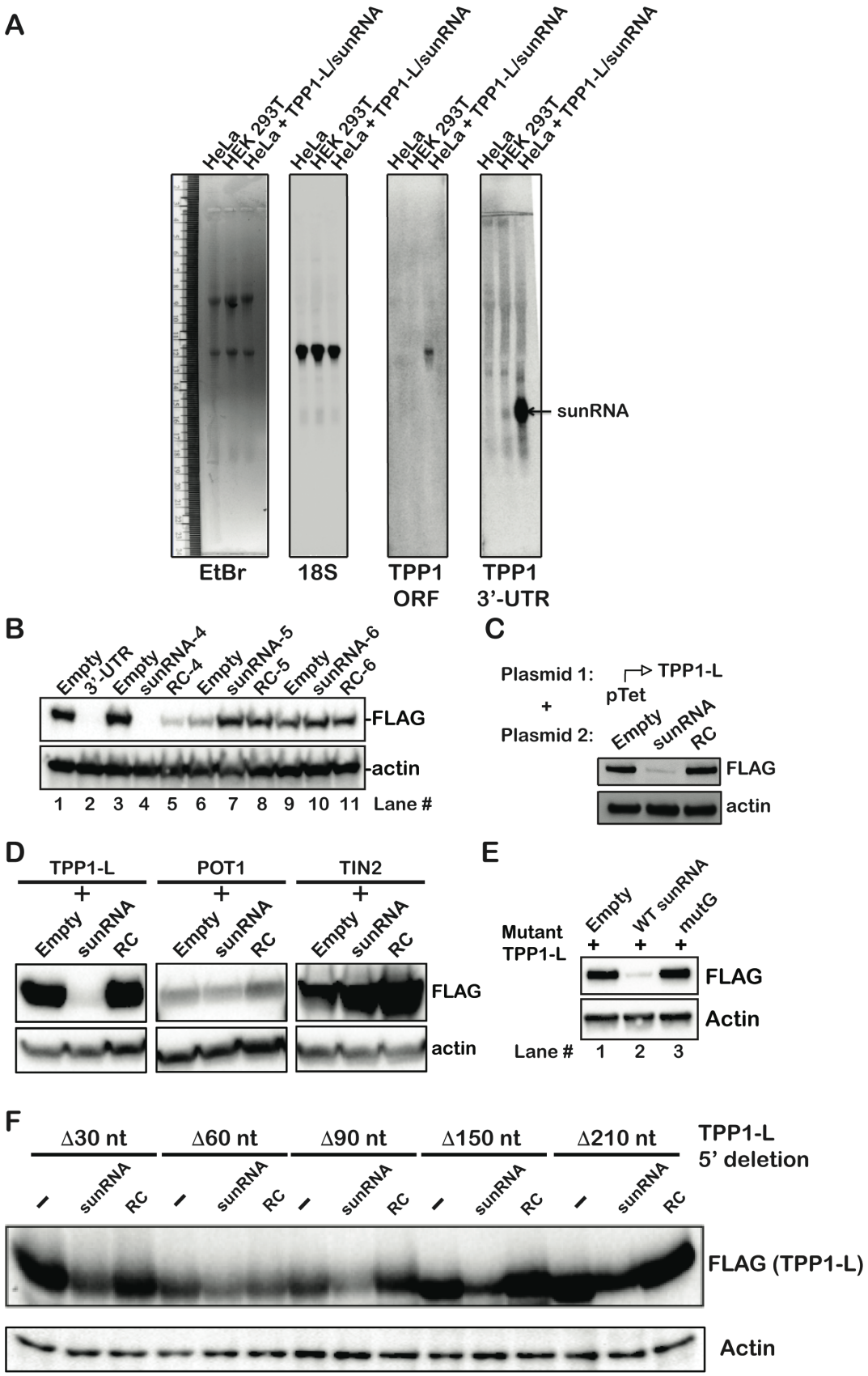

A
HeLa
HEK 293T
HeLa + TPP1-L/sunRNA
sunRNA
EtBr
18S
TPP1 ORF
TPP1 3'-UTR
B
Empty
3'-UTR
sunRNA-4
RC-4
sunRNA-5
RC-5
sunRNA-6
RC-6
FLAG
actin
1 2 3 4 5 6 7 8 9 10 11 Lane #
C
Plasmid 1:
TPP1-L
pTet
+
Plasmid 2:
Empty
sunRNA
RC
FLAG
actin
D
TPP1-L
POT1
TIN2
+
Empty
sunRNA
RC
FLAG
actin
E
Mutant TPP1-L
Empty
WT sunRNA
mutG
+
FLAG
Actin
Lane # 1 2 3
F
Δ30 nt
Δ60 nt
Δ90 nt
Δ150 nt
Δ210 nt
TPP1-L 5' deletion
-
sunRNA
RC
FLAG (TPP1-L)
Actin

Figure 3.10 Silencing of the sunRNA is specific for TPP1-L.
(A) Northern blot of sunRNA from whole-cell RNA isolated from indicated cell lines using TRIzol (Thermo Fisher). EtBr indicates ethidium bromide staining. “18S”, “TPP1-ORF”, and “TPP1 3’-UTR” indicate Northern blots with probes complementary to these RNAs. The “HeLa + TPP1-L/sunRNA” lanes indicate HeLa cell line transfected with the plasmid co-expressing FLAG-TPP1-L and sunRNA. The 3’-UTR probe should in principle detect endogenous TPP1 mRNA in addition to sunRNA. However, despite a detectible signal for sunRNA, we did not observe a signal for endogenous TPP1 mRNA, possibly because its detection is obscured by the abundance of 18S RNA that migrates at a similar position on the gel. (B) Western blot to detect silencing of TPP1-L by sunRNA constructs 4, 5, and 6. (C) Anti-FLAG immunoblots were performed on lysates from cells co-transfected with two plasmids, one of which codes for dox-inducible FLAG-TPP1-L, and the other for dox-inducible sunRNA, RC, or empty constructs. (D) Anti-FLAG immunoblots showing that the sunRNA of TPP1 silences TPP1-L but not human POT1 or TIN2. (E) Compensatory mutations in TPP1-L that restore complementarity to MutG sunRNA fail to rescue silencing. (F) Anti-TPP1 blots from lysates transfected with plasmids co-expressing the indicated TPP1-L 5’ deletions and either no sunRNA (“-”), sunRNA, or RC. Note that the deletions were designed to maintain the frame of TPP1-L translation (i.e., deletions of precisely 10, 20, 30, 50, or 70 aa from the N-terminus of the translated protein).

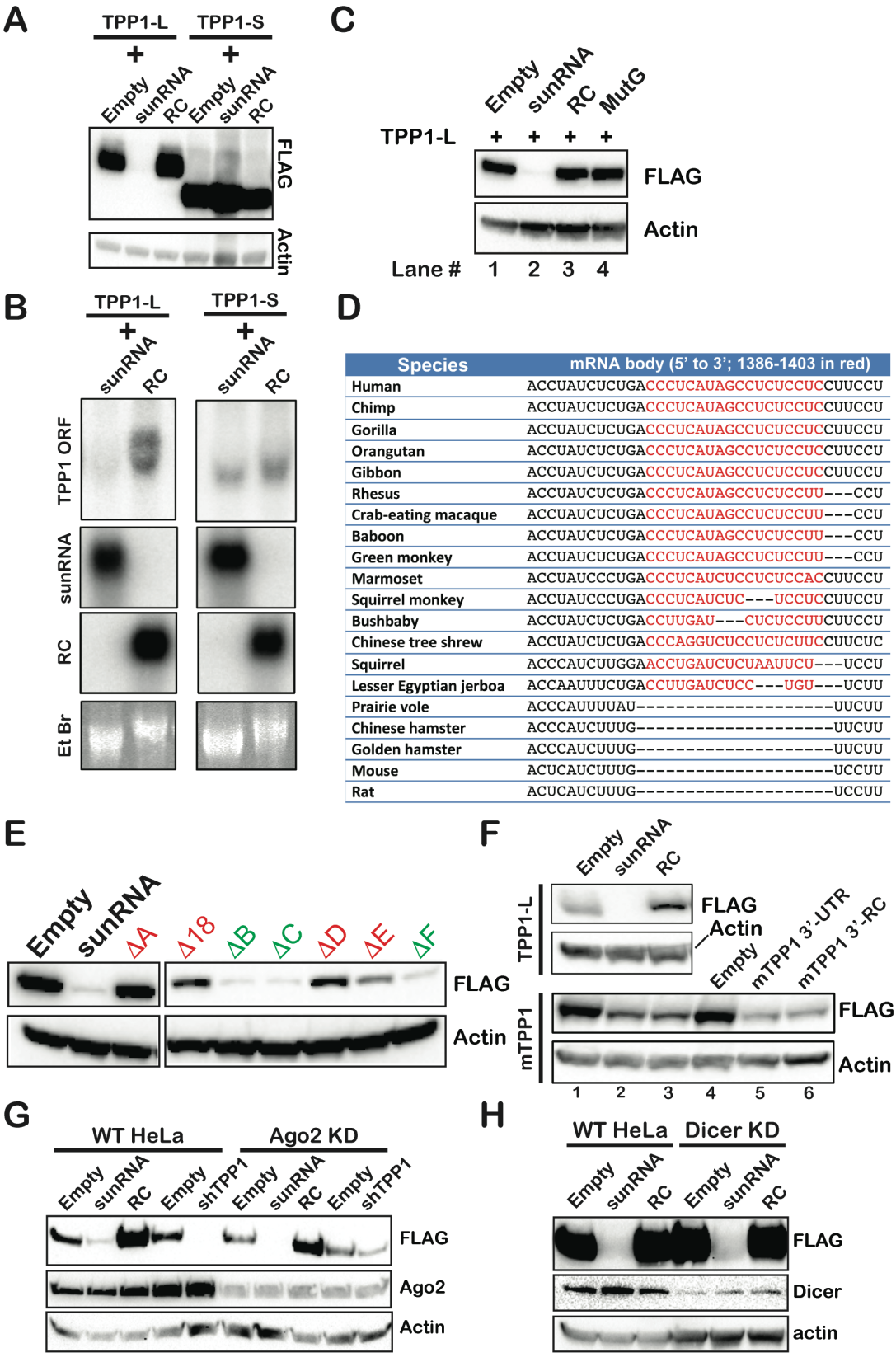

A
TPP1-L
TPP1-S
+
+
Empty
sunRNA
RC
Empty
sunRNA
RC
FLAG
Actin
B
TPP1-L
TPP1-S
+
+
sunRNA
RC
sunRNA
RC
TPP1 ORF
sunRNA
RC
Et Br
C
Empty
sunRNA
RC
MutG
TPP1-L
+ + + +
FLAG
Actin
Lane # 1 2 3 4
D
Species
mRNA body (5' to 3'; 1386-1403 in red)
Human ACCUAUCUCUGACCCUCAUAGCCUCUCCUCCUUCCU
Chimp ACCUAUCUCUGACCCUCAUAGCCUCUCCUCCUUCCU
Gorilla ACCUAUCUCUGACCCUCAUAGCCUCUCCUCCUUCCU
Orangutan ACCUAUCUCUGACCCUCAUAGCCUCUCCUCCUUCCU
Gibbon ACCUAUCUCUGACCCUCAUAGCCUCUCCUCCUUCCU
Rhesus ACCUAUCUCUGACCCUCAUAGCCUCUCCUU---CCU
Crab-eating macaque ACCUAUCUCUGACCCUCAUAGCCUCUCCUU---CCU
Baboon ACCUAUCUCUGACCCUCAUAGCCUCUCCUU---CCU
Green monkey ACCUAUCUCUGACCCUCAUAGCCUCUCCUU---CCU
Marmoset ACCUAUCCCUGACCCUCAUCUCCUCUCCACCUUCCU
Squirrel monkey ACCUAUCCCUGACCCUCAUCUC---UCCUCCUUCCU
Bushbaby ACCUAUCUCUGACCUUGAU---CUCUCCUUCUUCCU
Chinese tree shrew ACCUAUCUCUGACCCAGGUCUCCUCUCUUCCUUCUC
Squirrel ACCCAUCUUGGAACCUGAUCUCUAAUUCU---UCCU
Lesser Egyptian jerboa ACCAAUUUCUGACCUUGAUCUCC---UGU---UCUU
Prairie vole ACCCAUUUUAU----------UUCUU
Chinese hamster ACCCAUCUUUG----------UUCUU
Golden hamster ACCCAUCUUUG----------UUCUU
Mouse ACUCAUCUUUG----------UCCUU
Rat ACUCAUCUUUG----------UCCUU
E
Empty
sunRNA
ΔA
Δ18
ΔB
ΔC
ΔD
ΔE
ΔF
FLAG
Actin
F
Empty
sunRNA
RC
TPP1-L
FLAG
Actin
Empty
mTPP1 3'-UTR
mTPP1 3'-RC
mTPP1
FLAG
Actin
1 2 3 4 5 6
G
WT HeLa
Ago2 KD
Empty
sunRNA
RC
Empty
shTPP1
Empty
sunRNA
RC
Empty
shTPP1
FLAG
Ago2
Actin
H
WT HeLa
Dicer KD
Empty
sunRNA
RC
Empty
sunRNA
RC
FLAG
Dicer
actin

Figure 3.11 sunRNA silences TPP1-L but not TPP1-S.
(A) Immunoblot showing that sunRNA does not silence TPP1-S. (B) Northern blot analysis confirming that the sunRNA diminishes TPP1-L mRNA levels but not TPP1-S mRNA levels. (C) Anti-FLAG immunoblot analysis showing that mutations in sunRNA region complementary to TPP1-L abolish silencing. (D) Sequence alignment of TPP1 homologs reveals a second primate specific element in TPP1 ORF (first element is aa 1-86 of TPP1-L) that is absent in rodent TPP1 proteins. This region includes a site complementary to the sunRNA (shown in red). (E) Anti-FLAG immunoblot analysis of lysates from cells transiently transfected with sunRNA mutants (defined in Figure 3.9A) and FLAG-TPP1-L show that disruption of the predicted complementarity between TPP1-L and the sunRNA reduces TPP1-L silencing. (F) *Top:* Replicate data for the sunRNA silencing FLAG-TPP1-L. *Bottom:* Bi-directional constructs for co-expression of FLAG-tagged mouse TPP1 with Empty vector (lane 1), human TPP1 3'-UTR (sunRNA; lane 2), human TPP1 RC (lane 3), Empty vector (lane 4), mouse TPP1 3'-UTR (lane 5), or the reverse complement (RC) of the mouse TPP1 3'-UTR (lane 6) were transfected into HeLa cells and analyzed using anti-FLAG immunoblotting. (G and H) Immunoblots from HeLa transfected with FLAG-TPP1-L and Empty/sunRNA/RC constructs; and siRNAs for Ago2 (F) or Dicer (G) knockdown. shTPP1: same as in Figure 3.7B.

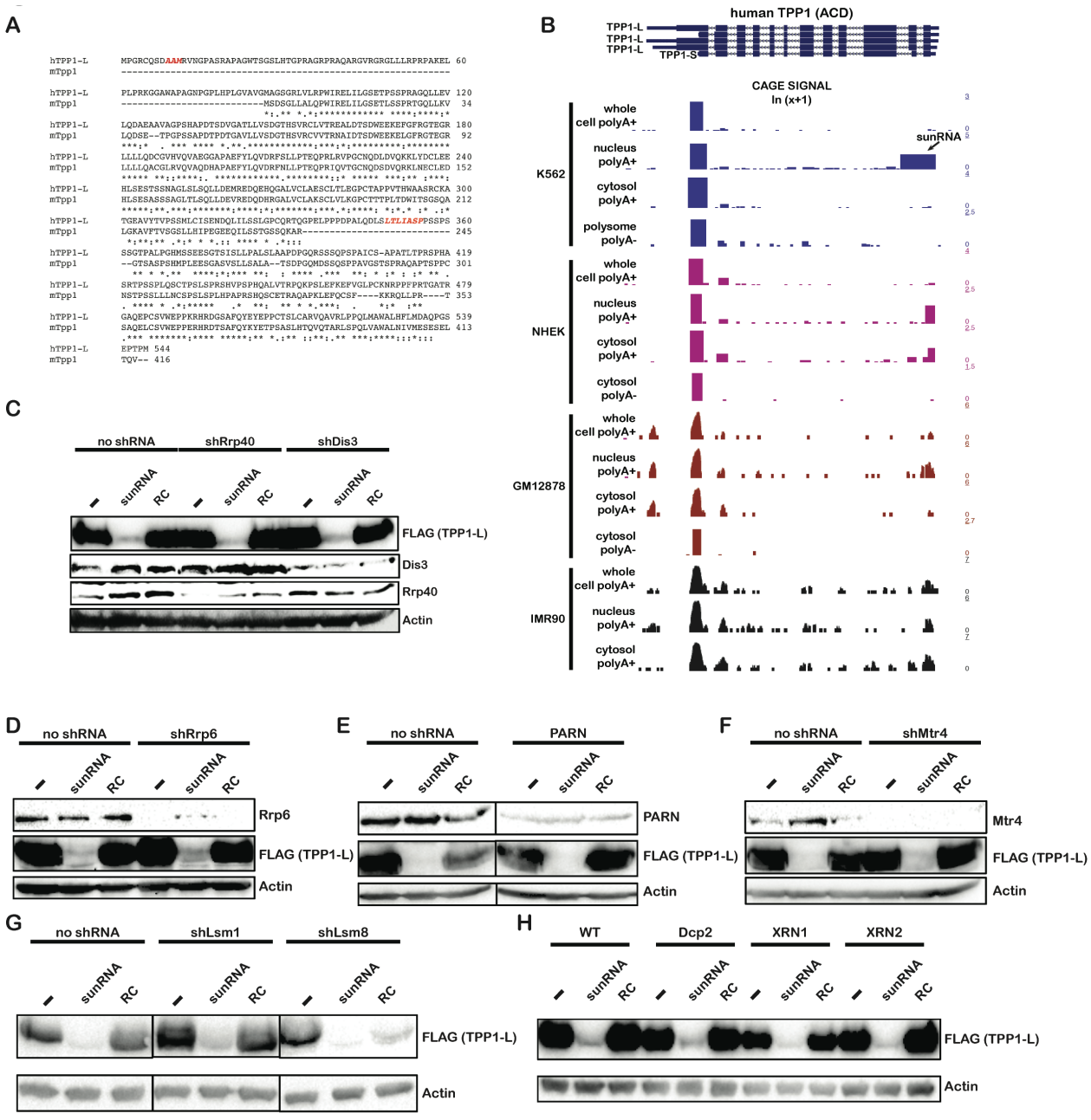

A
hTPP1-L MPGRCQSDAAMRVNGPASRAPAGWTSGSLHTGPRAGRPRAQARGVRGRGLLLRPRPAKEL 60
mTpp1 ---------- 120
hTPP1-L PLPRKGGAWAPAGNPGPLHPLGVAVGMAGSGRLVLRPWIRELILGSETPSSPRAGQLLEV 120
mTpp1 MSDSGLLALQPWIRELILGSETLSSPRTGQLLKV 34
hTPP1-L LQDAEAAVAGPSHAPDTSDVGATLLVSDGTHSVRCLVTREALDTSDWEEKEFGFRGTEGR 180
mTpp1 LQDSE--TPGPSSAPDTPDTGAVLLVSDGTHSVRCVVTRNAIDTSDWEEKELGFRGTEGR 92
hTPP1-L LLLLQDCGVHVQVAEGGAPAEFYLQVDRFSLLPTEQPRLRVPGCNQDLDVQKKLYDCLEE 240
mTpp1 LLLLQACGLRVQVAQDHAPAEFYLQVDRFNLLPTEQPRIQVTGCNQDSDVQRKLNECLED 152
hTPP1-L HLSESTSSNAGLSLSQLLDEMREDQEHQGALVCLAESCLTLEGPCTAPPVTHWAASRCKA 300
mTpp1 HLSESASSSAGLTLSQLLDEVREDQDHRGALVCLAKSCLVLKGPCTTTPLTDWITSGSQA 212
hTPP1-L TGEAVYTVPSSMLCISENDQLILSSLGPCQRTQGPELPPPDPALQDLSLYLIASPPSSPS 360
mTpp1 LGKAVFTVSGSLLHIPEGEEQILSSTGSSQKAR---------- 245
hTPP1-L SSGTPALPGHMSSEESGTSISLLPALSLAAPDPGQRSSSQPSPAICS-APATLTPRSPHA 419
mTpp1 --GTSASPSHMPLEESGASVSLLSALA--TSDPGQMDSSQSPPAVGSTSPRAQAPTSPPC 301
hTPP1-L SRTPSSPLQSCTPSLSPRSHVPSPHQALVTRPQKPSLEPKEFVGLPCKNRPPFPRTGATR 479
mTpp1 NSTPSSLLLNCSPSLSPLHPAPRSHQSCETRAQAPKLEPQCSF----KKRQLLPR----T 353
hTPP1-L GAQEPCSVWEPPKRHRDGSAFQYEYEPPCTSLCARVQAVRLPPQLMAWALHFLMDAQPGS 539
mTpp1 SAQELCSVWEPPERHRDTSAFQYKYETPSASLHTQVQTARLSPQLVAWALNIVMESESEL 413
hTPP1-L EPTPM 544
mTpp1 TQV-- 416
B
human TPP1 (ACD)
TPP1-L
TPP1-L
TPP1-L
TPP1-S
CAGE SIGNAL
ln (x+1)
sunRNA
K562
NHEK
GM12878
IMR90
whole cell polyA+
nucleus polyA+
cytosol polyA+
polysome polyA-
cytosol polyA-
C
no shRNA
shRrp40
shDis3
sunRNA
RC
FLAG (TPP1-L)
Dis3
Rrp40
Actin
D
no shRNA
shRrp6
Rrp6
FLAG (TPP1-L)
Actin
E
no shRNA
PARN
PARN
FLAG (TPP1-L)
Actin
F
no shRNA
shMtr4
Mtr4
FLAG (TPP1-L)
Actin
G
no shRNA
shLsm1
shLsm8
FLAG (TPP1-L)
Actin
H
WT
Dcp2
XRN1
XRN2
FLAG (TPP1-L)
Actin

Figure 3.12 sunRNA mediated regulation of TPP1 is specific to human TPP1 sequence but not dependent on any unique RNA surveillance or silencing machinery.
(A) Amino acid sequence alignment of human TPP1-L and mouse TPP1. Residues indicated in red are coded by nucleotides predicted to be complementary to the human sunRNA sequence. Both these motifs reside in regions that are absent in mouse TPP1 protein (and mRNA). (B) Detailed analysis of publically available RIKEN CAGE data for the *ACD* locus encoding TPP1-L mRNA, TPP1-S mRNA, and sunRNA in the indicated fractions of the indicated cell lines. Combined, these data suggest that sunRNA is 5' capped, 3' polyadenylated, nuclear, and not actively engaging polysomes. Moreover, the 3' end of the sunRNA CAGE peak is represented in all the tested cell lines, while the 5' end is less conserved, suggesting that shorter sunRNAs (like sunRNA 1-4) are more abundant than longer sunRNAs (like sunRNA-5 and sunRNA-6. (C-H) sunRNA-mediated silencing of TPP1-L prevails despite siRNA-mediated knockdown of Rrp40, Dis3 (C); Rrp6 (D); PARN (E); Mtr4 (F); LSM1, LSM8 (G); Dcp2, XRN1, or XRN2 (H). "-" indicates Empty vector control.

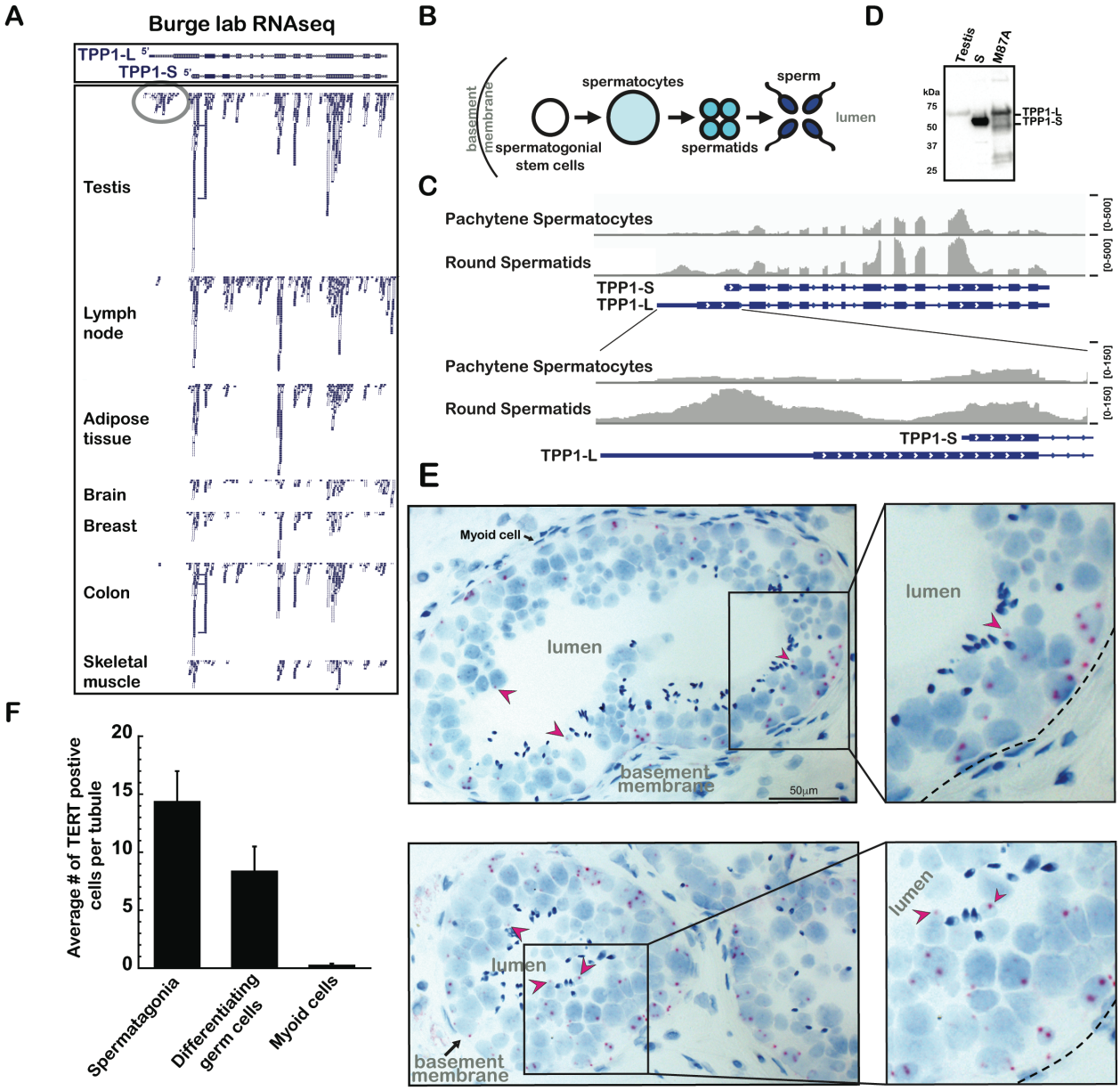

A
Burge lab RNAseq
TPP1-L 5'
TPP1-S 5'
Testis
Lymph node
Adipose tissue
Brain
Breast
Colon
Skeletal muscle
B
basement membrane
spermatocytes
sperm
lumen
spermatogonial stem cells
spermatids
D
Testis
S
M87A
kDa
75
50
37
25
TPP1-L
TPP1-S
C
Pachytene Spermatocytes
Round Spermatids
TPP1-S
TPP1-L
[0-500]
[0-150]
E
Myoid cell
lumen
basement membrane
50μm
F
Average # of TERT postive cells per tubule
20
15
10
5
0
Spermatagonia
Differentiating germ cells
Myoid cells

Figure 3.13 TPP1-L is upregulated and TERT persists throughout spermatogenesis.
(A) Deposited Burge lab RNA-Seq data from the indicated tissues/organs was visualized on the UCSC genome browser. The conspicuous abundance of TPP1-L specific reads in testes relative to other tissues is highlighted with a grey oval. (B) Schematic for the different stages of spermatogenesis. (C) Archived RNA-Seq data reveal an increase in TPP1-L abundance in human round spermatids relative to spermatocytes. (D) Immunoblot analysis using an anti-rabbit TPP1 polyclonal antibody (Bethyl Laboratories) of the chromatin-associated fraction of cells from a biopsy of a human testis reveals a band corresponding to TPP1-L. FLAG-tagged TPP1-L and TPP1-S overexpression cell extracts were run in parallel as size markers. (E) RNAscope staining for TERT mRNA in Formaldehyde Fixed-Paraffin Embedded (FFPE) sections of human testes tissue. *Blue:* hematoxylin staining; *red:* TERT mRNA. Pink arrows: TERT positive nuclei lining the lumen (indicative of meiotic spermatocytes and post-meiotic spermatids). Enlarged views of the boxed areas are shown on the right. (F) Quantitation of the average number of TERT positive nuclei per tubule for cells localized to the basement membrane (spermatogonia), cells between the basement membrane and the lumen (differentiating meiotic or post-meiotic germ cells), and non-germ (myoid) cells. For each of two biological replicates, 38 tubules were scored and the mean percentage (bar) and standard error of the mean (error bar) of TERT positive cells was plotted.

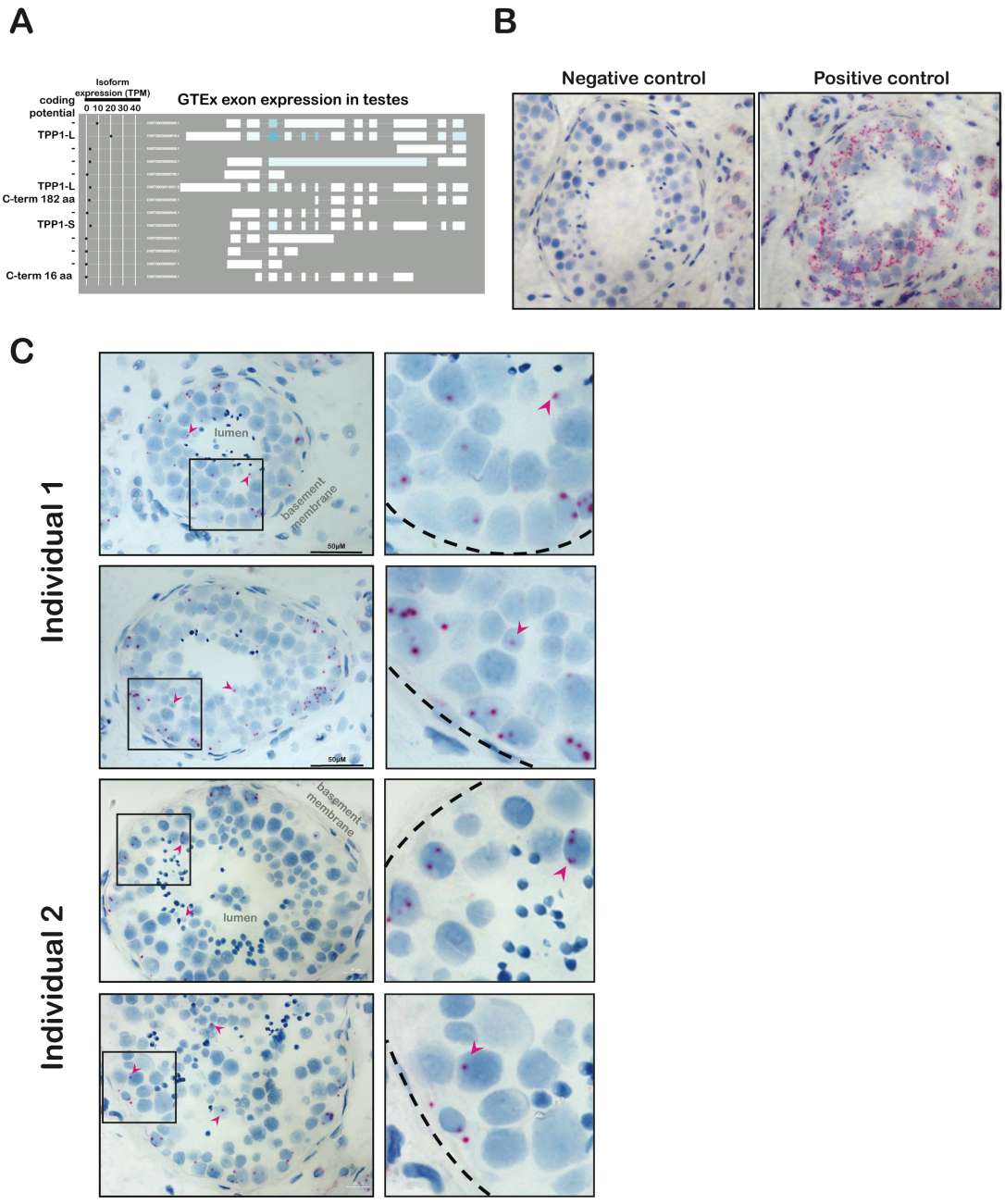


Figure 3.14 RNAscope data in human testes.
(A) Tissue/Organ-specific isoform annotation by GTex transcriptome analysis reveals a greater abundance of TPP1-L than TPP1-S mRNA in testes. (B) *Left:* Negative control for RNAscope in human testes. The probe (RNAscope® Negative Control Probe-DapB; cat. # 310043) recognizes *Bacillus subtilis* strain SMY methylglyoxal synthase (mgsA) gene, dihydrodipicolinate reductase (dapB) gene, and YpjD (ypjD) gene, which are all absent in humans. *Right:* Positive control for RNAscope in human testes. The probe (RNAscope® Positive Control Probe - Hs-PPIB; cat. # 313901) recognizes peptidylprolyl isomerase B (cyclophilin B) (PPIB) mRNA, which is moderately expressed in human cells. (C) Replicates of data shown in Figure 3.13.

CHAPTER IV

Defining the Complete TPP1-TIN2 Interface

4.1 Abstract

Telomeres are specialized nucleoprotein complexes that cap the ends of linear chromosomes to prevent their misrecognition as DNA breaks within the genome. The six-protein complex shelterin relies on a network of interactions between the individual shelterin proteins to protect telomeric DNA from undergoing unwanted ligation, resection, or recombination. TIN2 is the central mediator of these interactions, bridging the proteins that bind to double stranded telomeric DNA, TRF1 and TRF2, to the POT1-TPP1 heterodimer that binds the single stranded overhang. Here we show using a homology-based alanine scanning mutagenesis screen that the interface between TIN2 and TPP1 is larger than suggested by existing structural information. We describe this additional surface on TPP1 as the TPP1 TIN2-binding motif extension, or $TPP1_{TBM\text{-}ext}$ and find that the $TPP1_{TBM\text{-}ext}$ is essential for promoting the TIN2-TRF2 interaction.

Together, our results provide a more complete picture of the TPP1-TIN2 interface and its importance in telomere end protection.

4.2 Introduction

The ends of linear chromosomes are capped by specialized nucleoprotein complexes called telomeres. Telomeres protect chromosome ends from being misrecognized by DNA damage repair proteins as double stranded breaks in need of repair (Palm and de Lange, 2008). Telomeres are made up of two components, a repetitive DNA sequence (GGTTAG in humans) that begins with a long double stranded region and ends in a short single stranded overhang, and a six-protein complex called shelterin. Shelterin consists of six proteins (TRF1, TRF2, Rap1, TIN2, TPP1, and POT1) which together prevent the activation of the ATM and ATR kinases. To carry out this protective function, specific shelterin components and sub-complexes bind to telomeric DNA and sequester it away from DNA damage response and repair machinery, blocking DNA damage repair pathways from acting at chromosome ends (Palm and de Lange, 2008). The POT1-TPP1 heterodimer is one such shelterin sub-complex that carries out end protection by binding to the single stranded overhang of telomeric DNA. Using two N-terminal oligosaccharide/oligonucleotide (OB) domains, POT1 binds with high affinity and specificity to the G-rich single stranded overhang. However in the presence of TPP1, POT1 has an even greater affinity for the single stranded telomeric DNA (Wang et al., 2007). Consistent with this, POT1 is responsible for excluding RPA from the telomeric overhang, thus preventing the activation of ATR kinase (de Lange, 2018). TRF1 and TRF2 bind the double stranded telomeric DNA, with the latter blocking the

ATM kinase response (Broccoli et al., 1997; Denchi and de Lange, 2007). Rap1 interacts with TRF2 to increase its specificity for GGTTAG repeats (Janouskova et al., 2015). Finally, TIN2 plays a central role in shelterin by bridging the POT1-TPP1 heterodimer at the single stranded telomeric DNA to TRF1 and TRF2 at the double stranded telomeric DNA (Kim et al., 1999; Takai et al., 2011). TPP1 promotes this interaction by increasing TIN2's affinity for TRF2 (O'Connor et al., 2006), allowing TIN2 to interact simultaneously with TRF2 and TPP1 (Hu et al., 2017). Thus, TIN2 inhibits both ATM and ATR signaling at chromosome ends (de Lange, 2018).

The POT1-TPP1 heterodimer is essential for repressing the ATR mediated DNA damage response at chromosome ends (Hockemeyer et al., 2007). Two structural studies depicting the C-terminus of POT1 in complex with the POT1 binding domain (PBD) of TPP1 gave important insights into the formation of this complex (Chen et al., 2017; Rice et al., 2017). In these, the C-terminus of POT1 forms an OB domain which makes extensive contacts with the TPP1 PBD, commensurate with their high affinity interaction. Consistent with this, mutations within the TPP1-binding region of POT1 are associated with chronic lymphocytic and familial leukemias, suggesting that perturbation of this interface results in genome instability (Chen et al., 2017; Rice et al., 2017).

TPP1 not only binds POT1, but also interacts with TIN2 through its C-terminus (Palm and de Lange, 2008). A structure of the $TIN2_{TRFH}$ domain depicts how TIN2 can bind TPP1 and TRF2 simultaneously (Hu et al., 2017). In it, the TPP1 TIN2-binding motif ($TPP1_{TBM}$; aa 510-544) was shown to be necessary and sufficient for the TPP1-TIN2 interaction, while the TRF2 TIN2-binding motif ($TRF2_{TBM}$; aa 350-366) was able to bind to the $TIN2_{TRFH}$ simultaneously, forming a stable ternary complex. However, this

structure does not explain how TPP1 promotes the interaction between TIN2 and TRF2. As the $TPP1_{TBM}$ and $TRF2_{TBM}$ peptides do not directly interact in this structure, it fails to provide an obvious mechanism for the cooperativity. To fill in the gaps in our understanding of the TPP1-TIN2-TRF2 interaction, we conducted a homology directed mutagenesis screen to determine which residues on TPP1 were important for POT1 or TIN2 binding. We found that while the POT1-TPP1 interface was relatively difficult to disrupt, the TIN2-TPP1 interface was relatively fragile. We determined that the TPP1-TIN2 interface is larger than previously appreciated, and we find that the extended region of this interface is essential for TPP1-mediated cooperativity. Finally, we discover additional residues in $TIN2_{TRFH}$ which are critical for TPP1 and TRF2 binding.

4.3 Results

4.3.1 Homology-assisted site-directed mutagenesis of TPP1 reveals extended TPP1-TIN2 interaction interface

Using a homology-directed alanine scanning mutagenesis screen we engineered one single amino acid mutant and ten double mutants in TPP1's POT1-binding domain (TPP1-PBD) as well as six single amino acid mutants and six double mutants in the C-terminus of TPP1 (Figure 4.1). We conducted co-immunoprecipitation experiments with FLAG tagged TPP1 and Myc tagged POT1 or TIN2 to determine how these mutants affected TPP1 binding. Only one double mutant (L279A/L281A) in the TPP1-PBD showed reduced POT1 binding, while all other mutants pulled down POT1 to similar levels as WT TPP1 (Figure 4.2A). This is consistent with the two published structures

which depict the TPP1 PBD in complex with the POT1 C-terminus (Chen et al., 2017; Rice et al., 2017). Within the structures, residues L279 and L281 lie at the end of the TPP1 PBD alpha helix 1 which is buried in a hydrophobic pocket on the surface of the POT1 Holliday junction resolvase-like (HJRL) domain (Figure 4.1). Therefore, it is not surprising that mutation of these residues would interrupt the POT1-TPP1 interface. Continued POT1-TPP1 binding in the presence of most of the introduced mutations highlights the robustness of this interface and corroborates the notion of POT1-TPP1 being an obligate heterodimer at mammalian chromosome ends. In contrast to the extensive interface between POT1 and TPP1, the TPP1-TIN2 interface was comparatively easy to disrupt. Six different TPP1-TBD mutants completely abrogated binding to TIN2 (D496A, F500/Q501A, Y502/Y504A, P507A, R519A/L520A, L524A/W527A) and one mutant showed reduced binding (L511A) (Figure 4.2B). The residues mutated R519A/L520A and L524/W527A are observed in the protein-protein interface of the published structure of the $TIN2_{TRFH}$ domain bound to the $TPP1_{TBM}$ (Figure 4.1) (Hu et al., 2017). Residues R519 and L520 reside in the loop of the $TPP1_{TBM}$ helix-loop-helix, while residues L524 and W527 reside in helix 2. R520, L524, and W527 together form part of an extended hydrophobic interface between the $TPP1_{TBM}$ and the $TIN2_{TRFH}$ domain. The extensive contacts between side chains of these residues and the residues in helixes one and two of TIN2 contribute to the specificity of this interaction. Interestingly, residues D496, F500, Q501, Y502, Y504, and P507 are not included in the definition of $TPP1_{TBM}$ and thus are absent in the structure. However, our co-immunoprecipitation analysis indicates that these residues are indistinguishably as important as the $TPP1_{TBM}$ residues for the TPP1-TIN2

interaction, suggesting that the TPP1-TIN2 interface extends beyond $TPP1_{TBM}$ (Figure 4.2B). Therefore, we call this region the TBM extension, or $TPP1_{TBM\text{-}ext}$.

4.3.2 TPP1 enhances the TIN2-TRF2 interaction

TPP1 has previously been shown to promote the interaction between TIN2 and TRF2 (Hu et al., 2017; O'Connor et al., 2006). We asked if mutations in the $TPP1_{TBM}$ or $TPP1_{TBM\text{-}ext}$ prevent this cooperativity. Co-immunoprecipitation experiments revealed that Myc-tagged TIN2 efficiently pulled down transiently co-expressed WT TPP1 and TRF2 on anti-MYC bound beads (Figure 4.3 lane 3). However, TIN2 was unable to pull down WT TRF2 in the absence of TPP1 (Figure 4.2 compare lanes 1-3). Co-expression of $TPP1_{TBM}$ mutant R519A/L520A was unable to rescue TRF2 binding, as TIN2 did not pull down $TPP1_{R519A/L520A}$ or WT TRF2 (Figure 4.2 lane 6). Similarly, mutants in the $TPP1_{TBM\text{-}ext}$ blocked this cooperativity as TIN2 did not pull down WT TRF2 with either $TPP1_{F500A/Q501A}$ or $TPP1_{Y502A/Y504A}$ (Figure 4.2 lanes 4-5). Together these data suggest that both the $TPP1_{TBM}$ and the $TPP1_{TBM\text{-}ext}$ are critical for promoting the TIN2-TRF2 interaction.

4.3.3 Extending the definition of the TPP1-binding region of $TIN2_{TRFH}$

We asked how the $TPP1_{TBM\text{-}ext}$ interacts with the TRFH domain of TIN2. Using the solved crystal structure of the TPP1-TIN2-TRF2 complex, we conducted a second site-directed mutagenesis screen of four conserved and surface exposed residues in the TRFH domain of TIN2 proximal to the N-terminus of the $TPP1_{TBM}$ peptide (Figure 4.4A). We asked how mutation of these residues to alanine affected binding to WT TPP1 and TRF2. As expected, the previously characterized $TIN2_{A15R}$ mutation, which abrogates the interaction between TIN2 and TPP1, was unable to pull down WT TPP1

or TRF2 (Figure 4.4B lane 3). In contrast, $TIN2_{S160A}$ pulled down TPP1 and TRF2 as well as WT TIN2, suggesting that this residue is not important for the TPP1 or TRF2 interaction (Figure 4.4B lanes 2 and 7). Interestingly, mutation of either TIN2 F152 or TIN2 E153 to alanine resulted in a complete loss of TPP1 and TRF2 binding (Figure 4.4B lanes 4-5). These residues reside on the opposite face of the $TIN2_{TRFH}$ from the region that interacts with the $TPP1_{TBM}$ peptide, about 20 Å away from the N-terminus of the $TPP1_{TBM}$ (Figure 4.4C). Therefore, these residues likely contact the $TPP1_{TBM\text{-}ext}$ but not the $TPP1_{TBM}$. Finally, mutation of residue W198 of $TIN2_{TRFH}$ resulted in a partial loss of TPP1 and TRF2 binding (Figure 4.4B lane 7). Together these data expand on our understanding of the TIN2-TPP1 interface and suggest that it is more extensive than suggested by previous studies.

4.4 Discussion

Using a homology directed mutagenesis screen we discovered that the TPP1-TIN2 interface is larger than previously appreciated. We found that TPP1 amino acids 496-507, as well as the previously characterized amino acids 510-544 of the $TPP1_{TBM}$, are critical for facilitating the interaction between TPP1 and TIN2. Therefore, we call this region the TPP1 TIN2-binding motif extension, or $TPP1_{TBM\text{-}ext}$. The $TPP1_{TBM\text{-}ext}$ is not only critical for TIN2 binding, but it is also essential for bolstering the TIN2-TRF2 interaction. Finally, we discover additional residues within TIN2 which likely interact with the $TPP1_{TBM\text{-}ext}$.

Previous studies have implicated TPP1 in promoting the TIN2-TRF2 interaction (O'Connor et al., 2006), yet a molecular mechanism for this cooperativity remains unknown. Future structural studies which aim to characterize the full TPP1-TIN2

interaction in the absence of TRF2 as well as the complete TPP1-TIN2-TRF2 complex may give significant insight into the mechanistic underpinnings of this cooperativity. However, the TPP1-TIN2 interface is only half of this mechanism. As a result of a TIN2-TPP1 interaction, the $TIN2_{TRFH}$-TRF2 interaction makes up the second half of this cooperative binding. Future studies should work to biochemically characterize this shelterin interface in detail, determining the minimum portion of TRF2 that is needed to confer cooperativity as well as the change in affinity that occurs between TIN2 and TRF2 upon TPP1 binding.

We found that TIN2 mutations at F152 and E153 abrogate TPP1 binding. These residues do not bind the $TPP1_{TBM}$ or $TRF2_{TBM}$ in the published structure, but instead point outward on the opposite face of the $TIN2_{TRFH}$ structure from the surface which interacts with the $TPP1_{TBM}$ peptide. Future structural studies will determine if these residues interact with the $TPP1_{TBM\text{-}ext}$ or an uncharacterized portion of the TPP1-TIN2 interface. Although the $TPP1_{TBM}$ and $TRF2_{TBM}$ do not directly interact, it is intriguing to envision that a larger TPP1-TIN2-TRF2 complex may exhibit further structural contacts that have not previously been characterized. TIN2 residue E153 is less than 10 Å from the C-terminus of the $TRF2_{TBM}$ peptide. If the $TPP1_{TBM\text{-}ext}$ interacts with $TIN2_{E153}$, it is reasonable to propose that the $TPP1_{TBM\text{-}ext}$ would be within close proximity to TRF2, invoking a protentional direct interaction between the two proteins when they are both bound to TIN2. This would not only give a molecular mechanism to their cooperativity, but it would also describe a new shelterin interaction.

Together, these data describe an additional segment of the TPP1-TIN2 interface and suggest novel interfaces within shelterin that have evaded previous inspection.

Future studies which characterize each shelterin interaction will be essential for deciphering the importance of these shelterin subcomplexes in telomere function.

4.5 Materials and methods

4.5.1 Molecular cloning and site-directed mutagenesis

All TPP1 mutations were introduced into the p3x-FLAG-TPP1-cDNA6/Myc-HisC vector using QuikChange® Site-Directed Mutagenesis Kit (Agilent Technologies) and complementary mutagenic primers (Integrated DNA Technologies). The resulting FLAG-TPP1 plasmids were sequenced to confirm the presence of the intended mutation and the absence of errors that may have been introduced during PCR amplification. 3x-FLAG-tagged TRF2 and 6x-MYC-tagged POT1 and TIN2 for human cell expression were cloned into the pTET-IRES-eGFP-Bi4 vector for use in co-immunoprecipitation experiments. Additionally, 6x-MYC-tagged TIN2 and 3x-FLAG-tagged TIN2 and TRF2 were cloned into a pcDNA3 derived vector. This pcDNA3 derived 3x-FLAG-tagged TIN2 vector was then used for subsequent site-directed mutagenesis exactly as described above to generate FLAG-TIN2$_{A15R}$, FLAG-TIN2$_{F152A}$, FLAG-TIN2$_{E153A}$, FLAG-TIN2$_{S160A}$, and FLAG-TIN2$_{W198A}$.

4.5.2 Co-immunoprecipitation

Co-immunoprecipitation experiments were performed exactly as described previously (Grill et al., 2018). Briefly, HeLa-EM2-11ht cells were transfected with 1 µg of each plasmid. 24-48 hours after transfection, cells were trypsinized, resuspended in medium containing 50% FBS, and pelleted. Cells were then resuspended in 400 µL of

lysis buffer (50 mM Tris-Cl (pH 7.6), 20% glycerol, 1 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 0.02% SDS, 1 mM dithiothreitol, 2 mM phenylmethylsulfonyl fluoride, complete protease inhibitor cocktail [Roche]) and kept on ice. 33 µL of 4 M NaCl, and 433 µL of water was added and lysates were spun down using centrifugation. 40 µL of supernatant was then added to SDS gel loading buffer and kept aside for analysis of input samples. Remaining lysate was used directly for immunoprecipitation. For FLAG immunoprecipitation, lysate was added to 30 µL of pre-washed anti-FLAG M2 affinity gel (Sigma; A2220) and nutated overnight at 4°C. For Myc immunoprecipitation, 5 µL of c-Myc antibody (DSHB; 9E 10) was added and lysates were incubated for 2-4 hours at 4°C. After antibody incubation, lysate was transferred to tubes containing 30 µL of pre-washed protein A/G agarose (Pierce; 20421) and nutated overnight at 4°C. After overnight incubation, beads were washed and protein was eluted from the beads by adding 60 µL of 2X SDS gel loading buffer. All samples were analyzed by SDS-PAGE followed by immunoblotting with HRP conjugated FLAG or Myc antibodies.

4.5.3 Immunoblotting

Immunoblotting was performed using standard procedures and exactly as described previously (Grill et al., 2018). The following antibodies were used for detection with chemiluminescence by ECL plus reagents (Pierce ECL Western Blotting Substrate; Thermo Scientific): mouse monoclonal anti-FLAG M2-HRP conjugate (Sigma; A8592; 1:10,000) and mouse monoclonal anti-c-Myc (9E10) HRP conjugate (Santa Cruz; sc-40 HRP; 1:10,000). The data were visualized using a gel-documentation system (ChemiDoc™ MP System; BioRad).

4.5.4 HeLa cell culture

HeLa-EM2-11ht cells were cultured exactly as described previously (Grill et al., 2018) at 37°C in the presence of 5% CO_2 and propagated in modified DMEM (Dulbecco's Modified Eagle Medium; Gibco 11995-065) medium containing 100 U/mL penicillin, 100 µg/mL streptomycin, and 10% fetal bovine serum.

4.6 Figures and Tables

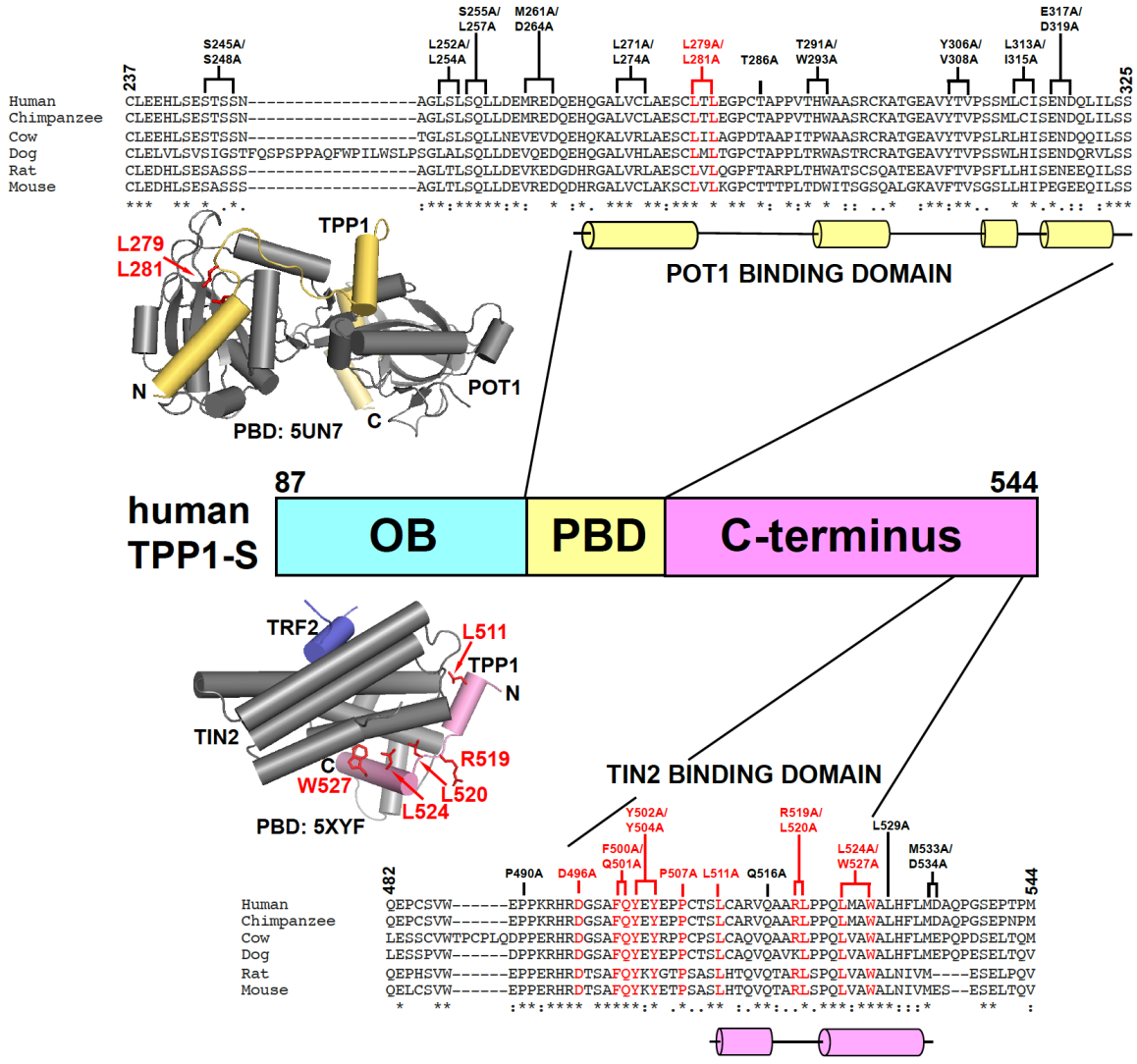


Figure 4.1 Mutagenesis screen to identify the TPP1-POT1 and TPP1-TIN2 interaction surfaces. Sequence alignment of the human TPP1 POT1- and TIN2-binding domains along with additional mammalian orthologs. Mutated residues are shown in red. Brackets indicate two residues simultaneously mutated (double mutant) and amino acid numbers represent the human sequence. Asterisks, colons, and periods beneath the sequence lineups represent identical residues, strongly conserved residues, and weakly conserved residues respectively, as described by the MUSCLE algorithm. Predicted α helices are shown as cylinders underneath the sequence alignment as predicted by SWISS-MODEL (Waterhouse et al., 2018). The structure of the POT1 C-terminus bound to the TPP1 PBD (PBD: 5UN7) is shown above the TPP1 domain diagram with POT1 shown in grey and TPP1 shown in yellow. TPP1 mutant L179A/L181A is highlighted in red. Structure of the $TIN2_{TRFH}$-$TPP1_{TBM}$-$TRF2_{TBM}$ complex (PBD: 5XYF) is shown below the TPP1 domain diagram with TIN2 represented in grey, $TRF2_{TBM}$ represented in purple, and $TPP1_{TBM}$ represented in pink. $TPP1_{TBD}$ mutants are highlighted in red.

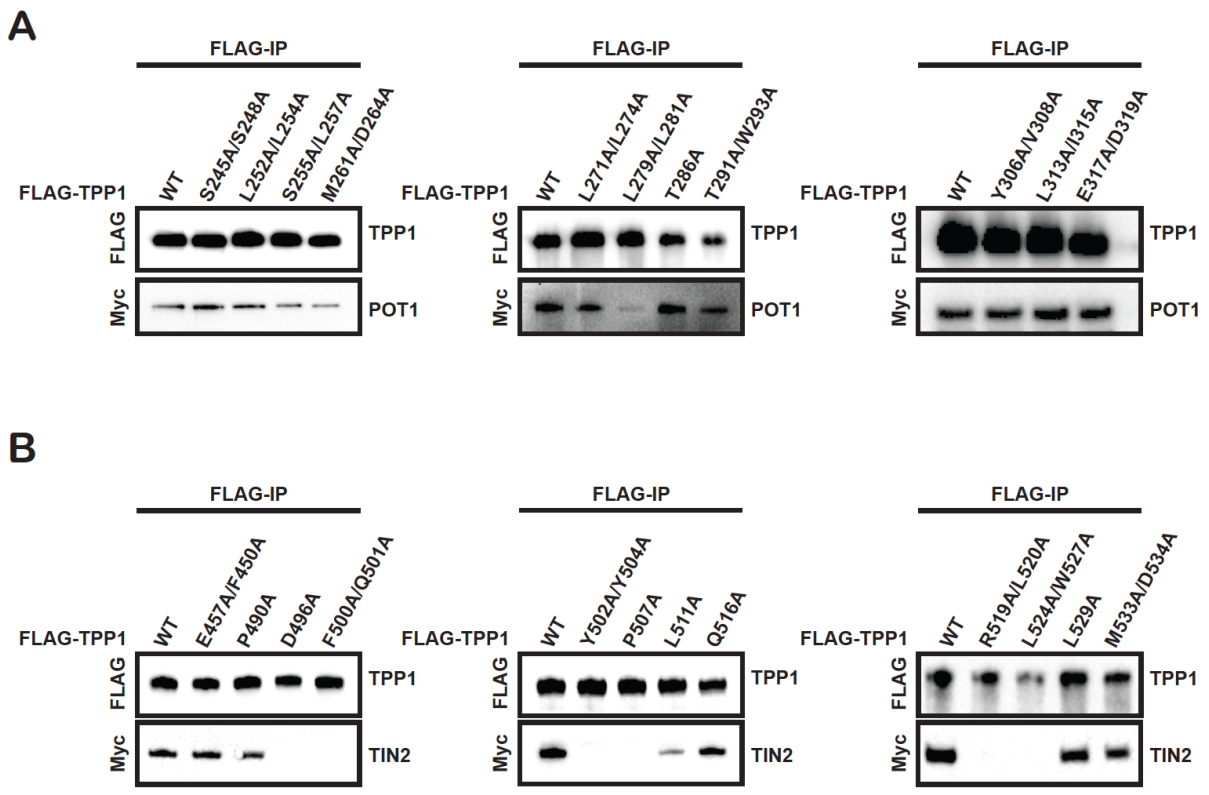


Figure 4.2 Mutations in the TPP1$_{PBD}$ and the TPP1 C-terminus disrupt POT1 or TIN2 binding. A) Pulldown of transiently expressed FLAG-TPP1$_{PBD}$ mutants on anti-FLAG conjugated beads with Myc-POT1. B) Pulldown of transiently expressed FLAG-TPP1 C-terminus mutants on anti-FLAG conjugated beads with Myc-TIN2.

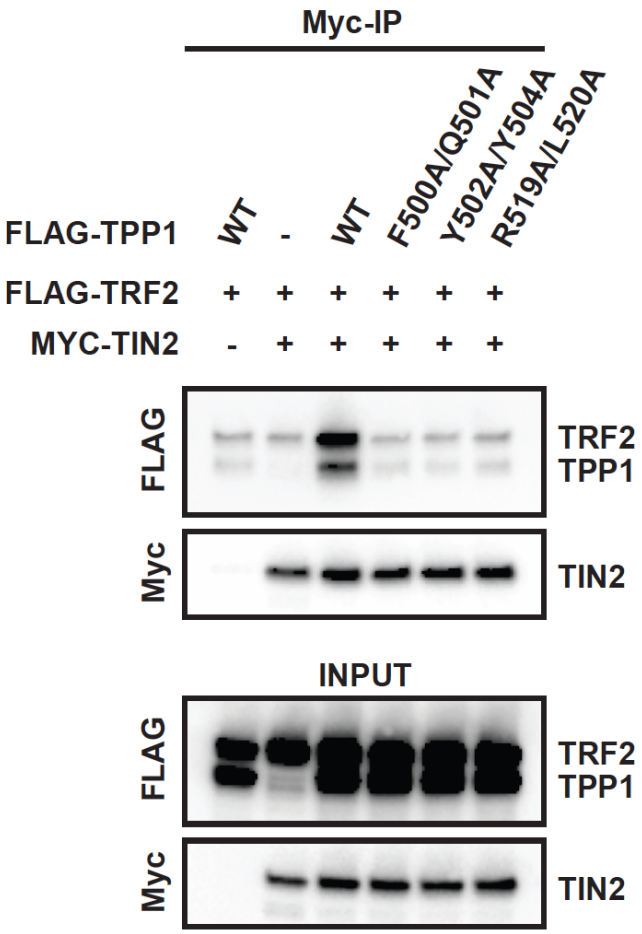


Figure 4.3 Mutations in the $TPP1_{TBM}$ and $TPP1_{TBM\text{-}ext}$ disrupt cooperativity of the TPP1-TIN2-TRF2 interaction.
Anti-Myc antibody bound Myc-TIN2 was pulled down on protein A/G agarose beads with transiently expressed FLAG-TRF2 and indicated FLAG-TPP1 construct.

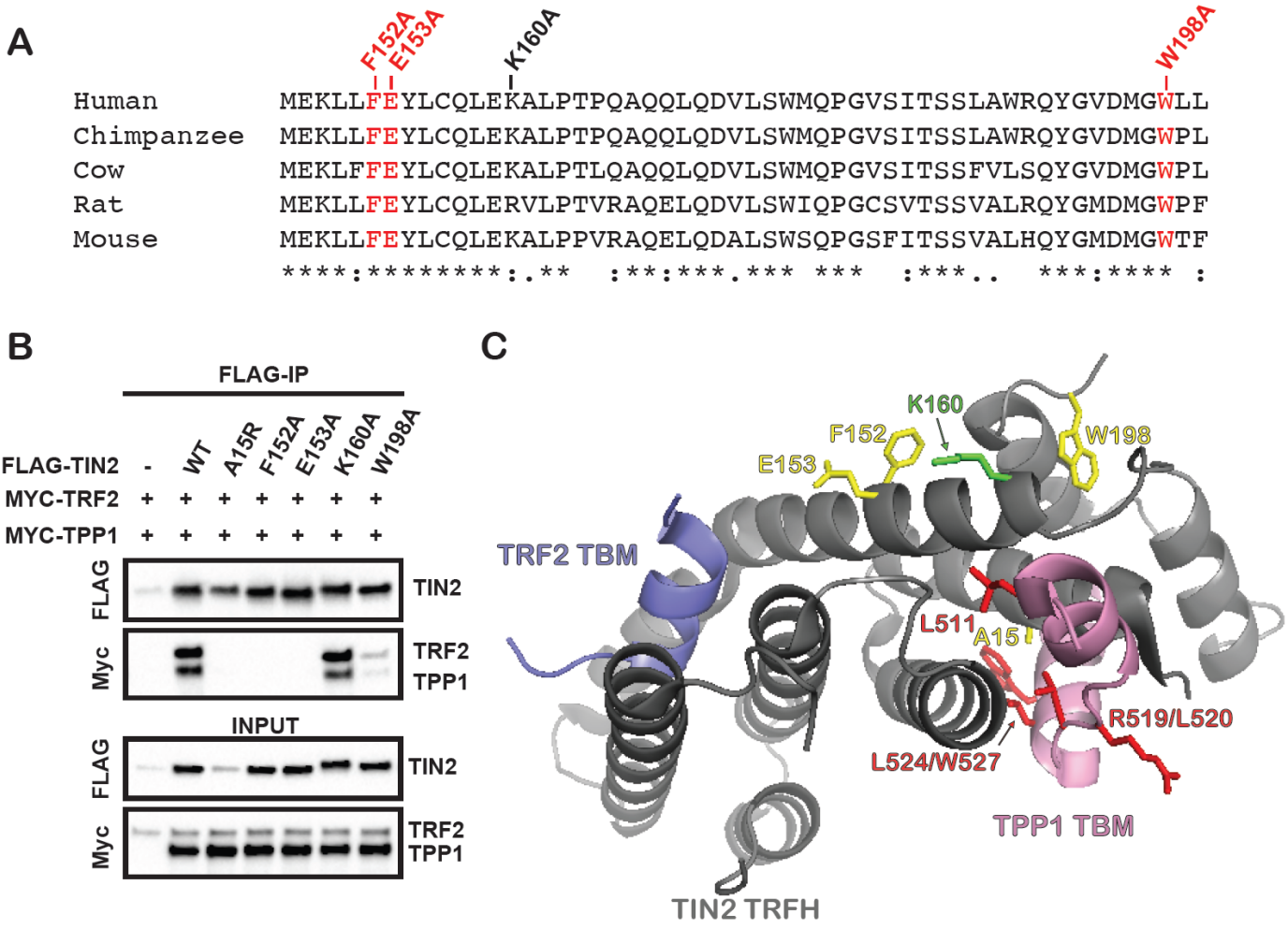


Figure 4.4 Mutations in TIN2 alpha helix 7 result in loss of TPP1 and TRF2 binding.
A) Sequence conservation of the TIN2 TRFH domain. Residues examined in this study are labeled above the aligned sequence with red denoting amino acids whose mutation impaired the TPP1-TIN2 interaction and black denoting amino acids whose mutation did not affect TPP1-TIN2 binding. B) Pulldown of indicated FLAG-TIN2 construct on anti-FLAG conjugated beads with WT Myc-TRF2 and Myc-TPP1. C) Structure of the $TIN2_{TRFH}$ (grey)-$TPP1_{TBM}$ (pink)- $TRF2_{TBM}$ (lilac) complex (PDB: 5xyf; (Hu et al., 2017)). TIN2 residues whose mutation impact binding to TPP1 and which likely contact the $TPP1_{TBM-ext}$ are shown in yellow, $TIN2_{K160A}$ is shown in green, and $TPP1_{TBM}$ mutants are shown in red.

CHAPTER V

Concluding Remarks and Future Directions

5.1 Overview

In this thesis, I have introduced the mechanisms by which telomerase and the shelterin protein complex solve the end protection and end replication problems in human cells (chapter one). I have discovered a region on the N-terminus of the OB domain of TPP1 that is critical for telomerase action and contributes to the species-specificity of the TPP1-telomerase interaction (chapter two). I have discussed the properties of two separation-of-function isoforms of TPP1 and shown how these isoforms regulate telomerase in somatic and germ cells (chapter three). Finally, I have detailed the POT1 and TIN2 binding surfaces on TPP1 and identified the entire TPP1-TIN2 interface that confers cooperativity with TRF2 (chapter four). In this final chapter, I reflect on these findings and extrapolate on their implications, highlighting unresolved questions and suggesting future directions.

5.2 The N-terminus of the OB domain of TPP1 and its role in telomerase action

In chapter two I discussed the importance of the N-terminus of the OB (NOB) of TPP1 for telomerase function. Deletion of the NOB drastically reduces TPP1-mediated telomerase processivity *in vitro*, results in loss of telomerase recruitment to the telomere, and shortens telomeres in cells. Unlike the TPP1 TEL patch, which is conserved between human and mouse TPP1, the NOB region in mice is distinct. Simply swapping out the mouse NOB for that of the human allows mouse TPP1 to stimulate human telomerase, suggesting that the NOB region of TPP1 contributes to the species-specificity of the TPP1-telomerase interaction. Mapping the NOB region and the TEL patch on the TPP1 OB crystal structure suggests that these regions exist as two distinct binding clefts that interact with telomerase. Together, these distinct surfaces on TPP1 are critical for telomerase function.

The primarily acidic TEL patch has already been implicated in binding the highly basic TEN domain of TERT (Schmidt et al., 2014), thus we initially speculated that the primarily hydrophobic NOB may contact a similarly hydrophobic portion of telomerase. However our view of this interaction has evolved with the publication of the 4.8 Å cryo-EM structure of *Tetrahymena* telomerase (Jiang et al., 2018). Prior to this structure, the 9 Å cryo-EM *Tetrahymena* structure showed that the unmodeled region ascribed to the insertion in fingers domain (IFD), which has previously been implicated in telomerase recruitment to the telomere, is located proximal to the TEN domain and p50, the *Tetrahymena* ortholog of human TPP1. In the higher resolution structure, the modeled IFD coalesces with the TEN domain to form a single interaction surface. Although solved at lower resolution, the most recent human telomerase cryo-EM structure is

consistent with the *Tetrahymena* structure, depicting similarly juxtaposed TEN and IFD domains. Together these structural models and biochemical data suggest that the NOB and TEL patch may reorient to bind TERT not as two individual interactions, but instead as a single surface that contacts both the TEN domain and IFD of TERT. Future structural and functional studies focused on determining the residues on TERT that are important for the telomerase-TPP1 interaction will be essential for understanding this multifaceted interface.

The TPP1-telomerase interaction is species-specific, as mouse TPP1 cannot stimulate human telomerase (Zaug et al., 2010). Simply swapping in the human NOB allows mouse TPP1 to stimulate human telomerase, albeit not to wild-type levels. This partial rescue suggests that other residues in TPP1 that contribute to this species-specificity remain to be discovered. By simply aligning the mouse and human TPP1 protein sequences and determining which residues are mouse-specific, swapping them with the equivalent human residues, and determining their effect on human telomerase stimulation, we may be able to pinpoint other residues that are essential for the human TPP1-telomerase interaction. Importantly, this same process can be adopted to probe for residues in TERT that contribute to species-specific TPP1 binding. By exploiting these differences, along with structural studies, we can define the entire human TPP1-telomerase interface.

Since our publication defining the NOB region of TPP1, a study was published describing two unrelated individuals with dyskeratosis congenita-like phenotypes with novel homozygous mutations in the NOB (Tummala et al., 2018). Both patients had very short telomeres, with one patient's telomere lengths falling below the first percentile

for his age. Similar to DC patients with the TPP1ΔK170 mutation, patient's with NOB mutations also exhibit genetic anticipation, as this patient's parents also had short telomeres. Together with our biochemical and cell-based studies, these clinical data underscore the significance of the NOB region as being essential for telomerase action and provide evidence of its biological importance in human health.

5.3 The separation-of-function of TPP1-L and TPP1-S and its effect on germline biology

In chapter three I examined two isoforms of TPP1: TPP1-S which initiates at methionine 87, and TPP1-L which initiates at methionine one. Although TPP1-L and TPP1-S differ in only 86 amino acids at their N-terminus, overexpression of TPP1-S in human cells leads to telomere hyperelongation while overexpression of TPP1-L results in telomere shortening. I found that three arginine residues in the N-terminus of TPP1-L specifically restrict telomerase activation at telomeres. While TPP1-S is the major isoform in all somatic cells, TPP1-L is specifically upregulated in the differentiating germ cells of the testes, where telomerase expression persists.

To my knowledge this is the first report of a TPP1 variant which can fully recruit telomerase to the telomere but leads to a loss of telomerase activity. While I show that three arginine residues specific to TPP1-L are essential for this loss, mutation of them to alanine does not fully rescue telomerase activity or telomere length back to TPP1-S levels. Consequently, there are likely to be other residues within the first 86 amino acids of TPP1-L that are also important for blocking telomerase action. Interestingly an N-terminal FLAG tag on TPP1-S was recently shown to be detrimental to telomere elongation, suggesting that any bulk that is N-terminal to the TPP1 OB domain will

negatively impact TPP1's ability to stimulate telomerase processivity (Sandhu et al., 2019). However our finding that mutation of the three arginine residues partially rescues this phenotype suggests that bulk is not the only factor mediating this process. Future studies should aim to pinpoint the residues that are important for this inhibition and fully characterize the region of TPP1-L that is essential for this function.

The mechanism by which TPP1-L inhibits telomerase remains to be elucidated. Future structural analysis of TPP1-L bound to telomerase would greatly facilitate our understanding of this inhibition. The recently solved cryo-EM structure of *Tetrahymena* telomerase bound to the TPP1 ortholog p50 reveals that the N-terminus of p50 is located in a pocket bordered by components of TERT (including the TEN domain, IFD, and catalytic core) as well as the TEB proteins that are essential for processivity of telomerase in *Tetrahymena* (Figure 5.1) (Jiang et al., 2018). The size of this pocket is small, with the C-α backbone of the N-terminus of p50 only 8 Å away from the TEN domain or the IFD, and 20 Å from the telomerase RNA or the catalytic core of TERT. Based on the *Tetrahymena* structure, we created a homology model of human telomerase bound to the human TPP1 OB domain. In this homology model, the N-terminus of the TPP1 OB domain is similarly situated adjacent to the TEN-IFD interface. Given our current understanding of how TPP1 and telomerase interact, it is hard to envision how the N-terminal 86 amino acids specific to TPP1-L could be accommodated in the current structural model. I propose that the N-terminus of TPP1-L is sterically precluded from residing in this cavity between TPP1 and telomerase. Instead, I predict that the structure of TPP1-L bound to telomerase would be distinctly different from our current model of the TPP1-S-telomerase interaction. Nonetheless, TPP1-L recruits

telomerase to telomeres with equal efficiency as TPP1-S, implying that the TEL patch and NOB regions of TPP1-L remain able to interact with their corresponding binding partners on TERT. While this insinuates that the TPP1 OB domain would be positioned equivalently in both scenarios, a true understanding of the similarities and differences in how TPP1-S and TPP1-L bind and regulate telomerase will have to await the structural characterization of both complexes.

We found that TPP1-S is the major isoform in all somatic cells and embryonic stem cells, while TPP1-L is specifically expressed in the differentiating germ cells of the testes. Our analysis in HeLa cells suggests that TPP1-L specifically blocks telomerase action at telomeres, however more work is needed to fully understand if this function translates to cells in the testes. The first step in understanding this function is to confirm the expression and regulation of telomerase in these cells. Here we show that TERT mRNA persists throughout spermatogenesis, however follow-up experiments should determine the amount of functional telomerase RNP that persists after differentiation of spermatagonial stem cells. While TERT mRNA expression can be used as a crude indicator of functional telomerase, experiments which detect telomerase activity are essential for clearly determining the amount of assembled and active telomerase that exists in any given system. Biochemical experiments which probe telomerase expression and activity throughout spermatogenesis could give great insight into the role of telomerase in differentiating germ cells. If telomerase expression remains throughout spermatogenesis, yet telomerase activity is reduced in spermatids and other differentiating germ cells, this could evoke the role of TPP1-L in preventing telomerase activity in these cells.

While a decrease in telomerase activity would provide significant correlative evidence that TPP1-L blocks telomerase in germ cells, true causative evidence would require cell specific knockout of TPP1-L. Hypothetically, if we were able to edit the genome of human spermatogonial stem cells to prevent expression of TPP1-L, we could monitor the telomere lengths of edited cells as they differentiate. While we would expect the telomere lengths in wild-type cells to stay relatively stable, those cells which lack TPP1-L would show telomere length increases throughout differentiation. This would not only provide direct evidence of TPP1-L's function in germ cells, but it would also provide additional credence for continued telomerase expression throughout spermatogenesis. Interestingly, human sperm telomere length has been shown to increase with age (Antunes et al., 2015). This telomere elongation is especially unusual in nature, as even stem cells which stably express telomerase only just maintain telomere length. While the mechanism for this is unknown, it is appealing to hypothesize that consistent telomerase expression during spermatogenesis would allow for modest telomere length increases over the course of a species' lifetime. TPP1-L may function to block hyperelongation of these telomeres, as excessive telomere elongation could compromise telomere stability (Rivera et al., 2017), but low-levels of TPP1-S protein in these cells (translated perhaps from the TPP1-L mRNA itself) may be enough to lead to the relatively small increase in telomere length seen in sperm as men age. Unfortunately, current technical challenges have prevented the further characterization of TPP1-L and telomerase in human testes. These experiments require accurate sorting of human testes cells to separate spermatagonial stem cells from the differentiating germ cells, a process that is not currently optimized for the

human system. Once sorted, the spermatagonial stem cells would need to be cultured and manipulated for genome editing, cell biology, or biochemistry. Even if these experiments were technically feasible, a large amount of fresh human sample would be needed to even begin answering these questions. Because of these challenges, it might be most prudent for future research to focus on TPP1-L and telomerase in the germ cells of other primates, such as monkeys.

Interestingly, while telomerase activity is readily detectable in spermatagonial stem cells, no telomerase activity is detected in human sperm (Wright et al., 1996). Yet our studies demonstrate that TERT expression persists in the differentiating germ cells of the testes, suggesting that telomerase may be present, but not active, in these cells. One potential explanation for this phenomenon would be the presence of TPP1-L, and possibly telomerase, at chromosome ends in human sperm cells. This would allow telomerase to remain bound to TPP1-L throughout spermatogenesis, while preventing telomerase action at telomeres in both the actively differentiating germ cells and the mature sperm. This telomerase would then be poised at chromosome ends in sperm to elongate telomeres once fertilization takes place. Curiously, ninety percent of the human spermatid genome is packaged by protamines, small arginine-rich proteins that condense the genome into a genetically inactive state (Balhorn, 2007). However telomeric DNA is part of the ten percent of the human spermatid genome which is still packaged by histones (Zalenskaya et al., 2000). This suggests that while the majority of the human spermatid genome is bound only by protamines, shelterin could still reside on the histone packaged telomeric DNA of mature sperm. Future studies should

determine if shelterin stays bound to telomeric DNA in human sperm, and if telomerase continues to interact with TPP1-L in this context.

It is interesting to consider the reasons why TPP1-L may have evolved to function specifically in the germ cell niche. The absence of TPP1-L and the lack of telomerase regulation in mice gives credence to the idea that TPP1-L became an evolutionary necessity only when telomerase regulation became essential for long lived mammals. While mice express telomerase in all cells, higher primates must block telomerase expression in most somatic cells to prevent unwanted cell division. Germ cells are no exception to this, as unwanted cell division during oogenesis or spermatogenesis could lead to the same cancerous outcomes as unwanted cell division in somatic cells. Yet germ cells must contend with an added complication. They must be able to quickly reactivate telomerase after fertilization to support the replicative capacity of an emerging embryo. Thus, we hypothesize that telomerase is not shut down in germ cells, but instead temporarily restricted from acting at telomeres until fertilization. TPP1-L elegantly solves this problem by blocking telomerase action while still allowing telomerase expression to remain unchanged. Our study focuses solely on this process in testes, however future studies should determine if TPP1-L is similarly upregulated during oogenesis to block telomerase action in ovaries. Understanding these processes in spermatogenesis and oogenesis has far reaching implications in both telomere biology and human health. Does mutation or repression of TPP1-L result in a greater propensity for germ cell tumors? Does increased expression of TPP1-L in germ stem cells result in telomere shortening and a decrease in reproductive capacity? Understanding the TPP1-telomerase relationship in this biological niche can provide

unparalleled insight into human reproductive telomere biology that has been greatly under studied until now.

5.4 The end-protection surfaces of TPP1

In chapter four I conducted a homology directed mutagenesis screen to determine the residues on TPP1 that are important for POT1 or TIN2 binding. I engineered eleven mutants in the POT1-binding domain (PBD) of TPP1 and twelve mutants in the TIN2-binding domain (TBD) of TPP1. Only one double mutant in the PBD affected the POT1-TPP1 interaction, while six mutants in the TBD completely abrogated the TPP1-TIN2 interaction. These mutants show that the TPP1-TIN2 interface is larger than previously appreciated. This region is not only important for TIN2 binding, but it also enhances the interaction between TIN2 and TRF2. Mutants in both the newly discovered portion of the interface, as well as the previously characterized portion of the interface, prevent this cooperativity. Using the published structure of the TPP1-TIN2-TRF2 complex, I conducted a second mutagenesis screen that targeted conserved surface exposed residues in TIN2 that were proximal to the known TPP1 binding region. I found three TIN2 residues that are essential for TPP1 binding but were not previously implicated in the TPP1-binding region of TIN2.

TPP1 has three known protein interaction domains, the OB domain which binds telomerase, the PBD which binds POT1, and the TBD which interacts with TIN2. Previous mutagenesis of the TPP1 OB domain led to the discovery of the TEL patch, which is the surface on TPP1 that is responsible for interacting with TERT (Nandakumar et al., 2012). In both the OB domain and the TBD, multiple single or double mutants were capable of disrupting the interaction between TPP1 and either TERT or TIN2. Of

the fourteen OB mutants that were analyzed, eight resulted in a defect in TERT binding, while six mutants in the TBD resulted in loss of TIN2 binding. These screens suggest that both the TPP1-TERT interface and the TPP1-TIN2 interface are relatively easy to disrupt. In stark contrast, the POT1-TPP1 interaction was extremely difficult to disrupt, as only one of the eleven mutants showed even a partial defect in POT1 binding. This extensive interaction surface is consistent with POT1-TPP1 acting as an obligate heterodimer in cells.

While it is reasonable to expect the TPP1-TERT interaction to be relatively weak because of its transiency, it is less obvious as to why the TPP1-TIN2 interaction is similarly easy to disrupt. In cells, both TPP1 and POT1 rely on TIN2 for telomere localization, making this interaction essential for both the end protection function of shelterin and the end replication function of TPP1. Deletion of either TIN2 or POT1 results in a similar ATR mediated DNA damage response at telomeres, suggesting both proteins play essential roles in the shelterin complex. Yet the POT1-TPP1 interaction surface is significantly more extensive than that of TPP1-TIN2. One explanation for this disparity is the need to form of shelterin sub-complexes at chromosome ends. TIN2 is significantly more abundant at telomeres than either POT1 or TPP1, suggesting that a majority of TIN2 molecules at the telomere are not bound to the TPP1-POT1 heterodimer (Takai et al., 2010). In contrast, POT1 and TPP1 have a similar abundance, supporting the idea of an obligate POT1-TPP1 heterodimer at the single stranded tail, which represents only a small fraction of the total telomeric DNA. While the extensive interaction between POT1 and TPP1 may preclude formation of any shelterin sub-complexes that deplete TPP1 from the single stranded telomeric DNA, the

weaker interaction between TPP1 and TIN2 may support the formation of shelterin sub-complexes which allow TIN2 to reside at the double stranded telomeric DNA without POT1 or TPP1.

The high affinity interaction between POT1 and TPP1 is easily explained by the need for TPP1 to increase POT1's stability on single stranded telomeric DNA, as well as recruit telomerase to the single stranded telomeric DNA end. Additionally, the weaker TIN2-TPP1 interaction may be central to the formation of shelterin sub-complexes on both the single stranded and double stranded portions of the telomere. However the underlying need for cooperativity between TRF2, TIN2, and TPP1 is more difficult to explain. This cooperativity suggests that the formation of a TRF2-TIN2-TPP1 sub-complex is favored over a TIN2-TRF2 sub-complex. However TRF2 has not been implicated in any end replication function and has not been shown to be important for the telomeric localization of TIN2, TPP1, or POT1. The main function of TRF2 is to prevent an ATM mediated DNA damage response at telomeres by binding to the double stranded telomeric DNA, yet it can fully perform this function in the absence of TPP1. Thus, it is unclear why a TRF2-TIN2-TPP1 complex would be preferential for end protection or end replication.

Despite TPP1 being sufficient for enhancing the TRF2-TIN2 interaction, a TRF2-TIN2-TPP1 sub-complex is unlikely to exist at the telomere. Because TPP1 is rarely found without POT1, and similarly TRF2 is thought to always interact with Rap1, the likely shelterin sub-complex that would form is that of Rap1, TRF2, TIN2, TPP1 and POT1. This complex could bridge both the single stranded and double stranded regions of the telomeric DNA through binding of POT1 and TRF2. Interestingly, this importance

of the single stranded/double stranded intersection has also been noted in the formation of t-loops, which are dependent on TRF2 binding the double stranded telomeric DNA near the junction (Stansel et al., 2001). Therefore, it is intriguing to ask if this shelterin sub-complex may be important for the formation or resolution of t-loops, which would provide a binding site for both TRF2 and POT1 in close proximity. In support of this theory, POT1 also contains a Holliday junction resolvase-like domain which may be important for resolution of T-loops, however currently there is no biochemical evidence to suggest that this domain retains resolvase activity (Chen et al., 2017; Rice et al., 2017).

Finally, it is intriguing to speculate about the importance of these shelterin sub-complexes in the recruitment and processivity of telomerase. Although TPP1 is the only shelterin protein known to directly recruit telomerase to the telomere, how telomerase specifically finds the far 3' end of the single stranded tail remains unknown. The single stranded telomeric DNA spans 300 nucleotides on average in humans, allowing for the binding of up to 25 POT1-TPP1 heterodimers. To add to this, an average of 65 POT1-TPP1 heterodimers are found at human telomeric DNA (Takai et al., 2010), suggesting that some of these reside in the double stranded region. Each of these TPP1 molecules can, in principle, recruit telomerase to its location on the telomeric DNA. Yet telomerase must find the far 3' end in order to extend the telomeres. It is possible that this process occurs through simple one-dimensional diffusion of telomerase as it moves from one TPP1 to another along the length of telomere until it contacts its substrate at the 3' end. However the relative sparsity of telomerase compared to telomere ends and the short time frame with which telomerase must act suggests that this process is not mediated

solely by diffusion from one TPP1 to the next, but instead may be a targeted process by which telomerase is recruited specifically to the far 3' tail of the telomere. This could be achieved through the formation of larger shelterin sub-complexes, such as the TRF2-TIN2-TPP1 based sub-complex, which could have a greater impact on telomerase recruitment than TPP1 alone. Indeed, a complex containing Rap1, TRF2, TIN2, TPP1, and POT1 increased stimulation of telomerase processivity *in vitro* compared to POT1-TPP1 alone (Lim et al., 2017), suggesting that larger shelterin sub-complexes may be important for mediating telomerase recruitment and processivity.

Although it is possible for shelterin to exist as a six-membered complex, it is indisputable that shelterin sub-complexes must also form on the telomeric DNA. Previous studies in TPP1 have determined separation-of-function mutants which disrupt the TPP1-TERT interaction but leave the TPP1-POT1 and TPP1-TIN2 interactions unaffected (Nandakumar et al., 2012). Now, with the additional mutagenesis screens I have conducted, it will be possible to probe each individual protein binding surface on TPP1. Future research should be dedicated to finding and characterizing separation-of-function mutations within every shelterin interaction surface. Using this detailed analysis, researchers can parse out the end-protection and end-replication functions of each of these important proteins and their effect on telomere biology.

5.5 Figures and Tables

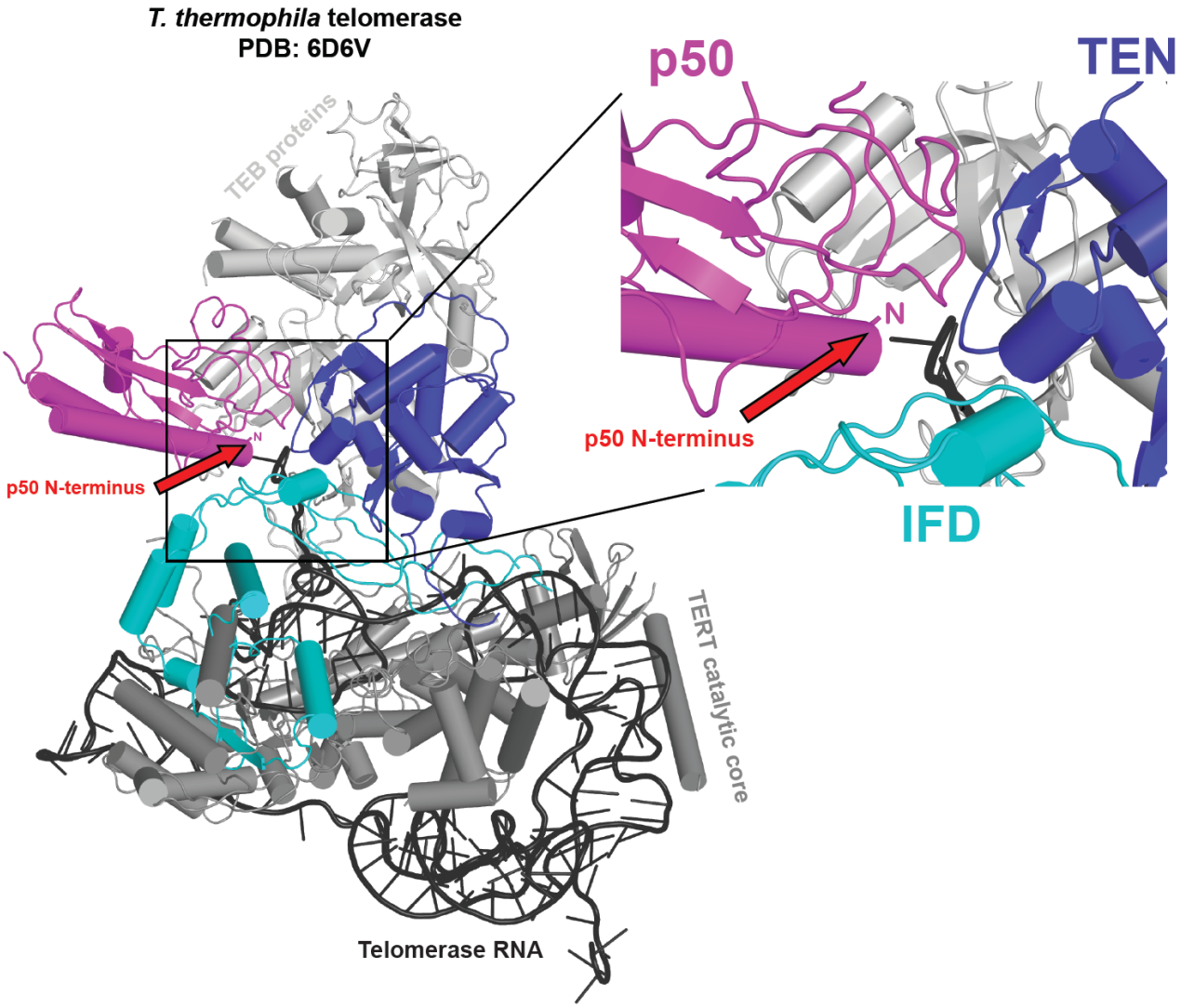


Figure 5.1 Structure of *Tetrahymena thermophila* telomerase.
Depiction of the cryo-EM structure of *Tetrahymena* telomerase, with an enlarged view of the boxed area shown on the right. p50, the TPP1 ortholog in *Tetrahymena*, is shown in pink. A red arrow points to Met1 in *Tetrahymena* p50, which corresponds to Met87 of the human TPP1 protein. The TEN domain of TERT is shown in dark blue and the IFD is shown in cyan. The catalytic core of TERT is shown in dark gray while the telomerase RNA is represented in black. The TEB proteins are depicted in light grey.

Appendix A

A Lentivirus-Free Inducible CRISPR-Cas9 System for Efficient Targeting of Human Genes[3]

A.1 Abstract

CRISPR-Cas9 is a cutting-edge tool for modifying genomes. The efficacy with which Cas9 recognizes its target has revolutionized the engineering of knockouts. However this efficacy complicates the knocking out of important genes in cultured cells. Unedited cells holding a survival advantage within an edited population can confound the knockout phenotype. Here we develop a HeLa-based system that overcomes this limitation, incorporating several attractive features. First, we use Flp-recombinase to generate clones stably integrated for Cas9 and guide RNAs, eliminating the possibility

[3] A modified version of this appendix is published in Analytical Biochemistry (DOI: 10.1016/j.ab.2017.05.001). I helped create the pG1G2-FLAG-Cas9-F3 vector with Kamlesh Bisht and Jacqueline Graniel. Kamlesh Bisht performed all experiments using the final vector including stable cell line generation, *AGO2* and *DICER* targeting, and surveyor assays.

of unedited cells. Second, Cas9 can be induced uniformly in the clonal cultures using doxycycline to measure the knockout phenotype. Third, two genes can be simultaneously knocked out using this approach. Finally, by not involving lentiviruses, our method is appealing to a broad research audience. Using this methodology, we generated an inducible *AGO2*-knockout cell line showing normal RNA interference in the absence of doxycycline. Upon induction of Cas9, the *AGO2* locus was cleaved, the AGO2 protein was depleted, and RNA interference was compromised. In addition to generating inducible knockouts, our technology can be adapted to improve other applications of Cas9, including transcriptional/epigenetic modulation and visualization of cellular DNA loci.

A.2 Introduction

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) system found in several bacteria and archaea embodies an adaptive immune mechanism that relies on faithful recognition of specific nucleic acid sequences (Barrangou et al., 2007; Hsu et al., 2014; Ishino et al., 1987; Jansen et al., 2002; Kunin et al., 2007; Labrie et al., 2010; Makarova et al., 2011; Mojica et al., 2000; Wright et al., 2016). Detailed investigation of the mechanism by which the CRISPR-Cas system recognizes and degrades foreign DNA sequences has fueled the emergence of a new era of genome editing (Gasiunas et al., 2012; Hale et al., 2008; Hale et al., 2012; Hale et al., 2009; Jinek et al., 2012; van der Oost et al., 2014). Several clades of CRISPR-Cas mediated interference systems exist, yet Cas9 from *Streptococcus pyogenes* (SpCas9) is the most extensively used member of the Cas9 endonuclease family (Anders et al., 2014; Jiang and Marraffini, 2015; Nishimasu et al., 2015;

Nishimasu et al., 2014). Although the natural function of SpCas9 (hereby referred to as Cas9) is to cleave infecting phage DNA, Cas9 and its accessory RNA elements have been engineered to recognize and/or cleave DNA, both *in vitro* (Gasiunas et al., 2012; Jinek et al., 2012) and in several eukaryotic model organisms (Sander and Joung, 2014). Cas9 is an RNA-guided DNA endonuclease that forms a complex with a pair of RNA molecules: a guide or CRISPR RNA (crRNA) and an accessory trans-activating CRISPR RNA (tracrRNA). This ribonucleoprotein complex binds to the genomic target via Watson-Crick base pairing through information provided by the guide RNA, and Cas9 cleaves the double stranded DNA target. For genome-editing applications, the crRNA and the tracrRNA can be fused to create a single RNA molecule termed as sgRNA (Jinek et al., 2012). The superior ability of Cas9 to recognize its DNA target, in even the most complex of eukaryotic genomes, is what qualifies Cas9 as a powerful tool for editing genomes. In addition to representing an excellent advancement in biotechnology, CRISPR-Cas9 also holds immense promise for the cure of genetically defined diseases that remain intransigent to other forms of therapy (Cox et al., 2015; Suzuki et al., 2016).

Although the potential of CRISPR-Cas9 in genome editing is clear, there are several other important applications of this technology. Many such applications utilize a version of Cas9 that is catalytically dead (dCas9), but fully capable of binding the DNA target in an sgRNA-dependent manner (Qi et al., 2013). For example, dCas9 can be directed to promoters for regulating gene expression (Gilbert et al., 2013; Konermann et al., 2013; Maeder et al., 2013; Mali et al., 2013; Perez-Pinera et al., 2013). dCas9 can also be tethered to chromatin-modifying enzymes to affect site-specific epigenetic

changes (Seth and Harish, 2016). Furthermore, dCas9 fused to fluorophores such as GFP may be used to directly visualize specific DNA sequences in living (and fixed) cells (Chen et al., 2013); a technique which previously required cell fixation-based fluorescence *in situ* hybridization (FISH) approaches.

Establishment of gene knockouts depends on error-prone non-homologous end joining (NHEJ) of double stranded DNA breaks created by Cas9. Given that the major determinant of knockout efficiency is activity of Cas9, it is not surprising that this enzyme has been successful in knocking out genes in various biological contexts (Cong et al., 2013). Although the ability of CRISPR-Cas9 technology to efficiently knockout genes is an extremely attractive characteristic, it also poses disadvantages compared to existing RNA-knockdown technologies. For example, attempts to knock out an essential gene in cultured human cells will result in selection for cells/clones that are unedited (or edited but still preserve gene function). Therefore, the measured cellular phenotype/s of the surviving cells will not be representative of a true gene knockout. Inducing Cas9 uniformly in all cultured cells can circumvent this problem and allow for the detection of an immediate knockout phenotype. Indeed, there are methods to induce Cas9 both in mouse tissues (Dow et al., 2015) and human induced pluripotent (iPS) cells (Gonzalez et al., 2014). However these methods require injection/transfection/transduction of guide RNAs into the cell population, leading to the caveat mentioned above. Specifically, in the case of Cas9 targeting essential genes, cells that do not receive guide RNAs and thus remain unedited have a potential growth advantage over edited cells. Finally, most of the published methods involve lentiviral-based approaches for delivering Cas9 and/or the guide RNAs (Cao et al., 2016; Shalem

et al., 2014). Although advantageous in several respects, lentiviral approaches require stricter biosafety considerations that pose an additional obstacle for laboratories that are not equipped/approved to conduct such experiments.

Given that the HeLa cell line remains the most popular experimental tool for studying human gene function in cell culture (Landry et al., 2013), we set out to develop a robust methodology to apply the CRISPR-Cas9 system in this cell line. We developed a system with the ability to: (i) generate clones integrated stably with single copies of *Cas9* and guide RNA genes to eliminate the possibility of unedited cells; (ii) prevent Cas9 induction and allow for the propagation/storage of clones until the time of the knockout experiment; (iii) knockout multiple genes simultaneously; (iv) provide a simple, transient transfection-based, lentivirus-free protocol; and (v) provide an economical method for gene disruption obviating the need for repeated use of siRNA/guide RNA or transfection reagents. By using a combination of a HeLa-based clonal cell line [HeLa EM2-11ht (Weidenfeld et al., 2009)] and a new vector that we describe in this study (pG1G2-FLAG-Cas9-F3), we have developed a method that successfully fulfills all these criteria. As proof-of-principle, we generated an inducible knockout of the *AGO2* gene in HeLa cells, and successfully shutdown RNA interference as a function of Cas9 induction. Our choice of *AGO2* knockout was driven by our inability to generate clones of this knockout using standard CRISPR-Cas9 protocols, possibly because of the importance of RNAi for cell growth and function. We believe that our newly developed system holds promise not only for the gene-editing functions of Cas9, but also for the numerous other applications of Cas9 that have emerged.

A.3 Results

A.3.1 Generation of a Flp recombinase-targetable vector for stable expression of two guide RNAs and a dox-inducible FLAG-Cas9 protein in HeLa cells

To build a system that allowed for virus-free, stably inducible, genome editing in HeLa cells, we first constructed a vector that encodes for the components that are key to attaining these characteristics (see Materials and Methods for detailed information). The pBi-F3 vector contains a bi-directional dox/tet-inducible promoter (Ptet) as well as a wild-type/mutated F/F3 pair of Flp recombinase targeting sites flanking the expression cassette (Weidenfeld et al., 2009). We and others have previously used this bidirectional cassette to express a protein-coding gene and a non-coding RNA (e.g., shRNA/miRNA) gene simultaneously from CMV promoters in a dox-dependent manner (Nandakumar et al., 2012; Weidenfeld et al., 2009). We set forth to construct a plasmid that encoded both Cas9 and a cloning site for guide RNAs in the pBi-F3 backbone. However, Cas9 guide RNAs have been expressed in human cells driven by a U6 promoter. The U6 promoter allows for transcription by PolIII, which unlike PolII (which recognizes CMV promoters), does not append a polyA tail at the 3' end of the guide RNA/spacer RNA sequence. To adapt the pBi-F3 vector to co-express two Cas9 guide RNAs, we first amplified a U6-guide cloning RNA cassette using the pX333 plasmid as a template (Maddalo et al., 2014) (Figure A.1). Ligation of the dual guide RNA cloning module with the pBi-F3 backbone resulted in an intermediate vector termed here as pG1G2-F3 (Figure A.1). We note that the selection of cloning sites to introduce the

guide RNA cassette into the pBi-F3 vector resulted in the removal of one Ptet element. Thus, the guide RNAs cloned into pG1G2-F3 will be expressed constitutively in cultured human cells. This did not detract from the overarching goal of obtaining an inducible Cas9 vector system.

The next step in generating our proposed vector was to insert the FLAG-Cas9 gene into the pG1G2-F3 vector downstream of the remaining Ptet element. We first proceeded to amplify the FLAG-Cas9 gene from the pX330 vector (Cong et al., 2013) and inserted it into the pBi4 backbone using conventional restriction site-based cloning to yield the pFLAG-Cas9-Bi4 plasmid backbone (Figure A.1). Next the FLAG-Cas9 fragment was sub-cloned from the pBi4 backbone into the pG1G2-F3 vector to yield the pG1G2-FLAGCas9-F3* plasmid (Figure A.1). The resulting plasmid encodes most of the desired characteristics of the final vector, but was not compatible with cloning of the second guide RNA (g2) using the *BbsI* restriction enzyme. This is because pG1G2-FLAG-Cas9-F3* contained three additional *BbsI* sites elsewhere on the vector backbone. The unwanted *BbsI* sites were mutated sequentially using mutagenesis strategies (detailed in Materials and Methods) to yield the final pG1G2-FLAG-Cas9-F3 vector (Figure A.1). The final vector was sequenced across the guide RNA cassette and the Cas9 gene to rule out any unwanted mutations that may have been introduced during the process of cloning.

A.3.2 Cas9 in pG1G2-FLAG-Cas9-F3 can be induced to efficiently edit the HeLa genome

To test the functionality of pG1G2-FLAG-Cas9-F3 we constructed three derivatives of the vector, one encoding a guide RNA against human *AGO2*, another

encoding a guide RNA against human *DICER*, and a third vector encoding one guide RNA for *AGO2* and one guide RNA for *DICER* (Figure A.2A). To test the dox-dependent induction of FLAG-Cas9, we transiently transfected HeLa EM2-11ht cells. We selected these cells for two reasons. First, this clonal HeLa cell line constitutively expresses the reverse tetracycline controlled transactivator gene coding for rtTA2S-M2 (Weidenfeld et al., 2009). This allows us to switch on genes driven by the Ptet promoter in the presence of dox (50-200 ng/ml). Second, this cell line contains a single, genomically annotated, F/F3 FRT locus for integration of the guide RNA/Cas9 cassette from pG1G2-FLAG-Cas9-F3 to facilitate stable cell line generation (Weidenfeld et al., 2009) (described later). Immunoblotting analysis revealed that transient transfection of pAGO-G1-FLAG-Cas9-F3 and pDICER-G1-FLAG-Cas9-F3 resulted in robust expression of FLAG-Cas9 in HeLa EM2-11ht cells, but only in the presence of dox (Figure A.2B). Thus, as previously reported (Weidenfeld et al., 2009), the Ptet promoter is induced strongly (several-fold above background levels) in the presence of dox in cells also expressing rtTA2S-M2.

Next, we asked if the expressed Cas9 protein in combination with the guide RNA was able to cleave the human genome in a sequence-specific manner. To measure genome editing via CRISPR-Cas9 action, we performed the Surveyor nuclease assay that we and others have reported previously (Bisht et al., 2016; Cong et al., 2013). This assay depends on Surveyor nuclease-catalyzed cleavage of mismatched regions in an otherwise perfect DNA duplex. In human cells, efficient cleavage by Cas9 is generally followed by rejoining of broken ends through NHEJ. This error-prone DNA ligation pathway frequently introduces variable insertions/deletions (indels) at the site of Cas9

cleavage. With such a differentially edited cell population, PCR amplification of the DNA spanning the Cas9 recognition sequence followed by denaturation/renaturation will result in DNA duplexes containing mismatches around the Cas9 cut site. These mismatches, which serve as a scar of Cas9 action, become a substrate for the Surveyor nuclease. Indeed, induction of Cas9 for 72 h post-transfection resulted in robust cleavage of the human *AGO2* and *DICER* loci with guide RNAs targeting these loci (Figure A.2C). No cleavage was detected in the absence of dox, further verifying the inducible nature of the CRISPR-Cas9 knockout using this system.

Next we compared the genome editing capacity of the pG1G2-FLAG-Cas9-F3 vector to that of a well-established Cas9 vector, pX333. For this, we cloned the *AGO2* guide RNA into the pX333 vector either individually or in combination with the *DICER* guide RNA. Cleavage of *AGO2* and *DICER* in the cells expressing the appropriate guide RNAs encoded by the pG1G2-FLAG-Cas9-F3 vector backbone was comparable to that of transfections of pX333-derived plasmids encoding the same guide RNAs. These data suggest that neither our cloning strategy, the nature of the Ptet promoter, nor the characteristics of the HeLa EM2-11ht cell line compromised the efficacy of CRISPR-Cas9 in genome editing to any measurable extent (Figure A.2D). Additionally, the ability of the pAGO-DCR-FLAG-Cas9-F3 plasmid to simultaneously orchestrate *AGO2* and *DICER* cleavage validates the use of this vector for efficiently co-expressing two separate guide RNAs.

A.3.3 Virus-free generation of a cell line stably encoding a guide RNA targeting *AGO2* and an inducible Cas9 gene

The final step towards developing a lentivirus/retrovirus-free system to stably integrate guide RNAs and a dox-inducible *Cas9* gene involves integrating the guide RNA-Cas9 module into the unique F/F3 FRT locus of the HeLa EM2-11ht cell line. The unique integration site ensures uniform expression of inserted genes in a clonal population, compared to lentiviral methods which integrate inserts randomly across the host genome. As proof-of-principle, we decided to invoke two pG1G2-FLAG-Cas9-F3 constructs (separately), one that encodes a guide RNA targeting *AGO2* (to yield AGO2-g1-Cas9-F3 HeLa), and another that is devoid of any guide RNA sequence (to yield Empty-Cas9-F3 HeLa). HeLa EM2-11ht cells were co-transfected with the appropriate FLAG-Cas9-F3-guide RNA construct, and a second plasmid encoding Flp recombinase and a puromycin resistance marker (Figure A.3A). Transfected cells were selected for by including puromycin in the growth medium for ~24 h. Flp recombination-driven integration of the guide RNA-Cas9 locus between the F/F3 sites in the HeLa EM2-11ht genome will result in loss of the thymidine kinase gene that normally separates the F and F3 sites in the parental genome (Figure A.3A). Thymidine kinase expression in the parental cell line renders it sensitive to ganciclovir (gan), allowing for selection of successful recombinant clones by including gan in the growth medium. After ~10 days of gan selection, clones were readily visible under a light microscope. 24 such clones were initially passaged into a 96 well plate, from which duplicate cultures were established. Cells from one of the duplicate cultures (still in 96-well format) were induced briefly (~24 h) with dox and harvested for screening (Figure A.3A). We

performed Western spot-blotting for the FLAG epitope (present on the FLAG-Cas9 construct) to screen for successfully recombined clones (Figure A.3A, B). As shown in Figure A.3B, positive clones were easily recognizable due to the substantially higher than background anti-FLAG signal. Only the un-induced replicate cultures were utilized for downstream analysis. We proceeded to characterize one clone each of AGO2-g1-Cas9-F3 HeLa and Empty-Cas9-F3 HeLa.

A.3.4 Induction of Cas9 results in cleavage of the genomic target, reduction of target protein levels, and loss of target protein function

AGO2 is a central component of the RNA interference machinery in human cells (Hutvagner and Simard, 2008; Liu et al., 2004c; Meister et al., 2004; Rand et al., 2004). Deletion of the *AGO2* gene is therefore expected to diminish siRNA/shRNA-mediated RNA silencing in cultured human cells (Figure A.4D). We proceeded to test this hypothesis using our newly established AGO2-g1 Cas9-F3 HeLa cell line. First, we investigated the ability of dox to switch on FLAG-Cas9 expression in the experimental cells as well as in the "Empty Cas9-F3 HeLa" cells. Indeed, robust Cas9 expression was established within 1 day of induction with dox in both the control and the experimental cell lines (Figure A.4A). Although we did notice a faint Cas9 signal (which was more prominent in the control cell line) at day 0, we believe this reflects the high sensitivity of the FLAG-antibody rather than the Cas9 protein levels necessary for detectible biological function (see below). Given that maximal/stable Cas9 levels are obtained within the first tested time point, our method provides for rapid Cas9 induction.

Next, we performed the Surveyor nuclease assay to investigate the ability of the genomically encoded guide RNA and Cas9 protein to edit the *AGO2* locus. Consistent

with the kinetics of Cas9 induction, maximal *AGO2* cleavage was observed at the earliest time-point after induction (Figure A.4B). Although it seemed as though only a small fraction of the amplicon was cleaved by the Surveyor nuclease, we envision the genome editing to be almost (or fully) complete in the cultured population for the following reasons. First, we believe that it is not possible to observe 100% cleavage in the Surveyor nuclease assay (even with the manufacturer's positive control; data not shown), because the Surveyor nuclease is limiting under these conditions. Second and more importantly, the highest running band in the AGO2-g1-Cas9 HeLa cells runs slightly, but consistently, lower than the corresponding band (representative of Cas9 un-cleaved DNA) in the empty vector control (Figure A.4B). This is consistent with the observation of a greater accumulation of deletions (versus insertions) as a result of NHEJ-based repair (Zhuang et al., 2009). The third and most conclusive evidence for efficient knockout of *AGO2* in the CRISPR-edited cells comes from immunoblot analysis of endogenous human AGO2 protein. Whereas AGO2 protein levels are unaffected in control cells or in AGO2-g1-Cas9 HeLa cells before dox treatment, AGO2 protein levels rapidly decline upon Cas9 induction (Figure A.4C).

In our previous studies on the telomere protein TPP1, we have described an shRNA construct (sh^{TPP1}) that efficiently knocks down recombinant FLAG-TPP1 (as well as endogenous TPP1) in HeLa EM2-11ht cells (Nandakumar et al., 2012). We hypothesized that inducing *AGO2* cleavage will diminish RNA interference (RNAi) in these cells, leading to a restoration of TPP1 protein levels despite the presence of its targeting shRNA (Figure A.4D). Fully consistent with this idea, sh^{TPP1} efficiently knocked down transiently expressed FLAG-TPP1 in Empty-Cas9-F3 HeLa cells, but

was not effective in reducing TPP1 protein levels in AGO2-g1-Cas9-F3 HeLa cells (Figure A.4E). In summary, we have developed a plasmid-based, virus-free system for conducting genome editing in an inducible manner uniformly across cells in a HeLa culture. Using this system, we efficiently induced a human *AGO2* knockout. As a result, we were able to assess the immediate biological consequences of *AGO2* deletion, namely, the disruption of RNAi.

A.4 Discussion

CRISPR-Cas9 technology has revolutionized the field of genome editing. Recent and rapid developments have allowed for significant improvement in precision, efficiency, as well as versatility of this technology. Our study provides a significant improvement for the application of CRISPR-Cas9 technology in cultured human cells. Although technologies exist for inducing Cas9 expression, engineering cell lines to stably integrate Cas9, and utilizing recombinogenic methods to insert *Cas9* into the genome, our study is the first (to our knowledge) to integrate all these features in one system for use in cultured human cells. Specifically, we developed a Flp-recombinase targetable vector that contains a dox-inducible *Cas9* gene as well as the capacity to express two guide RNAs. Transient transfection of this vector results in efficient Cas9 expression and genome editing, but in a strictly dox-dependent manner. More importantly, integration of this module into the genome of HeLa cells provides the ability to edit genomes in an inducible manner. As proof-of-principle, we knocked out *AGO2* in a clonal culture stably co-expressing Cas9 protein and guide RNAs against *AGO2* to attenuate RNAi. The lack of lentiviral intervention in our methodology makes it attractive to a wider spectrum of human cell biologists. Our method requires ~one

month from start to completion, providing a swift avenue for generating inducible knockout cell lines of potentially any gene in the human genome (Figure A.5).

AGO2 is a gene critical for mammalian development and growth (Liu et al., 2004c; Morita et al., 2007). Our ability to knock this gene out in cultured human cells highlights an important application of our methodology. Our method involves co-expression of the relevant guide RNAs and Cas9 uniformly in clonal culture, severely reducing the possibility of unedited cells in the population. It can be argued that although our method will uniformly edit all cells, cells where editing did not disrupt gene function substantially (e.g., short in-frame mutations) would hold a survival advantage in culture. However, this concern can be overcome easily using two guide RNAs to excise a large region of the gene-of-interest. Hence, we believe that our method holds the potential to reveal the immediate consequences of knockout of any essential human gene in cell culture.

We envision several other advantages as well as potential applications of our newly developed technology. First, the use of stably inducible cell lines in our methodology provides an obvious economical advantage. Once established for any particular gene-of-interest, it eliminates the need for repeated transfections. Second, the number of guide RNAs that can be expressed using our method is not limited to two. The pG1G2-FLAG-Cas9-F3 can be easily adapted to insert multiple copies of the dual-guide RNA cassette to facilitate multiple gene knockouts (e.g., of all/several redundant genes in a pathway) using a single vector. Third, our strategy is not only restricted to generation of gene knockouts, but can also be very useful in generating gene knock-ins. Knock-in experiments involve co-transfection of the guide RNA/Cas9 vector and the

DNA repair template, which is often a single stranded oligodeoxynucleotide (ssODN) containing the intended mutation. Our method eliminates the necessity of transfecting multiple DNA/RNA constructs, and avoids the complications that accompany the co-transfection of nucleic acids of different types (single stranded vs. double stranded). Using our method, only the repair ssODN needs to be introduced into the Cas9-guide RNA encoded cells prior to induction of Cas9 expression.

Fourth, our method will allow for a vast improvement in the exploitation of dCas9-GFP as a replacement over conventional fluorescence *in situ* hybridization techniques. For example, dCas9-GFP has been used to visualize telomeres and other genomic loci in human cells (Chen et al., 2013). Although these experiments serve as proof-of-principle for application of CRISPR-Cas9 technology, they suffer from potential drawbacks. For example, constitutive localization of dCas9-GFP and guide RNAs at the genomic target might have unanticipated physiological consequences in the cell. At telomeres, such binding could result in unwanted chromosome end-deprotection and/or telomerase misregulation (Palm and de Lange, 2008). Our method could be easily adapted to attenuate such side effects. A pBi-F3-based dCas9 vector can be stably integrated into the HeLa genome without risk of constitutive expression of the dCas9-GFP protein. A brief pulse of dCas9-GFP expression induced with dox can be used immediately prior to imaging of the telomeres (or other DNA loci), minimizing the risk of unwanted affects from dCas9-GFP association with its genomic target. Similarly, our methodology could improve the application of dCas9 as a modulator of transcription and epigenetic modulation. The inducible (and stable) nature of our Cas9 expression platform will allow for the generation of cell lines where a particular endogenous gene-

of-interest may be activated/repressed or epigenetically modified with great temporal control. In summary, our newly developed method provides several improvements over existing CRISPR-Cas9 technology, and holds great promise for the study of cellular function in the context of cultured human cells.

A.5 Materials and Methods

A.5.1 Reagents and kits for molecular biology

Oligonucleotides for PCR priming, Cas9 guide RNA cloning, and Sanger sequencing were purchased from Integrated DNA Technologies. All restriction enzymes were purchased from New England Biolabs (NEB). Purification of plasmid DNA and other cloning intermediates was performed using DNA purification kits from Qiagen. Genomic DNA from cultured human cells was isolated using the GenElute Mammalian Genomic DNA Miniprep kit from Sigma. PCR reactions for cloning purposes were performed with either Pfu Turbo DNA polymerase (Agilent) or Phusion High-Fidelity DNA Polymerase (NEB) using the manufacturers' protocols. Ligations of DNA vectors with inserts were performed using the Quick Ligation Kit (NEB). Calf intestinal alkaline phosphatase (CIP) for removing the 5'-phosphate of vectors prior to ligation was purchased either from Promega or from NEB. Site-directed mutagenesis was performed using QuikChange II (Agilent). Reagents for CRISPR-Cas9 experiments are described separately below.

A.5.2 Sanger sequencing

The DNA sequences of all inserts as well as plasmid regions involved in site-directed mutagenesis were verified using Sanger sequencing conducted at the University of Michigan DNA Sequencing Core.

A.5.3 Parental plasmids

The bicistronic vector pX330-U6-Chimeric_BB-CBhhSpCas9 for site-specific genome editing in cultured human cells was a kind gift from Dr. Feng Zhang, Broad Institute of MIT and Harvard, McGovern Institute for Brain Research, and Departments of Brain and Cognitive Sciences and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA (Addgene plasmid # 42230) (Cong et al., 2013), and was obtained upon signing a material transfer agreement (MTA). The pX333 vector that allows for cloning of two tandem U6 promoter-driven guide RNAs was a kind gift from Dr. Andrea Ventura, Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY (Addgene plasmid # 64073) (Maddalo et al., 2014), and was obtained upon signing an MTA. The pBI-F3-miRNA-d1GFP-loxP-F (abbreviated here as pBi-F3) and pBI4-miRNA-d1GFP (abbreviated here as pBi4) plasmids were obtained from TET Systems GmbH & Co. KG, Heidelberg, Germany, upon signing an MTA, and have been described previously (Bisht et al., 2016; Nandakumar et al., 2012; Weidenfeld et al., 2009). The FLAG-TPP1 plasmid and the plasmid encoding an shRNA that targets the gene coding for human TPP1 have been described previously (Nakashima et al., 2013; Nandakumar et al., 2012).

A.5.4 Cloning of the g1/g2-inducible Cas9 F3 vector

To obtain a FLAG-Cas9 cassette driven by a doxycycline (dox)-inducible promoter, the pBI4-miRNA-d1GFP vector (6.4 kb) was first digested with *NotI* and *XhoI*. The fragment containing the dox-inducible promoter was treated with CIP and ligated to the FLAG-Cas9 insert obtained by PCR amplification of the FLAG-Cas9 fragment in the pX330 vector (containing codon-optimized *Streptococcus pyogenes Cas9* gene, a FLAG tag, and two nuclear localization sequences) to yield the pFLAG-Cas9-Bi4 plasmid.

To insert the tandem guide RNA expression cassette into the pBi-F3 backbone, the pBI-F3-miRNA-d1GFP-loxP-F vector was digested with *BglII* and *StuI*. The ~5.4 kb fragment obtained after gel-purification of the digestion reaction was treated with CIP and ligated to the tandem guide RNA expression cassette insert amplified using PCR of the pX333 vector template with to yield the pG1G2-F3 plasmid.

To sub-clone the FLAG-Cas9 fragment in the pG1G2-F3 backbone, the pG1G2-F3 and pFLAG-Cas9-Bi4 vectors were separately digested with *StuI* and *HpaI*. The restriction digestion of pG1G2-F3 yielded a ~1.4 kb fragment and a ~4.9 kb fragment. The ~4.9 kb fragment isolated after gel purification was treated with CIP prior to DNA ligation. The restriction digestion of pFLAG-Cas9-Bi4 yielded a ~4.8 kb fragment and a ~3.8 kb fragment. The ~4.8 kb fragment was gel-purified and ligated to the purified vector fragment to yield the pG1G2-FLAG-Cas9-F3* vector. This vector contains a dox-inducible FLAG-Cas9 gene and the sites for cloning in two guide RNA genes. However, the *BbsI* recognition site, which is the cloning site for one of the two possible guide RNAs, is repeated at three additional sites in the vector. While one extraneous *BbsI*

site (most proximal to the guide RNA cassette) was removed using site-directed mutagenesis (Agilent), the other unwanted *BbsI* sites were subsequently and sequentially mutated using two-step overlap extension PCR methodology (Higuchi et al., 1988). This yielded the final pG1G2-FLAG-Cas9-F3 vector that contains (i) a dox-inducible *FLAG-Cas9* gene; (ii) two tandem U6 promoter-driven cloning sites for guide RNA expression; and (iii) the F/F3 FRT sites that allow for integration into a homologous site in the HeLa EM2-11ht genome.

A.5.5 Cell Culture

All experiments in this study involving cultured human cells were performed with HeLa EM2-11ht cells or its clonal derivatives developed for this study. The HeLa EM2-11ht cell line was obtained from TET Systems GmbH & Co. KG, Heidelberg, Germany, upon signing a MTA. Cells were cultured in an incubator maintained at 37 °C and 5% CO_2 in growth medium containing DMEM, 10% FCS, 2 mM L-glutamine, 1 mM sodium pyruvate, 100 units/ml of penicillin, and 100 mg/ml of streptomycin. Induction of dox-inducible genes in this cell line was performed using doxycycline (Sigma) added to the growth media at a final concentration of 200 ng/ml.

A.5.6 Immunoblotting

Standard immunoblot protocols were used with the following primary antibodies at specified dilutions: mouse monoclonal anti-Flag M2-HRP conjugate (Sigma; A8592; 1:20,000), mouse monoclonal anti-β-actin antibody (Sigma; A5441; 1:10,000), and mouse monoclonal anti-AGO2 antibody (11A9) (Rudel et al., 2008), a kind gift from Dr. Gunter Meister, Center for Integrated Protein Science Munich (CIPSM), Germany. Secondary horseradish peroxidase-conjugated goat antibodies against rabbit IgG

(Santa Cruz Biotechnology; 1:10,000) and mouse IgG (Santa Cruz Biotechnology; 1:10,000) were used to recognize the primary antibodies and allow for detection of the protein-of-interest using chemiluminescence detection technology (Pierce ECL Western Blotting Substrate; Thermo Scientific). The data were visualized using a gel-documentation system (ChemiDoc MP System; BioRad).

A.5.7 Transient transfections

All transient transfection experiments were performed using Lipofectamine 2000 (Life Technologies) following the manufacturer's recommendations. For transfections, cells were seeded ~24 hours prior to transfection at 40-50% confluency. Transfections were performed in growth medium devoid of antibiotics and serum. Medium complete with antibiotics and serum was used to replace the transfection medium 5 h post-transfection.

A.5.8 Guide RNA design

The Zhang Lab CRISPR design algorithm was used to design guide RNAs against *AGO2* and *DICER* (http://crispr.mit.edu/). Based on a high CRISPR design tool score and a low probability of off target cleavage events, one or two of the highest scoring guide RNAs were used in the experiments. The selected guide sequence oligonucleotides were annealed to each other, phosphorylated with T4 polynucleotide kinase enzyme (NEB), and cloned into the *BbsI* and/or *BsaI* site/s of the pX333 vector or the pG1G2-FLAG-Cas9-F3 vector for guide RNA expression.

A.5.9 Surveyor nuclease assay

Surveyor nuclease experiments to detect genomic editing by Cas9 were performed by adapting protocols published previously (Bisht et al., 2016). For

experiments involving transient transfection of Cas9/guide RNA-containing plasmids, ~0.3 x 10^6 HeLa EM2-11ht cells were transfected in a 12-well growth format with 1 µg of Cas9/guide RNA-containing plasmids using 2.5 µl Lipofectamine 2000 (Life Technologies). After 72 h of transfection, cells were trypsinized and harvested by centrifugation. 150-200 ng of genomic DNA isolated from the harvested cells was used for PCR amplification with Surveyor primers in a 50 µl reaction volume. Upon completion of the PCR reaction, 4 µl of each reaction was loaded on a 1-2% agarose-TAE gel to check the uniformity and efficiency of amplification among different samples. The amplification products were purified using a PCR purification kit and subjected to the Surveyor nuclease assay according to the manufacturer's protocol (Surveyor Mutation Detection Kit; catalog #706025; Transgenomic). Upon completion of the Surveyor nuclease cleavage reaction, the products were visualized using ethidium bromide-stained, 2% GTG agarose-TAE gels.

A.5.10 Lentivirus-free production of clonal stable cell lines encoding dox-inducible Cas9

HeLa-EM2-11ht cells were transfected in a 6-well format using Lipofectamine 2000 (Life Technologies) with 1 µg of the indicated pG1G2-FLAG-Cas9-F3 and 1 µg of a puromycin resistance-encoding, Flp recombinase-expressing plasmid. ~24 h post-transfection, cells were treated with puromycin (5 µg/ml; Sigma-Aldrich). After ~24 h of selection under puromycin, fresh medium containing ganciclovir (50 µM; Sigma-Aldrich) was added and selection was conducted for ~10 days. 24 individual clones were picked and transferred to individual wells of a 96-well plate. Once confluent the cells in each well were passaged into duplicate 96-well plates. One of the duplicate plates was used

to identify positive clones. Positive clones were identified based on a strong signal in a spot-blot based immunoblotting assay against the FLAG tag (present on the Cas9 construct). For this, clones were grown in a 96-well format in the absence of doxycycline. The plate was trypsinized and split into two new 96-well plates (master plate and replica plate). The replica plate clones were treated for 24 h with doxycycline (200 ng/ml in culture medium) and lysed in 100 µl of 5 M urea-containing buffer. 50 µl of the lysate was blotted on a nitrocellulose membrane using a vacuum manifold-based dot blot apparatus. The membrane was stained with Ponceau-S staining solution (Sigma) and imaged to ensure uniform blotting. The blot was then blocked with StartingBlock (PBS) blocking buffer (Thermo Scientific) and incubated with anti-Flag M2-HRP conjugate (Sigma; A8592; 1:20,000). Positive clones marked on the replica plate were used to expand corresponding clones from the master plate. One positive clone for each pG1G2-FLAG-Cas9-F3 construct was expanded until it grew to confluency in 6 cm dishes. Doxycycline (200 ng/ml) was added to the growth medium to induce genome editing uniformly in cells of the clonal culture.

A.6 Figures and Tables

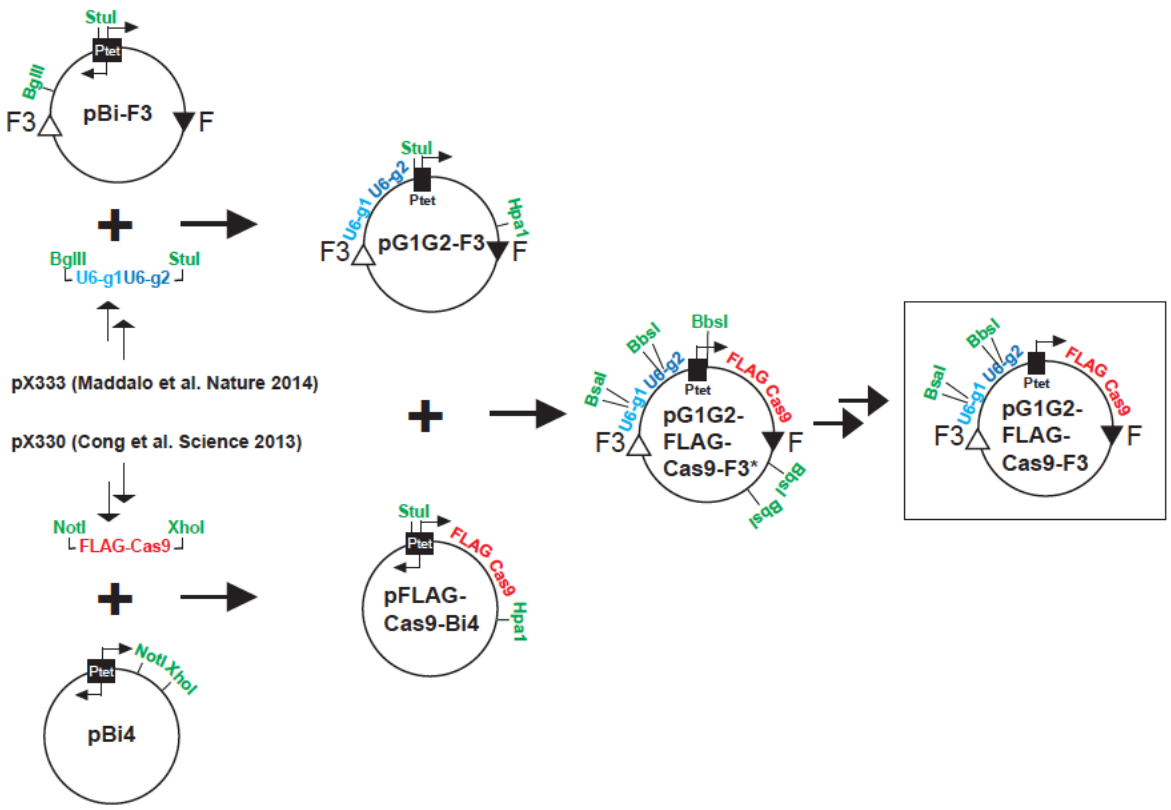


Figure A.1 Cloning strategy for generation of a vector allowing for site-specific virus-free integration into the genome, and stable co-expression of doxycycline-inducible Cas9 and two guide RNAs.
The stepwise cloning protocol starting from FLAG-Cas9 (from pX330) and tandem guide RNA gene inserts (from pX333) leading to the engineering of the final pG1G2-FLAG-Cas9-F3 vector is shown. Restriction sites pertinent to the subsequent cloning step are indicated. F3 indicates a mutated FRT site, while F indicates a wild-type FRT site. The F/F3 combination allows for directional integration into the analogous (and unique) F/F3 site in the HeLa EM2-11ht genome. Ptet indicates a promoter that is induced by addition of dox in cells (e.g., HeLa EM2-11ht cells) also expressing the rtTA2S-M2 protein. The U6 promoter allows for PolIII-driven transcription of guide RNA.

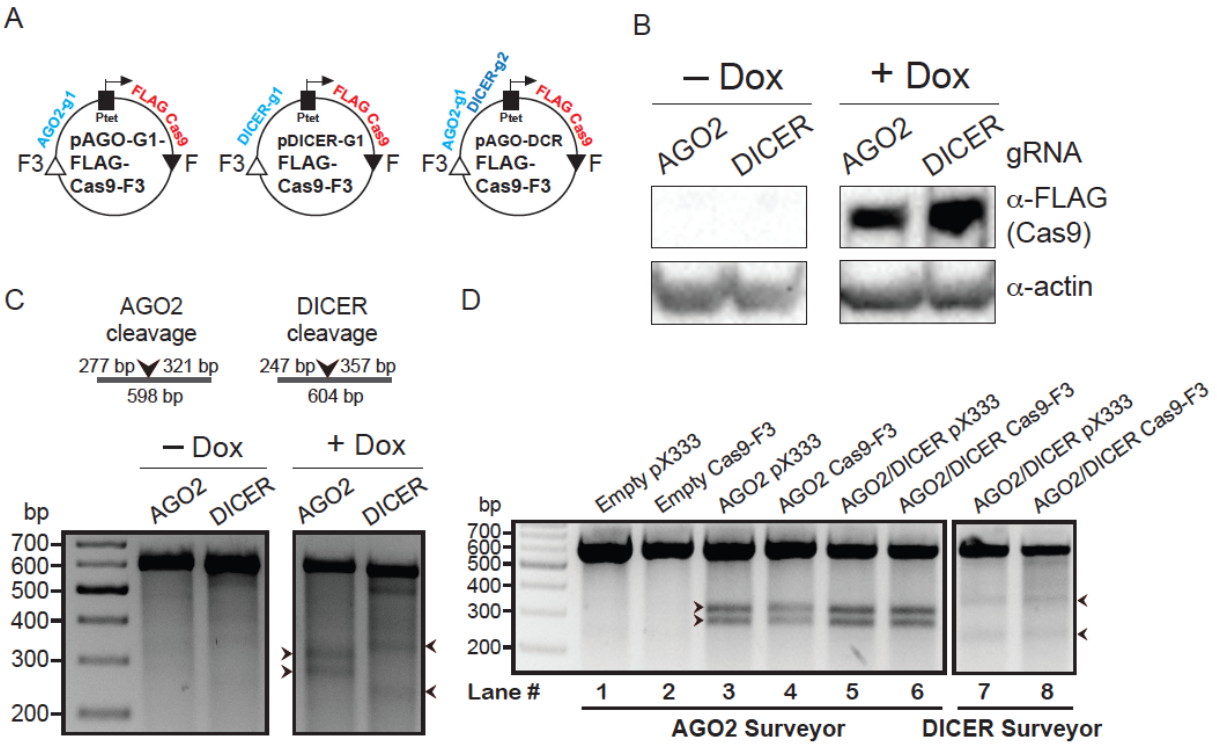


Figure A.2 pG1G2-FLAG-Cas9-F3 can be adapted to efficiently cleave the endogenous human *AGO2* and *DICER* loci.
(A) The vector diagrams of the three pG1G2-FLAG-Cas9-F3 derived plasmids encoding gRNAs against the indicated gene loci are shown. The pAGO-G1-FLAG-Cas9-F3 and pDICER-G1-FLAGCas9-F3 plasmids both contain one empty guide RNA cassette.
(B) HeLa EM2-11ht cells were transfected with plasmids encoding guide RNAs (as shown in panel A) targeting the indicated genes. Immunoblotting was performed on lysates of cells 24 h post-transfection using the indicated antibodies. “+ Dox” indicates transfections including 200 ng/ml of dox in the growth medium, while “- Dox” indicates transfections not supplemented with Dox. (C) The Surveyor nuclease assay was performed to detect cleavage of the indicated genomic targets by guide RNAs and Cas9 derived from the two plasmids shown in panel A. The sizes of the PCR amplicons that were subjected to Surveyor nuclease treatment are shown schematically as horizontal bars at the top. For each experiment, the sizes of the expected Surveyor reaction products, which correspond roughly to the distances (in bp) between the Cas9 cut site (shown as an arrowhead on the schematic) and the ends of the PCR amplicon are also indicated. The digested fragments obtained experimentally from Surveyor nuclease are indicated with arrowheads adjacent to the gel, and match closely the predicted sizes. (D) Surveyor assay results for the indicated guide RNAs (AGO2 or DICER) residing on the indicated plasmid backbones (pX333 or Cas9-F3) is shown. Cleavage efficiency of a given guide RNA was similar between the two tested vector backbones.

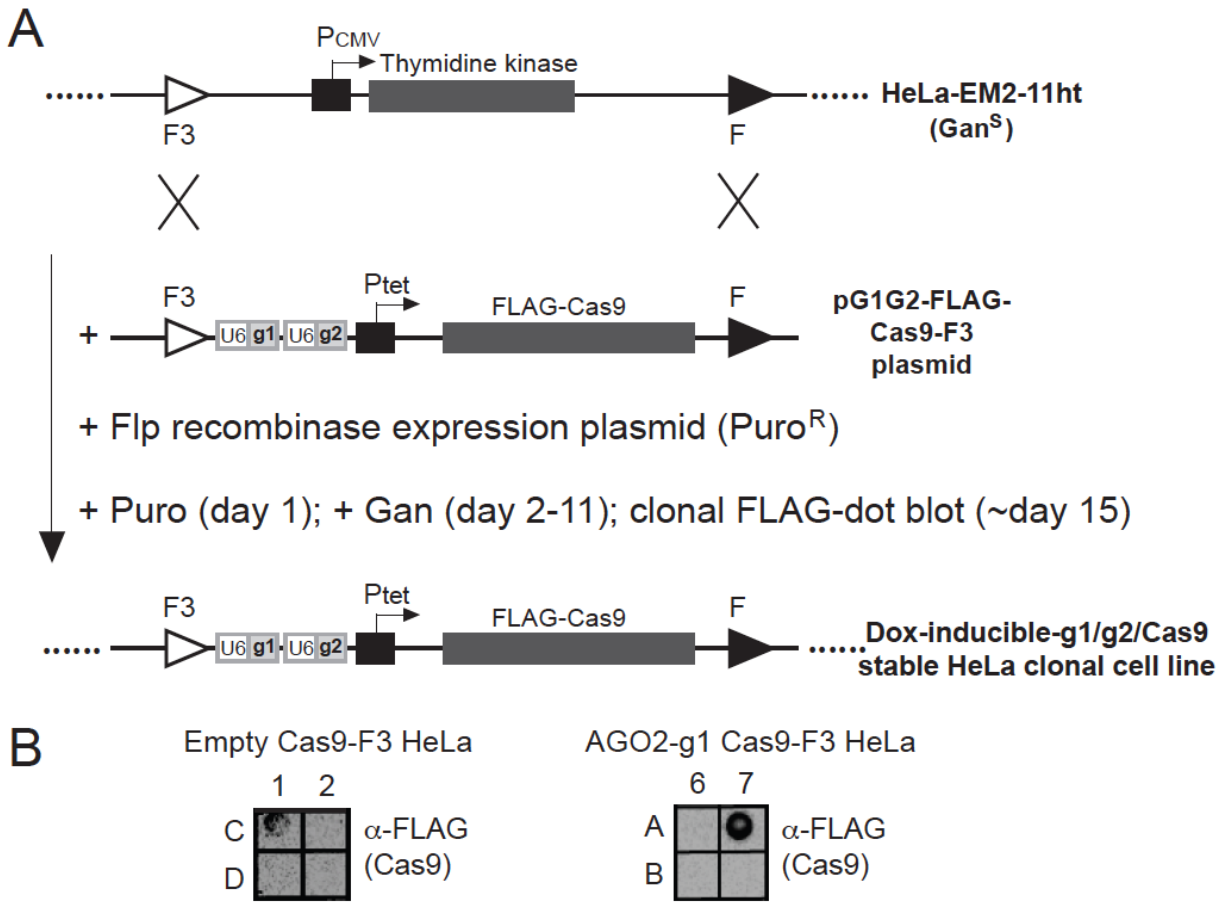


Figure A.3 Virus-free generation of a cell line stably expressing dox-inducible Cas9 and guide RNAs against human *AGO2*.
(A) A schematic for the engineering of HeLa EM2-11ht clonal cell lines stably integrated with the guide RNA/FLAG-Cas9 cassette using transient Flp recombinase expression. Transfected cells are enriched for using selection with puromycin, resistance to which is provided by the Flp-encoding plasmid. Stably integrated cells are enriched for by treatment with ganciclovir, which causes toxicity in parental cells because of continued thymidine kinase expression. Colonies surviving the selection procedure were expanded and analyzed for dox-dependent FLAG-Cas9 expression using a spot-blot based anti-FLAG immunoblot assay. (B) Anti-FLAG spot-immunoblot of positive clones for Empty-Cas9-F3 HeLa (well # C1 in a 96-well plate) and AGO2-g1-Cas9-F3 HeLa (well # A7 in a 96-well plate) are shown. Dox was added to induce Cas9 expression.

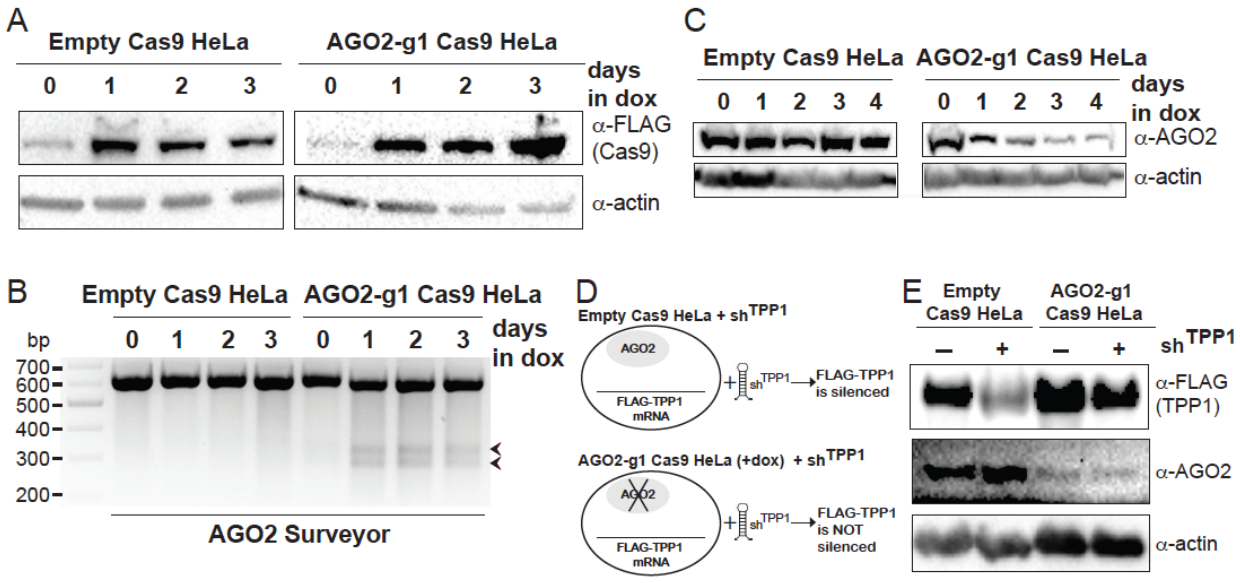


Figure A.4 Induction of Cas9 in the presence of an *AGO2*-targeting guide RNA results in robust cleavage of the human *AGO2* locus, rapid depletion of AGO2 protein, and loss of RNA interference capacity.
(A) Empty-Cas9-F3 HeLa and AGO2-g1-Cas9-F3 HeLa cells were treated with dox (200 ng/ml) and induction of FLAG-Cas9 was monitored as a function of days of induction using anti-FLAG immunoblotting. “0 days in dox” indicates a time-point immediately prior to dox treatment. (B) Surveyor analysis as described in Figure 2C and D was performed with genomic DNA derived from the indicated stable cell lines after indicated durations of dox treatment. *AGO2* cleavage was absent in the Empty vector cell line, while clearly present in AGO2-g1-Cas9-F3 HeLa cells at all time-points measured after dox treatment. Arrowheads indicate the Surveyor cleavage products, which match the predicted sizes shown in the schematic in Figure 2C. (C) Endogenous AGO2 protein levels were detected using immunoblotting; AGO2 levels sharply declined upon dox addition in AGO2-g1-Cas9-F3 HeLa cells. (D) Schematic of the prediction of consequences of *AGO2* knockout on FLAG-TPP1 expression. (E) Results of the sh^{TPP1} knockdown experiment in control and AGO2-g1-Cas9-F3 HeLa cells induced with dox. While the shRNA efficiently knocks down TPP1 in the control cells, it is unable to do so in cells AGO2-deficient AGO2-g1-Cas9-F3 cells.

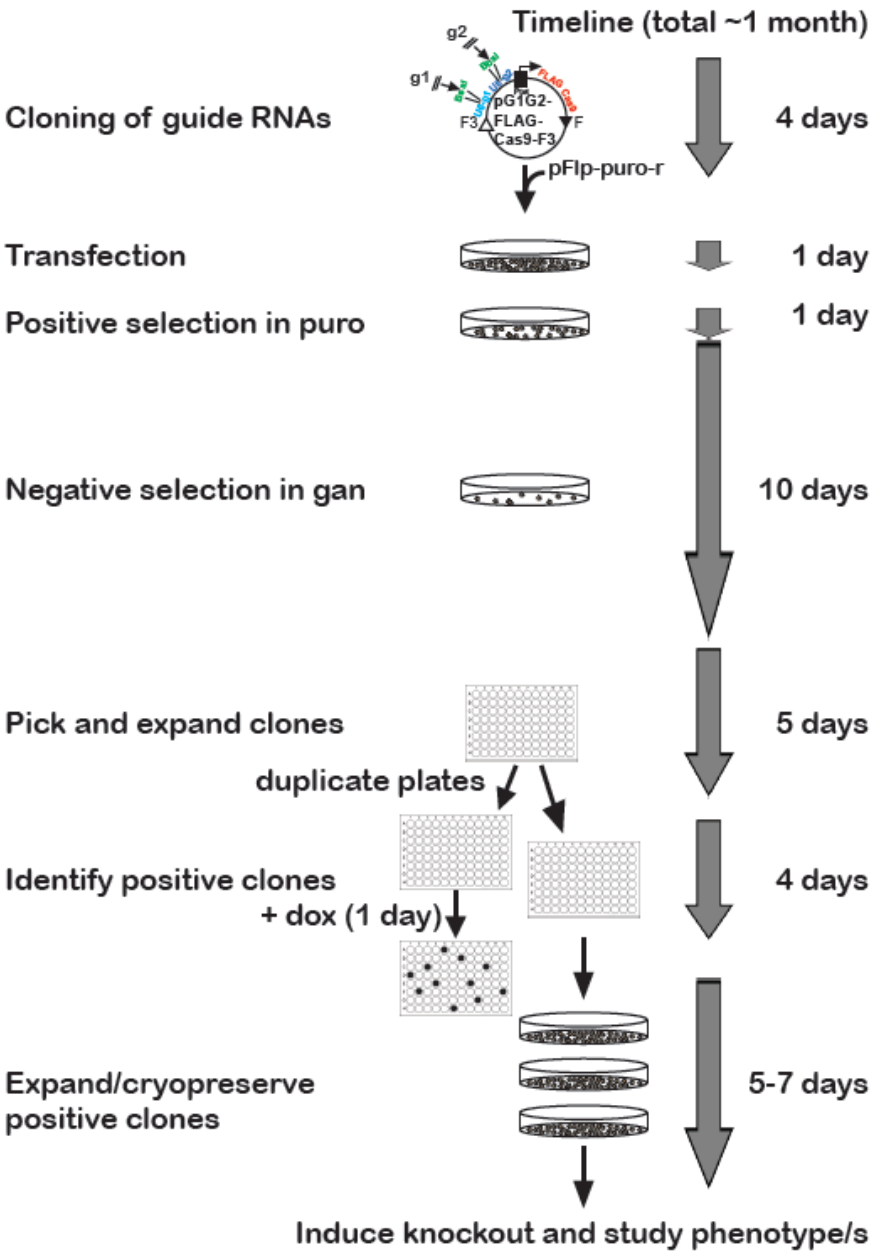


Figure A.5 Timeline for generating stable cell lines expressing dox-inducible Cas9 and guide RNAs.

Appendix B

Novel Interaction Between Proteins Responsible for Chromosome End Protection

B.1 Introduction

The six-protein complex shelterin is responsible for protecting chromosome ends from erroneous DNA damage response and repair machinery. Within shelterin, TRF2 binds double stranded telomeric DNA to prevent an erroneous ATM-mediated DNA damage response at telomeres, while POT1 binds single stranded telomeric DNA, thus preventing an ATR-mediated DNA damage response. TPP1 interacts with POT1 to increase POT1's affinity for single stranded telomeric DNA and recruit telomerase to chromosome ends. TRF1 contributes to the repression of an ATR-mediated DNA damage response and supports telomere replication by recruiting helicases to the double stranded telomeric DNA it binds (Sfeir et al., 2009; Zimmermann et al., 2014). TIN2 bridges the single stranded shelterin proteins with the double stranded shelterin proteins by interacting with TPP1, TRF1, and TRF2, while Rap1 binds TRF2 to help

repress unwanted homology directed repair (Palm and de Lange, 2008; Rai et al., 2016).

While each shelterin protein contributes to a distinct function at chromosome ends, their higher-order interactions are also essential for carrying out their individual roles. TPP1 and POT1 function as an obligate heterodimer, yet TPP1 must also interact with TIN2 to localize to telomeric DNA, making the TPP1-TIN2 interaction essential for POT1-mediated end protection (Frescas and de Lange, 2014a). TIN2 also interacts with both TRF1 and TRF2, thereby bridging POT1 and TPP1 to the double stranded telomeric DNA binding proteins (O'Connor et al., 2006). Loss of TIN2 elicits both an ATR and ATM-mediated DNA damage response at telomeres, suggesting that the TRF2-TIN2 interaction, as well as the TIN2-TPP1 interaction, is essential for chromosome end protection. Finally, loss of TRF1-TIN2 binding results in a significant reduction of POT1, TPP1, and TIN2 at the telomere (Frescas and de Lange, 2014b), underscoring the importance of each individual shelterin protein-protein interaction for telomere integrity.

The bulk of our knowledge regarding protein-protein interactions within shelterin comes from co-immunoprecipitation experiments. While these studies have clearly identified robust binary interactions, they are inept at quantifying the strength of individual interactions or identifying transient/weak interactions without the use of a cross-linker. However a major impediment to biophysically characterizing shelterin interactions is the difficulty in obtaining large quantities of purified shelterin components.

To develop a quantitative understanding of the various protein-protein interactions within shelterin, we performed a flow cytometry protein interaction assay

(FCPIA) that allows for quantitation of direct and competitive protein-protein binding events using small quantities of purified recombinant proteins. Using this methodology, we verify and quantitatively characterize known interactions within shelterin, including the TRF1-TIN2 interaction and the TRF2-Rap1 interaction. To our surprise we also observe a previously unreported interaction between the shelterin proteins TRF1 and TRF2 and demonstrate how addition of either TIN2 or Rap1 inhibits this interaction. However we are unable to identify a specific region in TRF2 that is responsible for this interaction or unequivocally demonstrate this interaction using other conventional approaches, suggesting that more work is required to evaluate the physiological significance of our observation.

B.2 Results

B.2.1 Describing shelterin protein-protein interactions *in vitro*

Using a previously described flow cytometry protein interaction assay (FCPIA) we characterized the TRF2-Rap1 interaction (Blazer et al., 2010; Pendlebury et al., 2017). For this, we first purified full length human Rap1 and TRF2 proteins (Figure B. 1B). Rap1 was biotinylated and immobilized on uniformly sized streptavidin beads while TRF2 was labeled with Alexa Fluor 488. Biotinylated Rap1 on streptavidin beads was then incubated with increasing concentrations of Alexa Fluor labeled TRF2 and the median fluorescence on the beads was measured by flow cytometry (Figure B. 1C). We determined a dissociation constant (K_d) of 0.5 $\pm$ 0.07 nM for the Rap1-TRF2 interaction (Figure B. 1D,G). We then asked how this compared with the TIN2-TRF1 interaction.

We purified human TIN2 and TRF1 proteins from insect cells and used the FCPIA to determine how TRF1 bound TIN2 (Figure B. 1B). Alexa Fluor labeled TRF1 bound biotinylated TIN2 with a dissociation constant of 12.9 $\pm$ 8 nM (Figure B. 1E, G). To further investigate the interaction, we asked how the K_d of the TRF1-TIN2 interaction changed when TRF1 was biotinylated and TIN2 was Alexa Fluor labeled (Figure B. 1F). We found that the K_d for this interaction was 55 $\pm$ 26, slightly higher than the reverse configuration (Figure B. 1G). To further test the specificity of this interaction, we purified a TIN2 truncation protein which stops at amino acid 218 ($TIN2_{218}$) and thus does not contain the TRF1 binding motif. Alexa Fluor labeled $TIN2_{218}$ did not bind biotinylated TRF1, further confirming the specificity of the FCPIA and the TRF1-TIN2 interaction (Figure B. 1F). Together these data establish the FCPIA as powerful tool for characterizing shelterin protein-protein interactions.

B.2.2 TRF1 directly binds shelterin protein TRF2

While optimizing the FCPIA for the TRF1-TIN2 interaction, we used Alexa Fluor labeled TRF2 as a negative control for binding biotinylated TRF1, as an interaction between these two proteins had not been previously detected (Broccoli et al., 1997). While we observed the expected TRF1-TIN2 interaction in this assay, we were surprised to find that Alexa Fluor labeled TRF2 also interacted with biotinylated TRF1 with a K_d of 74 $\pm$ 31 nM (Figure B. 2A,C), as previous studies had failed to see an interaction between these two proteins in co-immunoprecipitation assays (Broccoli et al., 1997). To further test this interaction, we biotinylated TRF2 and Alexa Fluor labeled TRF1 and observed that TRF1 interacted with TRF2 with a K_d of 102 $\pm$ 21 nM (Figure 2B,C). Both TRF1 and TRF2 bind double stranded telomeric DNA, so we asked if this

interaction could be bridged by contaminating DNA (that co-eluted with the protein/s during chromatographic purification). To answer this, we added benzonase, a nuclease which degrades both RNA and DNA, to our Alexa Fluor labeled TRF1 in our flow cytometry assays. Alexa Fluor labeled TRF1 along with benzonase bound biotinylated TRF2 with a K_d of 101 nM, suggesting that the TRF1-TRF2 interaction is not mediated by RNA or DNA (Figure B. 2B,C). In our direct binding FCPIA experiments either TRF1 or TRF2 is labeled using Alexa Fluor 488- maleimide, which reacts with the cysteines in these proteins and thus could affect their binding properties. Therefore we utilized a competition based flow cytometry assay described previously to quantitatively assess TRF1-TRF2 binding in the absence of labeling artifacts (Pendlebury et al., 2017). For this, Alexa Fluor labeled TRF2 was pre-bound to biotinylated TRF1 on streptavidin beads. Unlabeled TRF2 was then titrated in and allowed to compete for TRF1 binding. Median fluorescence on the beads was measured as a function of the titrated unlabeled TRF2 (Figure B. 2D). Using this competition-based flow cytometry assay, we found that unlabeled TRF2 successfully competed for TRF1 binding, further confirming our direct binding FCPIA results (Figure 2E). Together these data suggest a novel direct interaction between the shelterin components TRF1 and TRF2.

B.2.3 Binding of either Rap1 or TIN2 inhibits the TRF1-TRF2 interaction.

TRF1 and TRF2 have previously been suggested to homodimerize, but not heterodimerize, through their TRFH domains (Broccoli et al., 1997). Therefore, we asked what differences may explain both our FCPIA results as well as this previously published data. While our binding studies were conducted with purified protein, previous studies utilized in cell co-immunoprecipitation assays. We hypothesized that

binding of other shelterin components, such as Rap1 and TIN2, which would occur in cells, may affect the interaction between TRF1 and TRF2. To that end, we asked how binding of TIN2 to TRF1 affected the TRF1-TRF2 interaction. We utilized our competition-based flow cytometry assay by using Alexa Fluor labeled TRF2 pre-bound to biotinylated TRF1. As expected, cold TRF2 competed with Alexa Fluor labeled TRF2 for TRF1 binding (Figure B. 3A, blue). Excitingly, cold TIN2 also out competed Alexa Fluor labeled TRF2 for TRF1 binding, suggesting that the TRF1-TIN2 and TRF1-TRF2 interactions are mutually exclusive (Figure B. 3A, red). Notably, $TIN2_{218}$, which does not contain the TRF1 binding motif but does contain the TRF2 binding region, did not compete off TRF2 from TRF1 (Figure B. 3A, green). We asked if Rap1 binding to TRF2 similarly inhibited the TRF1-TRF2 interaction. For this, we used our direct binding FCPIA with biotinylated TRF1 immobilized on streptavidin beads and Alexa Fluor labeled TRF2. As expected, Alexa Fluor labeled TRF2 bound biotinylated TRF1 (Figure B. 3B, blue), however when unlabeled Rap1 was pre-bound to Alexa Fluor labeled TRF2, no TRF1 binding occurred (Figure B. 3B, red). Together these data describe how the shelterin proteins TIN2 and Rap1 inhibit the direct interaction between TRF1 and TRF2.

B.2.4 Determining the region on TRF2 which is responsible for interacting with TRF1

Previous studies have suggested that the TRF2 basic domain promotes the formation of telomeric RNA:DNA hybrids, while the acidic domain of TRF1 counteracts this function at telomeres (Lee et al., 2018). Therefore, we hypothesized that the basic domain of TRF2 may be responsible for directly interacting with TRF1. To that end, we

engineered a deletion mutant of TRF2 which lacked the N-terminal basic domain (TRF2ΔN) and asked how that affected TRF1 binding. Alexa Fluor labeled Rap1 bound biotinylated TRF2ΔN on streptavidin beads, suggesting that deletion of the basic domain did not result in gross protein folding defects (Figure B. 4B, red). Surprisingly, deletion of the basic domain did not impact TRF1 binding, as TRF1 bound TRF2ΔN with a K_d of 65 nM (Figure B. 4B, blue). Similarly to wild-type TRF2, addition of Rap1 prevented the TRF2ΔN-TRF1 interaction (Figure B. 4B, green). Together these data suggest that the TRF2 basic domain is not responsible for the TRF1-TRF2 interaction. Therefore, to determine which portions of TRF2 were important for interacting with TRF1, we engineered an additional four TRF2 deletion constructs: TRF2ΔTRFH (Δ aa 42-245), TRF2ΔRap1 (Δ aa 286-299), TRF2ΔTIN2 (Δ aa 352-366), and bTRFH2 (aa 1-245). We recombinantly expressed and purified wild-type TRF2 and each TRF2 truncation mutant and asked how these mutants affected TRF1 binding. Using our competition-based flow cytometry assay we found that TRF2ΔTRFH, bTRFH2, and TRF2ΔTIN2 successfully competed for TRF1 binding as well as wild-type TRF2 (Figure B. 4C). In contrast, deletion of the Rap1 binding domain or the N-terminal basic domain of TRF2 drastically reduced the binding affinity for TRF1. Interestingly, while TRF2ΔN bound TRF1 as well as wild-type TRF2 in our direct binding flow cytometry assay (Figure B. 4B), it did not compete for TRF1 binding as well as wild-type TRF2 in our competition assay (Figure B. 4C). Additionally, while TRF2ΔRap1 did not compete for TRF1 binding as well as wild-type TRF2, bTRFH2, which similarly lacks the Rap1 binding domain, was able to compete for TRF1 as well as wild-type TRF2 (Figure B.

4C). These apparently contradictory results suggest that multiple different domains in TRF2 may contribute to TRF1 binding.

B.2.5 TRF1-TRF2 weakly interact *in vivo*

We wondered if artifacts from protein labeling or our FCPIA assay could be preventing further characterization of the TRF1-TRF2 interaction. Therefore, we conducted GST-pulldown experiments using GST-TRF2 as bait and untagged wild-type TRF1 or Rap1 as prey. While GST-TRF2 efficiently pulled down Rap1 on glutathione beads (Figure B. 5A, compare lanes 2 & 7), we could not conclusively detect TRF1 pulled down by GST-TRF2 (Figure B. 5A, compare lanes 2 & 4) because degradation products of GST-TRF2 occluded visualization of TRF1. Therefore, we conducted in cell co-immunoprecipitation experiments using transiently overexpressed C-terminally FLAG or Myc-tagged TRF1 and TRF2. Although endogenous TIN2 and Rap1 would likely bind some proportion of tagged TRF1 and TRF2 and prevent their interaction, we hypothesized that over-expressing these proteins would prevent saturation with TIN2 or Rap1. While FLAG-tagged TRF2 did not pull down Myc-tagged TRF1 (Figure B. 5B lane 4), FLAG-tagged TRF1 weakly pulled down Myc-tagged TRF2 (Figure B. 5B lane 3). Together these data suggest a weak TRF1-TRF2 interaction can be detected in cells, however further research is necessary to confirm this interaction *in vitro* and validate its significance in cells.

B.3 Discussion

Using both a direct and competition-based flow cytometry protein interaction assay we discover a novel interaction between the shelterin proteins TRF1 and TRF2. We find that this interaction is not mediated by RNA or DNA, and addition of TIN2 or

Rap1 blocks TRF1-TRF2 binding. We find that TRF2 co-immunoprecitates with TRF1, however *in vitro* we were unable to determine a specific domain in TRF2 that is responsible for this interaction.

During the course of this work, a study was published that suggests the basic domain of TRF2 stimulates the formation of telomeric RNA:DNA hybrids, known as telR-loops, by specifically binding to telomeric RNA. In contrast, the N-terminal acidic domain of TRF1 counteracts this TRF2-mediated telR-loop formation, preserving telomere integrity *in vivo* (Lee et al., 2018). However, the mechanism for this repression remains unknown. It is possible that the novel interaction between TRF1 and TRF2 that we describe here contributes to the mechanism by which the acidic domain of TRF1 prevents unwanted TRF2-mediated telR-loop formation. While the basic domain of TRF2 is suggested to mediate telR-loop formation, we find that the basic domain of TRF2 is not solely responsible for binding TRF1, as deletion of the basic domain did not affect TRF1 binding in our direct binding flow cytometry protein interaction assay. However, TRF2ΔN was unable to compete for TRF1 binding as well as wild-type TRF2 in our competition-based flow cytometry assay, suggesting that these assays may not be fully complementary. Instead, it is possible that biotin labeling TRF2ΔN for our direct binding assays artificially increases the interaction between TRF2ΔN and TRF1, while use of unlabeled TRF2ΔN in our competition-based assays results in a more accurate representation of the binding affinity.

More troubling still, our contradictory results with the bTRFH2 and TRF2ΔRap1 constructs suggest that TRF2 may be incompatible with our FCPIA and/or chemical labeling. It is possible that purified TRF2 has a proclivity to non-specifically interact with

complementarily charged protein surfaces (TRF2 is highly basic) which are normally bound to other proteins, such as its constitutive binding partner Rap1. Thus, all of the purified TRF2 variants, as well as full length TRF2, would have some propensity to non-specifically bind other proteins (or the bead matrix) that they come in contact with. While it is exciting to imagine that addition of Rap1 blocks the TRF1-TRF2 interaction, it is also prudent to ask if addition of Rap1 allows for the stable formation of a TRF2-Rap1 complex which is no longer prone to erroneously binding beads or other proteins, thus preventing an artifactual TRF1 interaction.

We also find that TIN2 prevents the TRF1-TRF2 interaction. Purified TIN2 and TRF1 pose none of the problems that we found with TRF2, as these proteins were stable and less likely to nonspecifically interact with proteins or beads. Thus, this repression does not pose the same caveats as the Rap1-TRF2 complex. If the TRF1-TRF2 complex arose solely because of a nonspecific interaction with the TRF2 protein, then one would expect that TRF2 would bind to TRF1-TIN2 just as it did to TRF1 alone. Yet addition of TIN2 blocks the TRF1-TRF2 interaction, giving credence to the authenticity of this result. Therefore, future studies should focus on understanding the mechanism of this inhibition.

To that end, it is interesting to consider that the TRF1 TRFH domain could be responsible for interacting with both TIN2 and TRF2. The TRF1 TRFH domain interacts with an F-X-L-X-P-basic motif on TIN2, however previous studies have shown that variations on this sequence, such as an I-X-L-X-P-basic motif, also interact with the TRF1 TRFH domain (Pendlebury et al., 2017). Curiously, TRF2 contains an S-X-L-X-P-basic sequence between its Rap1 binding domain and TIN2 binding domain. It is

interesting to think that this motif might bind the TRF1 TRFH domain, docking in the same location as the TIN2 F-X-L-X-P-basic motif and providing an elegant mechanism for the TIN2-mediated repression of the TRF1-TRF2 interaction.

Our co-immunoprecipitation studies with overexpressed TRF1 and TRF2 suggest that FLAG-TRF1 successfully bound Myc-TRF2 *in vivo*. However the ability of Rap1 and TIN2 to block this interaction calls in to question the evolutionary and functional advantage of conserving the binding between these two proteins. Rap1 and TIN2 are significantly more abundant than TRF1 and TRF2, suggesting that a TRF1-TRF2 interaction at telomeres is unlikely. Future studies should ask if TRF1-TRF2 binding occurs *in vivo*, and if so, how and when this interaction takes place. Once these points are addressed, the most exciting questions, such as what affect this interaction has on telomere biology, can begin to be answered.

B.4 Materials and Methods

B.4.1 Molecular cloning

For bacterial expression of TRF2, full-length TRF2 sequence was cloned in to pET-smt3 vector using standard restriction endonuclease-mediated cloning. For expression of TRF2ΔN (aa 42-500) and bTRFH2 (aa 1-245), appropriate sequences were PCR amplified from full length TRF2 cDNA and cloned into the pET-Smt3 vector using standard restriction endonuclease-mediated cloning. For expression of TRF2 variants TRF2ΔTRFH (Δ aa 43-245), TRF2ΔRap1 (Δ aa 286-299), and TRF2ΔTIN2 (Δ aa 352-366), deletion constructs were cloned using ExSite PCR-based mutagenesis on

the TRF2 pET-smt3 parental vector. For expression of $TIN2_{218}$, TIN2 aa 1-218 were PCR amplified and cloned into the pET-smt3 vector using standard restriction endonuclease-mediated cloning. For insect cell expression of full length TIN2, Rap1, and TRF1, the pFBHTb-Smt3star vector was used to engineer pFBHTb-Smt3star-TIN2, pFBHTb-Smt3star-Rap1, and pFBHTb-Smt3star-TRF1 using as strategy described previously (Kocak et al., 2014). For GST-TRF2 bacterial cell expression, full length TRF2 was cloned and ligated into the pGEX-6P-2 vector (GE Life Sciences) using standard restriction endonuclease-mediated cloning. 3x-FLAG-tagged TRF1 and TRF2 and 6x-MYC-tagged TRF1 and TRF2 for human cell expression in co-immunoprecipitation experiments were cloned into the pTET-IRES-eGFP-Bi4 vector as described previously (Grill et al., 2019). All oligonucleotides used in this study were purchased from IDT.

B.4.2 Purification of protein expressed in insect cells

Protein purification was conducted as described previously (Grill et al., 2018). Full-length human TRF1, Rap1, and TIN2 were expressed as SUMOstar-$(His)_6$ fusion proteins in baculobirus-infected High Five cells (Thermo Fisher Scientific). Cells were lysed in 25 mM Tris pH 7.6, 500 mM NaCl, 10 mM BME, 0.1 mM EDTA, 1 mM PMSF, and 1 complete protease inhibitor tablet (Roche) and the resulting proteins were purified from soluble cellular lysate using Ni-NTA agarose resin (Qiagen). The SUMOstar tag was then cleaved with SUMOstar protease (LifeSensors) and the untagged protein was dialyzed overnight in 250 mM NaCl, 50 mM tris pH 7.6, and 2 mM BME. After dialysis, proteins were further purified by Superdex 200 size-exclusion chromatography (GE Healthcare Life Sciences), analyzed by SDS-PAGE, and stored in -80°C for future use.

B.4.3 Purification of protein expressed in bacteria

Protein purification was conducted as described previously (Grill et al., 2018). Full length TRF2 and all TRF2 truncation constructs were purified using an identical method, as well as truncated $TIN2_{218}$. Briefly, proteins were expressed in *E. coli* strain BL21 (DE3) as a Sumo-$(His)_{10}$-fusion protein. Cells were harvested, lysed in 50 mM Tris pH 7.6, 500 mM NaCl, 2 mM BME, 1 mM PMSF, and 1 complete protease inhibitor tablet (Roche), and centrifuged. Resulting soluble cellular lysates were purified with Ni-NTA agarose resin (Qiagen) and eluted from the column using 250 mM imidazole. The sumo tag was then cleaved using Ulp1 protease (MTA with Cornell University) and the resulting untagged protein was further purified by Superdex 200 size-exclusion chromatography (GE Healthcare Life Sciences) in 25 mM Tris-Cl (pH 8), 250 mM NaCl, and 2 mM DTT for all TRF2 constructs, and 25 mM Tris-Cl (pH 8), 150 mM NaCl, and 2 mM DTT for $TIN2_{218}$. For purification of GST-tagged TRF2, protein was expressed in *E. coli* strain BL21 (DE3) and purified using similar protocols to those previously described (Pendlebury et al., 2017) using Glutathione Sepharose 4B beads (GE Healthcare) following manufacturer's instructions. GST-TRF2 was then eluted off the glutathione beads after 10 minute incubation with glutathione. GST-TRF2 was then further purified by Superdex 200 size-exclusion chromatography (GE Healthcare Life Sciences) in 25 mM Tris-Cl (pH 8), 250 mM NaCl, and 2 mM DTT.

B.4.4 Flow-cytometry based assay for quantifying protein-protein interactions

Flow cytometry was conducted as described previously (Pendlebury et al., 2017). Briefly, proteins were labeled by reacting a 2-fold excess of label over protein concentration for 20 minutes at room temperature in the dark. Bait protein was

biotinylated using biotin maleimide (Sigma; B1267) and prey protein was fluorophore-labeled with Alexa Fluor488-C5 (Life Technologies). Unlabeled biotin or Alexa Fluor label was removed with a Micro Bio-spin 6 spin column (Bio-Rad; cat#7326002) and protein concentration was measured using Bio-Rad protein assay dye reagent (Bio-Rad; cat#5000006). Once labeled, proteins we stored at -80°C for future use.

B.4.5 Direct binding assays

150 μM of Biotin-labeled bait protein was immobilized on streptavidin beads (spherotech; SVP-20-5). Alexa Fluor labeled prey protein was then added at varying concentrations and mixed with the bead-bound bait protein in flow buffer (20 mM Tris-Cl (pH 8.0), 100 mM NaCl, 2 mM DTT, 1% bovine serum albumin, and 0.1% lubrol). Negative control lanes containing identical prey proteins with beads that were not bound to prey protein were used to calculate background signal. Proteins were then incubated on ice for 30 minutes and analyzed on an Accuri C6 flow cytometer (BD biosciences) using a Hypercyt autosampler (IntelliCyt). The median fluorescence intensity on the beads was calculated on the flow cytometer and the mean value of duplicate measurements on each plate was fitted to a single site binding model. Dissociation constants were calculated using Prism 7.0 (Graphpad) after subtracting background signal.

B.4.6 Competition experiments

The two interacting proteins were labeled with biotin or Alexa Fluor488 exactly as described above. Both biotin and Alexa Fluor labeled proteins were held at a constant concentration of 150 nM throughout the experiment and were allowed to bind streptavidin beads before incubation with unlabeled competitor protein. Unlabeled

competitor protein was then titrated into the mixture and fluorescence intensity was calculated exactly as described above. Median fluorescence data was fitted to a one-site competition curve and IC_{50} values were calculated in Prism 7.0 (Graphpad).

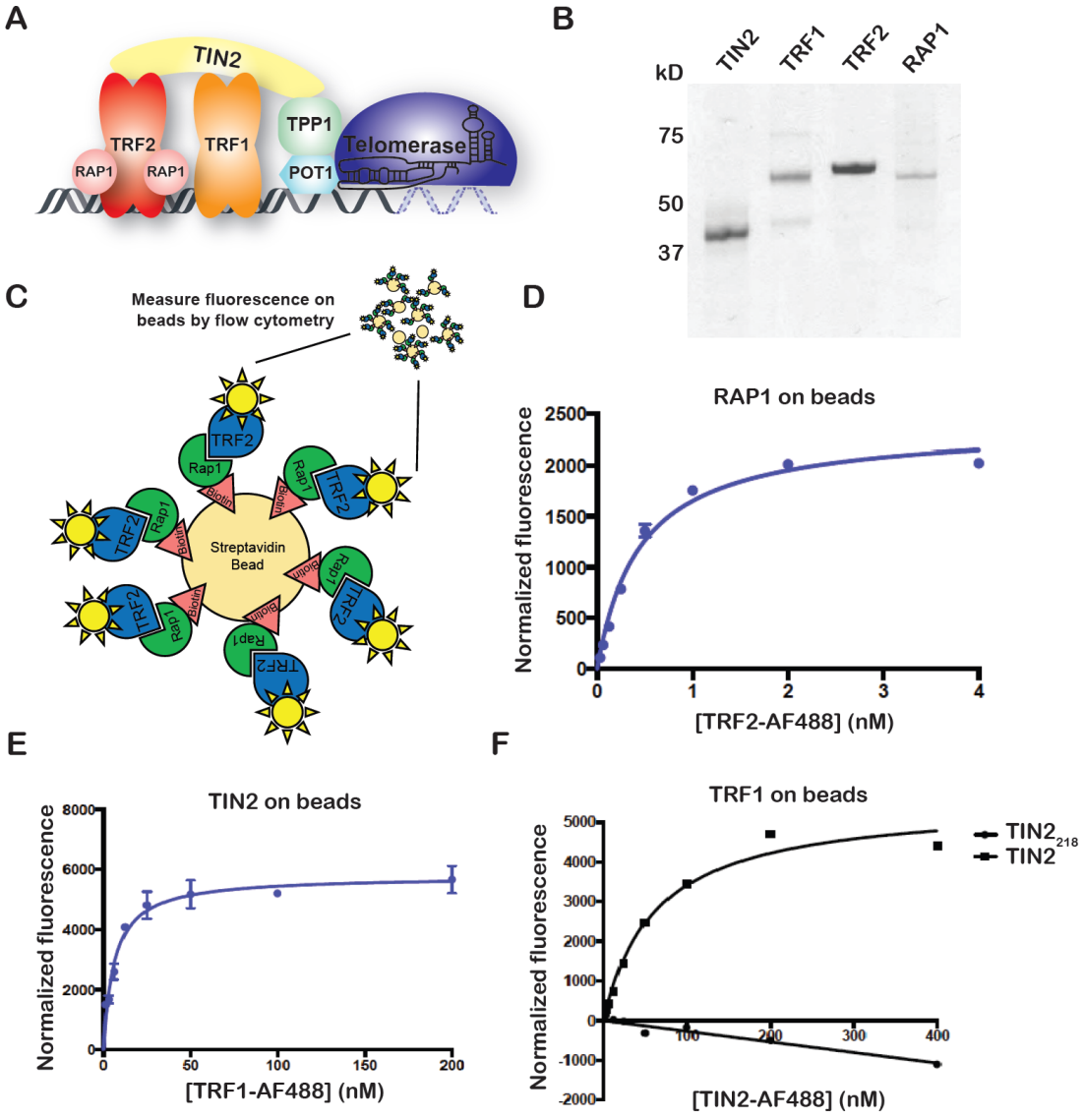
B.4.7 Co-immunoprecipitation assay for protein-protein interaction

Co-immunoprecipitation experiments were done exactly as described previously (Grill et al., 2018) using 1 µg each of plasmids containing Flag-TRF1, Flag-TRF2, Myc-TRF1, and Myc-TRF2.

B.4.8 GST pulldown assay

GST-glutathione bead pulldown experiments were done exactly as described previously (Pendlebury et al., 2017).

B.5 Figures and Tables

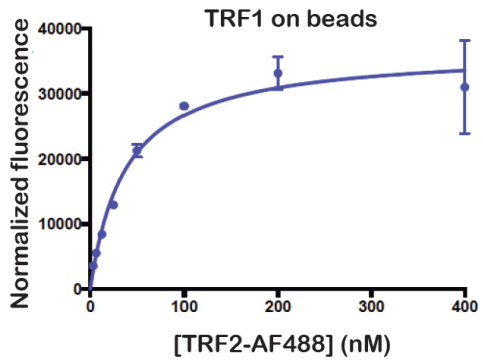


G

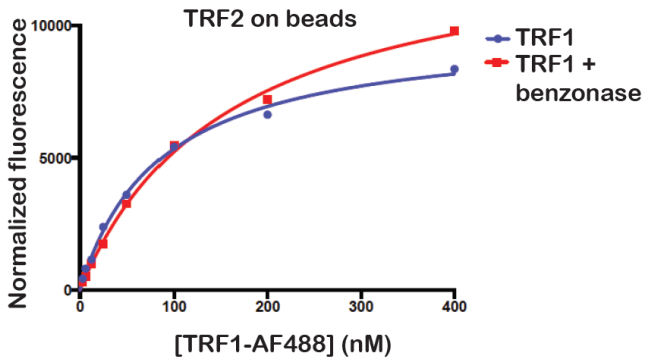
On beads	Alexafluor 488 labeled	Number of replicates	mean Kd (nM)	Std. Deviation
RAP1	TRF2	3	0.501	0.069
TIN2	TRF1	5	12.92	8.185
TRF1	TIN2	3	55.05	25.9

Figure B.1 Describing shelterin protein-protein interactions *in vitro* using a flow cytometry protein interaction assay.
A) Schematic of the six shelterin proteins along with telomerase bound to double and single stranded telomeric DNA. B) Coomassie-blue-stained SDS-PAGE gel of purified full-length TIN2, TRF1, TRF2, and Rap1 proteins used in this study. C) Schematic of a direct binding flow cytometry protein interaction assay. Alexa Fluor labeled TRF1 binds biotin labeled Rap1 on streptavidin beads and the median fluorescence intensity on the beads is measured. D) Direct association of Alexa Fluor labeled TRF2 with biotin-labeled Rap1 on streptavidin beads was scored with a flow cytometer. Fluorescence signal was background corrected with reactions containing Alexa Fluor labeled protein and unbound streptavidin beads. Mean of technical duplicates is plotted. E) Direct association of Alexa Fluor labeled TRF1 with biotin labeled TIN2 on streptavidin beads was scored with a flow cytometer exactly as in D. F) Direct association of Alexa Fluor labeled TIN2 or $TIN2_{218}$ with biotin labeled TRF1 was scored with a flow cytometer exactly as in D. G). Affinity constants related to panels D, E, and F are shown as the mean and standard deviation from the indicated number of replicates.

A



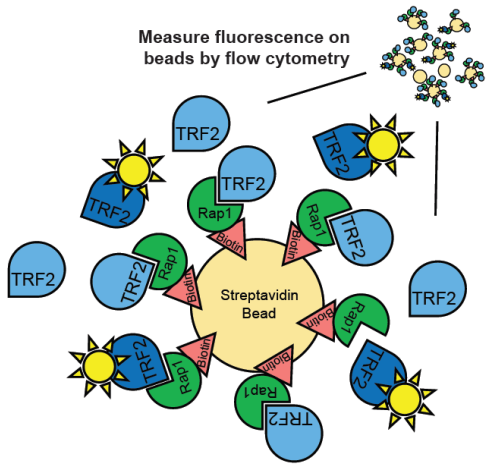
B



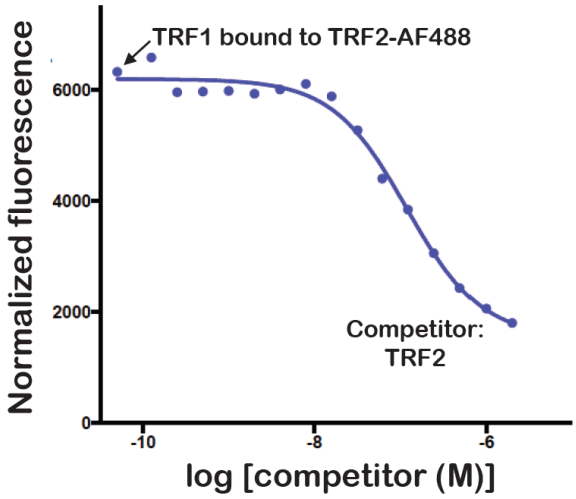
C

On beads	Alexafluor 488 labeled	Number of replicates	mean Kd (nM)	Std. Deviation
TRF1	TRF2	4	73.99	31.39
TRF2	TRF1	4	101.975	20.52
TRF2	TRF1+ benzonase	2	100.5	N/A

D



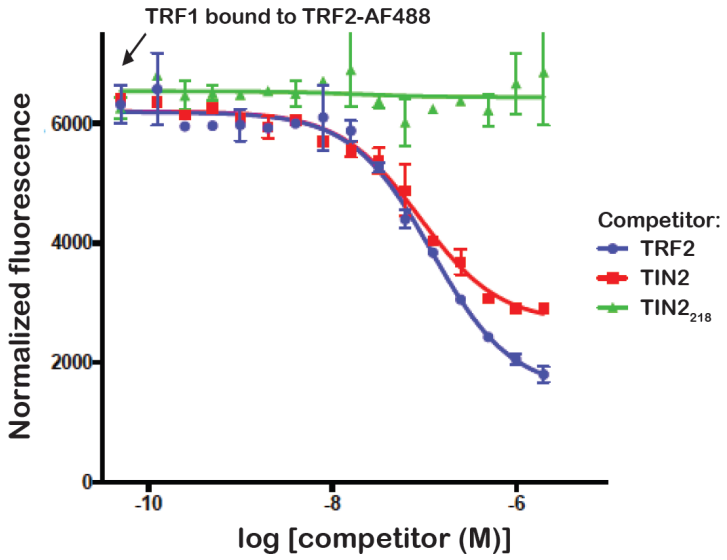
E



On beads	Alexafluor 488 labeled	Competitor	Number of replicates	Mean $\log IC_{50}$	Std. Deviation
TRF1	TRF2	TRF2	3	-7.059	0.0166

Figure B.2 TRF1 directly binds shelterin protein TRF2.
A) Direct association of Alexa Fluor labeled TRF2 with biotin labeled TRF1 on streptavidin beads was scored with a flow cytometer. Fluorescence signal was background corrected with reactions containing Alexa Fluor labeled protein and unbound streptavidin beads. Mean of technical duplicates is plotted. B) Direct association of Alexa Fluor labeled TRF1 (Blue) or Alexa Fluor labeled TRF1 + benzonase (red) with biotin labeled TRF2 on streptavidin beads was scored with a flow cytometer exactly as in A. C) Affinity constants related to panels A and B are shown as the mean and standard deviation from the number of replicates indicated. D) Schematic of the competition-based flow cytometry protein interaction assay. Alexa Fluor labeled TRF2 is prebound to biotin labeled Rap1 immobilized on streptavidin beads. Unlabeled TRF2 is then titrated in and allowed to compete for binding to biotin labeled Rap1. The median fluorescence on the beads is measured by flow cytometry. E) Fluorescence-based competition experiments using Alexa Fluor labeled TRF2 (150 nM) prebound to biotin-labeled TRF1 on streptavidin beads titrated with varying concentrations of unlabeled TRF2. Mean of technical duplicates is plotted and the mean and standard deviation of 3 biological replicates (of technical duplicates) is shown below.

A



On beads	Alexafluor 488 labeled	Competitor	Number of replicates	Mean $logIC_{50}$	Std. Deviation
TRF1	TRF2	TRF2	3	-7.059	0.0166
TRF1	TRF2	TIN2	2	-7.43	N/A

B

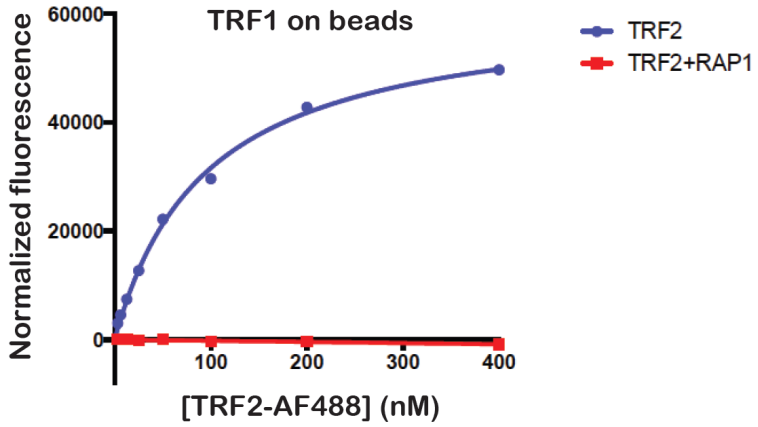
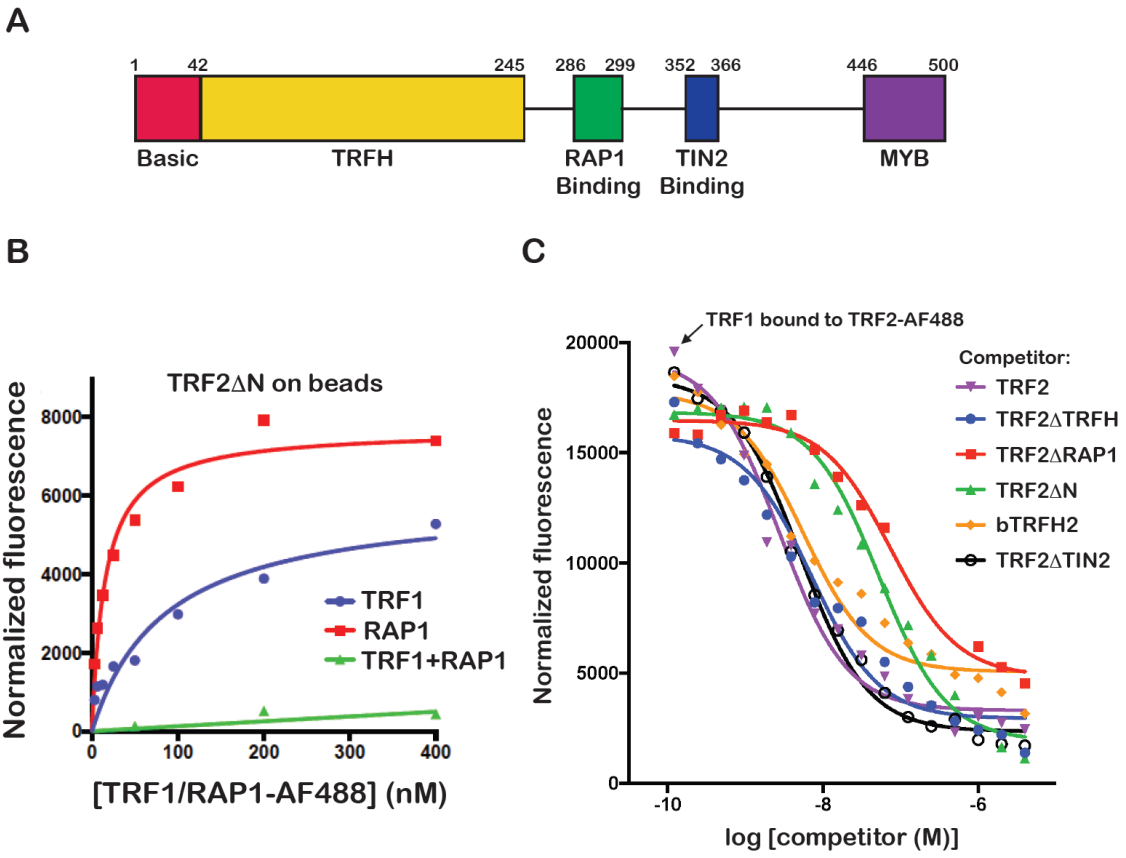


Figure B.3 Rap1 and TIN2-mediated inhibition of the TRF1-TRF2 interaction.
A) Fluorescence-based competition experiments using Alexa Fluor labeled TRF2 (150 nM) prebound to biotin-labeled TRF1 on streptavidin beads titrated with varying concentrations of unlabeled TRF2 (blue), TIN2 (red), or $TIN2_{218}$ (green). Mean of technical duplicates is plotted and the mean and standard deviation of indicated number of biological replicates (of technical duplicates) is shown below. B) Direct association of Alexa Fluor labeled TRF2 with biotin labeled TRF1 (blue) or TRF1 and Rap1 (red) on streptavidin beads was scored with a flow cytometer. Fluorescence signal was background corrected with reactions containing Alexa Fluor labeled protein and unbound streptavidin beads. Mean of technical duplicates is plotted.



On beads	Alexafluor 488 labeled	Number of replicates	mean Kd
TRF2ΔN	TRF1	2	65.2
TRF2ΔN	RAP1	1	15.48

Competitor	$\log IC_{50}$
TRF2	-8.554
TRF2ΔTRFH	-8.141
TRF2ΔRAP1	-7.13
TRF2ΔN	-7.276
bTRFH2	-8.259
TRF2ΔTIN2	-8.295

Figure B.4 Determining the TRF1 binding region of TRF2.
A) Domain diagram depicting the basic, TRFH, Rap1 binding, TIN2 binding, and MYB domains of human TRF2. B) Direct association of Alexa Fluor labeled TRF1 (blue), Rap1 (red), or TRF1 and Rap1 (green) with biotin-labeled TRF2ΔN on streptavidin beads. Fluorescence signal was background corrected with reactions containing Alexa Fluor labeled protein and unbound streptavidin beads. Mean of technical duplicates is plotted and the mean and standard deviation of indicated number of biological replicates (of technical duplicates) is shown below. B) Fluorescence-based competition experiments using Alexa Fluor labeled TRF2 (150 nM) prebound to biotin-labeled TRF1 on streptavidin beads titrated with varying concentrations of the indicated unlabeled TRF2 variant. Mean of technical duplicates is plotted and the $\log IC_{50}$ of the indicated competitor is shown below.

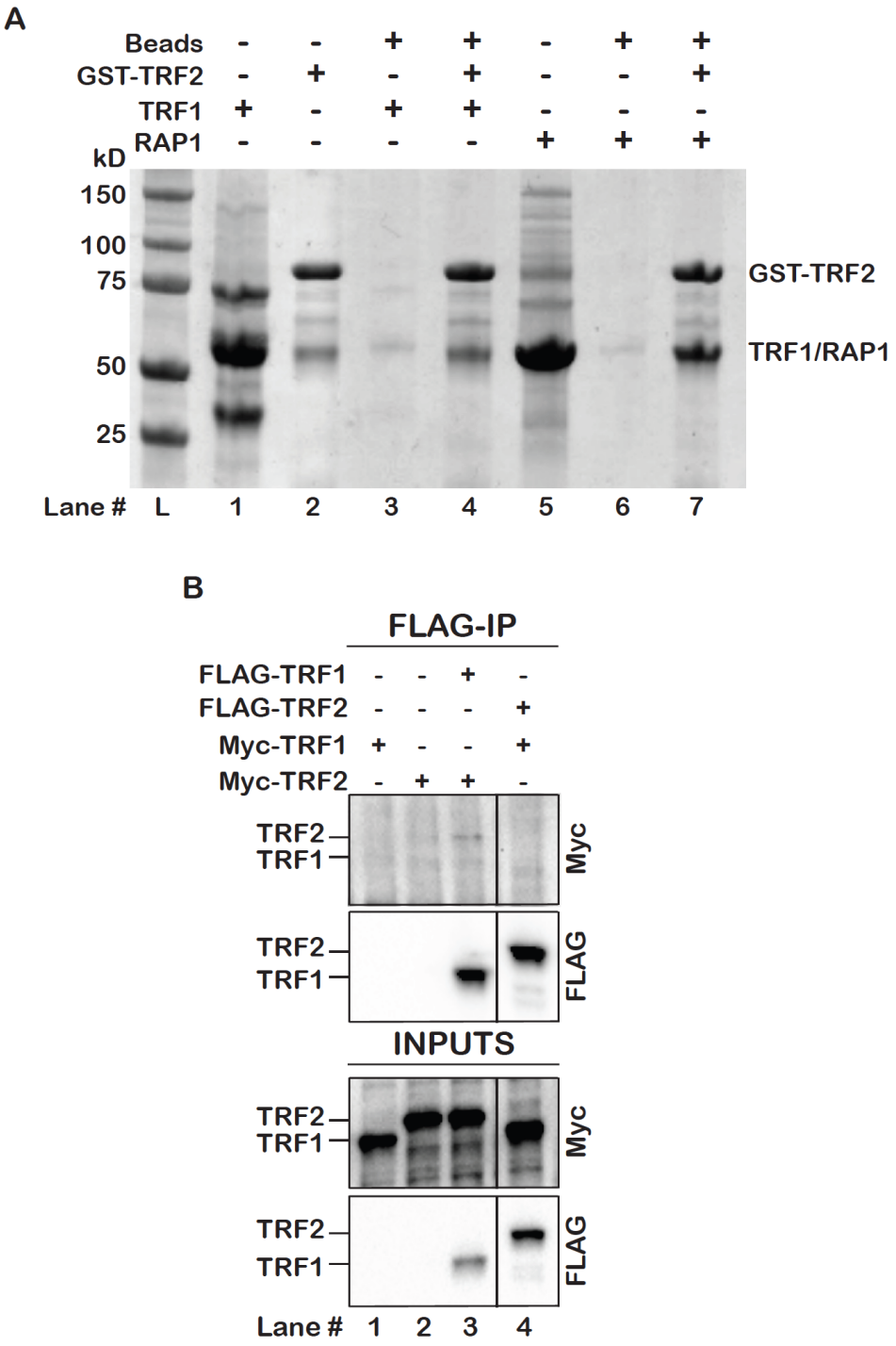


Figure B.5 TRF2 binds TRF1 *in vivo*.
A) Pulldown of TRF1 or RAP1 using GST-TRF2 on glutathione Sepharose beads. B) Pulldown of either FLAG-TRF1 or FLAG-TRF2 and Myc-TRF2 or Myc-TRF1 transiently expressed in HeLa cells on anti-FLAG conjugated beads. "Input" and "FLAG-IP" immunoblots refer to soluble lysates before overnight incubation with anti-FLAG beads and anti-FLAG beads after overnight immunoprecipitation, respectively.

Bibliography

Abreu, E., Aritonovska, E., Reichenbach, P., Cristofari, G., Culp, B., Terns, R.M., Lingner, J., and Terns, M.P. (2010). TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. Mol Cell Biol *30*, 2971-2982.

Affymetrix, E.T.P., and Cold Spring Harbor Laboratory, E.T.P. (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature *457*, 1028-1032.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Antunes, D.M., Kalmbach, K.H., Wang, F., Dracxler, R.C., Seth-Smith, M.L., Kramer, Y., Buldo-Licciardi, J., Kohlrausch, F.B., and Keefe, D.L. (2015). A single-cell assay for telomere DNA content shows increasing telomere length heterogeneity, as well as increasing mean telomere length in human spermatozoa with advancing age. J Assist Reprod Genet *32*, 1685-1690.

Aubert, G. (2014). Telomere dynamics and aging. Prog Mol Biol Transl Sci *125*, 89-111.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science *318*, 798-801.

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. Genome Biol *8*, 227.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Baumann, P., and Cech, T.R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. Science *292*, 1171-1175.

Bell, R.J., Rube, H.T., Kreig, A., Mancini, A., Fouse, S.D., Nagarajan, R.P., Choi, S., Hong, C., He, D., Pekmezci, M.*, et al.* (2015). Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. Science *348*, 1036-1039.

Bianchi, A., Smith, S., Chong, L., Elias, P., and de Lange, T. (1997). TRF1 is a dimer and bends telomeric DNA. EMBO J *16*, 1785-1794.

Bianchi, A., Stansel, R.M., Fairall, L., Griffith, J.D., Rhodes, D., and de Lange, T. (1999). TRF1 binds a bipartite telomeric site with extreme spatial flexibility. EMBO J *18*, 5735-5744.

Bisht, K., Grill, S., Graniel, J., and Nandakumar, J. (2017). A lentivirus-free inducible CRISPR-Cas9 system for efficient targeting of human genes. Anal Biochem *530*, 40-49.

Bisht, K., Smith, E.M., Tesmer, V.M., and Nandakumar, J. (2016). Structural and functional consequences of a disease mutation in the telomere protein TPP1. Proc Natl Acad Sci U S A *113*, 13021-13026.

Blackburn, E.H., and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. Cold Spring Harb Perspect Biol *3*.

Blackburn, E.H., and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol *120*, 33-53.

Blackburn, E.H., Greider, C.W., Henderson, E., Lee, M.S., Shampay, J., and Shippen-Lentz, D. (1989). Recognition and elongation of telomeres by telomerase. Genome *31*, 553-560.

Blazer, L.L., Roman, D.L., Muxlow, M.R., and Neubig, R.R. (2010). Use of flow cytometric methods to quantify protein-protein interactions. Curr Protoc Cytom *Chapter 13*, Unit 13.11.11-15.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. Science *279*, 349-352.

Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. Nat Genet *17*, 231-235.

Bryan, C., Rice, C., Hoffman, H., Harkisheimer, M., Sweeney, M., and Skordalakes, E. (2015). Structural Basis of Telomerase Inhibition by the Highly Specific BIBR1532. Structure *23*, 1934-1942.

Campisi, J. (1997). The biology of replicative senescence. Eur J Cancer *33*, 703-709.

Cao, J., Wu, L., Zhang, S.M., Lu, M., Cheung, W.K., Cai, W., Gale, M., Xu, Q., and Yan, Q. (2016). An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. Nucleic Acids Res *44*, e149.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M.*, et al.* (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics *37*, 327-336.

Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Karolchik, D.*, et al.* (2018). The UCSC Genome Browser database: 2018 update. Nucleic Acids Res *46*, D762-D769.

Celli, G.B., and de Lange, T. (2005). DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. Nat Cell Biol *7*, 712-718.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S.*, et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

Chen, C., Gu, P., Wu, J., Chen, X., Niu, S., Sun, H., Wu, L., Li, N., Peng, J., Shi, S.*, et al.* (2017). Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer. Nat Commun *8*, 14929.

Chen, J.L., and Greider, C.W. (2003a). Determinants in mammalian telomerase RNA that mediate enzyme processivity and cross-species incompatibility. EMBO J *22*, 304-314.

Chen, J.L., and Greider, C.W. (2003b). Template boundary definition in mammalian telomerase. Genes Dev *17*, 2747-2752.

Chen, J.L., Opperman, K.K., and Greider, C.W. (2002). A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. Nucleic Acids Res *30*, 592-597.

Chen, L.Y., Liu, D., and Songyang, Z. (2007). Telomere maintenance through spatial control of telomeric proteins. Mol Cell Biol *27*, 5898-5909.

Chen, Y., Rai, R., Zhou, Z.R., Kanoh, J., Ribeyre, C., Yang, Y., Zheng, H., Damay, P., Wang, F., Tsujii, H.*, et al.* (2011). A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. Nat Struct Mol Biol *18*, 213-221.

Chen, Y., Yang, Y., van Overbeek, M., Donigian, J.R., Baciu, P., de Lange, T., and Lei, M. (2008). A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. Science *319*, 1092-1096.

Chiba, K., Lorbeer, F.K., Shain, A.H., McSwiggen, D.T., Schruf, E., Oh, A., Ryu, J., Darzacq, X., Bastian, B.C., and Hockemeyer, D. (2017). Mutations in the promoter of the telomerase gene. Science *357*, 1416-1420.

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., and de Lange, T. (1995). A human telomeric protein. Science *270*, 1663-1667.

Chong, P.A., Vernon, R.M., and Forman-Kay, J.D. (2018). RGG/RG Motif Regions in RNA Binding and Phase Separation. J Mol Biol *430*, 4650-4665.

Chu, T.W., D'Souza, Y., and Autexier, C. (2016). 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 *36*, 210-222.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Conrad, M.N., Wright, J.H., Wolf, A.J., and Zakian, V.A. (1990). RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. Cell *63*, 739-750.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet *46*, 1311-1320.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005). How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. EMBO Rep *6*, 39-45.

Cox, D.B., Platt, R.J., and Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. Nature medicine *21*, 121-131.

Cristofari, G., and Lingner, J. (2006). Telomere length homeostasis requires that telomerase levels are limiting. EMBO J *25*, 565-574.

de Lange, T. (2018). Shelterin-Mediated Telomere Protection. Annu Rev Genet *52*, 223-247.

de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M., and Varmus, H.E. (1990). Structure and variability of human chromosome ends. Mol Cell Biol *10*, 518-527.

Denchi, E.L., and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. Nature *448*, 1068-1071.

Diolaiti, M.E., Cimini, B.A., Kageyama, R., Charles, F.A., and Stohr, B.A. (2013). In situ visualization of telomere elongation patterns in human cells. Nucleic Acids Res *41*, e176.

Dokal, I. (2011). Dyskeratosis congenita. Hematology Am Soc Hematol Educ Program *2011*, 480-486.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nature biotechnology *33*, 390-394.

Egan, E.D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. RNA *18*, 1747-1759.

Else, T., Theisen, B.K., Wu, Y., Hutz, J.E., Keegan, C.E., Hammer, G.D., and Ferguson, D.O. (2007). Tpp1/Acd maintains genomic stability through a complex role in telomere protection. Chromosome Res *15*, 1001-1013.

Fairall, L., Chapman, L., Moss, H., de Lange, T., and Rhodes, D. (2001). Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. Mol Cell *8*, 351-361.

Fayomi, A.P., and Orwig, K.E. (2018). Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. Stem cell research *29*, 207-214.

Frank, A.K., Tran, D.C., Qu, R.W., Stohr, B.A., Segal, D.J., and Xu, L. (2015). The Shelterin TIN2 Subunit Mediates Recruitment of Telomerase to Telomeres. PLoS Genet *11*, e1005410.

Frescas, D., and de Lange, T. (2014a). Binding of TPP1 protein to TIN2 protein is required for POT1a,b protein-mediated telomere protection. J Biol Chem *289*, 24180-24187.

Frescas, D., and de Lange, T. (2014b). TRF2-tethered TIN2 can mediate telomere protection by TPP1/POT1. Mol Cell Biol *34*, 1349-1362.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R.*, et al.* (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91-100.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Gillis, A.J., Schuller, A.P., and Skordalakes, E. (2008). Structure of the Tribolium castaneum telomerase catalytic subunit TERT. Nature *455*, 633-637.

Glousker, G., Touzot, F., Revy, P., Tzfati, Y., and Savage, S.A. (2015). Unraveling the pathogenesis of Hoyeraal-Hreidarsson syndrome, a complex telomere biology disorder. Br J Haematol *170*, 457-471.

Gomes, N.M., Ryder, O.A., Houck, M.L., Charter, S.J., Walker, W., Forsyth, N.R., Austad, S.N., Venditti, C., Pagel, M., Shay, J.W.*, et al.* (2011). Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. Aging Cell *10*, 761-768.

Gonzalez, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell stem cell *15*, 215-226.

Gottschling, D.E., and Zakian, V.A. (1986). Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. Cell *47*, 195-205.

Greider, C.W. (1991). Telomerase is processive. Mol Cell Biol *11*, 4572-4580.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell *43*, 405-413.

Greider, C.W., and Blackburn, E.H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell *51*, 887-898.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331-337.

Griffith, J., Bianchi, A., and de Lange, T. (1998). TRF1 promotes parallel pairing of telomeric tracts in vitro. J Mol Biol *278*, 79-88.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. Cell *97*, 503-514.

Grill, S., Bisht, K., Tesmer, V.M., Shami, A.N., Hammoud, S.S., and Nandakumar, J. (2019). Two Separation-of-Function Isoforms of Human TPP1 Dictate Telomerase Regulation in Somatic and Germ Cells. Cell Rep *27*, 3511-3521.e3517.

Grill, S., Tesmer, V.M., and Nandakumar, J. (2018). The N Terminus of the OB Domain of Telomere Protein TPP1 Is Critical for Telomerase Action. Cell Rep *22*, 1132-1140.

Guo, X., Deng, Y., Lin, Y., Cosme-Blanco, W., Chan, S., He, H., Yuan, G., Brown, E.J., and Chang, S. (2007). Dysfunctional telomeres activate an ATM-ATR-dependent DNA damage response to suppress tumorigenesis. EMBO J *26*, 4709-4719.

Guo, Y., Kartawinata, M., Li, J., Pickett, H.A., Teo, J., Kilo, T., Barbaro, P.M., Keating, B., Chen, Y., Tian, L.*, et al.* (2014). Inherited bone marrow failure associated with

germline mutation of ACD, the gene encoding telomere protein TPP1. Blood *124*, 2767-2774.

Günes, C., and Rudolph, K.L. (2013). The role of telomeres in stem cells and cancer. Cell *152*, 390-393.

H.J., M. (1938). The remaking of chromosomes. (Collecting Net), pp. 182-195.

Hahn, W.C., Counter, C.M., Lundberg, A.S., Beijersbergen, R.L., Brooks, M.W., and Weinberg, R.A. (1999). Creation of human tumour cells with defined genetic elements. Nature *400*, 464-468.

Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. RNA *14*, 2572-2579.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V., 3rd, Graveley, B.R., Terns, R.M.*, et al.* (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol Cell *45*, 292-302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanaoka, S., Nagadoi, A., Yoshimura, S., Aimoto, S., Li, B., de Lange, T., and Nishimura, Y. (2001). NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains. J Mol Biol *312*, 167-175.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S.*, et al.* (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res *22*, 1760-1774.

HAYFLICK, L. (1965). THE LIMITED IN VITRO LIFETIME OF HUMAN DIPLOID CELL STRAINS. Exp Cell Res *37*, 614-636.

HAYFLICK, L., and MOORHEAD, P.S. (1961). The serial cultivation of human diploid cell strains. Exp Cell Res *25*, 585-621.

Higuchi, R., Krummel, B., and Saiki, R.K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res *16*, 7351-7367.

Hockemeyer, D., Palm, W., Else, T., Daniels, J.P., Takai, K.K., Ye, J.Z., Keegan, C.E., de Lange, T., and Hammer, G.D. (2007). Telomere protection by mammalian Pot1 requires interaction with Tpp1. Nat Struct Mol Biol *14*, 754-761.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K.*, et al.* (2013). TERT promoter mutations in familial and sporadic melanoma. Science *339*, 959-961.

Horvath, M.P., Schweiker, V.L., Bevilacqua, J.M., Ruggles, J.A., and Schultz, S.C. (1998). Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA. Cell *95*, 963-974.

Houghtaling, B.R., Cuttonaro, L., Chang, W., and Smith, S. (2004). A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. Curr Biol *14*, 1621-1631.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Hu, C., Rai, R., Huang, C., Broton, C., Long, J., Xu, Y., Xue, J., Lei, M., Chang, S., and Chen, Y. (2017). Structural and functional analyses of the mammalian TIN2-TPP1-TRF2 telomeric complex. Cell Res *27*, 1485-1502.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957-959.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Hwang, H., Buncher, N., Opresko, P.L., and Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. Structure *20*, 1872-1880.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of bacteriology *169*, 5429-5433.

Izadyar, F., Wong, J., Maki, C., Pacchiarotti, J., Ramos, T., Howerton, K., Yuen, C., Greilach, S., Zhao, H.H., Chow, M.*, et al.* (2011). Identification and characterization of repopulating spermatogonial stem cells from the adult human testis. Human reproduction *26*, 1296-1306.

Jafri, M.A., Ansari, S.A., Alqahtani, M.H., and Shay, J.W. (2016). Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. Genome Med *8*, 69.

Janouskova, E., Necasova, I., Pavlouskova, J., Zimmermann, M., Hluchy, M., Marini, V., Novakova, M., and Hofr, C. (2015). Human Rap1 modulates TRF2 attraction to telomeric DNA. Nucleic Acids Res *43*, 2691-2700.

Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565-1575.

Jiang, J., Wang, Y., Sušac, L., Chan, H., Basu, R., Zhou, Z.H., and Feigon, J. (2018). Structure of Telomerase with Telomeric DNA. Cell *173*, 1179-1190.e1113.

Jiang, W., and Marraffini, L.A. (2015). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annual review of microbiology *69*, 209-228.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jones, M., Bisht, K., Savage, S.A., Nandakumar, J., Keegan, C.E., and Maillard, I. (2016). The shelterin complex and hematopoiesis. J Clin Invest *126*, 1621-1629.

Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. J Cell Biol *164*, 647-652.

Kabir, S., Hockemeyer, D., and de Lange, T. (2014). TALEN gene knockouts reveal no requirement for the conserved human shelterin protein Rap1 in telomere protection and length regulation. Cell Rep *9*, 1273-1280.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L.*, et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and de Lange, T. (1999). p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. Science *283*, 1321-1325.

Karlseder, J., Hoke, K., Mirzoeva, O.K., Bakkenist, C., Kastan, M.B., Petrini, J.H., and de Lange, T. (2004). The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. PLoS Biol *2*, E240.

Karlseder, J., Kachatrian, L., Takai, H., Mercer, K., Hingorani, S., Jacks, T., and de Lange, T. (2003). Targeted deletion reveals an essential function for the telomere length regulator Trf1. Mol Cell Biol *23*, 6533-6541.

Keegan, C.E., Hutz, J.E., Else, T., Adamska, M., Shah, S.P., Kent, A.E., Howes, J.M., Beamer, W.G., and Hammer, G.D. (2005). Urogenital and caudal dysgenesis in

adrenocortical dysplasia (acd) mice is caused by a splicing mutation in a novel telomeric regulator. Hum Mol Genet *14*, 113-123.

Kelleher, C., Kurth, I., and Lingner, J. (2005). Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro. Mol Cell Biol *25*, 808-818.

Kibe, T., Zimmermann, M., and de Lange, T. (2016). TPP1 Blocks an ATR-Mediated Resection Mechanism at Telomeres. Mol Cell *61*, 236-246.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science *266*, 2011-2015.

Kim, S.H., Beausejour, C., Davalos, A.R., Kaminker, P., Heo, S.J., and Campisi, J. (2004). TIN2 mediates functions of TRF2 at human telomeres. J Biol Chem *279*, 43799-43804.

Kim, S.H., Kaminker, P., and Campisi, J. (1999). TIN2, a new regulator of telomere length in human cells. Nat Genet *23*, 405-412.

Kiss, T., Fayet-Lebaron, E., and Jády, B.E. (2010). Box H/ACA small ribonucleoproteins. Mol Cell *37*, 597-606.

Klobutcher, L.A., Swanton, M.T., Donini, P., and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci U S A *78*, 3015-3019.

Kocak, H., Ballew, B.J., Bisht, K., Eggebeen, R., Hicks, B.D., Suman, S., O'Neil, A., Giri, N., Maillard, I., Alter, B.P.*, et al.* (2014). Hoyeraal-Hreidarsson syndrome caused by a germline mutation in the TEL patch of the telomere protein TPP1. Genes Dev *28*, 2090-2102.

Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. Nature *500*, 472-476.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology *8*, R61.

Kunne, T., Swarts, D.C., and Brouns, S.J. (2014). Planting the seed: target recognition of short guide RNAs. Trends in microbiology *22*, 74-83.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nature reviews Microbiology *8*, 317-327.

Landry, J.J., Pyl, P.T., Rausch, T., Zichner, T., Tekkedil, M.M., Stutz, A.M., Jauch, A., Aiyar, R.S., Pau, G., Delhomme, N.*, et al.* (2013). The genomic and transcriptomic landscape of a HeLa cell line. G3 *3*, 1213-1224.

Lee, Y.W., Arora, R., Wischnewski, H., and Azzalin, C.M. (2018). TRF1 participates in chromosome end protection by averting TRF2-dependent telomeric R loops. Nat Struct Mol Biol *25*, 147-153.

Lei, M., Podell, E.R., Baumann, P., and Cech, T.R. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. Nature *426*, 198-203.

Lei, M., Podell, E.R., and Cech, T.R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat Struct Mol Biol *11*, 1223-1229.

Lesch, B.J., Silber, S.J., McCarrey, J.R., and Page, D.C. (2016). Parallel evolution of male germline epigenetic poising and somatic development in animals. Nat Genet *48*, 888-894.

Li, B., and de Lange, T. (2003). Rap1 affects the length and heterogeneity of human telomeres. Mol Biol Cell *14*, 5060-5068.

Li, B., Oestreich, S., and de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. Cell *101*, 471-483.

Lim, C.J., Zaug, A.J., Kim, H.J., and Cech, T.R. (2017). Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. Nat Commun *8*, 1075.

Lindsey, J., McGill, N.I., Lindsey, L.A., Green, D.K., and Cooke, H.J. (1991). In vivo loss of telomeric repeats with age in humans. Mutat Res *256*, 45-48.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science *276*, 561-567.

Liu, D., O'Connor, M.S., Qin, J., and Songyang, Z. (2004a). Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. J Biol Chem *279*, 51338-51342.

Liu, D., Safari, A., O'Connor, M.S., Chan, D.W., Laegeler, A., Qin, J., and Songyang, Z. (2004b). PTOP interacts with POT1 and regulates its localization to telomeres. Nat Cell Biol *6*, 673-680.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004c). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Loayza, D., Parsons, H., Donigian, J., Hoke, K., and de Lange, T. (2004). DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. J Biol Chem *279*, 13241-13248.

Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194-1217.

Lubin, J.W., Rao, T., Mandell, E.K., Wuttke, D.S., and Lundblad, V. (2013). Dissecting protein function: an efficient protocol for identifying separation-of-function mutations that encode structurally stable proteins. Genetics *193*, 715-725.

Lundblad, V., and Szostak, J.W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell *57*, 633-643.

Lustig, A.J., Kurtz, S., and Shore, D. (1990). Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. Science *250*, 549-553.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E.*, et al.* (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423-427.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat Methods *10*, 977-979.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F.*, et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nature reviews Microbiology *9*, 467-477.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature biotechnology *31*, 833-838.

McClintock, B. (1939). The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. Proc Natl Acad Sci U S A *25*, 405-416.

McCloskey, A., Taniguchi, I., Shinmyozu, K., and Ohno, M. (2012). hnRNP C tetramer measures RNA length to classify RNA polymerase II transcripts for export. Science *335*, 1643-1646.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Mercer, T.R., Wilhelm, D., Dinger, M.E., Solda, G., Korbie, D.J., Glazov, E.A., Truong, V., Schwenke, M., Simons, C., Matthaei, K.I.*, et al.* (2011). Expression of distinct RNAs from 3' untranslated regions. Nucleic Acids Res *39*, 2393-2403.

Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q.*, et al.* (1997). hEST2, the

putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell *90*, 785-795.

Michel, A.M., Fox, G., M Kiran, A., De Bo, C., O'Connor, P.B., Heaphy, S.M., Mullan, J.P., Donohue, C.A., Higgins, D.G., and Baranov, P.V. (2014). GWIPS-viz: development of a ribo-seq genome browser. Nucleic Acids Res *42*, D859-864.

Mitchell, J.R., and Collins, K. (2000). Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. Mol Cell *6*, 361-371.

Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H., and Skordalakes, E. (2010). Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. Nat Struct Mol Biol *17*, 513-518.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015). Regulatory small RNAs from the 3' regions of bacterial mRNAs. Current opinion in microbiology *24*, 132-139.

Miyoshi, T., Kanoh, J., Saito, M., and Ishikawa, F. (2008). Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length. Science *320*, 1341-1344.

Mojica, F.J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol *36*, 244-246.

Morgan, T.H. (1911). RANDOM SEGREGATION VERSUS COUPLING IN MENDELIAN INHERITANCE. Science *34*, 384.

Morita, S., Horii, T., Kimura, M., Goto, Y., Ochiya, T., and Hatada, I. (2007). One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics *89*, 687-696.

Mossessova, E., and Lima, C.D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell *5*, 865-876.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J.R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A *85*, 6622-6626.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. Science *277*, 955-959.

Nakashima, M., Nandakumar, J., Sullivan, K.D., Espinosa, J.M., and Cech, T.R. (2013). Inhibition of telomerase recruitment and cancer cell death. J Biol Chem *288*, 33171-33180.

Nandakumar, J., Bell, C.F., Weidenfeld, I., Zaug, A.J., Leinwand, L.A., and Cech, T.R. (2012). The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. Nature *492*, 285-289.

Nandakumar, J., and Cech, T.R. (2013). Finding the end: recruitment of telomerase to telomeres. Nat Rev Mol Cell Biol *14*, 69-82.

Nandakumar, J., Podell, E.R., and Cech, T.R. (2010). How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. Proc Natl Acad Sci U S A *107*, 651-656.

Nikitina, T., and Woodcock, C.L. (2004). Closed chromatin loops at the ends of chromosomes. J Cell Biol *166*, 161-165.

Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., and Nureki, O. (2015). Crystal Structure of Staphylococcus aureus Cas9. Cell *162*, 1113-1126.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

O'Connor, M.S., Safari, A., Xin, H., Liu, D., and Songyang, Z. (2006). A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. Proc Natl Acad Sci U S A *103*, 11874-11879.

Opresko, P.L., and Shay, J.W. (2017). Telomere-associated aging disorders. Ageing Res Rev *33*, 52-66.

Palm, W., and de Lange, T. (2008). How shelterin protects mammalian telomeres. Annu Rev Genet *42*, 301-334.

Pendlebury, D.F., Fujiwara, Y., Tesmer, V.M., Smith, E.M., Shibuya, H., Watanabe, Y., and Nandakumar, J. (2017). Dissecting the telomere-inner nuclear membrane interface formed in meiosis. Nat Struct Mol Biol *24*, 1064-1072.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W.*, et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973-976.

Podlevsky, J.D., and Chen, J.J. (2012). It all comes together at the ends: telomerase structure, function, and biogenesis. Mutat Res *730*, 3-11.

Prowse, K.R., and Greider, C.W. (1995). Developmental and tissue-specific regulation of mouse telomerase and telomere length. Proc Natl Acad Sci U S A *92*, 4818-4822.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Qiao, F., and Cech, T.R. (2008). Triple-helix structure in telomerase RNA contributes to catalysis. Nat Struct Mol Biol *15*, 634-640.

Rai, R., Chen, Y., Lei, M., and Chang, S. (2016). TRF2-RAP1 is required to protect telomeres from engaging in homologous recombination-mediated deletions and fusions. Nat Commun *7*, 10881.

Rai, R., Hu, C., Broton, C., Chen, Y., Lei, M., and Chang, S. (2017). NBS1 Phosphorylation Status Dictates Repair Choice of Dysfunctional Telomeres. Mol Cell *65*, 801-817.e804.

Rajavel, M., Orban, T., Xu, M., Hernandez-Sanchez, W., de la Fuente, M., Palczewski, K., and Taylor, D.J. (2016). Dynamic peptides of human TPP1 fulfill diverse functions in telomere maintenance. Nucleic Acids Res *44*, 10467-10479.

Rand, T.A., Ginalski, K., Grishin, N.V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci U S A *101*, 14385-14389.

Rao, T., Lubin, J.W., Armstrong, G.S., Tucey, T.M., Lundblad, V., and Wuttke, D.S. (2014). Structure of Est3 reveals a bimodal surface with differential roles in telomere replication. Proc Natl Acad Sci U S A *111*, 214-218.

Ravindranathan, A., Cimini, B., Diolaiti, M.E., and Stohr, B.A. (2018). Preliminary development of an assay for detection of TERT expression, telomere length, and telomere elongation in single cells. PLoS One *13*, e0206525.

Rice, C., Shastrula, P.K., Kossenkov, A.V., Hills, R., Baird, D.M., Showe, L.C., Doukov, T., Janicki, S., and Skordalakes, E. (2017). Structural and functional analysis of the human POT1-TPP1 telomeric complex. Nat Commun *8*, 14928.

Rivera, T., Haggblom, C., Cosconati, S., and Karlseder, J. (2017). A balance between elongation and trimming regulates telomere stability in stem cells. Nat Struct Mol Biol *24*, 30-39.

Rudel, S., Flatley, A., Weinmann, L., Kremmer, E., and Meister, G. (2008). A multifunctional human Argonaute2-specific monoclonal antibody. RNA *14*, 1244-1253.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology *32*, 347-355.

Sandhu, R., Wei, D., Sharma, M., and Xu, L. (2019). An N-terminal Flag-tag impairs TPP1 regulation of telomerase function. Biochem Biophys Res Commun *512*, 230-235.

Sasa, G.S., Ribes-Zamora, A., Nelson, N.D., and Bertuch, A.A. (2012). Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood. Clin Genet *81*, 470-478.

Savage, S.A. (2014). Human telomeres and telomere biology disorders. Prog Mol Biol Transl Sci *125*, 41-66.

Savage, S.A. (2018). Beginning at the ends: telomeres and human disease. F1000Res *7*.

Savage, S.A., and Bertuch, A.A. (2010). The genetics and clinical manifestations of telomere biology disorders. Genet Med *12*, 753-764.

Savage, S.A., Giri, N., Baerlocher, G.M., Orr, N., Lansdorp, P.M., and Alter, B.P. (2008). TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. Am J Hum Genet *82*, 501-509.

Schmidt, J.C., and Cech, T.R. (2015). Human telomerase: biogenesis, trafficking, recruitment, and activation. Genes Dev *29*, 1095-1105.

Schmidt, J.C., Dalby, A.B., and Cech, T.R. (2014). Identification of human TERT elements necessary for telomerase recruitment to telomeres. Elife *3*.

Seth, K., and Harish (2016). Current status of potential applications of repurposed Cas9 for structural and functional genomics of plants. Biochem Biophys Res Commun *480*, 499-507.

Sexton, A.N., Regalado, S.G., Lai, C.S., Cost, G.J., O'Neil, C.M., Urnov, F.D., Gregory, P.D., Jaenisch, R., Collins, K., and Hockemeyer, D. (2014). Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. Genes Dev *28*, 1885-1899.

Sexton, A.N., Youmans, D.T., and Collins, K. (2012). Specificity requirements for human telomere protein interaction with telomerase holoenzyme. J Biol Chem *287*, 34455-34464.

Sfeir, A., Kabir, S., van Overbeek, M., Celli, G.B., and de Lange, T. (2010). Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. Science *327*, 1657-1661.

Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell *138*, 90-103.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Shay, J.W., and Wright, W.E. (2010). Telomeres and telomerase in normal and cancer stem cells. FEBS Lett *584*, 3819-3825.

Shay, J.W., and Wright, W.E. (2011). Role of telomeres and telomerase in cancer. Semin Cancer Biol *21*, 349-353.

Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N.B., and Tzfati, Y. (2007). A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. Mol Cell Biol *27*, 2130-2143.

Shibuya, H., Ishiguro, K., and Watanabe, Y. (2014). The TRF1-binding protein TERB1 promotes chromosome movement and telomere rigidity in meiosis. Nat Cell Biol *16*, 145-156.

Stansel, R.M., de Lange, T., and Griffith, J.D. (2001). T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. EMBO J *20*, 5532-5540.

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z.*, et al.* (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144-149.

Takai, H., Smogorzewska, A., and de Lange, T. (2003). DNA damage foci at dysfunctional telomeres. Curr Biol *13*, 1549-1556.

Takai, K.K., Hooper, S., Blackwood, S., Gandhi, R., and de Lange, T. (2010). In vivo stoichiometry of shelterin components. J Biol Chem *285*, 1457-1467.

Takai, K.K., Kibe, T., Donigian, J.R., Frescas, D., and de Lange, T. (2011). Telomere protection by TPP1/POT1 requires tethering to TIN2. Mol Cell *44*, 647-659.

Takai, K.K., Kibe, T., Donigian, J.R., Frescas, D., and de Lange, T. (2017). Telomere Protection by TPP1/POT1 Requires Tethering to TIN2. Mol Cell *67*, 162.

Tejera, A.M., Stagno d'Alcontres, M., Thanasoula, M., Marion, R.M., Martinez, P., Liao, C., Flores, J.M., Tarsounas, M., and Blasco, M.A. (2010). TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. Dev Cell *18*, 775-789.

Theimer, C.A., Blois, C.A., and Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol Cell *17*, 671-682.

Theimer, C.A., and Feigon, J. (2006). Structure and function of telomerase RNA. Curr Opin Struct Biol *16*, 307-318.

Tummala, H., Collopy, L.C., Walne, A.J., Ellison, A., Cardoso, S., Aksu, T., Yarali, N., Aslan, D., Fikret Akata, R., Teo, J.*, et al.* (2018). Homozygous OB-fold variants in

telomere protein TPP1 are associated with dyskeratosis congenita-like phenotypes. Blood *132*, 1349-1353.

Tycowski, K.T., Shu, M.D., Kukoyi, A., and Steitz, J.A. (2009). A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Mol Cell *34*, 47-57.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nature reviews Microbiology *12*, 479-492.

van Steensel, B., and de Lange, T. (1997). Control of telomere length by the human telomeric protein TRF1. Nature *385*, 740-743.

Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P., and Artandi, S.E. (2009). A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. Science *323*, 644-648.

Venteicher, A.S., and Artandi, S.E. (2009). TCAB1: driving telomerase to Cajal bodies. Cell Cycle *8*, 1329-1331.

Walne, A.J., Vulliamy, T., Beswick, R., Kirwan, M., and Dokal, I. (2008). TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. Blood *112*, 3594-3600.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R., and Lei, M. (2007). The POT1-TPP1 telomere complex is a telomerase processivity factor. Nature *445*, 506-510.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y.*, et al.* (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res *22*, 1798-1812.

Wang, J., Zhuang, J., Iyer, S., Lin, X.Y., Greven, M.C., Kim, B.H., Moore, J., Pierce, B.G., Dong, X., Virgil, D.*, et al.* (2013). Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res *41*, D171-176.

Wang, Y., and Feigon, J. (2017). Structural biology of telomerase and its interaction at telomeres. Curr Opin Struct Biol *47*, 77-87.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L.*, et al.* (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res *46*, W296-W303.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136*, 615-628.

Weidenfeld, I., Gossen, M., Low, R., Kentner, D., Berger, S., Gorlich, D., Bartsch, D., Bujard, H., and Schonig, K. (2009). Inducible expression of coding and inhibitory RNAs from retargetable genomic loci. Nucleic Acids Res *37*, e50.

Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29-44.

Wright, W.E., Pereira-Smith, O.M., and Shay, J.W. (1989). Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. Mol Cell Biol *9*, 3088-3092.

Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W., and Shay, J.W. (1996). Telomerase activity in human germline and embryonic tissues and cells. Developmental genetics *18*, 173-179.

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

Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M.S., and Songyang, Z. (2007). TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. Nature *445*, 559-562.

Yang, D., He, Q., Kim, H., Ma, W., and Songyang, Z. (2011). TIN2 protein dyskeratosis congenita missense mutants are defective in association with telomerase. J Biol Chem *286*, 23022-23030.

Ye, J.Z., Donigian, J.R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A.N., Chait, B.T., and de Lange, T. (2004a). TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. J Biol Chem *279*, 47264-47271.

Ye, J.Z., Hockemeyer, D., Krutchinsky, A.N., Loayza, D., Hooper, S.M., Chait, B.T., and de Lange, T. (2004b). POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. Genes Dev *18*, 1649-1654.

Zalenskaya, I.A., Bradbury, E.M., and Zalensky, A.O. (2000). Chromatin structure of telomere domain in human sperm. Biochem Biophys Res Commun *279*, 213-218.

Zaug, A.J., Podell, E.R., Nandakumar, J., and Cech, T.R. (2010). Functional interaction between telomere protein TPP1 and telomerase. Genes Dev *24*, 613-622.

Zhang, Y., Chen, L.Y., Han, X., Xie, W., Kim, H., Yang, D., Liu, D., and Songyang, Z. (2013). Phosphorylation of TPP1 regulates cell cycle-dependent telomerase recruitment. Proc Natl Acad Sci U S A *110*, 5457-5462.

Zhong, F.L., Batista, L.F., Freund, A., Pech, M.F., Venteicher, A.S., and Artandi, S.E. (2012). TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. Cell *150*, 481-494.

Zhu, X.D., Küster, B., Mann, M., Petrini, J.H., and de Lange, T. (2000). Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. Nat Genet *25*, 347-352.

Zhu, Y., Tomlinson, R.L., Lukowiak, A.A., Terns, R.M., and Terns, M.P. (2004). Telomerase RNA accumulates in Cajal bodies in human cancer cells. Mol Biol Cell *15*, 81-90.

Zhuang, J., Jiang, G., Willers, H., and Xia, F. (2009). Exonuclease function of human Mre11 promotes deletional nonhomologous end joining. J Biol Chem *284*, 30565-30573.

Zimmermann, M., Kibe, T., Kabir, S., and de Lange, T. (2014). TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR signaling. Genes Dev *28*, 2477-2491.