

IDENTIFYING, CHARACTERIZING AND INHIBITING THE TELOMERASE  
REGULATORY NETWORK

APPROVED BY SUPERVISORY COMMITTEE

Jerry Shay, PhD

---

Woodring Wright, MD, PhD

---

David Corey, PhD

---

Rolf Brekken, PhD

---

Melanie Cobb, PhD

---

---

## DEDICATION

Dedicated to my parents, Mark and Tracy Holohan, as well as my siblings, Kelly Nudleman and Kyle Holohan for their support and encouragement.

IDENTIFYING, CHARACTERIZING AND INHIBITING THE TELOMERASE  
REGULATORY NETWORK

by

BRODY CHRISTOPHER HOLOHAN

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2015

Copyright

by

Brody Christopher Holohan, 2015

All Rights Reserved

## ACKNOWLEDGEMENTS

I would like to thank my mentors, Jerry Shay and Woodring Wright for their continuous support, advice and encouragement, as well as allowing me time and resources to pursue my ideas. They are a fantastic team, and I could not have asked for better mentors.

I would also like to thank my lab-mates, particularly Dr. Andrew Ludlow, Dr. Yong Zhou, Dr. Tsung-Po Lai, Dr. Oliver Delgado, Dr. Wanil Kim, Dr. Guido Stadler, Kimberly Batten, Ryan Laranger and Neal Jones. I would also like to thank Kevin Kennon for administrative support.

Without my collaborators much of this would have been impossible, so I would like to thank Dr. Tim De Meyer, Dr. Abraham Aviv, Dr. Fowzan Aklayura, Dr. Steve Hunt, Dr. Tim Spector, Dr. Massimo Mangino, Dr. Daphne Friedman and Dr. Daniel Eisenberg for their help, intellectual discussions and willingness to pursue challenging ideas.

I thank the University of Texas Southwestern Medical Center and the Department of Cell Biology for the opportunity to earn a PhD, as well as the Cancer Prevention and Research Institute of Texas and National Institute of Health for funding.

Further, I thank my fellow students that made these years enjoyable and acted as my support network in tough times.

IDENTIFYING, CHARACTERIZING AND INHIBITING THE TELOMERASE  
REGULATORY NETWORK

Brody Christopher Holohan, PhD

The University of Texas Southwestern Medical Center at Dallas, 2015

Supervising Professor: Jerry W. Shay, PhD

Telomeres, which are structures that cap the ends of linear chromosomes are maintained by telomerase, a reverse transcriptase. Telomere length limits the self-renewal capacity for telomerase negative cells, and nearly all tumors circumvent this limitation through telomerase expression; as such, telomerase is an attractive target for cancer therapy. In order to identify new targets for anti-telomerase therapy, I demonstrate that a number of candidate genes are required for telomere maintenance *in vitro* through shRNA-mediated knockdown and telomere length analysis. Further, I show that Perifosine, a drug identified upstream of a number of the candidates can act as a telomerase inhibitor in a majority of cell lines evaluated *in vitro* as well as induce shortening of the shortest telomeres in tumors from

human patients treated with Perifosine in a phase II clinical trial. Additionally, I identify a trans-generational trend in telomere length at birth in human populations that may bias estimates of telomere shortening rate that has public health implications. Lastly, using data from a large twin study, I have identified a network of genes that regulate the rate of telomere shortening in humans that may be used to clarify the association between telomere length, aging and age-related disease.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>PRIOR PUBLICATIONS</b> .....	<b>xii</b>
<b>LIST OF FIGURES</b> .....	<b>xiii</b>
<b>LIST OF TABLES</b> .....	<b>xvii</b>
<b>LIST OF APPENDICES</b> .....	<b>xix</b>
<b>LIST OF DEFINITIONS</b> .....	<b>xx</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
Historical overview of telomere biology .....	1
Modern overview of telomeres and telomerase .....	5
Molecular Biology of the telomere .....	7
Molecular Biology of telomerase .....	9
Heritability of telomere length .....	12
Telomeres and cancer .....	13
Telomeropathies .....	15
Summary and hypothesis .....	21
<b>CHAPTER 2: VALIDATION OF AN IN VITRO SCREEN FOR POSITIVE</b>	
<b>REGULATORS OF TELOMERASE</b> .....	<b>22</b>
Introduction .....	22
Knockdown of the hits chosen for validation leads to telomere shortening over time ..	26

Reduction in telomerase activity, telomere shortening and altered telomerase mRNA splicing upon knockdown of larp7 .....	30
Humans with a loss-of-function mutation in larp7 suffer a classical telomeropathy ....	34
Discussion .....	37
Methods .....	46

**CHAPTER 3: CHARACTERIZATION OF PERIFOSINE AS A POSSIBLE**

<b>TELOMERASE INHIBITOR .....</b>	<b>51</b>
Introduction .....	51
AKT inhibitors induce telomere shortening in some cell lines in vivo .....	52
Long-term treatment with perifosine inhibits colony formation in soft agar .....	56
The HCC38 breast cancer cell line responds most robustly to perifosine of the cell lines tested .....	58
Perifosine may act as a telomerase inhibitor in a xenograft model, but does not reduce metastatic tumor burden .....	60
Perifosine induces shortening of the shortest telomeres and reduces telomerase enzymatic activity in CLL samples from a phase II clinical trial .....	63
Perifosine’s effects on cell growth, telomerase activity and telomere length are uncorrelated with one another .....	68
Expression data indicates stress response networks may alter Perifosine sensitivity ....	72
Perifosine activates p38 activity in some but not all of the sensitive cell lines .....	81
Perifosine treatment alters TRF2 expression in some cell lines .....	83
Discussion .....	85

Methods .....	95
<b>CHAPTER 4: IDENTIFICATION OF A TIME-DEPENDENT TREND IN TELOMERE LENGTH .....</b>	<b>100</b>
Introduction .....	100
Results: Cross-sectional estimates of telomere shortening rate are consistently lower than longitudinal estimates .....	108
PBY predicts telomere length better than FAB and models of cross-sectional data which do not assume fixed initial telomere length are consistent with longitudinal studies .....	109
Collinearity between Age, FAB and PBY .....	113
Mediation analysis reveals heterogeneity in the PBY effect and a larger FAB effect than previously reported .....	115
The age-associated telomere shortening rate from mediation analysis is more consistent with the pseudo-longitudinal shortening rate than cross-sectional estimates .....	119
A secular trend in initial telomere length exists after accounting for the effects of age, FAB and PBY .....	122
Discussion .....	124
Methods .....	132
<b>CHAPTER 5: IDENTIFICATION OF THE PHYSIOLOGICAL RATE OF TELOMERE SHORTENING (ROTS) NETWORK VIA TWIN VARIANCE ANALYSIS .....</b>	<b>136</b>
Introduction .....	136
Monozygotic but not dizygotic twins diverge in telomere length over time .....	137

The telomeric response to two different environmental stimuli depends on gene- environment interactions .....	142
Surveying the genome for MZ-like behavior in DZ twins can reveal the sites that control telomere dynamics .....	145
The environmental response Rate of Telomere Shortening locus near LARP7 may alter the response to smoking .....	153
The intrinsic Rate of Telomere Shortening loci can predict telomere length .....	158
Twin Variance Analysis on permuted phenotypes yields information about the false discovery rate .....	162
The telomere length predictions from the RoTS loci replicate in another population .	167
Discussion .....	168
Methods .....	190
<b>CONCLUSION .....</b>	<b>191</b>
<b>APPENDIX A: TRF DATA .....</b>	<b>192</b>
<b>APPENDIX B: CUSTOM SOFTWARE .....</b>	<b>195</b>
<b>BIBLIOGRAPHY .....</b>	<b>215</b>

## PRIOR PUBLICATIONS

Holohan B, Wright WE, Shay JW (2014). Cell biology of disease: Telomeropathies: an emerging spectrum disorder. *The Journal of Cell Biology*. **205**, 289-299.

Holohan B, De Meyer T, Batten K, Mangino M, Hunt SC, Bekaert S, De Buyere ML, Rietzschel ER, Spector TD, Wright WE, Shay JW (2015). Decreasing initial telomere length in humans intergenerationally understates age-associated telomere shortening. *Aging Cell*.  
(In press)

Holohan B, Hagiopian MM, Lai TP, Friedman DR, Wright WE, Shay JW (2015). Perifosine as a potential novel anti-telomerase therapy. (In preparation)

Holohan B, Mangino MM, Spector TD, Wright WE, Shay JW (2015). Twin variance analysis identifies a network that regulates the rate of telomere shortening in humans. (In preparation)

## LIST OF FIGURES

Figure 2.1: Screening method to identify positive regulators of telomerase .....	23
Figure 2.2: Representative TRF images .....	29
Figure 2.3: Knockdown quantitation of LARP7 .....	31
Figure 2.4: Reduction in telomerase activity and telomere shortening in LARP7 knockdown cells .....	32
Figure 2.5: LARP7 knockdown decreases full-length and –beta TERT splicing .....	33
Figure 2.6: LARP7 mutants and their descendants have very short telomeres and generationally decreasing initial telomere length .....	35
Figure 2.7: Pedigree of the LARP7 mutant family .....	36
Figure 2.8: The telomere maintenance network .....	38
Figure 3.1: AKT inhibitor IV induces dose-dependent telomere shortening .....	53
Figure 3.2: AKT inhibitor IV inhibited cell growth .....	54
Figure 3.3: Perifosine causes telomere shortening in telomerase positive cells .....	55
Figure 3.4: Long term treatment with perifosine reduces soft agar colony formation .....	57
Figure 3.5: Perifosine induces telomere shortening in HCC38 and MDA-MB-435 cells ...	59
Figure 3.6: Perifosine treatment did not alter metastatic tumor burden or telomere length in a xenograft .....	61
Figure 3.7: Primary tumors exhibit much longer telomere length than subclones of the parental HCC38 population .....	62
Figure 3.8: Mouse DNA produces substantially biased measurements of telomere length in human samples .....	63

Figure 3.9: Reduction in ddTRAP activity in Perifosine treated CLL without a change in mean TL .....	65
Figure 3.10: Universal STELA reveals shortening of the shortest telomeres .....	67
Figure 3.11: Telomere length in cells treated with Perifosine .....	68
Figure 3.12: Telomerase activity in Perifosine-treated cell lines .....	70
Figure 3.13: Gene sets enriched in cell lines insensitive to Perifosine-induced telomere shortening .....	73
Figure 3.14: Gene sets enriched in cell lines sensitive to Perifosine-induced telomere shortening .....	74
Figure 3.15: Expression of the genes most enriched in the locomotory behavior category in the static lines .....	75
Figure 3.16: Expression of the genes most enriched in DNA repair pathways in the shortening lines .....	76
Figure 3.17: Expression of the genes most enriched in the chromosome organization network in shortening lines .....	77
Figure 3.18: The 50 genes most enriched in both groups of cell lines .....	78
Figure 3.19: Expression of the candidate positive regulators of telomerase in Perifosine-sensitive and resistant cells .....	80
Figure 3.20: p38 activity in three chronically Perifosine-treated cell lines .....	81
Figure 3.21: p38 activity in three additional chronically Perifosine-treated cell lines .....	82
Figure 3.22: TRF2 levels in Perifosine-treated cells .....	84
Figure 4.1: Schematic of how a secular trend could bias cross-sectional measurements ..	101

Figure 4.2: Multicollinearity in the datasets investigated between age and PBY, PBY and age and FAB .....	114
Figure 4.3: Graphical representation of the PBY effect .....	118
Figure 4.4: The pseudo-longitudinal telomere shortening assay in the UK Twins cohort	121
Figure 4.5: The secular trend in initial telomere length .....	123
Figure 4.6: Mediation analysis .....	135
Figure 5.1: Inter-twin variance in telomere length vs. age .....	138
Figure 5.2: Demonstration how GxEs could generate the difference between MZ and DZ twins .....	141
Figure 5.3: Gene-environment interactions control the response to smoking .....	142
Figure 5.4: Gene-environment interactions control the response to caffeine intake .....	143
Figure 5.5: Gene-environment interaction control the response to caffeine total exposure	144
Figure 5.6: The premise of Twin Variance Analysis .....	146
Figure 5.7: The procedure used to find loci that modulate the rate of telomere shortening	147
Figure 5.8: Flow chart of the procedure used .....	148
Figure 5.9: Example output from the Twin Variance Analysis .....	149
Figure 5.10: Bootstrapping analysis of the bias introduced by small numbers of discordant twins .....	151
Figure 5.11: $-\log_{10}$ -transformed rateP values for the 10,000 most significant SNPs before filtering .....	152
Figure 5.12: Manhattan plot of the results of the twin variance analysis .....	153
Figure 5.13: TRFITV vs age by discordance at a RoTS locus near LARP7 .....	154

Figure 5.14: Telomere dynamics by genotype at rs10029516 and tobacco exposure .....	155
Figure 5.15: TRFITV by discordance at rs10029516 and cigarette exposure .....	156
Figure 5.16: TRFITV by discordance at rs10029516 and cigarette exposure excluding TT individuals .....	157
Figure 5.17: Three example intrinsic RoTS loci .....	158
Figure 5.18: Individuals with more risk genotypes exhibit faster telomere shortening .....	159
Figure 5.19: Telomere length predicted by the RoTS loci compared with observed telomere length in the UK Twins cohort .....	161
Figure 5.20: Manhattan plots of the results of the twin variance analysis on permuted phenotypes .....	163
Figure 5.21: qq plots of the results of the rateP values from twin variance analysis compared to the permuted controls .....	164
Figure 5.22: qq plots of specific subsets of the results of the twin variance analysis .....	166
Figure 5.23: Predicted telomere length vs. observed telomere length in the Family Heart Study .....	167
Figure 5.24: Intrinsic RoTS loci over MDM4 and FMN2 .....	171
Figure 5.25: Environmental RoTS loci near FMN1, FLNB and AhRR .....	174
Figure 5.26: Mechanistic model linking initial length and telomere shortening rate .....	175
Figure 5.27: RoTS loci near MIR4283-1 and ABR .....	180
Figure 5.28: Intrinsic RoTS locus near RYR2 .....	181
Figure 5.29: A number of RoTS loci near genes identified in Chapter 2 .....	182

Figure 5.30: Overlaps between loci identified with age as a proxy for the environment and loci identified with specific environmental information ..... 185

Figure 5.31: Using the same method to find GxEs for BMI and Caffeine identifies the HLA locus and RYR3, a caffeine receptor ..... 186

Figure 5.32: The CSMD1 locus correlates with both exposure and response to caffeine . 188

## LIST OF TABLES

Table 2.1: The genes selected for validation via shRNA knockdown .....	26
Table 2.2: Summary of the results of knockdown, long-term tissue culture and telomere measurements .....	28
Table 3.1: Summary of responses to Perifosine .....	69
Table 3.2: Growth rate, telomere dynamics and changes in telomerase activity in the lung cancer cell lines evaluated .....	71
Table 4.1: Summary of age-associated telomere shortening rates in the literature .....	109
Table 4.2: Contrasting models of Age and PBY vs Age and FAB in all three datasets ....	111
Table 4.3: Multicollinearity between the variables investigated in the combined datasets	113
Table 4.4: Mediation-adjusted models .....	115
Table 4.5: Gender-segregated mediation-adjusted models .....	117
Table 5.1: Analysis of the relationship between telomere length and telomere shortening rate in the literature .....	169

LIST OF APPENDICES

APPENDIX A: TRF DATA ..... 192

APPENDIX B: CUSTOM SOFTWARE..... 195

## LIST OF DEFINITIONS

DNA – Deoxyribonucleic acid

RNA – Ribonucleic acid

TRAP – Telomere Repeat Amplification Protocol

hTR – Human telomerase RNA component

hTERT – Human telomerase reverse transcriptase

TERT – Telomerase reverse transcriptase

ALT – Alternative lengthening of telomeres

TRF1/2 – Telomere repeat binding factor 1/2

TIN2 – TRF1 interacting nuclear protein 2

Rap1 – Repressor/activator protein 1

TPP1 – TIN2 interacting protein 1

POT1 – Protection of telomeres 1

RTEL1 – Regulator of telomere elongation helicase 1

CTC1 – CST telomere maintenance complex component 1

STN1 – Oligonucleotide/oligosaccharide-binding fold containing 1 (also known as OBFC1)

TEN1 – TEN1 CST complex subunit

TERRA – Telomere repeat containing RNA

IPF – Idiopathic Pulmonary Fibrosis

DKC – Dyskeratosis Congenita

HHS – Hoyeraal Hriedarsson syndrome

shRNA – Short Hairpin RNA

TRF – Terminal Restriction Fragment (a measurement of telomere length)

LARP7 – La Autoantigen Related Protein 7

TOP – 5'-terminal oligopurine

PCR – Polymerase Chain Reaction

GFP – Green Fluorescent Protein

qPCR – Quantitative Polymerase Chain Reaction

LUC – Luciferase

Universal STELA – Universal Single Telomere Length Analysis

GSEA – Gene Set Enrichment Analysis

LTL – Lymphocyte Telomere Length

TL – Telomere Length

FAB – Father's Age at the Birth of his offspring

PBY – Paternal Birth Year

NHLBI – National Heart, Lung and Blood Institute

FHS – Family Heart Study

MAB – Mother's Age at the Birth of her offspring

MBY – Maternal Birth Year

CE – Common Era

GxE – Gene-Environment Interaction

MZ – Monozygotic (Identical)

DZ – Dizygotic (fraternal)

ITV – Inter-Twin Variance

TRFITV – Inter-Twin Variance in telomere length (TRF)

ANCOVA – Analysis of Covariance

SNP – Single Nucleotide Polymorphism

RateP – p-value used to test if the trend in concordant and discordant twins is the same

rs(a number) – Refseq IDs for SNPs

RoTS – Rate of Telomere Shortening

GWAS – Genome-Wide Association Study

qq plot – Quantile-Quantile plot

PC – Permuted Control

HSC – Hematopoietic Stem Cell

MPP – Multipotent Progenitor Cells

miRNA – MicroRNA

# CHAPTER ONE

## INTRODUCTION

### *Historical overview of telomere biology*

Telomeres, the ends of linear chromosomes, were first identified in their modern form in 1978 (Blackburn & Gall 1978), but were predicted to exist as early as 1938 by Herman Muller. Initially referred to as “natural ends” (McClintock 1941) to differentiate the ends of the chromosomes that exist in undamaged cells from the many free ends that result from breakage of DNA, they were later renamed “Telomeres”, from the Greek roots telos (end) and meres (part). The discovery of telomeres and their function would eventually reconcile a number of longstanding problems within molecular and cellular biology and provide critical insight into the two processes at the evolutionary heart of every organism; reproduction and death.

Leonard Hayflick and Paul Moorhead made the original observations that cells had an internal, cell-autonomous limit on the number of times they could divide in tissue culture (the “Hayflick limit”) (Hayflick & Moorhead 1961). These investigators overcame a long-held belief that all cells in culture are inherently immortal and showed that normal human cells exhibit a limit to their ability to divide in culture and that they stopped proliferating after a very predictable number of cell doubling. When cells reached the Hayflick limit (so named by the Nobel laureate, Sir Macfarlane Burnett), they entered a permanent growth arrest state known as replicative senescence wherein they stopped dividing. Following the discovery of DNA’s structure and its identification as the physical mechanism for heredity, early work on DNA replication indicated that the ends of linear DNA molecules presented unique problems

(Watson & Crick 1953). The 5'-most region of the lagging strand cannot be replicated during DNA synthesis because of the directional nature of DNA replication, resulting in a loss of DNA in any given replication event of at least the length of the final Okazaki fragment, a phenomenon known as the “end-replication problem”. Both James Watson and Alexey Olovnikov independently predicted the end-replication problem would present a block to cell division when the ends of chromosomes became critically short (Watson 1972; Olovnikov 1996), forming the basis for the modern understanding of replicative senescence. Elizabeth Blackburn, who would later win the Nobel Prize in 2009 for her work on telomere and telomerase biology with Carol Greider and Jack Szostak, characterized the telomere sequence of the model organism *Tetrahymena thermophila*, as repetitive DNA of the sequence TTGGGG (which differs from the telomere sequence of most multicellular organisms) (Blackburn & Gall 1978). She presented the idea that the telomere repeats represented a biological solution to the end-replication problem; the DNA lost from the ends of the chromosomes with cell division would be lost from this repeat array, and *Tetrahymena* must have some mechanism for restoring this repeat array between cell divisions. Blackburn and her student, Carol Greider, would later identify the ribonucleoprotein enzyme activity (a complex including protein and RNA components) that adds new repeats onto the telomeres in *Tetrahymena* (Greider & Blackburn 1987), eventually named telomerase, when it was cloned many years later. As predicted, telomerase compensated for the loss of telomeric repeats from the end-replication problem through the de-novo synthesis of new telomere repeats templated from the RNA component of the enzyme rather than the existing DNA, circumventing the requirement of an RNA primer for DNA synthesis. Jack Szostak (who shared the Nobel with Elizabeth Blackburn and Carol Greider) and Victoria Lundblad showed that in yeast, addition of telomeres from *Tetrahymena* to a linearized yeast plasmid

allowed the linear DNA to persist, indicating that the function of telomeres was conserved between too very distantly related species (Szostak & Blackburn 1982). They also later identified the telomerase components of yeast as well as a large number of accessory factors required for telomere maintenance, termed “ever-shorter telomeres” (EST) genes (Lundblad & Szostak 1989).

The relevance of telomere biology to human health started to take shape in the 1980's as various researchers determined that the mammalian and human sequence was TTAGGG and that different human organs have different telomere lengths (Cooke & Smith 1986). The first isolation of human telomerase enzymatic activity occurred in 1989 when Greg Morin identified telomerase activity in HeLa (a cervical cancer cell line) cell lysates (Morin 1989). Calvin Harley and Carol Greider then showed that normal somatic cells in humans do not express telomerase, and telomeres shorten over time in humans (Harley et al. 1990). In addition, the Robin Allshire lab showed that human reproductive tissues and fetal tissues had fairly long telomeres but that adult somatic tissue had significantly shorter telomeres that became even shorter with increased age (Hastie et al. 1990). The sum of this initial work presented the idea that telomere shortening via the end-replication problem, coupled with the absence of telomerase activity in most somatic cells could explain the loss of replicative potential over time in tissue culture and in aged humans; this generated a great deal of excitement in both academia and the lay media in spite of the fact that at this point there was only a correlation between telomere shortening and replicative senescence.

Catalyzed by the introduction of the Telomere Repeat Amplification Protocol (TRAP), a sensitive assay for telomerase enzymatic activity, in 1994 (Kim et al. 1994), identification of the specific components of telomerase in humans rapidly progressed. The RNA component, hTR, was identified in 1995 (Feng et al. 1995), followed by the protein

component, hTERT, in 1997 (Nakayama et al. 1997). This initial work also showed that in human cells the telomerase protein component was limiting for activity, and that introduction of TERT alone was sufficient to induce telomerase activity in those cell lines tested (Weinrich et al. 1997). Jerry Shay and Woodring Wright concurrently demonstrated through cell fusion experiments that telomere length limited cell division in hybrid cells (Wright et al. 1996a). Definitive proof that telomere length limited division in cell culture came in 1998 with the experiment that in cell lines identical in all respects other than the introduction of TERT, telomerase activity was sufficient to allow cells to overcome the Hayflick limit and continue to grow after cells that did not express telomerase had arrested their growth (Bodnar et al. 1998).

The early work on telomeres and telomerase triggered an enormous spike in interest in telomere biology, marked by the award of the Lasker award to Blackburn, Greider and Szostack in 2006 and the Nobel Prize in 2009. Between the discovery of telomerase enzyme activity (originally called terminal telomere transferase) and 1999, there were less than 1000 cumulative papers using the term “telomerase”, whereas there were over 6,000 by 2005, and at the time of this writing in 2015 there are over 13,000. This interest led to a number of discoveries important to human health. Telomere length has been associated with a host of complex diseases from cardiovascular disease and diabetes to degenerative neurological syndromes such as Alzheimer’s and Parkinson’s disease (Farzaneh-Far et al. 2010; Kume et al. 2012; Jiang et al. 2013; Raschenberger et al. 2013; Albrecht et al. 2014; Huzen et al. 2014; Rode et al. 2014; Zhao et al. 2014), reviewed in (Bojesen 2013). Mutations in core telomerase components also cause a spectrum of genetic diseases known by a number of names (Telomeropathies, Dyskeratosis Congenita, Hoyeraal-Hriedarsson syndrome, Revesz syndrome). These genetic diseases are collectively characterized by

impaired telomere maintenance, either through reduced ability to protect and recognize telomeres or in compromised activity in telomerase itself (reviewed in (Holohan et al. 2014)). It is becoming increasingly evident that telomeres act as an integrated biomarker of organismal stress, and a number of emerging discoveries indicate telomere length and telomerase may play a causal role in these pathologies (Bhayadia et al. 2015).

#### *Modern overview of telomeres and telomerase*

Telomeres are structures on the termini of chromosomes composed of hexameric TTAGGG repeats and a protein complex called the shelterin complex that protect the ends of linear chromosomes from recognition as double-stranded breaks (de Lange 2010).

Telomeres shorten with every cell division due to the fact that DNA synthesis is unable to replace the RNA primer of the most 5' region of the lagging strand during S-phase, a phenomenon known as the "end-replication problem" (Levy et al. 1992). In addition to the end-replication problem, telomeres shorten as a result of processing events following DNA replication, and they are exquisitely sensitive to a number of DNA-damaging insults from reactive oxygen species to thymine dimers and replication errors, resulting in a telomere shortening rate in telomerase negative cells of roughly 50-100 base pairs per cell division (Wu et al. 2012) in tissue culture, and 20-60 base pairs per year in peripheral blood mononuclear cells in adulthood (Daniali et al. 2013). This progressive telomere shortening imposes a limit on the number of cell divisions any given cell can undergo before undergoing senescence generally thought to be a result of a p16-driven DNA damage signal (Wright et al. 1989), or a p53/p21-driven unreparable DNA damage signal if the p16 pathway is bypassed (Chin et al. 1999; Takai et al. 2003).

This progressive telomere shortening eventually results in an unreparable DNA damage signal that induces cells to enter senescence, the M1 stage of crisis, also known as the

Hayflick limit. Precancerous cells may avoid M1 by abrogating p16/INK4A signaling; these cells may continue to divide with concomitant telomere shortening until their telomeres are so short that they trigger progressive chromosomal breakage/fusion/breakage events that almost invariably lead to cell death via the p53/p21 pathway, termed crisis or stage M2 (Chin et al. 1999; Takai et al. 2003; Shay & Wright 2011). This extended life period between the Hayflick limit and crisis was initially observed by virologists who were expressing viral genes (such as SV40 large T-antigen) in normal cells. A very small number of cells that enter crisis re-establish telomere maintenance and escape crisis by expressing telomerase or the ALT (alternative lengthening of telomeres) pathway. The tight regulation of telomerase activity is believed to present one of the most important checks to cancer incidence (Shay & Wright 2011).

Germline cells, embryonic stem cells and a subset of somatic stem cells partially counteract this progressive telomere shortening via telomerase, a ribonucleoprotein reverse transcriptase that can maintain or extend telomere length during cell division. Telomerase adds new telomere repeats to the 3' overhang of telomeres during S phase of the cell cycle in order to maintain telomere integrity (Wright et al. 1996b). Adult stem cells generally undergo telomere shortening in spite of telomerase expression, though this telomere shortening rate is substantially lower than in telomerase negative cells.

Because telomere shortening reduces the number of times a daughter cell can divide, progressive telomere shortening in somatic stem cells may represent selective optimization of a tradeoff between regenerative capacity of a tissue and cancer risk within that tissue. Telomere shortening leading to senescence may diminish the cancer risk for a given daughter cell, even though that also reduces the ability of that daughter cell to divide and continue to function nominally (Pereira & Ferreira 2013). Thus, as an organism ages and

the cancer risk per cell division increases because of increased numbers of mutations per cell, telomere length is progressively shortened in order to impose an initial blockage or check on the growth of cancer incidence. While telomerase itself does not act as an oncogene, the absence of telomerase activity acts as a tumor suppressor.

#### *Molecular biology of the telomere*

Telomeres are composed of the TTAGGG repeats as well as the protein factors that are required for adequate protection of the ends of the chromosome from recognition as a DNA break. The core protein group that executes this function is called the Shelterin complex, which is composed in mammals of Telomere repeat binding factor 1 and 2 (TRF1, TRF2), TRF1-interacting nuclear protein 2 (TIN2), repressor/activator protein 1 (Rap1), TIN2-interacting protein 1 (TPP1) and Protection of Telomeres 1 (POT1), reviewed in (de Lange 2010). TRF1 and TRF2 can both independently associate with telomere DNA, and they nucleate the recruitment of the rest of the Shelterin complex via the TRF-homology domains present on the proteins, which bind to conserved F/YxLxP motifs present on TIN2 and Rap1, as well as a number of other proteins (Chen et al. 2008), though both TRF1 and TRF2 can associate with telomeric repeats without the presence of the rest of the Shelterin complex (Lin et al. 2013). After it is recruited by either TRF1 or TRF2, TIN2 binds to TPP1 and TPP1 recruits POT1 to the complex (Houghtaling et al. 2004; Xin et al. 2007).

The combined function of the Shelterin complex proteins protect the telomeres from pathological DNA repair activities, and loss of any Shelterin component can lead to genome instability, chromosomal deletion/duplication events (Sfeir & de Lange 2012) and cell death. The critical function of the Shelterin complex is most clearly demonstrated in the embryonic or perinatal lethal phenotypes observed in mice with mutations in Shelterin complex components (Tejera et al. 2010; Beier et al. 2012). The Shelterin complex's ability to

prevent the telomere from triggering a DNA damage signal arises from a structure known as a T-loop, in which the 3' overhang of the single-stranded G-strand overhang invades the telomere duplex DNA intramolecularly to produce a structure similar to a Holiday recombination intermediate which does not resemble a double strand break because it has no exposed 3' end (Griffith et al. 1999; Doksanı et al. 2013). This structure must be dissociated during replication in order to copy the terminus of the telomere, and this is accomplished RTEL1, a DNA helicase. If this dissociation cannot occur, such as in the case of RTEL1 mutant cells, the T-loop can be completely excised by nucleases such as SLX4, which causes rapid shortening of the telomere via terminal deletion events (Vannier et al. 2012).

Because of the 5' to 3' directionality of DNA synthesis, the telomere that used the C-strand as a template (the leading strand) can be synthesized to a blunt-ended duplex at the 3' end. Formation of a T-loop requires a single-stranded overhang, and this is accomplished in the leading strand through resection of the C-strand by Apollo and Exo1 nucleases (Chen et al. 2008; Wu et al. 2012). Resection of the C-strand is followed by partial fill-in of the resected region, which is mediated by the CST complex (composed of CTC1, STN1 and TEN1). The CST complex is also involved in promoting the restart of stalled replication forks that result from the formation of G-quadruplexes during telomere DNA replication (Tang et al. 2008; Stewart et al. 2012; Wang et al. 2012).

The repetitive nature of telomeric sequences and their heterochromatin status in humans led to the initial idea that telomeres were not transcribed, however recently transcription of the telomere from sub-telomeric non-canonical promoters has been detected (Feuerhahn et al. 2010). These long non-protein coding telomeric transcripts, (TERRA) are RNA polymerase II transcripts which are sometimes 5' 7'methyl-G capped and polyadenylated; transcription

begins in the subtelomere and continues into telomeric sequences (Azzalin et al. 2007; Schoeftner & Blasco 2008). TERRA modulates the heterochromic state of telomeres by promoting methylation of telomeric histones. TERRA also regulates telomeric DNA damage signaling and homologous recombination-mediated terminal deletion of T-loops by competing with the telomeric overhang for TRF2 binding. TRF2 prevents action of the homologous recombination repair machinery on the T-loop and competitive inhibition of TRF2 by TERRA can allow resolution of the T-loop structure via excision (Porro et al. 2014). Furthermore, loss of TRF2 induces upregulation of TERRA transcription, which leads to further inhibition of TRF2 via competitive inhibition with TERRA transcripts, further deprotection of the telomere and eventual replicative senescence. However, other studies have shown that TERRA can play a protective role under certain circumstances (de Silanes et al. 2014), suggesting that there may be a stoichiometric interaction between TERRA, telomeric overhangs, TRF2 and other unidentified factors that yields telomere protection or de-protection depending on relative levels of the relevant component parts.

The function and diversity of telomeric transcripts is a very active area of research, and a number of recent studies have shown that TERRA levels alter transcription of a large number of genes, and that many of these alterations can be recapitulated with the addition of an oligonucleotide bearing G-quadruplex forming regions (Porro et al. 2014; Hirashima & Seimiya 2015). The ability of form G-quadruplexes is especially relevant to telomere biology, since G-quadruplexes are a feature of the c-myc oncogene's promoter (Simonsson et al. 1998; Siddiqui-Jain et al. 2002), and c-myc can activate telomerase (Flores et al. 2006); TERRA may down-regulate telomerase indirectly via modulation of c-myc transcription.

*Molecular biology of telomerase*

In humans, the core of the telomerase holoenzyme is composed of two components; telomerase reverse transcriptase (hTERT) protein subunit that contains the catalytic protein components and the telomerase RNA component (hTR) which templates the addition of new telomere repeats onto the terminus of the chromosome. In addition, hTR operates as a structural element within the larger ribonucleoprotein holoenzyme complex (Nakamura et al. 1997). In addition to hTR and hTERT, there are a number of accessory factors that are associated with telomerase, including NHP2, GAR1, dyskerin and NOP10 (Egan & Collins 2010). The telomerase complex is assembled in the Cajal body through unclear means, though it requires recruitment of telomerase components by TCAB1 (Stern et al. 2012). Several chaperone proteins are required for telomerase assembly and in vitro reconstitution, including p23, HSP70 and HSP90, indicating they may be involved in assembly of telomerase, though dyskerin and hTR spontaneously associate via an interaction between dyskerin and the H/ACA box region of hTR (Forsythe et al. 2001; Trahan & Dragon 2009). Coilin, a Cajal body protein with RNase activity processes the 3' end of hTR, and is believed to be required for telomerase assembly (Stern et al. 2012; Broome & Hebert 2013). Additionally, the chaperonin TRiC is required for proper folding of TCAB1, and loss of TRiC can lead to impaired telomere maintenance due to failure to assemble telomerase (Freund et al. 2014).

#### *Telomere dynamics in humans*

At birth in humans, telomere length in nucleated blood cells is in the range of 10-15 kilobases, and this initial telomere length undergoes progressive shortening throughout an individual's life span (Aviv et al. 2003). Telomere shortening is fastest early childhood and the pubertal growth spurt, before leveling off at a slower rate in adulthood; estimates of

telomere shortening rate vary by methodology and populations studied, but lie in the range of 20 to 60 base pairs per year.

Early work in telomere biology indicated that differing tissues have different telomere lengths, however more recent work has determined that telomere lengths in different organs strongly correlate and shorten at roughly the same rates (Daniali et al. 2013). These observations suffer from small sample sizes and infrequent replication because analysis of telomere length requires a sample of the organs under investigation, which is difficult and invasive to obtain for most organs other than blood or skin. Studies in primates and humans (Gardner et al. 2007) indicate that organs shorten telomeres asynchronously, though deterministic differences between tissues (skeletal muscle telomeres are longer than skin and blood) and strong correlations within an individual between tissues indicate shared initial telomere length and that the differences are the result of different turnover rates or telomerase activity within the tissues.

A host of environmental factors impact telomere length, including air pollution, tobacco exposure, stress, urban garbage, chronic infections and asthma (Farzaneh-Far et al. 2010; Kume et al. 2012; Jiang et al. 2013; Raschenberger et al. 2013; Albrecht et al. 2014; Huzen et al. 2014; Rode et al. 2014; Zhao et al. 2014; Tahara et al. 2015). It is not fully understood how these exposures exert their effects on telomere length, though it is probable that any insult that induces a tissue to increase the rate of cell division will trigger telomere shortening. In the case of *Helicobacter pylori*, a bacteria that infects the stomach lining and can cause ulcers and stomach cancer, infection triggers gastritis (chronic inflammation of the stomach lining), and the extent of telomere shortening in the stomach lining is proportional to the degree of gastritis (Tahara et al. 2015). The turnover stimulated by the chronic inflammation and cell death closely associate with the degree of telomere

shortening, indicating that it is likely the increased rate of cell division responsible for the shortening.

Increased cell turnover is not the only insult that can alter the rate of telomere shortening. Certain lifestyle factors such as exercise (Osthus et al. 2012) and caloric intake (Kark et al. 2012) can effect telomere length in humans. While it is possible that exercise may stimulate cell turnover, endurance athletes have longer telomeres than the general population, indicating that any telomere shortening driven by the acute stress of the exercise itself is compensated by a poorly-understood telomere elongation stimulus. While the mechanism of exercise-mediated telomere increases is not understood in humans, in murine model studies exercise associated telomere increases may be mediated by alteration in expression of the Shelterin complex, or by stimulation of telomerase itself (Werner et al. 2008; Ludlow et al. 2012a; Ludlow et al. 2012b). Regulation of telomere length can also be highly tissue specific, as detailed study of the effect of exercise on different muscle tissues showed that a single exercise regimen exerted different effects on three different tissues (Ludlow et al. 2012b). The heterogeneity that results from these tissue-specific responses may be one of the mechanisms driving the substantial inter-individual variations in blood telomere length, as the variation around the age-mean may be due in part to a large number of overlapping and competing stimuli.

#### *Heritability of telomere length*

One unique aspect of telomere biology is that telomere length is hereditary through two distinct mechanisms; parents pass on genetic variants in telomerase components and genes that regulate the activity of telomerase that can effect telomere length in offspring, but they also directly pass on the telomeres themselves as physical entities (De Meyer et al. 2014). This leads to a “dual inheritance” modality in telomere length, in that it is heritable

both genetically and epigenetically (Chiang et al. 2010). Because of this dual inheritance and the importance of telomere biology to regenerative capacity and cancer risk, it is likely that telomere length represents a trans-generational mode of biological communication similar to genetic imprinting through methylation. Telomeres may be another mechanism that allows parents to program the genome of their offspring for the local environment in order to improve their reproductive fitness.

#### *Telomeres and cancer*

The replicative limit imposed by telomere length, the discovery that spontaneously-resolving neuroblastomas lacked telomerase and this could explain their regression (Hiyama et al. 1995), coupled with the observation that most cancers express telomerase gave hope that a telomerase inhibitor could act as a universal cancer therapy. If all cancers required telomerase in order to maintain telomere length and their immortalized state, then a treatment that inhibited telomerase would be predicted to eliminate the ability of cancer cells to divide indefinitely. In practice, use of telomerase inhibitors as a cancer therapy is vastly more complicated than initially hoped, and a small number of tumors are capable of maintaining telomeres through a poorly-understood process known as Alternative Lengthening of Telomeres (ALT), which does not utilize telomerase. However, over 90% of tumors depend on telomerase to maintain telomeres (Royle et al. 2008), and there is only one documented instance of a telomerase-positive population of cells switching to ALT upon telomerase inhibition (Bechter et al. 2004), suggesting such a change is biologically nontrivial.

A number of telomerase inhibitors are in clinical development, most notably Imetelstat (Chiappori et al. 2015), a 13-mer nucleic acid drug that competitively inhibits the template region of the RNA subunit of telomerase. Imetelstat has encountered a number of obstacles

to its widespread adoption chiefly resulting from the chemistry of its modified nucleic acid backbone, though it blocks telomerase activity well in tissue culture (Joseph et al. 2010). Different cancer cell lines have vastly different mean telomere lengths as well as different sensitivities to Imetelstat (Brennan et al. 2010). Sensitivity to Imetelstat only weakly correlates with telomere length and telomerase activity, suggesting that Imetelstat is subject to many of the same obstacles to therapy such as bioavailability and degradation/export-mediated resistance modalities from which other cancer therapies suffer. Imetelstat has failed in a number of large clinical trials in treatment of solid tumors, and its proponents are currently in the process of repurposing it to treat myeloproliferative diseases based on the observation that its main side effect is thrombocytopenia (pathologically low platelet counts) and thus it is being used to treat patients with thrombocythemia (too many platelets). In summary, if anti-telomerase cancer therapy is to become reality, novel telomerase inhibition approaches are required.

Inhibiting telomerase activity in order to remove the immortality of cancer cells is arguably the most straightforward method to utilize telomere biology to treat cancer, but it is not the only possibility. Because telomerase is the only reverse transcriptase normally encoded in the human genome (barring infection with a retrovirus or expression of human endogenous retroviruses) and its RNA-dependent DNA polymerase is distinct from other DNA polymerases, it may be possible to utilize an idiosyncrasy of this domain to find a treatment that is synthetically lethal with telomerase in order to directly kill telomerase positive cancer cells, or to use physical features of the reverse transcriptase that are different from normal DNA polymerases to design nucleoside analogs that are preferentially used by telomerase. Such an approach has been used in the past to treat Human Immunodeficiency Virus (HIV) via antiretroviral therapy using nucleoside/nucleotide

analogues that have higher affinity for HIV reverse transcriptase compared to DNA polymerase (Furman et al. 1986; Herman et al. 2012). Indeed, comparatively high concentrations of 3'-azido-3'-deoxythymidine triphosphate (AZT, a thymidine analog used in antiretroviral therapy for HIV patients) can inhibit telomerase, demonstrating that selective inhibition of telomerase is possible through a method that does not competitively inhibit the active site (Brown et al. 2003). It is possible that new guanosine analog reverse transcriptase inhibitors for HIV treatment would exhibit higher anti-telomerase activity and lower toxicity than AZT because guanosine composes half of the substrate used by telomerase and lower toxicity because the human genome's G/C content is below 50%; a normal cell will use fewer G's per unit DNA replication than it will use T's, whereas telomerase uses 3 G's per telomere repeat compared with 2 T's.

Synthetic lethality with telomerase activity has also been demonstrated in the recent past through the use of 6'-thio deoxyguanine, a modified nucleotide drug in very early clinical development which can selectively kill telomerase positive cells while sparing normal cells (Mender et al. 2015). Telomerase incorporates 6'-thio dG into telomeres during cell division, which interferes with their ability to inhibit the DNA damage response, leading to cell death. dGTP competes with 6'-thio dG for utilization by telomerase, and so it may be possible to further improve the clinical utility of 6'-thio dG by co-administering an inhibitor dGTP synthesis, such as Azothioprine, which would logically increase the rate of 6'-thio dG incorporation.

### *Telomeropathies*

Interference with any of the processes required to maintain telomeres in humans causes a spectrum of diseases collectively referred to as the Telomeropathies, which are characterized by the common underlying pathological mechanism, impaired maintenance of

the telomeres. Symptoms of telomeropathies include bone marrow failure, idiopathic pulmonary fibrosis, the “diagnostic triad” (oral leukoplakia, skin hyperpigmentation and nail dystrophy) as well as a number of very rare, incompletely penetrant symptoms including cerebral calcifications and exudative retinopathy (reviewed in (Holohan et al. 2014)).

The most common genes associated with telomeropathies are telomerase subunits, such as hTERT, hTR, Dyskerin (DKC1), NOP10, NHP2 and TCAB1 because investigation of a family with a telomeropathy is typically driven by the appearance of the diagnostic triad followed by a candidate-gene driven direct sequencing approach. (Heiss et al. 1998; Vulliamy et al. 2001; Vulliamy et al. 2002; Armanios et al. 2005; Yamaguchi et al. 2005; Armanios et al. 2007; Tsakiri et al. 2007; Walne et al. 2007; Vulliamy et al. 2008; Calado et al. 2009a; Calado et al. 2009b; Zhong et al. 2011). It is unclear at the present time if the reason that some mutations are more commonly observed than other mutations is due to an underlying difference in mutation frequency or removal via selection, or if fewer groups investigating telomeropathies examine the genes less commonly associated with telomeropathies.

Aside from telomerase holoenzyme components, a number of other genes have been found to be mutated in patients with telomeropathies. Mutations in TIN2, a shelterin component, are often reported in severe telomeropathies (Savage et al. 2008; Sasa et al. 2012; Vulliamy et al. 2012). Reports have linked dysfunction of RTEL1, the helicase that dissociates the t-loop, with Hoyeraal-Hreidarsson syndrome (HHS) (Ballew et al. 2013; Le Guen et al. 2013; Walne et al. 2013b). Additionally, a dominant-negative splice form of Apollo, a nuclease implicated in post-replication processing of the telomere, has been reported in a HHS patient. Though the mutation did not cause loss of the Apollo protein, introduction of this splice form into fibroblasts induced telomere dysfunction and end-to-end

fusions, indicating that it was likely responsible for this patient's symptoms (Touzot et al. 2010). Defects in the CST complex, which also involved in telomeric end processing and replication have been associated with telomeropathies. Patients with CTC1 mutations have been identified in patients with Coats Plus Syndrome (Anderson et al. 2012; Keller et al. 2012; Polvi et al. 2012; Walne et al. 2013a).

The most common symptom of a telomeropathy in adults is idiopathic pulmonary fibrosis (IPF), a progressive fibrotic failure of the lung coincident with inflammation (Armanios et al. 2007; Armanios & Blackburn 2012). Between 8% and 20% of familial cases of IPF can be traced to mutations in hTR or hTERT, and a small number of sporadic cases of IPF are due to de novo mutations in hTERT or hTR (Tsakiri et al. 2007). Up to 25% of sporadic IPF cases and 37% of familial cases occur in individuals in the shortest 10th percentile of the population, indicating that many of these cases are telomeropathies due to factors other than mutations in genes known to influence telomere length (Cronkhite et al. 2008). These IPF families exhibit autosomal dominant inheritance, which is consistent with the haploinsufficiency of telomerase in humans; in these families, IPF first appears in mutation carriers in old to middle age, at a median age of incidence of 51 years, while bone marrow failure and the classic triad of Dyskeratosis Congenita symptoms occurs in younger members of the cohort (Parry et al. 2011).

In addition to IPF, other adult-onset symptoms of a telomeropathy include liver cirrhosis, aplastic anemia, an acute myelogenous leukemia (AML) (Fogarty et al. 2003; Calado et al. 2009a; Calado et al. 2009b). These adult-onset symptoms, though incompletely penetrant, can co-occur, such that patients with TERT mutations sometimes present with both IPF and liver cirrhosis. It is also possible that proliferative insults can drive organ failure in other susceptible tissues, for example patients that acquire liver failure after

treatment of their IPF with azathioprine (azathioprine is used in the treatment of IPF because IPF is presumed to be autoinflammatory, and azathioprine can suppress the immune response by interfering with cell division through purine synthesis inhibition) (Calado et al. 2009b). Susceptibility to cell turnover-inducing insults in previously healthy organs indicates that telomere shortening occurs in all organs in these patients, but the proliferative failures tend to occur in a single tissue prior to the others.

The first disorder linked to a defect in telomere maintenance was Dyskeratosis Congenita in the late 1990's (Mitchell et al. 1999). DKC is conventionally linked to the "classic triad" of symptoms, which include oral leukoplakia (thick, white lesions inside the mouth characterized by infiltrating immune cells), skin hyperpigmentation (dark patches on the skin) and nail dystrophy (reduced or absent finger and toenail growth, leading to very small or lost nails). DKC is associated with a number of symptoms less common than the classic triad, most importantly proliferative failure of the bone marrow leading to either generalized anemia or specific lymphopenias (Dokal 2011; Young 2012). Because of this bone marrow failure, DKC patients are often treated with bone marrow transplants. Ablation of the existing bone marrow compartment, required for a bone marrow transplant, can induce pulmonary fibrosis illustrating the relatedness between DKC and the adult-onset forms of the disease (Goldfarb et al. 2013). In addition to the classic triad and organ failures, endo- and epithelial tissues fail sporadically in DKC patients, leading to enterocolitis, emphysema, premature hair greying, dental caries, short stature, esophageal stricture and osteoporosis (Dokal 2011). DKC is considered a less severe form of HHS, Coats Plus Syndrome and Revesz Syndrome.

HHS is characterized by growth-related defects so severe that they are apparent prenatally, such as a intrauterine growth retardation, cerebellar hypoplasia and

microcephaly in addition to all symptoms observed in DKC (Aalfs et al. 1995). HHS patients generally have shorter telomeres than DKC patients and much shorter telomeres than the general population, almost always in the bottom 1% of telomere length, and HHS patients have a very high mortality rate, rarely surviving into adulthood (Walne et al. 2013b). HHS is distinguished from a related syndrome, Revesz syndrome, by the addition of exudative retinopathy in addition to the HHS symptoms (Kajtar & Mehes 1994). Coats Plus Syndrome was recently linked to mutations in the CTC1 gene, and these patients can exhibit both exudative retinopathy as well as cerebral calcifications (Anderson et al. 2012).

The nomenclature to classify these diseases is highly polluted by legacy classifications, as it is increasingly clear that the severe telomeropathies (Coats Plus syndrome, Revesz Syndrome and Hoyeraal-Hreidarsson Syndrome (HHS)) all describe a single disorder, as individuals with the same causal mutation can fall into different disease categories depending on which of a suite of incompletely penetrant rare symptoms they exhibit. For example, genotypically identical individuals in the same pedigree can show very different symptoms; this difference is so stark that full siblings with the same causal alleles as an HHS patient may show only bone marrow failure or the DKC classic triad while their sibling suffers from all of the symptoms of HHS (Walne et al. 2013b). Though exudative retinopathy can in theory differentiate HHS patients from Revesz Syndrome patients and cerebral calcifications can distinguish between Coats Plus and Revesz Syndrome, in practice patients with cerebral calcifications and exudative retinopathy have been diagnosed with all three severe forms of DKC, demonstrating that not only are these diagnostic guidelines drawn on incompletely penetrant symptoms that are fully non-predictive of genotype, they are applied haphazardly (Scheinfeld et al. 2007; Ramasubramanian & Shields 2012).

A number of groups have noted that telomeropathies are probably a single spectrum disorder rather than distinct diseases (Carroll & Ly 2009; Diaz de Leon et al. 2010; Armanios & Blackburn 2012; Holohan et al. 2014). It is not yet clear to what extent the differences observed between patients with different mutations are due to differences in the functions of those proteins and which differences are due to underlying biological differences in those patient cohorts, such as initial telomere length, synthetically pathological co-occurring mutations in that specific genetic background, or differing environmental exposures because of the small size of each cohort and the rarity of these

In addition to the disease linked to impaired telomere maintenance because of the causal genes' known association with telomere biology, there are a number of genetic diseases that have been recently associated with telomere shortening or otherwise impaired telomere maintenance. It is as yet unclear if the impaired telomere maintenance is correlative or causal with regard to these diseases' symptoms, if they are diseases with telomere shortening, rather than diseases of telomere shortening. Because this distinction is not yet clear, they are referred to as the "peripheral telomeropathies" (Holohan et al. 2014). However, it is likely that it is the telomere maintenance itself that matters, rather than the mechanism through which it is compromised. To date, in addition to the genes conventionally associated with telomerase, impaired telomere maintenance has been observed in patients with mutations in FANCD2, RECQL4, MPN1, VHL and DNMT3B (Smith et al. 1998; Lyakhovich et al. 2011; Mason & Sekiguchi 2011; Ghosh et al. 2012; Joksic et al. 2012; Shchepachev et al. 2012; Yehezkel et al. 2013; Ning et al. 2014). It is likely that the list of peripheral telomeropathies will grow as high-throughput sequencing approaches identify more loss-of-function mutations in bone marrow failure patients with very short

telomeres and decreasing technical hurdles result in measurement of telomere length in a much greater variety of diseases.

### *Summary and hypothesis*

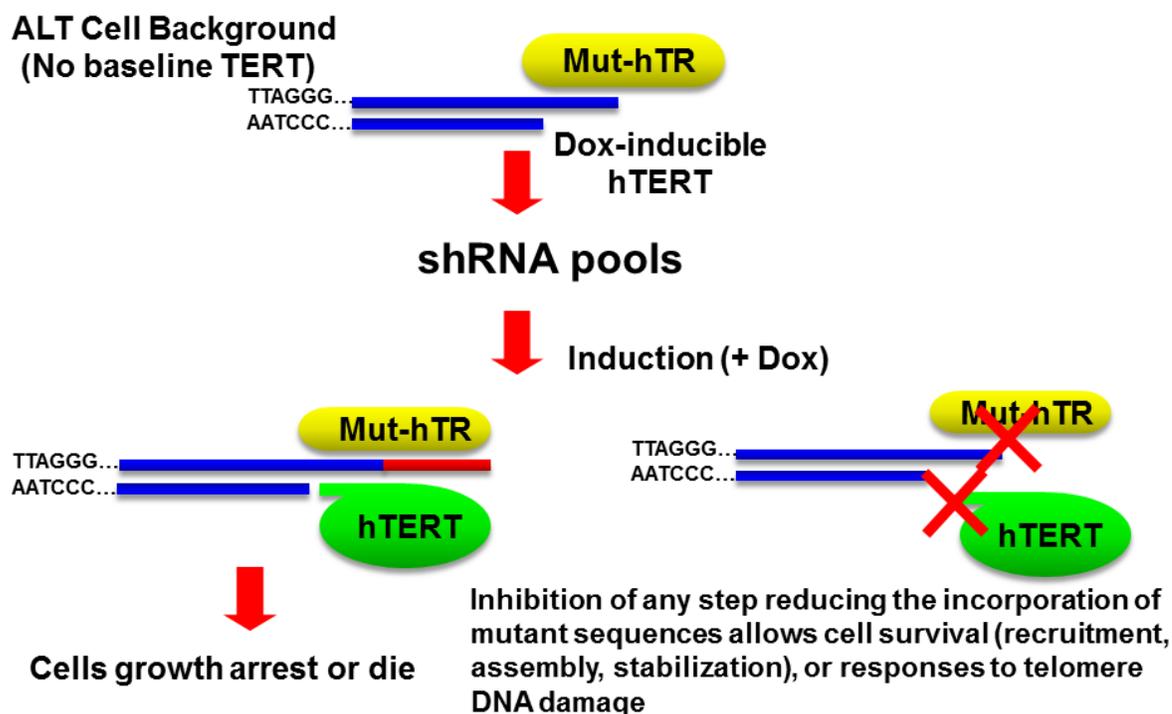
Telomeres function as an integrator of physiological stress at the organismal level and as a check to cancer progression and cellular immortality. As such, they represent an attractive target that no currently approved therapies are able to effectively leverage, and the identification of new telomerase inhibitors has the potential to dramatically improve cancer treatment. Further, the complex nature of telomere length regulation raises possibilities for intergenerational trends in telomere length that have not been investigated. In the following chapters, I will describe work to characterize the telomerase regulatory network, evaluate the clinical utility of a drug repurposed as a telomerase inhibitor, evaluate the observation of a trans-generational trend in telomere length, and utilize a large twin database to functionally interrogate the physiological telomerase regulatory network.

## **CHAPTER TWO**

### **Validation of an *in vitro* screen for positive regulators of telomerase**

#### **INTRODUCTION**

The goal of these studies is to identify new positive regulators of telomerase in order to discover new avenues for anti-telomerase cancer therapy. This initial work is based upon the results of a screen performed by a previous graduate student in the lab, Hirotooshi Hoshiyama. He performed a screen for positive regulators of telomerase following methods described in the literature, diagrammed in Figure 2.1 (Hoshiyama *et al.* 2012). In order to identify positive regulators of telomerase, the method exploited the templating function of the telomerase RNA component (hTR) in order to construct a cell line that was selectively killed by telomerase activity in order to use positive selection. The screening method used a system which included a doxycycline-inducible exogenous telomerase construct and an hTR construct with a mutated template region in the VA13 ALT cell line, which has no endogenous hTR and no baseline telomerase activity. This system resulted in cell death upon telomerase induction via doxycycline; telomerase activity resulted in the synthesis of non-canonical telomere repeats due to the mutant introduced and only source of hTR that could not be properly protected by the shelterin complex (Sfeir & de Lange 2012), triggering an unreparable DNA damage signal.



**Figure 2.1: Screening method to identify positive regulators of telomerase**

In this system, knockdown of a gene required for telomerase activity would allow a cell to survive the induction of telomerase activity; shRNAs that target positive regulators of telomerase should be enriched in the surviving cells

This system then allowed the use of a shRNA pooled screening method followed by parallel sequencing for barcode sequences attached to the shRNAs enriched in cells following telomerase induction in order to identify genes that are likely to be positive regulators of telomerase. shRNAs that interfere with telomerase activity are likely to target genes that are necessary for telomerase activity such as assembly, processing, transposition to telomeres during DNA synthesis, or inhibition of telomerase processivity. The initial stages of my work focused on the final results of this screening system and the *in vitro* validation of these results.

Hirotooshi performed several screens using this system because technical issues with the number of candidate shRNAs for each possible gene target in the genome-wide pool limited the utility of his initial results. The final analyses and the gene lists on which I based my initial work were derived from a second screen that eliminated DNA damage response factors via a counterscreen with DNA damaging agents which resulted in a much smaller library of 534 candidate genes at a density of 30 shRNAs per target. The genes were selected from a list that were identified from a literature search for all genes known at the time to interact, however distally, with a protein or process known to regulate telomerase or the DNA damage response. The hits chosen for final validation were drawn from a subset of the most significant hits from the genome-wide screen for positive regulators of telomerase as well as those candidate genes that were most highly enriched in the candidate screen without corresponding enrichment in the DNA damage controls. In short, the initial list was not an unbiased view of the telomerase regulatory network, but rather the result of a “screen-informed candidate gene approach”.

Information from model organisms, particularly *Saccharomyces cerevisiae*, is particularly useful for interpreting the results of the genome-wide screen. Following identification of a subunit of telomerase in *S. cerevisiae*, Ever Shorter Telomeres 1 (EST1) (Lundblad & Szostak 1989), yeast mutant screening eventually revealed that roughly 7% of all *S. cerevisiae* protein coding genes resulted in changes in telomere length upon deletion; over 400 different telomere length maintenance (TLM) genes have been identified to date (Askree *et al.* 2004; Gatbonton *et al.* 2006; Ungar *et al.* 2009; Harari *et al.* 2013). If the same proportion of human genes were important to telomere length regulation in humans, roughly 2,000 genes should be important to telomere length regulation, and if they were evenly distributed between positive and negative regulators of telomerase (as they appear

to be in the yeast studies), up to 1,000 hits from the *in vitro* screens could be real positives. Moreover, the number of identified genes in yeast jumped from 270 to over 400 when the ~1400 essential genes were screened via hypomorphic and temperature-sensitive screening. This indicates that genes involved in processes such as DNA repair and replication, core metabolic processes, protein degradation and transcription are more likely to be TLM genes than a given random gene; the shRNA-based method for screening used here most closely resembles a hypomorphic model, such that essential genes are less likely to result in false negatives from cell death.

The gene list chosen for validation via shRNA knockdown and long-term tissue culture is shown in Table 2.1. The list includes 3 positive controls (TERT, DKC1 and NHP2) as well as 56 genes derived from the most significant results of the genome-wide screen in addition to those genes that were most highly enriched in the secondary screen while avoiding enrichment in the DNA damage counterscreens.

In this chapter, I demonstrate how nearly all of the genes chosen for validation from the list of candidate positive regulators of telomerase induce a decrease in telomere length after extended tissue culture upon knockdown via shRNA. Detailed analysis of one of the candidates, LARP7 reveals a corresponding drop in telomerase activity. Additionally, in collaboration with other individuals in the lab and a group in Saudi Arabia, I show that LARP7 causes a telomeropathy in humans with naturally occurring loss-of-function mutations, and that loss of LARP7 alters the transcription and splicing patterns of the hTERT mRNA *in vitro*.

<b>TERT</b>	RAD52	LSM8	PRPF4B
<b>DKC1</b>	TNKS1BP1	ORC1	SFRS9
<b>NHP2</b>	FANCD2	RFC1	RAD54L
<u>LARP7</u>	FANCN	MRE11A	XRCC4
<u>CDK9</u>	EIF5A2	SMARCA1	RPA2
<u>LARP1</u>	PINX1	APEX1	PPP1R10
<u>CDC14C</u>	POLK	NIPBL	POLG
<u>DBF4</u>	CBX5	POLL	H2AFX
<u>BAIAP2L1</u>	POLH	POLI	RB1
<u>PGBD5</u>	POLD2	YWHAE	CLIC1
<u>DNAJB11</u>	E2F1	BMI1	CHEK1
<u>LGALS7</u>	HDGF	MCM2	CCNT1
<u>ZC3HAV1L</u>	CSRP1	RECQL4	ERCC2
POLR1E	ERCC3	ORC4	XPA
HELQ	BLM	ALKBH3	

**Table 2.1: The genes selected for validation via shRNA knockdown**

Positive controls are shown in bold, while the underlined genes comprise those genes that were highly significant in the genome-wide screen. The other candidates were selected because they were enriched in the secondary screen without enrichment in the DNA damage control.

## RESULTS

### *Knockdown of the hits chosen for validation leads to telomere shortening over time*

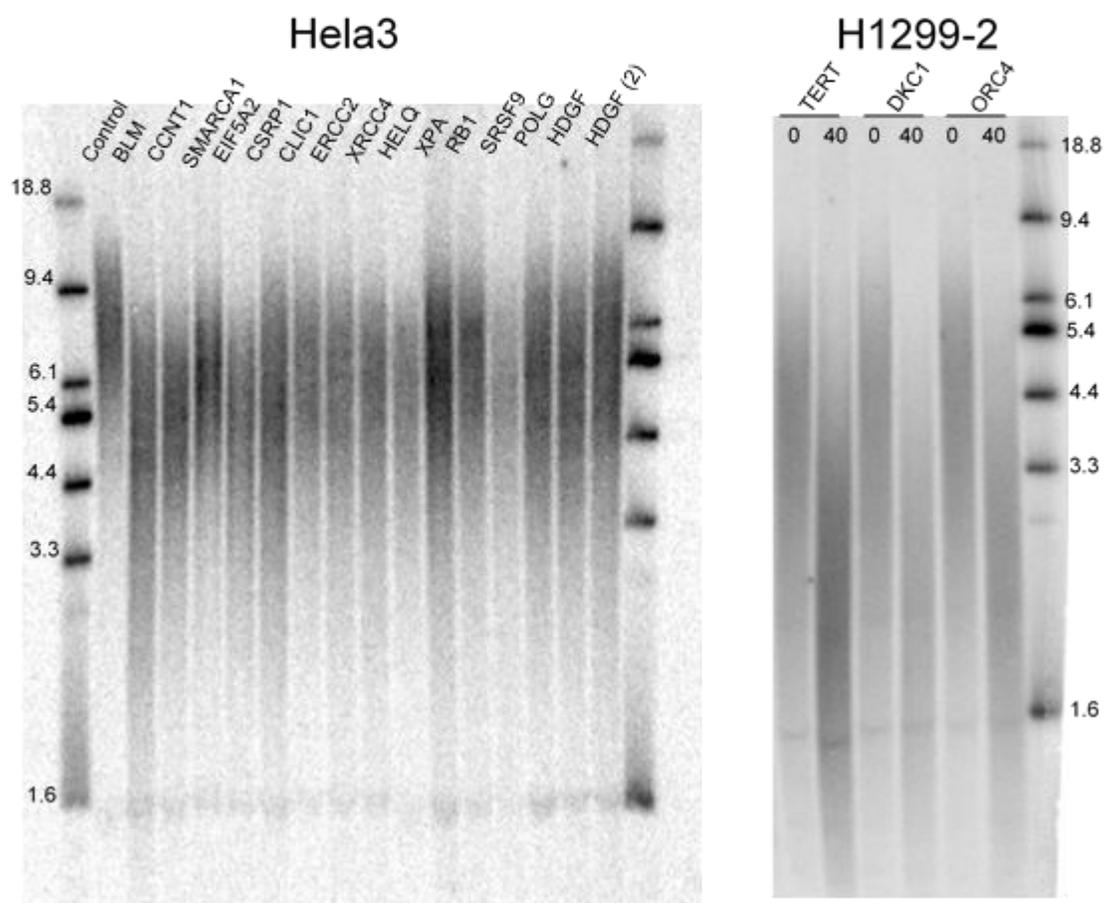
Clonal populations derived from the H1299 and HeLa cell lines derived by Hirotohi, H1299 clone 2 (H1299-2) and HeLa clone 3 (HeLa3), were selected as the validation models because of their intermediate telomere length and telomerase activity. I used clonal populations in order to avoid shifts in telomere length resulting from potential founder effects.

Because telomere length decreases comparatively slowly in the absence of telomerase, roughly 50-100 base pairs per cell division, prolonged tissue culture is required in order to assay if a genetic perturbation altered telomere length. As such, knockdown populations of these genes were culture for 40-60 population doublings in order to assay for a change in telomere length. Table 2.2 shows the aggregated results of the knockdown and telomere measurement. Stable knockdown of 31 of the candidate genes including all three positive controls resulted in telomere shortening in both lines, while a further 15 induced telomere shortening in one line but were not evaluated in the other line due to poor knockdown efficiency, 6 candidates induced telomere shortening on knockdown in one cell line but did not do so in the other cell line, and 3 did not induce telomere shortening in either context. Four of the candidates were not investigated (RECQ4L, CHEK1, ALKBH3 and LSM8) because of repeated failure of the knockdown. Representative Terminal Restriction Fragment (TRF) assays are shown in Figure 2.2.

Telomere shortening in both lines		Telomere shortening in one line	Mixed results
TERT	SMARCA1	MCM2	MRE11A
DKC1	POLI	TNKS1BP1	BMI1
NHP2	YWHAE	FANCD2	ORC4
LARP7	RAD52	POLK	POLH
CDK9	FANCN	CBX5	HDGF
LARP1	CSRP1	POLD2	CCNT1
CDC14C	SFRS9	PRPF4B	<b>Total: 6</b>
DBF4	XRCC4	RAD54L	<b>Negative</b>
BAIAP2L1	POLG	RPA2	ORC1
PGBD5	H2AFX	PPP1R10	RFC1
DNAJB11	ERCC2	CLIC1	POLR1E
LGALS7	XPA	NIPBL	
ZC3HAV1L	RB1	EIF5A2	
PINX1	ERCC3	APEX1	
E2F1	BLM	POLL	
HELQ	<b>Total: 31</b>	<b>Total: 15</b>	<b>Total: 3</b>

**Table 2.2: Summary of results of knockdown, long-term tissue culture and telomere measurements**

Genes that caused telomere shortening on knockdown in both cell lines assayed (left column), caused telomere shortening in one line but were not evaluated in the other line (middle column) caused telomere shortening in one cell line but did not cause shortening in the other (top right), and did not cause telomere shortening in either line (bottom right) are shown.



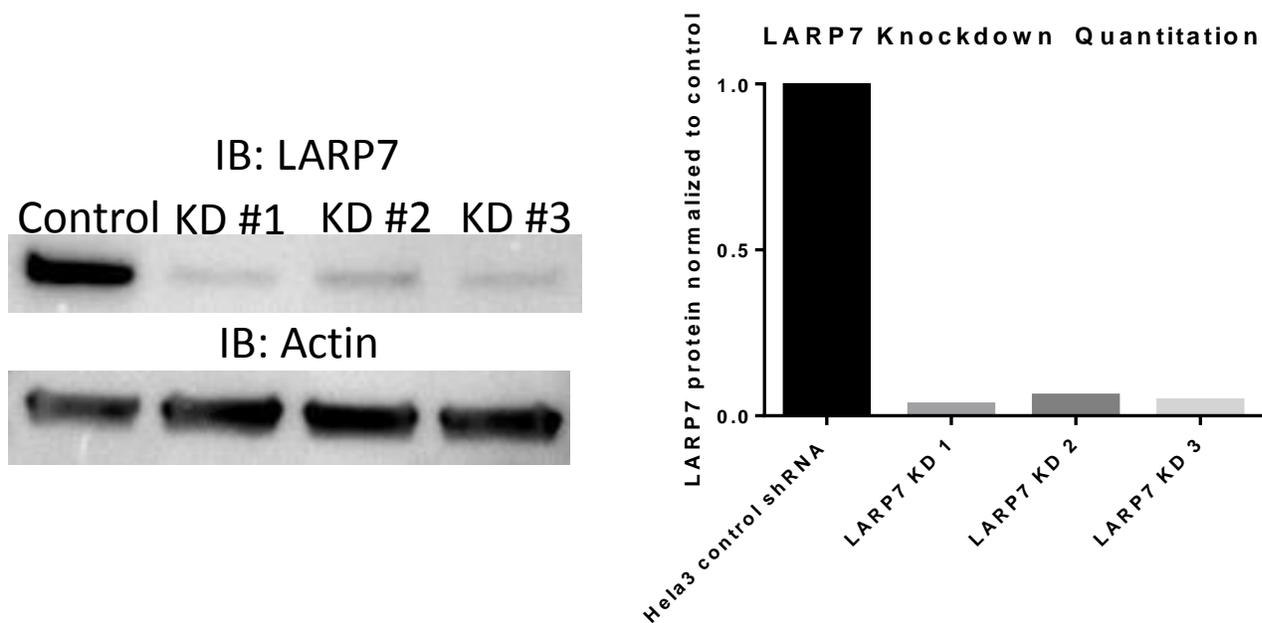
### Figure 2.2: Representative TRF images

Cells with shRNA constructs targeting the candidates are compared with HeLa3 cells with a scrambled shRNA construct (left panel, control) at population doubling 60. On the right panel telomere length is compared in H1299-2 population doubling (PD) 0 and 40 after selection for puromycin resistance.

In total, this experiment indicates that 43 of the 56 genes selected for validation regulate telomere length *in vitro*. Further, it illustrates that regulation of telomere length is a context-dependent process, as six of the candidates altered telomere length upon knockdown in one cell line but not in another.

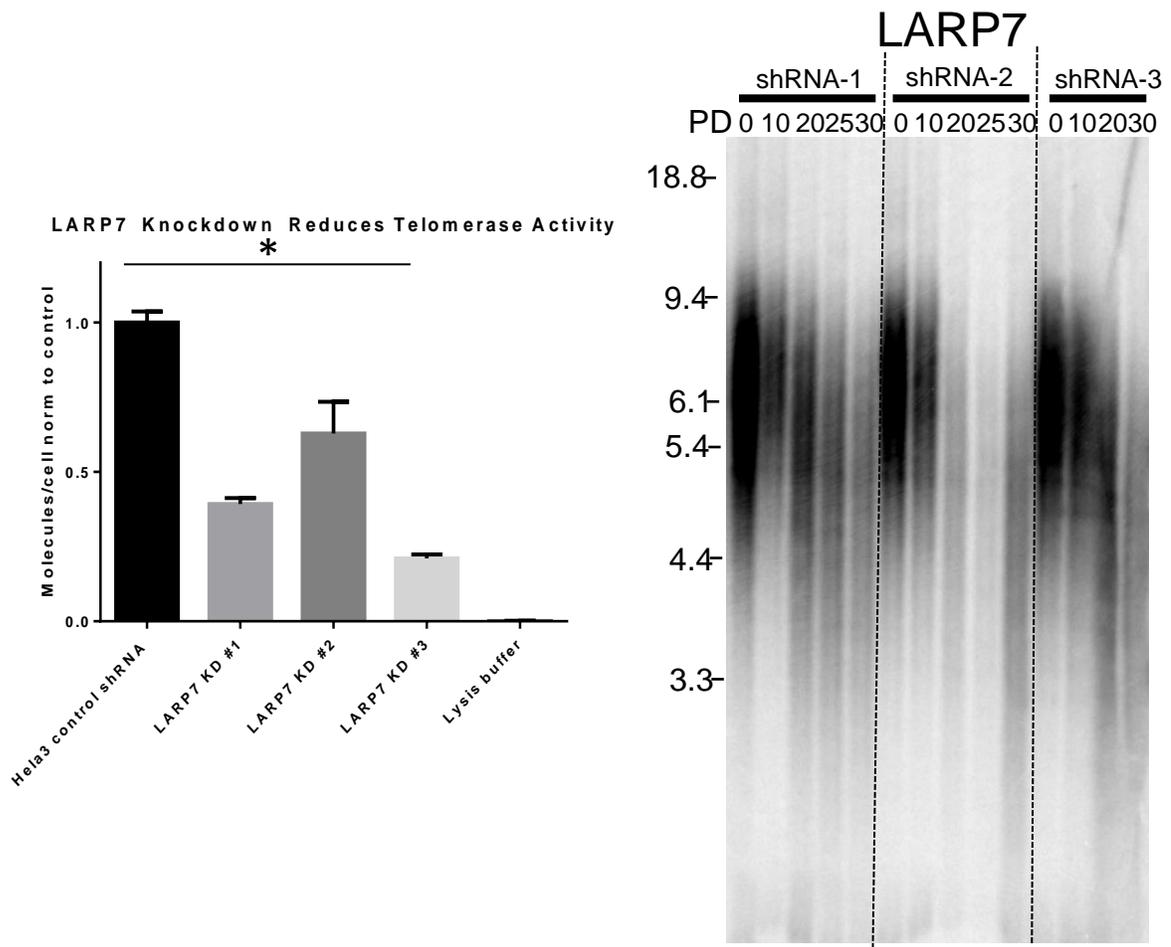
*Reduction in telomerase activity, telomere shortening and altered telomerase mRNA splicing upon knockdown of LARP7*

Based on a number of interesting observations in the literature, one candidate, LARP7, was selected for more detailed analysis. Knockdown of LARP7 was achieved with three different shRNAs and quantitated by western blot, shown in Figure 2.3. Knockdown of LARP7 resulted in a two to threefold reduction in telomerase activity in all three knockdowns, in addition to telomere shortening upon extended tissue culture, shown in Figure 2.4.



**Figure 2.3: Knockdown quantitation of LARP7**

Western blot for LARP7 in HeLa3 cells infected with a control shRNA and shRNAs targeting LARP7 indicates successful knockdown with all three constructs.

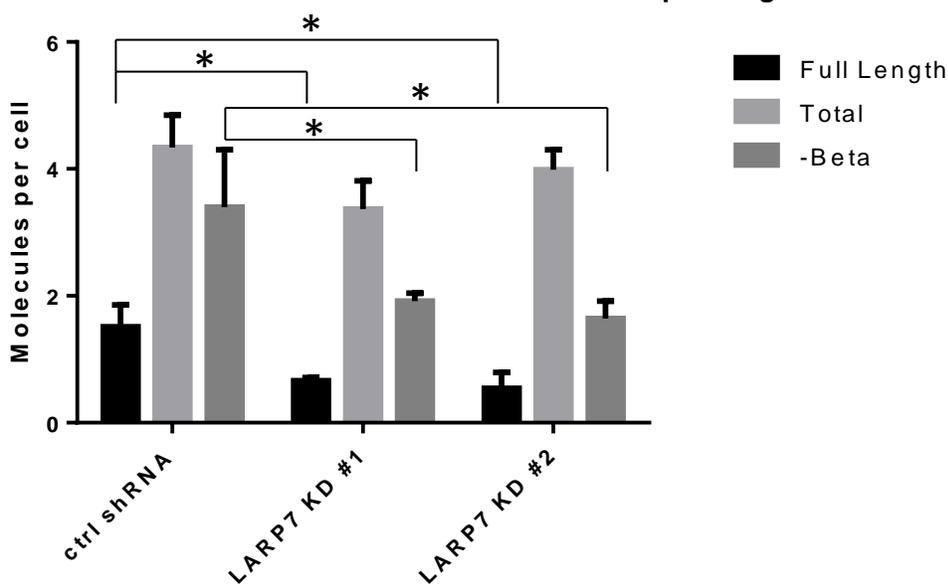


**Figure 2.4: Reduction in telomerase activity and telomere shortening in LARP7 knockdown cells**

Knockdown of LARP7 resulted in decreased telomerase enzymatic activity (left panel,  $p < 0.005$ ) and progressive telomere shortening over time in all three knockdowns examined.

In collaboration with Wanil Kim, a post-doctoral researcher in the lab, the splicing of the TERT mRNA in HeLa3 cells with LARP7 knockdown shRNAs, resulted in a reduction in the abundance of full-length TERT mRNAs, consistent with LARP7's role as a transcriptional regulator (Figure 2.5). LARP7 knockdown also caused a reduction in the abundance of the nonfunctional –beta isoform (removal of exons 7 and 8, which eliminates the reverse transcriptase domain) without a statistically significant reduction in total TERT transcripts, suggesting that the reduction in full-length was predominantly mediated by increased splicing of TERT pre-mRNA into a splice form not measured.

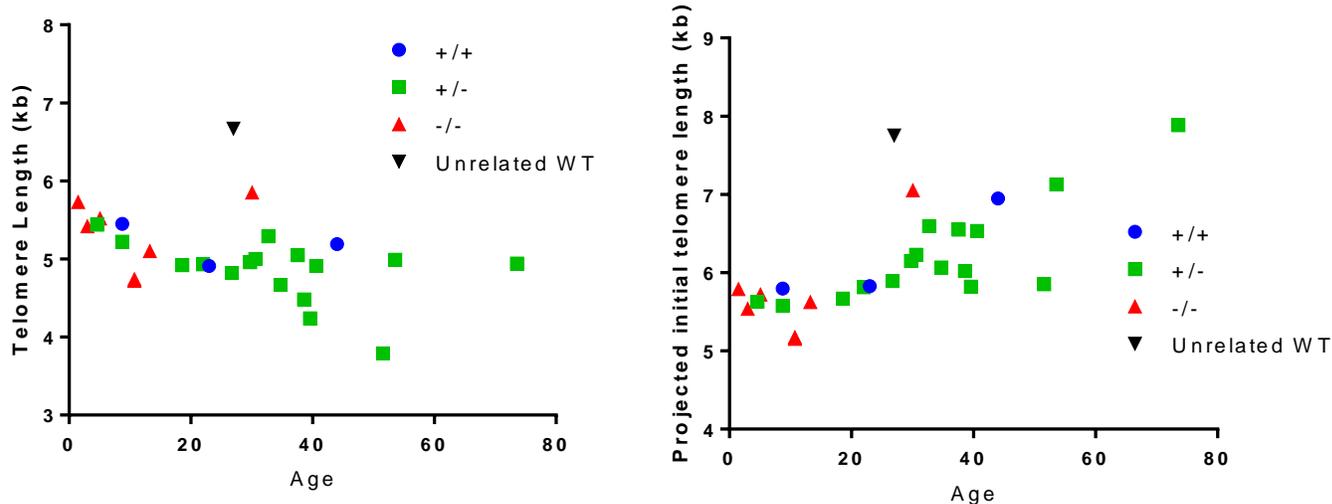
#### LARP7 Knockdown Alters Telomerase Splicing



**Figure 2.5: LARP7 knockdown decreases full-length and –Beta TERT splicing**  
LARP7 knockdown resulted in a decreased abundance of full-length and –beta splice forms of the TERT mRNA ( $p < 0.05$ ), though it did not induce a statistically significant reduction in total TERT transcripts.

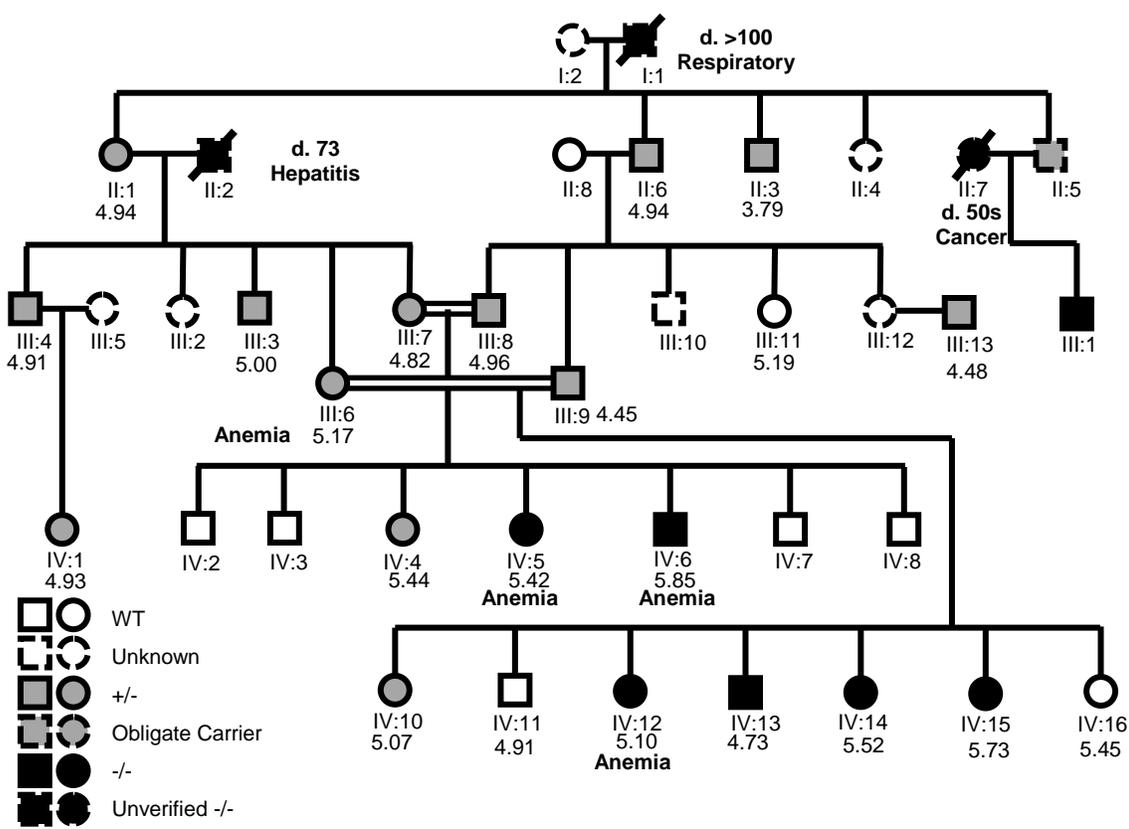
*Humans with a frameshift loss-of-function mutation in LARP7 suffer a classical telomeropathy*

Based on the *in vitro* data on LARP7's importance to telomerase activity, it seemed likely that humans with loss-of-function mutations in LARP7 should display symptoms of a telomeropathy. A consanguineous cohort in Saudi Arabia with a primordial dwarfism phenotype had recently been described in Saudi Arabia, and the causal mutation was traced to a 7 base pair duplication in LARP7 that caused a frameshift and early termination codon (Alazami *et al.* 2012). I contacted Fowzan Akluraya, the senior author of that study, and he sent me DNA samples extracted from peripheral blood mononuclear cells from the LARP7 mutant family to test if the family exhibited very short telomeres and a pattern of telomere length anticipation consistent with telomeropathies. TRF analysis of the LARP7 mutant family revealed very short telomeres, in length ranges consistent with TERT mutants, in all individuals in the family regardless of genotype (Figure 2.6, primary TRFs shown in Appendix A). This is not entirely unexpected, as all individuals in this family are descended from a homozygous LARP7 mutant individual, and the wild-type individuals examined were descended from at least two generations of LARP7 mutation carriers (pedigree shown in Figure 2.7). A follow-up interview conducted by the Saudi team, after I shared the results of the TRF analysis, revealed that multiple individuals in the LARP7 mutant family exhibited generalized anemia and that this anemia was concentrated in the youngest two generations of the family, consistent with what is observed in other telomeropathies.



**Figure 2.6: LARP7 mutants and their descendants have very short telomeres and generationally decreasing initial telomere length**

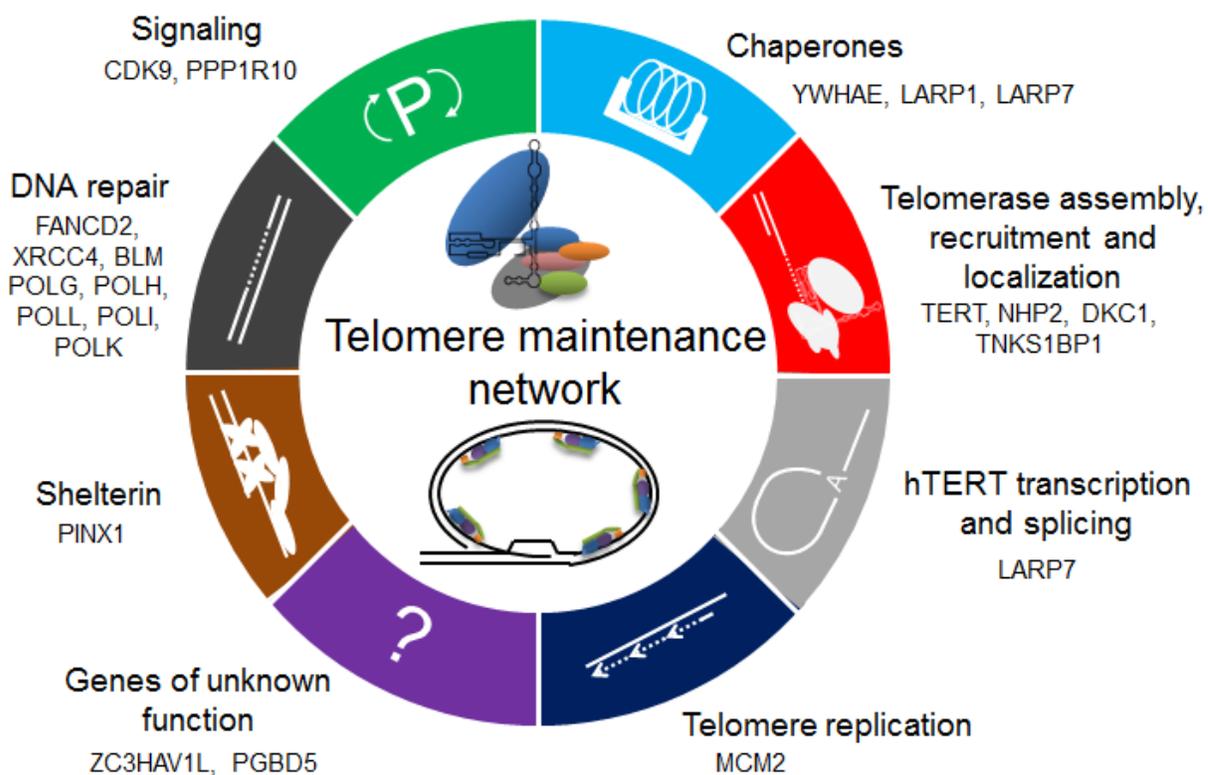
Telomere length in the LARP7 mutant cohort is very short compared to an unrelated wild-type individual and the general population regardless of genotype (left). Projected telomere length at birth (measured telomere length + age\*40 bp) decreases with age in this family, suggesting inheritance of short telomeres. Wild-type individuals are shown in blue, heterozygotes are in green, homozygous mutants are in red, and an unrelated wild-type individual is in black.



**Figure 2.7: Pedigree of the LARP7 mutant family**  
 Very short telomeres appear in all measured individuals regardless of genotype. Telomere lengths in kilobases are shown in addition to genotype and familial relationships. The pedigree illustrates a pattern of very short telomeres, pathology and a pattern of anticipation observed in telomeropathies

## **DISCUSSION**

The high positive rate observed upon knockdown of the candidate genes is consistent with observations from the *S. cerevisiae* literature, in that 43 of the 56 candidates appeared to be necessary for telomere length maintenance. This rate is not surprising since most of the identified candidates were pre-selected as important because of their links to processes known to be important to telomere biology. Telomerase activity is very closely tied to cell division, and prepositioning of telomerase is required for correct telomere maintenance (Zhao *et al.* 2011). Thus, perturbation of some important aspect of DNA replication or cell division would likely impact this careful prepositioning of telomerase, which could result in decreased telomere length over time. Evaluation of the general function of the validated hits in the literature indicates they fall into a number of general categories, detailed in Figure 2.8. It is probable that this network does not include all processes required for telomerase activity and telomere maintenance because of the hypothesis-driven mechanism of candidate selection.



**Figure 2.8: The telomere maintenance network**

Processes impacted by the validated hits are shown along with those genes implicated in their function.

LARP7 is part of a family of La autoantigen related proteins (LARPs) which include LARP3, which was previously shown to bind to the telomerase RNA component and act as a negative regulator of telomerase activity (Ford *et al.* 2001) and LARP1, another of the validated hit proteins. The La domain in general acts as a fairly promiscuous RNA-binding motif, associating with coding RNA polymerase II transcripts as well as pre-tRNAs (Bayfield *et al.* 2010), RNA polymerase III transcripts (Martino *et al.* 2015) and structural RNAs such as the 7SK snRNA (Liu *et al.* 2014) as well as the telomerase RNA component in humans (Ford *et al.* 2001) and in model organisms (Singh *et al.* 2013). LARP7 is an RNA-binding protein that contains the La domain as well as another RNA-recognition motif, RRM1, and

its canonical functions include an association with the 7SK small nuclear ribonucleoprotein (7SKsnRNP, or 7SK), which is a negative regulator of the positive transcription elongation factor, P-TEFb (Uchikawa *et al.* 2015). The primordial dwarfism observed in human LARP7 loss-of-function mutants may result from increased differentiation of stem cells due to decreased levels of the Lin28 mRNA (Dai *et al.* 2014) which is in agreement with the loss of self-renewal in primordial germ cells observed in LARP7 knockout mice (Okamura *et al.* 2012). LARP7 is the most likely human ortholog of p65 in *Tetrahymena thermophila*, which is an integral component of *Tetrahymena* telomerase, required for the assembly of the holoenzyme (Singh *et al.* 2013). Thus, LARP7 knockdown could result in telomere shortening through a number of pathways, and that a different combination of these possibilities is responsible for the phenotypes observed in tissue culture and in humans.

One of the more interesting observations in the LARP7 mutant family is that even wild-type individuals have very short telomeres. This observation is consistent with what is observed in TERT mutant families, in that even wild-type offspring of TERT mutant heterozygotes have telomere lengths considerably shorter compared to the general population (Chiang *et al.* 2010). Since all individuals evaluated were descended from at least two generations of LARP7 mutant carriers, they should all have very short telomeres if this inherited short telomeres effect was in operation. Inheritance of short telomeres within this cohort is further supported by the anticipation in telomere shortening shown by correcting the measured telomere lengths for age-associated telomere shortening. I accounted for the effects of age by using an estimate of 40 base-pairs per year of telomere shortening, which is derived from a review of the longitudinally-measured telomere shortening rate in the literature (Table 4.1), which will be discussed at greater length in Chapter 4. This estimate is a conservative estimate of telomere shortening rate for these

purposes, since the individuals evaluated here are not all post-pubescent, and are therefore longer than they would appear if they were evaluated after the faster telomere shortening period in childhood. This family displays starkly decreasing initial telomere length over time, a difference of on average 560 base pairs in initial telomere length between generation III and generation IV individuals, or roughly 30 base-pairs per year. The very short telomeres observed in the family, coupled with decreasing initial telomere length over generational time and the anemia apparent in younger generations are consistent to LARP7 deficiency as a classical telomeropathy.

Following my presentation of the LARP7 findings at the American Society of Human Genetics in October of 2014, Steve Meyn approached me because an anemic patient in practice in Canada had a frameshift mutation in LARP7; measurement of this patient's telomeres is pending, but Dr. Meyn performed a second physical examination of this patient, and the patient exhibits all the characteristics described in the LARP7 mutant family evaluated here (Alazami syndrome). If this patient's telomeres are considerably shorter than the general population it would lend considerable credence to the telomeropathy-like properties of the LARP7 mutants, since it would greatly decrease the odds of error as a result of a generally lower telomere length in the population of Saudi Arabia, which is unknown but formally possible.

Another validated candidate positive regulator of telomerase contains the La domain, LARP1. LARP1's known functions are very different from LARP7, and it is most prominently implicated in the regulation of 5'-terminal oligopurine (TOP)-containing mRNAs (Tcherkezian *et al.* 2014). TOP-containing mRNAs include the core translational machinery, such as the ribosome components (Avni *et al.* 1997). TOP transcript expression is tightly tied to cell

cycle progression, and LARP1 has recently been described binding to the mTOR complex 1 (mTORC1), as well as regulating a large number of transcripts implicated in cancer progression such as BCL2, AKT, BAX, EGFR, VEGFR, Beta-catenin, and SMAD3 (Mura *et al.* 2014). Telomere shortening observed in LARP1 knockdown cells is thus likely a result of a very large number of overlapping processes, and its association with AKT/mTOR signaling is instructive for experiments with AKT inhibitors and telomere length to be discussed in greater depth in chapter 3.

Telomere shortening upon knockdown of a number of high-level transcriptional regulators such as LARP1, LARP7, and CDK9 (a component of the P-TEFb complex which is negatively regulated by the LARP7-containing 7SK complex (Han *et al.* 2014)) as well as protein and RNA chaperones including YWHAE (14-3-3 epsilon) is consistent with the idea that appropriate telomere maintenance is the sum of many highly complex interrelated molecular processes, and that a broad-perturbation in the homeostatic state of a cell may interfere with proper telomere maintenance indirectly. Protein chaperones, such as p23, HSP70 and HSP90 are required for proper telomerase assembly (Forsythe *et al.* 2001), probably because TERT itself is a very large (127kD) protein that must stably associate with many other telomerase subunit proteins (DKC1, NHP2, GAR1, NOP10, TCAB1, etc.) as well as a large, structurally complicated RNA component that is itself associated with RNA chaperones (Egan & Collins 2012). A broad perturbation in transcription influencing various components of this large holoenzyme with an even larger molecular support network is likely to cause impaired telomere maintenance in much the same manner as proteotoxic stress preferentially kills aneuploid cancer cells. Aneuploid cancer cells are subject to stoichiometric stress as a result of their altered chromosomal copy number, and the addition of further proteotoxic stress overloads an already stressed system (Tang *et al.* 2011; Chen

*et al.* 2015). Similarly, because of the size, poor stability and complexity of the telomerase holoenzyme, stress to or greater demand for chaperones or general transcriptional systems is likely to result in a stoichiometric imbalance in one of the many necessary components of proper telomere maintenance, resulting in decreased telomere length over time via compromised telomerase activity. It is possible that this property of the network may be an adaptive feature of the telomerase regulatory network in multicellular organisms, as it would impose another check on cancer progression by lowering the odds that a given cell with copy number changes would be able to recapitulate adequate telomere maintenance. The large number of processes that must be behaving nominally for adequate telomere maintenance may impose a stoichiometric check on cellular immortality. An interesting test of this hypothesis would be to evaluate telomerase activity in LARP1, LARP7 or CDK9 overexpressing cells; if the telomere shortening observed on knockdown was the result of a direct down-regulation of some an important telomerase component, telomere length should increase, whereas the converse would be observed if the telomere shortening was the result of a stoichiometric imbalance in telomerase holoenzyme components.

Both FANCD2 and FANCN knockdown cells exhibited telomere shortening over time, which is consistent with the literature documenting short telomere length in a subset of Fanconi Anemia (FA) patients (Pavesi *et al.* 2009). Since FA patients as a group do not exhibit telomere shortening compared to other inherited bone marrow failure syndromes (Gadalla *et al.* 2010), short telomeres are not a general feature of FA, rather a subset of FA patients have very short telomeres, it seems most likely that the telomere shortening observed here is independent of the DNA repair functions shared between the various Fanconi anemia family DNA repair proteins. In particular, Apollo is required for the localization of FANCD2 to telomeres, indicating that FANCD2 may be involved in 5' end

resection or another processing event required for telomere maintenance. This function could be the reason that FANCD2 knockdown cells exhibit telomere shortening over time, though it is unclear if the telomere shortening observed *in vitro* and the telomere damage foci in human FANCD2 patients are involved with the phenotypic characteristics of FA, since FANCD2 patients have similar clinical characteristics as other FA patients (Mason & Sekiguchi 2011; Joksic *et al.* 2012). Furthermore, FANCD2 interacts with telomeric DNA and tankyrase I (TNKS1 regulates TRF1 telomeric binding) and FANCD2 mutant cells exhibit downregulation of TRF1, indicating that FANCD2 may be involved in several processes required for telomere maintenance (Smith *et al.* 1998; Lyakhovich *et al.* 2011).

Less is known about the association with FANCN and telomeres, however lymphoblastoid and fibroblast cell lines derived from FANCN heterozygous mutant patients exhibit copy number aberrations and concomitant increase in telomere fluorescent in situ hybridization (FISH) signals, suggesting that haploinsufficiency of FANCN may lead to genomic instability and/or telomere elongation (Wark *et al.* 2013). It is not known if these cells had longer telomeres on average, since increases in chromosomal copy number will be associated with a greater number of total telomeres, and increased telomere FISH signal could result from a larger number of average shorter telomeres. FANCN was originally identified as a BRCA2-interacting protein, and later found to be linked to FA; patients with biallelic mutations in FANCN have high rates of breast cancer, as well as high rates of pediatric cancers (Tischkowitz & Xia 2010). Importantly, long telomere length is a risk factor for non-hereditary breast cancer (Pellatt *et al.* 2013), and BRCA2 mutant individuals have longer telomeres than their non-BRCA2 mutant relatives (Pooley *et al.* 2014). It is possible

that FANCN mediates the BRCA2 mutation's effect on telomere length, though it is not known why telomere shortening would be observed *in vitro* on knockdown while dysfunction of FANCN is associated with longer telomere length *in vivo*. However, intergenerational analysis of BRCA1 and BRCA2 mutant, as well as BRCAX families (families without BRCA1/2 mutants with hereditary breast cancer) indicates anticipation in age of onset of breast cancer concomitant with progressively shorter telomeres (Martinez-Delgado *et al.* 2011), with intergenerational telomere dynamics highly comparable to that observed in this chapter in the LARP7 mutant family. Because FANCN is an important interaction partner with BRCA2, and FANCN mutants would be classified as a BRCAX family, it is possible that FANCN mutant families will display comparable telomere dynamics, which would be consistent with the telomere shortening observed here.

The DNA polymerases evaluated here (POLG, POLH, POLL, POLI and POLK) have a number of interesting biological features that may explain why their loss induces progressive telomere shortening. DNA polymerase gamma, POLG, is the only mitochondrial DNA polymerase encoded in the human genome, and mutations in POLG are associated with diverse mitochondrial diseases, including Alpers syndrome (progressive childhood failure of the central nervous system), Ataxia Neuropathy Spectrum disorders (broad dysfunction of the skeletal muscle and nervous system), Epilepsy and Myocerebrohepatopathy (failure of muscle and nervous systems as well as liver disease) (Wong *et al.* 2008). While mitochondrial dysfunction is unlikely to directly alter telomere length, buildup of reactive oxygen species from perturbed respiration can lead to telomere shortening via damage to the telomeres. This mechanism could be differentiated from a telomerase-dependent mechanism through POLG knockdown in a telomerase negative cell

line; greater telomere shortening per cell division should occur upon POLG knockdown compared with the unperturbed cells.

POLH (DNA polymerase eta) and POLI (DNA polymerase iota) are involved in the translesion DNA synthesis required for repair of UV-induced DNA damage, principally thymine dimers. Mutations in POLH are associated with Xeroderma Pigmentosum-variant (Masutani *et al.* 1999), a pathological sensitivity to UV light caused by an inability to repair UV-induced DNA damage. Recent work has shown that POLH localizes to telomeres in response to DNA damage, and that loss of POLH results in impaired telomere replication (Pope-Varsalona *et al.* 2014), providing a simple explanation for the telomere shortening observed upon knockdown. POLI performs functions very similar to POLH, though it has a higher error rate (Frank & Woodgate 2007), and it is possible that POLI knockdown induces telomere shortening through a mechanism very similar to POLH.

Though the network identified here is based on a screen-informed candidate gene approach, it demonstrates that a very substantial fraction of the genes highly enriched in the screening system are required for telomere maintenance, suggesting that the telomerase regulatory network is much larger than initially anticipated. Additionally, the LARP7 mutant family's very short telomere phenotype, comparable to telomerase mutants, illustrates that at least one of these candidate proteins is relevant to human health, and it is likely that other families with telomeropathies due to defects in the function of these genes are likely to be discovered. Further, though many of these genes are involved in pathways very central to the function of the cell, it is possible that drugs targeting these genes may act as telomerase inhibitors.

## **METHODS**

### **pGipZ vector system, cloning**

Knockdown of the candidate hits was accomplished through the introduction of two custom shRNAs targeting each candidate inserted into a pGipZ construct, which included genes encoding Green Fluorescent Protein (GFP), puromycin resistance and ampicillin resistance. The pGipZ vector was linearized via digestion with XhoI and EcoRI restriction enzymes (37° C for 1 hour, 65° C 30 minutes to deactivate enzymes) and then gel purified from a 1% agarose gel with the QIAquick gel extraction kit (Qiagen). Custom shRNA sequences were synthesized (Sigma) containing XhoI and EcoRI restriction sites and PCR amplified using the high-fidelity KOD polymerase (Millipore) for 20 cycles according to the manufacturer's instructions. PCR products were then purified with the QIAquick gel extraction kit and digested with EcoRI and XhoI restriction enzymes as previously. Ligation of PCR products and linearized vectors was accomplished with T4 DNA ligase at 16° C overnight, followed by inactivation at 65° C for ten minutes. Competent *E. coli* were transformed via the addition of the ligated plasmid followed by a 30 minutes incubate at 4° C and then a 45 second heat shock. Cells recovered on ice for 2 minutes and were then amplified through the addition of 250 uL prewarmed SOC media and a 1 hour incubation at 37° C with agitation at 225 RPM. The cells were plated on ampicillin-bearing bacterial agarose plates and grown overnight. The next morning, resistant colonies were picked and grown in 50mL ampicillin-bearing Luria broth and the plasmids were extracted with the QIAprep spin miniprep kit (Qiagen). Plasmids were assayed on a 1% agarose gel to verify they were the appropriate size, then used in viral packaging.

### **Viral Packaging, Infection**

The resulting plasmids were packaged into lentiviral vectors, integrating replication-incompetent retroviruses. 5 million 293FT cells were plated into 10cm plates in standard culture conditions. Finished plasmids were mixed with the pasPAX2 structural vector and pMDG2 envelope vector, as well as CaCl<sub>2</sub> and HEPES-buffered saline. The mixture was then vortexed 30s to generate bubbles, and allowed to sit for 15 minutes before being applied to viral packaging cells. 24 hours following the calcium chloride transfection, cells were checked for plasmid expression via fluorescence imaging, and if GFP positive viral the media was changed and viral supernatants were obtained 24, 48 and 72 hours later.

H1299-2 and Hela3 populations were infected with the lentiviruses supernatants, and the cells were puromycin selected and cultured for 40-60 population doublings.

### **Tissue culture**

All cells were cultured at 37° C in 5% CO<sub>2</sub>, in Media X (HyClone, Logan UT) with the addition of 10% calf serum (HyClone, Logan, UT). Cells were passaged 1:32 when they approached 90% confluence; they were washed 2x with phosphate buffered saline and then harvested by trypsinization (5 minutes at 37° C). Selection for puromycin resistance was accomplished via 5 days culture with puromycin (5 ug/mL). Cells were puromycin selected after viral infection and again at PD 20. DNA for TRF analysis was harvested using the DNeasy blood and tissue kit (Qiagen).

### **Knockdown validation**

Quantitation of the knockdown efficiency for each gene was accomplished with primer pairs and hydrolysis probes designed with the tools available from the manufacturer of the universal probe library (Roche). RNA was extracted from infected and puromycin

selected cells with the RNeasy mini kit (Qiagen), and reverse transcribed to cDNA with the iScript cDNA synthesis kit (Bio-Rad). qPCR analysis of RNA abundance was measured using the Ssofast Evagreen qPCR mix (Bio-Rad) and quantified following amplification using a Roche Lightcycler 480II qPCR platform (Roche) in reference to three housekeeping genes, GAPDH, PPIA and HPRT.

### **Terminal restriction fragment assays**

DNA for TRF analysis was extracted with the DNease blood and tissue kit (Qiagen), and then 2 ug of DNA from each sample was digested with the mixture of six restriction enzymes as described in (Herbert *et al.* 2003) for 16 hours at 37° C, and then samples were run on a 0.7% TAE gel in the presence of gel red for 18 hours at 70 volts. The resulting gel was imaged to assay for complete sample digestion, before the was denatured in a solution containing 1.5M NaCl<sub>2</sub> and 0.5 M NaOH on a shaker at room temperature for 30 minutes. The gel was then dried for 2 hours at 56° C using a SGD4050 slab gel dryer under vacuum with the addition of two large textbooks (~2kg each) of pressure, with two layers of whatman paper between the gel and the dryer. The gel was then washed 2x with deionized water to remove fragments of whatman paper attached to the gel and larger fragments were removed manually with forceps, then the gel was neutralized with a 1.5 M NaCl<sub>2</sub>, 0.5M Tris-HCl, pH 8 solution for 30 minutes at room temperature on a shaker. The gel was then placed in a hybridization tube with 15 mL rapid hybridization buffer (GE healthcare) for one hour at 42° C. Following pre-hybridization, 15 mL fresh pre-warmed rapid hybridization buffer was added to the hybridization tube and the c-strand telomere sequence probe (described below) was

added to the solution; hybridization of the probe occurred overnight at 42° C. The following day, the probe mixture was preserved for later use and the gel was washed 1x with wash buffer 1 (2x SSC, 0.1% SDS), 2x with wash buffer 2 (0.5x SSC, 0.1% SDS), and 2x with wash buffer 3 (0.5x SSC, 1% SDS). Each wash step was 15-30 minutes in duration. Following the final wash, the gel was placed between two layers of saran wrap in an exposure cassette and a storage phosphor screen (Molecular dynamics) was used to obtain an overnight exposure. The screen was imaged using a Typhoon TRIO variable mode imager (Amersham Biosciences).

#### **Western blot for LARP7**

Cells were harvested for western assays via trypsinization (5 minutes at 37° C), and then counted with the TC automated cell counter (Bio-Rad). Cells were then pelleted by centrifugation at 700 rpm for 3 minutes, then washed with cold (pre-chilled on ice) PBS 2x. A final pelleting step at 2000 rpm for 3 minutes occurred before the cell pellets were resuspended in RIPA bugger (150 mM NaCl, 50 mM Tris-HCl pH 7.4, 1mM EDTA, 1% TritonX-100, 1% sodium deoxycholic acid, 0.1% SDS) at a ratio of 50uL per 1 million cells. Cells were lysed through 3 cycles of 20 second vortex steps and 5 minutes on ice, then centrifuged at 20000 rpm for 10 minutes. Samples were then denatured by incubation at 95° C for 5 minutes. Protein concentration was quantitated with the BCA protein assay kit (Pierce) according to the manufacturer's instructions. Samples were loaded into 10 well preformed Mini Protein TGX gradient polyacrylamide gels (Bio-Rad) and run at 100 volts until the lowest ladder band reached the bottom of the gel. Proteins were

transferred to a PVDF membrane using a trans-blot turbo pack (Bio-Rad). The membrane was then washed 2x with PBST, then blocked with a 5% milk in PBST solution for 1 hour at room temperature on a shaker. The LARP7 primary antibody (Abcam) was used at a concentration of 1:5000 following blocking and allowed to hybridize overnight at 4° C. The membrane was then rinsed 3x with PBST at room temperature on a shaker. The anti-rabbit secondary antibody was used at a concentration of 1:5000 in a 5% milk PBST solution and incubated at room temperature for 1 hour on a shaker. After administration of the secondary antibody, the membrane was washed 3x with PBST, and signal was detected with ECL plus western blotting detection reagents (Amersham). Signal was obtained by with the G Box imaging system (Syngene).

## **CHAPTER THREE**

### **Characterization of Perifosine as a Possible Telomerase Inhibitor**

#### **INTRODUCTION**

AKT, or Protein Kinase B is a Serine/Threonine kinase downstream of PI3K, and regulation of this kinase is important for a very large number of signaling events with regard to cell growth, survival and division (Ocana *et al.* 2014). A number of links between the protein subunit of telomerase, TERT and AKT exist, including two consensus AKT phosphorylation motifs at S227 and S824 of the telomerase protein. These RXRXXS/T sites appear to be involved in the nuclear localization of the telomerase holoenzyme (Kang *et al.* 1999; Chung *et al.* 2012). Further, complexes containing AKT, HSP90, mTOR, S6 kinase and TERT have been observed (Haendeler *et al.* 2003), and treatment with Rapamycin has been reported to reduce telomerase activity, consistent with AKT/mTOR-based regulation of telomerase activity (Bae-Jump *et al.* 2006).

Many of the validated candidates examined in Chapter 2 are downstream of the AKT or mTOR pathways, including LARP1, LARP7, and LGALS7. Inhibition of AKT caused at least a two-fold reduction in phosphopeptides derived from these hits by mass spectrometry (Andersen *et al.* 2010). Since the AKT/mTOR pathway was implicated as important to telomerase activity in the literature and a number of the newly-identified proteins that regulate telomere maintenance are downstream (and in the case of LARP1 also upstream) of AKT, I tested if a number of AKT inhibitors could inhibit telomerase activity and induce telomere shortening over time in order to evaluate this approach for anti-telomerase therapy. Initial work focused on readily available small molecules only suitable for *in vitro* studies, but later examined Perifosine, an orally-available alkylphospholipid drug that is in phase III

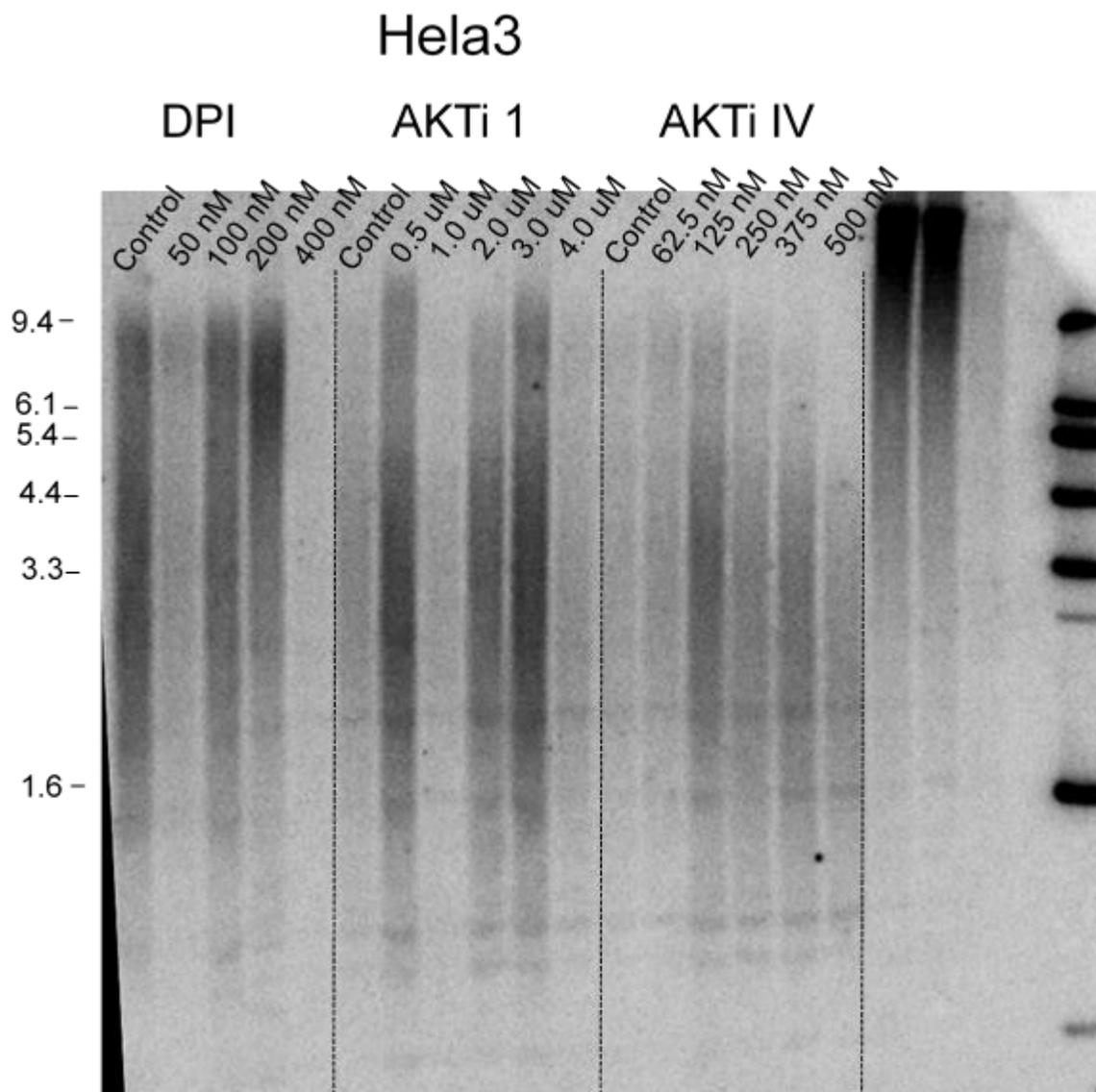
clinical trials for multiple myeloma. Perifosine interferes with the pleckstrin homology domain of AKT, which prevents its membrane localization and phosphorylation (Gills & Dennis 2009). Perifosine is an attractive candidate for anti-telomerase therapy because it has a long half-life in blood (~100 hours) and a low side effect profile; anti-telomerase therapy must be tolerable for long periods of time because telomere shortening generally occurs during cell replication and gaps in telomerase inhibition could allow cancer cells to re-extend telomeres and thwart the treatment (Unger *et al.* 2010).

In this chapter I describe the initial experiments examining the effect of various AKT inhibitors on telomere length *in vitro*, a metastatic xenograft model testing Perifosine's clinical utility in inducing telomere shortening, evaluation of purified tumor samples taken from chronic lymphocytic leukemia (CLL) patients treated with Perifosine for a long period of time as part of a Phase II clinical trial, and conclude by describing a number of experiments regarding the mechanistic details of Perifosine's effect on telomere maintenance.

## **RESULTS**

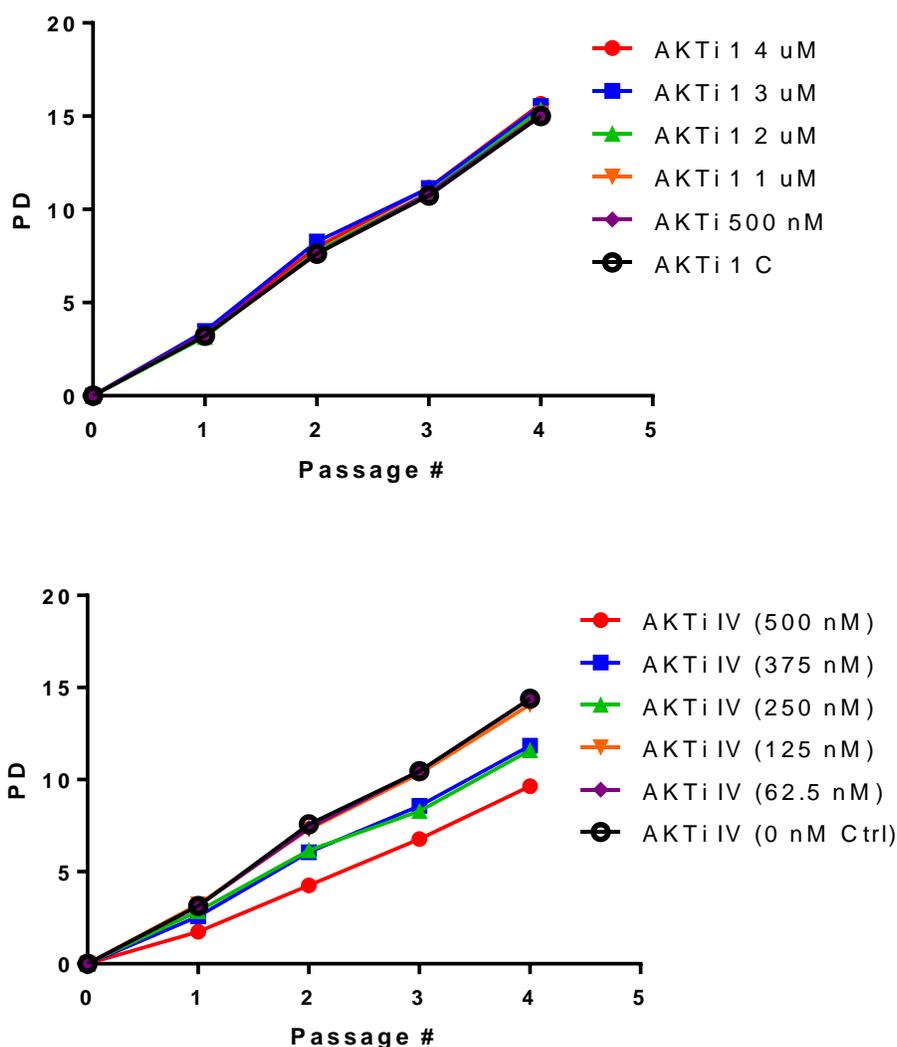
### *AKT inhibitors induce telomere shortening in some cell lines in vitro*

To test the initial hypothesis that AKT inhibitors may also function as telomerase inhibitors, I treated HeLa3 cells with low concentrations (less than 50% of their IC<sub>50</sub> value in HeLa cells) for twenty population doublings to assay if sub-lethal concentrations of these drugs could impact telomere biology. Of the three AKT inhibitors initially evaluated, AKT inhibitor IV induced telomere shortening in a dose-dependent manner (Figure 3.1).



**Figure 3.1: AKT inhibitor IV induces dose-dependent telomere shortening**  
HeLa3 cells were treated with the indicated concentrations of drug for 10-15 population doublings, and their telomeres were measured via the TRF assay.

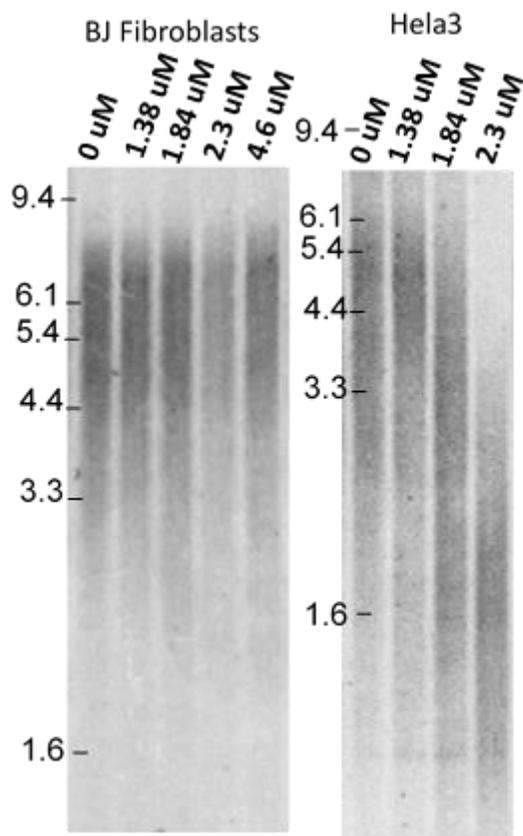
Concentrations of AKT inhibitor IV that induced telomere shortening also slowed down cell growth, though this growth inhibition did not fully correlate with telomere shortening; the 375 nM dose induced more telomere shortening compared to the 250 nM dose, even though they caused roughly comparable changes in growth rate (Figure 3.2) suggesting some off target effects.



**Figure 3.2: AKT inhibitor IV inhibited cell growth**

Hela3 cells treated with the indicated concentrations of AKT inhibitor 1 and IV were counted when they were passaged. Cells treated with higher concentrations of AKT inhibitor IV exhibited a slower growth rate, demonstrated by lower population doublings (PD) at each passage.

The rapid telomere shortening observed with AKT inhibitor IV treatment was encouraging; however AKTi IV itself is not a promising candidate for anti-telomerase therapy because of its pharmacological properties. In order to evaluate a small molecule AKT inhibitor much more clinically relevant, I repeated this experiment using Perifosine in HeLa3 cells as well as a BJ fibroblasts, telomerase negative cells, in order to determine if effects on telomere length was telomerase-dependent (Figure 3.3). Again, concentrations up to one half of the reported IC50 value in HeLa cells were used, though cell death in the HeLa3 population at this value (4.6  $\mu\text{M}$ ) prevented analysis of that dose.

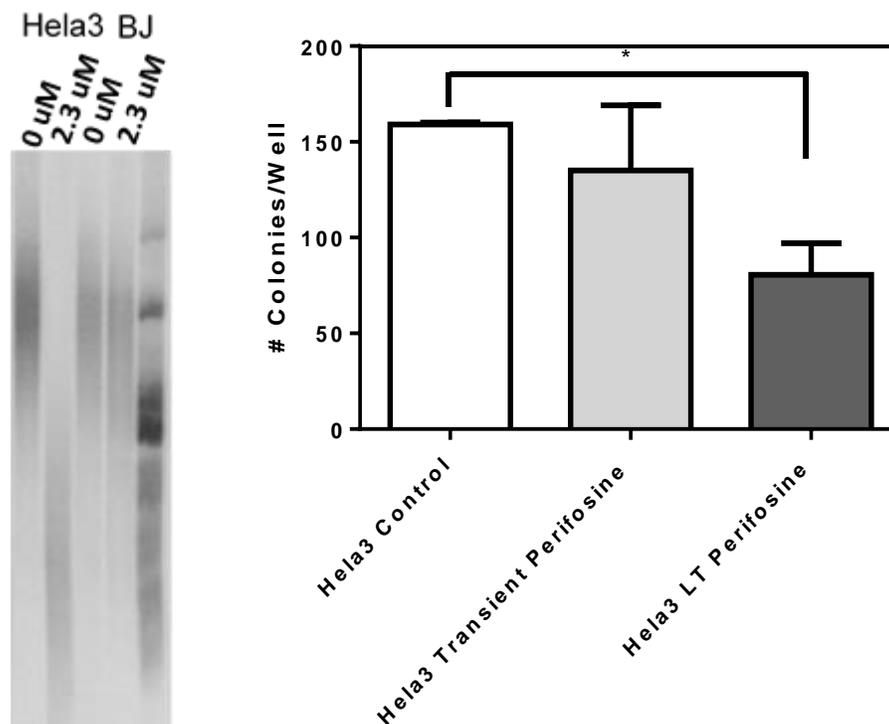


**Figure 3.3: Perifosine causes telomere shortening in telomerase positive cells** BJ fibroblast and HeLa3 cells were treated with the indicated concentrations of Perifosine for 20 population doublings and their telomeres were measured by TRF.

Perifosine caused telomere shortening in the HeLa3 population without accelerating telomere shortening in the BJ fibroblast line, indicating that the shortening observed is likely to be telomerase dependent.

*Long-term treatment with Perifosine inhibits colony formation in soft agar*

To determine if long-term Perifosine treatment will lead to loss of cancer associated changes in long-term treated cancer cells, HeLa3 cells (treated for 70 population doublings) were evaluated via the soft-agar colony formation assay, a measurement of growth and invasiveness. Long-term treatment with Perifosine, but not transient treatment with Perifosine (administration of Perifosine during the soft agar colony formation assay to cells previously naïve to Perifosine) resulted in a significant drop in colony number, indicating that Perifosine interfered with colony formation through a mechanism independent of AKT-mediated growth inhibition, which would be apparent in the cells treated transiently (Figure 3.4).

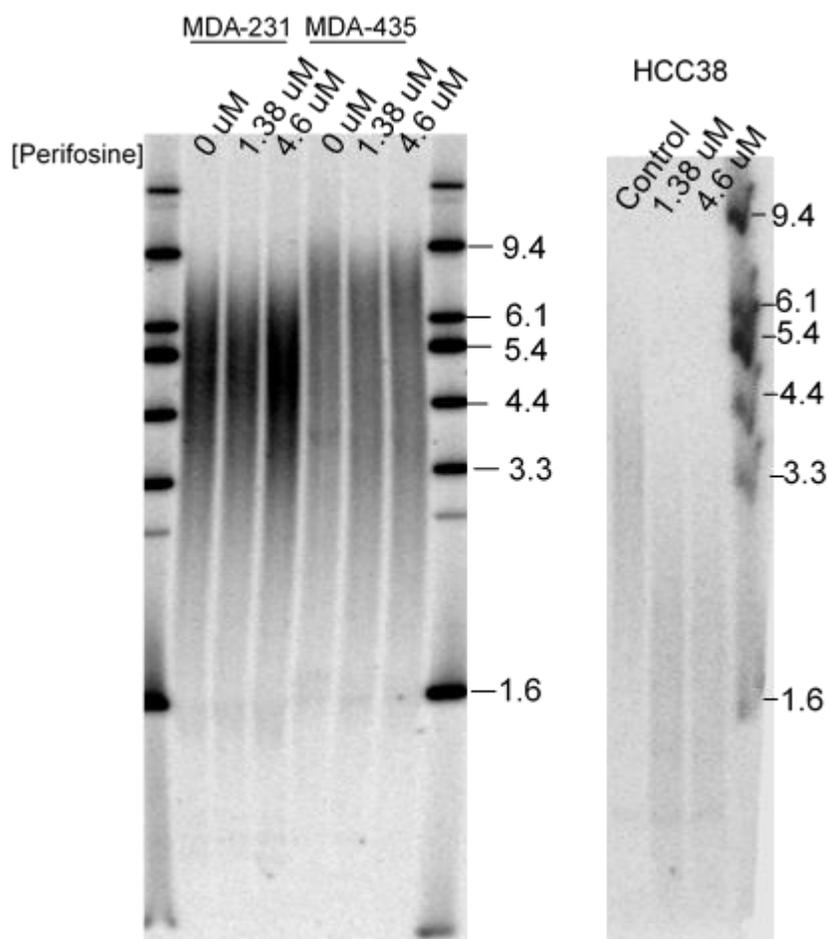


**Figure 3.4: Long term treatment with Perifosine reduces soft agar colony formation**

HeLa3 cells were treated with 2.3uM Perifosine for 70 population doublings and their telomeres were measured by TRF analysis (left). Long-term (LT) treated HeLa3 cells and cells naïve to Perifosine were examined via the soft agar colony formation assay; cells treated with Perifosine for 70 population doublings prior to the assay exhibited a statistically significant ( $p=0.04$ ) reduction in colony formation with treatment during the assay, while HeLa3 cells treated only during the assay did not.

*The HCC38 breast cancer cell line responds most robustly to Perifosine of the cell lines tested*

Concurrent with mechanistic experiments evaluating the molecular changes induced by Perifosine, I began to test Perifosine's clinical utility as a telomerase inhibitor. The most likely application for a telomerase inhibitor is in the context of minimal residual disease—telomerase inhibitors will likely need a fair amount (months to years) of time to work because they induce telomere shortening in a cell division dependent manner, which means they would be used best if treatment began after removal of a primary tumor. Treatment would continue while any cells not removed during surgical removal of a primary tumor are growing, such that presumably any recurrent disease would encounter a telomere length-mediated block to continued cell division. In order to test this, a model that includes excision of a primary tumor followed by telomere length analysis in recurrent primary tumor or metastatic disease would be the ideal test of a putative telomerase inhibitor. Orthotopic mammary fat pad xenograft models fit this profile; primary tumor cells are injected into the mammary fat pad of immune compromised mice, and metastatic growth after removal of the primary tumor can be assayed. Both of the cell lines conventionally used for this assay had telomeres too long to realistically shorten to a sufficient extent upon Perifosine treatment to limit tumor growth in this assay, however the HCC38 breast cancer line is an ideal target, as HCC38 displays very short baseline telomere length and its telomeres shorten rapidly upon administration of low doses of Perifosine (Figure 3.5).



**Figure 3.5: Perifosine induces telomere shortening in HCC38 and MDA-MB-435 cells**

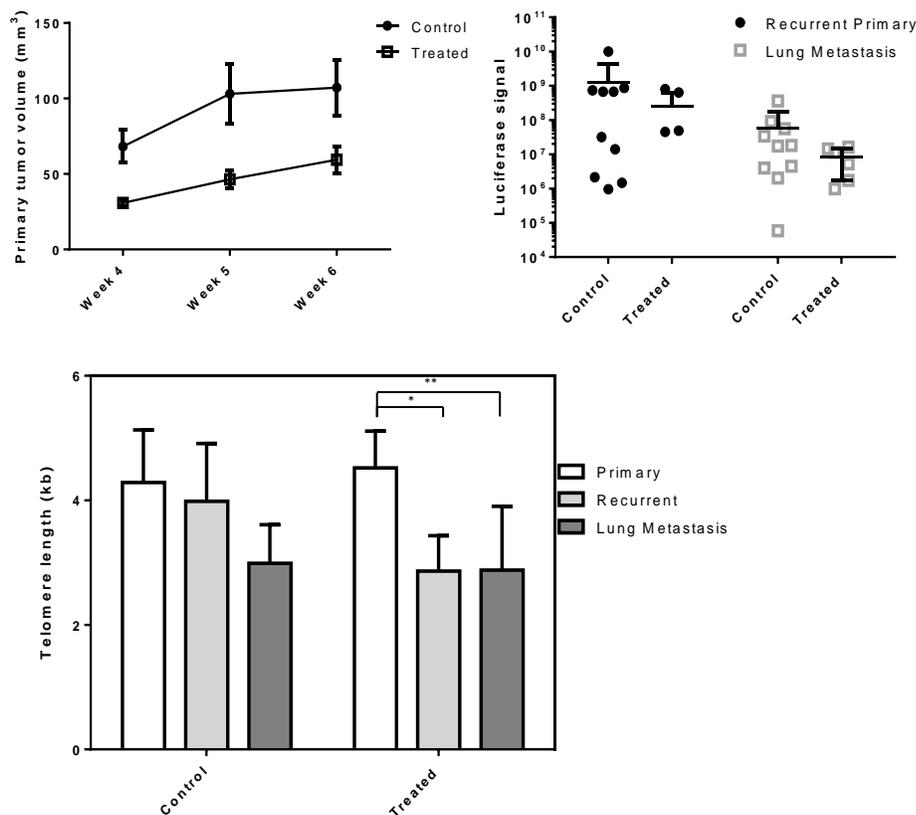
Cells were treated with the indicated concentrations of Perifosine for 20 population doublings, and their telomeres were measured by TRF. HCC38 cells had lower baseline telomere length and a stronger response to Perifosine than the two cell

Subsequent xenograft experiments used the HCC38 line because it was the cell line most likely to show an important response to Perifosine; because of its short telomeres and rapid shortening upon Perifosine treatment, it is an *in vitro* model of a best-case clinical scenario.

*Perifosine may act as a telomerase inhibitor in a xenograft model, but does not reduce metastatic tumor burden*

Other telomerase inhibitors have been evaluated in xenograft models (Joseph *et al.* 2010; Marian *et al.* 2010; Barszczyk *et al.* 2014; Hu *et al.* 2014b), however measurement of telomere length in xenograft tumors is complicated by intense telomere signals from the much longer telomeres of the mouse support cells (vascular epithelium, fibroblasts) that are present in a xenograft. The cells used in this experiment are HCC38 cells bearing a plasmid that encodes Luciferase, Green Fluorescent Protein (GFP) and puromycin resistance (HCC38+GFP+LUC), so mouse cells can be killed after tumor removal via a short *in vitro* treatment with puromycin prior to telomere length analysis. Treatment with Perifosine by oral gavage began in week 2 after injection of the tumor cells, primary tumors were surgically removed in week 6, and the mice were sacrificed and evaluated in week 12, totaling 10 weeks of Perifosine treatment.

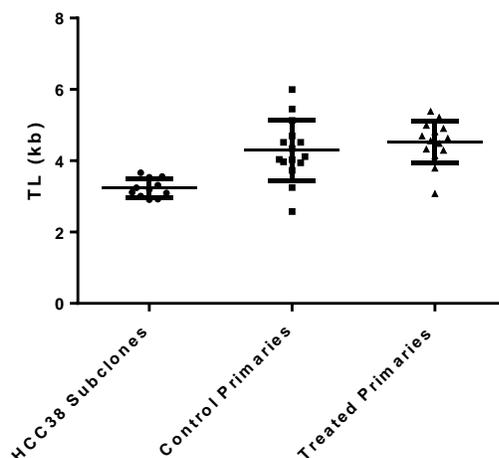
Though Perifosine reduced the growth rate of primary tumors when they were measurable, a finding reported in the past, treatment with this dose did not induce a statistically significant reduction in either recurrent primary or metastatic tumor burden. There was not a significant difference in telomere length between control and treated tumors of any type, but metastatic and recurrent treated tumors had significantly shorter telomeres compared to primary tumors in the treated group (Figure 3.6). These results can be interpreted to suggest that Perifosine has a modest but detectable effect, though it may also reflect the tighter distribution of telomere lengths in treated primary tumors.



### Figure 3.6: Perifosine treatment did not alter metastatic tumor burden or telomere length in a xenograft

Perifosine treated primary tumors were smaller on average than control tumors at all timepoints measured (top left). However, neither recurrent primary tumor nor metastatic tumor load were significantly reduced by Perifosine treatment (top right). Telomere length in treated recurrent primary and metastatic tumors was significantly shorter in treated mice (bottom), though there was not a significant difference between control and treated tumors of any type.

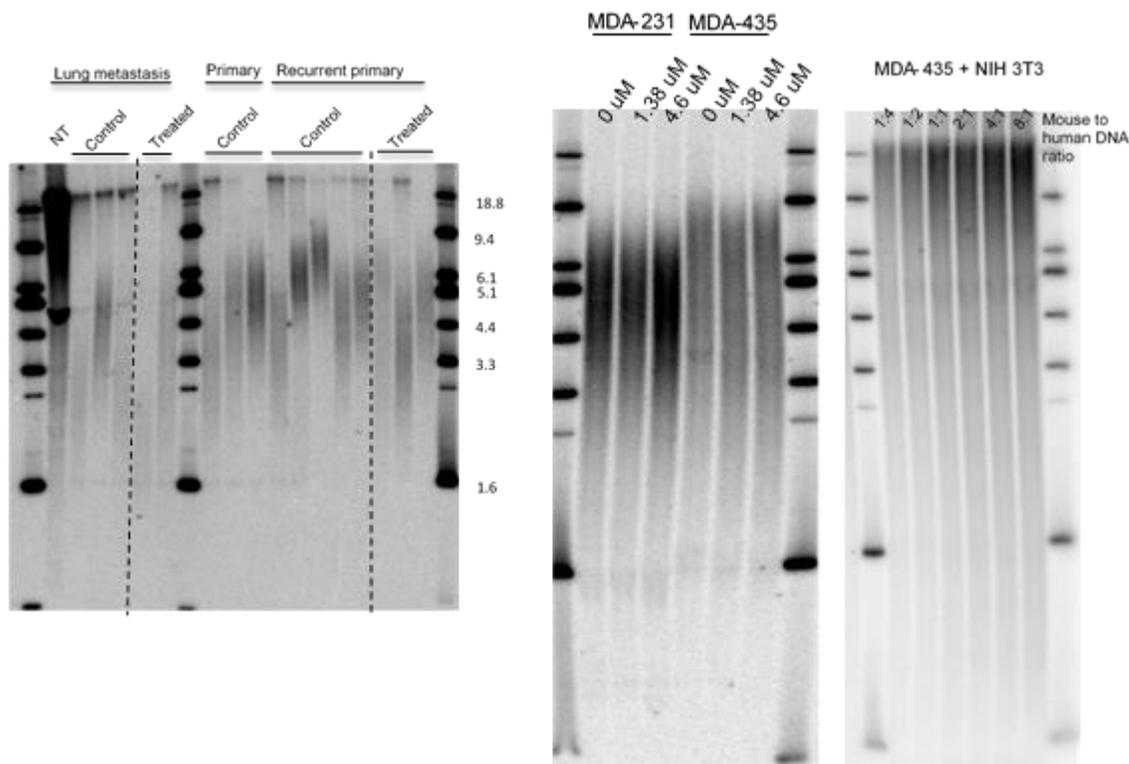
The failure of Perifosine to induce significant telomere shortening in this xenograft model may reflect some difference between a tissue culture dish and the mouse physiological environment, which is supported by the fact that the primary tumors had telomere lengths very substantially longer than the telomere lengths of subclones of the HCC38+GFP+LUC cell line, potentially ruling out founder effects as the source of this difference (Figure 3.7, primary TRFs in Appendix A).



**Figure 3.7: Primary tumors exhibit much longer telomere length than subclones of the parental HCC38 population**

Telomere length (TL) was measured in subclones of the HCC38+GFP+LUC population and the primary tumors after removal of mouse support cells. Primary tumors exhibited longer telomeres than parental subclones ( $p < 0.05$ ) in both groups, indicating that the founder effect is an unlikely source for the increased telomere length observed.

While Perifosine failed to induce significant telomere shortening in this xenograft model, this is still the first attempt to measure telomere length in a mouse tumor model using a clinical tested drug, while removing the effects of mouse support cells. Optimization work conducted prior to the xenograft model indicates that as little as 20% by mass contamination of mouse DNA is sufficient to produce a dramatically biased interpretation of telomere length (Figure 3.8). Other groups that have attempted to quantify telomere length from xenografts without removing contaminating mouse DNA clearly suffer from mouse DNA contamination (Barszczyk *et al.* 2014), as the band above 18.8 kilobases present in non-tumor bearing mouse lung telomeres (Figure 3.8) occurs in all of the samples measured in that study. Consideration of the effect of telomerase inhibitors on telomere length over time is an absolutely vital preclinical step, and this method allows that for the first time.



**Figure 3.8: Mouse DNA produces substantially biased measurements of telomere length in human samples**

The band present above 18.8 kilobases in the non-tumor bearing mouse lung (NT, left) indicates the presence of mouse DNA in a TRF (left). A mixture of NIH 3T3 DNA (a mouse cell line) and MDA-MB-435 DNA containing as little as 20% by mass mouse DNA produces a dramatically biased interpretation of telomere length (right, middle TRF presents MDA-MB-435 DNA *in vitro* with no mouse contamination).

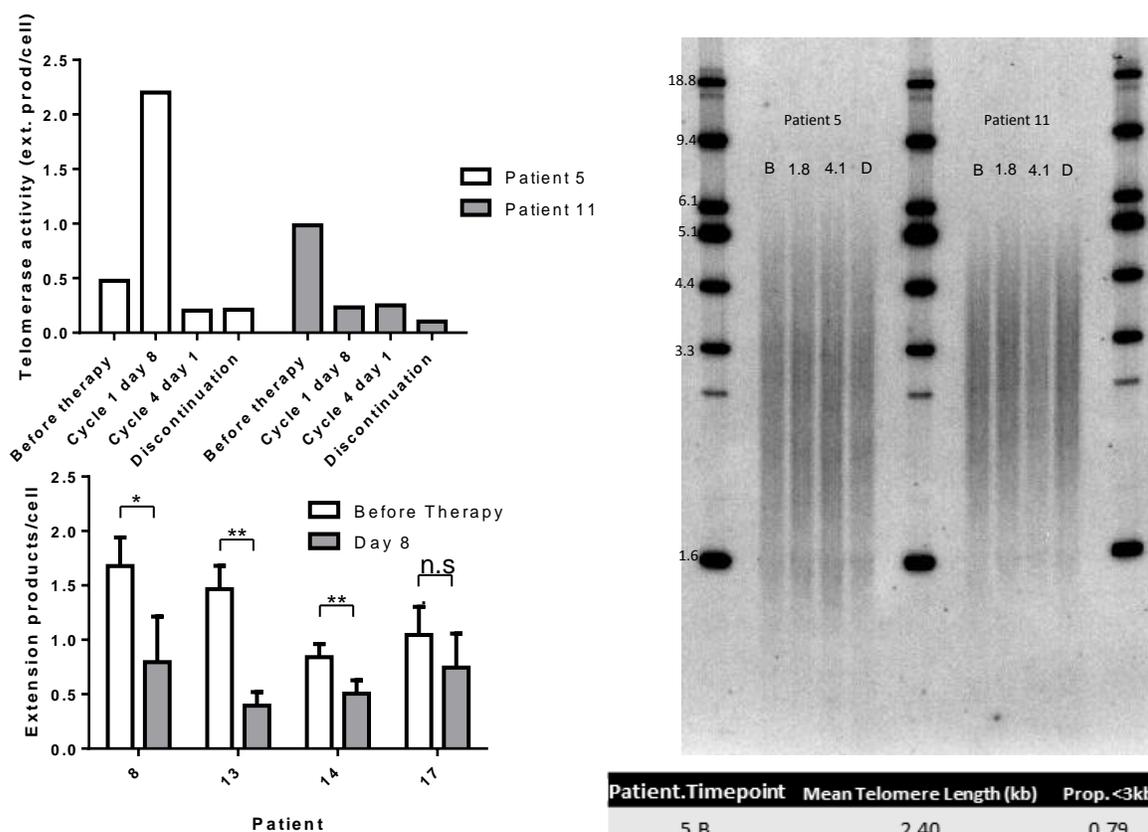
*Perifosine induces shortening of the shortest telomeres and reduces telomerase enzymatic activity in CLL samples from a phase II clinical trial*

While the xenograft experiment was underway, I contacted Dr. Daphne Friedman at Duke University about her recent clinical trial using Perifosine (Friedman *et al.* 2014).

Unlike most other studies evaluating Perifosine, this protocol used a continuous dosing regimen that would not allow tumor cells time to re-extend their telomeres between treatment cycles. She had purified tumor samples left over from the trial that she was willing

to send some of them to me to evaluate for changes in telomere biology. Samples were available for two patients from before the initiation of therapy, cycle 1 day 8 (day 8 of treatment), cycle 4 day 1 (day 84) and at discontinuation of therapy (day 124 and day 183 of treatment for the two patients). Another four patients had samples remaining from before the initiation of therapy and at cycle 1 day 8, which were evaluated for changes in telomerase enzymatic activity.

Mean telomere length did not change in the two patients treated for extended periods of time, though telomerase enzymatic activity was decreased at cycle 1 day 8 compared to before therapy in four of the six patients evaluated (Figure 3.9).



**Figure 3.9: Reduction in ddTRAP activity in Perifosine treated CLL without a change in mean TL**

Four of six patients exhibited a decrease in telomerase activity as measured by the droplet-digital TRAP assay (left) while there was no change in mean telomere length in long-term treated patients (right). B: Before therapy, 1.8: cycle 1, day 8, 4.1: cycle 4, day 1, D: Discontinuation of therapy

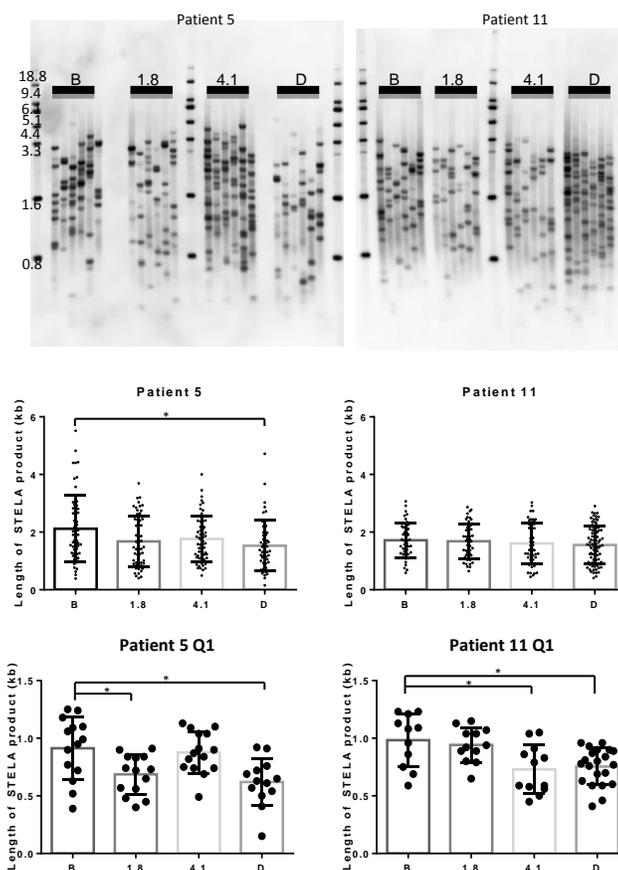
Patient.Timepoint	Mean Telomere Length (kb)	Prop. <3kb
5.B	2.40	0.79
5.1.8	2.40	0.79
5.4.1	2.38	0.80
5.D	2.44	0.77
11.B	2.38	0.80
11.1.8	2.44	0.77
11.4.1	2.41	0.77
11.D	2.45	0.76

These results indicate that it is possible that Perifosine could act as a telomerase inhibitor in human patients because of the change in telomerase enzymatic activity, but either the length of treatment was not long enough or tumor cells in some way compensated for the decrease in telomerase activity observed. CLL tumor cells are a population composed of varying proportions of proliferating and non-proliferating cells, and the factors that dictate the balance between dividing and non-dividing cells are not fully understood. In

most cancers, the bulk of the tumor population is both non-dividing and telomerase negative, though it is possible for non-dividing, telomerase negative cells to revert to a telomerase positive dividing state (Damle *et al.* 2007). The low telomerase activity observed in the population and lack of telomere shortening observed in the bulk tumor could result from a non-dividing, and thus non-shortening, bulk tumor population with a dividing and shortening cell pool too small to measure by TRF.

To test this hypothesis, I worked with Tsung-Po Lai, a postdoctoral researcher in the lab to perform Universal Single Telomere Length Analysis (Universal STELA), a PCR-based assay that can quantitate the length of individual telomeres. Because Universal STELA involves long-range PCR on a G-rich, repetitive substrate, using primer design that requires a subtelomeric end and a normal telomere overhang (not internal telomere sequence) the products of this reaction are biased toward the shortest telomeres. Further, since the length of the shortest telomeres is the most important factor in determining telomere-driven replicative senescence (Hemann *et al.* 2001), the analysis was focused on the shortest quartile of the Universal STELA products.

Patient 5 exhibited a statistically significant reduction in the length of the total STELA product population, while there was no change in the total population in patient 11. A statistically significant reduction in the length of the shortest quartile of Universal STELA products occurred in both patients, illustrating that the shortest telomeres were shortening in both patients treated with Perifosine continuously for long periods of time (Figure 3.10).

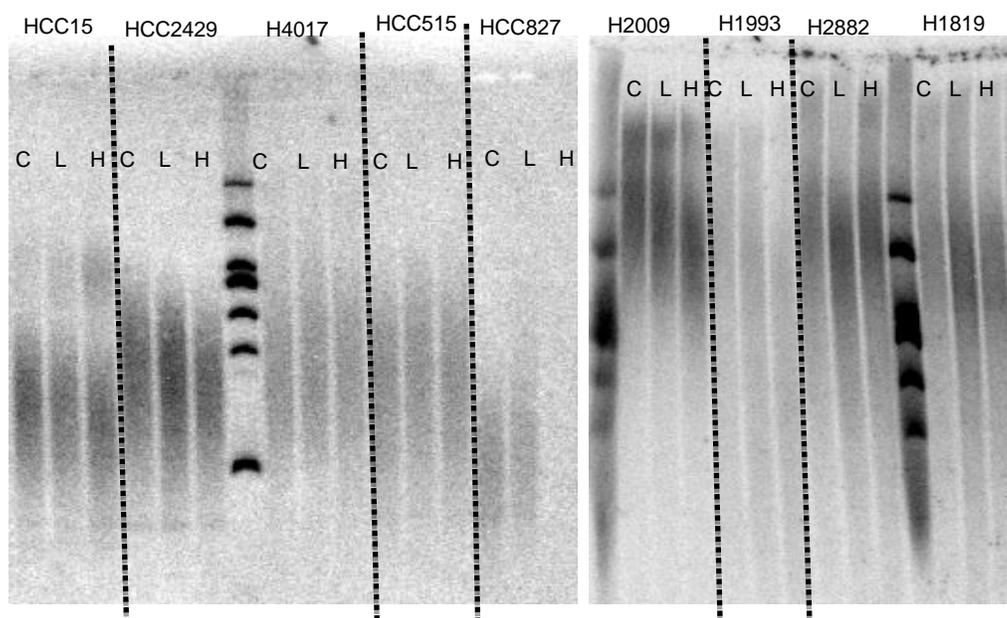


**Figure 3.10: Universal STELA reveals shortening of the shortest telomeres**

Universal STELA primary data is shown in the top panel, quantitated in the middle panel. Patient 15 had shorter Universal STELA products at discontinuation compared to before therapy, while both patients had shorter telomeres in the bottom quartile of Universal STELA products at discontinuation (bottom panel). B: Before therapy, 1.8: cycle 1, day 8, 4.1: cycle 4, day 1, D: discontinuation.

*Perifosine's effects on cell growth, telomerase activity and telomere length are uncorrelated*

In order to understand the capacity for heterogeneity in response to Perifosine, I treated a panel of lung cancer cell lines that had been rigorously evaluated by the Minna lab (Kim *et al.* 2013) with Perifosine with the intention of using the data available on these cell lines to inform mechanistic experiments about the link between Perifosine and telomerase. Each cell line was treated with two different doses of Perifosine and cultured for twenty population doublings in order to determine the effect of Perifosine on that cell line's telomeres over time (Figure 3.11, all primary TRFs in appendix A). A summary of the effect of Perifosine on telomere length and telomerase activity is shown in Table 3.1.



**Figure 3.11: Telomere lengths in cells treated with Perifosine**

A panel of cell lines was treated with Perifosine for 20 population doublings and their telomeres were measured by TRF. C: Control, L: Low dose, 1.84 uM Perifosine, H: High dose, 4.6 uM Perifosine.

The response to Perifosine was highly heterogeneous, with 12 of 20 cell lines evaluated exhibiting telomere shortening, and two cell lines lengthened their telomeres upon Perifosine treatment.

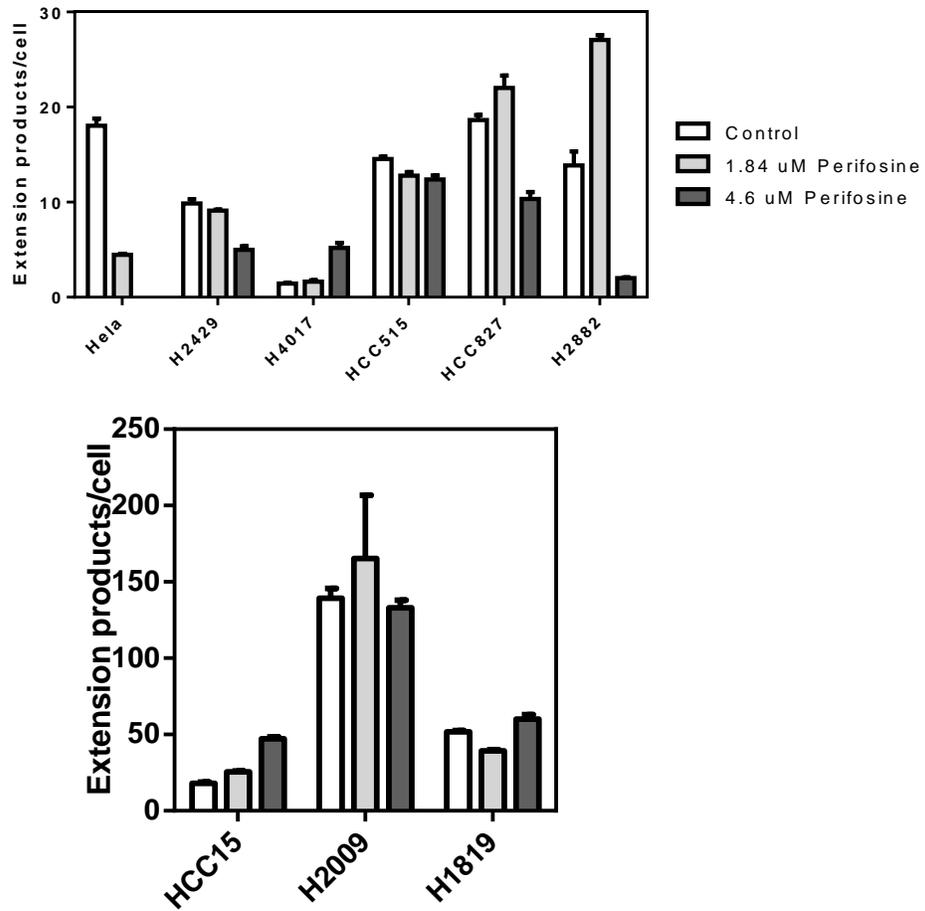
Cell Line	Telomere Length	Telomerase activity
HCC4017	Unchanged	Increased
H2087	Lengthening	Increased
HCC827	Unchanged	Decreased
H2009	Shortening	Unchanged
H2882	Shortening	Decreased
HCC15	Shortening	Increased
H1819	Shortening	Unchanged
HCC515	Unchanged	Unchanged
HCC2429	Shortening	Decreased
Hela	Shortening	Decreased
HT1080	Lengthening	N/A
H1993	Shortening	N/A
H2073	Shortening	N/A
HCC95	Unchanged	N/A
MDA-MB-231	Shortening	N/A
MDA-MB-435	Shortening	N/A
DLD1	Shortening	N/A
A375	Unchanged	N/A
HCC38	Shortening	N/A
H1299	Unchanged	N/A

**Table 3.1: Summary of responses to Perifosine**

12 of 20 total cell lines assayed decreased mean telomere length upon perifosine treatment (middle column), while four of ten cell lines decreased telomerase activity when treated with 4.6uM Perifosine (left column).

Telomerase activity was measured in a subset of the cell line panel (Figure 3.12), and four of the ten lines evaluated had a reduction in telomerase activity upon Perifosine treatment, though one of the cell lines with decreased telomerase activity did not exhibit telomere shortening (HCC827). Furthermore, the disconnect between changes in telomerase activity and what would be predicted from the change in telomere length (increased activity in HCC15, for instance, which shortened its telomeres) suggests that in

some cell lines the change in telomerase activity is not driving the change in telomere length.



**Figure 3.12: Telomerase activity in Perifosine-treated cell lines**

Cells were treated with the indicated concentrations of Perifosine for twenty population doublings. Changes in telomerase enzymatic activity (y-axis, telomerase extension products per cell) did not predict changes in telomere length.

Growth rates for a subset of the panel were tracked during a 41 day Perifosine treatment during which cells were split 1:32 and counted when they approached 90% confluency in order to determine if toxicity or cytostasis as a result of Perifosine treatment was correlated with changes in telomere length or telomerase activity (Table 3.2).

Cell Line	Population doublings/Day			Telomere Length	Telomerase activity
	Control	1.84 $\mu$ M	4.6 $\mu$ M		
HCC4017	0.52	0.46	0.42	Static	Increased
H2087	0.77	0.54	0.42	Lengthening	Increased
HCC827	0.49	0.45	0.21	Static	Decreased
H2009	0.73	0.72	0.72	Shortening	Unchanged
HT1080	0.61	0.55	0.53	Lengthening	Not assayed
H2882	0.38	0.38	0.38	Shortening	Decreased
HCC15	0.67	0.67	0.55	Shortening	Increased
H1993	0.53	0.51	0.37	Shortening	Not assayed
H1819	0.26	0.26	0.26	Shortening	Unchanged
HCC515	0.38	0.38	0.37	Static	Unchanged
HCC2429	0.74	0.71	0.61	Shortening	Increased
H2073	0.26	0.23	0.19	Shortening	Not assayed
HCC95	0.24	0.24	0.24	Static	Not assayed

**Table 3.2: Growth rate, telomere dynamics and changes in telomerase activity in the lung cancer cell lines evaluated**

Growth rate in population doublings per day is shown for each dose of Perifosine used, color coded by growth rate (red = slower, blue = faster). Changes to telomere length and telomerase activity in each cell line are shown in the right two panels; change in growth rate was not predictive of change in telomere length or telomerase activity.

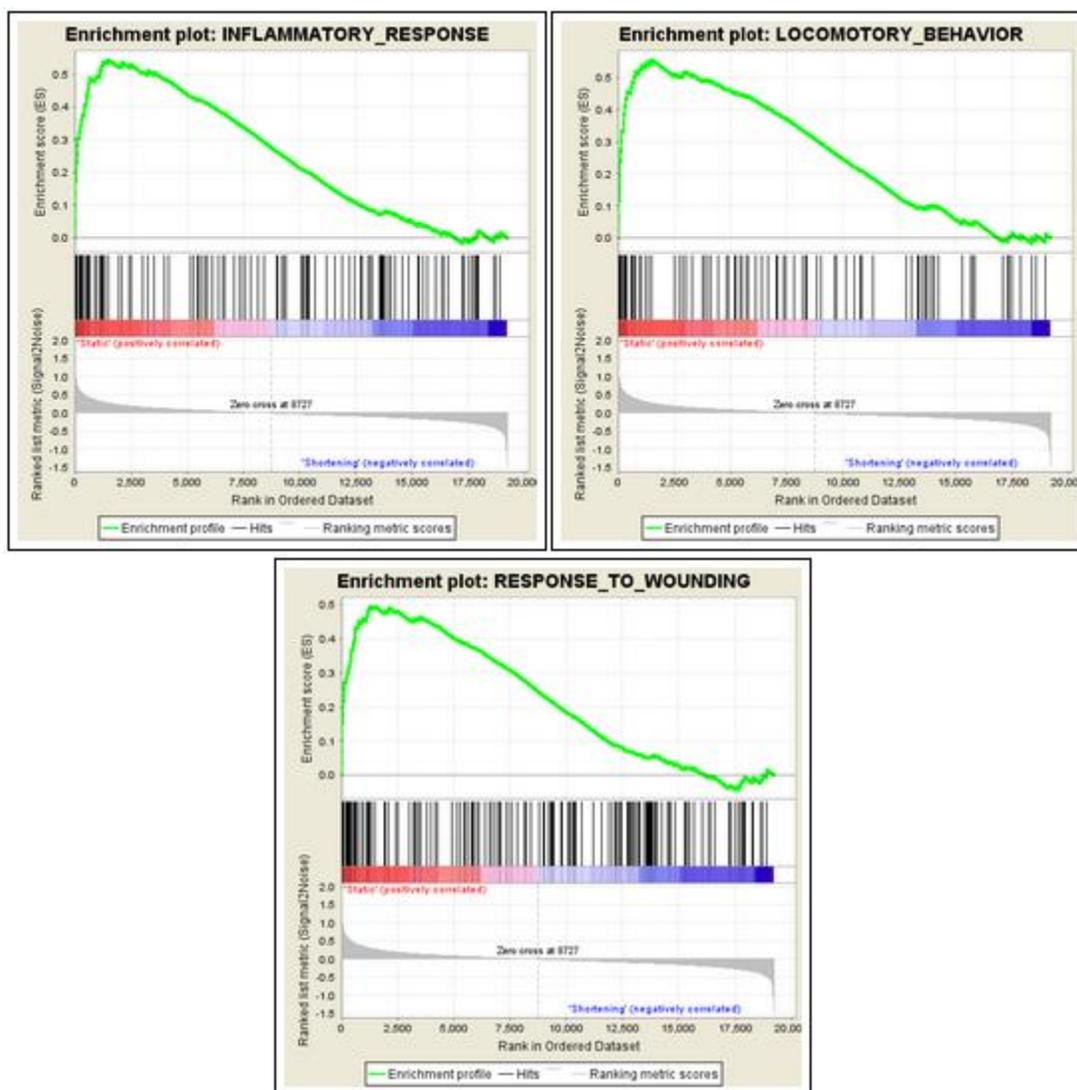
Cell lines that exhibited telomere shortening in response to Perifosine exhibited a range of changes in their growth rate, from H2882 and H2009 that did not slow their growth rate in response to Perifosine to H1993, which showed a reduction in growth rate when treated.

These experiments demonstrate that there is a high degree of heterogeneity in response to Perifosine both between cell lines and within cell lines with regard to different

phenotypes, and changes in telomere length are not correlated with changes in telomerase activity or changes in growth rate. This suggests that the three different phenotypes measured are driven by at least partially non-overlapping molecular networks that respond differently to Perifosine depending on the molecular state of the cell, for example gene expression or mutation status.

*Expression data indicates that stress response networks may alter Perifosine sensitivity*

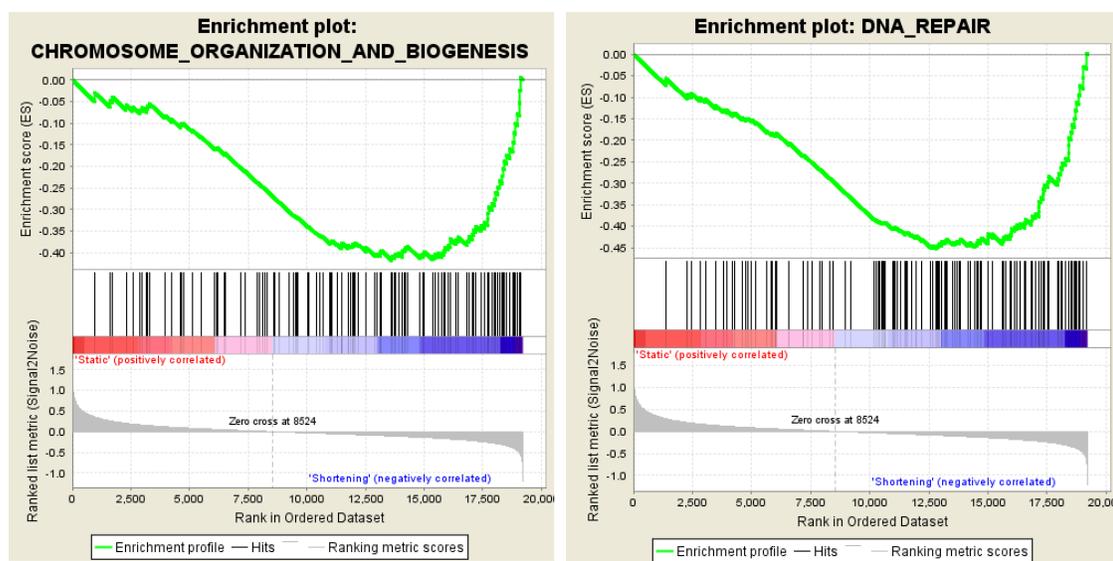
Gene-set enrichment analysis using expression data on these cell lines obtained by the Minna lab revealed differences in expression between the most unambiguously responding and non-responding cell lines (in terms of telomere length). Cell lines that were resistant to Perifosine's effect on telomeres had greater levels of expression of genes involved in inflammatory responses, locomotion, and response to wounding (Figure 3.13).



**Figure 3.13: Gene sets enriched in cell lines insensitive to Perifosine-induced telomere shortening**

Expression levels of genes involved in these networks indicated that on average the cell lines insensitive to Perifosine had higher levels of genes involved in the inflammatory response, response to wounding and locomotory behavior. Enrichment score (a function of relative expression in static vs. shortening) indicates the odds of observing a difference in this network by chance.

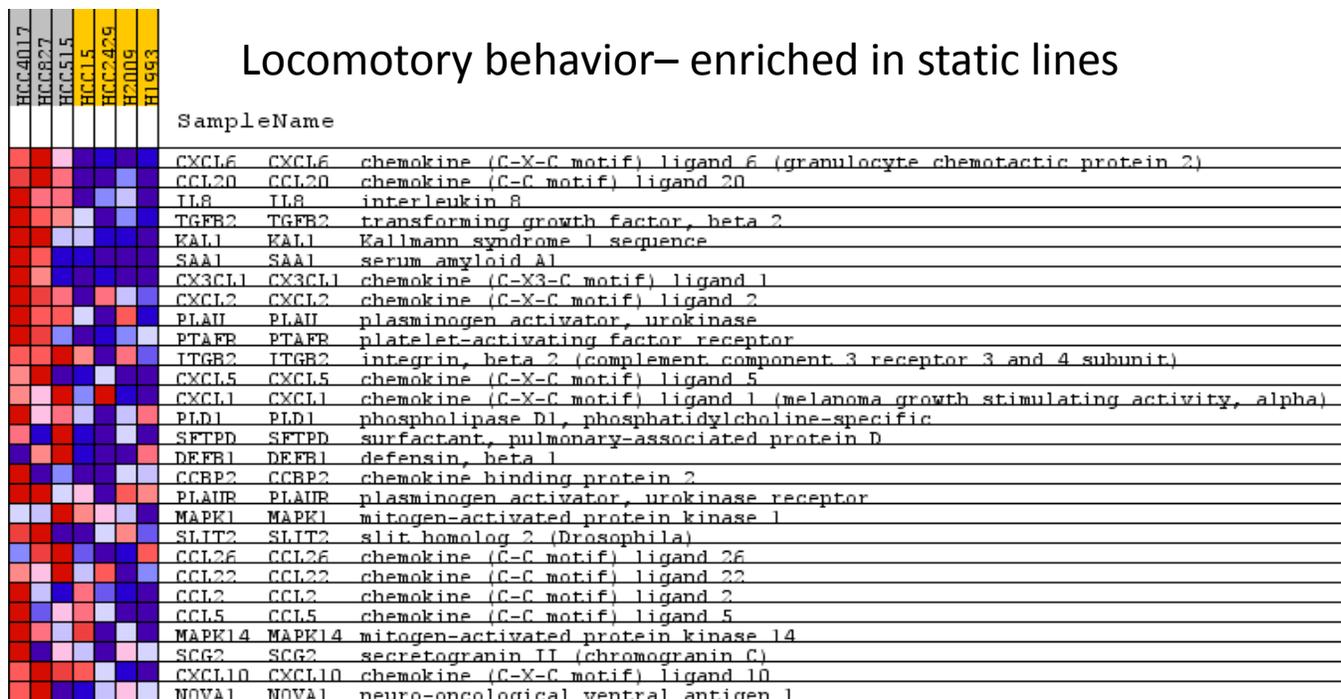
Further, cell lines sensitive to Perifosine had greater expression of genes involved in chromosome organization and DNA repair (Figure 3.14).



**Figure 3.14: Gene sets enriched in cell lines sensitive to Perifosine-induced telomere shortening**

The cell lines that were sensitive to Perifosine had greater expression of genes involved in these two networks. Enrichment score (top) is a function of the odds of observing this bias by chance.

Investigation of the genes specifically overexpressed within these networks strongly implicated the molecular response to stress and the p38 pathway in particular as upregulated in the cell lines that were insensitive to Perifosine's effect on telomeres (static lines). For example, the locomotory behavior category (Figure 3.15) included upregulation of IL8, a number of chemokines associated with aggressive migration, MAPK1, MAPK14, TGFB2, PLA1 (Plasminogen activator, Urokinase) and PLAUR (the receptor for PLA1).

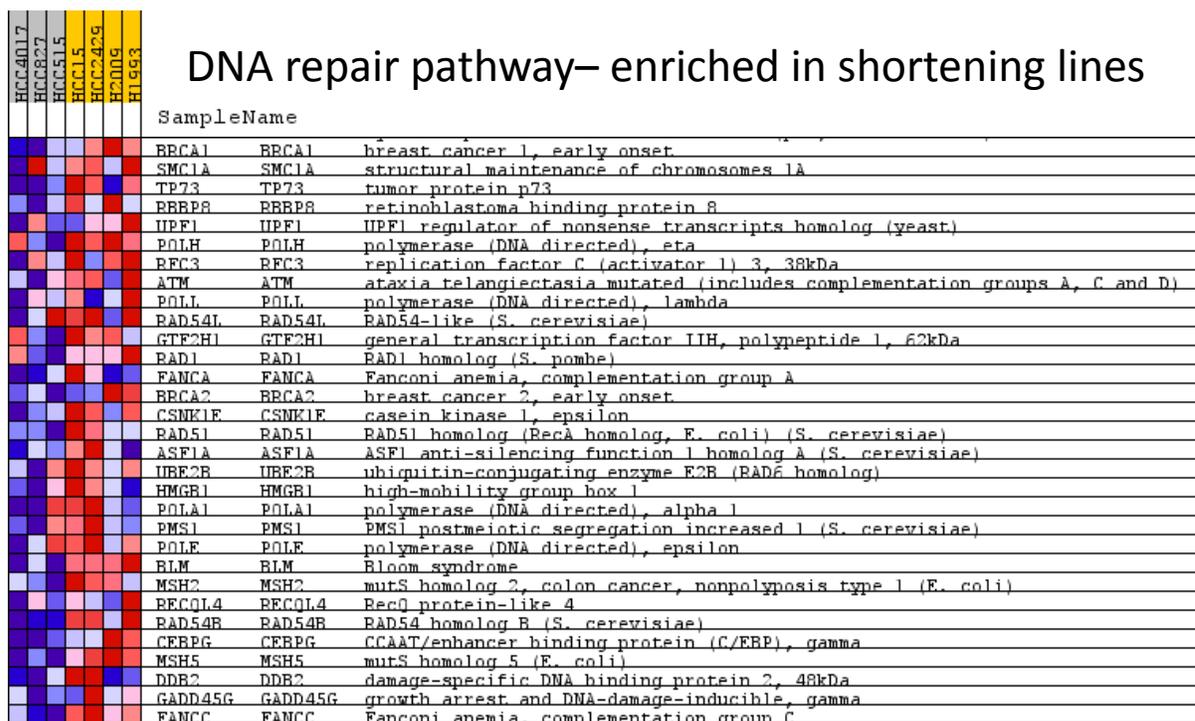


**Figure 3.15: Expression of the genes most enriched in the locomotory behavior category in the static lines**

Expression (normalized to mean expression in that cell line) is indicated by color (red = more expression, blue = less expression) for each gene indicated in the static lines (grey, HCC4017, HCC827, HCC515) compared with the shortening lines (HCC15, HCC2429, H2009, H1993). Genes are ranked by enrichment in the static lines.

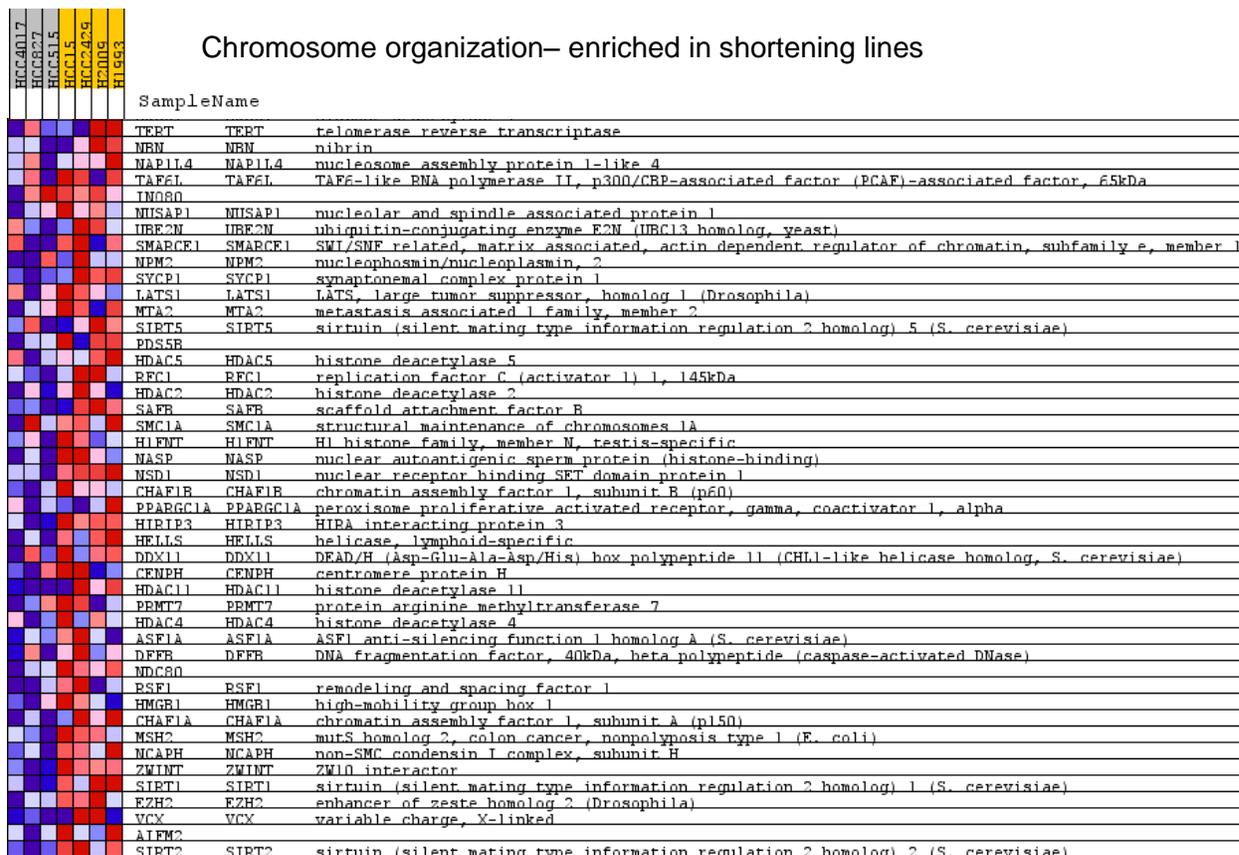
Further, the DNA repair pathways upregulated in the cell lines sensitive to Perifosine's effects on telomeres (shortening lines), shown in Figure 3.16, include a number of hits from the candidate gene approach discussed in Chapter 2, including POLH, POLL and BLM, as well as FANCC and FANCA, which form a complex that regulates the activation of FANCD2, a hit examined in Chapter 2 (Kupfer *et al.* 1999; D'Andrea & Grompe 2003). This network also includes ATM, BRCA1 and BRCA2 with substantially increased expression in sensitive lines compared with static lines.

The chromosomal organization group included a number of interesting genes that were enriched in the sensitive lines, such as SIRT1, SIRT2, SIRT5, a number of other histone deacetylases, helicases and methyltransferases (Figure 3.17).



**Figure 3.16: Expression of the genes most enriched in DNA repair pathways in the shortening lines**

Expression (normalized to mean expression in that cell line) is indicated by color (red = more expression, blue = less expression) for each gene indicated in the static lines (grey, HCC4017, HCC827, HCC515) compared with the shortening lines (HCC15, HCC2429, H2009, H1993). The genes most enriched in shortening lines are shown on the bottom.

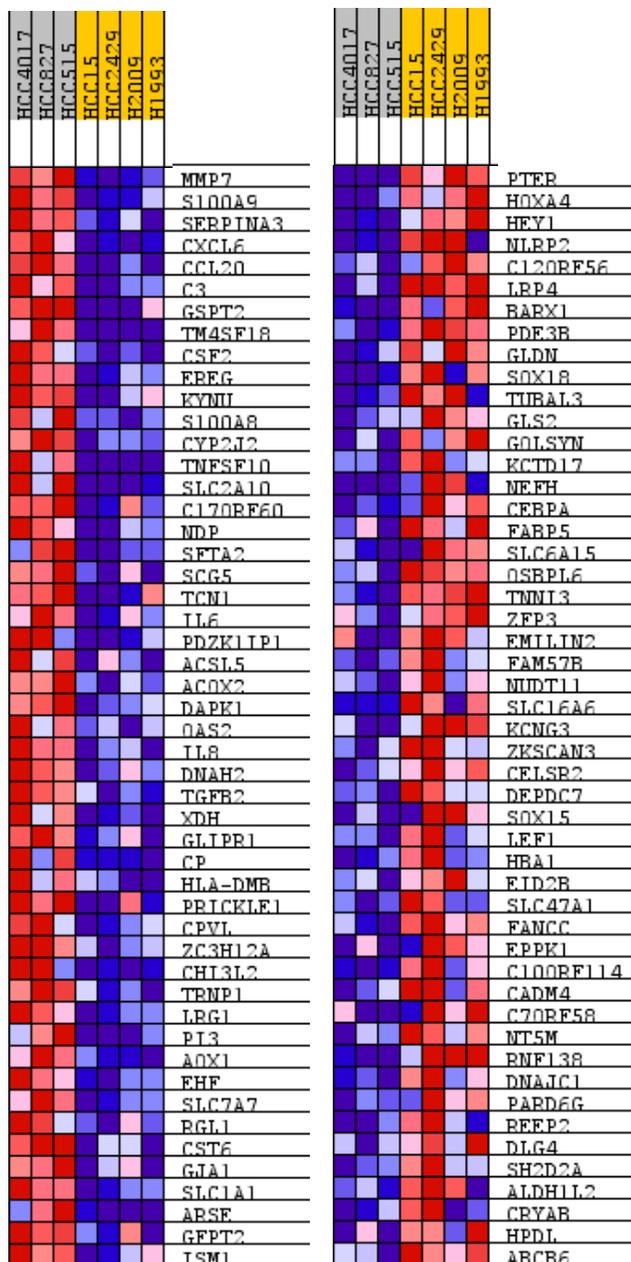


**Figure 3.17: Expression of the genes most enriched in the chromosome organization network in shortening cell lines**

Expression (normalized to mean expression in that cell line) is indicated by color (red = more expression, blue = less expression) for each gene indicated in the static lines (grey, HCC4017, HCC827, HCC515) compared with the shortening lines (HCC15, HCC2429, H2009, H1993). The genes most enriched in shortening lines are shown on the bottom.

In addition to the gene sets enriched in the shortening and static lines, a number of genes were very highly enriched in the two groups of cell lines that were not part of the enriched gene sets (Figure 3.18). Notably, another cytokine important for inflammation and the stress response, IL6, was highly upregulated in the static lines compared with the shortening lines. Further, the gene most highly upregulated in the shortening lines compared to the static lines, ABCB6, is a mitochondrially localized ATP-Binding Cassette protein, and dysfunction of this protein *in vivo* has been linked to increased CXCL5,

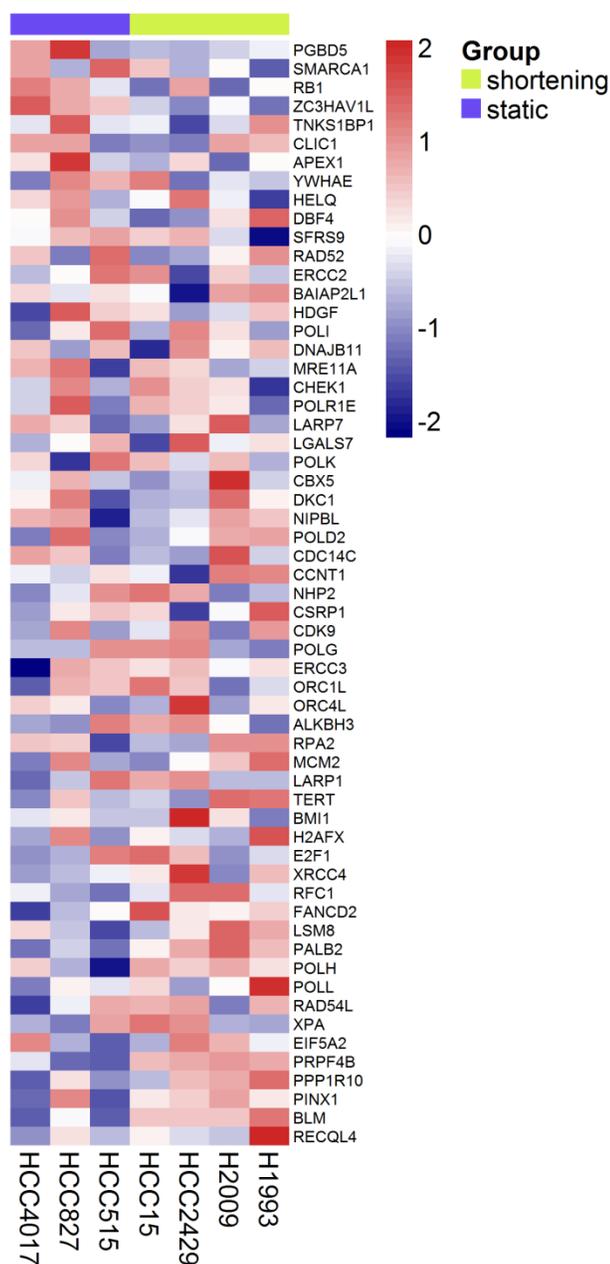
proliferation of megakaryocyte progenitors, and accelerated atherosclerosis (Murphy *et al.* 2014). CXCL5 is upregulated in the static lines (Figure 3.15, 12<sup>th</sup> from the top), consistent with this regulation.



**Figure 3.18: The 50 genes most enriched in both groups of cell lines**

Expression is indicated by color (red >) for each gene indicated in the static lines (grey, HCC4017, HCC827, HCC515) compared with the shortening lines (HCC15, HCC2429, H2009, H1993). The 50 genes most enriched in the static lines are on the left, the 50 genes most enriched in the shortening lines are on the right.

Lastly, a number of the hits from Chapter 2 were differentially expressed in the sensitive lines compared with the static lines (Figure 3.19). PGBD5, SMARCA1, RB1 and ZC3HAV1L were more expressed by the static lines compared to the shortening lines, while XRCC4, RFC1, FANCD2, LSM8, PALB2 (FANCN), POLH, POLL, RAD54L, XPA, EIF5A2, PRPF4B, PPP1R10, PINX1, BLM and RECQ4L were all more expressed in the shortening lines. Interestingly, LARP1, LARP7 and LGALS7 showed no strong enrichment in either category, nor did TERT or the other telomerase subunits, DKC1 and NHP2. These data in aggregate can be interpreted to suggest that the p38 pathway and the inflammatory response are more highly activated in cell lines that are resistant to the telomeric effects of Perifosine, while an intact DNA damage response and chromatin state regulation network confers sensitivity to Perifosine.

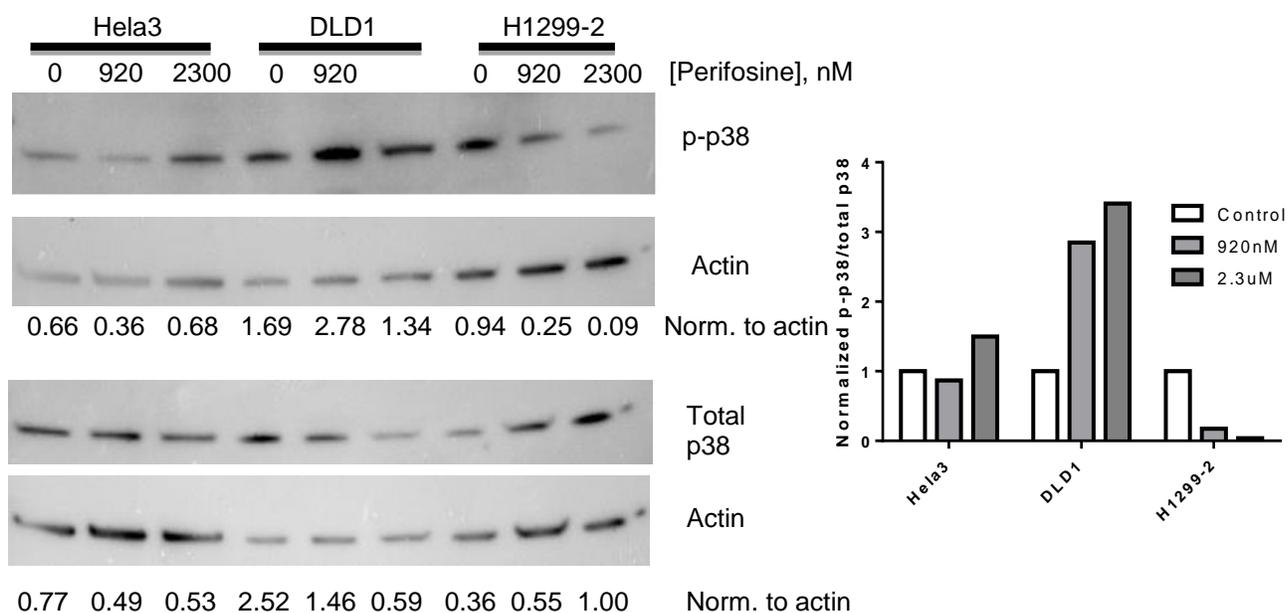


**Figure 3.19: Expression of the candidate positive regulators of telomerase in Perifosine-sensitive and resistant cells**

Relative expression in each cell line is indicated by color (red = increased, blue = decreased). Genes involved in DNA repair pathways are more enriched in shortening lines, while the genes enriched in static cells are predominantly of unknown function. Genes are ranked by difference in expression between static and shortening lines (most enriched in static at the top, most enriched in shortening at the bottom).

*Perifosine activates p38 activity in some but not all of the sensitive cell lines*

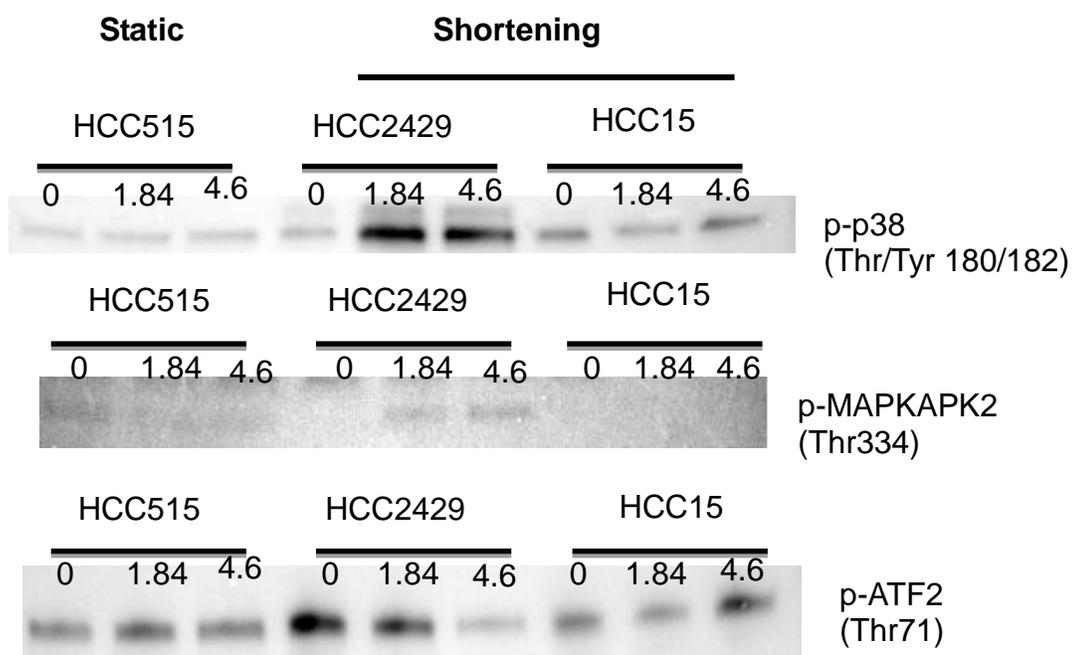
The results indicating inflammatory pathways and the p38 pathway in particular were not completely surprising, given that in addition to the AKT pathway, Perifosine is implicated in p38 activity modulation; some reports indicate that Perifosine activates p38 (Tzarum *et al.* 2012), while others claim that it inhibits p38 (Li *et al.* 2006; Wang *et al.* 2012). To test if differential activation of the p38 pathway explained the difference in sensitivity to Perifosine, I measured phospho-p38 in cell lines continuously treated with Perifosine for 20 population doublings. I evaluated four sensitive (Hela3, DLD1, HCC2429, HCC15) and two static lines (H1299-2 and HCC515) (Figure 3.20, 3.21).



**Figure 3.20: p38 activity in three chronically Perifosine-treated cell lines**

Phospho-p38 (Thr/Tyr 180/182) and actin are shown on top left, total p38 and loading control are on the bottom left. Signal normalized to actin is quantitated below the blot and normalized phospho-p38:total p38 is quantitated at the right.

This experiment indicated that in some cell lines that were sensitive to Perifosine-induced telomere shortening, p38 was activated by Perifosine. Neither static line had an increase in p38 activation with Perifosine treatment, and H1299-2 exhibited inhibition of p-p38. The heterogeneity in response to Perifosine in both p-p38 and total p38 may explain why the literature offers so many conflicting observations; each report on Perifosine's effect on p38 was only in the context of a single background, and if these experiments had been conducted in one cell line they would offer a similarly conflicted view.



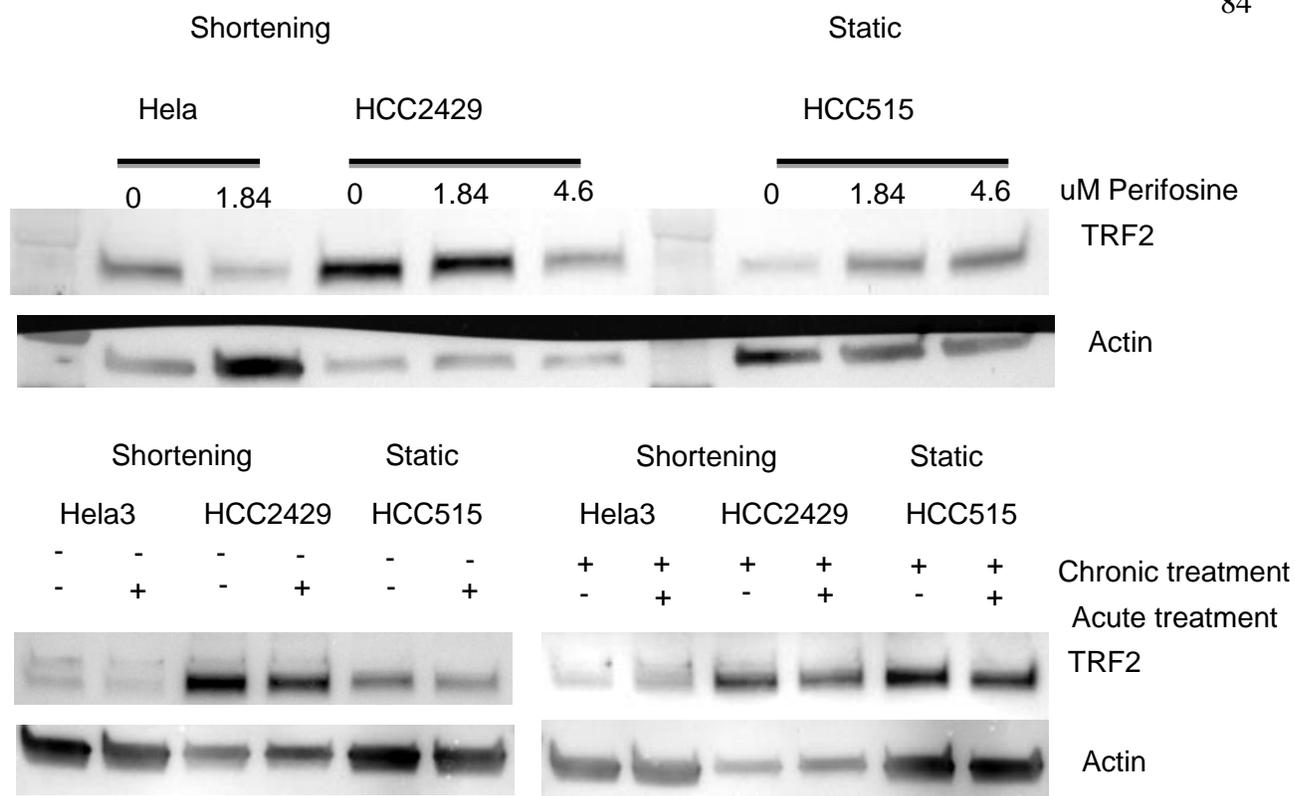
**Figure 3.21: p38 activity in three additional chronically Perifosine-treated cell lines**

P-p38 activity (top) increased in one of the two shortening lines measured here, while it did not increase in the static line. P-MAPKAPK2 (a downstream target of p38) increased in HCC2429 and may have decreased in HCC515. P-ATF2, quantitated at the bottom, behaved differently in all three cell lines.

These experiments indicate that activation of p38 may explain why some of the shortening lines are sensitive to the telomeric effects of Perifosine, but they also show that this is not a universal mechanism. HCC15, the shortening line that does not activate p38 in response to Perifosine also most resembled the static lines in terms of MAPK1 and MAPK14 expression (Figure 3.15), suggesting a difference in the baseline state of the MAPK network as a mechanism for the difference in p38 activation between it and the other three shortening lines evaluated.

*Perifosine treatment alters TRF2 expression in some cell lines*

P38 activity has been linked to changes in expression of TRF2, a Shelterin component (Ludlow *et al.* 2012), and changes in the expression and localization of the Shelterin complex can induce a DNA damage response that would explain why the shortening cell lines had higher expression of DNA damage response proteins. To determine if Perifosine treatment and concomitant p38 activity resulted in a decrease in TRF2 levels, I evaluated TRF2 by western blot in three chronically treated cell lines, Hela3, HCC2429 and HCC515 (Figure 3.22).



### Figure 3.22: TRF2 levels in Perifosine-treated cells

Top, TRF2 and Actin levels in cells treated with the indicated concentrations of Perifosine. Bottom, TRF2 and Actin levels in cells either naïve to Perifosine treatment (- in the chronic treatment) or continuously treated with 1.84uM Perifosine for 20 population doublings (+ chronic treatment) either with the addition of Perifosine for 3 days (+ acute treatment) or cultured without Perifosine for 3 days (- acute treatment).

Long-term Perifosine treatment led to a reduction in TRF2 levels in both cell lines that exhibit telomere shortening in response to Perifosine (Hela3 and HCC2429), while an increase was observed in the HCC515 static line treated with Perifosine. The decreases in TRF2 levels in the shortening line are consistent with the model of p38-mediated reduction in TRF2 levels driving telomere shortening, however the increase observed in the static line indicates that other processes are at work.

I also conducted an experiment to test if the observed changes in TRF2 were the cause or one of the effects of a change in telomere length. Cells that had not been

previously treated with Perifosine were treated for three days, and cells that had been continuously treated with Perifosine were released from treatment for three days, followed by measurement of TRF2 protein levels in chronically treated cells treated during the experiment (+ chronic, + acute), chronically treated cells not treated during the experiment (+ chronic, - acute), naïve cells treated during the experiment (- chronic, + acute), and naïve cells not treated (- chronic, - acute). TRF2 levels decreased in all three cell lines treated transiently with Perifosine including the HCC515 static cell line, while release from chronic treatment resulted in increased TRF2 levels in all three cell lines as well. These experiments suggest that changes in TRF2 protein levels in response to chronic Perifosine are not established immediately following exposure, and therefore they are likely impacted by the changes to telomere length occurring over chronic treatment.

## **DISCUSSION**

Of the four AKT inhibitors evaluated, only two (Perifosine and AKT inhibitor IV) induced telomere shortening in HeLa3 cells. Though subsequent analysis indicated that the p38-mediated effects of Perifosine are more likely to be important to its function in altering telomere length (GSEA indicated inflammatory and stress response pathways rather than growth pathways dictated response to Perifosine), AKTi IV and Perifosine may share some mechanistic characteristics. For example, AKTi IV can block the activity of CCL2 in sensitizing TRPV1 sodium channels to capsaicin treatment (Kao *et al.* 2012), and CCL2 was upregulated in cell lines resistant to Perifosine. Pretreatment with AKTi IV can also block the release of CCL2 from vascular smooth muscle cells in response to NAD(+)/NADP(+) (Kim *et al.* 2011), suggesting that it may act both upstream and downstream of elements

that appear relevant to the response to Perifosine. Further, AKTi IV can be detected in mitochondria, and AKTi IV treatment leads to defects in respiration and increases in ROS levels in HeLa cells (Meinig & Peterson 2015), which could indicate that AKTi IV treatment would induce a DNA damage response similar to what the GSEA indicates occurs in cell lines sensitive to Perifosine-mediated telomere shortening. Another study found that AKTi IV can inhibit phosphorylation of ERK2 in response to IL-1B in human fibroblast-like synoviocytes (Tsuji *et al.* 2012), which would indicate that AKTi IV effects the p38 pathway, and the similarity between its effect on telomeres and Perifosine's may be due to partially overlapping responses to the drugs.

Though Perifosine did not accelerate telomere shortening in BJ fibroblasts (Figure 3.3), it may alter telomere length in other normal cells, particularly because an intact DNA damage response pathway is important to the telomeric response to the drug (Figure 3.14). While it is possible that DNA repair pathways in these cells interfere with the localization and activity of telomerase in cells treated with Perifosine, it is also possible that the fibroblasts assayed have some as-yet undefined molecular property that confers resistance to the telomeric effects of Perifosine like that observed in many of the cancer lines measured. To clarify this, observation of Perifosine's effect on telomere length in other telomerase negative cells would be required before a definitive conclusion about the telomerase-dependence of this effect is reached. Importantly, the cell lines that shortened their telomere length in response to Perifosine did not have lower telomerase activity or expression than cell lines resistant to Perifosine, such that high levels of telomerase do not appear to confer either resistance or sensitivity to Perifosine's effects on telomere length.

Unexpectedly, Perifosine did not alter either metastatic tumor burden or telomere length in the xenograft model. Though there was a difference between primary, recurrent primary and metastatic tumor telomere length within the treated group in the xenograft model, there was not a significant difference in mean telomere length between control tumors and treated tumors, indicating that Perifosine was not sufficient to induce dramatic telomere shortening in this xenograft model. It is possible that the mouse microenvironment included some stimulus that antagonized the function of Perifosine, such as estrogen or normoxia, both of which have been associated with increased telomerase activity (Guan *et al.* 2012; Zhou *et al.* 2013a). It is also possible that the observed small differences in telomere length between tumor sites are the result of founder effects from the parent population, though this seems unlikely because telomere length in the primary tumors was longer than subclones of the parental HCC38+GFP+Luc cell line *in vitro* (Figure 3.7).

One worrying aspect of these experiments is that the cell line selected, HCC38, was used because it was the cell line most likely to yield positive results of those lines compatible with the model. HCC38 cells have both very short telomeres and a robust, rapid response to Perifosine in terms of telomere length; it represents a best-case scenario for the use of Perifosine clinically. However, even in this context the drug did not significantly alter the important outcome of the experiment (metastatic tumor burden), casting doubt on the applicability of this treatment toward human health. It is possible that longer treatment with Perifosine or the inclusion of more mice in the experiment would generate a statistically significant effect rather than the trend observed, but the near-equal telomere lengths observed in the tumors measured here suggests that some difference between the mouse microenvironment and tissue culture results in a context-dependent response to the drug.

Fifteen mice each in the control and treated group began the xenograft experiment; however mice in each group had a high mortality rate as a result of complications such as post-surgical infections, hemorrhage during primary tumor removal and death from unclear causes particularly in the treated mice that was most common in a number of mice in the treated group that displayed an emaciated, cachexia-like appearance 1-2 days before death (data not shown). The cachexic appearance and mortality events occurred in weeks 8 and 9 (6 and 7 weeks of treatment, 2-3 weeks following primary tumor removal), but the experimental design did not include measurements of body mass or fat integrity because cachexia had not been reported in xenograft models of Perifosine treatment before. Prior to the xenograft experiment, calculations of group size indicated that with 15 mice in each group a change in telomere length of 200 base pairs (what I considered a lower bound for an effect considering what was observed *in vitro*) with a standard deviation of 200 base pairs (a high estimate of SD from what is observed in repeated analysis of TRF samples) would be detectable at an alpha of 0.05, however I did not adequately account for the mortality rate observed. If another xenograft test of Perifosine were conducted, an increased number of mice in both groups, particularly the treatment group to account for this higher mortality rate and monitoring of phenotypes relevant to fat integrity and cachexia may yield more information and more significant results.

Somatic telomerase activity in mouse cells coupled with the fact that laboratory mouse telomeres are an order of magnitude larger than human telomeres (Chadeneau *et al.* 1995) have complicated pre-clinical evaluation of telomerase inhibitors in the past. In these experiments, I mostly circumvented the problem of mouse cell contamination of tumor samples through brief tissue culture and the induction of mouse cell death with puromycin treatment since the human tumor cell line was puromycin resistant. Since telomere length

predominantly changes during cell division, short periods of time in tissue culture are unlikely to alter telomere length, though small changes as a result of 1-4 population doublings in tissue culture may have occurred. A recent study (Barszczyk *et al.* 2014) measured telomere length in xenografts treated with Imetelstat without removing mouse cells, and it is possible that their telomere length data could have been biased by mouse telomere contamination. Indeed, the band present above the 18.8 kilobase ladder band in the non-tumor-bearing mouse lung tissue sample in Figure 3.8 appears in all xenograft samples in that study, and optimization experiments conducted prior to this work indicate that the presence of 20% by mass mouse contamination of a human DNA sample is sufficient to produce a dramatically biased measurement of telomere length (Figure 3.8). Improved methods to distinguish between murine and human telomeres in xenograft samples such as that presented here should improve understanding of telomere biology in physiological contexts in the future. The band indicative for mouse DNA contamination was present in a number of the tumor samples examined here as well, which opens the possibility that the results of this xenograft were also biased because of the presence of mouse DNA, however the extent of this contamination is far lower following puromycin treatment in this experiment than observed in other studies of xenograft tissue, and some of the samples examined did not have a very significant band above 18.8 kb. Knowing the origin of this signal, I excluded all signal above the 18.8 kb ladder band from quantitation in all samples though mouse contamination could still result in biased-high telomere lengths because of lower molecular weight products contaminating the rest of the sample.

Evaluation of the CLL samples' telomere length is the first experiment with a putative telomerase inhibitor, Perifosine, used in the clinic to evaluate changes in telomere length over the course of treatment. Telomere length is a vital biomarker of the efficacy of

treatment with a telomerase inhibitor, since the aim of anti-telomerase therapy is to impose a replicative limit on tumors through progressive telomere shortening. Changes in mean telomere length were modest, 300-400bp reduction in length of the shortest quartile, in response to Perifosine. Treatment beginning earlier in the course of the disease and continuing for longer periods of time would be predicted to cause further reduction in telomere length, which is viable given Perifosine's generally mild toxicity profile at the doses used. Shortening of the shortest telomeres is the most important indicator of a replicative limit (Hemann *et al.* 2001), and the reduction in length in the shortest quartile of telomeres observed in these patients is encouraging. It is possible that Perifosine provided in a setting of minimal residual disease may prevent relapses in specific subsets of patients by imposing a limit on replication.

The reduction in telomerase enzymatic activity demonstrated in purified tumor samples from Perifosine-treated CLL patients, coupled with the observation of a reduction in the length of the shortest telomeres indicates that Perifosine may be a viable anti-telomerase therapy. Interestingly, mild thrombocytopenia and leukopenia were observed in the clinical trials of Perifosine, toxicities it shares with Imetelstat (Thompson *et al.* 2013), as well as genetic disorders of impaired telomere maintenance such as dyskeratosis congenita (Holoan *et al.* 2014). Therefore this overlap may be due to "on target" effects of the drugs via dysregulated hematopoietic stem cell maintenance via telomerase inhibition, rather than specific idiosyncrasies of the drugs themselves. However, the most important question regarding the clinical tractability of anti-telomerase therapy remains to be answered; would a hypothetical perfect anti-telomerase therapy impose a replicative limit on most tumors in time to impact the course of the disease? That answer will in part depend on the initial telomere length in the tumor in question, but it is important to note that neither of the two

patients treated for very long periods of time with Perifosine continuously exhibited a complete response (one had a partial response and one had stable disease), suggesting that the shortening observed here did not hugely impact the course of their disease.

Inhibition of telomerase activity occurred in some lines that exhibited telomere shortening in response to Perifosine, though a change in telomerase activity did not universally predict an alteration in telomere length. For example, both H2009 and H1819 exhibited telomere shortening without a reduction in telomerase activity, suggesting that Perifosine may impair telomere maintenance in some backgrounds in a telomerase enzymatic activity independent manner, such as by regulating telomerase nuclear localization. Telomere shortening, changes in telomerase activity and changes in growth rate did not correlate with one another in response to Perifosine, indicating that Perifosine has many context-dependent effects on target cells and that each of these phenotypes is regulated by at least partially non-overlapping response networks.

The gene-set enrichment analysis (GSEA) indicated that the p38 pathway or other stress pathways may be more important than the AKT pathways in determining response to Perifosine in terms of telomere length. Phosphorylation of p38 in response to Perifosine in three of four shortening lines evaluated and neither of the static lines evaluated suggests that the propensity to increase the activity of this pathway in response to Perifosine may predict the telomere response, though it is probable that downstream effects of p38 are responsible for the actual change in telomere length. GSEA indicated that the static lines exhibit a generally inflammatory, migratory and stressed phenotype before administration of Perifosine; if a similar change is responsible for the telomere shortening observed in the sensitive lines, it is possible that the static lines are insensitive to further activation of this network because they are adapted to constitutive signaling through these pathways, and

have developed regulatory states that maintain telomeres even in the presence of these stimuli. Indeed, many of the characteristics of the static cell lines are in common with the networks observed enriched specifically in the context of senescence, rather than quiescence or radiation-induced growth arrest (Lackner *et al.* 2014). Decreased p38 activity has been associated with increased telomerase enzymatic activity and decreased DNA damage signaling in senescent CD8+ lymphocytes (Henson *et al.* 2015), which could be circumvented through the addition of CD45RA to activate p38 signaling. Together, the evidence from the literature suggests that in many contexts activation of p38 can drive a reduction in telomerase activity and telomere shortening, which logically must have been circumvented through some compensatory mechanism in the static lines because of their constitutively higher activity in this pathway in the presence of stable telomere length.

The GSEA generated many hypotheses, and the hypothesis evaluated, that p38 activity induced a change in telomere length in the shortening lines through a telomerase enzymatic activity independent mechanism, appears to be true in some contexts. However, another important experiment that has not yet been performed is a measurement of global changes in gene expression in the shortening lines compared to the static lines; based on these results, I would predict that the downstream transcriptional targets of p38, inflammation and DNA damage signaling would increase in the shortening lines, while the static lines would be more insensitive to this treatment. Further, since most of the genes most highly expressed in the static lines compared with the shortening lines are endocrine or paracrine excreted diffusible factors (chemokines, IL8/IL6, urokinase), chronic treatment with conditioned media from the static lines with the addition of Perifosine may abrogate the telomere shortening observed in the shortening lines if the activation of these factors provides a protective milieu. However, if these factors are merely the result of constitutive

stress signaling, conditioned media may increase the telomere shortening observed in the sensitive lines. These secreted factors may be useful predictors of telomeric response to Perifosine because they can be measured in blood without highly technical assays, and because of my collaboration with Dr. Friedman and the clinical trial group at Duke it will be possible to measure levels of these factors in stored blood from human patients in their clinical trial to determine if the levels of these factors before therapy correlate with clinical response (assuming such measurements are justified by *in vitro* data). This work comprises the beginning of an understanding of the mechanism through which Perifosine induces telomere shortening in some lines, but definitive experiments remain to be conducted.

AKT inhibitors and Perifosine were initially tested because of the association between AKT and a number of the candidate genes evaluated in Chapter 2. However, subsequent analysis revealed that the p38 pathway and other stress pathways appeared more important to the telomeric effects of Perifosine than the AKT pathway, and the absence of a pattern in expression in the three hits specifically noted as AKT-responsive (LARP1, LARP7 and LGALS7, Figure 3.19) is consistent with the rest of the GSEA. Unexpectedly, DNA damage/DNA repair pathways appear to be more highly expressed in the Perifosine-sensitive lines, and because a large number of the candidate proteins were selected from the DNA damage response pathways they appear to be enriched in the shortening lines. BLM, a candidate helicase upregulated in all four shortening lines compared with the static lines has been implicated in telomeric recombination in yeast (Lillard-Wetherell *et al.* 2005), and both WRN and BLM can unwind telomere-specific secondary structures (Popuri *et al.* 2014). Further, the two error-prone polymerase candidates upregulated in shortening lines, POLH and POLL are implicated in translesion DNA synthesis and non-homologous end joining (NHEJ), respectively. In addition to

thymine dimers, POLH can bypass 8-oxo guanine damage and abasic sites, both of which can be caused by increased reactive oxygen species which would be predicted to result from activated stress pathways (Makridakis & Reichardt 2012). 8-oxo guanine lesions are specifically enriched with concomitant decreases in telomerase activity in chronically inflamed states, such as Barrett's esophagus (Cardin *et al.* 2013). Though it is counterintuitive that higher levels of DNA repair proteins and thus better maintenance of the genome correlated with sensitivity to a stress-inducing agent, it is possible that the DNA damage response pathways in these lines are partially intact, such that the activity of these pathways prevents telomerase activity, while cell division and DNA replication occurs in spite of DNA damage because these are cancerous cells. One interesting followup experiment would be to treat shortening lines with Perifosine and inhibitors of the DNA damage response such as an ATM inhibitor to determine if the DNA damage response is required for the telomerase/telomere length phenotype observed.

In total, this chapter demonstrated that two of four drugs selected because they were AKT inhibitors could induce telomere shortening in Hela cells. Detailed analysis of one of them, Perifosine, indicates that it induces telomere shortening in roughly 60% of cell lines evaluated, possibly through activation of the p38 activity and a reduction in TRF2 abundance; Perifosine also led to changes in growth rate and telomerase enzymatic activity that were not correlated with the change in telomere length. A xenograft experiment treating a cell line known to be sensitive to Perifosine *in vitro* did not produce a statistically significant reduction in metastatic tumor burden or telomere length, though technical issues and the unexpectedly high mortality rate of treated mice may mask a subtle effect. Human samples treated with Perifosine exhibit a reduction in telomerase activity and shortening of the shortest telomeres but this effect was modest and only detectable through the use of

specialized, highly sensitive methods. Perifosine or other AKT inhibitors/p38 activators may have clinical utility as telomerase inhibitors, but additional pre-clinical work is required.

## **METHODS**

### **Cell culture and *in vitro* Perifosine treatment**

All cells were cultured at 37° C in 5% CO<sub>2</sub>, in Media X (HyClone, Logan UT) with the addition of 10% calf serum (HyClone, Logan, UT). Cells were treated with the indicated concentrations of Perifosine with media changes and fresh Perifosine added every 3 days; cells were passaged 1:32 when they approached 100% confluence.

HCC38 cells expressing a Luciferase (HCC38 + Luc) construct were obtained from the Brekken lab (UT Southwestern). HCC38 + Luc cells were infected with a lentiviral vector bearing an empty pGipZ plasmid containing GFP and puromycin resistance genes (HCC38+Luc+GFP).

### **Soft agar colony formation assay**

The soft agar colony formation was performed as in (Roig *et al.* 2010) with a number of modifications. The assay was performed in triplicate in 12-well plates, at a density of 1500 cells per well. In addition to the two agar layers, 0.5mL Media X +10% calf serum was added either with the addition of 1.84uM Perifosine (long-term treated and transiently treated cells) or without the addition of Perifosine (control cells). Cells were allowed to grow under these conditions for four weeks, replacing the media with fresh media every three days. At the conclusion of week 4, colonies were imaged and counted with the ImageJ analysis software.

## **Xenograft design**

The second mammary fat pads of 30 (15 control and 15 treated) 6-8 week old female NOD/SCID IL2G mice (obtained from an on-campus supplier) were injected with  $3 \times 10^6$  HCC38 cells, and tumors were allowed to grow for 2 weeks before Perifosine was administered to the treated group. Primary tumors were measured once weekly in weeks 4 to 6 by digital calipers. Primary tumors were removed after measurement in week 6. Bioluminescent measurements of recurrent primary tumor burden were obtained in week 12. The mice were subsequently sacrificed and lung tumor burden was measured by luciferase activity. Tumor samples were disassociated and cultured for telomere analysis at primary removal in week 6 as well as from both recurrent primary and lung metastatic tumors in week 12. All xenograft experiments were conducted under UT Southwestern IACUC approval.

The treated group was given an initial loading dose of 62.6 mg/kg Perifosine suspended in sterile PBS by oral gavage in the first week of treatment followed by weekly maintenance doses of 35.8 mg/kg Perifosine which are reported to maintain Perifosine blood concentrations between 12 uM immediately following a dose and 5 uM immediately prior to a dose based on previous studies on Perifosine pharmacokinetics in mice (Vink *et al.* 2005).

## **Explant culture**

Following surgical excision, tumor samples were placed into 750 uL of Gibco Minimum Essential Media (MEM; Life Technologies) on ice and transported to the tissue culture facility. Tumor samples were then chopped into small pieces with autoclaved surgical scissors and resuspended in 750 uL MEM containing 0.5 mg/mL Liberase DL research

grade (Roche). Tissue was incubated at 37° C for 60 minutes under agitation at 225 RPM. Fragments were further dissociated by repeat pipetting through 10, 5, and 1mL pipettes (5x each), and then filtered through a 200 micrometer nylon mesh. Samples were centrifuged at 1500 x g for 5 minutes, and then resuspended in 10mL Media X + 10% fetal calf serum + 5 ug/mL puromycin and plated into 10cm dishes for cell culture to remove mouse cells. Fresh Media X + 10% fetal calf serum + 5ug/mL puromycin was replaced 3 days following cell culture to remove dead cells. On day 7 of puromycin selection (to remove mouse contaminating cells), culture dishes were washed 2x with 10 mL phosphate buffered saline, and then harvested for TRF analysis by trypsinization (5 minutes).

### **Surgical procedure**

For orthotopic mammary fat pad xenograft injections, mice were anesthetized with 2% continuous isoflurane and the area of the incision was thoroughly cleaned with 70% ethanol and betadine. A small incision was made over the right axillary fat pad and  $3 \times 10^6$  HCC+Luc+GFP cells suspended in 50uL autoclaved PBS+EDTA were injected into the fat pad using a 30-gauge needle. The incision was closed with wound clips, and mice were placed in a recovery cage on a heating pad. Mice were given 0.1mg/kg intraperitoneal buprenorphine analgesia following breathing stabilization postoperatively and 24 hours later. All surgical instruments were autoclaved prior to operating and a hot bead sterilizer was used to sterilize tools between subjects.

The primary tumors were removed using the same anesthesia and pre-surgical conditions as the xenograft injection procedure. A small incision was made in the skin directly overlying the tumor, and the skin/subcutaneous tissue surrounding the tumor was carefully dissected free from the tumor surface circumferentially using both blunt and sharp

dissection. The entire macroscopic primary tumor was removed when possible; however in cases where the primary tumor was involving surrounding structures (brachial artery, chest wall or forelimb musculature) a small amount of tumor was left in situ. Incisions were closed with wound clips, and mice were placed in a recovery cage on a heating pad, given buprenorphine analgesia, and monitored until ambulatory.

### **Bioluminescence imaging**

40 mg/mL D Luciferin (Gold Biotech, MO) suspended in autoclaved PBS was injected subcutaneously in mice anesthetized with 2% continuous isoflurane. Mice were placed on a heated imaging stage for ten minutes, and then imaged using the IVIS Lumina imaging system (Perkin Elmer, MA). Luciferase signals were quantified using the manufacturer's software. Mice were sacrificed immediately following acquisition of the recurrent primary tumor image in week 12; lungs were removed and imaged under the same conditions.

### **Animal care**

Animals were housed in a pathogen-free facility and the animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Texas Southwestern Medical Center.

### **Telomere length and telomerase enzymatic activity assays**

Terminal restriction fragment assays were performed as in (Herbert *et al.* 2003). The droplet-digital TRAP (ddTRAP) assay was performed as described in (Ludlow *et al.* 2014). Universal STELA was performed as in (Bendix *et al.* 2010), with the exception that a pre-lengthened, pre-annealed form of the panhandle primer pair was used in order to omit the

fill-in step. Panhandle oligos were pre-annealed by heating at 95°C for 5 minutes, followed by cooling to 25°C at a rate of 1°C per minute in New England Biolabs buffer 2 (New England Biolabs, MA). The sequence of the 42-mer panhandle primer was 5'-TGTAGCTGAAGACGACAGAAAGGGCGTGGTGC GGACGCGGG-3', and the 44-mer was 5'-TACCCGCGTCCGCACCACGCCCTTTCTGTCGTCTTCACGCTACA-3'.

### **Human samples**

CLL samples (10 million cells/pellet) were purified as part of the previous IRB approved clinical study at Duke University Medical Center (Friedman *et al.* 2014). Patients in that study were treated with 50 mg twice daily oral Perifosine for up to six months. Unused samples from the human clinical trial protocol were used for these experiments.

### **Statistical analysis**

Two-sided t-tests performed with Graphpad Prism 7 software were used to compare group means. Gene-set enrichment analysis was performed using the default parameters of software available from the Broad institute (<http://www.broadinstitute.org/gsea/index.jsp>) on expression data obtained from the Minna lab on the cell lines of interest (Mootha *et al.* 2003; Subramanian *et al.* 2005).

## **CHAPTER FOUR**

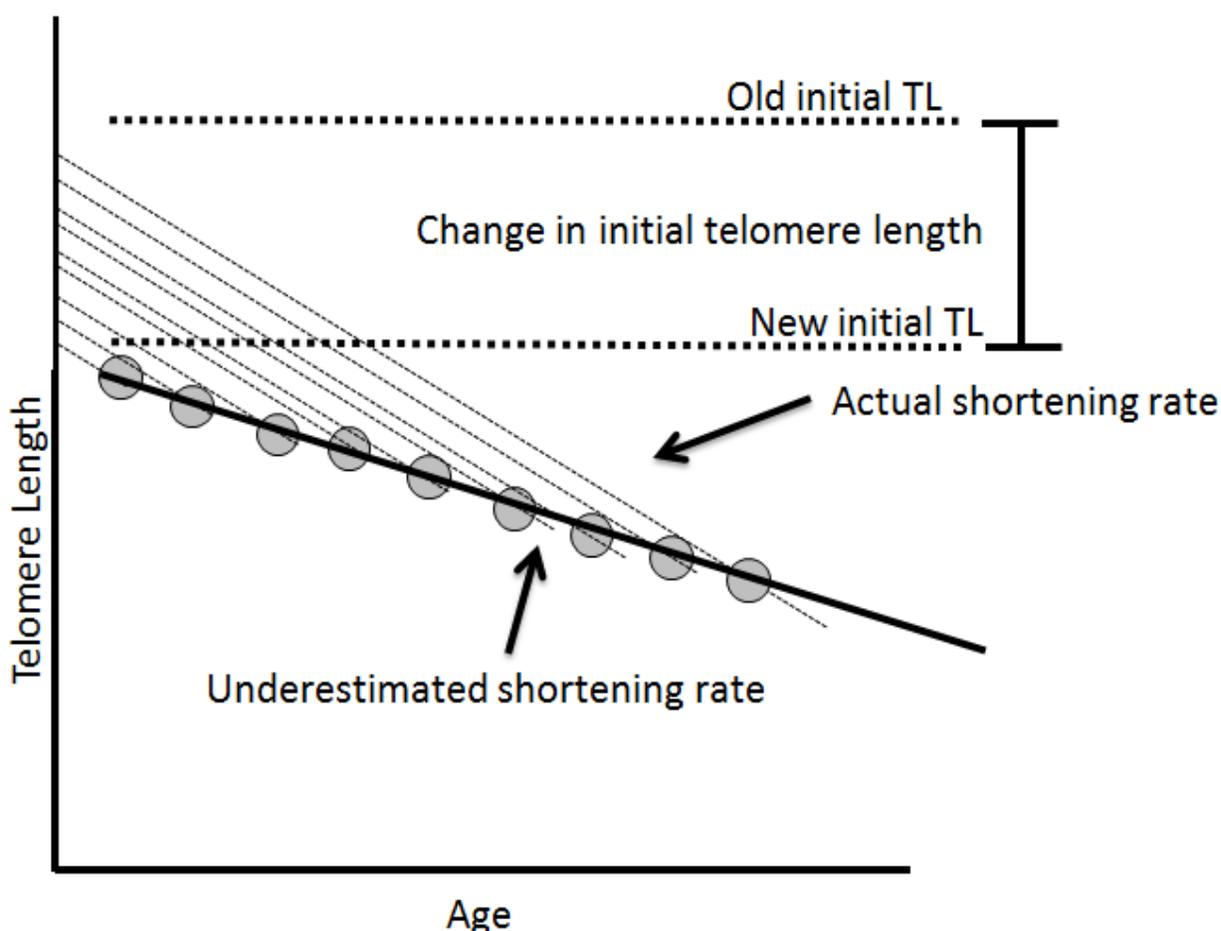
### **Identification of a time-dependent trend in telomere length**

#### **INTRODUCTION**

In humans, telomere length in peripheral blood nucleated cells, Leukocyte Telomere Length (LTL) shortens with age. In addition to the telomeropathies (diseases of telomere shortening) described in Chapter 1, shorter LTL has been linked to numerous complex diseases, including hypertension, cardiovascular disease, dementia, diabetes, stroke and asthma as well as environmental exposures associated with these pathologies such as stress, infections, smoking and obesity ((Farzaneh-Far *et al.* 2010; Kume *et al.* 2012; Jiang *et al.* 2013; Raschenberger *et al.* 2013; Albrecht *et al.* 2014; Huzen *et al.* 2014; Rode *et al.* 2014; Zhao *et al.* 2014), reviewed in (Bojesen 2013)). This leads to the idea that telomere length in general may be an integrator of total organismal stress, and could in theory be useful as an indicator of general health.

Facilitating this, measurements of LTL correlate strongly with measurements of TL in other organs (Daniali *et al.* 2013). A trend in the literature observed during the composition of (Holohan *et al.* 2014) is that estimates of telomere shortening rate give widely varying answers depending on the method used to evaluate it. Most studies that measure age-associated telomere shortening rates utilize a cross-sectional design, in which telomeres of multiple individuals of varying ages are measured and the shortening rate is computed from a linear regression of LTL on age. This method relies on a number of assumptions, though the most important assumption is that all individuals within the population studied started at the same average telomere length, and that the initial telomere length has no relationship

with age or time. If this assumption is not true, all differences in telomere length that arise from this age-associated trend in initial length will be ascribed to age even if such an underlying shift occurred, and this will lead to a bias in perceived shortening rate. If a time-dependent trend in initial length, what is known as “secular or time-dependent trend” in epidemiology, exists, cross-sectional measurements of telomere shortening rate will be biased in the opposite direction from the secular trend. This bias will lead to an incorrect shortening rate, and a discrepancy between the rate measured cross-sectionally and the rate measured through other study designs.



**Figure 4.1: Schematic of how a secular trend could bias cross-sectional measurements**  
 Cross sectional analyses rely on the assumption that initial telomere length is not changing. In a cross-sectional estimate of telomere shortening rate (black line), measurement of telomere length in multiple individuals of different ages (grey circles) will yield an underestimate of the actual telomere shortening rate (constant for all samples, dotted lines) if initial telomere length (y-intercept) is decreasing with time.

The bias produced by such a shift will depend on its direction. If initial length were increasing, cross-sectional measurement will produce an erroneously high rate of telomere shortening (younger individuals started longer than older individuals, so the slope of such a curve will appear steeper), whereas if initial length were decreasing cross-sectional measurement will produce an incorrectly low estimate of telomere shortening rate (younger individuals started shorter than older individuals, so the curve will appear less steep). Prior to this work, the possibility of a secular trend in initial telomere length had not been explored.

An alternative methodology, longitudinal analysis, measures telomere shortening rate by measuring a group of individuals at more than one time point several years apart. Shortening rate is then computed as the average difference between the first and second time point. Since longitudinal analysis constitutes a direct observation of the telomere shortening rate (the study design gives two or more concrete observations for each individual), longitudinal analysis relies on many fewer assumptions than a cross-sectional study, and so likely represents a much more correct picture of telomere shortening rate. While most studies utilize a cross-sectional design because of the daunting technical challenges imposed by telomeres (TRF)-based measurements of thousands of samples, there are a small number of studies that have examined telomere shortening rates longitudinally utilizing TRF analyses (Gardner *et al.* 2005; Aviv *et al.* 2009; Ehrlenbach *et al.* 2009; Farzaneh-Far *et al.* 2010; Chen *et al.* 2011; Houben *et al.* 2011). These studies show age-associated telomere shortening rates consistently ~20 base pairs per year higher compared to the cross-sectional studies.

While the longitudinal studies are generally smaller than cross-sectional studies, this difference even appears within a dataset. For example Chen (Chen *et al.* 2011) noted this difference in telomere shortening rate measured at each time point in their cohort cross-sectionally compared with the longitudinal measurement of telomere shortening rate, demonstrating that this is not a feature of population selection or some artifact of DNA degradation or methodological refinement over time, rather it is a trend that appears due to the inherent differences in cross-sectional versus longitudinal study design.

Furthermore, the telomere dynamics of cell lines with perturbed telomere biology described in Chapter 1 cast doubt on the interpretation of another trend described in the literature, the positive correlation between a father's age at the birth of their offspring and telomere length in those offspring (Unryn *et al.* 2005; De Meyer *et al.* 2007; Aston *et al.* 2012; Eisenberg *et al.* 2012; Prescott *et al.* 2012; Broer *et al.* 2013). Groups observing this trend cite the observation that cross-sectional analysis of sperm telomere length indicates a strong positive correlation with male age—older men have longer telomeres in sperm, which translates into perceived elongation of roughly 57 base pairs per year (Aston *et al.* 2012), which could explain why older fathers have offspring with longer telomeres. However, there are no cell lines with a perturbation to telomere biology that results in an indefinite linear upward trend in telomere length; cells with increasing telomere length invariably approach a new equilibrium telomere length following exponential kinetics typical of idealized chemical reactions approaching equilibrium (also observed in the literature in (Smith & de Lange 2000; Lee *et al.* 2009)). There are only a handful of perturbations known to create an indefinite downward trend, and they directly interfere with the telomerase holoenzyme, such as the loss of DKC1 (Batista *et al.* 2011). The linear nature of the trend identified indicates that either sperm precursors behave differently from cells in tissue culture (which is entirely

possible but lacks any evidence other than this observation), or that something other than progressive telomere elongation explains the trend.

In the developed world, the age of a father at the birth of his offspring (FAB) has been increasing over time (Bray *et al.* 2006), and the trends described above in increasing TL in older men would logically dictate that the average telomere length in the population would be increasing. That change would provoke an overestimate of telomere shortening rate cross-sectionally compared to longitudinally, which is the opposite of the bias observed. However, if sperm telomere length was fixed at a given point in an individual's development and telomerase positive sperm precursors merely maintained their telomere lengths at this equilibrium point (behavior observed often in tissue culture), observation of increasing telomere length in older men could be produced by a downward trend in this equilibrium length, which would logically be related to initial telomere length. In short, the sum of the literature is internally inconsistent, as the rise in initial telomere length predicted by the combined observations of increasing FAB and the relationship between FAB and TL is not observed, and the difference between longitudinal and cross-sectional measurements of shortening rate does not provide supporting evidence for such a theory. Rather, a downward trend in initial telomere length combined with a fixed equilibrium length related to initial telomere length in sperm cell precursors would explain both the discrepancy between longitudinal and cross-sectional measurements of shortening rate, the perceived lengthening in sperm with age and the relationship between FAB and TL in offspring, while maximizing parsimony with *in vitro* observations about telomere dynamics; this hypothesis invokes no novel telomere biology on the part of sperm cell precursors.

These observations in the literature led to the hypothesis that some underlying secular trend was biasing cross-sectional measurements of telomere shortening rate and producing the perceived lengthening of sperm telomere length. If such a trend were responsible, it would indicate that initial telomere length is decreasing over time, since cross-sectional measurements seem to be indicating erroneously low rates of telomere shortening (the curve is less steep than it should be) and that older men have longer telomeres (in this context because their initial telomere length is longer than that of younger men).

The average telomere length in humans has changed in the past over comparatively short timescales; differing mean LTL has been reported between different ethnic groups (Aviv *et al.* 2009) as well as between different countries within Europe (Eisenberg *et al.* 2011). Such changes could be accomplished via founder effects given the large inter-individual variation in LTL within a population (1-2 kb around the age-mean), but founder effects are not required. Telomere length, while hereditary, is predisposed to feed-forward changes in telomere length due to the nature of its inheritance; telomere length is inherited through two main pathways. Similar to almost any trait, variants in genes responsible for telomere maintenance, such as TERT, can alter telomere length (Bojesen *et al.* 2013). However, unlike most traits, the telomere repeats themselves are directly passed to the offspring physically via the telomeres present in germ cells from each parent, such that telomere length is also heritable epigenetically (De Meyer *et al.* 2014). This “dual inheritance” modality is readily apparent in the telomeropathies, as wild-type offspring of TERT mutant parents inherit the short telomeres from their parents even though they do not inherit the disease-causing (and telomerase compromising) variant, resulting in telomere lengths substantially shorter than the rest of the population, an effect which is apparent even two generations removed from telomerase dysfunction (the grandchildren of TERT mutants)

(Chiang *et al.* 2010). Furthermore, because of this feed-forward shortening, most telomeropathies exhibit anticipation via increasingly short initial telomere length within a cohort, leading to progressively more severe and far-reaching organ failure in younger generations (Armanios *et al.* 2005).

It is interesting to note that though the sample sizes for these families are typically too small to demonstrate this definitively, a cross-sectional analysis of telomere shortening rate in these families is likely to yield a flat relationship between age and telomere length because of the strong downward trend in initial telomere length as a result of the well-documented anticipation (for example the behavior of TERT mutants in (Chiang *et al.* 2010), Figure 1). Such an observation could be taken to indicate that these families do not exhibit age-associated telomere shortening under the same assumptions that underlie the cross-sectional estimates of telomere shortening. While telomere length is almost certainly constrained by selection at the lower and upper ranges (by telomeropathies and increased risk of certain cancers (Anic *et al.* 2013), respectively), it appears that this dual inheritance modality predisposes TL to highly directional random walks within that range, which may explain the high inter-individual variation in LTL.

Concurrently, many other developmental and physiological parameters are changing in a time-dependent manner. Secular trends include decreasing age at pubertal onset, increasing rates of obesity, diabetes, cardiovascular disease and asthma, increasing height and decreasing smoking are all operating at the present time (Komlos & Breitfelder 2008; Toppari & Juul 2010; Romero *et al.* 2012). Many of these phenomena (e.g. smoking) are known to impact telomere length, and a plethora of interactions between growth/insulin signaling, sex hormone signaling, cardiovascular health and cellularity and telomere biology

have already been described (Chen *et al.* 2014; Vasunilashorn & Cohen 2014; Boyer *et al.* 2015; Entringer *et al.* 2015); any of these trends are likely to produce some change in observed mean telomere length, and it is likely that other time-dependent trends in human health impact telomere length. Given these recent, large-scale trends, an alteration in initial population-scale telomere length is possible.

In this chapter the methods utilized to test the hypothesis that a secular trend in telomere length produces the internally inconsistent findings in telomere shortening rate are detailed. I introduce Paternal Birth Year (PBY) as a variable that can be used to correct for changes in initial telomere length in cross-sectional analysis, and demonstrate that a secular trend in telomere length exists even after correcting for the paternal effects described, indicating that something other than demographic changes are driving this change. These experiments use telomere length and demographic data provided as a collaboration with the UK Twins, National Heart, Lung and Blood Institute (NHLBI) Family Heart study and Asklepios study, which included information on age, telomere length, parental demographic information and smoking status from 2710, 2177, and 2434 individuals, respectively. Each cohort has been used prior to this work to study telomere biology (Higgins *et al.* 1996; Andrew *et al.* 2001; Rietzschel *et al.* 2007; Kimura *et al.* 2008), however this is the first instance of analysis for this purpose.

## **RESULTS**

*Cross-sectional estimates of telomere shortening rate are consistently lower than longitudinal estimates*

In order to understand the difference in cross-sectional and longitudinal measurements of telomere length, a list of studies that utilized TRF-based measurements of telomere length was compiled, because other telomere measurement techniques yield output in arbitrary units; these units are experiment-specific, and so inter-study comparisons become extremely unreliable. Since cross-sectional studies were considerably more abundant than longitudinal studies, only those with greater than one thousand participants were considered. All longitudinal studies are reported. The weighted average shortening rate reported in cross-sectional analysis is 22.82 bp/year, whereas the weighted average shortening rate in longitudinal analysis is 41.19 bp/year ( $p < 0.001$ ), shown in Table 4.1.

<b>Cross-sectional study</b>	<b>N</b>	<b>Shortening rate (bp/year)</b>
Valdes, 2005	1122	27
Aviv, 2006	1517	20.5
Cherkas, 2006	1552	19.8
Steer, 2007	1327	22
Bataille, 2007	1897	27
Barwell, 2007	1768	22
Cherkas, 2008	2152	22
Richards, 2008	1207	18.5
Hunt, 2008	1395	20
Fitzpatrick, 2011	1136	26
Weighted average		22.82
<b>Longitudinal study</b>	<b>N</b>	<b>Shortening rate (bp/year)</b>
Gardner, 2005	70	31.3
Ehrlenbach, 2009	510	45.5
Aviv, 2009	685	40.7
Farzaneh-Far, 2010	608	42
Chen, 2011	271	31.6
Houben, 2011	75	40.2
Weighted average		41.19

**Table 4.1: Summary of age-associated telomere shortening rates in the literature**

Telomere shortening rate as measured cross-sectionally is consistently ~20 bp/year slower than directly observed longitudinally.

*PBY predicts telomere length better than FAB and models of cross-sectional data which do not assume fixed initial telomere length are consistent with longitudinal studies*

To examine if a change in initial telomere length were occurring with time, the ideal experiment would be to measure telomere length in newborns over time (5-10 years at a minimum, for at least several hundred newborns per year); since that was labor and funding prohibitive, evaluation of existing datasets seemed like the most prudent method available.

For a study with a narrow window of sample collection, an individual's birth year and that person's age at the time of measurement will generally be nearly perfectly collinear, with the only variation in the age to birth year relationship arising from the sample collection time window. This almost perfectly collinear relationship would dramatically complicate analysis of the relationship between birth year and telomere length, and most studies had narrow enough collection windows that this would be unlikely to yield informative results. However, since many studies had evaluated father's age at birth of the offspring (FAB) with relation to telomere length, and if one were to perform the converse of that analysis, instead examining paternal birth year (PBY), this would allow the paternal variable to accommodate a shift in the underlying initial telomere length, freeing age in such a model to evaluate the telomere shortening rate without being subject to underlying time-dependent trends. Under this paradigm Tim de Meyer, Tim Spector and Steve Hunt (who manage the Asklepios, UK Twins and NHLBI-FHS telomere data, respectively) agreed to collaborate in order to test for an underlying shift in initial telomere length. With their assistance I was able to gain access to data on telomere length, age at time of sample collection, birth year, paternal and maternal birth year and age at birth of the individuals for these populations.

Model	Study	Paternal Effect (95% CI) (bp/year)	Age Effect (95% CI) (bp/year)	Paternal Effect Partial R <sup>2</sup>	Age Effect Partial R <sup>2</sup>	Model R <sup>2</sup>
Paternal Birth Year (PBY)	UK Twins	-19.48 (-23.03 to -15.93)	-40.94 (-45.04 to -36.82)	0.0681	0.1144	0.1825
	NHLBI-FHS	-14.66 (-18.18 to -11.14)	-36.42 (-40.54 to -32.31)	0.0847	0.1114	0.1962
	Asklepios	-17.73 (-21.97 to -13.49)	-46.05 (-52.76 to -39.33)	0.0257	0.0450	0.0708
	Combined	-17.22 (-19.40 to -15.04)	-39.44 (-42.08 to -36.80)	0.1058	0.0581	0.3967
Father's Age at Birth (FAB)	UK Twins	13.39 (9.63 to 17.14)	-21.38 (-23.25 to -19.51)	0.0067	0.1560	0.1627
	NHLBI-FHS	14.56 (11.04 to 18.08)	-21.69 (-23.63 to -19.75)	0.0184	0.1774	0.1958
	Asklepios	17.17 (12.90 to 21.43)	-28.30 (-33.03 to -23.57)	0.0238	0.0450	0.0689
	Combined	14.85 (12.63 to 17.08)	-22.18 (-23.53 to -20.84)	0.0116	0.1629	0.3911

**Table 4.2: Contrasting models of Age and PBY vs Age and FAB in all three datasets**

Multiple linear regression models using Paternal Birth Year (PBY) in addition to the age, top panel, produce age-associated telomere shortening rates more consistent with longitudinal measurements of telomere shortening than models that utilize Father's Age at Birth (FAB), bottom panel in all three datasets individually and combined. Further, models using PBY instead of age produce higher R<sup>2</sup> values and have a lower fractional contribution from age.

The first test performed was to evaluate if PBY was better than FAB at predicting telomere length in a model that includes age, and if these models became consistent with longitudinal results when initial telomere length is allowed to vary through inclusion of PBY. If a model utilizing PBY were better (had a higher R<sup>2</sup> value) than a model that included FAB instead of PBY, one could reject the null hypothesis that PBY (and thus an underlying shift in telomere length) does not occur. If PBY were worse than FAB (lower R<sup>2</sup>), which could occur if FAB was the underlying driving force and application of PBY instituted an unnecessary time dimension to this effect, I could rule out the idea that initial telomere length were changing. Because of the very large size of the datasets, the p-values for all linear regression models were less than 0.001.

Table 4.2 shows the results of this test; models that include PBY are better (higher  $R^2$  values) than models that include FAB in all three datasets and in all datasets combined, showing that PBY is a better predictor than FAB in all three populations. Further, PBY explains a great deal more of the variation in telomere length than FAB does (10.58% explained by PBY compared with 1.16% explained by FAB in the combined datasets), indicating that it is more important than FAB in determining telomere length. Interestingly, when PBY is included in the model, age becomes less important to the variation in telomere length, explaining 5.81% of the variance in the PBY model compared with 16.29% in the FAB model in the combined datasets, which may explain why the models' total  $R^2$  values are much closer together than the difference in  $R^2$  between PBY and FAB would otherwise indicate.

As predicted, including PBY in the model in order to allow for changes in initial telomere length reconciled the discrepancy between longitudinal and cross-sectional estimated shortening rate—in these cross-sectional datasets, inclusion of PBY in the linear model produced an estimate of age-associated shortening rate consistent with longitudinal studies, whereas models that included FAB were consistent with cross-sectional studies (39.44 bp/year in the PBY model vs. 22.18 bp/year in the FAB model).

The results also indicated that the effects of PBY and FAB were at least partially distinct from one another since FAB increased the predictive power of the model over a model that only included age without reducing the relative contribution of age to the model while PBY improved the model while impacting the effect of age. Because of these distinct effects, it was likely that the FAB effect was real and not just an inverted PBY effect, and a more accurate picture of the change in initial telomere length with time as well as age-

associated telomere shortening rate could be constructed through evaluation of the three variables (age, PBY and FAB) together.

#### *Collinearity between Age, FAB and PBY*

The largest problem with this approach is that in all three datasets, age was collinear with paternal birth year, and paternal birth year was collinear with age and father's age at birth (Table 4.3, graphically demonstrated in Figure 4.2). In order to evaluate these three variables together, some accommodation with their multi-collinearity would have to be reached. In the statistical literature and a number of genome-wide association studies, a method known as Mediation analysis is used in order to remove the effect of a known complicating variable (the mediator) from the effect of an independent variable of interest on a dependent variable (Preacher & Hayes 2008).

Pearson Correlation				
Variable	Age	PBY	FAB	MAB
PBY	-0.848*			
FAB	0.061*	-0.546*		
MAB	0.021 (NS)	-0.415*	0.796*	
MBY	-0.869*	0.957*	-0.416*	-0.469*

\*:  $p < 0.01$

NS: Not Significant

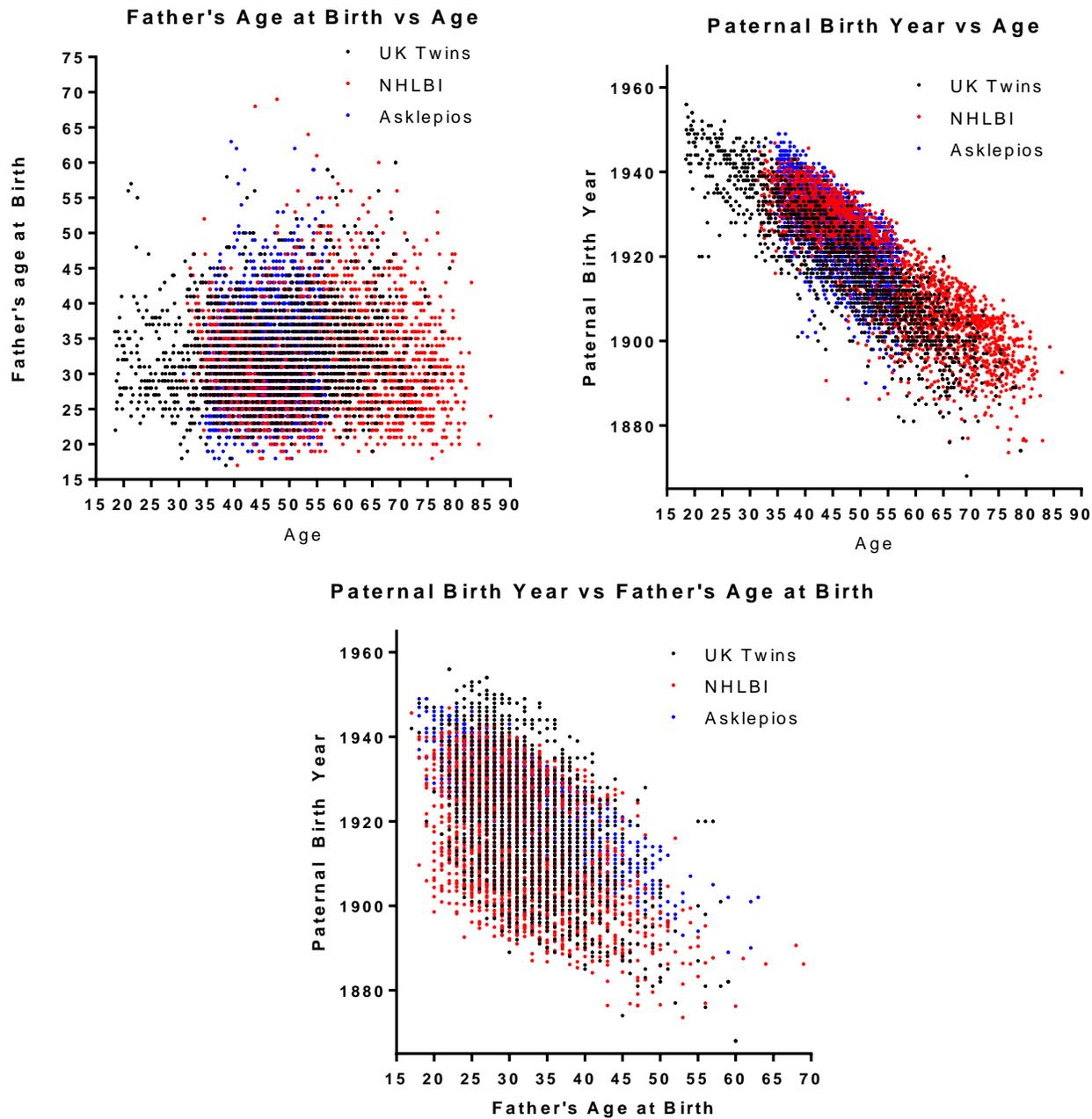
PBY: Paternal Birth Year

FAB: Father's Age at offspring's Birth

MAB: Mother's Age at offspring's Birth

MBY: Maternal Birth Year

**Table 4.3: Multicollinearity between the variables investigated in the combined datasets**



**Figure 4.2: Multicollinearity in the datasets investigated between Age and PBY, and PBY and Age and FAB.**

Mediation analysis reveals heterogeneity in the PBY effect and a larger FAB effect than previously reported

UK Twins					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	20.26	0.70	20.95	-41.64
PBY	31.58	*	-7.10	24.47	-12.38
FAB	-4.50	-23.19	*	-27.69	36.58
NHLBI-FHS					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	15.12	0.39	15.51	-36.81
PBY	27.47	*	-8.72	18.75	-5.94
FAB	-1.90	-23.41	*	-25.31	37.97
Asklepios					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	20.12	2.37	22.49	-48.42
PBY	20.82	*	-24.02	-3.20	6.29
FAB	-14.08	-41.06	*	-55.14	58.23
Combined					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	18.00	0.74	18.75	-40.19
PBY	28.10	*	-10.04	18.06	-7.18
FAB	-3.05	-23.77	*	-26.82	38.62

**Table 4.4: Mediation-adjusted models.**

The effect of each independent variable (Age, PBY and FAB) is adjusted for the effects of each other independent variable on telomere length and their collinearity within each dataset. The adjusted coefficients (far right) illustrate age-associated telomere shortening rates consistent with longitudinal observations, heterogeneity in PBY effects and reveal larger FAB effects than previously reported.

Mediation analysis allowed clarification of the effect of each independent variable investigated (Age, PBY, FAB) by correcting for the indirect effect of the other independent variables within these datasets (see the methods section of this chapter for a detailed

explanation for mediation). Mediation-corrected values for each of these variables are also consistent with longitudinal observations of telomere shortening rate, indicating an age-associated shortening rate of 40.19 bp/year in the combined datasets (Table 4.4). Additionally, the FAB effect after accounting for the indirect action of age any PBY was substantially larger than reported in the literature or derived from models that only include FAB and age (an increase of 38.62 bp/year in the combined datasets compared with 14.85 bp/year in the linear model of age and FAB). After accounting for the effect of the other variables, the PBY effect appeared highly population-specific and heterogeneous, demonstrating decreasing length with increasing year in the UK Twins and NHLBI cohorts (12.38 and 5.94 bp/year, respectively) and increasing length with time in the Asklepios set (6.29 bp/year).

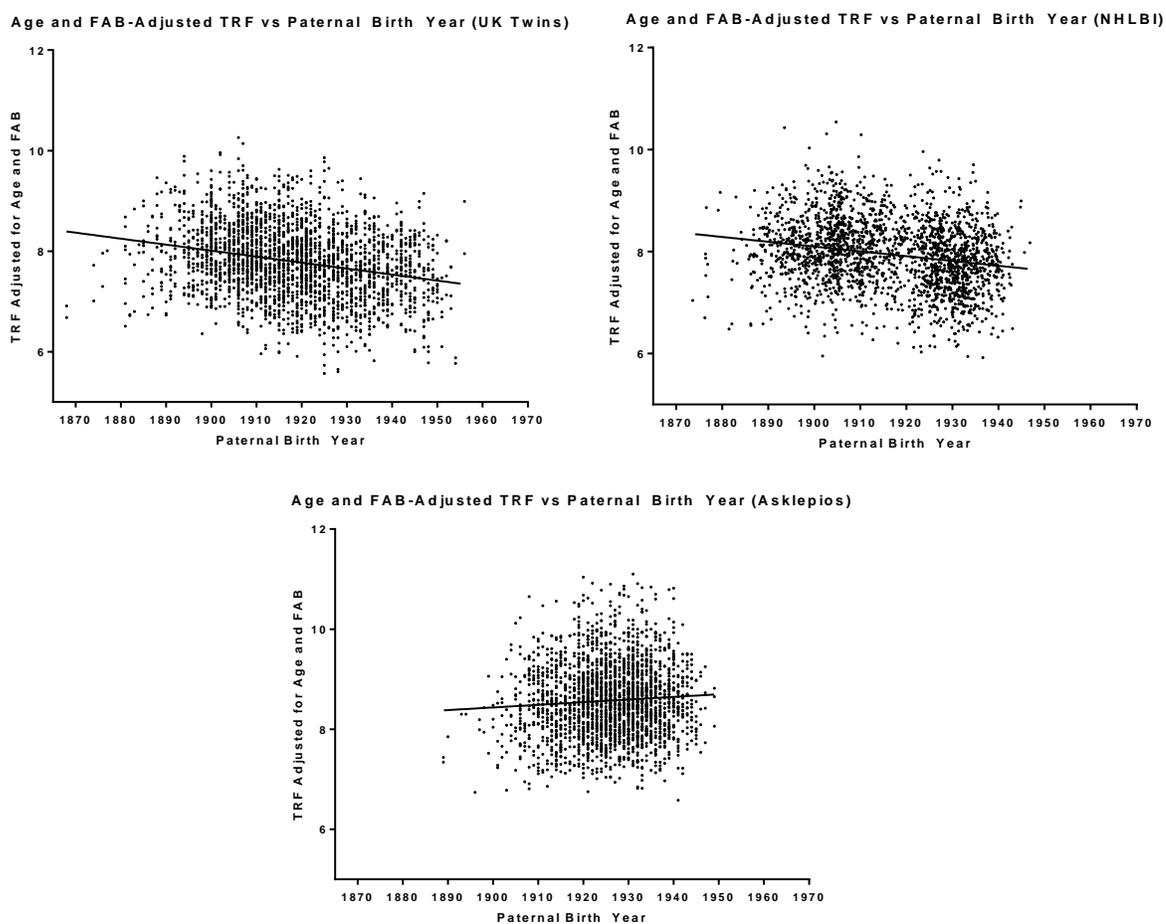
Because the literature identifying the FAB effect had reported a difference in magnitude of that effect by offspring gender, the mediation analysis was also performed on each gender separately (Table 4.5). In all cases except for males in the UK Twins, the direction and magnitude of the PBY and FAB effects were roughly similar. There were only 199 males in the UK Twins cohort, compared with 2511 females, so the gender discrepancy in that cohort may be due to a very small male sample size. The two studies with comparable numbers of males and females (NHLBI and Asklepios) generated conflicting observations about a gender-disparity in these effects. Males and females were almost exactly alike in all three variables examined in the NHLBI set, whereas the effects were larger in males for all three variables examined in the Asklepios cohort. Based on these observations, it is possible that gender plays a role in determining how these variables influence telomere length, but they do not appear to do so deterministically or strongly enough to appear in all datasets.

UK Twins (F)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	18.91	0.72	19.63	-40.62
PBY	30.80	*	-7.47	23.33	-10.70
FAB	-4.51	-24.17	*	-28.68	37.97
UK Twins (M)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	2.19	-0.33	1.86	-15.37
PBY	12.30	*	-0.90	11.40	-1.20
FAB	-3.66	-14.14	*	-17.80	8.26
NHLBI (F)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	15.14	0.50	15.64	-37.26
PBY	27.89	*	-8.63	19.26	-5.93
FAB	-2.64	-24.55	*	-27.19	39.09
NHLBI (M)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	14.88	0.25	15.13	-36.54
PBY	27.19	*	-8.91	18.28	-5.68
FAB	-1.15	-22.67	*	-23.82	37.12
Asklepios (F)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	22.12	2.89	25.01	-45.72
PBY	19.89	*	-22.31	-2.42	3.07
FAB	-3.06	-25.95	*	-29.01	45.07
Asklepios (M)					
Independent Variable	Mediation (bp/year)			Total Mediation	Adjusted Coefficient
	Age	PBY	FAB		
Age	*	18.37	1.892	37.29	-51.23
PBY	21.64	*	-25.92	-4.28	9.46
FAB	-3.09	-34.04	*	-37.13	49.49

**Table 4.5: Gender-segregated mediation-adjusted models.**

The effects of each variable were relatively constant between genders with the exception of the UK Twins, in which a small number of males complicates analysis.

Figure 4.3 shows a graphical representation of the PBY effect identified in all three datasets; TRF lengths are adjusted for the age and FAB effects in the mediation analysis, and these corrected telomere lengths are plotted against PBY of each individual in the three populations. This representation of the data shows the heterogeneity in the identified effect well, while illustrating that it is also quite linear.



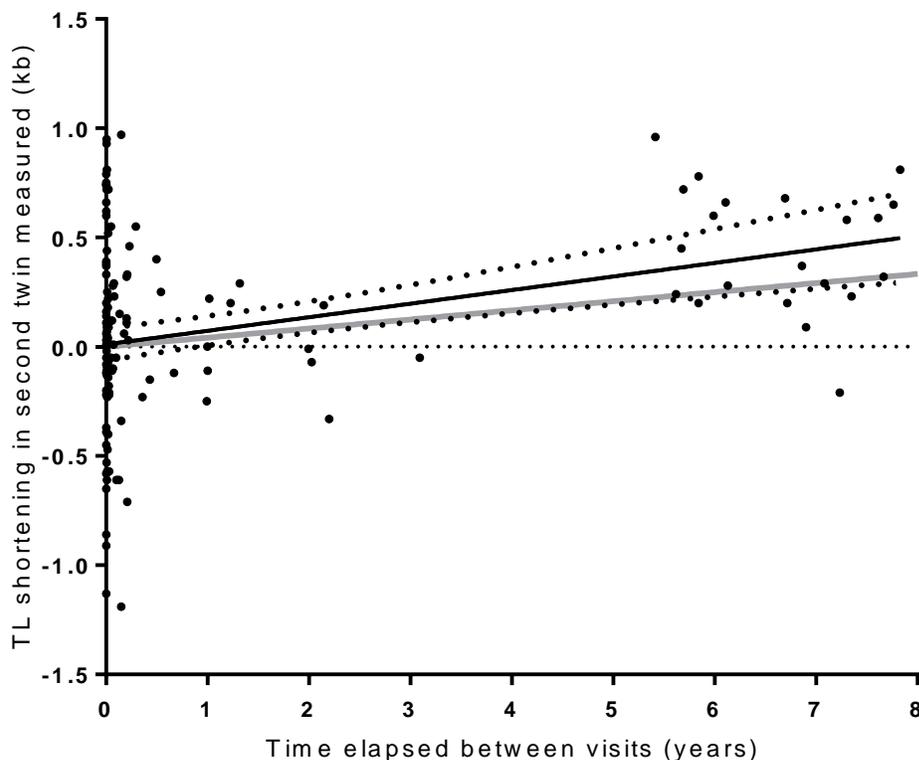
**Figure 4.3: Graphical representation of the PBY effect**

Paternal birth year (x-axis) was negatively correlated with telomere length after correcting for the effects of age and FAB (y-axis) in two of the three populations examined.

*The age-associated telomere shortening rate that results from the mediation analysis is more consistent with an independent measure of telomere shortening rate in the UK Twins than the shortening rate from cross-sectional analysis*

In order to perform another test to determine if the PBY effect and the age-associated telomere shortening rate observed when accounting for a shift in initial telomere length is more consistent with the actual telomere shortening rate, a method was devised to use the UK Twins cross-sectional data to perform a “pseudo-longitudinal” measurement of telomere shortening rate. Not every twin pair in the UK Twins population had their samples collected at the same point in time, and some twin pairs were measured up to eight years apart. For twins measured on the same day, on average the twin measured second should be no longer or shorter than the twin measured first barring some correlation between telomere length and assertiveness in a medical setting, which is possible but unlikely. However, if twins were measured on different days, on average the twin measured later should be shorter than the twin measured first, since the sample from the second twin is subjected to greater age-associated telomere shortening than the first twin—the time between the first twin’s sample collection and the second twin’s. The magnitude of this difference should be comparable to the age-associated telomere shortening rate, though it could be influenced by the factors that contribute to the delayed sampling of the second twin. This would be an independent measurement of the age-associated telomere shortening rate within one of the sample populations used in the mediation analysis, this method could be used to test if the age-associated shortening rate from the mediation analysis was more correct than the rate utilizing models of age and FAB.

To perform this analysis, the TRF length was subtracted of the second twin measured from the TRF length of the first twin measured, and this difference value was divided by the time between sample collection. It is important to note that this observation is biased by neither FAB nor PBV, since twins share both of these values (they have the same father at the same time).



**Figure 4.4: The pseudo-longitudinal telomere shortening assay in the UK Twins cohort**

Telomere shortening in the second twin measured compared to the first twin measured compared with the time elapsed between measurement of the twins can be used to determine age-associated telomere shortening rate in isolation from PBY and FAB. The black line is a linear regression of telomere shortening on time elapsed, with the dotted lines indicating the 95% confidence interval. The rate predicted from mediation analysis (41.64 bp/year, grey line) falls within this confidence interval. Each point is one twin-pair; only twin-pairs with different sample collection dates are shown.

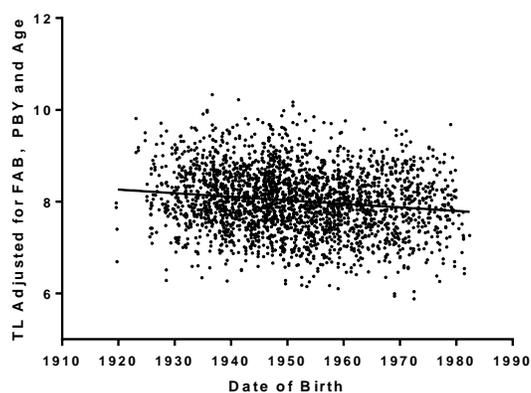
Pseudo-longitudinal analysis indicates a telomere shortening rate of 62.13 bp/year in the UK twins population—while this value is higher than predicted by mediation analysis, that rate falls within the 95% confidence interval of the pseudo-longitudinal regression, and is much closer to the pseudo-longitudinal value than that utilizing a model with only age and FAB (21.38 bp/year). This result indicates that an independent measurement of telomere shortening rate (pseudo-longitudinal analysis) estimates an age-associated telomere

shortening rate more consistent with the mediation-adjusted model including age, FAB and PBY than with a model including only age and FAB.

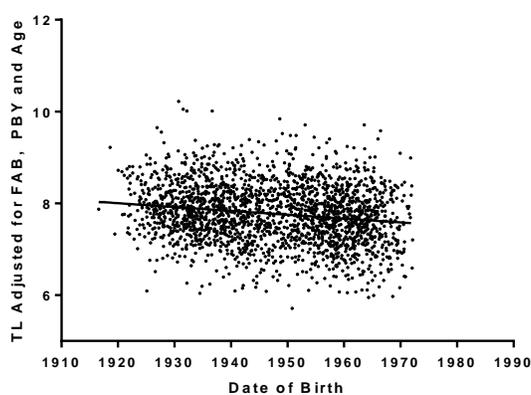
*A secular trend in initial telomere length exists after accounting for the effects of age, PBY and FAB*

Next, I tested if removing the mediation-adjusted effects of the three variables examined from the telomere data eliminated any relationship between date of birth and telomere length. Figure 4.5 shows telomere lengths adjusted for the mediation-adjusted values in Table 4.4 (a projection of initial telomere length if all individuals in that population had the same father's age at birth and paternal birth year) with regard to the date of birth of each individual in these datasets. There is a negative relationship between adjusted telomere length and date of birth in all three populations of 7.24, 14.97 and 23.80 base pairs per year in the UK Twins, NHLBI and Asklepios populations, respectively. This indicates that there is some time-associated variable that impacts telomere length that is not accounted for by the model that includes age, PBY and FAB; thus at least one important time-associated trend in telomere length remains to be discovered in these populations, and the heterogeneity of this effect suggests some environmental variable.

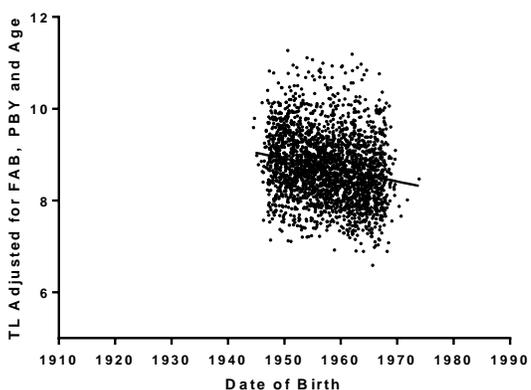
TRF length adjusted for FAB, PBY and Age vs date of birth (UK Twins)



TRF length adjusted for FAB, PBY and Age vs date of birth (NHLBI)



TRF length adjusted for FAB, PBY and Age vs date of birth (Asklepios)

**Figure 4.5: The secular trend in initial telomere length**

Telomere length after adjusting for FAB, age and PBY, projected initial length after correcting for paternal variables (y-axis) compared with date of birth of each individual (x-axis). A negative relationship between date of birth and adjusted initial telomere length exists in all three populations.

## **DISCUSSION**

In this series of experiments, I tested if a change in initial telomere length with time was the reason for the discrepancy between measurements of age-associated telomere shortening in cross sectional compared to longitudinal analyses. I demonstrated that inclusion of Paternal Birth Year (PBY) in a linear model with age resulted in an age-associated telomere shortening rate consistent with longitudinal data in cross-sectional datasets, and that models that include PBY instead of FAB were more robust. Further, I showed through mediation analysis that the PBY-correlated time-dependent trend in initial telomere length was heterogeneous in different populations, and that the FAB effect was larger than previously demonstrated after accounting for the age and PBY trends. Lastly, I show that a time-dependent (secular) trend in initial telomere length exists after adjusting for the effects examined.

While it was clear initially that change in initial and average telomere length is possible over time because different species (Gomes *et al.* 2011) and human ethnic groups possess different telomere lengths, it was unclear if such a trend was currently underway. The presence of a PBY effect after removing the effects of age and FAB, and the secular trend that remains after adjusting for all three variables shows that in humans, a large-scale change in telomere length may explain the discrepancy between longitudinal and cross-sectional measurements.

Though I initially hypothesized that a change in initial telomere length may explain the observation of increased telomere length in children of older fathers and of linearly increasing telomere length in sperm with age, it appears that the FAB effect is larger than

previously reported after accounting for changes in initial length with time and the collinearity between FAB and age via mediation analysis. Apparently heretofore uncharacterized telomere dynamics exist in sperm that allow them to behave in ways not observed in tissue culture, though it is also possible that some more linear process could explain the phenomenon, for example selection for sperm precursor cells with longer telomeres within the testis (which has been proposed before (Aviv & Susser 2013)).

Because each individual inherits half of their genome from each parent, the FAB effect in offspring should be equal to roughly half of the reported increase in telomere length in sperm (57 bp/year). The FAB effect reported here is more than half of this value, and this difference may arise from a number of phenomena. Firstly, the measurement of telomere elongation in sperm was based on a cross-sectional analysis, thus rendering it susceptible to the same biases as cross-sectional analysis of LTL. In this case, the secular trend and PBY effect will result in overestimation of the telomere elongation rate in sperm, since younger men started with shorter sperm TL than older men, indicating that the FAB effect is very substantially more than half of the magnitude of the sperm telomere elongation rate. Longitudinal analysis of sperm telomere dynamics would remove the bias from underlying shifts in initial telomere length, though that has not yet occurred.

It is possible that the epigenetic modality of telomere inheritance is largely driven by male gametes (De Meyer & Eisenberg 2014) because egg progenitor cells are far less proliferative per generation compared to male gamete precursors. The extent to which maternal age matters is not fully clear because of the aforementioned issues with collinearity between maternal and paternal variables, though some maternal age effect is certainly possible via telomere breakage or damage in meiotically arrested oocytes over time. If the

secular trend and PBY effect have been operating for more than one generation, a maternal effect is a logical consequence, as women would inherit shorter telomeres from their father, which would then be passed down as a maternal effect to their offspring. However, there is some evidence that paternal telomere length is more important for determining telomere length in offspring (Nordfjall *et al.* 2010), and it is more detrimental in terms of telomere length to inherit mutant telomerase alleles paternally rather than maternally; evidence from these families does indicate at least some maternal influence, as wild-type offspring of female TERT mutant heterozygotes still display telomere lengths substantially shorter than the rest of the population (Diaz de Leon *et al.* 2010), albeit to a lesser degree than individuals who inherited those alleles paternally.

In order to understand the implications of these findings, it is important to note that here PBY is used as a proxy variable for some unknown time-dependent variable correlated with PBY, rather than implying that PBY itself drives the effect. Between each population, the PBY effect was highly variable, possibly reflecting sensitivity to some difference in cumulative environment between the three study sites. The NHLBI Family Heart Study and UK Twins cohorts, the two datasets that indicated a negative relationship between PBY and telomere length, are both composed of individuals from urban centers in the US and UK, while the Asklepios cohort was composed of individuals in two smaller, more rural communities in Belgium (Erpe-Mere and Nieuwerkerken), presenting the possibility that the difference may result from some urban environmental exposure. Selection bias in individual recruitment may also explain why the Asklepios population behaved differently from the NHLBI and UK Twins groups, because the Asklepios study excluded individuals with overt cardiovascular disease, while the other two studies did not. Cardiovascular disease is associated with shorter telomere length (Fitzpatrick *et al.* 2011), and it is conceivable that

selecting only for individuals without cardiovascular disease may have selected for the individuals least exposed to the PBY effect; this hypothesis merits follow-up work that was not possible to perform with the current data.

The Asklepios cohort was roughly 1kb longer than equivalently aged individuals in the other two cohorts, and it is an intriguing possibility that the cumulative difference in PBY effect may explain why this population has longer telomeres than the other two groups. For example, if the Asklepios individuals and UK Twins cohort started at the same mean telomere length and they only diverged due to their contrasting PBY effects, a difference of 1kb in mean telomere length would exist in 53.5 years. A more general interpretation is that the PBY effect may in part explain why different populations in general have different mean telomere lengths, though there is not enough evidence to demonstrate this definitively at this juncture. In any case, it is likely that the PBY effect will exhibit both chronological and spatial variability due to yet-undetermined environmental, genetic and demographic drivers.

It will be interesting to observe how estimates of telomere shortening rate in existing cross-sectional cohorts change in response to these findings; because inclusion of PBY resolves the discrepancy in cross-sectional data compared to longitudinal data in these three populations, I expect this to occur in other populations as well. However, there is substantial variation in reported age-associated telomere shortening rates in cross-sectional studies, and it is not possible to understand the full scope of the PBY effect from evaluation of only these three populations, so the full variability in population-scale telomere dynamics is still relatively unknown.

Possibly the most important finding of these experiments is the identification of a time-dependent trend, or “secular trend” in telomere length after removal of the age and

paternal variables evaluated here. This indicates that there is at least one missing variable associated with time that is causing a reduction in telomere length. While this could be something relatively benign, such as a maternal effect that was not possible to evaluate here because of the extreme collinearity between maternal and paternal variables (Table 4.3), it could also represent a larger-scale environmental effect. It is becoming increasingly clear that environmental factors derived from industrial processes and urban living can impact telomere length (Hoxha *et al.* 2009; De Felice *et al.* 2012; Hou *et al.* 2012); exposure to these stimuli is increasing with time. Gradually increasing exposure to telomere shortening environmental stimuli is one possible explanation for the PBY effect and secular trend, particularly because of the dual inheritance modality in telomere length and the predisposition it generates toward feed-forward loops. For example, if individuals in a large city were exposed to a telomere shortening stimulus prior to reproduction, the initial telomere length in their children would be shorter (since they inherited the shortened molecules from their parents via germ cells). The second generation in the same environment would then be exposed to the same environmental conditions, and pass on even shorter telomeres to the next generation; presumably at some point natural selection will constrain this process, but it may manifest first as a large-scale public health problem. Notably, the FAB effect has been observed operating over two generations (Eisenberg *et al.* 2012), demonstrating that at least for organism-intrinsic processes, this feed-forward mechanism can occur.

One possibility is a recent environmental exposure as an explanation for the PBY effect and secular trend because the trends identified here would push these populations into pathological telomere length ranges, ~2-3kb shorter, in a few centuries. Given the secular trends in many other parameters that began in the modern era and the increased

rates of the industrial environmental exposures known to impact telomere length, it is possible that these trends are the result of many different evolutionarily novel insults to telomere biology. Projecting the trends shown in figure 5 back to their y-intercepts, in effect predicting initial telomere length in 0 CE, yielded unrealistic values in the range of 30kb; clearly, the trends began more recently than that. Telomere lengths of that size are only observed within Mammalia in Murine species, as well as some Marsupials that have long arrays of interspersed telomere and non-telomere DNA repeats (Gomes *et al.* 2011). Thus, the industrial revolution as the likely initiation time because of the many other large-scale changes in human reproductive life history, nutrient availability, density-dependent disease frequency, developmental timing and life span that occurred since the industrial revolution and drive the other secular trends.

It is interesting to note that differences between cross-sectional and longitudinal measurements of telomere shortening have been observed in other species in addition to humans, such as in *Meles meles*, the European badger (Beirne *et al.* 2014), and that this bias is in the same direction as in humans—cross-sectional observations show telomere shortening rates slower than longitudinal observations. Groups examining these species generally propose selective dropout of the shortest individuals from the population, in effect selecting for longer telomere lengths at the higher end of the age range, but it is also possible that if a large-scale environmental exposure were driving the secular trend in humans that these same exposures would lead to the same effect in other animals; while the selective dropout hypothesis is supported in some species, it is unclear to what extent the environment plays a role.

It is unlikely that the effects described here are driven by selective dropout of short individuals from the population in humans, as telomere length only weakly correlates with risk of death in humans, and only at extreme age ranges and extreme telomere values (the shortest quartile of the age-mean) (Fitzpatrick *et al.* 2011). Furthermore, short telomere length is most highly correlated with death from infections and cardiovascular disease, the two causes of death that have been most highly reduced by modern medicine. Rather, one possible explanation is that the secular trend and PBY effect arise from the lack of dropout of the shortest individuals in a population—because telomere length is highly heritable, individuals that survive and reproduce when they otherwise would not survive without medical intervention will have offspring with even shorter telomeres, who themselves will pass on short telomeres. This has been demonstrated in the literature in extreme cases, such as in the case of a TIN2 mutant patient with Dyskeratosis Congenita who was able to survive and reproduce following a bone marrow transplant; his offspring then went on to exhibit even shorter telomeres and Hoyeraal-Hriedarrsson syndrome, a more severe form of impaired telomere maintenance (Gleeson *et al.* 2012). Perhaps more commonly, there is a robust inverse correlation between asthma severity and telomere length (Albrecht *et al.* 2014); individuals with severe asthma that would normally not survive will have on average shorter telomere lengths, and offspring of these individuals would otherwise not enter the population. Since they are now able to survive and reproduce with medical care, this should decrease the average telomere length. Such a decrease would compound inter-generationally until some process arrested it, and the downward trend in initial telomere length may in part derive from modern medicine's contravention of the selective forces that had maintained the lower bound of telomere length in the past. At the very least, if selective dropout of individuals from the population were the cause of these trends, there should be

substantial nonlinearity in the observed trends because in humans mortality is very low until middle age, roughly age 50; these biases occur in populations of all age ranges.

While it is possible to make grim predictions about the implications of the PBY effect and secular trend described here (during peer review, one of the reviewers inquired about an extinction scenario), it is likely that the trends will not continue indefinitely, nor is it likely they have operated for a long period of time, and even if left unchecked, these trends will not cause humanity's extinction. In a hypothetical worst-case scenario in which these trends were to continue into a telomere range that began to increase the rates of disease associated with short telomeres, there would be a time lag measured in centuries between the time the shortest individuals are effected and the time the longest individuals are effected, which should be enough time for a selective sweep to solve the problem.

However, these trends are likely to have some relevance to public health, since a plethora of age-related diseases are correlated with short telomere length and at least some portion of those associations will be causal. In particular, it appears that the lung is the organ most sensitive to impaired telomere maintenance, as in addition to the association with asthma, idiopathic pulmonary fibrosis (IPF) is one of the first problems that occurs in the case of sporadic mutations in telomere maintenance genes (Diaz de Leon *et al.* 2011). The incidence of asthma in the developed world has been rising precipitously (Eder *et al.* 2006), and given the association between asthma and short telomeres as well as the trends identified here, speculation that progressively shorter telomeres may be in part driving the asthma epidemic may not be unwarranted. Short telomere length is a risk factor for some cancers, but it is also protective for other kinds of cancers, a large-scale trend in telomere length could also impact the rates of these diseases (Wentzensen *et al.* 2011; Anic *et al.*

2013; Walcott *et al.* 2013). Furthermore, there is a positive correlation between sperm telomere length and sperm count (Ferlin *et al.* 2013), and some developed countries have observed drops in sperm counts over time (Geoffroy-Siraudin *et al.* 2012) and a drop in teenage pregnancy rates without a corresponding change in sexual behavior (Jensen *et al.* 2002), suggesting that the trends identified in initial telomere length may also impact fertility. It is also possible that rates of age-related disease in the old may already be different from what they were pre-industrially without our knowledge because large-scale epidemiology occurred only very rarely prior to the Cold War.

To conclude, in this chapter a change in initial telomere length, represented by a paternal birth year driven effect is identified. These trends can reconcile a longstanding discrepancy in longitudinal compared to cross-sectional measurements of telomere shortening rate. In addition, the previously reported father's age at birth effect was clarified by accounting for the effects of the other variables within this dataset, showing that it is larger than previously reported. Lastly, a downward time-dependent trend in all three populations was identified after removing the effects of the other variables examined, and the public health implications of this trend were considered.

## **METHODS**

### **Mediation analysis**

Mediation analysis was performed with the multilevel R package (Bliese 2006), which uses the method described by Preacher and Hayes (Preacher & Hayes 2008). This method is intended to disentangle the effects of very collinear variables, and it allows quantitation of the extent to which a given variable (the mediator) biases the perceived effect of an

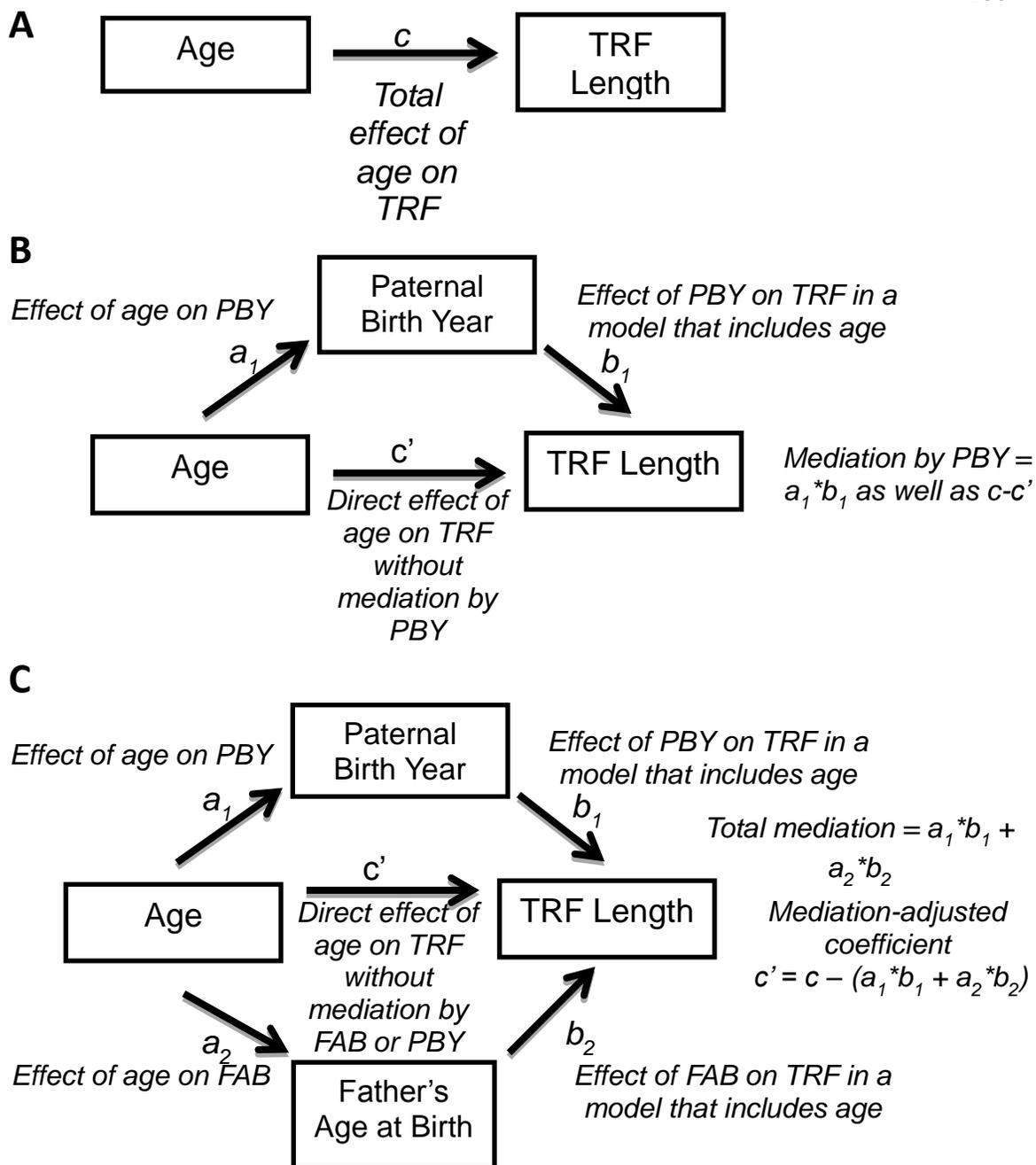
independent variable on a dependent variable. In what is known as the product of coefficients approach, the total effect of each independent variable (age, PBY and FAB in this case) on the dependent variable (TRF length) is determined via linear regression independently. Each independent variable is then regressed on each other independent variable, as well as the effect of each independent variable on the dependent variable in the context of a model including one other independent variable. The bias from each other independent variable is then computed as the product of the of the independent variable's effect on the mediator ( $a_1$ ) and the mediator's effect on the dependent variable in a model including the other independent variable ( $b_1$ ). This mediation will also correspond to the difference between the independent variable's effect on the dependent variable when considered by itself ( $c$ ) and that variable's effect on the dependent variable in a model that includes the mediator ( $c'$ ). Graphical demonstration of this method is shown in Figure 4.6.

## Datasets

All three datasets were obtained from the groups that originally collected the samples and measured TL (Higgins *et al.* 1996; Andrew *et al.* 2001; De Meyer *et al.* 2007; Rietzschel *et al.* 2007; Kimura *et al.* 2008). The datasets were selected because they were the largest populations that had the required demographic information as well as telomere length information, and all three of them used the Terminal Restriction Fragment (TRF) telomere measurement technique. TRF analysis was required because it gives output in non-arbitrary units (kilobases), and assays that give output in arbitrary units dramatically complicate inter-population comparisons.

## Statistics

Pearson correlations and linear regressions were in the Minitab 17 software package. When adjusting for paternal birth year, age and father's age at birth, in Figures 4.3 and 4.5, all samples were adjusted to age = 0, FAB = 0, PBY = 1900.



**Figure 4.6: Mediation analysis**

The effect of each other independent variable on each other independent variable can be determined through the product of coefficients approach

## **CHAPTER FIVE**

### **Identification of the Physiological Rate of Telomere Shortening (RoTS) Network via Twin Variance Analysis**

#### **INTRODUCTION**

Though telomere length (TL) is highly heritable, at any age range there is very large inter-individual variation of roughly 2 kilobases (kb) around the age-mean. Despite a number of well-powered genome-wide association studies (GWAS), very few loci that contribute to this inter-individual variation in TL have been identified. A recent comprehensive meta-analysis of over 26,000 individuals found support for only six loci; these loci could result in a difference in TL of at most 731bp (Pooley *et al.* 2013).

There are a number of ways that classical genomics could fail to detect important sites; if the number of loci is very large and the impact of individual alleles is small, even well-powered studies would fail to notice all but the largest effects. Previous studies did not account for the paternal birth year effect or secular trend identified in Chapter 4, which would result in increased noise in their data, potentially masking weaker effects. Further, if the variance were due to large numbers of gene-environment interactions (GxEs), environmental influences that alter telomere length in different ways depending on an individual's genotype, they may not be detected by classical GWAS studies. Methods of testing for GxEs exist, however they are specific hypotheses driven and therefore not compatible with high-throughput data. Because so little is known about telomere biology, to date only two studies have attempted to evaluate GxEs in the context of telomere length (Eshkoor *et al.* 2013; Mitchell *et al.* 2014), and both of these studies suffered from small

sample sizes and suboptimal PCR-based measurements of telomere length. The heterogeneity in response to Perifosine described in Chapter 3, the very large number of genes enriched in the genome-wide screen described in Chapter 2, and the implications of environmentally-driven changes to population-scale telomere length described in Chapter 4 all led to the hypothesis that a very large fraction of the genome is likely to be involved in GxEs for telomere length. In this chapter, I describe the initial experiments that led to the development of a novel method for exploring GxEs in a large twin database, as well as a number of a proof-of-concept experiments that demonstrate that GxEs are relevant to telomere length (and other phenotypes). These experiments demonstrate that GxEs are indeed common in humans, and how the application of this new method can identify a network of genes that can predict telomere length from age and genotype information and ultimately identify the Rate of Telomere Shortening (RoTS loci).

## **RESULTS**

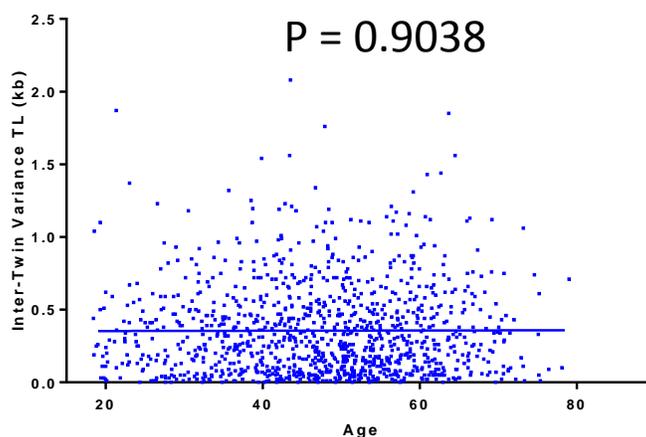
### *Monozygotic but not dizygotic twins diverge in telomere length over time*

As a control for the pseudo-longitudinal telomere shortening assay (Figure 4.4), I looked for an association between the difference in telomere length between twins in the UK Twins cohort and age. If twins diverged over time in terms of telomere length, I would have to correct for this effect in the pseudo-longitudinal shortening assay, and that would have explained why the result of the pseudo-longitudinal analysis yielded a telomere shortening rate substantially higher than predicted by the mediation-adjusted model. While performing this control, I noticed a very interesting difference in the behavior of monozygotic (MZ), genetically identical twins compared with dizygotic (DZ), genetically non-identical twins.

With increasing age, MZ twins exhibit increasing Inter-Twin Variance (ITV) in telomere length, the difference in telomere length between each twin measured. Strikingly, the DZ twins have no significant trend in TL ITV with age, even though there were more DZ twins in the UK Twins dataset. With 1102 DZ twin pairs and 517 MZ twin pairs, DZ twins did not diverge ( $p = 0.9038$ ), while MZ twins diverged by  $1.8 \pm 0.72$  bp per year ( $p = 0.0096$ ,  $r^2 = 0.01295$ ), graphically demonstrated in Figure 5.1.

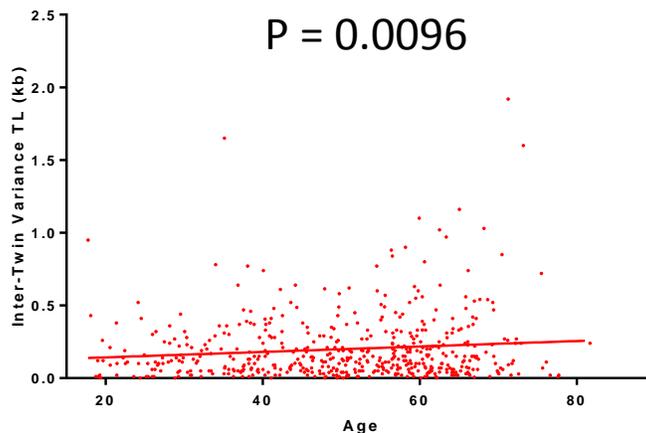
## DZ Twins

Inter-Twin Variance (ITV) in Telomere Length vs Age by zygosity



## MZ Twins

Inter-Twin Variance (ITV) in Telomere Length vs Age by zygosity



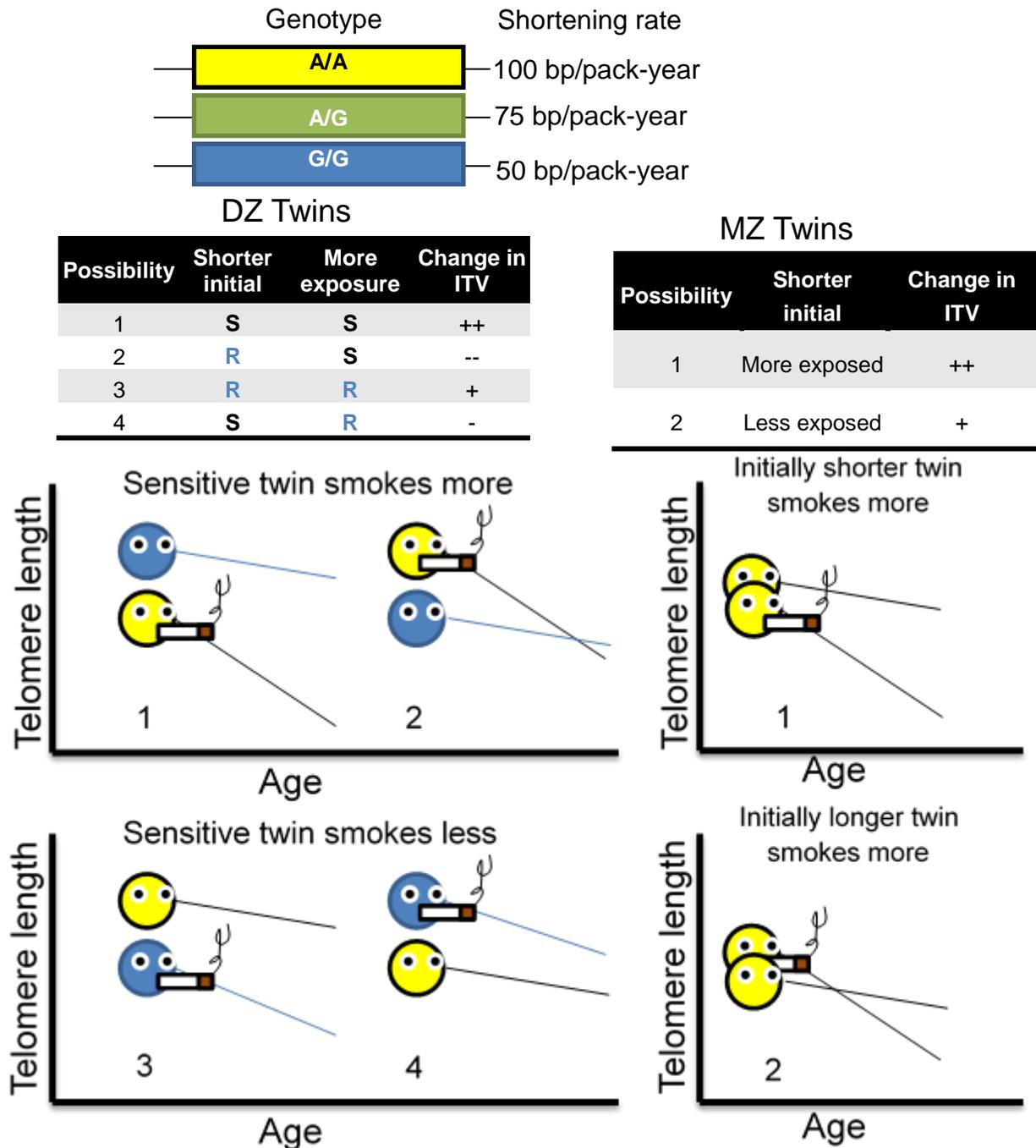
**Figure 5.1: Inter-twin variance in telomere length vs. age**

Dizygotic twins do not get farther apart in terms of telomere length as they age (top), while Monozygotic twins significantly diverge over time (bottom). Inter-twin variance in telomere length in kilobases is shown on the y-axis, while age in years of the twin pair is on the x-axis.

Though the effect observed in MZ twins is small (roughly 5% of the age-associated telomere shortening rate), the sample sizes used here render the chances of observing this effect in the MZ twins but not observing it in DZ twins very small. Considering that the principal difference between a DZ twin-pair and a MZ twin-pair is the fact that MZ twins are genotypically identical whereas DZ twins share 50% of their genome, this difference seemed most likely to arise from some genotype-dependent effect. If environmental variables altered telomere length in a genotype-independent manner, both varieties of twins should diverge as they age and their environments become cumulatively more different. However, if many environmental variables effect telomere length but the direction and magnitude of these affects depended on a large number of underlying genotypes, individuals that do not share a genome would behave randomly in relation to each other as their environments become more divergent; the logic underlying this hypothesis is shown graphically in Figure 5.2.

To demonstrate the logic of this hypothesis, consider a hypothetical gene that controls the rate of telomere shortening in response to smoking. In this model, genotype at a site within this gene controls the rate of telomere shortening in response to smoking, with “sensitive” individuals shortening at 100 base pairs per pack-year (smoking one pack of cigarettes per day for a year), whereas “resistant” individuals shorten 50 base pairs per pack-year. For a DZ twin pair in which one twin is sensitive and one twin is resistant, there are four possibilities for which twin starts with longer telomeres and which twin is more exposed to tobacco. Two of these possibilities (both of the situations where the twin that smokes more starts shorter) increase the difference in telomere length between the twins over time while two of these possibilities (the situations where the twin that starts longer smokes more) decrease the difference over time. For a population of DZ twins evenly

distributed over these four possibilities, the net change in the population ITV in TL over time will be zero, the sum of the four possibilities, together constituting a random walk in the population. Since MZ twins have the same genotype, there are two possibilities; the twin that starts with longer telomere smokes more or the twin that starts with shorter telomere smokes more. Because MZ twins start closer together, both of these scenarios increase the difference in TL with age.



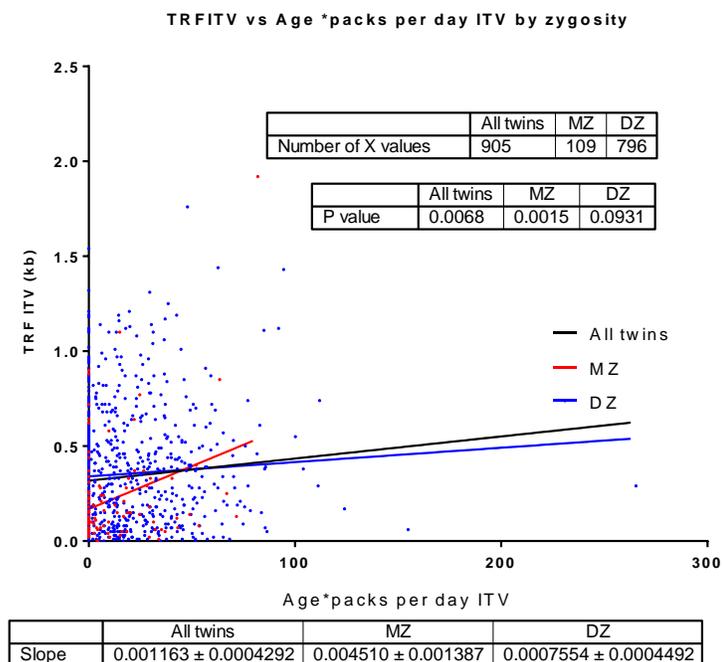
**Figure 5.2: Demonstration how GxEs could generate the difference between MZ and DZ twins**

Sensitive individuals (yellow, S) shorten faster than resistant individuals (blue, R) when exposed to tobacco (top). For DZ twins, the sum four possibilities for distribution of which twin smokes more and starts shorter (labeled as possibility 1-4 in the tables, middle, and the graphs, bottom) yield no change in ITV in TL over time. For MZ twins (middle and bottom right) both of the possibilities increase the ITV in TL over time.

In short, if a given environmental response is genotype dependent, the sum of the four different possibilities for greater sensitivity, greater exposure and initial telomere length is zero for non-identical twins, whereas both possibilities for identical twins result in an increase in the difference in telomere length between each twin with increasing age.

*The telomeric response to environmental stimuli depends on gene-environment interactions*

Because I had access to data on a number of exposures to environmental influences on telomere length from my previous work with the UK Twins cohort described in Chapter 4, I was able to conduct a proof-of-concept experiment by examining how the difference in lifetime tobacco exposure (imperfectly approximated by age\*packs of cigarettes smoked per day) affects the difference in telomere length (Figure 5.3).



**Figure 5.3: Gene-environment interactions control the response to smoking**

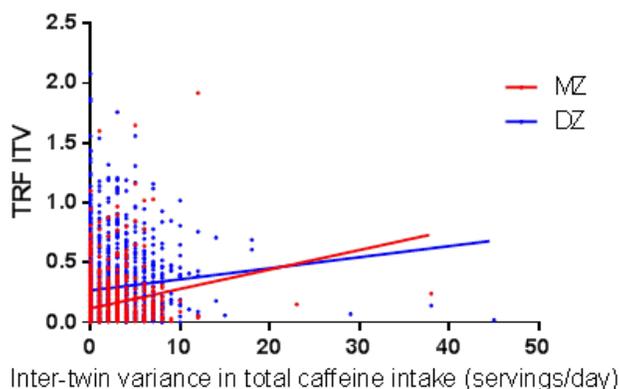
The difference in telomere length between each twin in a twin-pair (TRFITV) is shown on the y-axis and difference in lifetime cumulative tobacco exposure (age\*packs per day ITV) between each twin is shown on the x-axis. MZ twins become divergent in telomere length as they diverge in tobacco exposure much faster than DZ twins. Each point is 1 twin-pair.

Under the model shown in Figure 5.2, MZ twins should get farther apart in telomere length as they get farther apart in tobacco exposure, because they respond the same way to a differing environment, whereas DZ twins should diverge only according to the genotype-independent effects of smoking because they have different responses to differing stimuli.

These results are consistent with the model described in Figure 5.2; MZ twins' telomere lengths get farther apart as the twins' cumulative tobacco exposure becomes increasingly different much faster than DZ twins. In fact, the DZ trend is not significant ( $p=0.0931$ ), while the MZ twin trend is significant ( $p = 0.0015$ ) to such a degree that the trend is significant in all of the twins regardless of zygosity ( $p = 0.0068$ ). A similar trend is observed when examining the effects of different exposure to caffeine on the difference in telomere length (Figure 5.4)

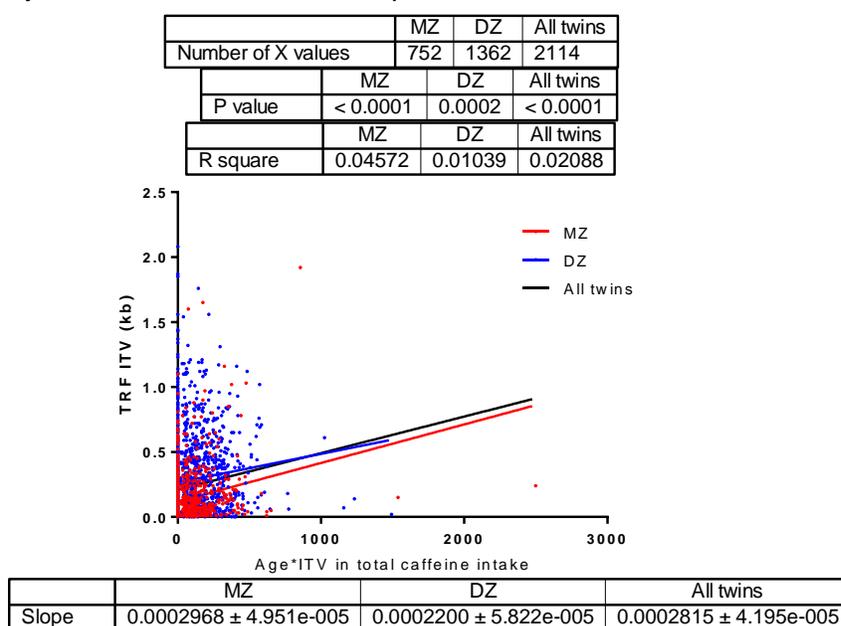
	MZ		DZ	
Slope	0.01628 ± 0.002894		0.009227 ± 0.002656	
R square	0.04047	0.008797	P value	< 0.0001
				0.0005

#### ITV in telomere length vs. ITV in caffeine intake



**Figure 5.4: Gene-environment interactions control the response to caffeine intake**  
The difference in telomere length between each twin in a twin-pair (TRFITV) is shown on the y-axis and difference in caffeine intake between each twin is shown on the x-axis. MZ twins become divergent in telomere length as they diverge in caffeine intake much faster than DZ twins. Each point is 1 twin-pair.

A difference in caffeine intake has a more significant effect on the difference in telomere length in MZ twins than it does in DZ twins (MZ  $p < 0.0001$ , DZ  $p = 0.0005$ ) and this difference explains a greater proportion of the difference in telomere length ITV (MZ  $r^2 = 0.0407$ , DZ = 0.008797), indicating that gene-environment interactions are important to the way that caffeine impacts telomere biology as well, and that the effect observed in cigarette smoke is not an idiosyncrasy of tobacco exposure. In this case, instantaneous caffeine exposure used because there is a strong correlation between age and caffeine intake in this dataset, however performing analysis on approximated exposure (age\* caffeine intake, Figure 5.5) also supports the trend; MZ twins diverge slightly faster and more deterministically with differential caffeine exposure than DZ twins.



**Figure 5.5: Gene-environment interactions control the response to caffeine total exposure**

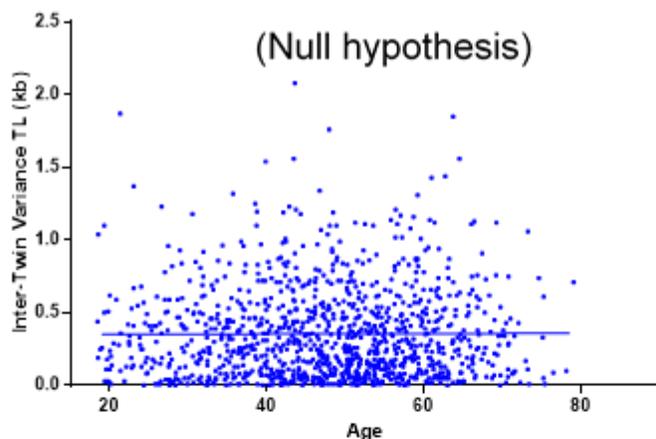
The difference in telomere length between each twin in a twin-pair (TRFITV) is shown on the y-axis and difference in caffeine exposure between each twin is shown on the x-axis. MZ twins become divergent in telomere length as they diverge in caffeine exposure much faster than DZ twins. Each point is 1 twin-pair.

These preliminary analyses indicate that the different relative behavior of MZ and DZ twins could be the result of many different types of gene-environment interactions, and that it is possible to determine if a given environmental stimulus impacts telomere biology by looking for this difference.

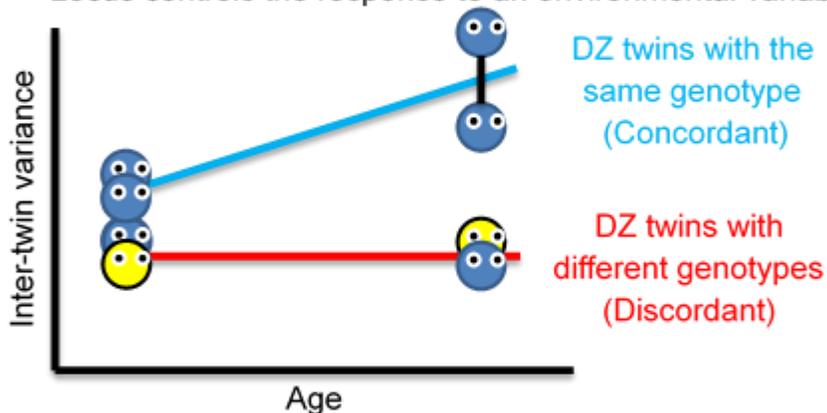
*Surveying the genome for MZ-like behavior in DZ twins can reveal the sites that control telomere dynamics*

From my initial work in DZ twins, there is no association between telomere ITV and age; therefore, if DZ twins that had the same genotype at a given locus in the genome (concordant twins) that controlled the response to some environmental factor, for instance the hypothetical gene examined in Figure 5.2, they should exhibit “monozygotic-like” behavior, an increase in the difference in telomere length between each twin with increasing age because of their cumulatively more different environments. Because of these known trends, one can treat the lack of association between telomere ITV and age in DZ twins as a whole as a null hypothesis, and then subset the DZ twin population into twin-pairs that have the same genotype at each locus (concordant) and twin-pairs that have different genotypes at each locus (discordant) twin-pairs, and look for the signal predicted to arise from a GxE (increasingly different telomere lengths in concordant DZ twins with age). Further, if a locus controlled the intrinsic rate of age-associated telomere shortening, discordant DZ twins should become increasingly different in TL as they age. Both of these predictions are graphically demonstrated in Figure 5.6.

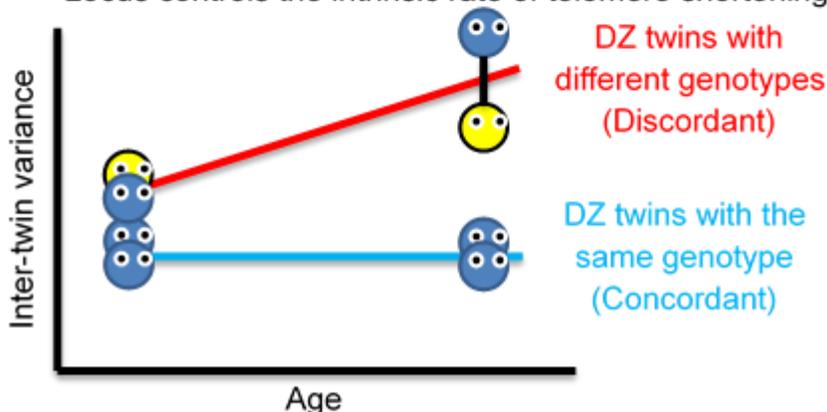
### Inter-Twin Variance (ITV) in Telomere Length vs Age by zygosity



Locus controls the response to an environmental variable



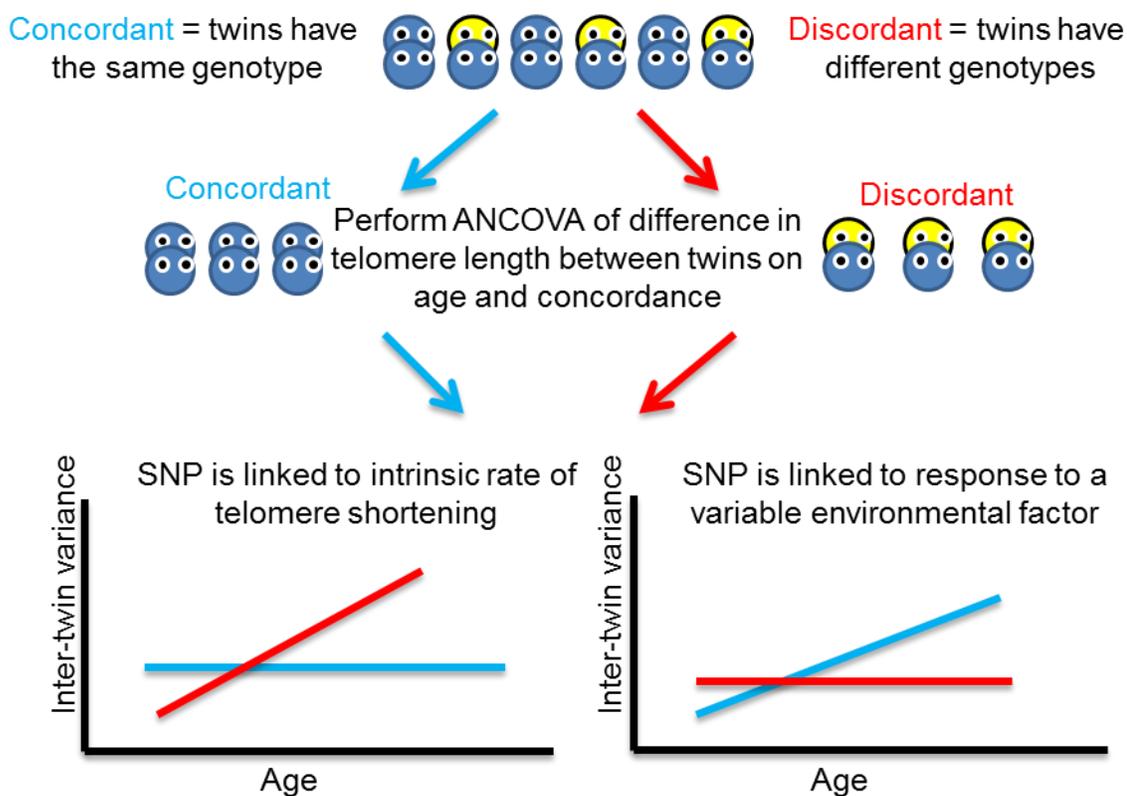
Locus controls the intrinsic rate of telomere shortening



### Figure 5.6: The premise of Twin Variance Analysis

In the DZ twin population as a whole, there is no trend in telomere ITV with increasing age (top). If a site on the genome controlled the response to some environmental variable, DZ twins with the same genotype at that site would become increasingly different in telomere length as they age and their environments become cumulatively more different (middle). If a site controls the intrinsic rate of telomere shortening, DZ twins with different genotypes will become increasingly different as they age (bottom).

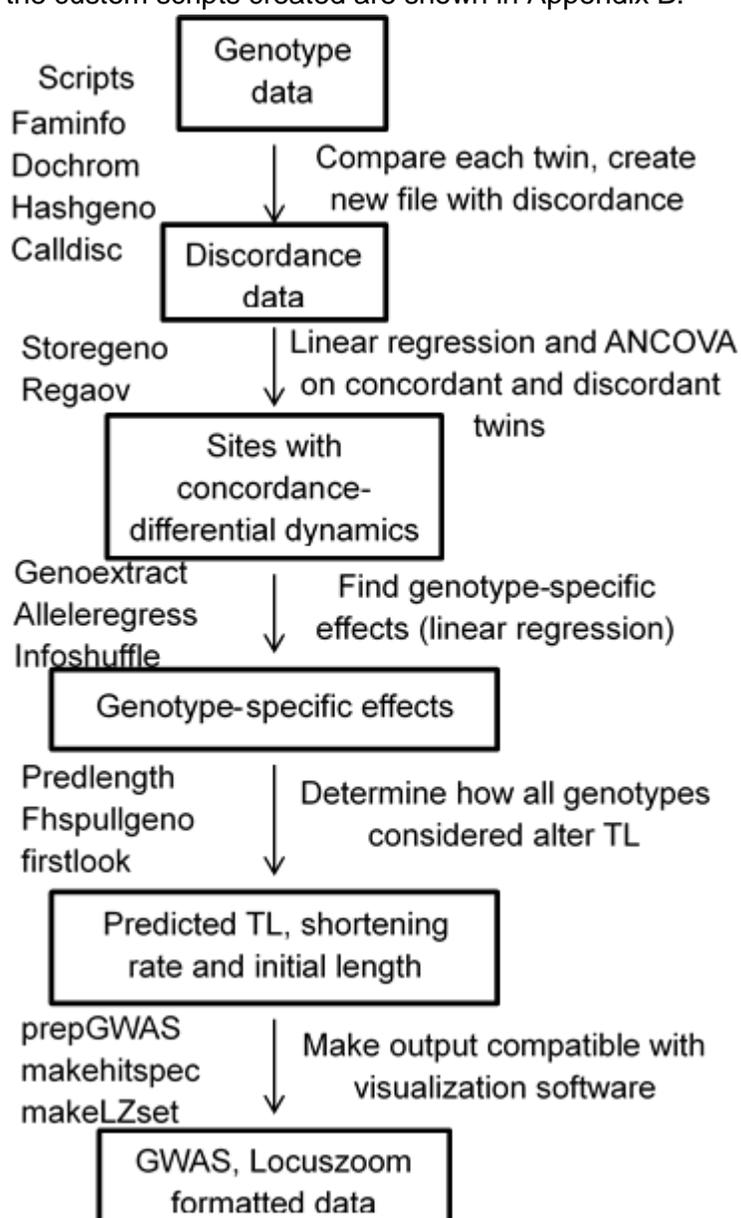
The procedure for evaluating these trends is graphically demonstrated in Figure 5.7. The DZ UK Twins population was flagged as either concordant or discordant at each position measured, and Analysis of Covariance (ANCOVA) was performed to test if the concordant twins and discordant twins had trends with parallel slopes with regard to the difference in telomere length between twins and age. Linear regression was also performed on each set (concordant and discordant) separately to test if the trend in either category of twin was significantly different from zero (the trend observed in the population at large). As this is a new method, I used a very conservative multiple testing correction, the Bonferroni cutoff, as the lower limit for statistical significance.



**Figure 5.7: The procedure used to find loci that modulate the rate of telomere shortening**

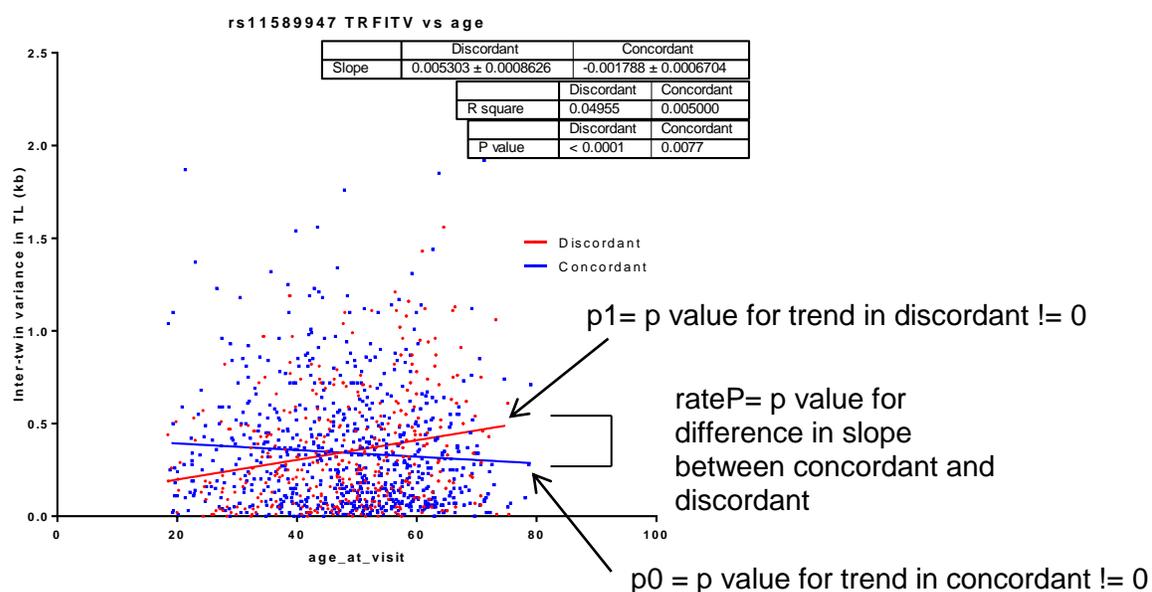
At each genotype measured, the DZ twins in the UK Twins cohort were separated into discordant and concordant twins. ANCOVA was used to test for a difference in TL ITV behavior with relation to age, and linear regression evaluated if concordance or discordance at that locus caused a deviation from the trend observed in the whole population.

In addition to a very large amount of data on behavior, environmental exposures and health outcomes, the UK Twins has genotyped their cohort, yielding measured and imputed information about genotype at over 2,000,000 Single Nucleotide Polymorphisms (SNPs) with minor allele frequency > 5% throughout the genome. I was able to access this information to perform this analysis; a flow chart for the computational analysis is shown in Figure 5.8, and the custom scripts created are shown in Appendix B.



**Figure 5.8: Flow chart of the procedure used**  
Custom R scripts used at each step of the analysis are shown along with relevant outputs.

This experiment is actually testing three different null hypotheses at each locus, shown from an example result in Figure 5.9. First, it tests if the concordant and discordant twins have the same relationship between telomere length ITV and age (recorded as rateP). Next, it tests if the discordant twins have a trend significantly different from zero ( $p_1$ ), and if the concordant twins have a trend significantly different from zero ( $p_0$ ). RateP is more sensitive than either of the two other tests and less subject to biases from small numbers of discordant twins (discussed below and in Figure 5.10), so it was used as the primary output.

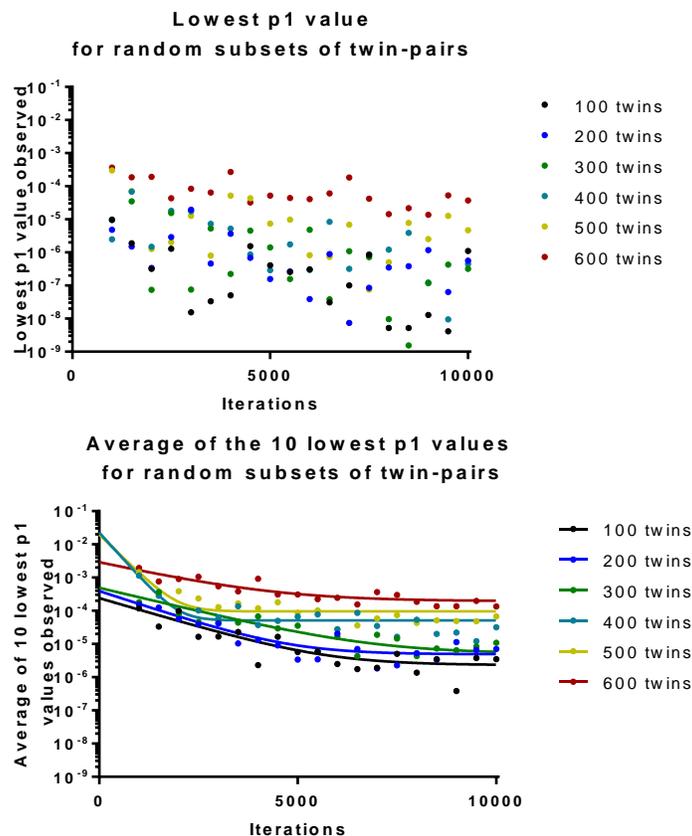


**Figure 5.9: Example output from the Twin Variance Analysis**

The test evaluated if the difference in TRF ITV on age trend in concordant and discordant twins was different from zero ( $p_0$  and  $p_1$ , respectively), as well as if the two types of twins were behaving in the same way (rateP). Data shown is from one of the more significant SNPs, discussed below.

At all sites considered there was a greater number of concordant twin pairs than discordant twin pairs, and sites with very small numbers of discordant twins often exhibited very low  $p$ -values that were potentially artifactual. Examining this in more detail (Figure 5.10) indicated that this is the result of sampling a finite population without replacement; up to a certain point, inclusion of more of the population in the discordant subset reduced the

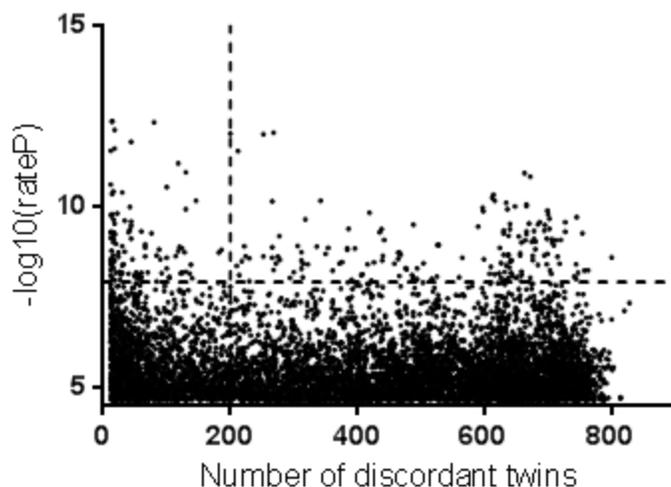
odds of obtaining highly significant results by chance. To quantify this point, I used a Monte Carlo-based approach, wherein a script used the same approach as the twin variance analysis on subsets of different numbers of twin-pairs selected randomly. Figure 5.10 shows that the lowest p1-value (the ability to reject the null hypothesis that there is no trend in TRFITV on age) obtained from random subsets of twins becomes more significant in smaller subsets of twins but reaches a plateau value after many iterations (number of random subsets of that size) that is similar for subsets of 100, 200 and 300 twin-pairs. In fact, the average of the 10 lowest p1 values obtained approaches a plateau value tightly fitting an exponential decay equation, and the plateau values for random subsets of 100, 200 and 300 twin-pair subsets are  $2.31 \times 10^{-6}$ ,  $4.89 \times 10^{-6}$  and  $5.14 \times 10^{-6}$ , indicating that they are roughly similar.



**Figure 5.10: Bootstrapping analysis of the bias introduced by small numbers of discordant twins**

The lowest p1 value observed in random subsets of twin-pairs of varying sizes (100 to 600) in a given number of random subsets (1000 to 10000) is roughly similar for subsets of 100, 200 and 300 twin-pairs (top). The average of the 10 lowest observed p1 values approaches a plateau value following exponential decay ( $r^2$  for 100 twin-pairs = 0.7587, 200 = 0.965, 300 = 0.6616) as the number of iterations increases that is in roughly the same place in subsets of 100, 200 and 300 twin-pairs (bottom).

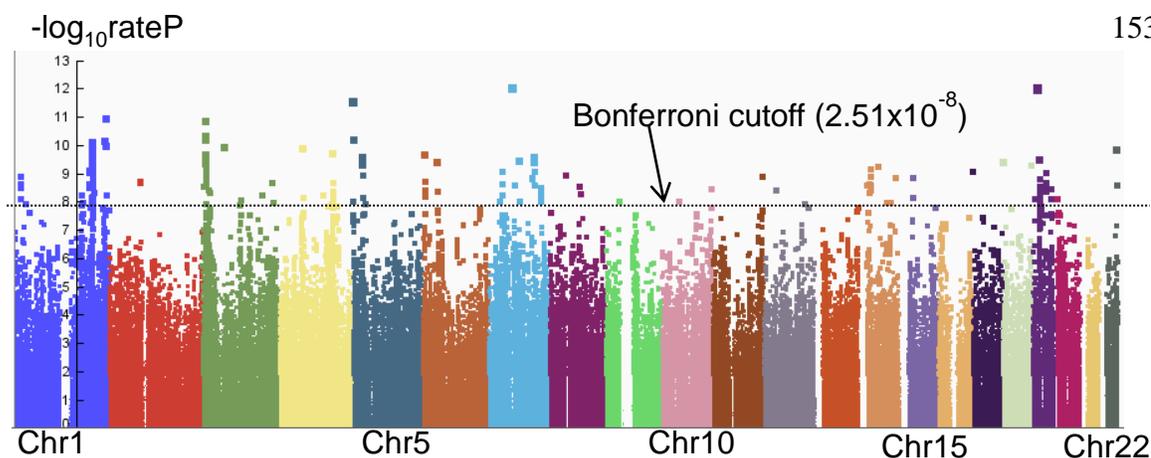
Figure 5.11 shows the rateP values from the 10,000 most significant SNPs evaluated before filtering compared to the number of discordant twins at that SNP. SNPs with less than 200 discordant twins (100 discordant twin-pairs) were not considered in later analyses to remove bias from small sample sizes, a conservative cutoff considering that the inflated significance appears to vanish at roughly half of that value (Figure 5.11).



**Figure 5.11:  $-\log_{10}$  transformed rateP values for the 10,000 most significant SNPs before filtering**

Negative  $\log_{10}$ -transformed rateP values (for example,  $13 = 1 \cdot 10^{-13}$ ) for the 10,000 most significant SNPs from the twin variance analysis are shown on the y-axis, while the number of discordant twins at that SNP are shown on the x-axis. The dashed line perpendicular to the x-axis is the cutoff used for later analysis that should exclude any observed inflation from small sample sizes. The dashed line perpendicular to the y-axis is the Bonferonni cutoff used for multiple testing correction, samples above this value were considered genome-wide significant.

After filtering out all results from SNPs with less than 100 discordant twin-pairs, 92 loci had genome-wide significant rateP values, and many of these loci included more than one genome-wide significant SNP. Figure 5.12 shows the  $-\log_{10}$  transformed rateP values in a Manhattan plot.

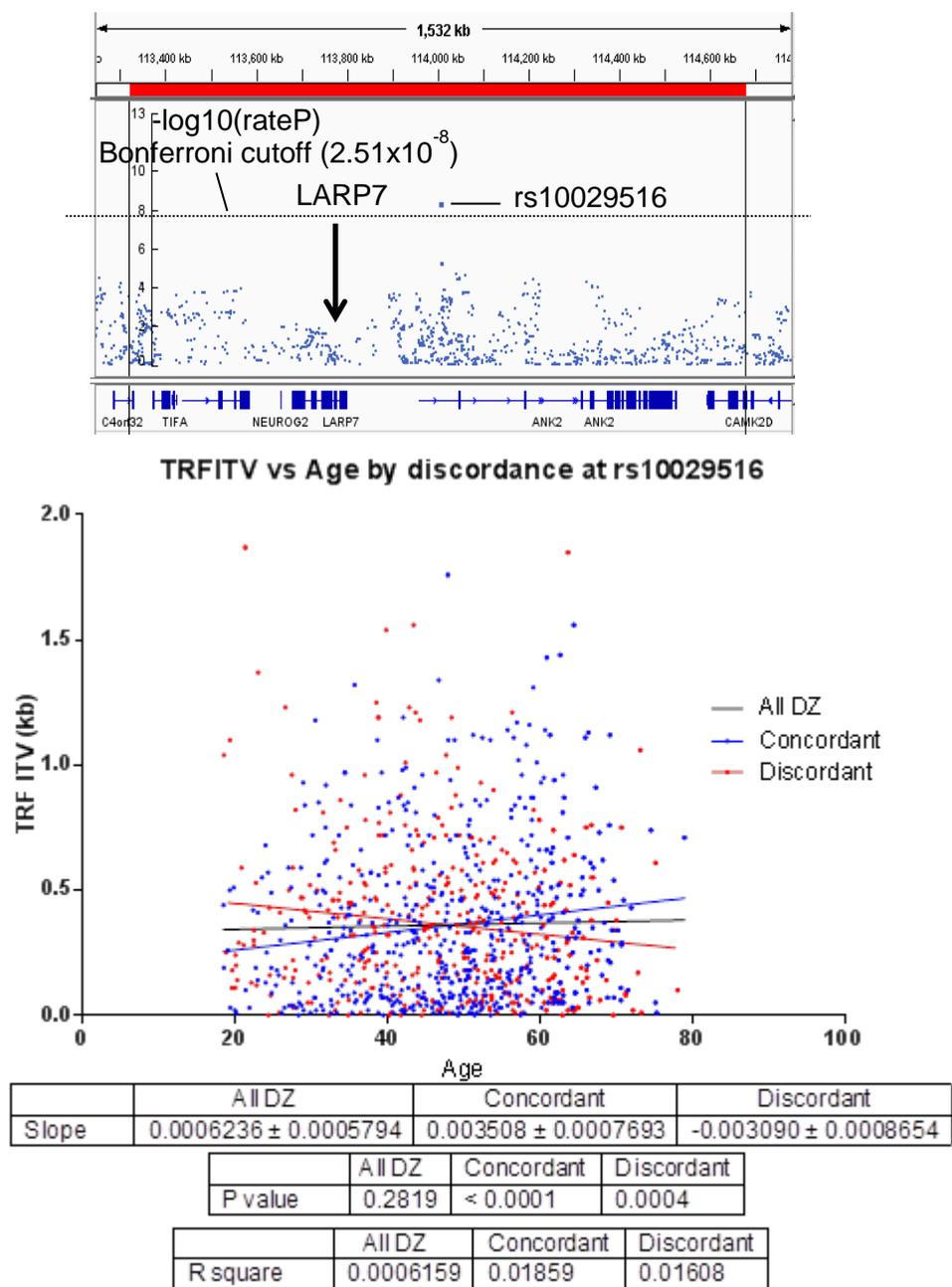


**Figure 5.12: Manhattan plot of the results of the twin variance analysis**

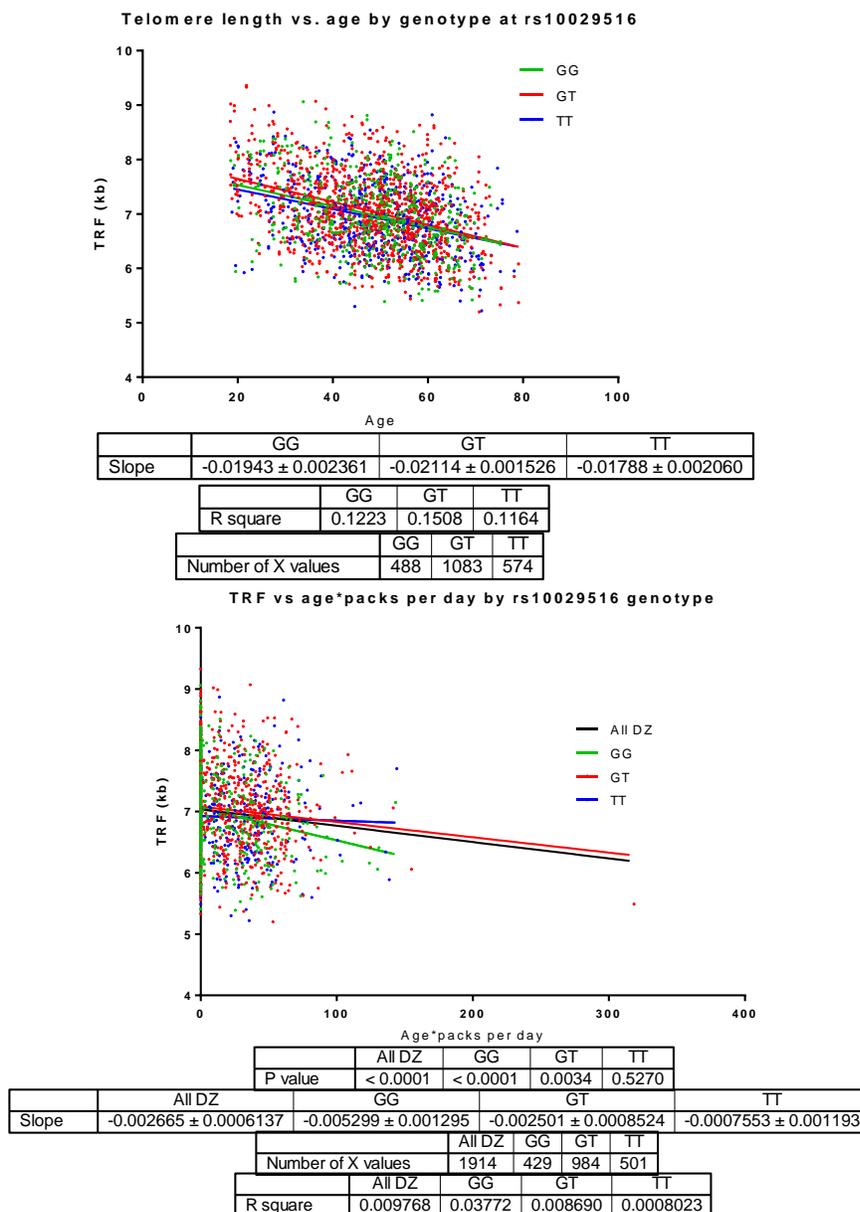
Negative log<sub>10</sub>-transformed rateP values for all SNPs with more than 100 discordant twin-pairs are shown on the y-axis, while their genomic coordinates are on the x-axis. Each chromosome is colored differently, and they are arranged in ascending order.

*The environmental response Rate of Telomere Shortening locus near LARP7 may alter the response to smoking*

Of the 92 loci identified, 44 were consistent with a signal resulting from the locus controlling the rate of telomere shortening in response to some unknown environmental factor; concordant DZ twins' telomere lengths diverged with increasing age. One such SNP, rs10029516, is within 200 kilobases of LARP7, one of the hits described in Chapter 2 (Figure 5.13). Concordant DZ twins at this locus appear to diverge by 3.498 base pairs per year, while discordant twins converge by roughly the same amount. This analysis uses age as a proxy variable for all possible environmental variables, so it was not immediately clear what was driving this effect. However, since I had data available on smoking, an environmental variable with one of the strongest known effect on telomere length (Valdes *et al.* 2005), I tested to see if individuals with different genotypes at rs10029516 were differentially sensitive to tobacco exposure (Figure 5.14).



**Figure 5.13: TRFITV vs age by discordance at a RoTS locus near LARP7**  
 TRF ITV is shown on the y-axis, while age of the twin-pair is on the x-axis. Concordant twins get farther apart with increasing age, while discordant twins get closer together.



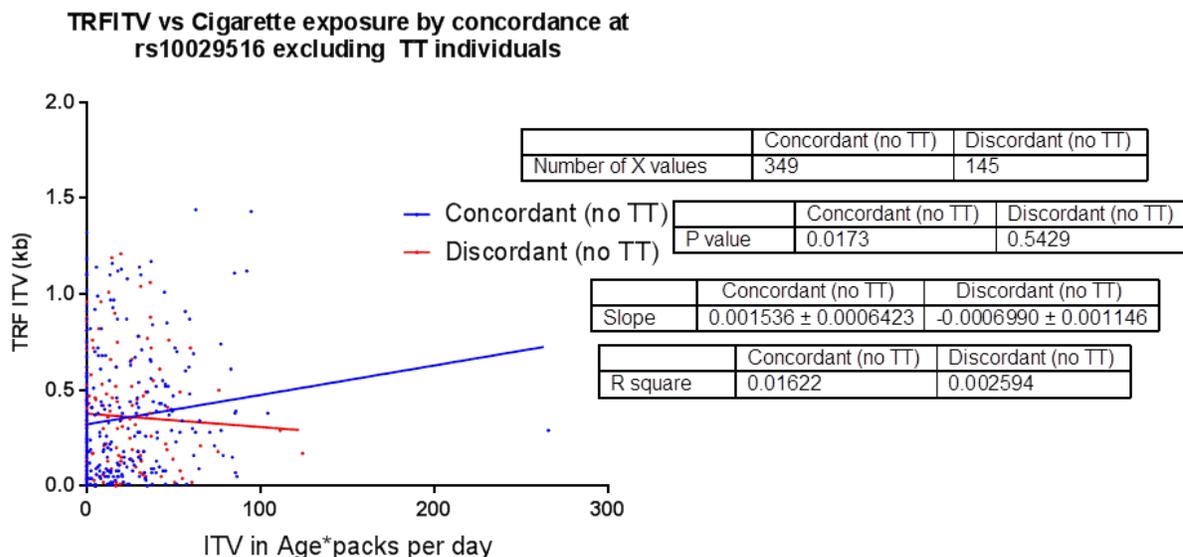
**Figure 5.14: Telomere dynamics by genotype at rs10029516 and tobacco exposure**

Small differences in cross-sectional age-associated telomere shortening rate are observed in individuals with different genotypes at rs10029516 (top), however a much more striking difference is observed in telomere length compared with tobacco exposure (bottom), wherein T allele carriers are allele dose-dependently resistant to tobacco exposure.

The two possible alleles at rs10029516, T and G, are in near-perfect Hardy-Weinberg equilibrium, and this SNP has not been associated with other diseases in the

past. The cross-sectional telomere shortening rates observed in individuals with different genotypes at rs10029516 were slightly different (Figure 5.14); however the discordance data (Figure 5.13) indicates that this is likely to be driven by external stimuli because twins with the same genotype diverge with age. T allele carriers (GT and TT individuals) were less sensitive than GG homozygotes to tobacco exposure (Figure 5.14) to a very dramatic extent; the effect of tobacco exposure on telomere length was not significant in TT individuals, while it was significant at  $p < 0.0001$  in GG individuals and the DZ population as a whole. Further, GT individuals were sensitive to tobacco exposure to an intermediate extent between the GG and TT individuals, suggesting that this is a gene dose-dependent effect. Lastly, the cross-sectional telomere shortening rate was lower in TT individuals than the rest of the population, and resistance to the telomere length effects of cigarette smoke may explain this difference. To further test if cigarette smoking may explain this signal, I examined the difference in telomere length between concordant and discordant DZ twins and the difference in tobacco exposure (Figure 5.15).

ITV in TL became increasingly larger as the concordant twins became more divergent in cigarette exposure while there was not a significant trend in discordant twins, providing strong evidence that this locus is important for the telomeric response to cigarette exposure. However, the genotype-specific analysis (Figure 5.14) indicates that TT individuals are almost fully resistant to cigarette smoke, so they would logically not get farther apart as they become increasingly different in smoking behavior since they are insensitive. If that were the case, excluding all twin-pairs that include a TT individual should make these trends more significant, since that would remove a population that was not following them. As predicted, removing TT individuals increases the significance of the trend in increasing TRF ITV with increasingly divergent smoking behavior ( $r^2=0.003563$  in all concordant,  $r^2=0.01622$  in concordant non-TT, Figure 5.16). This observation simultaneously supports the idea that rs10029516 may control a response to smoking and illuminates a source of false negatives inherent this method.

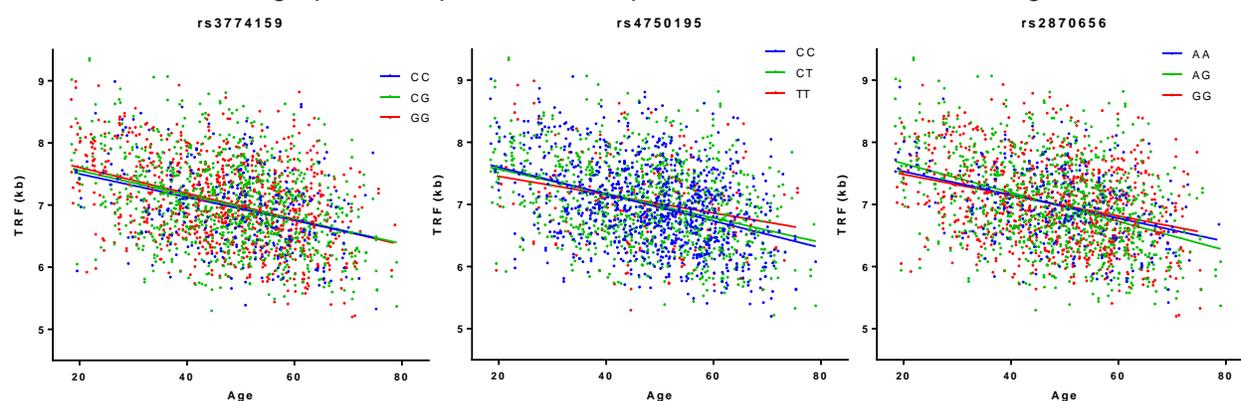


**Figure 5.16: TRF ITV by discordance at rs10029516 and cigarette exposure excluding TT individuals**

TRF ITV (y-axis) increases in concordant twins that do not have the TT genotype in response to differential tobacco exposure (x-axis) more significantly than in all concordant twins.

*The intrinsic Rate of Telomere Shortening loci can be used to predict rate of telomere shortening and telomere length*

The 48 intrinsic loci, SNPs where discordant DZ twins become increasingly different in telomere length over time, should logically be usable to predict telomere length, and individuals with different genotypes at these loci should shorten at different rates. To test this hypothesis, I determined the cross-sectional telomere shortening rate and predicted initial telomere length (y-intercept of the cross-sectional analysis) for each genotype at each of the RoTS loci, graphical output for three representative loci is shown in Figure 5.17.



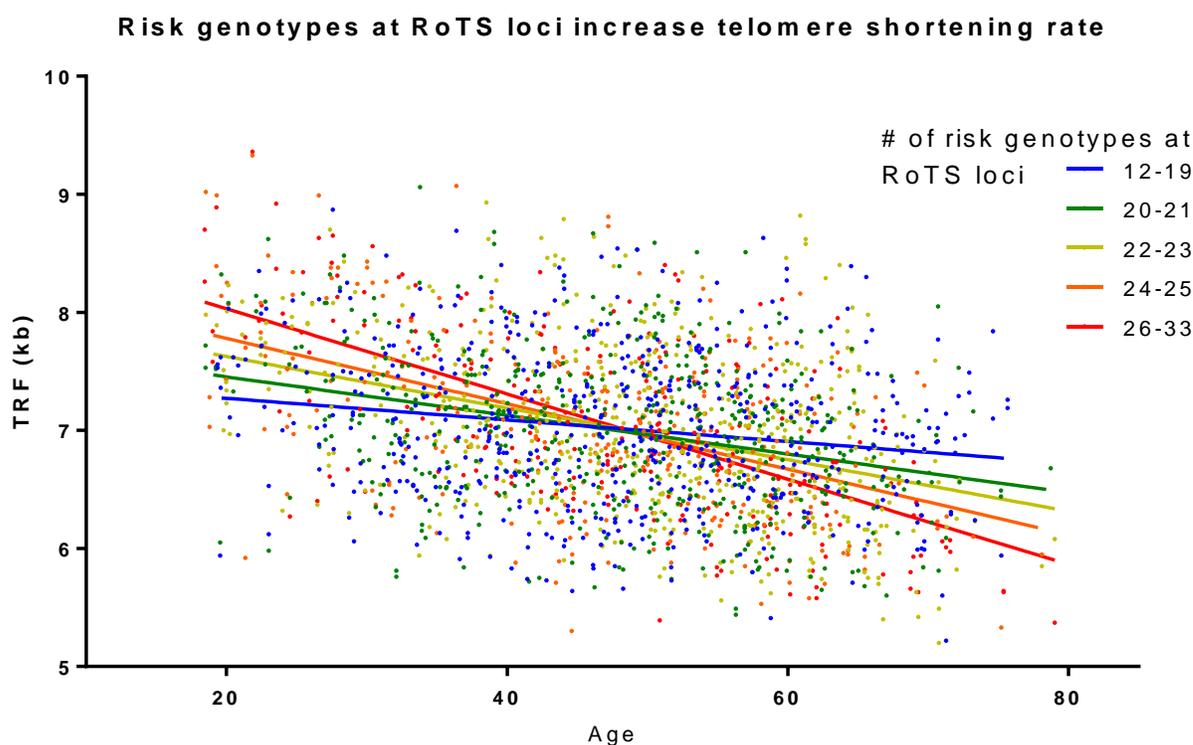
Sample size				Shortening Rate			
rsID	nAA	nAB	nBB	rsID	AA(bp/yr)	AB(bp/yr)	BB(bp/yr)
rs3774159	238	868	768	rs3774159	-18.51	-19.92	-22.81
rs2870656	236	864	745	rs2870656	-16.02	-25.06	-17.07
rs4750195	846	856	181	rs4750195	-23.22	-20.14	-12.55

rsID	Nearest gene	Site	Descriptor
rs3774159	ATP2B2	Intronic	Calcium pump
rs2870656	GRID2	Intronic	Glutamate receptor
rs4750195	PFKFB3	Intergenic	Glucose met., oncogene

**Figure 5.17: Three example intrinsic RoTS loci**

Telomere dynamics for individuals with different genotypes are shown in the three graphs (top) and summarized in the accompanying tables. Alleles coded A and B are C and G for rs3774159, C and T for rs4750195, and A and G for rs2870656.

To perform an initial test if these loci can together predict telomere shortening rate and telomere length, I quantitated the number of genotypes each individual in the UK Twins cohort had that shortened faster than the other two possible genotypes at that locus (the “risk” genotype), and compared how individuals with more risk genotypes behaved cross-sectionally compared to individuals with fewer risk genotypes (Figure 5.18).



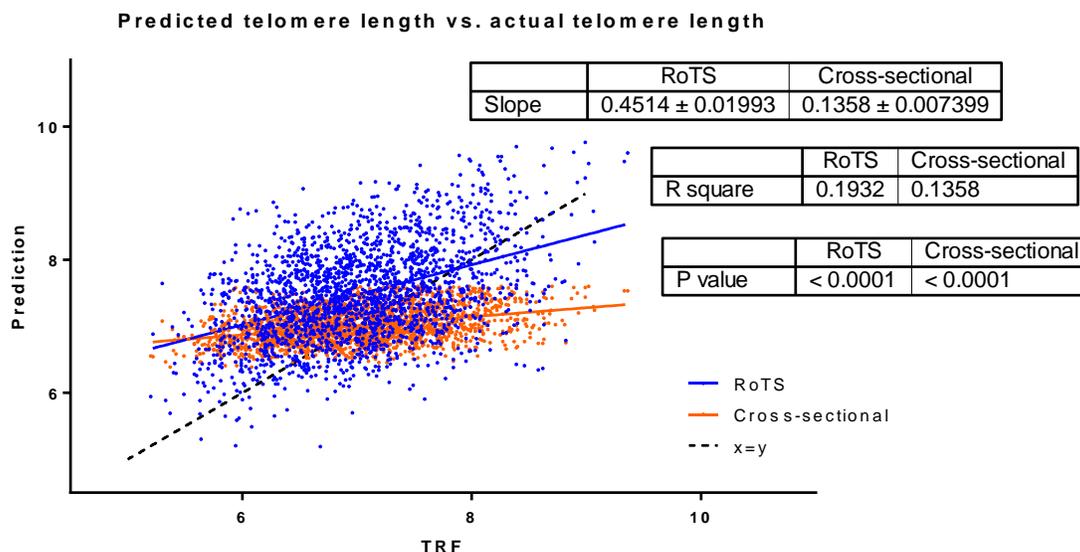
# genotypes	<20	20-21	22-23	24-25	>25
Slope (kb/yr)	-0.009160 ±	-0.01634 ±	-0.02185 ±	-0.02772 ±	-0.03610 ±
N	538	501	534	322	250
P value	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

**Figure 5.18: Individuals with more risk genotypes exhibit faster telomere shortening**  
Telomere length (y-axis, top) decreases faster cross-sectionally with age (x-axis) in individuals with a greater number of risk genotypes at RoTS loci.

As predicted, individuals with many risk genotypes shortened faster over time than individuals with fewer risk genotypes by more than a factor of three. Unexpectedly, an increased telomere shortening rate in individuals with many risk genotypes was

accompanied by an upward change in initial telomere length; the y-intercept of the population with many risk genotypes is higher than that of individuals with fewer risk genotypes. This insinuates that at least for these loci, initial telomere length and telomere shortening rate are intrinsically tied, a phenomenon I will address in greater depth in the discussion.

Next, I attempted to discern if genotype at these loci can predict telomere length as well as shortening rate and initial telomere length, and if these predictions can be extended to another population dataset from which the genotype-specific effects are not derived. I used the genotype quantitation performed for Figure 5.18 to calculate a predicted telomere shortening rate and predicted initial length for each person in the UK Twins cohort based on their genotypes at the intrinsic RoTS loci assuming that the effects of each allele were fully additive. This assumption is probably not fully correct, however as a first attempt I had no reason to assume other behavior. The difference between the cross-sectional shortening rate and genotype-specific telomere shortening rate was quantified for each genotype at these loci, as well as the difference between the genotype-specific y-intercept and the cross-sectional y-intercept. Each genotype's effect was added up for each locus in all individuals, excluding genotypes with less than 100 individuals, yielding a predicted initial telomere length and predicted telomere shortening rate. These were then used to predict telomere length at the age at which the individuals were measured (Figure 5.19).



**Figure 5.19: Telomere length predicted by the RoTS loci compared with observed telomere length in the UK Twins cohort**

Telomere length predicted by the RoTS loci, on the y-axis, is compared with observed telomere length (x-axis) for each individual. The RoTS loci's predictions (blue) are generally closer to reality (a perfect prediction would follow the  $x=y$  dashed line) than a prediction made only from the cross-sectional shortening rate and initial length (orange).

The predictions made from genotype at the RoTS loci and age were much closer to reality than a prediction made only from the cross-sectional data; the slope of the relationship between predicted telomere length and observed telomere length was 0.4514 for the RoTS loci compared with 0.1358 for the cross-sectional prediction, where a value of 1 would indicate a perfect relationship between prediction and reality. Further, observed telomere length explained a greater proportion of the variance in predicted telomere length for the RoTS loci than the cross-sectional prediction ( $r^2 = 0.1932$  for RoTS and 0.1358 for cross-sectional), indicating that at least within this dataset, genotype at these loci are considerably predictive of telomere length and telomere dynamics.

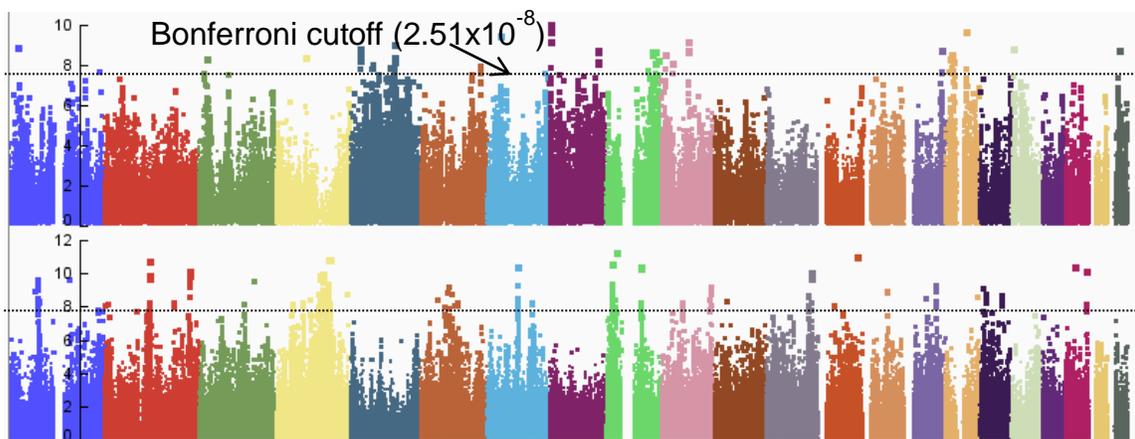
While this result was encouraging, it was also somewhat inevitable as the values used to predict telomere length in this population were in fact derived from this population; this result could be considered the final output of a very complex fit of the data rather than a

test of the predictive power of the loci. In order to test if these effects represent a biological reality rather than a statistical fluke within a single population, I contacted Steve Hunt, who I worked with on the experiments described in Chapter 4. I was able to obtain access to the genotype data available for the Family Heart Study, which has telomere length information as well as data on a number of environmental and behavioral risk factors for cardiovascular disease. Though this is not a twin dataset, I can use it to determine if the effect sizes derived from these analyses in the UK Twins can be used to predict telomere length in a different population.

*Twin Variance Analysis on permuted phenotypes yields information about false discovery rate*

A standard control in GWAS studies compares the observed results' significance against calculated expected significance values based on the sample sizes used and the distribution of the dependent variable in question. However, since this is not a standard GWAS method, I could not use these methods to determine what the expected significance values were for this data using the same methods as a GWAS study. In order to determine if these results were more significant than what I would expect to see if there were not signals to find, I performed a control analysis using the method described in Figure 5.8, but instead of using TRF ITV and Age, I randomly assigned each twin pair TRF ITV and age information from another twin pair without replacement, such that the distribution of TRFITV:Age pairs did not change, but any underlying biological information was removed by the data shuffling (accomplished with the Infoshuffle script, Appendix B and Figure 5.8). These shuffled TRFITV:Age pairs (three different shuffles with different random seeds were used) are referred to in later analyses as the permuted controls. Performing twin variance

analysis with the same method as with the real TRFITV and age data on the permuted controls yielded a number of loci that would be considered genome-wide significant using the Bonferonni correction (Figure 5.20), indicating that it is possible that this method's false discovery rate is nonzero.

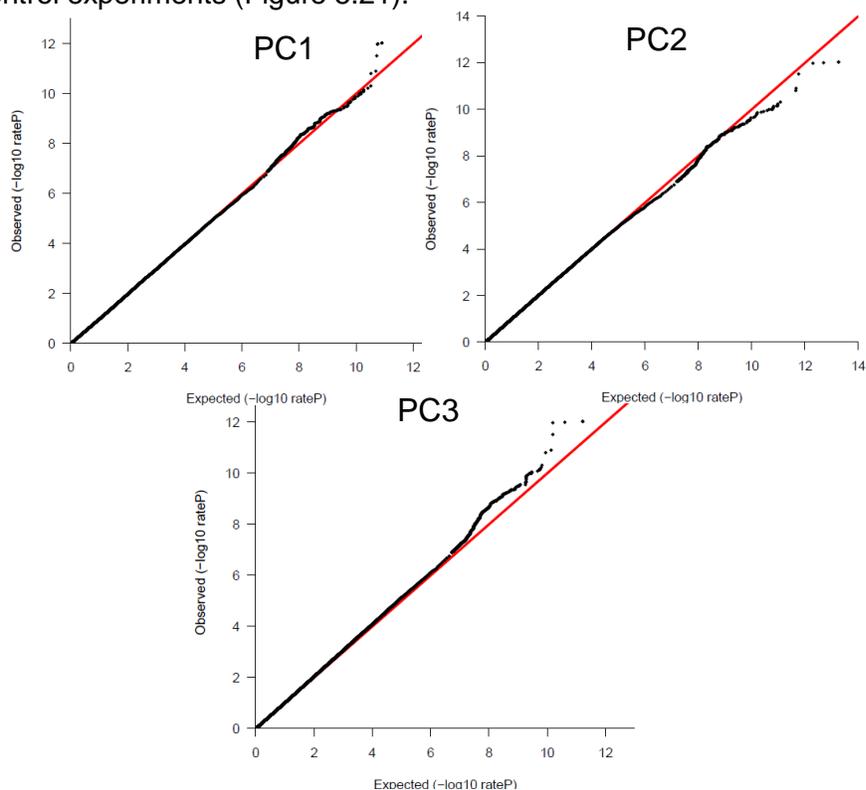


**Figure 5.20: Manhattan plots of the results of the twin variance analysis on permuted phenotypes**

Negative log<sub>10</sub>-transformed rateP values for all SNPs with more than 100 discordant twin-pairs are shown on the y-axis, while their genomic coordinates are on the x-axis. Each chromosome is colored differently, and they are arranged in ascending order (left to right, chromosome 1 to chromosome 22). Sites with diverging discordant twins are shown on top, sites with diverging concordant twins are shown on the bottom.

To further evaluate the results of the permuted control analyses, I generated a quantile-quantile (qq) plot, which is a test for large-scale inflation or deflation of significance values in an observed data set compared with an expected distribution. For this test, I used the results of the analysis on the real TRFITV and age data as the observed data set, and each of the three permuted control analyses as an expected distribution. The qq plots show a rank-ordered list of the rateP values in the observed distribution as a y-value for each point (the first value in the list is the most significant result, the second value the second most significant, etc.), and a rank-ordered list of the rateP values in the expected distribution as an x-value for each point, such that if a point lies above an x=y line of identity, then that

ranked result was more significant in the observed distribution than in the expected distribution. Thus, if the distribution observed lies above the line of identity, the results in the observed distribution were more significant than in the expected distribution as a whole. Performing this analysis showed that the rateP values derived from analysis of the actual TRFITV and age information were not more significant than what was observed in the permuted control experiments (Figure 5.21).



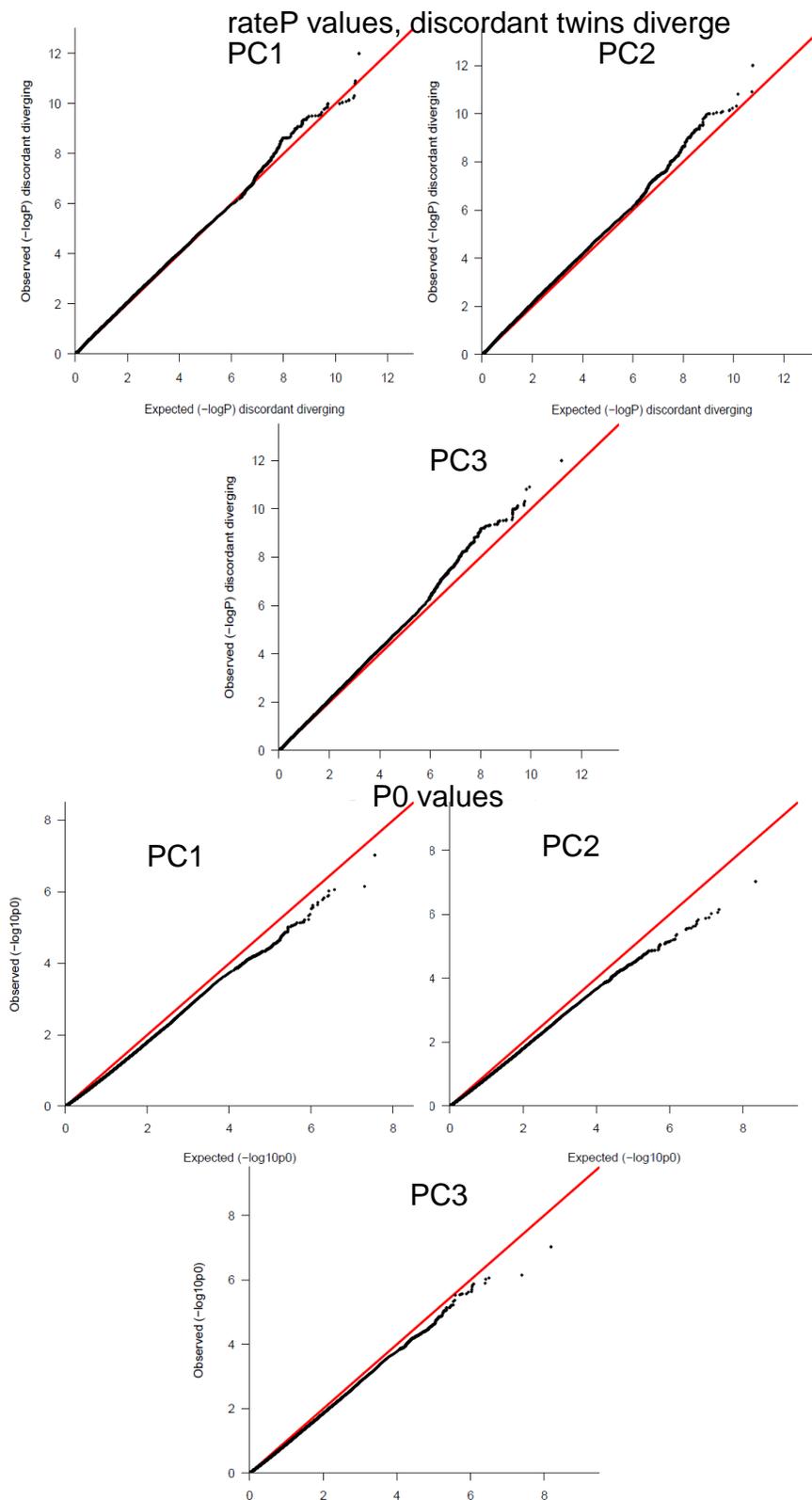
**Figure 5.21: qq plots of the results of rateP values from twin variance analysis compared to the permuted controls**

Rank-ordered  $-\log_{10}$  transformed rateP values from the twin variance analysis are on the y-axis, while rank-ordered  $-\log_{10}$  transformed rateP values from the permuted controls are on the x-axis. PC1 = Permuted control 1, PC2 = Permuted control 2, PC3 = Permuted control 3

The rateP values in the twin variance analysis of the real TRFITV and age information were roughly as significant as those in permuted control 1, slightly less significant than permuted control 2, and slightly more significant than permuted control 3, suggesting that on average the real analysis did not identify more highly significant results

than would be expected if there was not underlying biological information (as shown by the permuted controls, where the underlying information is removed via the shuffling of TRFITV and age values).

A closer comparison of the results from the real information and the permuted controls reveals a number of interesting differences. Examining the rateP values at only those sites where discordant twins get farther apart with age (intrinsic rate of telomere shortening loci) indicates that the real analysis found substantially more significant loci that seem to control the intrinsic rate of telomere shortening than would be expected without underlying information (Figure 5.22, top). Further, there were very substantially fewer loci with significant  $p_0$  values (sites where concordant twins have a nonzero trend in TRFITV on age) in the results compared with the permuted controls, indicating that there is a great deal of genetic control over the rate of intrinsic rate telomere shortening, since twins with the same genotype diverge with increasing age substantially less often than would be expected by chance (Figure 5.22, bottom). It is possible that the reason that the rateP values observed in the analysis of real information are not more significant than in the permuted controls is because there were fewer sites in the real data with diverging concordant twins and more sites in the data with diverging discordant twins, and these two overlapping effects canceled one another out. RateP tests for a significant difference in the behavior of concordant and discordant twins, and it will become less significant if the trend in either kind of twin becomes less significant; this is further supported because the permuted control with the least significant  $p_0$  values, PC3, was also the permuted control that had less significant rateP values compared with the analysis of real data (Figure 5.21 and Figure 5.22).

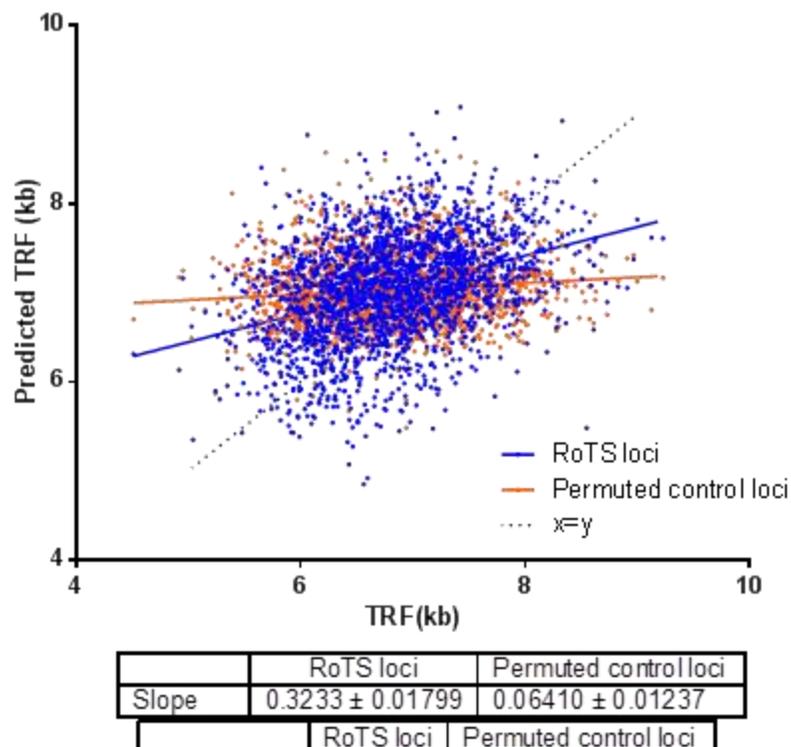


**Figure 5.22: qq plots of specific subsets of the results of twin variance analysis** Rank ordered  $-\log_{10}$  transformed rateP values for SNPs with diverging discordant twins (top) and rank-ordered  $-\log_{10}$  transformed p0 values for all SNPs (bottom) illustrate a great deal more genetic control of telomere shortening rate than expected by chance.

*The telomere length predictions from the RoTS loci replicate in another population*

To further evaluate if the RoTS loci identified represent real information rather than overfitting of the UK Twins data, I tested if the method used to predict telomere length in Figure 5.19 can be applied to the Family Heart Study cohort, using the changes to telomere shortening rate for each intrinsic RoTS genotype identified in the UK Twins cohort. To evaluate if this prediction is better than a prediction made by any other comparable group of SNPs, I compared the telomere predictions made by the RoTS loci to telomere predictions made with the same methods on the intrinsic loci (discordant twins diverge) identified by one of the permuted control analyses (Figure 5.23).

**Predicted telomere length vs. actual telomere length (FHS)**



**Figure 5.23: Predicted telomere length vs. observed telomere length in the Family Heart Study**

Telomere length predicted from age and genotype on the y-axis is compared with observed telomere length on the x-axis for each individual in the Family Heart Study for predictions made with the RoTS loci (blue) and the intrinsic loci identified from the permuted control analysis (orange).

The RoTS loci predicted telomere length substantially better than the loci identified with the same methods on the permuted phenotypes (slope of predicted TRF on observed TRF = 0.3233 in the RoTS predictions, 0.0641 for permuted control,  $r^2 = 0.1312$  for RoTS, 0.01241 for permuted control), indicating that the RoTS loci replicate in another population very substantially better than random loci selected with the same methodology.

## **DISCUSSION**

An interesting result of these experiments is that rate of telomere shortening and initial telomere length are intrinsically linked. This has been evaluated in the literature in the past; one longitudinal study evaluating individuals 20 to 40 years old at the first measurement and 25.7 to 48.2 years old at the second measurement found that individuals with longer telomere shortened faster than individuals with shorter telomeres (Aviv *et al.* 2009), consistent with the observations in this study that found a link between initial telomere length and telomere shortening rate (Figure 5.18). One other prediction from the results in Figure 5.18 is that the association between telomere length and telomere shortening rate should change based on the age range of the sample evaluated—old individuals with long telomeres should shorten slower than old individuals with short telomeres (they started shorter but shortened slower), whereas young individuals with long telomere should shorten faster than young individuals with short telomeres. Another study found that telomere length and telomere shortening rate correlated and that this effect was weakest in older individuals (Ehrlenbach *et al.* 2009). These results provide more evidence that the effect predicted by the results in Figure 5.18 may be correct (that there should be

age-dependence in the relationship between telomere length and telomere shortening rate due to underlying differences in telomere dynamics.)

One report examined four different longitudinal studies to determine if telomere length of an individual relative to other individuals of the same age changes over time (Benetos *et al.* 2013). They concluded that the relative telomere lengths of individuals are nearly fixed, however careful analysis of the data presented in this report suggested that they did not notice an age-dependent difference in the association between initial telomere length and follow-up telomere length (Table 5.1).

	LRC+BHS	LSADT
<b>Initial age</b>	30.30	75.00
<b>Final age</b>	43.00	86.00
<b>Range, +/-</b>	2.22	2.00
<b>Duration</b>	12.82	10.80
<b>m, initial:final</b>	0.92	0.96
<b>N</b>	891	80
<b>LTL</b>	7.30	5.85
<b>LTL range</b>	0.69	0.60
<b>5kb SR(kb/yr)</b>	0.0132	0.0259
<b>6kb SR(kb/yr)</b>	0.0198	0.0296
<b>7kb SR(kb/yr)</b>	0.0264	0.0333
<b>8kb SR(kb/yr)</b>	0.0330	0.0370
<b>9kb SR(kb/yr)</b>	0.0396	0.0407

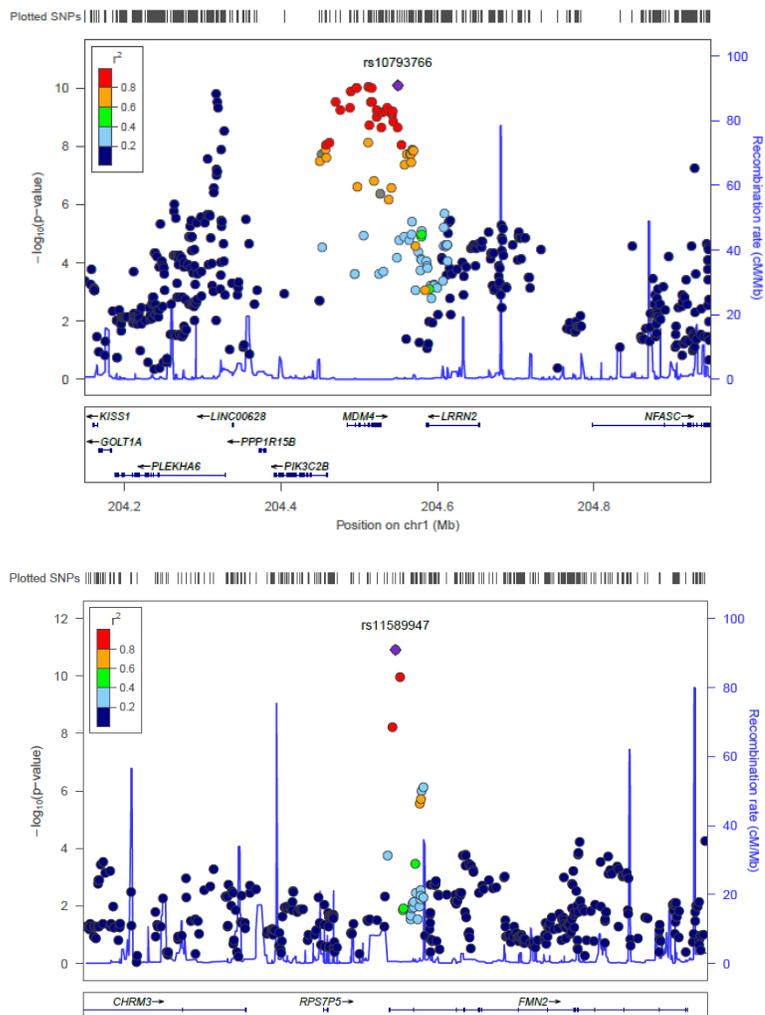
**Table 5.1: Analysis of the relationship between telomere length and telomere shortening rate in the literature**

Study characteristics for longitudinal datasets examining telomere shortening rate with summary statistics (top panel) indicates that the difference in shortening rate between long and short individuals gets less pronounced with age (bottom panel). The bottom panel shows the predicted shortening rate (kb/year) for an individual with telomere length at the start of the sampling period from 5-9 kb in each age range examined.

For example, in this comparative study, the investigators examined four cohorts (the Jerusalem Lipid Research Clinic, LRC, the Bogalusa Health Study, BHS, the Evolution de la

Rigidite Arterielle, ERA and the Longitudinal Study of Aging Danish Twins, LSADT) that spanned three age ranges (early 30's, late/middle age from 48-68, and mid /late 70's). I combined the results from the BHS and LRC in Table 5.1 because they spanned the same age range, and I did not evaluate the ERA because its study population included a much wider age range (58 years mean, +/- 10 years) and the unexplainable y-intercept (0.24, indicating that an individual with 0kb telomeres at the initial measurement could be expected to lengthen by 240 bp) in the data about the relationship between telomere length and shortening rate that make the data less interpretable. Using the effects reported in the paper (relationship between initial telomere measurement, m initial:final, y-intercept and the duration of the study) I calculated the change in telomere length per year for a hypothetical individual of 5 through 9 kilobases at the first measurement in each cohort. Comparing young individuals (BHS+LRC) to old individuals (LSADT), the difference in telomere shortening rate in short individuals and long individuals is much less pronounced, 26.4 bp/year faster shortening in a 9kb individual compared to a 5 kb individual in the young group, compared with a 14.8 bp/year difference in the old cohort.

The age-dependence in the relationship between telomere length and telomere shortening rate shown by the twin variance analysis which also finds support in the literature may inform on the biology underlying the intrinsic rate of telomere shortening loci. Indeed, though the causal variant at each locus has not been evaluated, recombination frequencies relative to the locations of the peaks give very strong hints about certain loci, for instance the MDM4 and FMN2 intrinsic RoTS loci (Figure 5.24), which strongly implicate those genes as likely to harbor the causal variant.



**Figure 5.24: Intrinsic RoTS loci over MDM4 and FMN2**

$-\log_{10}$ -transformed rateP values are shown on the left y-axis, while recombination rates in centimorgans per megabase (HAPMAPII) are shown in blue and on the right y-axis. Genomic coordinates are on the x-axis, and linkage to the index SNP (labeled) is shown for each SNP by color (redder=more linked).

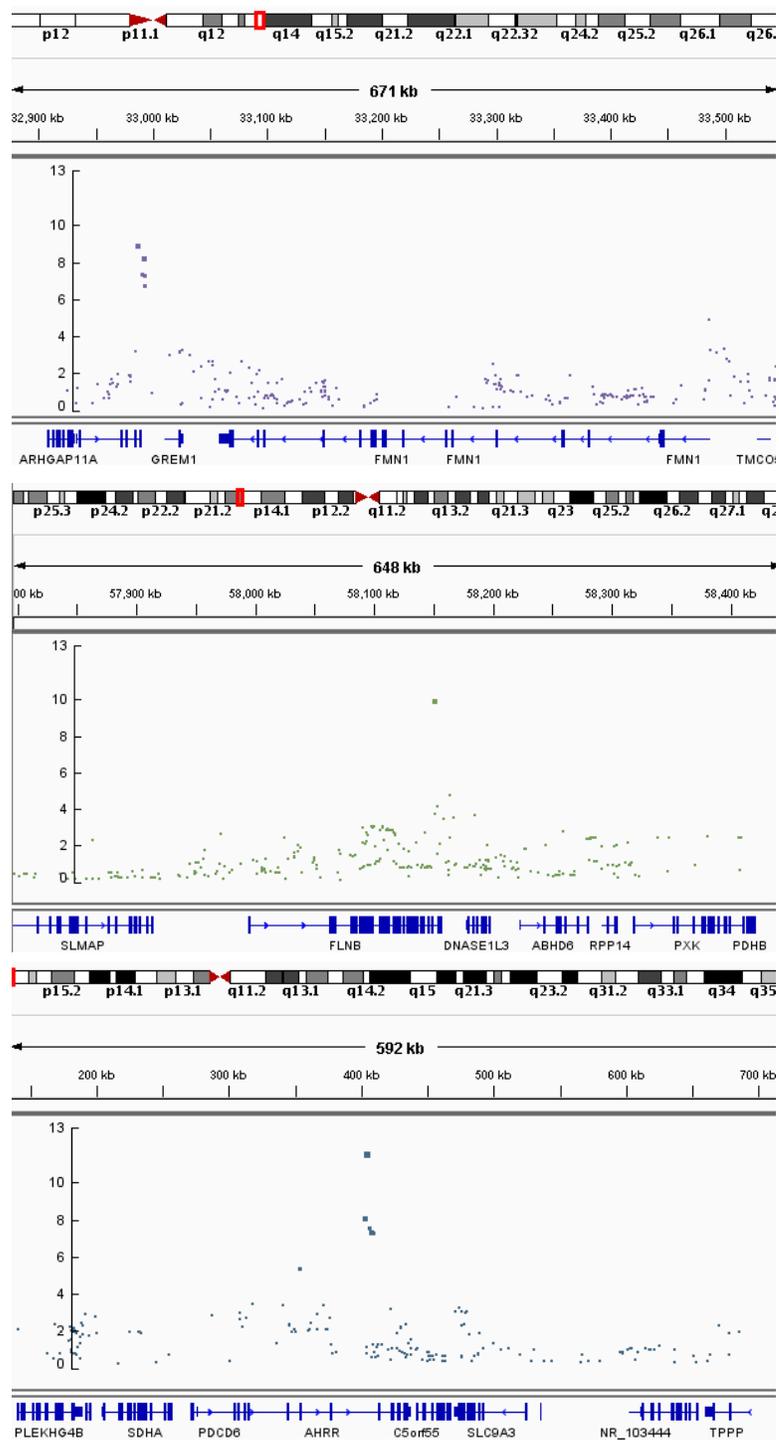
MDM4 modulates p53 activity in a manner similar to MDM2 and controls stem cell fate and differentiation in Embryonic Stem (ES) cells (Mendrysa *et al.* 2011; Menendez *et al.* 2011), and it can also trigger p53 activity in response to stress signaled by defects in splicing activity (Bezzi *et al.* 2013). Though MDM4 is typically studied in the context of

neuronal and developmental cell death, recent studies have shown that MDM4/MDM2/p53-mediated death of hematopoietic stem cells results from disruption of core ribosome components (Xiong *et al.* 2014) and that MDM4 cooperates with MDM2 to respond to DNA damage in the hematopoietic system in murine models (Pant & Lozano 2014). The RoTS peak relevant to MDM4 spans the entire gene, suggesting it is likely to harbor the causal variant, though there are other nearby peaks not linked to the most significant SNP over MDM4.

The peak over FMN2 is considerably clearer than the peak over MDM4, centered over intron 1 and exon 2 of the gene. FMN2 stabilizes p21 through a direct binding interaction, and FMN2 expression can be induced via a variety of insults including hypoxia and DNA damage (Yamada *et al.* 2013a; Yamada *et al.* 2013b). Upon p14ARF induction, FMN2 is cleaved and the N-terminus is retained in the nucleus, bound to p21 (Yamada *et al.* 2013a). FMN2 truncating mutations have been linked to intellectual disability in humans (Almuqbil *et al.* 2013), and the mouse knockout for FMN2 displays a conditioned fear learning response defect (Law *et al.* 2014). FMN2 is involved in cytoskeletal organization, particularly during meiosis, most probably because of the actin-binding formin domain present in the protein (Montaville *et al.* 2014). Loss of FMN2 causes infertility in female mice (Leader *et al.* 2002), though attempts to identify FMN2 defects in infertile humans have not met with success (Ryley *et al.* 2005). FMN2 is expressed strongly in developing and adult nervous tissue, as well as a number of other organs (Leader & Leder 2000). Upon review of the primary data from Hirotoshi's genome-wide screen (discussed in Chapter 2), FMN2 knockdown allowed cells to survive toxic telomerase induction more strongly than LARP7 (final adjusted p-value for FMN2 =  $3.36 \times 10^{-9}$  compared with  $1.24 \times 10^{-5}$  for LARP7),

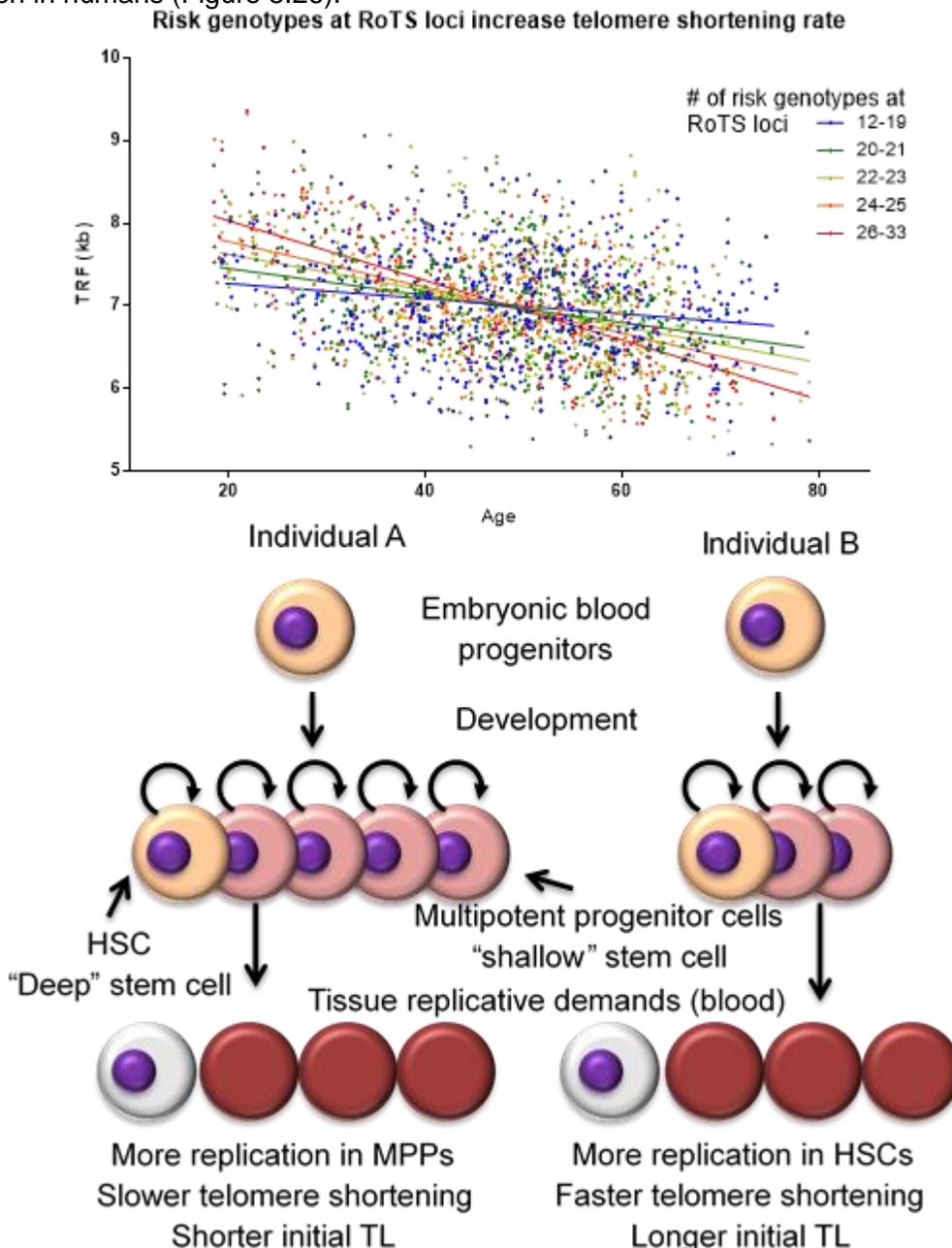
but it was not selected for follow-up, probably because there was no literature linking FMN2 and p21 at the time.

FMN1, a homolog of FMN2 associated with bone development (Hu *et al.* 2014a), is 79 kb from an environmental RoTS peak (Figure 5.25). FLNB, an actin-binding protein that mediates cytoskeletal signaling binds to FMN1 in the second formin homology domain (FHD2), which is conserved between FMN1 and FMN2 (Hu *et al.* 2014a), and a genome-wide significant environmental response RoTS peak also occurs over FLNB as well (Figure 5.25). FLNB promotes differentiation in muscle cell progenitors (Bello *et al.* 2009), and is involved in the molecular response that results in impaired bone marrow reconstitution following exposure to the dioxin TCDD (Casado *et al.* 2011) mediated through the Aryl Hydrocarbon Receptor (AhR). The Aryl Hydrocarbon Receptor Repressor (AhRR), expression of which can rescue hematopoietic stem cells from aromatic hydrocarbon insults (N'Jai A *et al.* 2011), is also under an environmental RoTS peak (Figure 5.25). In short, this analysis may have identified a network of stem cell fate regulating proteins that can influence the intrinsic rate of stem cell turnover as well as the turnover rate in response to environmental insults; as a result of this regulation, they appear to regulate the rate of telomere shortening.



**Figure 5.25: Environmental RoTS loci near FMN1, FLNB and AHRR**  
 Negative log<sub>10</sub>-transformed rateP values for SNPs with diverging concordant twins on the y-axis, while their genomic coordinates are on the x-axis.

The number of genes that seem to regulate stem cell differentiation, survival and growth, coupled with the observation the telomere shortening rate and initial telomere length appear to be intrinsically linked led me to propose a model that may explain this phenomenon as well as generate testable hypotheses about the nature of telomere length regulation in humans (Figure 5.26).



**Figure 5.26: Mechanistic model linking initial length and telomere shortening rate**  
 Top, the relationship between risk genotypes at RoTS loci, telomere shortening rate and initial telomere length. Bottom, differential stem cell dynamics that may explain this relationship.

All the data evaluated in this work is based on measurements of leukocyte telomere length (LTL), which is ultimately derived from the hematopoietic stem cell (HSC) compartment. The rate of cell division in the HSCs is not fixed over a lifetime, with division fastest in early childhood and development during which HSCs divide asymmetrically to give rise to a pool of multipotent progenitor cells (MPPs), which are partially committed progenitor cells with limited abilities to self-renew. These MPPs give rise to the differentiated blood cells (white blood cells, red blood cells, thrombocytes, etc.) (Kondo 2010). When this developmental period ends, HSC division rates slow to roughly one division per year, which corresponds to the slowest measured cross-sectional telomere shortening rate (Catlin *et al.* 2011). Consistent with these stem cell dynamics, telomere shortening in LTL is fastest in early childhood before plateauing at a lower rate in adulthood (Catlin *et al.* 2011).

It is not known how the ratio of MPP cells to HSCs is determined; hypothetically, if two individuals had different MPP:HSC ratios, they would have both different initial telomere length if measured after development as well as different telomere shortening rates. The individual with more expansion of HSCs to MPPs (a higher MPP:HSC ratio, individual A in Figure 5.26) during development would have shorter telomere length when measured in early adulthood (18-20, the youngest individuals in the UK Twins cohort) because the HSC compartment had to engage in more asymmetrical cell division to give rise to the MPPs. However, that individual would also exhibit slower telomere shortening because the HSC compartment would have to divide less often to replace the MPPs, since each cell in the MPP pool has less of a replicative burden. An individual with fewer MPPs relative to HSCs (a lower MPP:HSC ratio, Individual B in Figure 5.26) would have longer LTL in early adulthood because of the smaller amount of cell division in the HSCs, but over time they would exhibit faster telomere shortening because the replicative demands of the blood are

the same for both individuals and this individual's MPPs would have proportionally larger shares of that burden, requiring faster HSC division to compensate for MPP turnover. A testable hypothesis is that the intrinsic RoTS loci may control the MPP:HSC ratio because of the observed linkage between initial telomere length and telomere shortening rate and the identity of the genes most likely to harbor the causal variants. Further the environmental RoTS loci may also point to genes involved in survival and self-renewal of HSCs or MPPs, which would modify telomere shortening rate as a byproduct of modulating stem cell turnover.

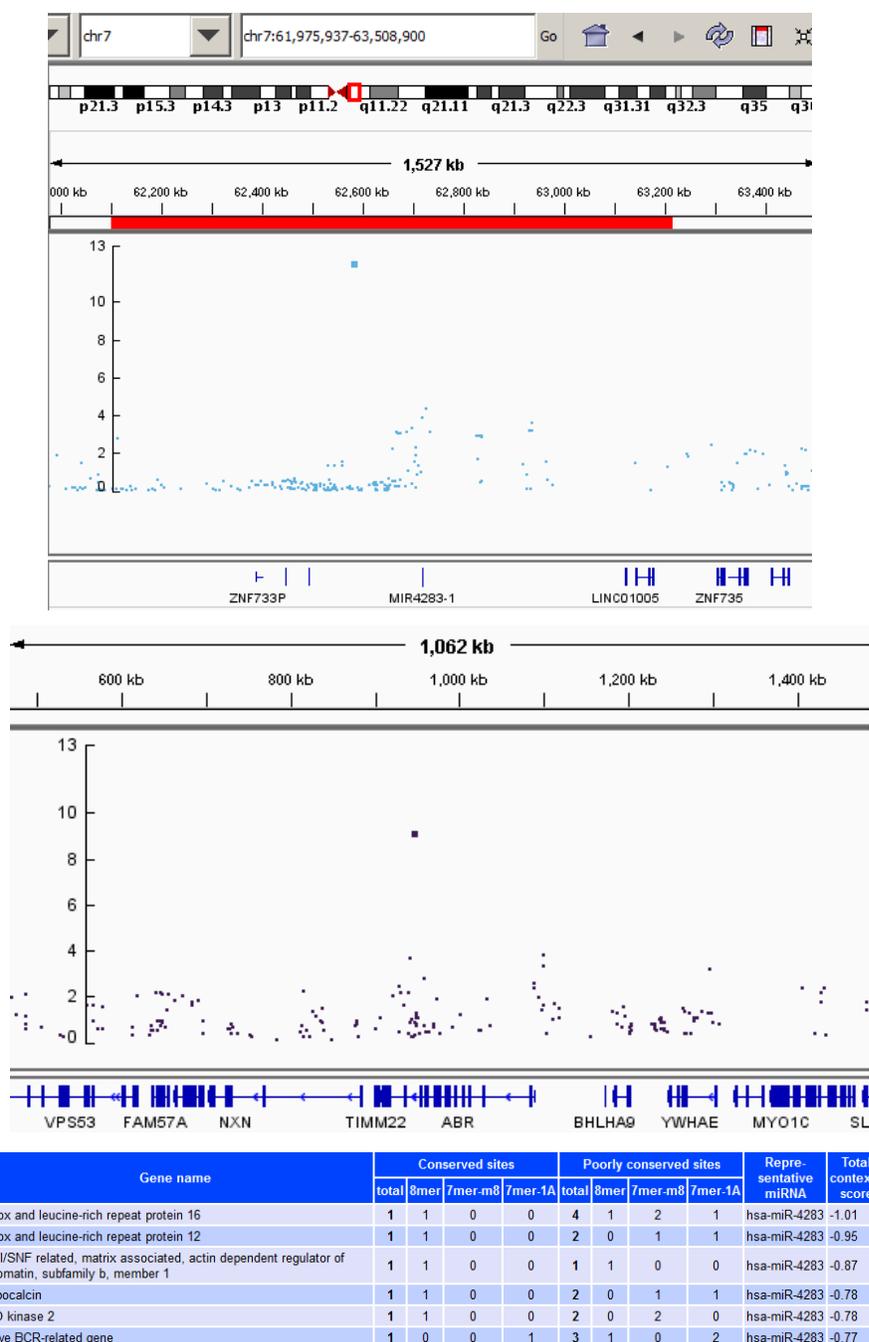
This model generates a number of interesting experiments. First, it is likely that if this is the reason that the RoTS loci appear relevant to telomere shortening rate then they should lose their predictive power or perhaps reverse their effects in a population that includes individuals that are still undergoing development. Most large population studies do not include individuals under the age of 18, but it is possible that large-scale measurement of neonatal telomere length may provide insight, since the predicted differences in initial telomere length would only apply if they occurred prenatally. Second, HSC turnover rate can be quantified, possibly without invasive measurements since several hundred members of the UK Twins cohort are now sequenced; HSC turnover rate and *de novo* mutation rate (which can be inferred from referencing an individual's twin) should correlate with the extent of DNA replication and therefore cell division. Banked blood and tissue samples from the cohort also render it possible to do *in vitro* experiments quantifying the twins' relative sensitivities to insults such as the dioxin TCDD that may be relevant to the RoTS loci, which could inform on the underlying physiological sensitivity. Banked blood also makes a FACS-based measurement of MPP:HSC ratio in blood possible, which may be the simplest test of this hypothesis; long initial length, fast shortening individuals should have a lower MPP:HSC

ratio than short initial length, slow shortening individuals. Finally, it is possible that the MPP:HSC ratio and the information it provides about the relative turnover rate and durability of the stem cell compartment may inform on the type, frequency and aggressiveness of blood-derived tumors in these individuals, though cancer rates may be low enough in the cohort that analysis of these factors by number of RoTS loci is not informative; this would not necessarily have to be performed in the UK Twins cohort, but rather could be performed in a much larger database.

Though the location and recombination of some of the RoTS loci provide strong circumstantial evidence for which genes are likely to harbor a causal variant, other loci are not as clear, for example an intrinsic RoTS locus on chromosome 7 more than a megabase away from the nearest protein-coding gene, itself a zinc-finger containing protein of no known function (ZNF735) (Figure 5.27). The nearest feature of any discernable function, microRNA 4283-1 is predicted to bind to the 3' UTR of ABR, Active BCR related by Targetscan and Targetminer, and ABR is under an environmental RoTS peak (Figure 5.27), however other miRNA binding site prediction programs (miRDB, microRNA.org) did not identify this site. Gene expression data for the UK Twins cohort is available, and in the future I plan to look for changes in gene expression in genes near each RoTS peak in groups composed of individuals with different genotypes at RoTS loci in order to narrow down the list of causal genes.

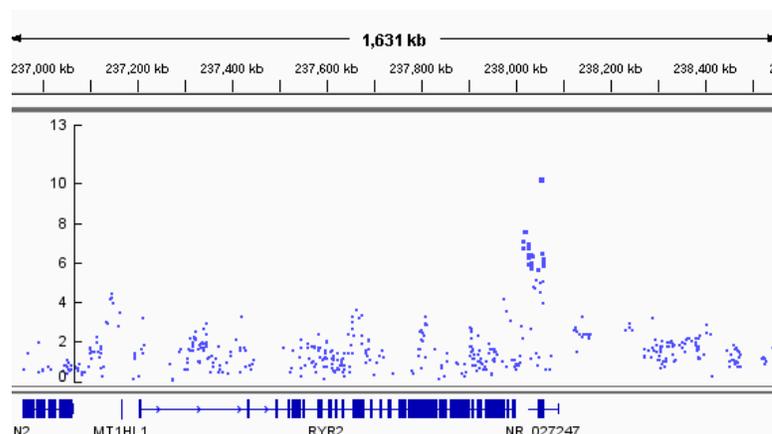
The environmental RoTS locus near LARP7 is particularly interesting because of the abundance of molecular, *in vitro* and *in vivo* data described here detailing LARP7's importance to telomere biology. The SNP near LARP7 is in the 5' untranslated region of the ANK2 gene roughly 200 kilobases from LARP7 (Figure 5.13). Though the *in vitro* case for

LARP7 as the locus likely to harbor the causal variant is strong, there is some evidence that ANK2 could be the source of the signal. ANK2 or Ankyrin B is a membrane-associated adaptor protein, and most Ankyrin proteins link membrane proteins to the spectrin membrane skeleton (Mohler *et al.* 2002). ANK2 knockout mice have defects in cardiac development, and it interacts with the cardiac ryanodine receptor, RYR2 (Mohler *et al.* 2002; Cunha & Mohler 2008). Importantly, RYR2 is very near an intrinsic RoTS peak (Figure 5.28), adding weight to the idea that the genome-wide significant peak near LARP7 may be due to ANK2 rather than LARP7, since an ANK2 interaction partner is also near a peak.



**Figure 5.27: RoTS loci near MIR4283-1 and ABR**

Negative log<sub>10</sub>-transformed rateP values for SNPs with diverging discordant (top) and concordant (middle) twins on the y-axis, while their genomic coordinates are on the x-axis. Targetscan predictions of the binding sites of miR4283-1 are shown on the bottom.

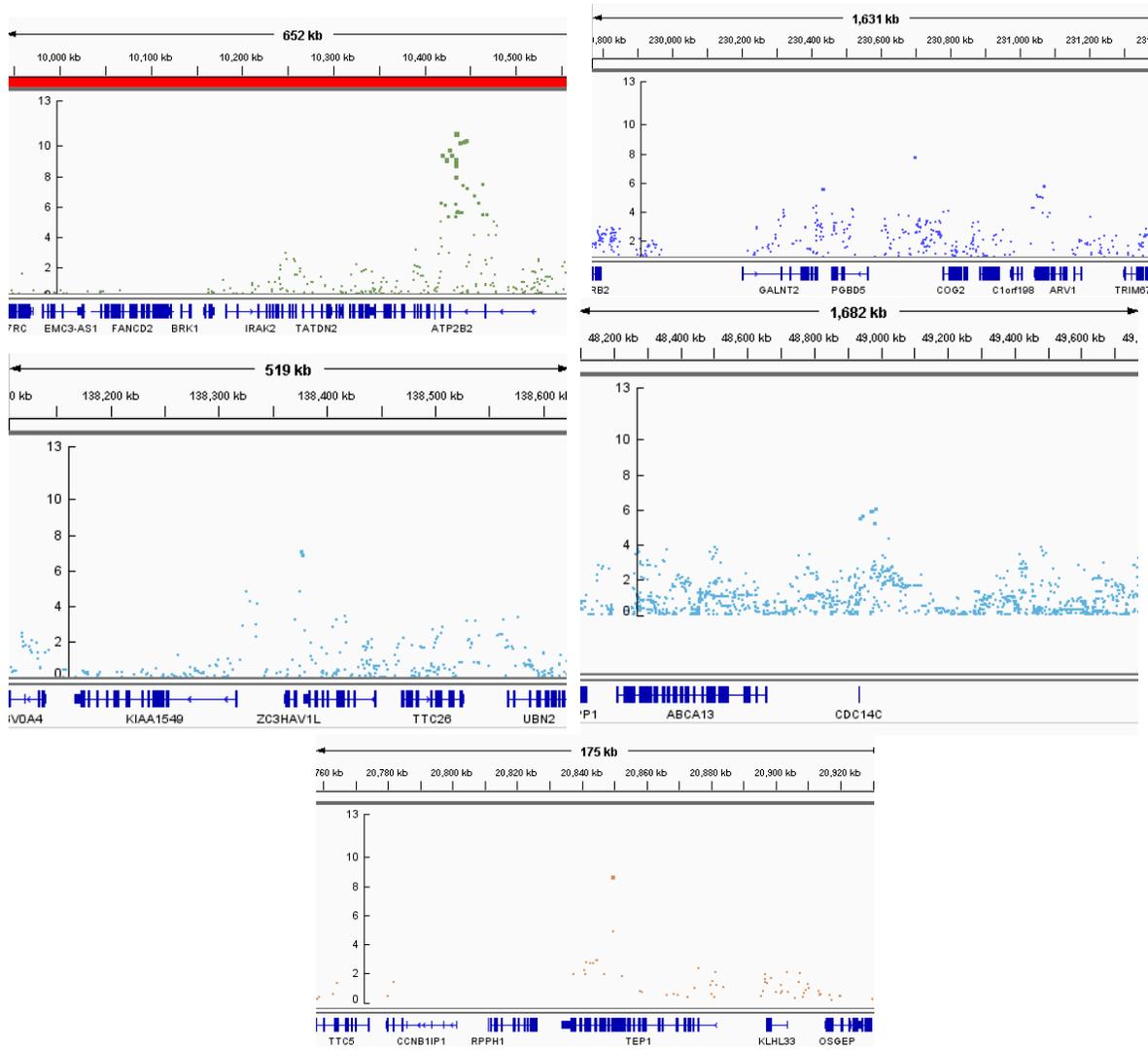


**Figure 5.28: Intrinsic RoTS locus near RYR2**

Negative log<sub>10</sub>-transformed rateP values for SNPs with diverging discordant twins on the y-axis, while their genomic coordinates are on the x-axis.

Given that cardiovascular health and telomere biology are very closely tied and both ANK2 and RYR2 are involved in cardiac development and maintenance, it is possible that LARP7 may not be driving the signal. It is also possible that a variant in that region may impact the expression of both ANK2 and LARP7, though evaluation of the gene expression data may yield more information on the important gene.

A number of other genes that cause telomere shortening upon knockdown *in vitro* are near RoTS peaks, for example YWHAE in Figure 5.27 is near the peak putatively associated with ABR. FANCD2 is within 300 kb of another hit, PGBD5 is within 200 kb, and both CDC14c and ZC3HAV1L are directly underneath peaks that are very near genome-wide significance (Figure 5.29). Further, while the nearest peak to TERT was 800 kilobases away over AhRR (Figure 5.25), there was a telomerase component, TEP1 (Harrington *et al.* 1997) under an environmental response peak (Figure 5.29). While it is possible that the screening methods described in Chapter 2 and the network identified here may overlap, they are addressing very different model systems and any comparison without more information for which genes near RoTS loci are driving the effect will be highly subjective.



**Figure 5.29: A number of RoTS loci near genes identified in Chapter 2**

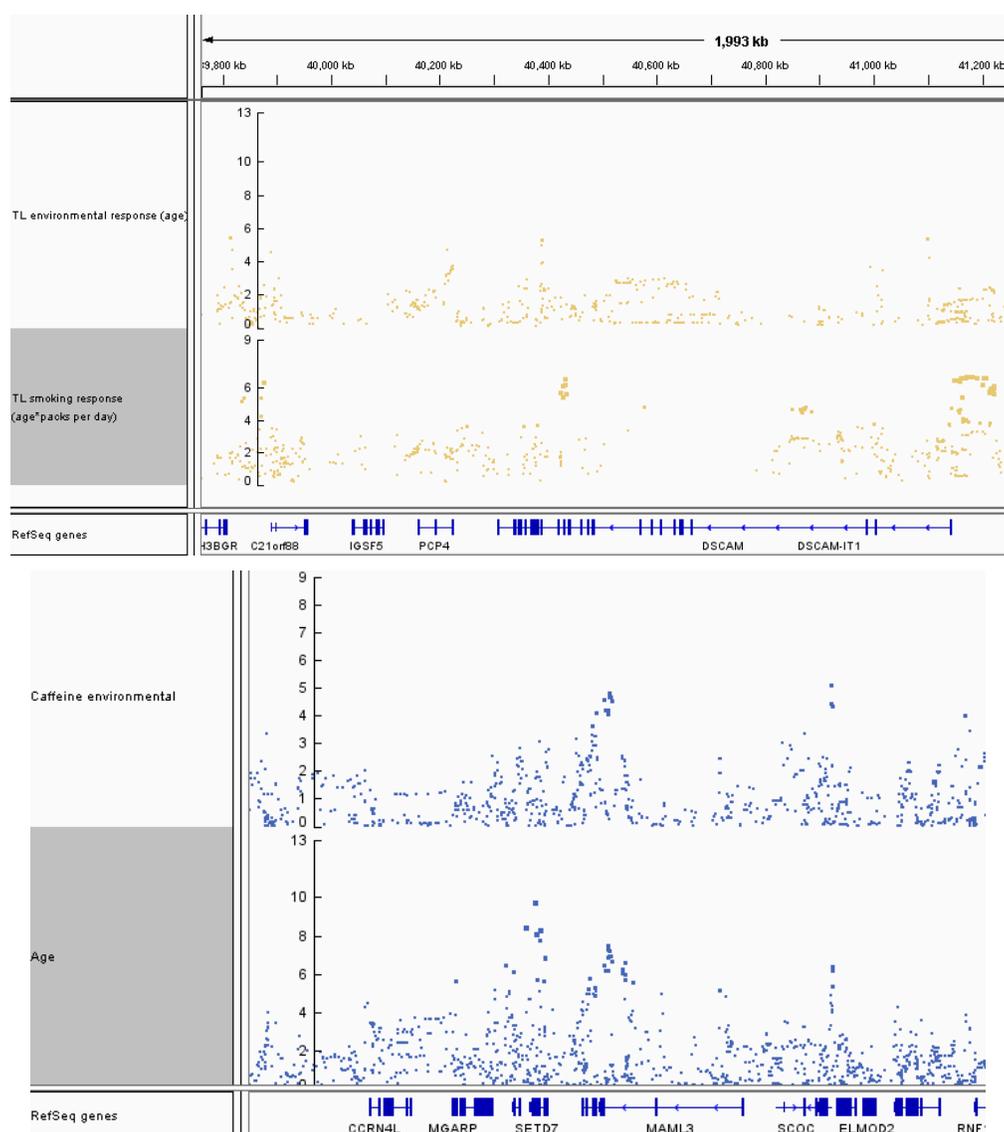
Negative log<sub>10</sub>-transformed rateP values for SNPs on the y-axis, while their genomic coordinates are on the x-axis. The rateP values for the peaks near ZC3HAV1L and CDC14c approach genome-wide significance, but they were not considered for telomere prediction.

The apparent insensitivity to the telomeric effects of smoking observed in the LARP7/ANK2 RoTS peak is interesting, but as already discussed it is possible that it was less significant than it would otherwise be because of the presence of a genotype that appears to render individuals fully insensitive to smoking-induced telomere shortening. Because concordant twin pairs will not diverge as they become increasingly divergent in

smoking behavior, a subset of the population will not obey the models outlined earlier in this chapter, and therefore this locus and others like it may generate false-negative results. Indeed, it is possible to perform this analysis replacing age on the x-axis, which is used as a proxy for all environmental variables, with specific environmental variables of interest. I have performed this analysis with the data available on two environmental factors (smoking and caffeine intake, as well as caffeine intake by soda, tea and coffee separately). These analyses are not yet fully evaluated, however there are some interesting phenomena; for one, though data for the LARP7/ANK2 locus for smoking is convincing in isolation, the rateP value derived from the difference in trend between concordant and discordant twins in smoking is not genome-wide significant (it is below the Bonferonni cutoff). However, other loci in the genome are significant, and of particular interest are those loci that indicate concordant twins diverge over time and also in response to difference in a specific environmental variable (two examples are shown in Figure 5.30).

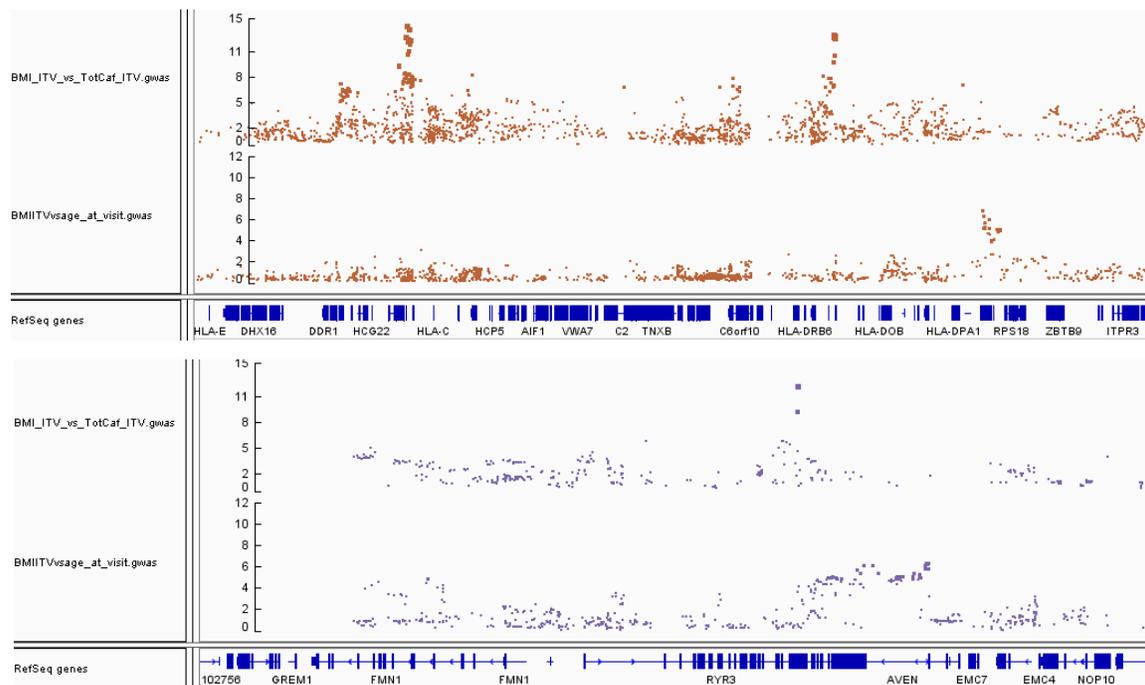
Down-syndrome cell adhesion molecule (DSCAM) has previously been associated with nicotine dependence (Uhl *et al.* 2008; Vink *et al.* 2009) and there was a genome-wide significant association between increasingly divergent telomere length in concordant twins as they became increasingly divergent in smoking behavior, as well as a near-significant association between TRF ITV and age in two separate regions of this large gene (Figure 5.30). There were also a large number of loci that were genome-wide significant in the smoking analysis that were not in the age data, including other smoking-associated genes such as FHIT (Yanagawa *et al.* 2011), HAS2 (Yatagai *et al.* 2014) and PRKCD (Zhou *et al.* 2013b).

Mastermind-like 3, MamL3 is a DNA binding protein not previously linked to the caffeine response, but it functions in the Notch and p53 pathways in a somewhat redundant fashion with MamL1 (Oyama *et al.* 2011). It is interesting that the peaks over MamL3 and ELMOD2 appeared in the caffeine analysis and the age analysis, but the peak near SETD7 does not. Because of their proximity, I used these peaks as a single locus (SETD7, wherein discordant twins diverge) in the telomere length prediction, but this data seems to indicate that they represent at least two distinct signals.



**Figure 5.30: Overlaps between loci identified with age as a proxy for the environment and loci identified with specific environmental information**  
 Negative log<sub>10</sub>-transformed rateP values for SNPs on the y-axis, with genomic coordinates on the x-axis. The top panel compares the association between divergence in smoking and TRFITV (top) and TRFITV and age (bottom). The bottom panel compares the divergence in caffeine exposure and TRFITV (top) compared with the age and TRFITV analysis (bottom).

It is also possible to perform analysis using this method replacing telomere length as the variable of interest. While these results are preliminary as well, I evaluated the effect of different caffeine intake on the difference in BMI, since the genetics of obesity are much better studied and I was attempting to validate the methodology; if classically obesity or caffeine-associated loci appeared in this analysis, I could place greater trust in the results. Again for brevity, two of many results are shown in Figure 5.31.



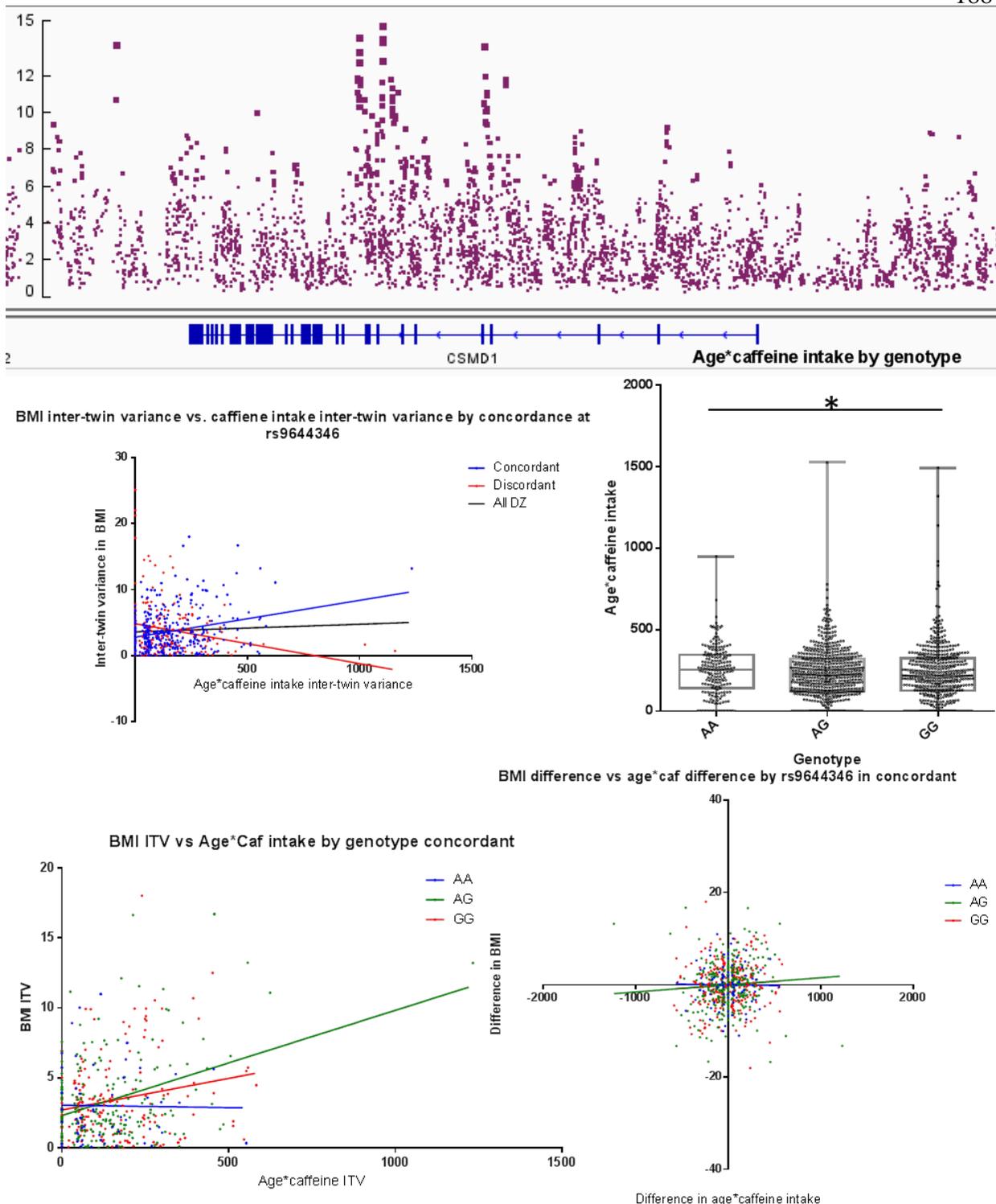
**Figure 5.31: Using the same method to find GxEs for BMI and Caffeine identifies the HLA locus and RYR3, a caffeine receptor**

Negative log<sub>10</sub>-transformed rateP values for SNPs on the y-axis, with genomic coordinates on the x-axis. The both panels compare the association between divergence in total caffeine intake and BMI ITV (top) and BMIITV and age (bottom).

The HLA locus has been associated with obesity since the late 1970's (Savi *et al.* 1977), and the association between difference in caffeine intake and difference in BMI at this locus is the strongest signal so far identified with this method; both HLA clusters in this vicinity were highly significant. The Ryanodine receptors are calcium channels that reside on the sarcoplasmic reticulum and respond to a variety of stimuli, including caffeine (Pessah

*et al.* 1987), though RYR3 is also expressed in non-excitabile cells (Ozawa 2001). It is encouraging that a known receptor for caffeine appears to be relevant to the association between ITV in caffeine intake and ITV in BMI. These preliminary findings provide more evidence that the method may be applicable to other areas of interest, and if it can make usable predictions about the response to specific environmental influences it may have direct applications in personalized medicine.

Twin variance analysis does have a number of shortcomings. One such shortcoming, the propensity to observe inflated significance at loci with very small numbers of discordant twins, has already been addressed through a filtering step. However, there are not many twin databases of this size, and that problem may preclude the use of this method in smaller studies. Further, the method can generate false negative results if one of the responses to an environmental variable at a locus includes a genotype fully insensitive to the variable in question, as observed in the LARP7/ANK2 locus. Additionally, the method cannot identify genes that are important to a process if there are no underlying genetic variants that confer changes in function in that gene, a problem it shares with GWAS studies. Lastly, if the assumptions outlined in Figure 5.2 are not true, the method can generate erroneously inflated or bizarre results. In particular, if a genotype at a locus is associated with different exposure to an environmental effect as well as a different response to that effect, behavior in DZ twins can become very erratic. An example of this is shown from the analysis of ITV in BMI on ITV in caffeine intake, Figure 5.32, the CSMD1 locus.



**Figure 5.32: The CSMD1 locus correlates with both exposure and response to caffeine** -log<sub>10</sub>-transformed rateP values for the CSMD1 locus (top) were inflated by greater exposure in the caffeine-resistant AA genotype. ITV in BMI vs ITV in caffeine exposure (middle left) indicates a strong decreasing association in discordant twins. AA individuals had significantly (p < 0.001) higher caffeine exposure than G allele carriers (middle right). Concordant G allele carriers diverged robustly in response to caffeine (bottom left). G allele carriers but not AA individuals increased BMI with increasing caffeine exposure relative to their twin (bottom right).

The CSMD1 locus initially generated a very strong signal, however the very strongly decreasing relationship between ITV in BMI and ITV in caffeine intake in the discordant twin-pairs had not been observed to that extent before. Closer analysis revealed that though AA genotype individuals were resistant to the BMI-related effects of caffeine, they had considerably higher exposure to caffeine than individuals with other genotypes; thus the discordant twin-pairs that included an AA individual violated the assumption that there would be an even distribution of DZ twins in the four different possible combinations of exposure and sensitivity (many more such pairs had the more resistant twin more exposed to the stimulus), generating an inflated rateP value.

These results are exciting and represent a first foray into an area of telomere biology not heavily evaluated before utilizing new methods. Much remains to be done to validate the findings from this analysis and verify that the results are non-spurious, however the agreement between much of the *in vitro* data presented earlier and the findings in this chapter suggests that at least some of them will be verified. While this method has a number of shortcomings that can lead to both false positive and false negative results the same can be said for any analysis of population-level data and the sources so far identified as the origin of these errors can be controlled. In conclusion, I have developed a novel method to explore gene-environment interactions with regard to telomere biology and possibly other factors, performed proof-of-concept experiments demonstrating that the telomeric response to two variables, caffeine exposure and smoking, depend on gene-environment interactions. I have used this method to identify a network of genomic sites that seem to regulate both the intrinsic rate of telomere shortening and the rate of telomere shortening in response to environmental stimuli, and this network seems most likely accomplish this telomere shortening regulation through hematopoietic stem cell dynamics

and I have also demonstrated that the telomere length predictions made by this method can be replicated in an unrelated population.

## **METHODS**

### **Datasets**

The UK Twins dataset and NHLBI Family Heart Study datasets were used in these analyses, obtained from the groups identified in Chapter 4. Genotype from the UK Twins cohort was obtained in IMPUTE format, whereas the NHLBI FHS used gene dosage format. Genotype formats were reconciled using custom software shown in Appendix B.

### **Statistics**

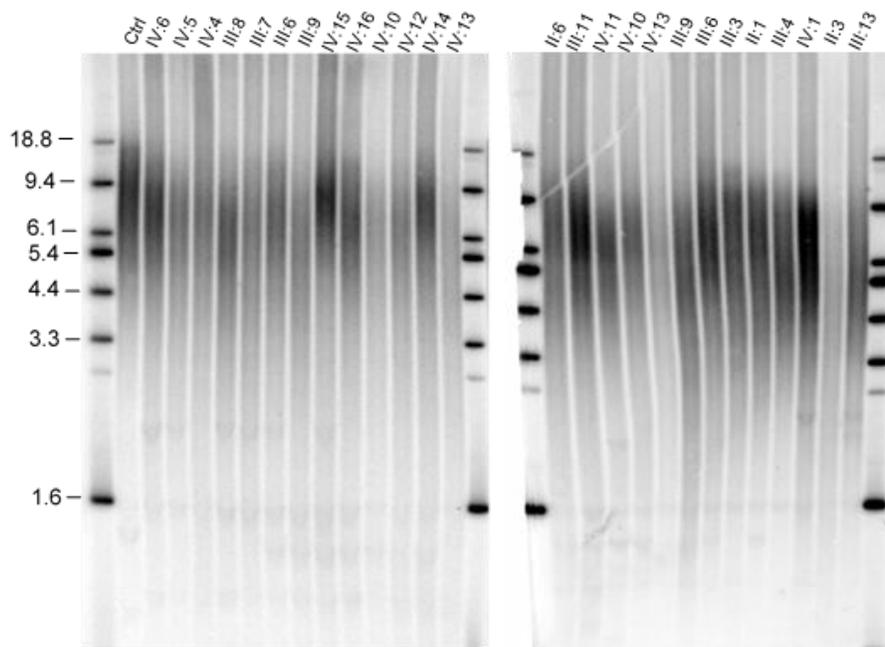
The R statistical software was used for the majority of these analyses using both the Data Table and Hash packages to facilitate handling of very large datasets. Custom scripts used for this analysis are shown in Appendix B. Graphics were made using Graphpad Prism 17, the Broad Institute's Integrative Genomics Viewer (Robinson *et al.* 2011) and Locuszoom (Pruim *et al.* 2010). The procedure used for the twin variance analysis was described in the results section and outlined in Figure 5.8.

## CONCLUSION

*In vitro* work described here demonstrates that the telomerase regulatory network is very large in human cells, and knockdown of many of the proteins involved in DNA repair and replication can lead to telomere shortening over time. The high positive rate indicates that the large number of genes implicated in telomerase regulation in the genome-wide screen may be real positives, suggesting that the telomerase regulatory network is vast and malleable. Telomere length analysis of a human family with a loss-of-function mutation in LARP7 indicates that LARP7 deficiency causes a telomeropathy, consistent with the *in vitro* results. Perifosine, a drug selected as a possible telomerase inhibitor because of its effects on the AKT pathway, which is upstream of a number of the validated candidates causes telomere shortening and a reduction in telomerase activity in a cell line dependent fashion. Perifosine may be useful clinically, as demonstrated by the shortening of the shortest telomers observed in human chronic lymphocytic leukemia tumors treated with Perifosine continuously as part of a phase II clinical trial. Analysis of three large population datasets reveals a time-dependent decrease in initial telomere length that may be due to environmental factors, consistent with the plasticity of the telomerase regulatory network observed *in vitro* and *in vivo*. Lastly, detailed analysis of a large twins database reveals a network of loci that can be used to predict telomere length from age and genotype that replicates in another dataset that implicates hematopoietic stem cell dynamics in regulation of human lymphocyte telomere length.

## APPENDIX A

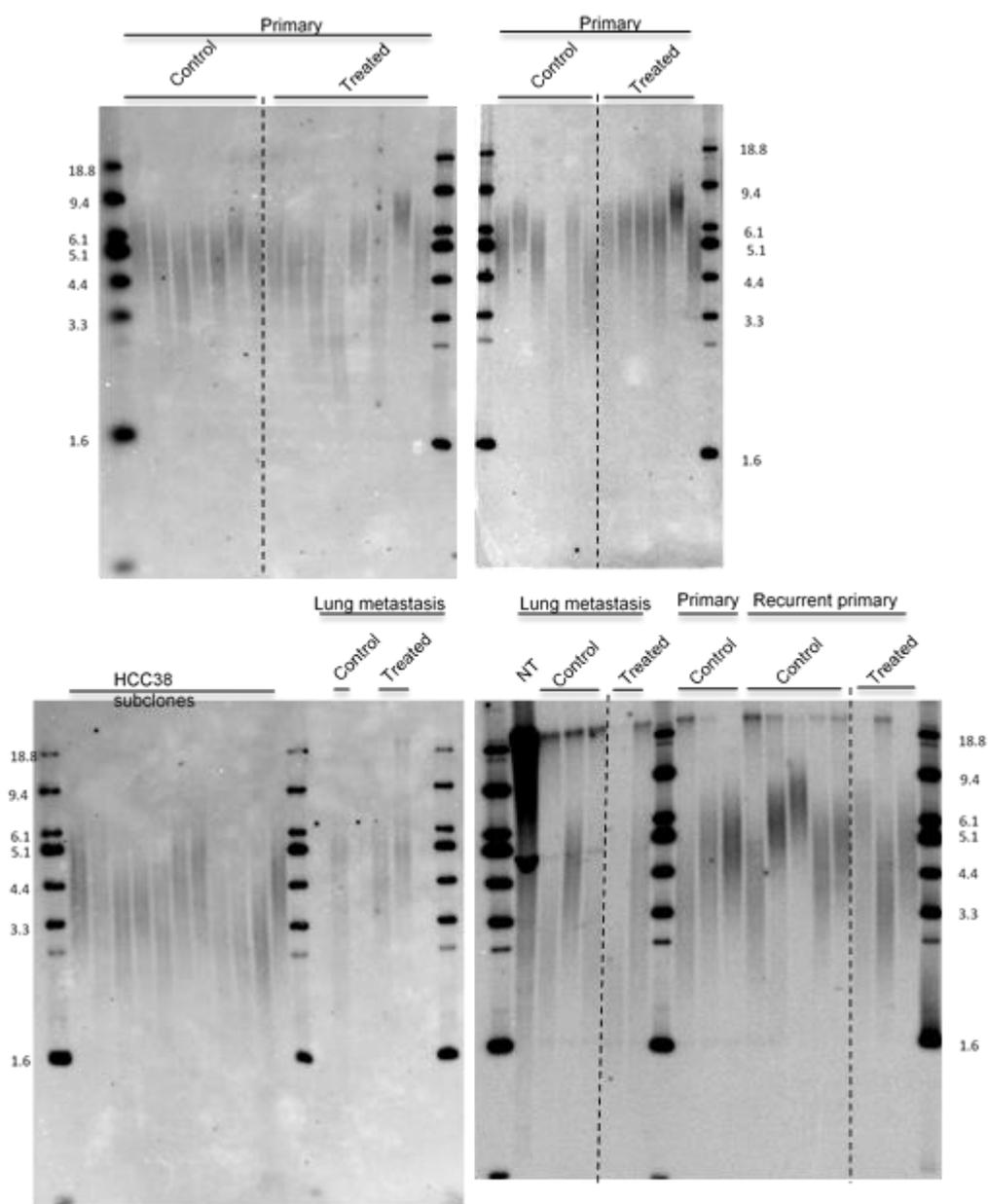
### TRF Data



**Figure A1: Raw TRF Data for the LARP7 Mutant family**

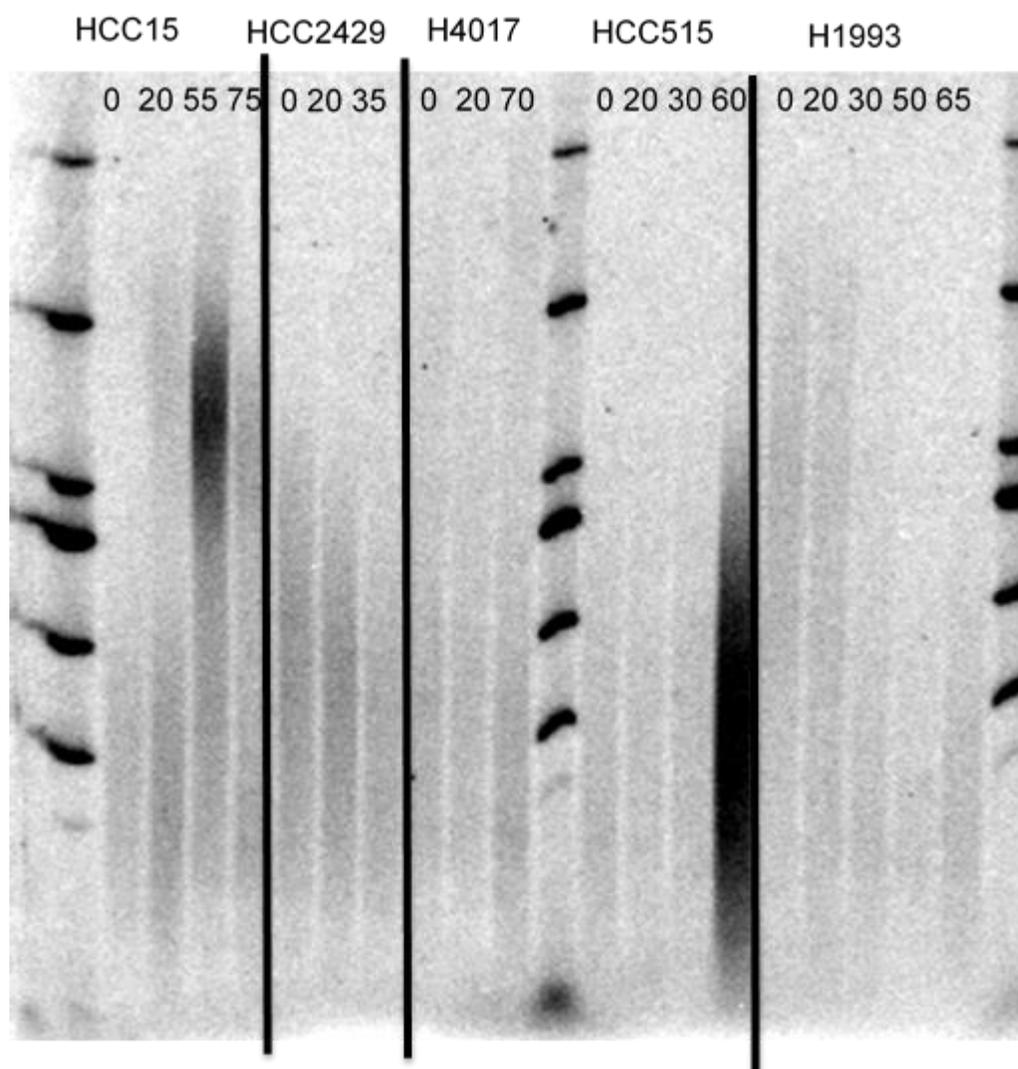
Terminal Restriction Fragment (TRF) gels are shown at the top with ladder bands in kilobases on the left. Individual numbers are in reference to the pedigree shown in Figure 2.7. Quantitation of these lengths are shown on the bottom right, with individual age in years. An unrelated wild-type individual is shown on the far left of the top left gel.

Individual	Age (years)	TL (kb)
IV:1	22.00	4.93
IV:4	4.58	5.44
IV:5	3.00	5.42
IV:6	30.08	5.85
IV:10	8.75	5.07
IV:11	23.00	4.91
IV:12	13.25	5.10
IV:13	10.75	4.73
IV:14	5.08	5.52
IV:15	1.50	5.73
IV:16	8.75	5.45
III:3	30.58	5.00
III:4	40.58	4.91
III:6	32.75	5.17
III:7	26.75	4.82
III:8	29.75	4.96
III:9	34.75	4.46
III:11	44.00	5.19
III:13	38.58	4.48
II:1	53.58	4.99
II:3	51.58	3.79
II:6	73.58	4.94



**Figure A2: TRF data for all xenograft samples analyzed**

Terminal Restriction Fragment (TRF) gels are shown at the top with ladder bands in kilobases on the left. Purified samples from untreated (control) and treated xenograft samples from primary, recurrent primary and metastatic tumors are shown, along with subclones of the HCC38+LUC+GFP parent population.



**Figure A3: TRF data for cell lines treated for extended periods of time with Perifosine**  
Telomere length analysis of cell lines treated for extended periods of time with Perifosine are shown along with the time point in population doublings assayed.

## APPENDIX B Custom Software

The following scripts are compatible with the R statistical programming language, presented here in text format for documentation purposes. They were written for the specific file architecture used in this analysis, and would need adaptation to work in the context of other datasets. Specific characteristics of each data type required occasional changes to the scripts not annotated here (NA handling, entry duplications, blanks) that were specific to the dataset being analyzed. The scripts are presented in the order used (Figure 5.7).

### **FAMINFO**

```
faminfo <- function (data) {
  j <- 1
  famdisc <- NULL
  famdisc <- data.frame()
  tempfamdisc <- data.frame()
  entry <- data.frame()
  for (i in 1:nrow(data)) {
    if (data[i,"twinrow"] != "." & data[i,"TRFITV"] != "." & data[i,"age_at_visit"] != ".")
  ){
      entry[j,1] <- data[i,"fam_ID"]
      entry[j,2] <- data[i,"TRFITV"]
      entry[j,3] <- data[i,"zyg"]
      entry[j,4] <- data[i,"sex_trf"]
      entry[j,5] <- data[i,"age_at_visit"]
      entry[j,6] <- data[i,"TRF"]
      entry[j,7] <- as.numeric(as.character(data[i,"twinrow"]))
      entry[j,8] <- data[i,"row"]
      entry[j,9] <- data[i,"ID_1"]
      entry[j,10] <- data[i,"cigperday"]
      j = j + 1
    }
  }
  names(entry) <-
c("fam_ID","TRFITV","zyg","sex_trf","age_at_visit","TRF","twinrow","row","ID_1","cigperday")
  assign("famdsc",entry,envir=globalenv())
  assign("famprob",entry,envir=globalenv())
}
```

### **DOCHROM**

```
dochrom <- function (rawdata,bsize) {
  #loads the second column of a chromosome file, then determines its length
  #tiles through the chromosome in blocks of length bsize to conserve memory
```

```

#outputs discordance files of length discsize (both calls and probabilities) for later regression
tempdata <- data.frame()
tempdata <- fread(rawdata,select=2,header=FALSE)
chromsize <- nrow(tempdata)
rm(tempdata)
blocknum <- floor(chromsize/bsize)
stragglers <- chromsize %% bsize
faminfo(twinfo)
print(paste(as.character(blocknum),"blocks detected. Calling discordance. "))
flush.console()

#read through each block and report concordance and probability
for (i in 1:blocknum) {
  tempdata <- fread(rawdata,header=FALSE,nrows=bsize,skip=((i-1)*bsize))
  assign("tempdata",tempdata,envir=globalenv())
  hdata <- hash(keys = (1:ncol(tempdata)), values = tempdata)
  assign("hdata",hdata,envir=globalenv())
  calldisc()
  discoutname=as.character((paste(((i-1)*bsize),"disc")),sep="_")
  proboutname=as.character((paste(((i-1)*bsize),"prob")),sep="_")
  save(famdisc,file=discoutname)
  save(famprob,file=proboutname)
  print(paste(as.character(i),"blocks completed"))
  flush.console()
}
#read the stragglers and report conc and probability
tempdata <-
fread(rawdata,header=FALSE,nrows=stragglers,skip=(blocknum*bsize))
assign("tempdata",tempdata,envir=globalenv())
hdata <- hash(keys = (1:ncol(tempdata)), values = tempdata)
assign("hdata",hdata,envir=globalenv())
calldisc()
i=blocknum+1
discoutname=as.character((paste(((i-1)*bsize),"disc")),sep="_")
proboutname=as.character((paste(((i-1)*bsize),"prob")),sep="_")
save(famdisc,file=discoutname)
save(famprob,file=proboutname)

}

```

### **HASHGENO**

```

hashgeno <- function(tempdata) {
  hdata <- hash(keys = (1:ncol(tempdata)), values = tempdata)
  assign("hdata",hdata,envir=globalenv())
} #Sends the tempdata to a hash called hdata

```

### **CALLDISC**

```

calldisc <- function() {

```

```

#this function retrieves genotype information from the hash, finds the highest register
#compares the twins to see if they have the same highest register
#tracks if one or both twins have this SNP successfully measured
#and makes a final discordance call based on relative probabilities
#famdisc must exist, tempdata must have been converted to a hash called hdata already

#pre-allocate space for data
famdisc <- famdisc[,1:9]
famdisc[10:(nrow(tempdata)+9)] <- NA
names(famdisc)[10:ncol(famdisc)] <- paste(values(hdata["2"]))
famprob <- famdisc
hdiscout <- hash(keys=seq(1:nrow(famdisc)), values= "NA")
hprobout <- hash(keys=seq(1:nrow(famdisc)), values= "NA")

for (fam in 1:nrow(famdisc)) {
#find the twin positions

    twin1entries <- (famdisc[fam,"row"]*3):((famdisc[fam,"row"]*3)+2)
    twin2entries <- (famdisc[fam,"twinrow"]*3):((famdisc[fam,"twinrow"]*3)+2)
    allent <- as.character(c(twin1entries,twin2entries)) #a character string
denoting the hash positions of the genotypes for each twin

#extract data for a twin pair from the main hash

    hout <- values(hdata[allent])
    hout <- as.data.frame(hout)

#reorder for discordance check to work
    hout <- hout[c(allent)]

#calculate which register is highest
    hout[7] <- max.col(hout[1:3])
    hout[8] <- max.col(hout[4:6])

#append column 9 and 10 which report the max probability (for later filtering)
    hout[9] <- pmax(hout[1],hout[2],hout[3])
    hout[10] <- pmax(hout[4],hout[5],hout[6])

#append column 11 which denotes if the twins are concordant (0) or discordant (1)
    hout[11] <- ceiling(abs(hout[7]-hout[8])/2)

#outputs data to hashes within the loop (should be faster and use less memory than
writing to the data frame each time)
    hdiscout[[as.character(fam)]] <- hout[11]
    hprobout[[as.character(fam)]] <- rowMeans(hout[9:10])
}

#transfer output from the hashes into the data frames
for (i in 1:nrow(famdisc)) {
    famdisc[i,10:(nrow(tempdata)+9)] <- t(hdiscout[[as.character(i)]])
    famprob[i,10:(nrow(tempdata)+9)] <- t(hprobout[[as.character(i)]])
}

```

```

#output as global variables
assign("famdisc",famdisc,envir=globalenv())
assign("famprob",famprob,envir=globalenv())
}

```

### **STOREGENO**

```

storegeno <- function (blocksize) {
  #read each chromosome in sequence and save them in chunks of size bsize as R
  objects for faster retrieval
  for (i in 1:22) {
    load(paste("/home/bholoh/Run 1 output/chr",i,"stats",sep="")) #load the
    master stats file from the TRFITV v age discordance to get SNP number per chrom
    blocknum <- floor(nrow(statlist)/blocksize) #find the number of full-sized
    blocks there are
    strag <- nrow(statlist) - (blocknum*blocksize) #find the number of SNPs in the
    last block
    print(paste(as.character(blocknum)," block(s) detected on chromosome ",i,".
    Storing data.",sep=""))
    flush.console()

    for (k in 1:blocknum) { #for each block
      tempdata <-
      fread(paste("/home/bholoh/discordance/Chr",i,"/",i,".gen",sep=""),header=FALSE,nrows=blocksize,skip=
      p=(blocksize*(k-1))) #read the block into RAM
      outname <- paste("chr",i,"b",k,"raw",sep="")
      save(tempdata,file=as.character(outname)) #save the block as an R
      object

      print(paste(as.character(k),"block(s) completed."))
      flush.console()
    }
    #repeat as above for the final block
    tempdata <-
    fread(paste("/home/bholoh/discordance/Chr",i,"/",i,".gen",sep=""),header=FALSE,nrows=strag,skip=(blocksize*
    blocknum))
    outname <- paste("chr",i,"b",(blocknum+1),"raw",sep="")
    save(tempdata,file=as.character(outname))
    print(paste(as.character(blocknum+1),"block(s) completed."))
    flush.console()
  }
}

```

### **REGAOV**

```

regaov <- function (depvar, indvar, cutoff, cutzero, chrom) {

  load(as.character(paste("/home/bholoh/discordance/Chr",chrom,"/chr",chrom,"stats",
  sep=""))) #load the TRF and age regression
  statlist <- as.data.frame(matrix(0,nrow=nrow(statlist),ncol=13)) #clear it (but leave it
  the same size)
  names(statlist) <-
  c("rsID","p0","p1",paste(as.character(depvar),"vs",as.character(indvar),"B0"),paste(as.character(depvar)

```

```

ar), "vs", as.character(indvar), "B1"), "mean0", "mean1", "mdif01", "n0", "n1", paste(as.character(indvar), "yint0"), paste(as.character(indvar), "yint1"), "ratep")
  blocknum = floor(nrow(statlist)/10000)
  print(paste(as.character(blocknum), "blocks detected. Regressing."))
  flush.console()

  for (w in 1:(blocknum+1)) {
    z = (w-1)*10000
    load(as.character(paste("/home/bholoh/discordance/Chr", chrom, "/", z, "disc", sep="")))
    load(as.character(paste("/home/bholoh/discordance/Chr", chrom, "/", z, "prob", sep="")))

    famdisc$zyg <- as.character(famdisc$zyg)
    famdisc$TRFITV <- as.numeric(as.character(famdisc$TRFITV))
    famdisc$age_at_visit <- as.numeric(as.character(famdisc$age_at_visit))
    famdisc <- famdisc[order(famdisc$ID_1),]
    famprob <- famprob[order(famprob$ID_1),]
    infodisc <- infodisc[order(infodisc$ID_1),]
    famdisc[, (ncol(famdisc)+1)] <- infodisc[, indvar]
    names(famdisc)[ncol(famdisc)] <- indvar
    famdisc[, (ncol(famdisc)+1)] <- infodisc[, depvar]
    names(famdisc)[ncol(famdisc)] <- depvar

    tempregress <- NULL

    for (x in 10:(ncol(famdisc)-2)) {
      i = (x-9)+z
      statlist[i,1] <- names(famdisc)[x]
      #regress the dependent variable on the independent variable for the
subset that are concordant (0) and have a mean probability call >= the cutoff
      if (length(which(famdisc[,x] == 1 & famprob[,x] >= cutoff) > 10)) {
        if (cutzero == 1) {
          sampop = which(famdisc[,x]==0 & famdisc[, "zyg"]=="DZ" &
famprob[,x] >= cutoff & famdisc[, depvar] != "." & famdisc[, indvar] != "." & is.na(famdisc[, indvar]) ==
FALSE & is.na(famdisc[, depvar]) == FALSE & famdisc[, indvar] != 0)
        } else {
          sampop = which(famdisc[,x]==0 & famdisc[, "zyg"]=="DZ" &
famprob[,x] >= cutoff & famdisc[, depvar] != "." & famdisc[, indvar] != "." & is.na(famdisc[, indvar]) ==
FALSE & is.na(famdisc[, depvar]) == FALSE)
        }

        if (length(sampop) > 10) { #avoids errors if null or very small N or too
many NAs
          tempregress <- lm(formula = eval(as.name(depvar)) ~
eval(as.name(indvar)), data = infodisc, subset = sampop)
          tempsum <- summary(tempregress)$coefficients
          if (ncol(tempsum) == 4 & nrow(tempsum) == 2) {
            statlist[i,2] <- tempsum[2,4] #should be the p-value
            statlist[i,4] <- tempsum[2,1] #should be effect size
            statlist[i,6] <-
mean(as.numeric(as.character(infodisc[sampop, depvar]))) #mean variance
            statlist[i,9] <- length(sampop) #sample size
            statlist[i,11] <- tempsum[1,1] #y-intercept

```

```

    } else {
      statist[i,2] <- 1
      statist[i,4] <- 0
      statist[i,6] <- 0
      statist[i,9] <- 0
      statist[i,11] <- 0
    }
  } else {
    statist[i,2] <- 1
    statist[i,4] <- 0
    statist[i,6] <- 0
    statist[i,9] <- 0
    statist[i,11] <- 0
  }
  #now regress as above but with discordant subsets (1)
  if (cutzero == 1) {
    sampop = which(famdisc[,x]==1 & famdisc[,"zyg"]=="DZ" &
famprob[,x] >= cutoff & famdisc[,depvar] != "." & famdisc[,indvar] != "." & is.na(famdisc[,indvar]) ==
FALSE & is.na(famdisc[,depvar]) == FALSE & famdisc[,indvar] != 0)
  } else {
    sampop = which(famdisc[,x]==1 & famdisc[,"zyg"]=="DZ" &
famprob[,x] >= cutoff & famdisc[,depvar] != "." & famdisc[,indvar] != "." & is.na(famdisc[,indvar]) ==
FALSE & is.na(famdisc[,depvar]) == FALSE)
  }
  if (length(sampop) > 10) { #avoids errors if null or very small N or too
many NAs
    tempregress <- lm(formula = eval(as.name(depvar)) ~
eval(as.name(indvar)), data = infodisc, subset = sampop)
    tempsum <- summary(tempregress)$coefficients
    if (ncol(tempsum) == 4 & nrow(tempsum) == 2) {
      statist[i,3] <- tempsum[2,4] #should be the p-value
      statist[i,5] <- tempsum[2,1] #should be effect size
      statist[i,7] <-
mean(as.numeric(as.character(infodisc[sampop,depvar]))) #mean variance
      statist[i,10] <- length(sampop) #sample size
      statist[i,12] <- tempsum[1,1] #y-intercept
    } else {
      statist[i,3] <- 1
      statist[i,5] <- 0
      statist[i,7] <- 0
      statist[i,10] <- 0
      statist[i,12] <- 0
    }
  } else {
    statist[i,3] <- 1
    statist[i,5] <- 0
    statist[i,7] <- 0
    statist[i,10] <- 0
    statist[i,12] <- 0
  }
  #calculate mean ITV differences
  if (statlist[i,6] * statist[i,7] != 0) {
    statist[i,8] = abs(statlist[i,6]-statlist[i,7])
  }

```

```

    } else {
      statist[i,8] = 0
    }

    #now an ANCOVA to get an indvar:genotype interaction p-value
    if (cutzero == 1) {
      sampop = which(famdisc[,"zyg"]=="DZ" & famprob[,x] >= cutoff &
famdisc[,depvar] != "." & famdisc[,indvar] != "." & is.na(famdisc[,indvar]) == FALSE &
is.na(famdisc[,depvar]) == FALSE & famdisc[,indvar] != 0)
    } else {
      sampop = which(famdisc[,"zyg"]=="DZ" & famprob[,x] >= cutoff &
famdisc[,depvar] != "." & famdisc[,indvar] != "." & is.na(famdisc[,indvar]) == FALSE &
is.na(famdisc[,depvar]) == FALSE)
    }
    if (length(sampop) > 10) { #avoids errors if null or very small N or too
many NAs
      snpname <- names(famdisc)[x]
      tempregress <- aov(formula = eval(as.name(depvar)) ~
eval(as.name(indvar))*eval(as.name(snpname)), data = famdisc, subset = sampop)
      if (length(summary(tempregress)[[1]][["Pr(>F)"]][3]) > 0) {
        statist[i,13] <-
summary(tempregress)[[1]][["Pr(>F)"]][3]
      } else {
        statist[i,13] <- 1
      }
    }
  }
}
print(paste(as.character(w),"block(s) completed."))
flush.console()
}
assign("statlist",statlist,envir=globalenv())
save(statlist,file=as.character(paste("/home/bholoh/discordance/Chr",chrom,"",as.ch
aracter(depvar),"on",as.character(indvar),"regressAOV",sep=""))) #save the whole chromosome file
when done
}

```

### **INFOSHUFFLE**

```

infoshuffle <- function(depvar,indvar,randseed,paired) {
  #takes indvar and depvar, both of which are twin variance values, and shuffles them to other
twin pairs that also had values for those variables
  #is then to be used in regaov for a multiple testing and stratification control
  #paired dictates if the twins get the indvar and depvar from the same (1) other pair or from
distinct pairs (2), or just the indvar (3) or depvar (4)

  if (paired == 1 | paired == 2 | paired == 3) {
    infodisc[,ncol(infodisc)+1] <- NA
    names(infodisc)[ncol(infodisc)] <- paste(indvar,"shuffle",sep="")
  }
  if (paired == 1 | paired == 2 | paired == 4) {
    infodisc[,ncol(infodisc)+1] <- NA
    names(infodisc)[ncol(infodisc)] <- paste(depvar,"shuffle",sep="")
  }
}

```

```

}

set.seed(randseed)
onetwin <- infodisc[which(infodisc[, "ID_1"] %% 2 == 1 & infodisc[, "zyg"] == "DZ"),]
#NOTE infoshuffle now excludes MZ twins!
hasdep <- which(is.na(onetwin[, depvar]) == FALSE & is.na(onetwin[, indvar]) == FALSE)
#find which twin pairs have the depvar
hasind <- which(is.na(onetwin[, indvar]) == FALSE & is.na(onetwin[, depvar]) == FALSE)
#find which twin pairs have the indvar

#go into infodisc and add values from depvar and indvar into dep/indvar shuffle
without replacement from a list of actual values
#currently set up to use indvar and depvar from the same individuals
for (i in 1:nrow(onetwin)) {
  if (is.na(onetwin[i, indvar]) == FALSE & is.na(onetwin[i, depvar]) == FALSE) {
#assigns a twin pair with both indvar and depvar a random indvar and depvar value from another pair
that had both values
    if (paired == 1) { #paired dictates if the twins get the indvar and depvar from
the same (1) other pair or from distinct pairs (2), or just the indvar (3) or depvar (4)
      #get the indvar and depvar from the same other pair
      indval <- sample(1:length(hasind), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(indvar, "shuffle", sep="")] <- onetwin[hasind[indval], indvar]
      depval <- sample(1:length(hasdep), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(depvar, "shuffle", sep="")] <- onetwin[hasind[indval], depvar]
      hasdep <- hasdep[-depval]
      hasind <- hasind[-indval]
    } else if (paired == 2) {
#get the indvar and depvar from other pairs
      indval <- sample(1:length(hasind), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(indvar, "shuffle", sep="")] <- onetwin[hasind[indval], indvar]
      depval <- sample(1:length(hasdep), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(depvar, "shuffle", sep="")] <- onetwin[hasdep[depval], depvar]
      hasdep <- hasdep[-depval]
      hasind <- hasind[-indval]
    } else if (paired == 3) {
#shuffles indvars and leaves depvars alone
      indval <- sample(1:length(hasind), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(indvar, "shuffle", sep="")] <- onetwin[hasind[indval], indvar]
      hasind <- hasind[-indval]
    } else if (paired == 4) {
#shuffles depvars and leaves indvars alone
      depval <- sample(1:length(hasdep), 1)
      infodisc[which(infodisc[, "fam_ID"] ==
onetwin[i, "fam_ID"]), paste(depvar, "shuffle", sep="")] <- onetwin[hasdep[depval], depvar]
      hasdep <- hasdep[-depval]
    }
  }
}
}

```

```

    assign("infodisc",infodisc,envir=globalenv())
}

```

### **GENOEXTRACT**

```

genoextract <- function (targets) {
  #genoextract will go into the genotype data and call genotypes much like calldisc,
  #except it will report the allele calls and output a seperate allele file for targets in question
  #it takes the output of varregress as input, though an index value must be added prior to this
  in column 11. The index is just the number the SNP appears in in the alldata file (the output of the
  aggregate() function).
  #The index can be added to the alldata data frame as: alldata[11] <- seq(1:nrow(alldata)) DO
  THIS BEFORE FILTERING OR YOU WILL GET DATA ON THE WRONG SNPS
  #it will load the incremental R objects instead of reading from the genotype data master files
  for execution time purposes
  #it will make another file that includes the number of minor alleles (alleldata) as well as SNP
  metadata called locidata

```

```

    load("/home/bholoh/discordance/Chr1/0 disc")
    alleldata <- famdisc[,1:9]
    alleldata[,10:(nrow(targets)+9)] <- 3
    alleleprob <- alleldata
    chrend <- 0
    numdone <- 0
    locidata <- data.frame(1,2,3,4,5)

    for (i in 1:22) {

      load(paste("/home/bholoh/Run 1 output/chr",i,"stats",sep=""))
      onchr <- targets[which(targets["index"] <= (nrow(statlist) + chrend) &
targets[,"index"] > chrend ),] #identifies the SNPs on the chromosome
      blocknum <- ceiling(nrow(statlist) / 50000) #the number of blocks in the
      chromosome
      print(paste(as.character(nrow(onchr))," target(s) on chromosome ",i,".
Retrieving genotypes.",sep=""))
      flush.console()

      if (nrow(onchr) > 0) {
        onchr[, (ncol(onchr)+1)] <- ceiling((onchr[,"index"] - chrend)/50000)
#which block the SNP is on

        for (j in 1:blocknum) {
          onblock <- onchr[which(onchr[,ncol(onchr)] == j),]
          if (nrow(onblock) > 0) { #if there are any SNPs in the block
            load(paste("/home/bholoh/genotype
data/chr",i,"b",j,"raw",sep=""))
            genoblock <- tempdata[which(tempdata[,V2] %in%
onblock[, "rsID"],)] #only pull out the SNPs of interest !!!NOTE THAT THE DATA TABLE PACKAGE
USES ALTERNATE COLUMN NOTATION

            names(alleldata)[(numdone+10):(numdone+nrow(genoblock)+9)] <- genoblock[[2]]

```

```

names(alleleprob)[(numdone+10):(numdone+nrow(genoblock)+9)] <- genoblock[[2]]
locidata <-
rbind(locidata,genoblock[,1:5,with=FALSE],use.names=FALSE)

for (fam in 1:nrow(alleledata)) {
  entries <-
(alleledata[fam,"row"]*3):((alleledata[fam,"row"]*3)+2)
  famentries <-
as.data.frame(genoblock[,entries,with=FALSE])
  famgeno <- (max.col(famentries[1:3])-1)
  fammaxprob <-
pmax(famentries[1],famentries[2],famentries[3])

  alleledata[fam,(numdone+10):(numdone+nrow(genoblock)+9)] <- t(famgeno)

  alleleprob[fam,(numdone+10):(numdone+nrow(genoblock)+9)] <- t(fammaxprob)
}
numdone <- numdone + nrow(genoblock)
}
}
}
chrend <- chrend + nrow(statlist) #stores the index of the last SNP on the
chromosome for later access
}
alleledata <- alleledata[order(alleledata$ID_1),]
assign("alleledata",alleledata,envir=globalenv())
assign("alleleprob",alleleprob,envir=globalenv())
assign("locidata",locidata,envir=globalenv())
}

```

### **ALLELEREGRESS**

```

alleler regress <- function (indvar, depvar, cutoff) {
  #this function will take an alleledata file (the family description data file plus the number of
  minor alleles for each individual
  #and regress the dependent variable on the independent variable for all individuals of each
  genotype above the quality cutoff argument (usually 0.9)
  #it obtains the information on the dependent and independent variables from a separate data
  frame called infodisc so it can also be used for non-telomere analyses
  #it will output this information into a data frame called allelestat

  allelestat <- as.data.frame(matrix(0,nrow=(ncol(alleledata)-9),ncol=24))
  alleledata <- alleledata[order(alleledata$ID_1),]
  infodisc <- infodisc[order(infodisc$ID_1),]
  names(allelestat) <-
c("rsID","p0","p1","p2",paste(as.character(depvar),"vs",as.character(indvar),"B0"),paste(as.character(
depvar),"vs",as.character(indvar),"B1"),paste(as.character(depvar),"vs",as.character(indvar),"B2"),"m
ean0","mean1","mean2","n0","n1","n2","yint0","yint1","yint2","ratep","yintp","intercept","ageB","aov1ge
noB","aov2genoB","aovB1","aovB2")
  alleledata[, (ncol(alleledata)+1)] <- infodisc[,indvar]
  names(alleledata)[ncol(alleledata)] <- indvar
  alleledata[, (ncol(alleledata)+1)] <- infodisc[,depvar]

```

```

names(alleledata)[ncol(alleledata)] <- depvar
alleleprob[, (ncol(alleleprob)+1)] <- infodisc[, indvar]
names(alleleprob)[ncol(alleleprob)] <- indvar
alleleprob[, (ncol(alleleprob)+1)] <- infodisc[, depvar]
names(alleleprob)[ncol(alleleprob)] <- depvar

assign("alleledata", alleledata, envir=globalenv())
assign("alleleprob", alleleprob, envir=globalenv())

for (x in 10:(ncol(alleledata)-3)) {
  i = x - 9
  allelestat[i, 1] <- names(alleledata[x])
  #subset out the homozygous major allele individuals
  sampop <- which(alleledata[, x] == 0 & alleledata[, "zyg"] == "DZ" &
alleleprob[, x] >= cutoff & infodisc[, depvar] != "." & infodisc[, indvar] != "." & infodisc[, indvar] != 0 &
infodisc[, depvar] != 0)
  if (length(sampop) > 10) { #avoids errors if null or very small N
    tempregress <- lm(formula = eval(as.name(depvar)) ~
eval(as.name(indvar)), data = infodisc, subset = sampop)
    tempsum <- summary(tempregress)$coefficients
    if (ncol(tempsum) == 4 & nrow(tempsum) == 2) {
      allelestat[i, 2] <- summary(tempregress)$coefficients[2, 4]
#should be the p-value
      allelestat[i, 5] <- summary(tempregress)$coefficients[2, 1]
#should be effect size
      allelestat[i, 8] <-
mean(as.numeric(as.character(infodisc[sampop, depvar]))) #mean of depvar
      allelestat[i, 11] <- length(sampop) #sample size
      allelestat[i, 14] <- summary(tempregress)$coefficients[1, 1]
#yint
    } else {
      allelestat[i, 2] <- 1
      allelestat[i, 5] <- 0
      allelestat[i, 8] <- 0
      allelestat[i, 11] <- 0
    }
  } else {
    allelestat[i, 2] <- 1
    allelestat[i, 5] <- 0
    allelestat[i, 8] <- 0
    allelestat[i, 11] <- 0
  }
  sampop <- which(alleledata[, x] == 1 & alleledata[, "zyg"] == "DZ" &
alleleprob[, x] >= cutoff & infodisc[, depvar] != "." & infodisc[, indvar] != "." & infodisc[, indvar] != 0 &
infodisc[, depvar] != 0)

  if (length(sampop) > 10) { #avoids errors if null or very small N
    tempregress <- lm(formula = eval(as.name(depvar)) ~
eval(as.name(indvar)), data = infodisc, subset = sampop)
    tempsum <- summary(tempregress)$coefficients
    if (ncol(tempsum) == 4 & nrow(tempsum) == 2) {
      allelestat[i, 3] <- summary(tempregress)$coefficients[2, 4]
#should be the p-value

```

```

allelestat[i,6] <- summary(tempregress)$coefficients[2,1]
#should be effect size
allelestat[i,9] <-
mean(as.numeric(as.character(Infodisc[sampop,deivar]))) #mean of deivar
allelestat[i,12] <- length(sampop) #sample size
allelestat[i,15] <- summary(tempregress)$coefficients[1,1]
#yint
    } else {
        allelestat[i,3] <- 1
        allelestat[i,6] <- 0
        allelestat[i,9] <- 0
        allelestat[i,12] <- 0
    }
} else {
    allelestat[i,3] <- 1
    allelestat[i,6] <- 0
    allelestat[i,9] <- 0
    allelestat[i,12] <- 0
}
sampop <- which(alleledata[,x] == 2 & alleledata[, "zyg"] == "DZ" &
alleleprob[,x] >= cutoff & Infodisc[,deivar] != "." & Infodisc[,indvar] != "." & Infodisc[,indvar] != 0 &
Infodisc[,deivar] != 0)

if (length(sampop) > 10) { #avoids errors if null or very small N
tempregress <- lm(formula = eval(as.name(deivar)) ~
eval(as.name(indvar)), data = Infodisc, subset = sampop)
tempsum <- summary(tempregress)$coefficients
if (ncol(tempsum) == 4 & nrow(tempsum) == 2) {
    allelestat[i,4] <- summary(tempregress)$coefficients[2,4]
#should be the p-value
    allelestat[i,7] <- summary(tempregress)$coefficients[2,1]
#should be effect size
    allelestat[i,10] <-
mean(as.numeric(as.character(Infodisc[sampop,deivar]))) #mean of deivar
allelestat[i,13] <- length(sampop) #sample size
allelestat[i,16] <- summary(tempregress)$coefficients[1,1]
#yint
    } else {
        allelestat[i,4] <- 1
        allelestat[i,7] <- 0
        allelestat[i,10] <- 0
        allelestat[i,13] <- 0
    }
} else {
    allelestat[i,4] <- 1
    allelestat[i,7] <- 0
    allelestat[i,10] <- 0
    allelestat[i,13] <- 0
}

#now an ANCOVA to get an indvar:genotype interaction p-value
#This section is commented out because it is computationally intensive and not
relevant to most analysis, though interesting. Remove the comment markers to include it

```

```

#               sampop = which(alleledata["zyg"]=="DZ" & alleleprob[,x] >= cutoff &
infodisc[,depvar] != "." & infodisc[,indvar] != "." & infodisc[,indvar] != 0 & infodisc[,depvar] != 0)
#               alleledata[,x] <- factor(alleledata[,x])
#               if (length(levels(factor(alleledata[,x])) ) > 1) {
#               tempregress <- aov(formula = eval(as.name(depvar)) ~
eval(as.name(indvar))*eval(as.name(names(alleledata)[x])), data = alleledata, subset = sampop)
#               allelestat[i,17] <- summary(tempregress)[[1]][["Pr(>F)"]][3]
#interaction p-value
#               allelestat[i,18] <- summary(tempregress)[[1]][["Pr(>F)"]][2]
#changed intercept p-value
#               if (length(tempregress$coefficients) == 6) {
#               allelestat[i,19] <- tempregress$coefficients[[1]]
#intercept
#               allelestat[i,20] <- tempregress$coefficients[[2]] #age
coef
#               allelestat[i,21] <- tempregress$coefficients[[3]] #het
intercept
#               allelestat[i,22] <- tempregress$coefficients[[4]] #min
allele intercept
#               allelestat[i,23] <- tempregress$coefficients[[5]] #het
rate coef
#               allelestat[i,24] <- tempregress$coefficients[[6]] #min
allele rate coef
#               }
#               }
#               }
#               }
assign("allelestat",allelestat,envir=globalenv())
}

```

### **PREDLENGTH**

```

predlength <- function (famdata, loci, alleledata, alleleprob, cutoff, outstring) {
#predlength will take a list of target SNPs (loci), genotypes at those SNPs (alleledata) and
imputation probability (alleleprob)
#and compute a predicted telomere length for each person in famdata by using the effects in
the loci file with respect to a regression of TRF on age alone
#It will exclude risk/protective allele calls for those sites with imputation probability < cutoff
#Then it will fit a predicted telomere length : actual telomere length regression and report the
p-value

```

```

#calculate the age-associated shortening rate and initial length cross sectionally
lengthreg <- lm(formula = TRF ~ age_at_visit, data = famdata)
shortrate <- summary(lengthreg)$coefficients[2,1] #cross-sectional shortening rate
initlength <- summary(lengthreg)$coefficients[1,1] #computed initial telomere length for the
whole population

```

```

#calculate how each genotype at each SNP in loci affects the shortening rate by itself
loci[, (ncol(loci)+1)] <- 0
names(loci)[ncol(loci)] <- "B0"
loci[, (ncol(loci)+1)] <- 0
names(loci)[ncol(loci)] <- "B1"

```

```

loci[,ncol(loci)+1] <- 0
names(loci)[ncol(loci)] <- "B2"

loci["B0"] <- loci[5] - shortrate
loci["B1"] <- loci[6] - shortrate
loci["B2"] <- loci[7] - shortrate

#set up a column for predicted telomere length, predicted shortening rate and predicted initial
length
famdata[,ncol(famdata)+1] <- 0
names(famdata)[ncol(famdata)] <- "pTRF"
famdata[,ncol(famdata)+1] <- 0
names(famdata)[ncol(famdata)] <- "pSR"
famdata[,ncol(famdata)+1] <- 0
names(famdata)[ncol(famdata)] <- "pIL"

#set up a column that is the cross sectionally predicted length
famdata[,ncol(famdata)+1] <- (famdata["age_at_visit"] * shortrate) + initlength
names(famdata)[ncol(famdata)] <- "cspTRF"

#calculate a predicted age:TRF relationship for each individual based on their genotype
assuming all genotypes are additive
for (i in 1:nrow(famdata)) {
  #now add each effect in the loci file
  for (x in 1:nrow(loci)) {
    if (alleleprob[i,which(names(alleleprob) == loci[x,"rsID"])] >= cutoff) {
      if (loci[x,paste("n",alleldata[i,which(names(alleldata) ==
loci[x,"rsID"])],sep="")] > 100) {
        famdata[i,"pSR"] <- famdata[i,"pSR"] + loci[x,(ncol(loci) +
alleldata[i,which(names(alleldata) == loci[x,"rsID"])] - 2) #adds the effect of the allele for that site
        famdata[i,"pIL"] <- famdata[i,"pIL"] +
(loci[x,(paste("yint",alleldata[i,which(names(alleldata) == loci[x,"rsID"])],sep="")]) - initlength) #adds
the difference between the yint and the population yint
      }
    }
  }
  #compute predicted values for that individual
  famdata[i,"pSR"] <- famdata[i,"pSR"] + shortrate
  famdata[i,"pIL"] <- famdata[i,"pIL"] + initlength
  famdata[i,"pTRF"] <- famdata[i,"pIL"] + (famdata[i,"age_at_visit"] * famdata[i,"pSR"])
}

#output the famdata
assign(outstring,famdata,envir=globalenv())

#regress TRF on predicted TRF for the normal fit and the new fit
prediction <- lm(TRF ~ pTRF, data = famdata)
cprediction <- lm(TRF ~ cspTRF, data = famdata)

assign("prediction",prediction,envir=globalenv())
assign("cprediction",cprediction,envir=globalenv())
}

```

**FHSPULLGENO**

```

fhspullgeno <- function(targets,outinfo,outgeno) {
#FHS data is in chunks of varying sizes, split into splits of 10 thousand SNPs each
#the chunk sizes are varying, but splits default to 10k and can be smaller
#this function will go through each split in a chunk until all targets in that chunk have been
found
#it will then return the genotype calls for each target to a file called FHSAIID
#This file will be the same format as the source genotype files, but only include targets

#make a new column for chunk called whatchunk

targets[, (ncol(targets)+1)] <- 0
names(targets)[ncol(targets)] <- "whatchunk"

#make a new column for locating the same snp called inforow

targets[, (ncol(targets)+1)] <- 0
names(targets)[ncol(targets)] <- "inforow"

#make a new column for denoting features called codeallele

targets[, (ncol(targets)+1)] <- 0
names(targets)[ncol(targets)] <- "codeallele"

#make a new column for denoting features called ncallele

targets[, (ncol(targets)+1)] <- 0
names(targets)[ncol(targets)] <- "ncallele"

#make a new column for finding genotype data calls markname

targets[, (ncol(targets)+1)] <- 0
names(targets)[ncol(targets)] <- "markname"

targets[,"SNP"] <- as.character(targets[,"SNP"])

for (i in 1:22) {
#take the target list for each chromosome and find what chunk they're on, what the maj/min
coded alleles are, output this
  if (length(which(targets[,"chrnum"] == i)) > 0) {
    onchrom <- which(targets[,"chrnum"] == i)
    info <-
suppressWarnings(fread(paste("/home/bholoh/c",i,"/hybrid_info",i,".csv",sep="")))
    info <- as.data.frame(info)

    for (x in 1:length(onchrom)) {
      if (length(which(as.character(info[,"snp"]) ==
targets[onchrom[x], "SNP"])) > 0) {

```

```

                                targets[onchrom[x], "inforow"] <-
which(as.character(info[, "snp"]) == targets[onchrom[x], "SNP"]) #figure out which entry matches the
snp
                                targets[onchrom[x], "whatchunk"] <-
info[targets[onchrom[x], "inforow"], "whatchunk"] #pull whatchunk
                                targets[onchrom[x], "codeallele"] <-
as.character(info[targets[onchrom[x], "inforow"], "effect_all"]) #pull which trait is coded as 0
                                targets[onchrom[x], "ncallele"] <-
as.character(info[targets[onchrom[x], "inforow"], "non_effect_all"]) #pull which trait is coded as 2
                                targets[onchrom[x], "markname"] <-
as.character(info[targets[onchrom[x], "inforow"], "markname"]) #record how it is referred to in genotype
data
                                }
                                }
                                }
                                print(paste(as.character(length(onchrom)), " targets located and annotated on chromosome
", i, ".", sep=""))
                                flush.console()
                                rm(info)
                                }

                                assign(outinfo, targets, envir=globalenv())

                                print("Retrieving genotypes.")
                                flush.console()
                                load("subjgeno")

                                for (i in 1:22) { #i will be used to indicate chromosome number
                                #pull out genotypes for the targets
                                #library(data.table) These two lines are commented out because they are only necessary for
manually iterating over each chromosome. This was necessary for the initial look at the FHS data
because for some reason the fread() command is unstable and encounters segfaults when reading
the data stored as excel files. To get around this, I set i = each chromosome separately and ran each
chromosome one at a time in this loop so that only one chromosome was lost if a segfault occurred.
                                #load("subjgeno")
                                numonchrom <- length(which(targets[, "chrnum"] == i))
                                print(paste(numonchrom, " target(s) on chromosome ", i, ", retrieving genotypes.", sep=""))
                                flush.console()

                                if (numonchrom > 0) {
                                onchrom <- which(targets[, "chrnum"] == i)
                                chunkdist <- targets[which(is.na(targets[onchrom, "whatchunk"]) ==
FALSE), "whatchunk"]

                                for (z in (min(chunkdist):max(chunkdist))) { #z will be used to indicate chunk
number

                                if (length(which(targets[onchrom, "whatchunk"] == z)) > 0) { #if there
are any targets on this chunk

                                for (w in 1:20) { #w will be used to indicate split number. It
never goes higher than 20.

```



```

}
if (rots[x,(indgeno+15)] > 100) { #more than 100 individuals
with that genotype
(rots[x,(indgeno+12)] - shortrate)
(rots[x,(indgeno+18)] - intlength)
}
}
}
}
}
}

#recording of data cleanup for pc2
library(data.table)
locidata <- as.data.frame(locidata)
names(allelestat)[5] <- "AAeff"
names(allelestat)[6] <- "ABeff"
names(allelestat)[7] <- "BBeff"
names(allelestat)[11] <- "nAA"
names(allelestat)[12] <- "nAB"
names(allelestat)[13] <- "nBB"
names(allelestat)[14] <- "yintAA"
names(allelestat)[15] <- "yintAB"
names(allelestat)[16] <- "yintBB"
names(allelestat)[1] <- "SNP"

locidata <- locidata[-1,]
names(locidata)[2] <- "SNP"
names(locidata)[4] <- "A"
names(locidata)[5] <- "B"

cutsub <-
allelestat[,c("SNP", "AAeff", "ABeff", "BBeff", "nAA", "nAB", "nBB", "yintAA", "yintAB", "yintBB")]
loccut <- locidata[,c("SNP", "A", "B")]
pc1targ <- merge(loccut, cutsub, by="SNP")
pc1targ <- merge(targets, pc1targ, by="SNP")

####NOTE: PC2 target list got entered twice somehow, the following step should not be
necessary!!!
pc2cut <- pc2targ[seq(1, nrow(pc2targ), by=2),]
pc2targ <- pc2cut
####NOTE: PC2 target list got entered twice somehow, the preceding step should not be
necessary!!!

pc1targ[,21] <- 0
pc1targ[which(pc1targ[, "codeallele"] == pc1targ[, "A"]), 21] <- 1
names(pc1targ)[21] <- "match"

falldat[,572] <- NULL
falldat[,572] <- NULL

```

```
rots <- pc1targ
#Goto the first line in this document, cleanup is complete.
```

### **PREPGWAS**

```
prepGWAS <- function(depvar,indvar,numcut) {
  alldata[,14] <- seq(1:nrow(alldata))
  names(alldata)[14] <- "index"
  names(alldata)[4] <- paste("B0",indvar,sep="")
  names(alldata)[5] <- paste("B1",indvar,sep="")
  names(alldata)[13] <- "P"
  names(alldata)[1] <- "SNP"
  names(alldata)[11] <- "yint0"
  names(alldata)[12] <- "yint1"
  highn <- alldata[which(alldata[, "n1"] >= numcut),]
  concddiv <- highn[which(highn[,paste("B0",indvar,sep="")] > 0 &
highn[,paste("B1",indvar,sep="")] < 0),]
  discdiv <- highn[which(highn[,paste("B1",indvar,sep="")] > 0 &
highn[,paste("B0",indvar,sep="")] < 0),]

  concddiv <- merge(concddiv,snp,by="SNP")
  discdiv <- merge(discdiv,snp,by="SNP")

  for (i in 1:22) {
    concddiv[which(concddiv[, "CHR"] == paste("chr",i,sep="")),17] <- i
    discdiv[which(discdiv[, "CHR"] == paste("chr",i,sep="")),17] <- i
  }

  names(concddiv)[17] <- "chrnum"
  names(discdiv)[17] <- "chrnum"

  concddiv <- concddiv[order(concddiv$BP),]
  concddiv <- concddiv[order(concddiv$chrnum),]
  discdiv <- discdiv[order(discdiv$BP),]
  discdiv <- discdiv[order(discdiv$chrnum),]
  concddiv <- concddiv[which(is.na(concddiv[, "chrnum"]) == FALSE),]
  discdiv <- discdiv[which(is.na(discdiv[, "chrnum"]) == FALSE),]

  write.table(concddiv,file=paste(numcut,"n",depvar,"on",indvar,"conc.gwas",sep=""),quo
te=FALSE,row.names=FALSE)
  write.table(discdiv,file=paste(numcut,"n",depvar,"on",indvar,"disc.gwas",sep=""),quot
e=FALSE,row.names=FALSE)

  assign("concddiv",concddiv,envir=globalenv())
  assign("discdiv",discdiv,envir=globalenv())
  assign("highn",highn,envir=globalenv())
}
```

### **MAKEHITSPEC**

```
makehitspec <- function(locsum,flank) {
```

```

hitspec <- as.data.frame(matrix("na",nrow=nrow(locsum),ncol=7))
names(hitspec) <- c("Feature","chr","start","end","flank","plot","arguments")
hitspec[,1] <- locsum[,1]
hitspec[,5] <- paste(flank,"kb",sep="")
hitspec[,6] <- "yes"

}

```

### **MAKELZSET**

```

makeLZset <- function(locsum,flank,sourcedat) {
#takes a list of loci and builds flanking regions of size flank from SNPs in sourcedat
lzdat <- sourcedat[1,]
names(locsum)[1] <- "SNP"
  for (t in 1:nrow(locsum)) {
      flankloc <- sourcedat[which(sourcedat[,"BP"] > (locsum[t,"BP"]-(flank/2)) &
sourcedat[,"CHR"] == locsum[t,"CHR"] & sourcedat[,"BP"] < (locsum[t,"BP"]+(flank/2)) ),]
      lzdat <- rbind(lzdat,flankloc)
  }
lzdat <- lzdat[-1,]
assign("lzdat",lzdat,envir=globalenv())
}

```

## BIBLIOGRAPHY

- Alazami AM, Al-Owain M, Alzahrani F, Shuaib T, Al-Shamrani H, Al-Falki YH, Al-Qahtani SM, Alsheddi T, Colak D, Alkuraya FS (2012). Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. *Human mutation*. **33**, 1429-1434.
- Albrecht E, Sillanpaa E, Karrasch S, Alves AC, Codd V, Hovatta I, Buxton JL, Nelson CP, Broer L, Hagg S, Mangino M, Willemsen G, Surakka I, Ferreira MA, Amin N, Oostra BA, Backmand HM, Peltonen M, Sarna S, Rantanen T, Sipila S, Korhonen T, Madden PA, Gieger C, Jorres RA, Heinrich J, Behr J, Huber RM, Peters A, Strauch K, Wichmann HE, Waldenberger M, Blakemore AI, de Geus EJ, Nyholt DR, Henders AK, Piirila PL, Rissanen A, Magnusson PK, Vinuela A, Pietilainen KH, Martin NG, Pedersen NL, Boomsma DI, Spector TD, van Duijn CM, Kaprio J, Samani NJ, Jarvelin MR, Schulz H (2014). Telomere length in circulating leukocytes is associated with lung function and disease. *The European respiratory journal*. **43**, 983-992.
- Almuqbil M, Hamdan FF, Mathonnet G, Rosenblatt B, Srour M (2013). De novo deletion of FMN2 in a girl with mild non-syndromic intellectual disability. *European journal of medical genetics*. **56**, 686-688.
- Andersen JN, Sathyanarayanan S, Di Bacco A, Chi A, Zhang T, Chen AH, Dolinski B, Kraus M, Roberts B, Arthur W, Klinghoffer RA, Gargano D, Li L, Feldman I, Lynch B, Rush J, Hendrickson RC, Blume-Jensen P, Paweletz CP (2010). Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors. *Sci Transl Med*. **2**, 43ra55.
- Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin research : the official journal of the International Society for Twin Studies*. **4**, 464-477.
- Anic GM, Sondak VK, Messina JL, Fenske NA, Zager JS, Cherpelis BS, Lee JH, Fulp WJ, Epling-Burnette PK, Park JY, Rollison DE (2013). Telomere length and risk of melanoma, squamous cell carcinoma, and basal cell carcinoma. *Cancer epidemiology*. **37**, 434-439.
- Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, Greider CW (2005). Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proceedings of the National Academy of Sciences of the United States of America*. **102**, 15960-15964.
- Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, Coker C, Krauskopf A, Kupiec M, McEachern MJ (2004). A genome-wide screen for *Saccharomyces cerevisiae*

- deletion mutants that affect telomere length. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 8658-8663.
- Aston KI, Hunt SC, Susser E, Kimura M, Factor-Litvak P, Carrell D, Aviv A (2012). Divergence of sperm and leukocyte age-dependent telomere dynamics: implications for male-driven evolution of telomere length in humans. *Molecular human reproduction*. **18**, 517-522.
- Aviv A, Chen W, Gardner JP, Kimura M, Brimacombe M, Cao X, Srinivasan SR, Berenson GS (2009). Leukocyte telomere dynamics: longitudinal findings among young adults in the Bogalusa Heart Study. *American journal of epidemiology*. **169**, 323-329.
- Aviv A, Susser E (2013). Leukocyte telomere length and the father's age enigma: implications for population health and for life course. *International journal of epidemiology*. **42**, 457-462.
- Avni D, Biberman Y, Meyuhos O (1997). The 5' terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. *Nucleic acids research*. **25**, 995-1001.
- Bae-Jump VL, Zhou C, Gehrig PA, Whang YE, Boggess JF (2006). Rapamycin inhibits hTERT telomerase mRNA expression, independent of cell cycle arrest. *Gynecologic oncology*. **100**, 487-494.
- Barszczyk M, Buczkowicz P, Castelo-Branco P, Mack SC, Ramaswamy V, Mangerel J, Agnihotri S, Remke M, Golbourn B, Pajovic S, Elizabeth C, Yu M, Luu B, Morrison A, Adamski J, Nethery-Brokk K, Li XN, Van Meter T, Dirks PB, Rutka JT, Taylor MD, Tabori U, Hawkins C (2014). Telomerase inhibition abolishes the tumorigenicity of pediatric ependymoma tumor-initiating cells. *Acta neuropathologica*. **128**, 863-877.
- Batista LF, Pech MF, Zhong FL, Nguyen HN, Xie KT, Zaug AJ, Crary SM, Choi J, Sebastiano V, Cherry A, Giri N, Wernig M, Alter BP, Cech TR, Savage SA, Reijo Pera RA, Artandi SE (2011). Telomere shortening and loss of self-renewal in dyskeratosis congenita induced pluripotent stem cells. *Nature*. **474**, 399-402.
- Bayfield MA, Yang R, Maraia RJ (2010). Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). *Biochimica et biophysica acta*. **1799**, 365-378.
- Beirne C, Delahay R, Hares M, Young A (2014). Age-related declines and disease-associated variation in immune cell telomere length in a wild mammal. *PloS one*. **9**, e108964.
- Bello NF, Lamsoul I, Heuze ML, Metais A, Moreaux G, Calderwood DA, Duprez D, Moog-Lutz C, Lutz PG (2009). The E3 ubiquitin ligase specificity subunit ASB2beta is a novel regulator of muscle differentiation that targets filamin B to proteasomal degradation. *Cell death and differentiation*. **16**, 921-932.
- Bendix L, Horn PB, Jensen UB, Rubelj I, Kolvraa S (2010). The load of short telomeres, estimated by a new method, Universal STELA, correlates with number of senescent cells. *Aging cell*. **9**, 383-397.
- Benetos A, Kark JD, Susser E, Kimura M, Sinnreich R, Chen W, Steenstrup T, Christensen K, Herbig U, von Bornemann Hjelmberg J, Srinivasan SR, Berenson

- GS, Labat C, Aviv A (2013). Tracking and fixed ranking of leukocyte telomere length across the adult life course. *Aging cell*. **12**, 615-621.
- Bezzi M, Teo SX, Muller J, Mok WC, Sahu SK, Vardy LA, Bonday ZQ, Guccione E (2013). Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes & development*. **27**, 1903-1916.
- Bliese P (2006). Multilevel Modeling in R (2.2)—A Brief Introduction to R, the multilevel package and the nlme package<sup>^</sup>eds): October.
- Bojesen SE (2013). Telomeres and human health. *Journal of internal medicine*. **274**, 399-413.
- Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, Edwards SL, Pickett HA, Shen HC, Smart CE, Hillman KM, Mai PL, Lawrenson K, Stutz MD, Lu Y, Karevan R, Woods N, Johnston RL, French JD, Chen X, Weischer M, Nielsen SF, Maranian MJ, Ghoussaini M, Ahmed S, Baynes C, Bolla MK, Wang Q, Dennis J, McGuffog L, Barrowdale D, Lee A, Healey S, Lush M, Tessier DC, Vincent D, Bacot F, Australian Cancer S, Australian Ovarian Cancer S, Kathleen Cuninghame Foundation Consortium for Research into Familial Breast C, Gene Environment I, Breast C, Swedish Breast Cancer S, Hereditary B, Ovarian Cancer Research Group N, Epidemiological study of B, Carriers BM, Genetic Modifiers of Cancer Risk in BMC, Vergote I, Lambrechts S, Despierre E, Risch HA, Gonzalez-Neira A, Rossing MA, Pita G, Doherty JA, Alvarez N, Larson MC, Fridley BL, Schoof N, Chang-Claude J, Cicek MS, Peto J, Kalli KR, Broeks A, Armasu SM, Schmidt MK, Braaf LM, Winterhoff B, Nevanlinna H, Konecny GE, Lambrechts D, Rogmann L, Guenel P, Teoman A, Milne RL, Garcia JJ, Cox A, Shridhar V, Burwinkel B, Marme F, Hein R, Sawyer EJ, Haiman CA, Wang-Gohrke S, Andrulis IL, Moysich KB, Hopper JL, Odunsi K, Lindblom A, Giles GG, Brenner H, Simard J, Lurie G, Fasching PA, Carney ME, Radice P, Wilkens LR, Swerdlow A, Goodman MT, Brauch H, Garcia-Closas M, Hillemanns P, Winqvist R, Durst M, Devilee P, Runnebaum I, Jakubowska A, Lubinski J, Mannermaa A, Butzow R, Bogdanova NV, Dork T, Pelttari LM, Zheng W, Leminen A, Anton-Culver H, Bunker CH, Kristensen V, Ness RB, Muir K, Edwards R, Meindl A, Heitz F, Matsuo K, du Bois A, Wu AH, Harter P, Teo SH, Schwaab I, Shu XO, Blot W, Hosono S, Kang D, Nakanishi T, Hartman M, Yatabe Y, Hamann U, Karlan BY, Sangrajrang S, Kjaer SK, Gaborieau V, Jensen A, Eccles D, Hogdall E, Shen CY, Brown J, Woo YL, Shah M, Azmi MA, Luben R, Omar SZ, Czene K, Vierkant RA, Nordestgaard BG, Flyger H, Vachon C, Olson JE, Wang X, Levine DA, Rudolph A, Weber RP, Flesch-Janys D, Iversen E, Nickels S, Schildkraut JM, Silva Idos S, Cramer DW, Gibson L, Terry KL, Fletcher O, Vitonis AF, van der Schoot CE, Poole EM, Hogervorst FB, Tworoger SS, Liu J, Bandera EV, Li J, Olson SH, Humphreys K, Orlow I, Blomqvist C, Rodriguez-Rodriguez L, Aittomaki K, Salvesen HB, Muranen TA, Wik E, Brouwers B, Krakstad C, Wauters E, Halle MK, Wildiers H, Kiemeny LA, Mulot C, Aben KK, Laurent-Puig P, Altena AM, Truong T, Massuger LF, Benitez J, Pejovic T, Perez JI, Hoatlin M, Zamora MP, Cook LS, Balasubramanian SP,

Kelemen LE, Schneeweiss A, Le ND, Sohn C, Brooks-Wilson A, Tomlinson I, Kerin MJ, Miller N, Cybulski C, Henderson BE, Menkiszak J, Schumacher F, Wentzensen N, Le Marchand L, Yang HP, Mulligan AM, Glendon G, Engelholm SA, Knight JA, Hogdall CK, Apicella C, Gore M, Tsimiklis H, Song H, Southey MC, Jager A, den Ouweland AM, Brown R, Martens JW, Flanagan JM, Kriege M, Paul J, Margolin S, Siddiqui N, Severi G, Whittemore AS, Baglietto L, McGuire V, Stegmaier C, Sieh W, Muller H, Arndt V, Labreche F, Gao YT, Goldberg MS, Yang G, Dumont M, McLaughlin JR, Hartmann A, Ekici AB, Beckmann MW, Phelan CM, Lux MP, Permuth-Wey J, Peissel B, Sellers TA, Ficarazzi F, Barile M, Ziogas A, Ashworth A, Gentry-Maharaj A, Jones M, Ramus SJ, Orr N, Menon U, Pearce CL, Bruning T, Pike MC, Ko YD, Lissowska J, Figueroa J, Kupryjanczyk J, Chanock SJ, Dansonka-Mieszkowska A, Jukkola-Vuorinen A, Rzepecka IK, Pylkas K, Bidzinski M, Kauppila S, Hollestelle A, Seynaeve C, Tollenaar RA, Durda K, Jaworska K, Hartikainen JM, Kosma VM, Kataja V, Antonenkova NN, Long J, Shrubsole M, Deming-Halverson S, Lophatananon A, Siriwanarangsana P, Stewart-Brown S, Ditsch N, Lichtner P, Schmutzler RK, Ito H, Iwata H, Tajima K, Tseng CC, Stram DO, van den Berg D, Yip CH, Ikram MK, Teh YC, Cai H, Lu W, Signorello LB, Cai Q, Noh DY, Yoo KY, Miao H, Iau PT, Teo YY, McKay J, Shapero C, Ademuyiwa F, Fountzilias G, Hsiung CN, Yu JC, Hou MF, Healey CS, Luccarini C, Peock S, Stoppa-Lyonnet D, Peterlongo P, Rebbeck TR, Piedmonte M, Singer CF, Friedman E, Thomassen M, Offit K, Hansen TV, Neuhausen SL, Szabo CI, Blanco I, Garber J, Narod SA, Weitzel JN, Montagna M, Olah E, Godwin AK, Yannoukakos D, Goldgar DE, Caldes T, Imyanitov EN, Tihomirova L, Arun BK, Campbell I, Mensenkamp AR, van Asperen CJ, van Roozendaal KE, Meijers-Heijboer H, Collee JM, Oosterwijk JC, Hooning MJ, Rookus MA, van der Luijt RB, Os TA, Evans DG, Frost D, Fineberg E, Barwell J, Walker L, Kennedy MJ, Platte R, Davidson R, Ellis SD, Cole T, Bressac-de Paillerets B, Buecher B, Damiola F, Faivre L, Frenay M, Sinilnikova OM, Caron O, Giraud S, Mazoyer S, Bonadona V, Caux-Moncoutier V, Toloczko-Grabarek A, Gronwald J, Byrski T, Spurdle AB, Bonanni B, Zaffaroni D, Giannini G, Bernard L, Dolcetti R, Manoukian S, Arnold N, Engel C, Deissler H, Rhiem K, Niederacher D, Plendl H, Sutter C, Wappenschmidt B, Borg A, Melin B, Rantala J, Soller M, Nathanson KL, Domchek SM, Rodriguez GC, Salani R, Kaulich DG, Tea MK, Paluch SS, Laitman Y, Skytte AB, Kruse TA, Jensen UB, Robson M, Gerdes AM, Ejlertsen B, Foretova L, Savage SA, Lester J, Soucy P, Kuchenbaecker KB, Olswold C, Cunningham JM, Slager S, Pankratz VS, Dicks E, Lakhani SR, Couch FJ, Hall P, Monteiro AN, Gayther SA, Pharoah PD, Reddel RR, Goode EL, Greene MH, Easton DF, Berchuck A, Antoniou AC, Chenevix-Trench G, Dunning AM (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nature genetics*. **45**, 371-384, 384e371-372.

Boyer L, Chouaid C, Bastuji-Garin S, Marcos E, Margarit L, Le Corvoisier P, Vervoitte L, Hamidou L, Frih L, Audureau E, Covali-Noroc A, Andujar P, Saakashvili Z, Lino A, Ghaleh B, Hue S, Derumeaux G, Housset B, Dubois-Rande JL,

- Boczkowski J, Maitre B, Adnot S (2015). Aging-Related Systemic Manifestations in COPD Patients and Cigarette Smokers. *PloS one*. **10**, e0121539.
- Bray I, Gunnell D, Davey Smith G (2006). Advanced paternal age: how old is too old? *Journal of epidemiology and community health*. **60**, 851-853.
- Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, Albrecht E, Amin N, Beekman M, de Geus EJ, Henders A, Nelson CP, Steves CJ, Wright MJ, de Craen AJ, Isaacs A, Matthews M, Moayyeri A, Montgomery GW, Oostra BA, Vink JM, Spector TD, Slagboom PE, Martin NG, Samani NJ, van Duijn CM, Boomsma DI (2013). Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. *European journal of human genetics : EJHG*. **21**, 1163-1168.
- Cardin R, Piciocchi M, Tieppo C, Maddalo G, Zaninotto G, Mescoli C, Rugge M, Farinati F (2013). Oxidative DNA damage in Barrett mucosa: correlation with telomeric dysfunction and p53 mutation. *Annals of surgical oncology*. **20 Suppl 3**, S583-589.
- Casado FL, Singh KP, Gasiewicz TA (2011). Aryl hydrocarbon receptor activation in hematopoietic stem/progenitor cells alters cell function and pathway-specific gene modulation reflecting changes in cellular trafficking and migration. *Molecular pharmacology*. **80**, 673-682.
- Catlin SN, Busque L, Gale RE, Gutter P, Abkowitz JL (2011). The replication rate of human hematopoietic stem cells in vivo. *Blood*. **117**, 4460-4466.
- Chadeneau C, Siegel P, Harley CB, Muller WJ, Bacchetti S (1995). Telomerase activity in normal and malignant murine tissues. *Oncogene*. **11**, 893-898.
- Chen G, Mulla WA, Kucharavy A, Tsai HJ, Rubinstein B, Conkright J, McCroskey S, Bradford WD, Weems L, Haug JS, Seidel CW, Berman J, Li R (2015). Targeting the adaptability of heterogeneous aneuploids. *Cell*. **160**, 771-784.
- Chen S, Yeh F, Lin J, Matsuguchi T, Blackburn E, Lee ET, Howard BV, Zhao J (2014). Short leukocyte telomere length is associated with obesity in American Indians: the Strong Heart Family study. *Aging*. **6**, 380-389.
- Chen W, Kimura M, Kim S, Cao X, Srinivasan SR, Berenson GS, Kark JD, Aviv A (2011). Longitudinal versus cross-sectional evaluations of leukocyte telomere length dynamics: age-dependent telomere shortening is the rule. *The journals of gerontology. Series A, Biological sciences and medical sciences*. **66**, 312-319.
- Chiang YJ, Calado RT, Hathcock KS, Lansdorp PM, Young NS, Hodes RJ (2010). Telomere length is inherited with resetting of the telomere set-point. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 10148-10153.
- Chung J, Khadka P, Chung IK (2012). Nuclear import of hTERT requires a bipartite nuclear localization signal and Akt-mediated phosphorylation. *Journal of cell science*. **125**, 2684-2697.
- Cunha SR, Mohler PJ (2008). Obscurin targets ankyrin-B and protein phosphatase 2A to the cardiac M-line. *The Journal of biological chemistry*. **283**, 31968-31980.
- D'Andrea AD, Grompe M (2003). The Fanconi anaemia/BRCA pathway. *Nature reviews. Cancer*. **3**, 23-34.

- Dai Q, Luan G, Deng L, Lei T, Kang H, Song X, Zhang Y, Xiao ZX, Li Q (2014). Primordial dwarfism gene maintains Lin28 expression to safeguard embryonic stem cells from premature differentiation. *Cell reports*. **7**, 735-746.
- Damle RN, Temburni S, Calissano C, Yancopoulos S, Banapour T, Sison C, Allen SL, Rai KR, Chiorazzi N (2007). CD38 expression labels an activated subset within chronic lymphocytic leukemia clones enriched in proliferating B cells. *Blood*. **110**, 3352-3359.
- Daniali L, Benetos A, Susser E, Kark JD, Labat C, Kimura M, Desai K, Granick M, Aviv A (2013). Telomeres shorten at equivalent rates in somatic tissues of adults. *Nature communications*. **4**, 1597.
- De Felice B, Nappi C, Zizolfi B, Guida M, Di Spiezio Sardo A, Bifulco G, Guida M (2012). Telomere shortening in women resident close to waste landfill sites. *Gene*. **500**, 101-106.
- De Meyer T, Eisenberg DT (2014). Possible technical and biological explanations for the 'parental telomere length inheritance discrepancy' enigma. *European journal of human genetics : EJHG*.
- De Meyer T, Rietzschel ER, De Buyzere ML, De Bacquer D, Van Criekinge W, De Backer GG, Gillebert TC, Van Oostveldt P, Bekaert S, Asklepios i (2007). Paternal age at birth is an important determinant of offspring telomere length. *Human molecular genetics*. **16**, 3097-3102.
- De Meyer T, Vandepitte K, Denil S, De Buyzere ML, Rietzschel ER, Bekaert S (2014). A non-genetic, epigenetic-like mechanism of telomere length inheritance? *European journal of human genetics : EJHG*. **22**, 10-11.
- Diaz de Leon A, Cronkhite JT, Katzenstein AL, Godwin JD, Raghu G, Glazer CS, Rosenblatt RL, Girod CE, Garrity ER, Xing C, Garcia CK (2010). Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. *PloS one*. **5**, e10680.
- Diaz de Leon A, Cronkhite JT, Yilmaz C, Brewington C, Wang R, Xing C, Hsia CC, Garcia CK (2011). Subclinical lung disease, macrocytosis, and premature graying in kindreds with telomerase (TERT) mutations. *Chest*. **140**, 753-763.
- Eder W, Ege MJ, von Mutius E (2006). The asthma epidemic. *The New England journal of medicine*. **355**, 2226-2235.
- Egan ED, Collins K (2012). An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Molecular and cellular biology*. **32**, 2428-2439.
- Ehrlenbach S, Willeit P, Kiechl S, Willeit J, Reindl M, Schanda K, Kronenberg F, Brandstatter A (2009). Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled high-throughput assay. *International journal of epidemiology*. **38**, 1725-1734.
- Eisenberg DT, Hayes MG, Kuzawa CW (2012). Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. *Proceedings of the National Academy of Sciences of the United States of America*. **109**, 10251-10256.

- Eisenberg DT, Salpea KD, Kuzawa CW, Hayes MG, Humphries SE, European Atherosclerosis Research Study IIG (2011). Substantial variation in qPCR measured mean blood telomere lengths in young men from eleven European countries. *American journal of human biology : the official journal of the Human Biology Council.* **23**, 228-231.
- Entringer S, Epel ES, Lin J, Blackburn EH, Buss C, Simhan HN, Wadhwa PD (2015). Maternal estriol concentrations in early gestation predict infant telomere length. *The Journal of clinical endocrinology and metabolism.* **100**, 267-273.
- Eshkoo SA, Ismail P, Rahman SA, Adon MY, Devan RV (2013). Contribution of CYP2E1 polymorphism to aging in the mechanical workshop workers. *Toxicology mechanisms and methods.* **23**, 217-222.
- Farzaneh-Far R, Lin J, Epel E, Lapham K, Blackburn E, Whooley MA (2010). Telomere length trajectory and its determinants in persons with coronary artery disease: longitudinal findings from the heart and soul study. *PLoS one.* **5**, e8612.
- Ferlin A, Rampazzo E, Rocca MS, Keppel S, Frigo AC, De Rossi A, Foresta C (2013). In young men sperm telomere length is related to sperm number and parental age. *Human reproduction.* **28**, 3370-3376.
- Fitzpatrick AL, Kronmal RA, Kimura M, Gardner JP, Psaty BM, Jenny NS, Tracy RP, Hardikar S, Aviv A (2011). Leukocyte telomere length and mortality in the Cardiovascular Health Study. *The journals of gerontology. Series A, Biological sciences and medical sciences.* **66**, 421-429.
- Ford LP, Shay JW, Wright WE (2001). The La antigen associates with the human telomerase ribonucleoprotein and influences telomere length in vivo. *Rna.* **7**, 1068-1075.
- Forsythe HL, Jarvis JL, Turner JW, Elmore LW, Holt SE (2001). Stable association of hsp90 and p23, but Not hsp70, with active human telomerase. *The Journal of biological chemistry.* **276**, 15571-15574.
- Frank EG, Woodgate R (2007). Increased catalytic activity and altered fidelity of human DNA polymerase  $\epsilon$  in the presence of manganese. *The Journal of biological chemistry.* **282**, 24689-24696.
- Friedman DR, Lanasa MC, Davis PH, Allgood SD, Matta KM, Brander DM, Chen Y, Davis ED, Volkheimer AD, Moore JO, Gockerman JP, Sportelli P, Weinberg JB (2014). Perifosine treatment in chronic lymphocytic leukemia: results of a phase II clinical trial and in vitro studies. *Leukemia & lymphoma.* **55**, 1067-1075.
- Gadalla SM, Cawthon R, Giri N, Alter BP, Savage SA (2010). Telomere length in blood, buccal cells, and fibroblasts from patients with inherited bone marrow failure syndromes. *Aging.* **2**, 867-874.
- Gardner JP, Li S, Srinivasan SR, Chen W, Kimura M, Lu X, Berenson GS, Aviv A (2005). Rise in insulin resistance is associated with escalated telomere attrition. *Circulation.* **111**, 2171-2177.
- Gatbonton T, Imbesi M, Nelson M, Akey JM, Ruderfer DM, Kruglyak L, Simon JA, Bedalov A (2006). Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS genetics.* **2**, e35.

- Geoffroy-Siraudin C, Loundou AD, Romain F, Achard V, Courbiere B, Perrard MH, Durand P, Guichaoua MR (2012). Decline of semen quality among 10 932 males consulting for couple infertility over a 20-year period in Marseille, France. *Asian journal of andrology*. **14**, 584-590.
- Gills JJ, Dennis PA (2009). Perifosine: update on a novel Akt inhibitor. *Current oncology reports*. **11**, 102-110.
- Gleeson M, O'Marcaigh A, Cotter M, Brosnahan D, Vulliamy T, Smith OP (2012). Retinal vasculopathy in autosomal dominant dyskeratosis congenita due to TINF2 mutation. *British journal of haematology*. **159**, 498.
- Gomes NM, Ryder OA, Houck ML, Charter SJ, Walker W, Forsyth NR, Austad SN, Venditti C, Pagel M, Shay JW, Wright WE (2011). Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging cell*. **10**, 761-768.
- Guan JZ, Guan WP, Maeda T, Makino N (2012). Different levels of hypoxia regulate telomere length and telomerase activity. *Aging clinical and experimental research*. **24**, 213-217.
- Haendeler J, Hoffmann J, Rahman S, Zeiher AM, Dimmeler S (2003). Regulation of telomerase activity and anti-apoptotic function by protein-protein interaction and phosphorylation. *FEBS letters*. **536**, 180-186.
- Han Y, Zhan Y, Hou G, Li L (2014). Cyclin-dependent kinase 9 may as a novel target in downregulating the atherosclerosis inflammation (Review). *Biomedical reports*. **2**, 775-779.
- Harari Y, Romano GH, Ungar L, Kupiec M (2013). Nature vs nurture: interplay between the genetic control of telomere length and environmental factors. *Cell cycle*. **12**, 3465-3470.
- Harrington L, McPhail T, Mar V, Zhou W, Oulton R, Bass MB, Arruda I, Robinson MO (1997). A mammalian telomerase-associated protein. *Science*. **275**, 973-977.
- Hemann MT, Strong MA, Hao LY, Greider CW (2001). The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell*. **107**, 67-77.
- Henson SM, Macaulay R, Riddell NE, Nunn CJ, Akbar AN (2015). Blockade of PD-1 or p38 MAP Kinase signalling pathways enhances senescent human CD8 T-cell proliferation by distinct pathways. *European journal of immunology*.
- Herbert BS, Shay JW, Wright WE (2003). Analysis of telomeres and telomerase. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]*. **Chapter 18**, Unit 18 16.
- Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, Rao DC, Sprafka JM, Williams R (1996). NHLBI Family Heart Study: objectives and design. *American journal of epidemiology*. **143**, 1219-1228.
- Holohan B, Wright WE, Shay JW (2014). Cell biology of disease: Telomeropathies: an emerging spectrum disorder. *The Journal of cell biology*. **205**, 289-299.
- Hoshiyama H, Tang J, Batten K, Xiao G, Rouillard JM, Shay JW, Xie Y, Wright WE (2012). Development of methods for quantitative comparison of pooled shRNAs by mass sequencing. *Journal of biomolecular screening*. **17**, 258-265.

- Hou L, Wang S, Dou C, Zhang X, Yu Y, Zheng Y, Avula U, Hoxha M, Diaz A, McCracken J, Barretta F, Marinelli B, Bertazzi PA, Schwartz J, Baccarelli AA (2012). Air pollution exposure and telomere length in highly exposed subjects in Beijing, China: a repeated-measure study. *Environment international*. **48**, 71-77.
- Houben JM, Giltay EJ, Rius-Ottenheim N, Hageman GJ, Kromhout D (2011). Telomere length and mortality in elderly men: the Zutphen Elderly Study. *The journals of gerontology. Series A, Biological sciences and medical sciences*. **66**, 38-44.
- Hoxha M, Dioni L, Bonzini M, Pesatori AC, Fustinoni S, Cavallo D, Carugno M, Albetti B, Marinelli B, Schwartz J, Bertazzi PA, Baccarelli A (2009). Association between leukocyte telomere shortening and exposure to traffic pollution: a cross-sectional study on traffic officers and indoor office workers. *Environmental health : a global access science source*. **8**, 41.
- Hu J, Lu J, Lian G, Ferland RJ, Dettenhofer M, Sheen VL (2014a). Formin 1 and filamin B physically interact to coordinate chondrocyte proliferation and differentiation in the growth plate. *Human molecular genetics*. **23**, 4663-4673.
- Hu Y, Bobb D, Lu Y, He J, Dome JS (2014b). Effect of telomerase inhibition on preclinical models of malignant rhabdoid tumor. *Cancer genetics*. **207**, 403-411.
- Huzen J, Wong LS, van Veldhuisen DJ, Samani NJ, Zwinderman AH, Codd V, Cawthon RM, Benus GF, van der Horst IC, Navis G, Bakker SJ, Gansevoort RT, de Jong PE, Hillege HL, van Gilst WH, de Boer RA, van der Harst P (2014). Telomere length loss due to smoking and metabolic traits. *Journal of internal medicine*. **275**, 155-163.
- Jensen TK, Carlsen E, Jorgensen N, Berthelsen JG, Keiding N, Christensen K, Petersen JH, Knudsen LB, Skakkebaek NE (2002). Poor semen quality may contribute to recent decline in fertility rates. *Human reproduction*. **17**, 1437-1440.
- Jiang X, Dong M, Cheng J, Huang S, He Y, Ma K, Tang B, Guo Y (2013). Decreased leukocyte telomere length (LTL) is associated with stroke but unlikely to be causative. *PloS one*. **8**, e68254.
- Joksic I, Vujic D, Guc-Scekic M, Leskovac A, Petrovic S, Ojani M, Trujillo JP, Surralles J, Zivkovic M, Stankovic A, Slijepcevic P, Joksic G (2012). Dysfunctional telomeres in primary cells from Fanconi anemia FANCD2 patients. *Genome integrity*. **3**, 6.
- Joseph I, Tressler R, Bassett E, Harley C, Buseman CM, Pattamatta P, Wright WE, Shay JW, Go NF (2010). The telomerase inhibitor imetelstat depletes cancer stem cells in breast and pancreatic cancer cell lines. *Cancer research*. **70**, 9494-9504.
- Kang SS, Kwon T, Kwon DY, Do SI (1999). Akt protein kinase enhances human telomerase activity through phosphorylation of telomerase reverse transcriptase subunit. *The Journal of biological chemistry*. **274**, 13085-13090.
- Kao DJ, Li AH, Chen JC, Luo RS, Chen YL, Lu JC, Wang HL (2012). CC chemokine ligand 2 upregulates the current density and expression of TRPV1 channels and Nav1.8 sodium channels in dorsal root ganglion neurons. *Journal of neuroinflammation*. **9**, 189.
- Kim HS, Mendiratta S, Kim J, Pecot CV, Larsen JE, Zubovych I, Seo BY, Kim J, Eskiocak B, Chung H, McMillan E, Wu S, De Brabander J, Komurov K, Toombs

- JE, Wei S, Peyton M, Williams N, Gazdar AF, Posner BA, Brekken RA, Sood AK, Deberardinis RJ, Roth MG, Minna JD, White MA (2013). Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer. *Cell*. **155**, 552-566.
- Kim JI, Kim HY, Kim SM, Lee SA, Son YH, Eo SK, Rhim BY, Kim K (2011). Extracellular Nucleotides Can Induce Chemokine (C-C motif) Ligand 2 Expression in Human Vascular Smooth Muscle Cells. *The Korean journal of physiology & pharmacology : official journal of the Korean Physiological Society and the Korean Society of Pharmacology*. **15**, 31-36.
- Kimura M, Cherkas LF, Kato BS, Demissie S, Hjelmberg JB, Brimacombe M, Cupples A, Hunkin JL, Gardner JP, Lu X, Cao X, Sastrasin M, Province MA, Hunt SC, Christensen K, Levy D, Spector TD, Aviv A (2008). Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. *PLoS genetics*. **4**, e37.
- Komlos J, Breitfelder A (2008). Height of US-born non-Hispanic children and adolescents ages 2-19, born 1942-2002 in the NHANES samples. *American journal of human biology : the official journal of the Human Biology Council*. **20**, 66-71.
- Kondo M (2010). Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Immunological reviews*. **238**, 37-46.
- Kume K, Kikukawa M, Hanyu H, Takata Y, Umahara T, Sakurai H, Kanetaka H, Ohyashiki K, Ohyashiki JH, Iwamoto T (2012). Telomere length shortening in patients with dementia with Lewy bodies. *European journal of neurology : the official journal of the European Federation of Neurological Societies*. **19**, 905-910.
- Kupfer G, Naf D, Garcia-Higuera I, Wasik J, Cheng A, Yamashita T, Tipping A, Morgan N, Mathew CG, D'Andrea AD (1999). A patient-derived mutant form of the Fanconi anemia protein, FANCA, is defective in nuclear accumulation. *Experimental hematology*. **27**, 587-593.
- Lackner DH, Hayashi MT, Cesare AJ, Karlseder J (2014). A genomics approach identifies senescence-specific gene expression regulation. *Aging cell*. **13**, 946-950.
- Law R, Dixon-Salazar T, Jerber J, Cai N, Abbasi AA, Zaki MS, Mittal K, Gabriel SB, Rafiq MA, Khan V, Nguyen M, Ali G, Copeland B, Scott E, Vasli N, Mikhailov A, Khan MN, Andrade DM, Ayaz M, Ansar M, Ayub M, Vincent JB, Gleeson JG (2014). Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein Formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *American journal of human genetics*. **95**, 721-728.
- Leader B, Leder P (2000). Formin-2, a novel formin homology protein of the cappuccino subfamily, is highly expressed in the developing and adult central nervous system. *Mechanisms of development*. **93**, 221-231.
- Leader B, Lim H, Carabatsos MJ, Harrington A, Ecsedy J, Pellman D, Maas R, Leder P (2002). Formin-2, polyploidy, hypofertility and positioning of the meiotic spindle in mouse oocytes. *Nature cell biology*. **4**, 921-928.

- Lee TH, Tun-Kyi A, Shi R, Lim J, Soohoo C, Finn G, Balastik M, Pastorino L, Wulf G, Zhou XZ, Lu KP (2009). Essential role of Pin1 in the regulation of TRF1 stability and telomere maintenance. *Nature cell biology*. **11**, 97-105.
- Li X, Luwor R, Lu Y, Liang K, Fan Z (2006). Enhancement of antitumor activity of the anti-EGF receptor monoclonal antibody cetuximab/C225 by perifosine in PTEN-deficient cancer cells. *Oncogene*. **25**, 525-535.
- Lillard-Wetherell K, Combs KA, Groden J (2005). BLM helicase complements disrupted type II telomere lengthening in telomerase-negative sgs1 yeast. *Cancer research*. **65**, 5520-5522.
- Liu P, Xiang Y, Fujinaga K, Bartholomeeusen K, Nilson KA, Price DH, Peterlin BM (2014). Release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. *The Journal of biological chemistry*. **289**, 9918-9925.
- Ludlow AT, Lima LC, Wang J, Hanson ED, Guth LM, Spangenburg EE, Roth SM (2012). Exercise alters mRNA expression of telomere-repeat binding factor 1 in skeletal muscle via p38 MAPK. *Journal of applied physiology*. **113**, 1737-1746.
- Ludlow AT, Robin JD, Sayed M, Litterst CM, Shelton DN, Shay JW, Wright WE (2014). Quantitative telomerase enzyme activity determination using droplet digital PCR with single cell resolution. *Nucleic acids research*. **42**, e104.
- Lundblad V, Szostak JW (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*. **57**, 633-643.
- Lyakhovich A, Ramirez MJ, Castellanos A, Castella M, Simons AM, Parvin JD, Surralles J (2011). Fanconi anemia protein FANCD2 inhibits TRF1 polyADP-ribosylation through tankyrase1-dependent manner. *Genome integrity*. **2**, 4.
- Makridakis NM, Reichardt JK (2012). Translesion DNA polymerases and cancer. *Frontiers in genetics*. **3**, 174.
- Marian CO, Cho SK, McEllin BM, Maher EA, Hatanpaa KJ, Madden CJ, Mickey BE, Wright WE, Shay JW, Bachoo RM (2010). The telomerase antagonist, imetelstat, efficiently targets glioblastoma tumor-initiating cells leading to decreased proliferation and tumor growth. *Clinical cancer research : an official journal of the American Association for Cancer Research*. **16**, 154-163.
- Martinez-Delgado B, Yanowsky K, Inglada-Perez L, Domingo S, Urioste M, Osorio A, Benitez J (2011). Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS genetics*. **7**, e1002182.
- Martino L, Pennell S, Kelly G, Busi B, Brown P, Atkinson RA, Salisbury NJ, Ooi ZH, See KW, Smerdon SJ, Alfano C, Bui TT, Conte MR (2015). Synergic interplay of the La motif, RRM1 and the interdomain linker of LARP6 in the recognition of collagen mRNA expands the RNA binding repertoire of the La module. *Nucleic acids research*. **43**, 645-660.
- Mason JM, Sekiguchi JM (2011). Snm1B/Apollo functions in the Fanconi anemia pathway in response to DNA interstrand crosslinks. *Human molecular genetics*. **20**, 2549-2559.

- Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M, Araki M, Iwai S, Takio K, Hanaoka F (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature*. **399**, 700-704.
- Meinig JM, Peterson BR (2015). Anticancer/Antiviral Agent Akt Inhibitor-IV Massively Accumulates in Mitochondria and Potently Disrupts Cellular Bioenergetics. *ACS chemical biology*. **10**, 570-576.
- Mendrysa SM, Ghassemifar S, Malek R (2011). p53 in the CNS: Perspectives on Development, Stem Cells, and Cancer. *Genes & cancer*. **2**, 431-442.
- Menendez S, Goh AM, Camus S, Ng KW, Kua N, Badal V, Lane DP (2011). MDM4 downregulates p53 transcriptional activity and response to stress during differentiation. *Cell cycle*. **10**, 1100-1108.
- Mitchell C, Hobcraft J, McLanahan SS, Siegel SR, Berg A, Brooks-Gunn J, Garfinkel I, Notterman D (2014). Social disadvantage, genetic sensitivity, and children's telomere length. *Proceedings of the National Academy of Sciences of the United States of America*. **111**, 5944-5949.
- Mohler PJ, Gramolini AO, Bennett V (2002). The ankyrin-B C-terminal domain determines activity of ankyrin-B/G chimeras in rescue of abnormal inositol 1,4,5-trisphosphate and ryanodine receptor distribution in ankyrin-B (-/-) neonatal cardiomyocytes. *The Journal of biological chemistry*. **277**, 10599-10607.
- Montaville P, Jegou A, Pernier J, Compper C, Guichard B, Mogessie B, Schuh M, Romet-Lemonne G, Carlier MF (2014). Spire and Formin 2 synergize and antagonize in regulating actin assembly in meiosis by a ping-pong mechanism. *PLoS biology*. **12**, e1001795.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*. **34**, 267-273.
- Mura M, Hopkins TG, Michael T, Abd-Latip N, Weir J, Aboagye E, Mauri F, Jameson C, Sturge J, Gabra H, Bushell M, Willis AE, Curry E, Blagden SP (2014). LARP1 post-transcriptionally regulates mTOR and contributes to cancer progression. *Oncogene*.
- Murphy AJ, Sarrazy V, Wang N, Bijl N, Abramowicz S, Westerterp M, Welch CB, Schuetz JD, Yvan-Charvet L (2014). Deficiency of ATP-binding cassette transporter B6 in megakaryocyte progenitors accelerates atherosclerosis in mice. *Arteriosclerosis, thrombosis, and vascular biology*. **34**, 751-758.
- N'Jai A U, Larsen MC, Bushkofsky JR, Czuprynski CJ, Jefcoate CR (2011). Acute disruption of bone marrow hematopoiesis by benzo(a)pyrene is selectively reversed by aryl hydrocarbon receptor-mediated processes. *Molecular pharmacology*. **79**, 724-734.
- Nordfjall K, Svenson U, Norrback KF, Adolfsson R, Roos G (2010). Large-scale parent-child comparison confirms a strong paternal influence on telomere length. *European journal of human genetics : EJHG*. **18**, 385-389.

- Ocana A, Vera-Badillo F, Al-Mubarak M, Templeton AJ, Corrales-Sanchez V, Diez-Gonzalez L, Cuenca-Lopez MD, Seruga B, Pandiella A, Amir E (2014). Activation of the PI3K/mTOR/AKT pathway and survival in solid tumors: systematic review and meta-analysis. *PloS one*. **9**, e95219.
- Okamura D, Maeda I, Taniguchi H, Tokitake Y, Ikeda M, Ozato K, Mise N, Abe K, Noce T, Izipisua Belmonte JC, Matsui Y (2012). Cell cycle gene-specific control of transcription has a critical role in proliferation of primordial germ cells. *Genes & development*. **26**, 2477-2482.
- Oyama T, Harigaya K, Sasaki N, Okamura Y, Kokubo H, Saga Y, Hozumi K, Suganami A, Tamura Y, Nagase T, Koga H, Nishimura M, Sakamoto R, Sato M, Yoshida N, Kitagawa M (2011). Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo. *Development*. **138**, 5235-5246.
- Ozawa T (2001). Ryanodine-sensitive Ca<sup>2+</sup> release mechanism in non-excitable cells (Review). *International journal of molecular medicine*. **7**, 21-25.
- Pant V, Lozano G (2014). Dissecting the p53-Mdm2 feedback loop in vivo: uncoupling the role in p53 stability and activity. *Oncotarget*. **5**, 1149-1156.
- Pavesi E, Avondo F, Aspesi A, Quarello P, Rocci A, Vimercati C, Pigullo S, Dufour C, Ramenghi U, Dianzani I (2009). Analysis of telomeres in peripheral blood cells from patients with bone marrow failure. *Pediatric blood & cancer*. **53**, 411-416.
- Pellatt AJ, Wolff RK, Torres-Mejia G, John EM, Herrick JS, Lundgreen A, Baumgartner KB, Giuliano AR, Hines LM, Fejerman L, Cawthon R, Slattery ML (2013). Telomere length, telomere-related genes, and breast cancer risk: the breast cancer health disparities study. *Genes, chromosomes & cancer*. **52**, 595-609.
- Pessah IN, Stambuk RA, Casida JE (1987). Ca<sup>2+</sup>-activated ryanodine binding: mechanisms of sensitivity and intensity modulation by Mg<sup>2+</sup>, caffeine, and adenine nucleotides. *Molecular pharmacology*. **31**, 232-238.
- Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, Michailidou K, Tyrer JP, Benlloch S, Brown J, Audley T, Luben R, Khaw KT, Neal DE, Hamdy FC, Donovan JL, Kote-Jarai Z, Baynes C, Shah M, Bolla MK, Wang Q, Dennis J, Dicks E, Yang R, Rudolph A, Schildkraut J, Chang-Claude J, Burwinkel B, Chenevix-Trench G, Pharoah PD, Berchuck A, Eeles RA, Easton DF, Dunning AM, Nordestgaard BG (2013). A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. *Human molecular genetics*. **22**, 5056-5064.
- Pooley KA, McGuffog L, Barrowdale D, Frost D, Ellis SD, Fineberg E, Platte R, Izatt L, Adlard J, Bardwell J, Brewer C, Cole T, Cook J, Davidson R, Donaldson A, Dorkins H, Douglas F, Eason J, Houghton C, Kennedy MJ, McCann E, Miedzybrodzka Z, Murray A, Porteous ME, Rogers MT, Side LE, Tischkowitz M, Walker L, Hodgson S, Eccles DM, Morrison PJ, Evans DG, Eeles RA, Antoniou AC, Easton DF, Dunning AM, Embrace (2014). Lymphocyte telomere length is long in BRCA1 and BRCA2 mutation carriers regardless of cancer-affected status. *Cancer epidemiology, biomarkers & prevention : a publication of*

- the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* **23**, 1018-1024.
- Pope-Varsalona H, Liu FJ, Guzik L, Opresko PL (2014). Polymerase eta suppresses telomere defects induced by DNA damaging agents. *Nucleic acids research.* **42**, 13096-13109.
- Popuri V, Hsu J, Khadka P, Horvath K, Liu Y, Croteau DL, Bohr VA (2014). Human RECQL1 participates in telomere maintenance. *Nucleic acids research.* **42**, 5671-5688.
- Preacher KJ, Hayes AF (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods.* **40**, 879-891.
- Prescott J, Du M, Wong JY, Han J, De Vivo I (2012). Paternal age at birth is associated with offspring leukocyte telomere length in the nurses' health study. *Human reproduction.* **27**, 3622-3631.
- Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* **26**, 2336-2337.
- Raschenberger J, Kollerits B, Hammerer-Lercher A, Rantner B, Stadler M, Haun M, Klein-Weigel P, Fraedrich G, Kronenberg F (2013). The association of relative telomere length with symptomatic peripheral arterial disease: results from the CAVASIC study. *Atherosclerosis.* **229**, 469-474.
- Rietzschel ER, De Buyzere ML, Bekaert S, Segers P, De Bacquer D, Cooman L, Van Damme P, Cassiman P, Langlois M, van Oostveldt P, Verdonck P, De Backer G, Gillebert TC, Asklepios I (2007). Rationale, design, methods and baseline characteristics of the Asklepios Study. *European journal of cardiovascular prevention and rehabilitation : official journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology.* **14**, 179-191.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011). Integrative genomics viewer. *Nature biotechnology.* **29**, 24-26.
- Rode L, Nordestgaard BG, Weischer M, Bojesen SE (2014). Increased body mass index, elevated C-reactive protein, and short telomere length. *The Journal of clinical endocrinology and metabolism*, jc20141161.
- Roig AI, Eskiocak U, Hight SK, Kim SB, Delgado O, Souza RF, Spechler SJ, Wright WE, Shay JW (2010). Immortalized epithelial cells derived from human colon biopsies express stem cell markers and differentiate in vitro. *Gastroenterology.* **138**, 1012-1021 e1011-1015.
- Romero CX, Romero TE, Shlay JC, Ogden LG, Dabelea D (2012). Changing trends in the prevalence and disparities of obesity and other cardiovascular disease risk factors in three racial/ethnic groups of USA adults. *Advances in preventive medicine.* **2012**, 172423.
- Ryley DA, Wu HH, Leader B, Zimon A, Reindollar RH, Gray MR (2005). Characterization and mutation analysis of the human formin-2 (FMN2) gene in women with unexplained infertility. *Fertility and sterility.* **83**, 1363-1371.

- Savi M, Neri TM, Zavaroni I, Coscelli I (1977). HLA antigens and insulin dependent diabetes mellitus. A family study. *Diabete & metabolisme*. **3**, 259-262.
- Sfeir A, de Lange T (2012). Removal of shelterin reveals the telomere end-protection problem. *Science*. **336**, 593-597.
- Singh M, Choi CP, Feigon J (2013). xRRM: a new class of RRM found in the telomerase La family protein p65. *RNA biology*. **10**, 353-359.
- Smith S, de Lange T (2000). Tankyrase promotes telomere elongation in human cells. *Current biology : CB*. **10**, 1299-1302.
- Smith S, Giriat I, Schmitt A, de Lange T (1998). Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science*. **282**, 1484-1487.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. **102**, 15545-15550.
- Tang YC, Williams BR, Siegel JJ, Amon A (2011). Identification of aneuploidy-selective antiproliferation compounds. *Cell*. **144**, 499-512.
- Tcherkezian J, Cargnello M, Romeo Y, Huttlin EL, Lavoie G, Gygi SP, Roux PP (2014). Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5'TOP mRNA translation. *Genes & development*. **28**, 357-371.
- Thompson PA, Drissi R, Muscal JA, Panditharatna E, Fouladi M, Ingle AM, Ahern CH, Reid JM, Lin T, Weigel BJ, Blaney SM (2013). A phase I trial of imetelstat in children with refractory or recurrent solid tumors: a Children's Oncology Group Phase I Consortium Study (ADV1112). *Clinical cancer research : an official journal of the American Association for Cancer Research*. **19**, 6578-6584.
- Tischkowitz M, Xia B (2010). PALB2/FANCN: recombining cancer and Fanconi anemia. *Cancer research*. **70**, 7353-7359.
- Toppiari J, Juul A (2010). Trends in puberty timing in humans and environmental modifiers. *Molecular and cellular endocrinology*. **324**, 39-44.
- Tsuji F, Oh-Hashi K, Kiuchi K (2012). Differential effects of Akt pathway inhibitors on IL-1beta-induced protein phosphorylation in human fibroblast-like synoviocytes. *Journal of receptor and signal transduction research*. **32**, 22-28.
- Tzarum N, Eisenberg-Domovich Y, Gills JJ, Dennis PA, Livnah O (2012). Lipid molecules induce p38alpha activation via a novel molecular switch. *Journal of molecular biology*. **424**, 339-353.
- Uchikawa E, Natchiar KS, Han X, Proux F, Roblin P, Zhang E, Durand A, Klaholz BP, Dock-Bregeon AC (2015). Structural insight into the mechanism of stabilization of the 7SK small nuclear RNA by LARP7. *Nucleic acids research*.
- Uhl GR, Drgon T, Johnson C, Li CY, Contoreggi C, Hess J, Naiman D, Liu QR (2008). Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Annals of the New York Academy of Sciences*. **1141**, 318-381.

- Ungar L, Yosef N, Sela Y, Sharan R, Ruppin E, Kupiec M (2009). A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic acids research*. **37**, 3840-3849.
- Unger C, Berdel W, Hanauske AR, Sindermann H, Engel J, Mross K (2010). First-time-in-man and pharmacokinetic study of weekly oral perifosine in patients with solid tumours. *European journal of cancer*. **46**, 920-925.
- Unryn BM, Cook LS, Riabowol KT (2005). Paternal age is positively linked to telomere length of children. *Aging cell*. **4**, 97-101.
- Valdes AM, Andrew T, Gardner JP, Kimura M, Oelsner E, Cherkas LF, Aviv A, Spector TD (2005). Obesity, cigarette smoking, and telomere length in women. *Lancet*. **366**, 662-664.
- Vasunilashorn S, Cohen AA (2014). Stress responsive biochemical anabolic/catabolic ratio and telomere length in older adults. *Biodemography and social biology*. **60**, 174-184.
- Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI, Heath AC, Martin NG, Westendorp RG, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, Amin N, van Duijn C, Penninx BW, Boomsma DI (2009). Genome-wide association study of smoking initiation and current smoking. *American journal of human genetics*. **84**, 367-379.
- Vink SR, Schellens JH, van Blitterswijk WJ, Verheij M (2005). Tumor and normal tissue pharmacokinetics of perifosine, an oral anti-cancer alkylphospholipid. *Investigational new drugs*. **23**, 279-286.
- Walcott F, Rajaraman P, Gadalla SM, Inskip PD, Purdue MP, Albanes D, Orr E, De Vivo I, Savage SA (2013). Telomere length and risk of glioma. *Cancer epidemiology*. **37**, 935-938.
- Wang FZ, Fei HR, Li XQ, Shi R, Wang de C (2012). Perifosine as potential anti-cancer agent inhibits proliferation, migration, and tube formation of human umbilical vein endothelial cells. *Molecular and cellular biochemistry*. **368**, 1-8.
- Wark L, Novak D, Sabbaghian N, Amrein L, Jangamreddy JR, Cheang M, Pouchet C, Aloyz R, Foulkes WD, Mai S, Tischkowitz M (2013). Heterozygous mutations in the PALB2 hereditary breast cancer predisposition gene impact on the three-dimensional nuclear organization of patient-derived cell lines. *Genes, chromosomes & cancer*. **52**, 480-494.
- Wentzensen IM, Mirabello L, Pfeiffer RM, Savage SA (2011). The association of telomere length and cancer: a meta-analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. **20**, 1238-1250.
- Wong LJ, Naviaux RK, Brunetti-Pierri N, Zhang Q, Schmitt ES, Truong C, Milone M, Cohen BH, Wical B, Ganesh J, Basinger AA, Burton BK, Swoboda K, Gilbert DL, Vanderver A, Saneto RP, Maranda B, Arnold G, Abdenur JE, Waters PJ, Copeland WC (2008). Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Human mutation*. **29**, E150-172.

- Xiong X, Zhao Y, Tang F, Wei D, Thomas D, Wang X, Liu Y, Zheng P, Sun Y (2014). Ribosomal protein S27-like is a physiological regulator of p53 that suppresses genomic instability and tumorigenesis. *eLife*. **3**, e02236.
- Yamada K, Ono M, Bensaddek D, Lamond AI, Rocha S (2013a). FMN2 is a novel regulator of the cyclin-dependent kinase inhibitor p21. *Cell cycle*. **12**, 2348-2354.
- Yamada K, Ono M, Perkins ND, Rocha S, Lamond AI (2013b). Identification and functional characterization of FMN2, a regulator of the cyclin-dependent kinase inhibitor p21. *Molecular cell*. **49**, 922-933.
- Yanagawa N, Tamura G, Oizumi H, Endoh M, Sadahiro M, Motoyama T (2011). Inverse correlation between EGFR mutation and FHIT, RASSF1A and RUNX3 methylation in lung adenocarcinoma: relation with smoking status. *Anticancer research*. **31**, 1211-1214.
- Yatagai Y, Sakamoto T, Yamada H, Masuko H, Kaneko Y, Iijima H, Naito T, Noguchi E, Hirota T, Tamari M, Konno S, Nishimura M, Hizawa N (2014). Genomewide association study identifies HAS2 as a novel susceptibility gene for adult asthma in a Japanese population. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. **44**, 1327-1334.
- Zhao J, Zhu Y, Lin J, Matsuguchi T, Blackburn E, Zhang Y, Cole SA, Best LG, Lee ET, Howard BV (2014). Short leukocyte telomere length predicts risk of diabetes in american indians: the strong heart family study. *Diabetes*. **63**, 354-362.
- Zhao Y, Abreu E, Kim J, Stadler G, Eskiocak U, Terns MP, Terns RM, Shay JW, Wright WE (2011). Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions. *Molecular cell*. **42**, 297-307.
- Zhou C, Steplowski TA, Dickens HK, Malloy KM, Gehrig PA, Boggess JF, Bae-Jump VL (2013a). Estrogen induction of telomerase activity through regulation of the mitogen-activated protein kinase (MAPK) dependent pathway in human endometrial cancer cells. *PloS one*. **8**, e55730.
- Zhou MS, Chadipiralla K, Mendez AJ, Jaimes EA, Silverstein RL, Webster K, Raj L (2013b). Nicotine potentiates proatherogenic effects of oxLDL by stimulating and upregulating macrophage CD36 signaling. *American journal of physiology. Heart and circulatory physiology*. **305**, H563-574.