creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학박사학위논문

고유의 템플릿을 이용하는 대안적인 텔로미어 유지 기전에 대한 연구

Studies on the Alternative Lengthening of Telomeres with a Unique Template in Mouse Embryonic Stem Cells

2019년 2월

서울대학교 대학원

생명과학부

성 상 현

ABSTRACT

Studies on the Alternative Lengthening of Telomeres with a Unique Template in Mouse Embryonic Stem Cells

Sanghyun Sung

Dept. of Biological Sciences

The Graduate School

Seoul National University

In order for eukaryotes to maintain genomic information, they must solve specific problems caused by their linear chromosomes: the end-replication problem due to the biochemical limitations of DNA replication machinery and the end-protection problem that must distinguish between DNA damage and the very end of the genome. If the two problems are not properly solved, the information of the terminal region is endangered to gradually disappear or to be damaged by incorrect processing. The special structure to solve these problems of the end is telomere. Telomere consists mainly of specific repeat sequences and their binding proteins. It is crucial in determining cell life and death to maintain telomere length and structure. Each time a cell divides, telomere is lost due to the end replication problem. When telomere length reaches a certain limit, it acts as a molecular timer to stop cell division or kill the cell. As a cell continues to divide beyond the limit, the genomic instability increases excessively and the cell is more likely to develop into a cancer cell. In other words, in order for cancer cells to cross the limit imposed by telomere, it is necessary to acquire a functional mechanism to preserve the length of telomere. Most cancer cells reactivate the reverse transcriptase, telomerase, to maintain telomere length and gain the ability to grow indefinitely. However, in some

cancer cells, the phenomenon of maintaining telomere length without telomerase was observed, which is called alternative lengthening of telomere (ALT). It has been known that ALT can be widely present in nature, as there are examples of telomerase-independent telomere maintenance mechanism in some plants and animals. Although ALT has been known to be a DNA replication mechanism that relies on homologous recombination, the exact molecular mechanism remains elusive. In this study, I used mouse embryonic stem cell (mESC) model to investigate the features and the molecular mechanisms of ALT. Whole genome sequence analysis revealed the unique sequence and structure of the telomere of ALT mESC. The most important feature was that the unique template sequence present in the subtelomere constituted telomere. It is a reliable hypothesis that telomeres of the ALT mESC maintain a unique protective mechanism, as non-telomeric sequences are highly involved, rather than the normal telomere repeats. To confirm this hypothesis, transcriptome and quantitative proteomic analysis were performed. As a result, I identified the pattern of gene expression changes following activation of ALT. The fundamental pathways enriched in ALT were transcriptional regulatory network and chromatin remodeling. I focused on the function of HMGN1

(high mobility group nucleosome binding domain 1), a non-canonical histone, which was the most significantly increased in ALT mESC among the candidate genes presumed to be involved in the mechanism of ALT. HMGN1 can compete with the linker histone H1 and contribute to post-translational modifications of other histones, thereby loosening the chromatin structure. I confirmed that HMGN1 binds to the unique telomere of ALT mESC. In addition, the inhibition of the function of HMGN1 increased the degree of telomere damages, reduced the amount of transcripts produced in telomere, and shortened the telomere length. Therefore, it was confirmed that the change produced by HMGN1 is a necessary factor to keep ALT telomere stable. In this study, the molecular characteristics of ALT mESCs were examined in terms of genome, transcriptome, and proteome. The main feature is the distinct structure of the telomere produced by a unique template and HMGN1 protein acting as an epigenetic regulator. The phenomenon of using a unique template sequence for telomere preservation is evolutionarily conserved from yeast to mammal. The ALT model established in mESCs will contribute to understanding not only of the molecular mechanism of ALT but of the mechanism of evolution of the genome.

Keywords: telomere, mouse embryonic stem cells, alternative lengthening of telomeres, genome instability, DNA damage, chromatin remodeling, telomere repeat containing RNA, HMGN1

Student Number: 2011-20335

TABLE OF CONTENTS

Abstract.....i

Table of Contents vi

List of Figures ix

List of Tables xiii

Introduction 1

What is telomere for? 2

Importance of telomere maintenance 3

Discovery of alternative mechanism of telomere maintenance 5

Possible molecular mechanisms of ALT..... 7

Molecular markers of ALT 10

Characterizing ALT-specific factors..... 11

Characteristics of subtelomere 12

Cases of the recruitment of subtelomeric sequences in ALT 13

Heterochromatin in telomere and subtelomere 14

TERRA is produced and has specific roles in telomere 16

Role of TERRA and R-loop in ALT 19

Purpose of this study 20

Materials and Methods 22

Results 39

Telomere of ALT mESC uses subtelomeric template 40

mTALTs are duplicated with a regular rule and a structure 43

ALT activation involves changes of RNA and protein expression 49

Local and global chromatin remodeling in post-ALT 51

Distinct protection mode of ALT telomere 56

HMGN1 is the novel ALT telomere binding protein 58

Discussion 64

741 ALT mESC as a model for ALT maintenance 65

mTALT has unique features to be recruited to telomeres 66

Potential mechanisms of mTALT duplication 69

The distinct telomere maintenance pathway of 741 ALT mESC 70

Unresolved terminal structure of mTALT-driven telomere 72

ALT initiation may not depend on the accumulation of genetic mutations 73

Potential roles of HMGN1 in ALT maintenance 74

How stemness is related to telomere maintenance 77

TERRA has a positive role in the protection of ALT telomere 79

Conclusion and perspectives 80

References 155

Abstract in Korean 163

LIST OF FIGURES

Figure 1. Two kinds of problems in chromosomal ends. 82

Figure 2. Schematic figure of T-loop structure and sheleterin complex. 83

Figure 3. Model of activation process of telomere maintenance mechanisms. 84

Figure 4. ALT mESCs are telomerase (Terc) knock-out..... 85

Figure 5. Copy number analysis of ALT mESC genome..... 86

Figure 6. mTALT is located at subtelomere of chromosome13..... 88

Figure 7. mTALT copy number changes through generations..... 89

Figure 8. ALT mESC telomere includes non-telomeric sequences..... 90

Figure 9. mTALT is co-localized with telomere. 91

Figure 10. mTALT region is duplicated consistently along the length 92

Figure 11. mTALT was duplicated to other chromosomes 93

Figure 12. Calculated mean telomere length. 94

Figure 13. Variant telomere repeat contents in pre- and post-ALT. 95

Figure 14. mTALT is duplicated tandemly with a consistent interval 96

Figure 15. mTALT region is duplicated as a unit. 97

Figure 16. mTALT has been cis-duplicated before ALT activation. 98

Figure 17. Specific copy of mTALTs was selected for duplication. 99

Figure 18. mTALTs are duplicated with a uniform orientation to telomere repeat. 101

Figure 19. mTALT duplication is not a transposon-like mechanism. 104

Figure 20. C-circle is not increased in post-ALT. 105

Figure 21. APBs are not increased in post-ALT 106

Figure 22. Schematic model of ALT activation in mESC. 107

Figure 23. Genome-wide SNP is similar between pre-ALT and post-ALT. 108

Figure 24. Transcriptome analysis of ALT mESC 109

Figure 25. Quantitative proteomics of ALT cells. 110

Figure 26. GO analysis of quantitative proteomics result. 111

Figure 27. List of differentially expressed proteins..... 112

Figure 28. ATAC-seq analysis with relation to transcription start sites..... 113

Figure 29. ATAC-seq peaks in mTALT region. 114

Figure 30. ATAC-seq analysis of canonical telomere repeats 115

Figure 31. Subtelomeres are enriched with ATAC-seq peaks..... 116

Figure 32. TERRA expression is increased in post-ALT cells..... 118

Figure 33. Schematic model of TERRA transcription 120

Figure 34. post-ALT cells have increased telomere damages 121

Figure 35. Post-ALT cells showed increased T-SCE 123

Figure 36. Shelterin is required for post-ALT cells 124

Figure 37. TRF2 is co-localized with telomere. 126

Figure 38. Stemness factor knockdown induces telomere dysfunction..... 127

Figure 39. HMGN1 expression is increased in ALT cells 128

Figure 40. HMGN1 is localized to ALT telomere. 129

Figure 41. HMGN1 binds to mTALT 130

Figure 42. HMGN1 knockdown induces telomere dysfunction. 131

Figure 43. HMGN1 knock-down decreases telomere length 133

Figure 44. HMGN1 knock-down decreases fragile telomeres. 134

Figure 45. HMGN1 knock-down decreases TERRA expression 135

Figure 46. TERRA downregulation induces telomere DNA damages 136

Figure 47. R-loop formation is increased after ALT activation 137

Figure 48. HMGN1 knockdown decreases the formation of R-loop in ALT telomere... 139

Figure 49. Working model 140

Appendix. Sequences of mTALT in mm10 reference genome. 141

LIST OF TABLES

Table 1. Differentially expressed proteins in quantitative proteomics. 146

Table 2. Differentially expressed genes in RNA-seq. 151

Introduction

What is telomere for?

The problem of preserving genome integrity starts from the end of chromosomes. During the evolutionary process of eukaryotes, the first appearance of a linear chromosome from circular one was innovation and crisis at the same time. The end itself should be distinguished from the double-strand breaks (DSBs) of internal genome regions and suppress the activation of DNA damage response (Blackburn, 1991; Greider, 1993). This is "end-protection problem". Another issue is "end replication problem" to find a way for maintaining the information of terminal chromosomes because the canonical DNA replication machinery cannot ensure the complete replication of the end parts of linear chromosomes (Figure 1) (de Lange, 2010).

The nucleoprotein complex to solve the two end problems is telomere. Telomere usually comprises simple tandem repeat sequences, (TTAGGG)n in the human case, and specific binding proteins for DNA repeats (Chong et al., 1995; de Lange, 2005). Telomere binding proteins residing in telomere are called shelterin complex which carries out variable functions for maintaining genome stability: 1) to support the canonical DNA replication machinery to progress in telomere regions, 2) to suppress the undesirable

activation of DNA repair mechanisms, mainly homologous recombination (HR) and non-homologous end joining (NHEJ), 3) to facilitate the formation of end protective structure, T-loop, 4)to recruit accessory proteins for particular telomere functions (Figure 2) (de Lange, 2004, 2018; van Emden et al., 2018).

Importance of telomere maintenance

Homeostasis of telomere length is critical for the life and death of cells. Due to the end replication problem, replicating cells are subjected to telomere erosion if not supported by telomere restoring mechanisms. Gradual telomere shortening eventually leads to the uncapping or the deprotected state of telomere with a dysregulated stoichiometry of telomere binding proteins (Cesare et al., 2013; Sfeir and de Lange, 2012). Thus, telomere length and the protection from the DDR machinery are intrinsically connected.

Various checkpoint factors monitor and control cell cycle to ensure the complete preservation of the genomic information (de Lange, 1998; Shay and Wright, 2005). DNA damage is a major source of checkpoint activation, especially in G2/M. If telomere reaches a critical threshold length, the resultant telomere deprotection induces "replicative

senescence", which is a permanent state of cell cycle arrest (Kaul et al., 2011).

p53 pathway plays a critical role in senescence. The DDR pathway via ATM and ATR activates p53, which can upregulate a critical cyclin-dependent kinase p21, which suppresses inhibition of Rb to arrest the cell cycle (Wang et al., 2016). Telomere functions as a counting timer for the remaining number of cell divisions. This is the basic mechanism of tumor suppression by the cell-intrinsic regulation.

Some cells can bypass the limit of replicative senescence mostly with the checkpoint inactivation including the disruption of the p53-p21 axis (Greenberg, 2005; Harley and Sherwood, 1997). In this case, telomere dysfunction can lead to chromosome fusion, which is too severely catastrophic to preserve genome integrity. Chromosome fusion forms dicentric chromosome, which can be broken at the next cell division. The broken sites are perceived as DNA damage and repaired mainly by non-homologous end joining. Then, fused chromosomes suffer successive breakages at next division, so these consecutive fusion-breakage cycles promote unrepairable genomic instability. The fusion-induced telomere dysfunction gives rise to mitotic arrest, which ultimately leads to apoptosis. This state suffering extensive cell deaths and an uncontrolled genomic instability

is called crisis. So two-step surveillance system (senescence and crisis) adjudged by telomere state manages the replication potential and the mutagenicity arising from telomere (Cheung and Deng, 2008; Sen et al., 1999).

The barrier of cell death imposed by crisis also cannot guarantee a complete stop of cell division. The exact mechanism that can make cell bypass crisis has not been fully clarified. At this point cells freed from the replication limit accumulate more disastrous genomic rearrangement and become malignant in the human case. Importantly, most of the cells must acquire telomere maintenance mechanism (TMM) to break out of the crisis (Figure 3) (Shay and Wright, 2005). Stem cells or germ cells that should have unlimited replication capacity equip telomerase, which is a specialized reverse transcription based telomere healing enzyme. These cells seldom suffer from the senescence and definitely the crisis resulting from telomere. Most of the human cancer cells gain telomerase activity at some point, which is normally suppressed in normal somatic cells.

Discovery of alternative mechanism of telomere maintenance

Interestingly, about 15% of cancer cells do not reactivate telomerase but utilize

recombination based telomere maintenance mechanism (Heaphy et al., 2011; Londono-Vallejo et al., 2004). This telomerase-independent mechanism is called "alternative lengthening of telomeres (ALTs)" by definition (Figure 3). Since the identification of telomerase mutant survivor of yeast, ALT-like mechanisms have been announced in diverse eukaryotic organisms including plants and humans (Bryan et al., 1995; Lundblad and Blackburn, 1993; Reddel et al., 1997). Telomere consists of simple repeats, thus there are widespread homologies between all telomeres consequently. The monumental experiment proving ALT to involve homologous recombination (HR) was that specific sequences integrated into telomere of ALT human cell can be copied to other chromosomes by copy-and-paste manner (Cho et al., 2014; Dunham et al., 2000; Muntoni et al., 2009).

In the study using yeast, two types of telomerase-independent survivors have been identified (Lundblad and Blackburn, 1993; Nakamura et al., 1998). Type I survivor depends on Rad51-based homologous recombination and mostly duplicates subtelomeric X or Y' elements for preserving telomere length. In contrast, Type II survivor uses telomere sequences per se as templates and shows only subtle amplifications of subtelomeric sequences. Both survivors depend on Rad52 protein which has a central function in DSB

repair and homologous recombination and replication protein Pol32 which indicates the recombination process for telomere maintenance involves replication. (Chen et al., 2001; McEachern and Haber, 2006).

Up to now the most part of human ALT case has been regarded as similar to the Type II mechanism because only telomere repeats and telomere-like variant repeats are duplicated (Conomos et al., 2012; Varley et al., 2002). The variant repeats are telomere-like repeat sequences but have one or two nucleotide changes primarily. The variant repeats interspersed within human ALT telomere were originated from the most proximal region (~2kb) of human telomeres. The presence of variant repeats throughout the ALT telomeres may change the chromatin structure to be more favorable to recombination through removal of shelterin and recruitment of nuclear receptors.

Possible molecular mechanisms of ALT

Although the involvement of HR in ALT is obvious, there are several different pathways and outcomes to which HR can lead ultimately (Li and Heyer, 2008; Londono-Vallejo et al., 2004; Sung and Klein, 2006). When it comes to DSB repair by HR, there are

at least three possible pathways (Jasin and Rothstein, 2013). The initial processes are identical: DSB ends are recognized and processed to generate 3' single-strands. Rad51 forms a nucleoprotein filament onto single strand DNA, which can invade a homologous target to make displacement-loop (D-loop) intermediates. In synthesis-dependent strand annealing (SDSA), the invading strand is displaced after DNA synthesis and re-annealed with the second end.

In break-induced replication (BIR), the D-loop associates with full machinery of DNA replication to copy the long track of the target donor chromosome. The newly synthesized strands by BIR segregate with the broken chromosome and not with the donor chromosome, which indicates BIR is a conservative DNA synthesis process (Donnianni and Symington, 2013). In double-strand break repair (DSBR), both ends of the DSB are captured to assemble double Holliday junction. This structure can either be resolved by cleavage of specific ways or dissolved to be displaced from the target chromosome (Sobinoff et al., 2017).

Recent studies showed that ALT employed BIR-like recombination mechanism (Costantino et al., 2014; Dilley et al., 2016). BIR plays a major role in repairing the broken

replication forks and other DSBs from various sources. The critical time-point BIR should work is that only one end of DSB exists and cannot be cured by NHEJ or other HR pathways. DSBs arising from telomere erosion or uncapping fell exactly into the category because telomere is the end of the chromosome by nature. In addition, telomere is a hard-to-replicate region, which can lead to frequent replication fork collapse and the formation of one-end DSB. Surprisingly, Rad51 was dispensable for break-induced telomere synthesis in human ALT (Dilley et al., 2016). The involvement of BIR mechanism in other ALTs is still an open question.

It's possible that a spectrum of mechanisms are utilized for ALT, mainly because not any single factor has been revealed to work in every ALT (Apte and Cooper, 2017; Lovejoy et al., 2012). Moreover, various ALTs show the different appearance in the telomere-amplification state, the distinct template choice, and the varied susceptibility to the chemical drug in human ALT cancers. To trace the reason for the ALT diversity to its origin is not simple. Telomere dysfunction induced genomic rearrangement may make too many changes in the mutation profile and the chromosome architecture. In addition, cell or organism-specific characteristics of telomere and the genetic network regulating telomere

biology obscure the integration of ALT under a united causality.

Molecular markers of ALT

The definition of ALT is simple and clear: telomere is synthesized via a mechanism other than telomerase. However, the exact molecular mechanism of ALT is still elusive. Many researchers observed molecular markers of ALT, which seem to be highly correlated with ALT activity (Draskovic et al., 2009; Henson et al., 2009; Sobinoff and Pickett, 2017). 1) ALT-associated PML bodies (APBs): APBs are PML bodies with telomere signals and HR-associated proteins. APBs have been regarded to manage telomere synthesis process; 2) C-circles: C-circles are extrachromosomal telomeric repeats including partially single-stranded CCCTAA repeats. It can be detected with modified rolling circle amplification assays; 3) Telomere-sister chromatid exchange (T-SCE): ALT cells show a higher telomeric exchange between sister chromatids, implying a higher rate of recombination; 4) Telomere dysfunction-induced foci (TIFs): TIFs mean DNA damage sites localized in telomere. TIFs can be detected with co-localization of telomere DNA and 53BP1 or γ-H2AX DNA damage markers. ALT cells normally have an increased TIFs.

All those ALT markers provide partial evidence of ALT activity. However, every marker has exceptions. Telomerase-positive cells can also have ALT markers, and there are ALT cells that do not possess those markers. Although many studies still rely on the identification of these markers in ALT, it should be taken carefully to interpret the results.

Characterizing ALT-specific factors

The identification of ALT telomere-specific proteins was critical to revealing the molecular mechanism of ALT. At the first time the method called "proteomics of isolated chromatin loci" (PICh) that uses DNA sequence-specific probe and captures fixed DNA and protein complexes, discovered novel proteins associated with ALT telomeres (Dejardin and Kingston, 2009). Specific orphan nuclear receptors were validated to bind ALT telomeres and hypothesized to interact with variant telomere sequences which were originally dispersed within subtelomeres. Following studies showed that variant telomere repeats were duplicated to terminal telomeres upon ALT activation and had potential to recruit orphan receptors and other ALT-specific proteins to constitute protein hub for the ALT mechanism (Conomos et al., 2014; Conomos et al., 2012).

The principal protein factors in ALT include not only telomere binding protein but other proteins interacting with telomere temporarily or indirectly through regulating telomere-associated protein networks. The exact and complete protein requirement for ALT has not been unraveled. Moreover, the role of canonical shelterin proteins has not been tested in every respect. The differential DNA constituent of telomere in ALT may bring new protein players and push out original telomere occupants. Structure-specific regulators including helicases, nucleases, and recombinases can also interact with telomere in context-dependent manners (Clynes et al., 2015; Cox et al., 2016; Pan et al., 2017).

Characteristics of subtelomere

Subtelomere is generally defined as a genomic region between the central part of genome and telomere (Riethman et al., 2003; Riethman et al., 2005). In general, subtelomere includes repetitive sequences including telomere repeats and low levels of genes. However, there is no specific boundary between subtelomere and other regions, no clear sequence motif shared in subtelomeres, and no general feature conserved among various organisms. The limited amount of sequence homology among subtelomeres exists

in same species. Although different subtelomeres share homology in some cases, diverse polymorphisms dominate the overall identity of subtelomere. The main polymorphisms are represented in length and gene copy number. Other properties like chromatin state, recombination frequency, and gene density vary among individuals within the same species.

The intrinsic properties of subtelomere obstructed the complete construction of the genome (Tashiro et al., 2017). The repetitive nature of subtelomere left many gaps in the genome assemblies. As the sequencing and other molecular cloning techniques progressed, the subtelomere has been gaining the right representation. However, despite the achievement of a standard genome, there still remains a high degree of diversity among individuals.

Cases of the recruitment of subtelomeric sequences in ALT

Subtelomeric sequences can be used as templates to replace telomere repeats in some of the telomerase-negative condition. The symbolic example of subtelomere-telomere exchange is Type I survivor of yeast. Yeast subtelomeres are mainly composed of two kinds of sequence motif, X and Y' elements, but not all subtelomere have Y' element.

X element is more proximal to a central genomic region, so Y' element is the generally neighboring part of telomere. Although about one-third of chromosomes do not contain Y' elements, all chromosomal ends in the end obtain Y' elements with the formation of Type I survivor (Chen et al., 2001; Nakamura et al., 1998).

Another case is the ALT survivor of telomerase-deficient *C. elegans*. *C. elegans* ALT survivors utilized an internal genomic region for telomere maintenance. The so-called 'Template for ALT (TALT)' region *cis*-duplicated from internal locus to subtelomere during evolutionary process, and amplified to all chromosomal ends after ALT activation (Seo et al., 2015). These cases suggest that particular subtelomeric sequences can be recruited as telomere constituents. However, the exact rule of determining the potential of certain sequences as template for ALT has not been revealed.

Heterochromatin in telomere and subtelomere

Telomere has a basically silenced state of chromatin, which is important for the telomere function and length regulation in many organisms. In *S. cerevisiae*, telomere binding protein Rap1 recruits chromatin-silencing Sir proteins (Duan et al., 2016; Kupiec,

2014). *D. melanogaster* is a special case, because the retrotransposons cap the end of chromosomes (Berendes and Meyer, 1968; Casacuberta, 2017; Silva-Sousa et al., 2012). The 'terminin' complexes bind and protect *D. melanogaster* telomere. The stable maintenance of telomere depends on HP1 (heterochromatin protein 1), a critical component of heterochromatin. Mammalian telomeres are enriched with heterochromatin marker H3K9me3, which provides a docking site for HP1.

The silenced state of telomeres has been nearly regarded as the suppressing mechanism of telomere lengthening (Bisht et al., 2008; Voon et al., 2016). The compact state of chromatin inhibits telomerase recruitment and suppress the recombination in telomeres. The histone methyltransferases (HMTs)-deficient cells showed a remarkable lengthening of telomeres. However, there is also the case that shows telomere shortening in silencing disrupted condition in yeast. So, it will be fair to explain that chromatin silencing condition supports telomere length homeostasis.

Silencing histone markers are also enriched in mammalian subtelomere (Cubiles et al., 2018; Robin et al., 2014). High DNA methylations also exist in subtelomere. Many repeat regions including peri-centromere are highly DNA methylated, which is

considered to suppress the HR. The pioneer study showed the expression of a generally euchromatic gene inserted into subtelomere is silenced, which is called telomere position effect (TPE). The formation and maintenance of subtelomeric heterochromatin seem to be influenced by telomere, but the precise mechanism is still elusive.

Cells experiencing telomere shortening show dynamic changes of chromatin state (Cubiles et al., 2018; Schoeftner and Blasco, 2009a). Telomere shortening accompanies the loss of heterochromatin histone markers, decompaction of chromatin, lowered capacity to suppress DDR, and consequently, higher rate of telomerase recruitment and recombination. In telomerase-deficient cells, this condition can be ALT-favorable. The lowered telomere protection may permit various DNA repair pathways, which can lead to telomere maintenance via ALT mechanism (Benetti et al., 2007; Episkopou et al., 2014).

TERRA is produced and has specific roles in telomere

In spite of heterochromatic feature of telomeres, telomere sequences are transcribed to produce mostly long non-coding RNAs, TERRA (TElomere Repeat-containing RNA) (Cusanelli and Chartrand, 2015; Oliva-Rico and Herrera, 2017). The

promoter region and the transcription regulatory factors are still puzzling. Subtelomere has been regarded as main sources for TERRA promoters, but there are also reports that show TERRA transcription starts within telomeres. Telomere length seems to be inversely correlated with TERRA expression (Schoeftner and Blasco, 2009a). However, many facets of TERRA biogenesis still remains as questions.

One of the main roles of TERRA may be to inhibit the function of telomerase. TERRA may compete with telomerase RNA component for binding telomere DNA sequences, or directly inhibit telomerase RNA component (TERC) gene by binding to it. However, in vivo evidence supporting the prediction is still not sufficient. Otherwise, TERRA can mediate telomere lengthening by recruiting telomerase to telomeres. So, the function of TERRA depends on a specific context.

TERRA can interact with telomeres through shelterin or other telomere-functioning enzymes like histone methyltransferases. TERRA itself seems to support the localization of these enzymes to telomeres. TERRA depletion was associated with a decrease of H3K9me3 and other silencing histone markers. So, TERRA has been thought to regulate heterochromatin formation in telomere regions. The production of TERRA is

regulated by the chromatin state of telomere, which means telomere transcription and heterochromatinization are on the intimate relationship with a mutual feedback.

TERRA can also interact with telomere directly. Due to having complementary sequences for telomere, TERRA can pair with telomere repeats by displacing one of the double-stranded DNA and base-pairing with the other. This three stranded RNA-DNA hybrid structure is called R-loop (Toubiana and Selig, 2018). R-loop can have various functions and be closely connected to telomere stability. Basically, R-loop can hinder transcription or replication machinery by blocking the progress of the protein complexes. Telomere is a hard-to-replicate region with the repetitive nature, which R-loop can make worse (Toubiana and Selig, 2018).

R-loop formation can influence genome integrity by inducing replication stress, mutation, recombination, and other rearrangements (Graf et al., 2017; Schoeftner and Blasco, 2009b). The persistent replication fork stalling can induce single and double-strand breaks. DSBs occur in case of the replication fork collapse, which can lead to a drastic telomere shortening. Since R-loop is a critical source of endogenous DNA damage, the formation and removal of R-loop must be tightly controlled.

RNase H enzymes degrade RNA moiety of RNA-DNA hybrids, and helicases like Pif1 can unwind the pairing of RNA and DNA (Geronimo and Zakian, 2016). THO/TREX complex involved in mRNA transcription and export is reported to suppress R-loop formation (Wang et al., 2015). The role of BRCA2 was also implicated in R-loop suppression. BRCA2 binds to the branched structure of R-loop, which could lead to more access by RNase H enzyme or DNA-RNA helicases. These results revealed R-loop may be a critical source of replication stress implicated in cancer (Bhatia et al., 2014).

Role of TERRA and R-loop in ALT

ALT cells have generally upregulated TERRA. TERRA augmentation could be just a resultant phenomenon following chromatin change. However, TERRA can be a necessary player to make ALT telomere more susceptible to HR. The exact mechanism of this process are still elusive, but some hypotheses with supporting observation exist (Arora and Azzalin, 2015; Arora et al., 2014). First, DSBs occurred from persistent R-loop can be initial triggers and substrates for HR. Second, the displaced single strand telomere has key roles. 1) ss-telomere can be coated with RPA, which can recruit DDR factors to telomeres.

2) RPA on ss-telomere would be exchanged with RAD51 to promote HR. 3) ss-telomere can be good substrates for the strand invasion from other telomeres. Third, TERRA itself can recruit HR related factors to telomere directly or indirectly.

Purpose of this study

I got ALT mouse embryonic stem cells (mESCs) which were homozygous Terc knock-out (-/-) and survived from the crisis (Niida et al., 2000). Telomerase-deficient mESC suffered replicative senescence and crisis resulting from telomere shortening, and recovered to show an elevated growth rate and proliferation, which led to the identification of ALT cell. Interestingly, the previously reported phenotype of ALT mESCs was similar to that of *C. elegans* ALT survivors. *C. elegans* ALT survivors utilized a non-telomeric sequence for telomere maintenance and ALT mESCs seemed to show an analogous pattern. The duplication of a non-telomeric sequence in telomerase-deficient survivors may be conserved phenomena from yeast to mammal.

In this study, I took advantage of the novel ALT mESCs to establish the unprecedented characteristics of ALT. I tried 1) to clearly delineate the telomere state of

ALT mESC, 2) to identify the cellular changes associated with ALT activation, and 3) to examine the functions of critical molecular factors in ALT maintenance.

I analyzed ALT mESCs in various points including genome, transcriptome, proteome, and epigenome. With a whole genome sequencing (WGS), I identified a distinct telomeric structure of ALT mESCs. A unique template in subtelomere of a chromosome (I named this template as mTALT meaning mouse template for ALT) was duplicated to all other chromosomes. The distinct DNA composition of ALT mESCs implies a different mode of protection and regulation of telomere.

Next, I profiled the gene expression pattern of ALT in perspectives of RNA and protein. ALT activation leads to changes in transcription and chromatin regulation. Among the candidate genes that seemed to be related to the ALT mechanism, I focused on the function of HMGN1. A non-canonical histone protein HMGN1 was highly enriched in ALT mESC, which served as an epigenetic modulator. Telomere-specific non-coding RNAs were also highly expressed in ALT mESCs, which have a protective role. ALT mESCs seem to maintain telomere with distinct sequences and structures based on the presence of two main pillars, mTALT and HMGN1.

Materials and Methods

Cells

Terc knock-out mESCs were kind gifts from H. Niida. The generation and culture method was decribed previously (Niida et al., 1998). Briefly, the Terc targeting construct (TR-10) was transfected into E14 mESC by electroporation to exchange the first Terc allele. The targeting plasmid included PGK-neo gene flanked by loxP sites. The transfected cells were selected with G418 (200 mg/ml) and Ganciclovir (1 mM). The resistant Terc+/- cells were picked after eight days. The neo gene was removed by infection with the adenovirus of Cre recombinase. The Terc+/- cells were re-targeted with TR-10. Two lines of Terc-/- ESC, DKO301 and DKO741 were recovered. In the next paper, the authors found ALT-activated mESC after ~450 passagings. Both of DKO301 and DKO741 produced survivors with rare frequencies (Niida et al., 2000). I got those ALT survivor cells at each PDL timing (PD100, PD350, PD400, PD450) with control (Terc+/+) cells. For mESC maintenance and stock preparation, mouse embryonic fibroblast (MEF) cells were used as feeder cells. Feeder-free mESCs were used for all experiments including nucleic acids extraction, protein extraction, immunostaining, and fluorescent in situ hybridization.

Whole genome sequencing (WGS)

Whole genome DNA was extracted using DNeasy Blood & Tissue Kit (QIAGEN). Macrogen (South Korea) performed library preparation and sequencing process using Hiseq X (Illumina) and NovaSeq 6000 (Illumina) platform. All samples were sequenced with 150bps paired-end sequencing.

WGS read mapping and copy number analysis

Cleaned sequence reads were mapped to the mm10 reference genome using Burrows-Wheeler Aligner (BWA) mem (version 0.7.17). Marking of duplicate reads, indexing and sorting was performed with Picard-Tools (version 1.96) and SamTools (version 1.6). Copy numbers were calculated as 1 kbps binned read counts that were normalized with the total mapped reads.

Variant detection method

Single nucleotide polymorphisms (SNPs) and insertion and deletions (indels) were detected and genotyped according to the mm10 reference genome and the assembled E14

genome using Haplotyper Caller (GATK version 3.8-0; filtering option=QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0).

Total RNA extraction and RNA sequencing analysis

mES cells were harvested in Trizol. RNA was extracted using chloroform and isopropanol precipitation. Macrogen (South Korea) performed library preparation and sequencing process using HiSeq 4000 (Illumina). Biological duplicate samples were sequenced by 101-bp paired-end. The reads for this experiment were aligned to the mm10 reference genome using the STAR read aligner (version 2.5.3a). FPKM (Fragments Per Kilobase Million) and TPM (Transcripts Per Kilobase Million) were calculated with RSEM (version 1.3.0). Differentially expressed genes (DEG) were calculated by R/Bioconductor DESeq2 (version 1.18.1). Differentially enriched gene sets are computed by comparing GSEA (version 2.0) enrichment scores.

Monochrome multiplex quantitative PCR (mmqPCR)

mmqPCR was modified from a previously published paper (Cawthon, 2009). mmqPCR

was used for telomere and mTALT length measurement. This technique can measure Ct (cycle threshold) value for two amplicons in a reaction well with a single DNA-intercalating dye. The Ct for more abundant target can be measured at earlier cycles, and at later cycles the Ct for the second target can be collected at a temperature highly above the melting temperature of the first amplicon. Primers for second amplicon were placed with GC-clamps on both ends to increase the melting temperature. Reaction mixture was 1x HOT FIREpol EvaGreen qPCR Supermix (Solis biodyne), target primers (final concentrations 500 nM each), single copy gene control primers (final concentrations 500 nM each), template DNA (100 ng). The thermal cycling program for telomere was Stage 1: 15 min at 95°C; Stage2: 2 Cycles of 15 s at 94°C, 1 5s at 49°C; and Stage3: 33 cycles of 15 s at 94°C, 10 s at 62°C, 15 s at 74°C with signal capture, 10 s at 84°C, 15 s at 88°C with signal capture. The thermal cycling program for mTALT was Stage 1: 15 min at 95°C; Stage2: 35 Cycles of 15 s at 94°C, 20 s at 62°C, 15 s at 72°C with signal capture, 10 s at 84°C, 15 s at 88°C with signal capture.

Interphase TIF assay

The protocol for Immunofluorescence and telomere fluorescence in situ hybridization (FISH) was modified from a previous paper (Cesare et al., 2015). The cells were cultured in feeder-free state and trypsinized at confluency of 50~60%. The cells were diluted with PBS at concentration of 2 to 5×10^5 cells/ml. The diluted cells were cytocentrifuged with the appropriately assembled cytospin funnels at 800 rpm for 10 min. The cells on slide were fixed with 3.7% formaldehyde, and permeabilized with KCM buffer (120 mM KCL, 20 mM NaCl, 10 mM Tris-Cl, pH 7.5, 0.1% Triton X-100). ABDIL buffer containing RNaseA (100 ug/ml) were added and incubated for blocking. The primary antibody diluted in ABDIL (20 mM Tris-Cl, pH 7.5, 2% bovine serum albumine, 0.2 % fish gelatin, 150 mM NaCl, 0.1% Triton X-100, 0.1% sodium azide) was added and incubated overnight at 4°C. The slides were washed with PBST and incubated with ABDIL-diluted secondary antibody for 1hr at room temperature. The slides were washed with PBST again. The cells were fixed with 3.7% formaldehyde. The slides were ethanol dehydrated with a graded ethanol series: 70%, 90%, and 100% ethanol. After the slides were dried, telomere-PNA probe diluted in hybridization solution (70% formamide, 0.5% blocking reagent, 10 mM Tris-Cl, pH 7.5) at 0.3 ug/ml was added and the cells were denatured for 5min at 80C. The hybridization was

done overnight at room temperature. The slides were washed with PNA wash A (70% formamide, pH 7.5, 10 mM Tris-Cl, pH 7.5) and B (50 mM Tris-Cl, pH 7.5, 150 mM NaCl, 0.8% Tween 20) buffer. After ethanol dehydration and air-dry, the slides were mounted with Prolong Gold and store overnight before imaging.

Metaphase-immunostaining

Basically, metaphase-immunostaining was same as interphase assay except chromosome preparation step. The cultured cells were treated with 1 uM nocodazole for 1.5 hr before harvest at 50~60% confluency. The cells were trypsinized and washed with PBS. The cells were resuspended with 0.8% Sodium Citrate solution and incubated at 37°C for 5 min. The swelled cells were cytocentrifuged at 2000 rpm for 10 min. The remaining steps were same as interphase-immunostaining.

Quantitative-FISH (Q-FISH)

The cultured cells were treated with 1 uM nocodazole for 1.5 hr before harvest at 50~60% confluency. The trypsinized cells were resuspended with 0.8% sodium citrate solution and

incubated at 37°C for 10 min. After spin-down, the supernatant was aspirated and the cells were resuspended in ~300 ul of remaining solution. 1 ml of pre-cooled fixative (75% methanol and 25% acetic acid) was added dropwise and the tubes were tapped after each drop. 9 ml of fixative was added with consistent tapping. The cells were fixed at 4°C overnight and stored at -20°C. Slides were prepared by soaking in 100% ethanol for 1hr. The cleaned slides were air-dried and the fresh fixative was prepared and cooled. The fixed cells were spin-down and resuspended in fresh fixative at ~1×10^6 cells/ml. Five paper towels were soaked with water and the cleaned slides were placed onto the wet towels. The resuspended cells were dropped from a height of 2 cm onto the slides. The slides were incubated at 65°C for 1min, and allowed to air-dry. The slides were checked with bright-filed light microscope to look for good metaphase chromosomes. The slides were cured overnight at room temperature. The slides were fixed with 3.7% formaldehyde and incubated with PBS containing RNase A (250 ug/ml). The slides were re-fixed with 3.7% formaldehyde and washed with PBS. The slides were ethanol dehydrated with a graded ethanol series: 70%, 90%, and 100% ethanol. After the slides were dried, telomere-PNA probe diluted in hybridization solution at 0.3 ug/ml was added and the cells were denatured

for 10 min at 70°C. The remaining washing and mounting steps were same as the immunostaining protocols.

Chromosome orientation-FISH (CO-FISH)

Cells were seeded and incubated at 40~50% confluency. The cells were incubated for 12 hr in growth media with 10 uM 3:1 BrdU/BrdC. For the last 1.5hr with BrdU/BrdC, 1 uM nocodazole was added to media. After spin-down, the supernatant was aspirated and the cells were resuspended in ~300ul of remaining solution. The process of cell harvest, fixation, and dropping onto slides was same as Q-FISH. The cured slides were rehydrated with PBS for 5min and fixed with 3.7% formaldehyde for 5min at room temperature. After PBS wash, the slides were incubated with PBS containing RNase A (250 ug/ml) for 15 min at 37°C. The slides were washed with PBS and rinsed with 2X SSC, which is followed by incubation in 2X SSC containing 0.5 ug/ml Hoechst 33258 for 15 min in the dark. The slides were irradiated with ~365 nm UV light for 30 min and washed with distilled, deionized water. 80 ul of 10 U/ul of ExonucleaseIII were treated to the slides and incubated for 30 min at 37°C. During this step, the newly synthesized DNA strands were degraded.

After wash with PBS, 70% deionized formamide/30% 2X SSC solution was added and the slides were incubated on a heating block with 70°C for 10 min. After rinse with water, the slides were ethanol-dehydrated. TelG-Cy3 PNA probe was placed onto the slides and denatured at 70°C for 10 min, and then hybridization was done for 2 hr at room temperature in the dark. The slides were washe twice in PNA wash A buffer, and TelC-FITC PNA probe was loaded. The slides were hybridized for 2 hr at room temperature in the dark, and washed twice in PNA wash A buffer, and three times in PNA wash B buffer. The slides were rinsed with water and air-dried completely. The samples were mounted with Prolong Gold.

Lentivirus-based gene knock-down

HEK293FT cells were plated at 100 mm plates. HEK293FT cells were co-transfected with target shRNA-expressing transfer vector (PLKO.1 backbone), pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260) by PEI method. In 100mm plate, the actual ratio of each vector and PEI is that transfer vector: pMD2.G : psPAX2 : PEI =6:3:1.5:21 (ug). After 7-12hrs, media was changed with fresh one. After 36~48hrs, culture media was harvested

and filtered with 0.45 um PVDF syringe filter. Filtered media was incubated with Lenti-X concentrator (Takara) at 5:1 ratio at 4°C on rotator for 1~2 days. The virus soup with concentrator was centrifuged at 1500 xg, 45 min. The resulted virus pellet was resuspended with ~200 ul fresh culture media. The overall virus harvest process was repeated after 24 hrs from first harvest. ALT mESC was plated at 12 well plate. At ~50% confluency the resuspended virus pellet was added to target mESC with poly-b-rene at concentration of 4 ug/ml. After 24 hrs second harvested virus was infected in the same way. After 24 hrs of incubation the infected cells were transferred to new plate. When the transferred cells reached 80~90% confluency, puromycin was added to each well at 1 ug/ml. The selecting media was changed every day. The decreased expression of target genes were confirmed by qPCR or western blot.

Proteome analysis

~10^7 Cells were harvested at 70~80% confluency from 100 mm plates. The pelleted cells were incubated with hypotonic buffer A (10 mM HEPES, pH 7.5, 20 mM NaCl, 0.01% Triton X-100, 1 mM DTT, 1 mM PMSF, proteinase inhibitor cocktail) and pipetted up and

down ~100 times. The nuclei were pelleted and washed once with PBS. The washed nuclei were resuspended in 8 M Urea, and sonicated at maximum intensity for 30 min (Cosmobio bioruptor).

DNA-RNA hybrid immunoprecipitation and quantitative PCR (DRIP-qPCR)

Nucleic acid was extracted with phenol/choloroform in phase lock gel tubes followed by ethanol precipitation. I sonicated nucleic acid at maximum intensity for 15 min (Cosmobio bioruptor). For RNase H-treated control, 10 ug of DRNA was treated with RNase H overnight at 37°C, and then purified with phenol/chloroform and ethanol precipitation. 10 ug of DRNA was incubated with 5 ug of S9.6 R-loop specific antibody in 1x IP buffer (10 mM $NaPO_4$, pH7, 140 mM NaCl, 0.05% Triton X-100) overnight at 4°C. 50 ul of Dynabead (ThermoFisher) for each sample was equilibrated with IP buffer and added to IP tubes and incubated at 4°C for 2 hr. I washed the beads three times with IP buffer, and added elution buffer (50 mM Tric-Cl, pH 8, 10 mM EDTA, 0.5% SDS, proteinase K). The elution was performed for 3 hr at 65°C. DNA was purified lastly with phenol/chloroform and ethanol precipitation. qPCR was done with the sonicated input DNAs.

ATAC-seq and peak calling

A basic process is from a previous paper (Corces et al., 2017). 50,000~75,000 cells were pelleted at 500 xg, 4°C for 5 min. I added 50 ul of cold RSB (10 mM Tris-Cl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin and pipetted 3 times. The cells were incubated on ice for 3min, and washed with 1ml of cold RSB containing 0.1% Tween-20 but no other detergents. The cells were pelleted and resuspended wiht 50 ul of transposition mixture (25 ul 2X TD buffer, 2.5 ul transposase, 16.5 ul PBS, 0.5 ul 1% digitonin, 0.5 ul 10% Tween-20, 5 ul H_20) (2X TD buffer=20 mM Tris-Cl, pH 7.6, 10 mM $MgCl_2$, 20% dimethyl formamide). I incubated the reaction at 37°C for 30 min with consistent shaking, and stop the reaction by adding 50 ul of tagmentation stop buffer (10 mM Tris-Cl, pH 8.0, 20 mM EDTA, pH 8.0). The reaction was cleaned with Zymo DNA clean and concentrator-5 kits. DNAs were eluted in 21 ul of elution buffer, and amplified for 5 cycles using NEBnext high-fidelity 2x PCR master mix. Using 5 ul (10%) of the pre-amplified mixture, I run a 15 ul qPCR to determine the number of additional cycles. I plotted linear relative fluorescence versus cycle, and determine the cycle number

that fitted to 1/4 of maximum fluorescence intensity. The remaining 45 ul of PCR reaction was processed with qPCR with additional cycles determined. I purified the final PCR reaction using Zymo kits. Sequencing was performed with Novaseq 6000 platform by 100-bp paired-end. Raw genomic sequence reads were mapped to the mm10 reference genome using BWA mem (version 0.7.17). Alignments were further processed using GATK (version 3.8-0) and PicardTools for duplicate marking (version 1.103). Peak-calling was performed with MACS2 (version 2.1.1.20160309).

Primers used for the experiments described.

Experiment	Name	Primer F	Primer R
mTALT qPCR	1	AGGCATGGCTTAGATGTTGC	GACTGTGCCCTACATGCTCA
	2	GTTCTCGGGTCATGTCTGGT	CCTGCTACTTGCTTGGGAAG
	3	GCCTCCTTGGTGTGTTGTTT	CACACTGTGGGTGGAATTTG
	4	ACCAGCTGAAACATCCATCC	TGGGGAAAGTGCTAATGGAC
	5	TAGGGGATGTTAGGCCTCCT	AGGTGTCCCCGAGATCTTTT
	6	GGGCAACTTCCTGTGTGATT	GCATGGAACGTGGCTTAGTT
mTALT	36B4F_mmQP CR	CGGCGGCGGGCGGCGCGGGCTGGGCGG ACTGGTCTAGGACCCGAGAAG	CGGCGGCGGGCGGCGCGGGCTGGGCG GACTGGTCTAGGACCCGAGAAG
mmqPCR	TALT primer	ACCAGCTGAAACATCCATCC	TGGGGAAAGTGCTAATGGAC
	TALT primer2	AGCGTGTGAAGGAGGAAGAG	CTCCACAACCGCAGTAAGGT

telomere mmqpcr	telG/C	ACACTAAGGTTTGGGTTTGGGTTTGGGT TTGGGTTAGTGT	TGTTAGGTATCCCTATCCCTATCCCTAT CCCTATCCCTAACA
mTALT	*cis*-duplication 1	GGGCAACTTCCTGTGTGATT	CCTGCTACTTGCTTGGGAAG
***cis*-duplication**	*cis*-duplication 2	CCTGTTCTGGGGTCATGTCT	AGGTGTCCCCGAGATCTTTT
mTALT quantification	mTALT3	AGCGTGTGAAGGAGGAAGAG	CTCCACAACCGCAGTAAGGT
	mTALT4	GTTCTCGGGTCATGTCTGGT	CCTGCTACTTGCTTGGGAAG
mTALT	1R	TTCCCACTACAAGGCTGACC	AGGTGTCCCCGAGATCTTTT
***trans*-duplication**	5R	GCTGCTACCCTCTTGTGCTT	AGGTGTCCCCGAGATCTTTT
mTALT probe	mTALT probe 1	AGGCTTGCAACTTGGAGCTA	TGGTGTGCTCCTAAATGCTG
	mTALT probe 2	GGACAGCTTCGCCATACATT	CACACTGTGGGTGGAATTTG
shRNA cloning	LKO	TGGACTATCATATGCTTACCGTAAC	GTATGTCTGTTGCTATTATGTCTA

Targets and oligo lists for lentiviral cloning

Target		**Sequence**	**Forward Oligo**	**Reverse Oligo**
hmgn1(1)	**3UTR**	**GTACAATCCAGAGGAATATTT**	**CCGGGTACAATCCAGAGGAATATTTCTCGAGAAATATTCCTCTGGATTGTACTTTTTG**	**AATTCAAAAAGTACAATCCAGAGGAATATTTCTCGAGAAATATTCCTCTGGATTGTAC**
hmgn1(2)	**CDS**	**GAAAGAAGCTAAGTCCGACTA**	**CCGGGAAAGAAGCTAAGTCCGACTACTCGAGTAGTCGGACTTAGCTTCTTTCTTTTTG**	**AATTCAAAAAGAAAGAAGCTAAGTCCGACTACTCGAGTAGTCGGACTTAGCTTCTTTC**
Tcstv3(1)	**CDS**	**ACCAGGATCCTGCATCTTATA**	**CCGGACCAGGATCCTGCATCTTATACTCGAGTATAAGATGCAGGATCCTGGTTTTTTG**	**AATTCAAAAAACCAGGATCCTGCATCTTATACTCGAGTATAAGATGCAGGATCCTGGT**
Tcstv3(2)	**CDS**	**CTGAAATCCTACCGCAGATTG**	**CCGGCTGAAATCCTACCGCAGATTGCTCGAGCAATCTGCGGTAGGATTTCAGTTTTT**	**AATTCAAAAACTGAAATCCTACCGCAGATTGCTCGAGCAATCTGCGGTAGGATTTCA**

			G	G
TERF2	CDS	CGAACAGCTGTG ATGATTAAA	CCGGCGAACAGCTGTGATGATTAAACTC GAGTTTAATCATCACAGCTGTTCGTTTTT G	AATTCAAAAACGAACAGCTGTGATGATTA AACTCGAGTTTAATCATCACAGCTGTTCG
Pot1a	CDS	GTCTACGATAACC ACGTTCAA	CCGGGTCTACGATAACCACGTTCAACTCG AGTTGAACGTGGTTATCGTAGACTTTTTG	AATTCAAAAAGTCTACGATAACCACGTTC AACTCGAGTTGAACGTGGTTATCGTAGA C
Pot1b	CDS 3UTR	ATACCCTTCCCAG CGTTTATA	CCGGATACCCTTCCCAGCGTTTATACTCG AGTATAAACGCTGGGAAGGGTATTTTTTG	AATTCAAAAAATACCCTTCCCAGCGTTTA TACTCGAGTATAAACGCTGGGAAGGGTAT
Klf4	CDS	AGTTGGACCCAG TATACATTC	CCGGAGTTGGACCCAGTATACATTCCTCG AGGAATGTATACTGGGTCCAACTTTTTTG	AATTCAAAAAAGTTGGACCCAGTATACAT TCCTCGAGGAATGTATACTGGGTCCAACT
Atrx	3UTR, CDS	CCCACGGATGAG AATGTAAAT	CCGGCCCACGGATGAGAATGTAAATCTC GAGATTTACATTCTCATCCGTGGGTTTTT G	AATTCAAAAACCCACGGATGAGAATGTAA ATCTCGAGATTTACATTCTCATCCGTGGG
Nanog	3UTR	GCCAACCTGTACT ATGTTTAA	CCGGGCCAACCTGTACTATGTTTAACTCG AGTTAAACATAGTACAGGTTGGCTTTTTG	AATTCAAAAAGCCAACCTGTACTATGTTT AACTCGAGTTAAACATAGTACAGGTTGGC
Tbx3	CDS	GCTGACGACTGT CGATATAAA	CCGGGCTGACGACTGTCGATATAAACTC GAGTTTATATCGACAGTCGTCAGCTTTTT G	AATTCAAAAAGCTGACGACTGTCGATATA AACTCGAGTTTATATCGACAGTCGTCAGC
Pold3	CDS	ACCGACTATGACA TCCTTAAA	CCGGACCGACTATGACATCCTTAAACTCG AGTTTAAGGATGTCATAGTCGGTTTTTTG	AATTCAAAAAACCGACTATGACATCCTTA AACTCGAGTTTAAGGATGTCATAGTCGGT
BLM	3UTR, CDS	CGAAGGAAACTC ACGTCAATA	CCGGCGAAGGAAACTCACGTCAATACTC GAGTATTGACGTGAGTTTCCTTCGTTTTT G	AATTCAAAAACGAAGGAAACTCACGTCA ATACTCGAGTATTGACGTGAGTTTCCTTC G
SLX4	3UTR, CDS	GCCTCCCAAAGT GCCTATAAC	CCGGGCCTCCCAAAGTGCCTATAACCTC GAGGTTATAGGCACTTTGGGAGGCTTTTT G	AATTCAAAAAGCCTCCCAAAGTGCCTATA ACCTCGAGGTTATAGGCACTTTGGGAGG C
TERF1	CDS	GAACGCCTTATCG CAGTTAAA	CCGGGAACGCCTTATCGCAGTTAAACTC GAGTTTAACTGCGATAAGGCGTTCTTTTT G	AATTCAAAAAGAACGCCTTATCGCAGTTA AACTCGAGTTTAACTGCGATAAGGCGTTC
Suv39h1	CDS	CTCTGCATCTTCC	CCGGCTCTGCATCTTCCGCACTAATCTCG	AATTCAAAAACTCTGCATCTTCCGCACTA

		GCACTAAT	AGATTAGTGCGGAAGATGCAGAGTTTTTG	ATCTCGAGATTAGTGCGGAAGATGCAGAG
Setdb1	CDS	TGCTATCTGGGAACCATATTG	CCGGTGCTATCTGGGAACCATATTGCTCGAGCAATATGGTTCCCAGATAGCATTTTTG	AATTCAAAAATGCTATCTGGGAACCATATTGCTCGAGCAATATGGTTCCCAGATAGCA
Kat2b	CDS	CCGGGATATTATGAAGTTATA	CCGGCCGGGATATTATGAAGTTATACTCGAGTATAACTTCATAATATCCCGGTTTTTG	AATTCAAAAACCGGGATATTATGAAGTTATACTCGAGTATAACTTCATAATATCCCGG
P300	CDS	CCCTGGATTAAGTTTGATAAA	CCGGCCCTGGATTAAGTTTGATAAACTCGAGTTTATCAAACTTAATCCAGGGTTTTTG	AATTCAAAAACCCTGGATTAAGTTTGATAAACTCGAGTTTATCAAACTTAATCCAGGG
CBP	CDS	CGCGAATGACAACACAGATTT	CCGGCGCGAATGACAACACAGATTTCTCGAGAAATCTGTGTTGTCATTCGCGTTTTTG	AATTCAAAAACGCGAATGACAACACAGATTTCTCGAGAAATCTGTGTTGTCATTCGCG

Results

Telomere of ALT mESC uses subtelomeric template

In 2000, H. Niida et al., published a paper that showed Terc knock-out mouse ES cells can survive after about 400 population doublings (Figure 4) (Niida et al., 2000). During passaging cells, most of the ES cells entered replicative senescence and died from telomere crisis. However, after ~50 passages of the crisis-suffering cells a little fraction of cells overcame the crisis and became immortalized. These cells were morphologically different with the pre-senescence cells, but proliferated at a similar rate with wild type cells. The authors didn't use a term like ALT, but the survived cells were ALT cells by definition. The paper also showed non-telomeric sequences were recruited to telomeres in one of the ALT survivor cells (DKO 741). I got the ALT mESC from Dr. Niida at each critical time point; PD100, pre-ALT cells which is almost normal because of remaining long telomeres; PD350, pre-ALT cells just before an expected crisis point (~PD400); PD450, post-ALT cells just escaped from the crisis; PD800, post-ALT cells surviving stably. From now on, pre-ALT and post-ALT mean PD100 and PD800, respectively, if not stated in other ways specifically.

To examine which regions in the genome were amplified, I performed whole

genome sequencing (WGS) and checked changes of copy number of genomic regions. Scatter plot of copy number variation shows terminal regions of the right part (q arm) of chromosome 13 are amplified significantly in post-ALT (Figure 5). The length of the amplified regions is about 7.4 kbps and the region includes a gene named Tcstv3 (Figure 6 b). The region is located at the very subtelomere of the right part of chromosome 13 (Figure 6 a). I named the region as mTALT, which means the template of ALT in a mouse.

mTALT copy numbers were calculated based on WGS data. The coverage of mTALT was compared with that of single copy control region. The result showed that the copy number of mTALT was increased as cell progressed from pre-ALT state toward ALT activation and stabilization (Figure 7). In particular, post-ALT cells showed a highly increased copy number of mTALT. Another type of ALT mESC (301 line) did not show an increased mTALT pattern.

Next, I checked the pattern of mTALT amplification. I performed a terminal restriction fragment (TRF) assay with restriction enzymes which can cut mTALT. After cutting genomic DNA with the enzyme, southern blot was conducted with telomere repeat specific probes. Only post-ALT (A1 in Figure 8) showed a discrete band resulting from the

enzyme action. It means post-ALT telomere includes mTALT and the interval between neighboring mTALTs may be fixed (Figure 8). To check the co-localization of mTALT and telomere, I performed FISH with mTALT specific probe and telomere specific probe. mTALT and telomere sequences were well co-localized only in post-ALT (Figure 9).

Looking into mTALT region (Figure 8), some part around 120,317,000 seemed not duplicated. When I used BaeI enzyme which can cut the underrepresented region in southern blot process, telomere band of post-ALT cells showed a short band. This result indicates that the BaeI-cut region also exists in post-ALT telomere. To confirm the constant duplication of mTALT region, I used several primer sets to amplify various parts of mTALT. As shown in Figure 10, all regions inside mTALT showed a similar degree of amplification. Taken together, the ununiformed duplication of mTALT in WGS data may result from the difference between the reference genome (mm10) and the strain of ALT cells (129/Ola).

To check directly the translocation of mTALT to other chromosomes, I performed a PCR assay with subtelomeric primer and mTALT specific primer (Figure 11). When I used primers specific to chromosome 1 or chromosome 5, both of the PCR showed post-ALT specific production of PCR bands. This result indicates mTALTs were moved to other

chromosomes from the original site of chromosome 13.

I quantified the level of telomere repeats with specialized program Computel and Telseq, which can calculate the telomere length based on telomere repeat counts from WGS raw data (Figure 12) (Ding et al., 2014; Nersisyan and Arakelyan, 2015). The results showed that the telomere length decreased from PD100 to PD350, and gradually increased after ALT activation. The telomere lengths are similar between PD100 and PD800. Considering the extensive amplification of mTALT in post-ALT, most of telomere in post-ALT mESC is composed of mTALTs. Canonical telomere repeat is rescued to the level of pre-ALT, but not amplified much above that.

Many human ALT cells use variant telomere repeats located in subtelomeric region to maintain a terminal stability. To check whether ALT mESC utilizes the similar recruiting mechanism of variant telomere repeats, I used a program, Telomere hunter (Feuerbach et al., 2016). The result indicated that a composition and a relative proportion of variant telomere repeats were not different between pre-ALT and post-ALT (Figure 13).

mTALTs are duplicated with a regular rule and a structure

I validated mTALT was selected and employed for the maintenance of ALT mESC telomeres. I wondered whether there was any specific rule of mTALT duplication. To check the duplication pattern, I performed a PCR assay which can discriminate the directions of connected mTALTs (Figure 14). Specific primer pair orienting from inside of mTALT to outside was designed. This primer pair cannot produce an amplicon with just one copy of mTALT. With forward or reverse primer only PCR, no amplicon was detected. However, with both of primers, a PCR band of particular size was produced. It means that mTALTs are duplicated in a head-to-tail pattern very regularly and the intervals between adjacent mTALTs are highly constant. If combined with the southern blot result with telomere repeat probe showing a band of unique size, it's evident that the contents between mTALTs are telomere repeats.

WGS data can also corroborate the regularity of duplicated mTALTs. The paired reads of the reads containing telomere repeats were extracted and aligned to the reference genome (Figure 15 a). The aligned peaks were majorly located at both 5' and 3' end parts of mTALTs (Figure 15 b). It means in ALT mESC telomere, telomere repeats are bordering on the terminal regions of mTALTs and seldom integrate into the internal part of

7.4kbps of mTALTs. Thus, ALT mESC telomere entails a replicating template of fixed size which consists of the whole stretch of mTALT and the flanking telomere repeats. mTALT and telomere repeats are not interweaved randomly.

The reference genome shows there is just one copy of mTALT in chromosome 13. However, even in pre-ALT cell, mTALT seems to be two copies, not a single copy. It was first revealed in WGS data showing mTALT showed two-fold coverage than single copy control region in PD100 cells (Figure 7). To confirm the duplication state of mTALT, I used PCR assay shown in figure 12a. In the reference genome, there was a certain region ("A" in figure 16 a) between mTALT and canonical telomere repeats. The forward primer was picked inside A region, and the reverse primer was picked inside mTALT. If mTALT exists as in the reference genome, there should be no amplicon. However, the PCR assay showed the presence of amplicon of a specific size in all tested cells (Figure 16 b). Taken together, mTALT was cis-duplicated before ALT activation and the duplication persisted after turning on ALT.

There can be two potential pathways for mTALT duplication in post-ALT: one is that two-copies of mTALTs were both amplified with a specific frequency, and the other is

that one copy of mTALTs was selected and duplicated (Figure 17 a). To identify the actual phenomenon, I examined SNPs in the mTALT region in pre-ALT and post-ALT. With the SNPs, I calculated allele frequencies as ratios of variant SNPs to reference SNPs at each point of mTALT (Figure 17 b). The trend line in the PD100 denotes mTALT region retains 1:1 ratio of two SNPs, which means the presence of two types of mTALT. In contrast, the allele frequency of PD800 showed a strong bias toward value 1, which implies only one type of mTALTs exists in PD800 cells. During ALT activation, a specific copy of mTALTs was selected and amplified for telomere maintenance.

Next question was how accurately the mTALT was duplicated. I plotted SNP types according to cell specificity (Figure 17 c). At each position of mTALT, SNPs were categorized as common, PD100-specific, or PD800-specific. As shown in figure 13c, common and PD100-specific SNPs were prevalent, but PD800-specific SNPs were extremely rare; there were two SNPs of PD800-specific. This result also indicates a portion of mTALTs in PD100 was selectively duplicated. In addition, the replication mechanism of mTALT is distinctly accurate and operated in an error-free manner.

The rule of mTALT duplication in terms of orientation can be a clue for the ALT

mechanism. The possible cases are majorly two. Case 1 is that mTALT are duplicated with a constant orientation. Case 2 is that mTALTs are randomly oriented with amplification (Figure 18 a). I used a specific read mapping assay to identify the real case. First, from WGS data, telomere repeat containing reads (R1s) were extracted and sorted depending on repeat direction. Second, paired mates (R2s) of telomere reads (R1s), containing mTALT sequences were mapped to the genomic mTALT location. When R2s of R1s with (TTAGGG)n repeats were mapped to mTALT region, the left part of mTALT showed an enrichment of peaks (upper track of figure 18 b). In contrast, when the complementary R1s (CCCTAA)n were used, R2 peaks were highly enriched at the right part of mTALT (lower track of figure 18 b). These results indicate that telomere repeats are connected to mTALT sequences in a uniform directionality. Thus, mTALTs were duplicated in a regular orientation, not in a random way.

One of the possible mechanisms of mTALT duplication is transposon-like behavior. There are two types of transposon depending on transposition intermediates: DNA-based (DNA transposon, cut-and-paste) and RNA-based (retrotransposon, copy-and-paste). When I performed a PCR assay to check whether the original TALT was intact, the

result showed the unbroken state of mTALT before and after ALT activation (Figure 19 a). To check the possibility of depending on a retrotransposon-like mechanism, I treated reverse transcription inhibitor (3TC) and checked the change of mTALT. As a result, there was no significant alteration of mTALT copy numbers (Figure 19 b). Taken together, mTALT duplication is not a transposon-like mechanism.

I assayed well-known molecular markers of ALT, C-circle and APBs. However, both of them were not increased in post-ALT cells (Figure 20 and Figure 21). In another type of ALT mESC (301 PD800), APBs were increased slightly. These results may come from the imperfect correlation between ALT and molecular markers. Otherwise, 741 ALT mESC may involve a distinct mechanism.

Here's the brief summary of ALT activation in terms of mTALT amplification. In WT or pre-ALT cells at early passages, telomere was long enough to be functional to protect the genome integrity. After telomere was shortened below a critical threshold, most of the cells stop to proliferate and eventually died. Some of the cells can withstand the crisis and activate ALT mechanism. One of the two copies of mTALT has been selected and used as a template for telomere maintenance. mTALT is tandemly duplicated with a regular gap

(Figure 22).

ALT activation involves changes of RNA and protein expression

The cellular changes involved in the process of telomere crisis and ALT activation remain enigmatic. The widespread but not fully verified hypothesis is that a genome instability is aggravated in the course of increasing telomere crisis and various genomic changes can happen, which may be connected to the critical events that induce ALT. Mutations may accumulate at the level of a single nucleotide. At the same time, mutagenic chromosomal changes including gene deletion, duplication, translocation can happen and also epigenetic changes do.

To check whether there were genetic mutations generated during ALT activation, I compared SNPs between pre-ALT and post-ALT (Figure 23). SNPs of post-ALT were comparable to that of pre-ALT (2,707,593 vs. 2,823,502), which means mutations resulting from genome instability had not been accumulated much. Thus, the possibility that gene expression or functional changes caused by certain mutations initiated ALT was quite low. mRNA sequencing and transcriptome analysis were conducted to confirm the degree of

gene expression change. In principal component analysis (PCA), PD100, PD800, and PD800 WT showed distinct cluster formations (Figure 24 a). PC1 and PC2 retained 96% of the variance (PC1:49% and PC2:47%) of original data sets. Based on PC1, PD800 was most distinct from the others, which may imply ALT-specific pattern of PD800. In case of PC2, PD100 and PD800WT were on the most distinct position. The clustering analysis with top 30% of genes showing an expression change exhibited PD800WT was more distantly clustered than the others (Figure 24 b). When the genes were plotted on the basis of relative gene expression in post-ALT compared to pre-ALT, and cut-off with two-fold changes, there were lots of significantly changed genes. Taken together, despite the rare accumulation of mutations, post-ALT cell showed its own discrete RNA expression pattern.

To more deeply investigate the difference of gene expression between pre-ALT and post-ALT, I organized a quantitative proteomics analysis based on TMT-labelling technique. This method can identify the proteome with multiplexing up to 10 samples and compare the abundance of each peptide precisely. With pre-ALT and post-ALT cells, I can quantify about 8000 proteins (Figure 25). After cut-off with q-value and fold change, the list of the increased proteins in post-ALT was analyzed by DAVID and Gene set enrichment

assay (GSEA) (Figure 26). As a result, Gene ontology (GO) terms including "transcription", "transcription factor activity," and "sequence-specific DNA binding" were enriched in DAVID analysis. GSEA result also showed transcription factor activity and chromatin remodeling associated terms were enriched. Thus, the change of nuclear proteome may reflect an elaborate alteration of transcription networks with the change of transcription factors and chromatin states to maintain telomere and genome. However, the causality between the proteome and ALT mechanism should be carefully investigated. When I examined the RNA expressions of the most highly increased proteins, most of them showed a significant correlation between RNA and protein expression (Figure 27).

Local and global chromatin remodeling in post-ALT

Post-ALT cells showed the characteristic gene expression pattern in the point of RNA and protein. If SNPs were not the cause of the expression changes, the next candidate can be the change of chromatin state. In particular, GO analysis with proteome showed chromatin remodeling can be a meaningful pathway in mESC ALT. I carried out ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) to verify the presence

and effect of epigenetic changes. ATAC-seq can examine how much transposons are accessible to chromatin and determine the openness of chromatin. Pre-ALT and post-ALT cells were reacted with transposase mixture, and sequenced with Novaseq platform. The sequencing result showed post-ALT specifically accumulated peaks in various genomic regions. When the peaks were aligned based on the location of transcription start sites (TSSs), the near-TSS region didn't show the ATAC-seq peak accumulation (Figure 28). The regions apart from TSS 50kbs either in upstream or downstream showed significantly high ATAC-seq peaks. The chromatin state of those regions can affect gene expression if associated with enhancer function.

To examine the regional characteristics of the open chromatin, ATAC-seq peaks increased in post-ALT were arranged according to q-value and were checked for the aligned loci from the top of the list. Interestingly, the top 9 regions were all mapped to the subtelomere of chromosome 13. In particular, the top region (120,311,697-120,319,273) was same as mTALT locus and showed the highest pileup index. Pileup index is calculated from the ATAC-seq law data as the raw reads are converted to the original fragment sizes and piled up aligning to the reference genome. The outcome that mTALT region gathered

the highest ATAC-seq pileup may indicate the chromatin of mTALT turned open and accessible to Tn5 transposase. However, another possibility is that the increased copy number of mTALT in post-ALT could result in the accumulation of ATAC-seq peaks even with a consistent chromatin state.

To check the change of chromatin state in mTALT region clearly, I extracted ATAC-seq peaks of pre-ALT and post-ALT in mTALT region. Next, I binned each peak in 10bp scale and normalized the binned peaks with WGS reads counts corresponding to each bin. In this way, I can consider the copy number change of mTALT in pre-ALT and post-ALT. It was shown that normalized ATAC-seq peaks were generally higher in pre-ALT (Figure 29 a). If the peaks were represented as the ratio of post-ALT to pre-ALT, most of mTALT region showed value below one (Figure 29 b). However, one region (120,317,100-120,317,270) inside mTALT and terminal regions of mTALT showed the enriched peaks in post-ALT. Due to the repetitive nature of mTALT sequences, the interpretation of ATAC-seq peaks needs a careful investigation with additional specific assays for transcriptional activity or other read-outs of the chromatin state.

The chromatin state of telomere is a critical variable in telomere biology,

especially in ALT activation. I quantified ATAC-seq peaks of telomere repeat from raw data and normalized them with telomere repeat counts from WGS using computel (Figure 30). As telomere became shortened, ATAC-seq counts of telomere were increased (PD100 vs. PD350). After ALT activation (PD450), telomere counts were decreased to a similar level with PD100. At the point of ALT stabilization (PD800), telomere counts were slightly more decreased. Thus, the chromatin state of telomere experienced a dramatic change. As telomeres became shorter, an epigenetic opening occurred until the crisis. When the ALT was activated and telomeres were protected properly again, heterochromatin seemed to be rescued to the level of the normal telomere.

The location of regions having changed chromatin may have a specific trend. I extracted the top 500 peaks according to q-value and plotted the relative positions in the respective chromosome of each peak (Figure 31 a). In y-axis, value 0 represents the left end of chromosome and value 1 does the right end. As shown in figure 31a, the density of dots was increased as the observed region moved closer toward value 1. When the top 500 peaks were aligned based on the relative positions in the chromosome and plotted again, the tendency was also evident (Figure 31 b). This pattern indicates that the right arm region

of chromosomes was prone to be epigenetically open after ALT activation. The left ends were vulnerable to chromosomal fusions as centromeres were located close to the left ends, and the fusion between left ends can be easily stabilized.

Chromatin remodeling may lead to transcriptional changes. In addition to the transcriptome analysis, I checked the transcriptional change of telomere after ALT activation. Telomere of ALT mESC includes not only canonical telomere repeats but mTALTs, which indicates the content of TERRA should also contain mTALT transcripts. In the result of RNA-seq, the transcripts including mTALT and telomere repeat containing TERRA were both increased in post-ALT cells (Figure 32). Furthermore, there was a specific direction of transcription in both kinds of transcripts. mTALT and canonical telomere repeat transcripts were in reverse orientation, specifically from terminal to internal region. It was a surprising result that TERRA has been known to be transcribed from subtelomeric start sites and be forwardly oriented.

It may be due to the novel promoters and transcription start sites originated from duplicated mTALTs, which were cryptic before mTALT amplification (Figure 33). However, the copy number change of mTALT should be considered carefully when

interpreting the increased mTALT transcripts. Anyway, the quantity of telomere repeat is similar between pre-ALT and post-ALT, which can lead to the inference that the increased transcription of telomere repeat is derived from the change of transcriptional efficiency.

Distinct protection mode of ALT telomere

mESC ALT entails the unique mTALT sequences as well as canonical telomere repeats. ALT telomere has been known to be partially deprotected and show DNA damage markers. When telomere dysfunction-induced foci (TIFs) were observed in pre-ALT and post-ALT cells, post-ALT cells showed increased TIFs (Figure 34). The increased TIFs may mirror the spontaneous telomere damages which can be resulted from replication stress or other sources.

Next, I examined the recombination state of post-ALT cells with chromosome orientation fluorescence *in situ* hybridization (CO-FISH). CO-FISH can reveal the degree of recombination in terms of sister chromatid exchanges in telomere region. Post-ALT cells showed increased T-SCE, which means higher recombination events (Figure 35).

In telomerase positive normal cells or human ALT cancer cells, shelterin

manages a critical role for telomere protection and cellular survival. To check mESC ALT also depends on shelterin for telomere stability, the function of TRF2, the major component of shelterin, was depleted with the shRNA expressing construct. After the selection with antibiotics, ALT cells died after 3~4 passages. It means that telomere repeat binding TRF2 carried out a critical role for telomere maintenance in post-ALT cells. Before all of the TRF2-depleted cells had died, the remaining cells were collected and checked for TIFs. Compared to control group cells, TRF2-depleted post-ALT cells had significantly increased TIFs, which indicated the exaggerated telomere damage induced by TRF2 reduction lead to a cellular death (Figure 36).

TRF2 localization in telomere was checked with the immunostaining combined with telomere FISH. In metaphase, post-ALT cells showed TRF2 and telomere co-localization (Figure 37). Thus, TRF2 seems to have a critical role for terminal protection. Telomere repeats co-existing with mTALTs seem to be selected for a protective role, not by just a random incident. One of the main role of TRF2 is T-loop formation. The terminal structure of ALT mESC telomere has not been investigated completely. To fully understand the protection mode of ALT, next assays should verify the exact sequence composition of

the very end of telomere, the presence of single-strand overhang, and the T-loop formation.

Among the most significantly changed proteins, there were stemness-related transcription factors like Klf4 and Nanog. When these transcription factors were depleted respectively, each cell showed slow cell growth and the increased telomere DNA damages (Figure 38). Klf4 and Nanog are crucial transcription factors to maintain stemness. The upregulation of stemness factors in post-ALT cells may reflect a physiologically important change, but the extensive regulatory networks of them make it difficult to define a fundamental phenomenon.

HMGN1 is the novel ALT telomere binding protein

The most highly enriched protein in quantitative proteomics was HMGN1, a non-canonical histone protein. HMGN1 is a notable member of high mobility group family and competes with linker histone H1 for binding the nucleosome core particle (Birger et al., 2003; Gerlitz, 2010; Herrera et al., 2000). In contrast to H1 stabilizing the compact chromatin structures, HMGN1 can loosen the chromatin with antagonizing the function of H1. HMGN1 also can affect the levels of histone post-translational modifications and

regulate the chromatin remodeling machinery. In this manner, HMGN1 can play an important role in controlling chromatin structure and function, which results in the dynamics of DNA metabolism including transcription, replication, and repair (Kugler et al., 2012; Postnikov et al., 2012).

When the expression levels of HMGN1 were checked in pre-ALT and post-ALT with western blot, HMGN1 was highly expressed after ALT activation (Figure 39). Next, I wanted to analyze whether the increased HMGN1 showed the characteristic binding pattern with the genome. After chromatin immunoprecipitation with HMGN1 antibody both in pre-ALT and post-ALT, telomere and mTALT were tested for HMGN1 interaction with qPCR. HMGN1 was enriched in telomere repeat of post-ALT (Figure 40). The result of mTALT region was similar, but mTALT may be too much duplicated to be interpreted with %input value (Figure 41). Canonical telomere repeat contents were comparable between pre-ALT and post-ALT, which means the interpretation based on telomere qPCR is rather appropriate. In the experiment which combined Immunostaining with HMGN1 antibody and FISH with telomere-specific probe, the co-localization of HMGN1 and telomere were increased in post-ALT (Figure 40 b). Taken together, HMGN1

is up-regulated and seems to interact highly with telomere in ALT mESC.

The increased HMGN1 may have a role in a cellular physiology. I depleted HMGN1 in post-ALT with shRNA expressing lentivirus. The cellular growth rate was slightly decreased after HMGN1 depletion. When the telomere damage was analyzed, HMGN1 knock-down induced the increased TIFs in interphase (Figure 42). Next, I checked telomere length of the HMGN1-depleted cell. The cells were passaged for about two months (~40 passages) to check an effect on telomere length clearly. The mmqPCR with telomere or mTALT primers showed about 30% decrease of telomere length (Figure 43). These results denote the increased HMGN1 in post-ALT is necessary for the maintenance of ALT telomere, and the HMGN1 dysfunction can lead to telomere damage and shortening.

Next, I checked the role of HMGN1 in recombination with CO-FISH. HMGN1 knock-downed cells didn't show any difference in T-SCE with control cells (Figure 44 a). Surprisingly, HMGN1 knock-down decreased fragile telomere phenotype (Figure 44 b). These result indicate that the upregulated HMGN1 is mostly linked to replication stress which is a major source of telomere fragility. T-SCE can only capture the post-replicative

telomere exchanges. Interchromosomal recombination and/or BIR events especially in G1 should be investigated.

Lastly, I checked whether the increased HMGN1 can regulate telomere transcription. RNA was extracted from HMGN1-depleted cells and reverse-transcribed to make cDNA. Using mTALT specific primers, the levels of transcripts containing mTALT sequences were measured. The result showed that HMGN1 depletion indeed induced the decrease of mTALT transcripts (Figure 45). Thus, the increased HMGN1 in ALT mESC affected telomere transcription positively.

HMGN1 could influence the chromatin state and DNA metabolism in various ways. Whether the process that HMGN1 affects telomere depends on the presence and function of TERRA should be tested. TERRA was specifically targeted for down-regulation without affecting HMGN1 level. If cells were transfected with the LNA gapmers specific to telomere repeat, RNase H can degrade the target RNA base-pairing with the LNA gapmers. When the LNA gampers complementary to TERRA were transfected (Figure 46), TIFs were increased. Thus, the presence of TERRA seems to promote a favorable environment for telomere stability. TERRA is not a byproduct of the

increased HMGN1, but an effective player in ALT mESC.

Considering the upregulated HMGN1 increased TERRA production and the interaction between HMGN1 and nucleosomes unfolded the chromatin, I reasoned TERRA should interact with telomere directly to form R-loop. I conducted DNA-RNA hybrid immunoprecipitation (DRIP) with R-loop specific antibody and quantified telomere content with a slot-blot assay. As a result, post-ALT cells showed higher DRIP signal than pre-ALT cells in telomere region (Figure 47 a, b). Next, I performed DRIP-qPCR to confirm the presence of R-loop, which identified post-ALT telomere contained more R-loop than pre-ALT telomere (Figure 47 c). When the qPCRs were done with mTALT specific primers, the result was reverse (mTALT3) or not significant between pre- and post-ALT cells (mTALT4). The dramatic copy number change of mTALT in post-ALT cells may have caused an inconsistent result with telomere repeat. Although the quantity of R-loop averaged with total mTALT contents was lower in post-ALT than pre-ALT, the local accumulation of R-loop may have a specific impact on that telomere.

Interestingly, HMGN1 depleted post-ALT cells showed the lower level of R-loop (Figure 48). This result is free of the copy number issue of mTALT, as the change of

mTALT DNA is relatively small. Taken together, HMGN1 upregulated in post-ALT induced TERRA increase and potentially the change of local telomere chromatin. The overall changes lead to increased R-loop formation in telomere, which may be a good target for DNA repair machinery to initiate telomere synthesis (Figure 49).

Discussion

741 ALT mESC as a model for ALT maintenance

Two types of ALT mESC were characterized previously, and 741 type ALT line was a main target of this study. Post-ALT PD800 cell was compared with pre-ALT PD100 cells in quantitative proteomics and transcriptome analysis to differentiate the distinct gene expression patterns. PD800 cell was passaged extensively from the point of ALT activation and stabilized ALT cells. PD100 cell showed a relatively normal growth rate than PD350 cell which would suffer telomere crisis soon. The comparison between PD100 and PD800 has a limitation to reveal the exact molecular difference regarding to ALT initiation, because those two cells were too severely separated on the basis of time point. The list of differentially expressed genes in PD800 will be useful for understanding the maintenance mechanism of ALT.

The longitudinal tracking of the process of ALT activation will be appropriate to study the direct changes related to ALT initiation. If combined with single cell sequencing, to trace the intra-population heterogeneity and the selection process of ALT cells will reveal potentially causal events for ALT activation. The frequency of ALT survivor appearance can also be calculated. In the previous study, only ~10^4 cells were passaged with one plate

for each cell type. If the frequency of ALT appearance can be quantified with upscaling, it will be an important starting point to assess the impact of ALT initiation-related factors.

There are various kinds of ALT other than human ALT cancers in that non-telomeric sequences can be used for telomere maintenance. Type I yeast survivor, *C. elegans* TALT survivor, and *S. pombe* HAATI (heterochromatin amplification-mediated and telomerase-independent) survivor (Jain et al., 2010) are all good examples. The genetic factors required for each case may differ, so the general categorization has no meaning other than differentiating them from the canonical telomere repeat-based ALT. ALT mESC case is a similar one.

mTALT has unique features to be recruited to telomeres

The essence of ALT mESC in this study is that mTALT has been substituted for canonical telomere repeat. Telomere repeat also co-exists with a highly constant size in the middle of mTALTs. Among numerous subtelomeric sequences, mTALT may entail specific features suitable for telomere maintenance. These features may lie in mTALT itself (DNA sequence) or the comprehensive environment surrounding mTALT including the

interacting proteins and epigenetic states. Telomere repeat binding proteins including shelterin and other transient interactors may make specific touch with mTALT region. Until now, mTALT-binding proteins have not been identified. CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9-based protein enrichment of target DNA sequences can be possible. The list of mTALT binding proteins and their interacting partners will shed light on the identity of mTALT in telomere.

Even before the mTALT translocation to other chromosomes, two copies of mTALTs (PD100) has been doubled during telomere shortening (PD350). Whether this doubling of mTALT is connected to ALT mechanism is an open question. It might be a necessary condition of ALT activation. The locally increased copy number of mTALT may render it a suitable region as a replacement of telomere repeat. Mouse subtelomeres are highly dynamic than the core region of the genome. The major sources of dynamicity are repetitive nature of subtelomere and shared homologies between different chromosomes. Because of the incomplete reference genome, I couldn't examine completely the homologous regions of mTALT in different chromosomes. In a further study, I will focus on which properties of mTALT made it an ideal template.

To identify the direct role of mTALT sequences in ALT activation, the assay that shows the impact of the deletion or substitution of mTALT on ALT activation will be needed. There is another type of ALT mESC (301 line) similar with human ALT cancer cell which utilizes telomeric repeats for maintenance of chromosomal ends. mTALT deletion may lead to various results including the biased formation of 301-type ALT cells and the appearance of a new template sequence other than the original mTALT. The detailed substitution assay within mTALT may reveal a necessary motif of template sequences.

The dispersal of mTALT may follow a step-wise process. When telomerase mutant cells were experiencing crisis and some obtained ALT activity, mTALTs in chromosome 13 were employed by all other chromosomal ends. After mTALTs were integrated and stabilized in all telomeres, each ALT-rescued cell should carry on their own normal life. The major portion of telomere synthesis would be managed by DNA polymerase, and the ALT mechanism would be needed at critical time points. Thus, the duplication event of mTALT may involve at least two kinds of mechanism. One is for the dispersal of mTALT from chromosome 13 to all other chromosomes, and the other is the recurrent replication of mTALTs after ALT activation. The two processes may share a core

machinery or be independent to each other.

Potential mechanisms of mTALT duplication

The observation that mTALT is duplicated as a unit and the intervening telomere repeat has constant length implies some facets of the telomere synthesis mechanism. When a telomere synthesis process initiated, the synthesis may keep replicating the template for at least 7.4kbps. If there was a stopping point inside mTALT region during telomere synthesis, the following next synthesis event must create a delicate junction.

The accuracy of duplication mechanism of mTALT was surprising, as a bit of mutability could bring about substantial mutation accumulation. There was no accumulation of mutations in duplicated mTALTs. An elaborate and accurate mechanism would be necessary for the controlled synthesis to prevail all over telomeres. The error-free amplification has no other way than something based on HR. HR-based replication mechanism spans a number of subordinate mechanisms, including double-strand break repair (DSBR), synthesis-dependent strand annealing (SDSA), single strand annealing and break-induced replication. The elevated T-SCE frequency in post-ALT cells seems to result

from telomeric crossover events. However, this result is not sufficient for pinpointing the exact HR pathway involved in mESC ALT. The detailed mechanism of mTALT replication might be a mixture of those pathways. Different HR processes might take on their own part among various steps of ALT.

The ALT mechanism seems to be stabilized in a short time scale, as there was no inversion in repeated mTALTs in PD800 cells. If telomeres which acquired mTALTs after ALT activation would suffer telomere crisis again, there may be chromosome fusions and inverted arrangements of mTALTs. Although there is a need for more longitudinal investigation of telomere length, it seems that telomere length has been restored steadily after ALT activation . The dynamics of telomere length after ALT stabilization will be an interesting issue.

The distinct telomere maintenance pathway of 741 ALT mESC

Previously, a paper showed that Zscan4 is involved in recombination-based telomere maintenance and genome stability in mouse ES cells (Zalzman et al., 2010). The authors said that at a given time point only a minor portion (~5%) of ES cells expressed

Zscan4, but Zscan4 is expressed in almost all cells at least once during nine passages. Meiosis-specific HR genes are involved in Zscan4-mediated telomere recombination. Zscan4 pathway seems to be related to ALT in that recombination factors work for telomere lengthening. However, the role of Zscan4 has not been studied in telomerase knock-out ES cells.

741 ALT mESC showed a significantly decreased expression of Zscan4. It indicates that 741 ALT mESC may depend on different sets of genes for telomere lengthening. However, a little fraction of 741 ALT mESC might upregulate Zscan4 in a short period. Single cell RNA-seq will reveal the expression change of Zscan4 in a population of ALT cells.

Another distinction between 741 mESC ALT and canonical human ALT lies in the role of ‘alpha thalassemia/mental retardation syndrome X-linked’ (ATRX). ALT cancers are highly correlated with mutations in ATRX, which is a chromatin remodeling protein belonging to the helicase of the SWI/SNF family (Clynes et al., 2015; Lovejoy et al., 2012). With histone chaperone ‘death-domain associated protein’ (DAXX), ATRX facilitates replication-independent nucleosome assembly with histone variant H3.3. The

lowered ATRX/DAXX activity in ALT cancers seems to increase replication stress and promote telomere lengthening. However, 741 ALT mESC showed an increased expression of ATRX. The defective nucleosome assembly at telomeric region seems not to be a critical source of replication stress in 741 ALT mESC. Genetic depletion of ATRX will clear up the point.

Unresolved terminal structure of mTALT-driven telomere

Normal telomere has distinctive structural marks: 3' overhang and T-loop. Both of them are required for telomere protection and maintenance. Shelterin and telomere-interacting proteins cooperate to dissociate and re-assemble T-loop in every cell cycle. 3'-overhangs are also actively reconstructed after telomere DNA synthesis by DNA polymerase and telomerase. T-loop has been regarded as a key structure of the end-protection problem with hiding the terminals from the DDR machinery. TRF2 takes on a major role in T-loop formation. In ALT mESC, the exact terminal sequence and structure have not been elucidated. The presence of TRF2 at ALT mESC telomere and the role of TRF2 for cellular survival were confirmed. However, the involvement of TRF2 in T-loop

formation remains as a question. Telomere imaging with super-resolution microscopy will reveal the exact structure of ALT mESC telomere.

ALT initiation may not depend on the accumulation of genetic mutations

One of the interesting point in WGS result is that the numbers of SNPs were similar between pre-ALT and post-ALT. It seems that post-ALT didn't suffer from the mutational load in terms of DNA point mutation in spite of the dramatic change of karyotype. A good example for comparing with the activation of mESC ALT is tumor evolution. Tumor evolution is a complex biological process and still a big mystery. There are many hypothetical models to explain the process of tumor evolution. Some of the models include a concept that mutations were acquired gradually and contributed to divergent fitness of each cell. Environment-dependent selective pressure may decide which clones will be selected for the next steps. mESC ALT does not seem to follow a similar process with a gradual mutation accumulation.

In another point of view, post-ALT cells seem to be genetically homogeneous. It can be inferred from allele frequencies of variant SNPs. The plausible explanation of

homogeneity of post-ALT genome may be a few or some cells have been selected to revive from the cellular death of genomic catastrophe. In addition, there seems to be no additional mutability induced after ALT activation. Thus, there may be a selective sweep which was originated from events during telomere crisis and selected potential ALT-on cells. If mutation accumulation was not a critical event, structural or epigenetic changes may have played major roles at the point of ALT initiation.

One important value of ALT mESC is the availability of cells before and after ALT activation, which is unobtainable in many human ALT cases. A longitudinal study tracing ALT activation process will be an important basis for future research of ALT evolution. Mutation profiling with longitudinal samples will clarify the implication of mutations. If combined with single-cell sequencing, it will be easy to delineate the heterogeneity of pre-ALT cells in terms of transcriptional diversity and to get some insight of the conditions that promote ALT activation.

Potential roles of HMGN1 in ALT maintenance

HMGN1 was the main player of ALT mESC as revealed by quantitative

proteomics and genetic experiments. HMGN1 can induce chromatin decompaction for DNA metabolism regulatory factors to be recruited to nucleosomal DNA. DNA-dependent activities can be upregulated depending on recruited factors.

Post-ALT cells showed HMGN1-dependent fragile telomere phenotype, which seemed to reflect the replication stress of telomere. The molecular identity of fragility has not been uncovered. Interestingly, fragile telomere phenotype was correlated with moderate telomere elongation in telomerase-positive cells (Sfeir et al., 2009). Post-ALT cells also showed a correlation among HMGN1 upregulation, fragile telomere, and telomere lengthening. T-SCE was not affected by the level of HMGN1. The exact telomere lengthening mechanism mediated by HMGN1 should be investigated in detail.

Although HMGN1 can bind chromatin globally, the preferred targets of HMGN1 binding have not been fully elucidated. In a study using human T cells, HMGN1 co-localized with DNase I hypersensitive site, promoter, enhancer and transcription factor binding sites which are all transcriptionally active sites (Birger et al., 2003; Kugler et al., 2012). The causality of the co-localization remained as a big question. Knock-down or overexpression of HMGN1 changed transcriptome in specific ways, but the true meaning

of the changes was not interpreted. In ALT mESC, ChIP-seq approach of HMGN1 will be the first step for the interpretation of transcriptional changes.

Although there were a number of studies to reveal the biochemical properties of HMGN1, the physiological roles of HMGN1 have been highly difficult to catch. The study of HMGN1 knockout mice revealed that those mice showed an increased tumorigenicity, a hypersensitivity to DNA damaging condition, and mild developmental defects. Another study with HMGN1 -/- cells showed HMGN1 regulates ATM activation to control G2/M checkpoint (Birger et al., 2005). These results imply that HMGN1 has some traits of a tumor-suppressor gene.

Considering the presence and function of histone H1, the main opponent of HMGN1, the balance between HMGN1 and histone H1 may be a critical factor to determine DNA repair activity. HMGN1 have worked to efficiently repair DNA damages, and many studies showed histone H1 had an opposite effect. In a study with mESCs, the depletion of histone H1 induced increased DDR activation and higher HR (Herrera et al., 2000). Histone H1 can also inhibit histone acetylation regulated by HMGN1 and ATM activation in DNA damage condition. In my study, HMGN1 has a positive role in telomere

length stability. The question about the role of histone H1 in ALT mESC will also be interesting.

Normal telomere has been known to bind a lower amount of histone H1, whereupon the nucleosome in telomere has higher mobility than other genomic regions. Telomere seems to be a good substrate for HMGN1 binding with its own nature. However, the exact interaction between telomere and HMGN1 must be analyzed in ALT context. With ChIP-qPCR, I showed HMGN1 interact strongly with telomere in post-ALT. Next step is that the function of HMGN1 should be inspected in each step of events in ALT activation and maintenance.

How stemness is related to telomere maintenance

I found stemness-related transcription factors (Klf4 and Nanog) are upregulated in post-ALT. Pre-ALT cells are also mESCs with potentially more pluripotency than post-ALT cells considering karyotype and relatively earlier time point from telomerase knock-out. The increased stemness-related factors may reflect a distinct form of reprogramming which can help the ALT mechanism. When I reduced gene expressions of the factors,

telomere DNA damages were increased. Thus, these factors should have roles in telomere maintenance in ALT mESC.

Nanog promoter has been known to interact with many genomic regions which include open chromatin-related regions and pluripotency networks in ES cell or iPSC. Among binding sites of Nanog region, the target sites of Klf4 were highly enriched, which means those regions may be coregulated. The interactome changed dynamically with differentiation states. This study provided evidence that a key pluripotency gene interactome can undergo changes, which can control transcriptional patterns and dictate cellular identity.

The specific function of stemness factors at telomere still remains elusive, in particular at the point of ALT initiation. As telomere dysfunction became intensified following generations after telomerase knock-out, there may be a growing need for ALT activation to overcome the crisis. Stemness factors may be one of the initial factors upregulated to induce other essential processes to initiate ALT mechanism. To confirm the hypothesis, the ability to initiate ALT of stemness gene-depleted cells should be tested.

TERRA has a positive role in the protection of ALT telomere

The interaction between TERRA and telomere has been studied in various contexts. In ALT mESCs TERRA seems to have a protective role to maintain telomere stability. HMGN1 knockdown induced a decrease of TERRA, which ultimately led to telomere damages and telomere shortening. The direct down-regulation of TERRA with LNA gapmer also resulted in the same observation. The depletion of HMGN1 also lowered R-loop formation in telomere. Taken together, R-loop also acted as a positive regulator of telomere protection. R-loop may hinder chromatin-associated DNA metabolism including DNA replication and transcription. R-loop may greatly contribute to replication stress of telomere. To restore the stalling or collapsed replication forks, various proteins not related to DNA replication can be recruited including checkpoint and HR factors. These proteins ensure the stable maintenance of replisome at stalled forks. If replication fork failed to maintain the continual progression by the collision with R-loop, HR factors can help to rebuild the replisome. Break-induced replication underlies this repair process. (Llorente et al. 2008). BIR is a fundamental mechanism of telomere synthesis in human ALT cells. BIR may be involved in ALT mESC telomere synthesis in which TERRA is a major source of

replication stress. There may also be more layers of DNA replication/repair processes which can preserve the integrity of ALT telomere.

The distinction between mTALT transcript and canonical TERRA is an interesting issue. The direction of TERRA transcription in ALT mESC is opposite to that of most TERRA in other cells or organisms. It may indicate that rearranged telomere structure induced a cryptic transcription of TERRA, because most of the promoters of canonical TERRA have been regarded to be originated from proximal subtelomere.

Conclusion and perspectives

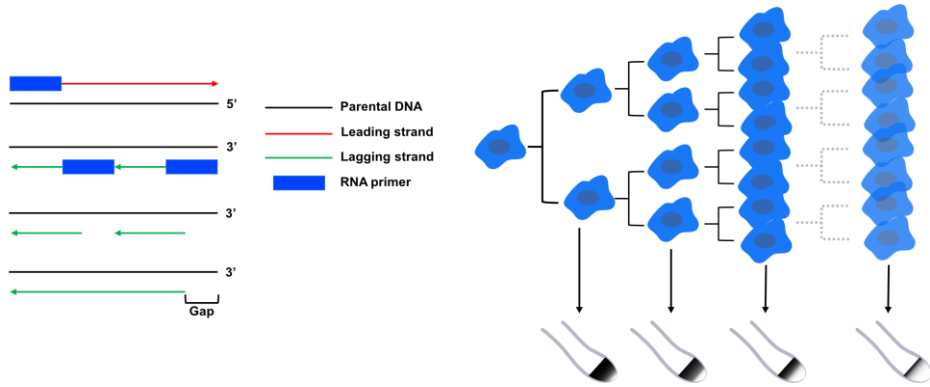
In this study, I established a novel ALT model which is conserved in eukaryotes at least in worm and mouse. *Trans*-duplication of mTALT from one chromosome to other terminals is analogous to the state of amplified TALT of *C. elegans*. This form of ALT seems to be similar with yeast type I survivors in that subtelomeric sequences are used as alternative terminal sequences. However, in ALT mESC the subtelomeric sequence is the very unique DNA block originated from the only one end of chromosome. There must be more steps to restore all the shortened telomeres in ALT mESC than yeast type I survivors

which harbor many copies of Y' subtelomeric sequences in different chromosomes.

HMGN1 functions to protect ALT telomeres, in part with upregulating TERRA and R-loop formation in telomere, which leads to the telomere lengthening pathway. HMGN1 may also work in a global way to affect chromatin state, transcriptional activity, and DNA repair efficiency. The role of HMGN1 was studied mainly focused on ALT maintenance. It will be an interesting question to analyze the relation between HMGN1 regulation and ALT activation.

The detailed molecular mechanism of mTALT duplication should further be investigated. GeCKO based genome-wide knock-out study or knock-down screening with mouse DECIPHER shRNA libraries will show the genetic basis of ALT with mTALT. The longitudinal study combined with single-cell sequencing will give critical clues to understand the evolution of ALT.

a



b

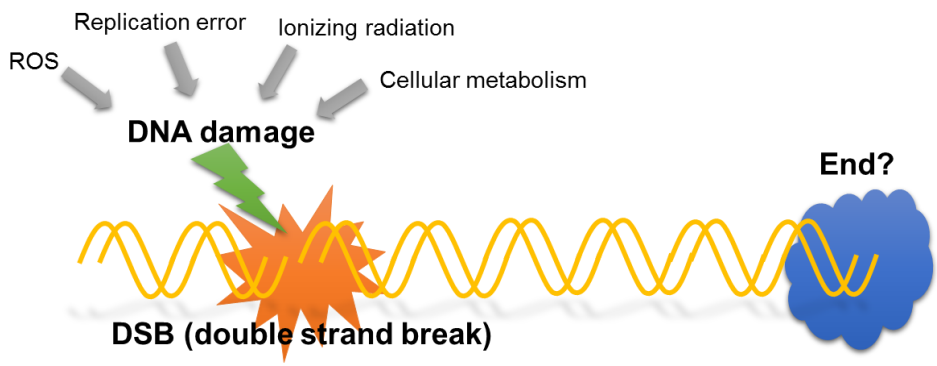


Figure 1. Two kinds of problems in chromosomal ends (a) Figure of end-replication problem. (b) Figure of end-protection problem.

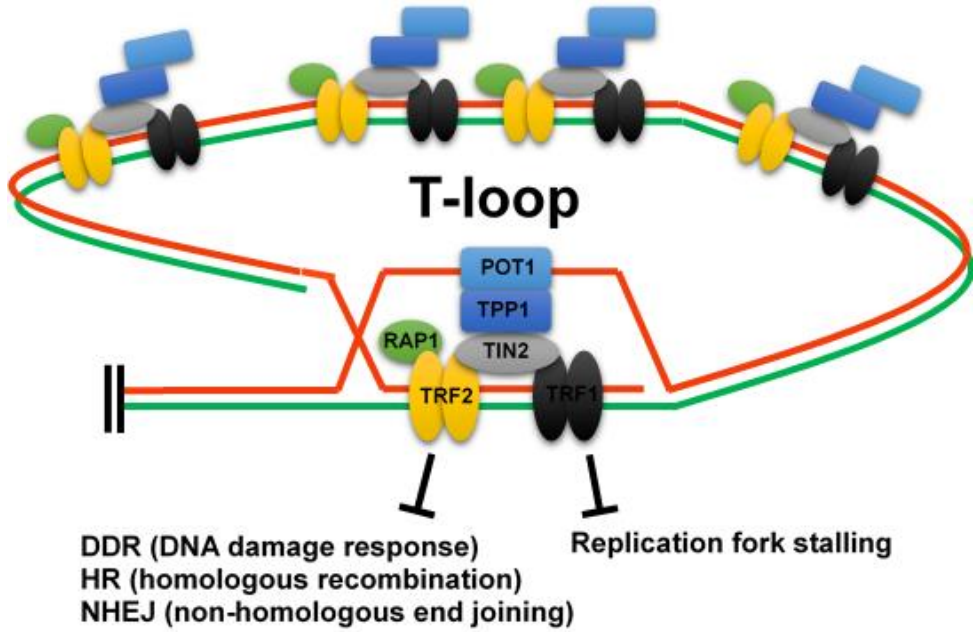


Figure 2. Schematic figure of T-loop structure and sheleterin complex

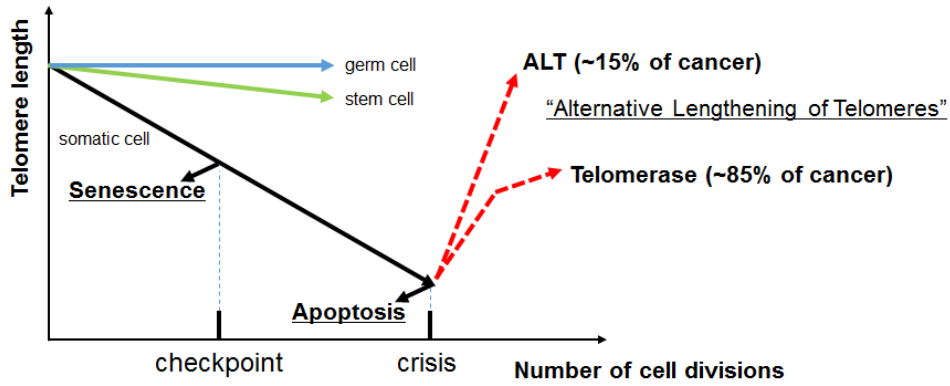
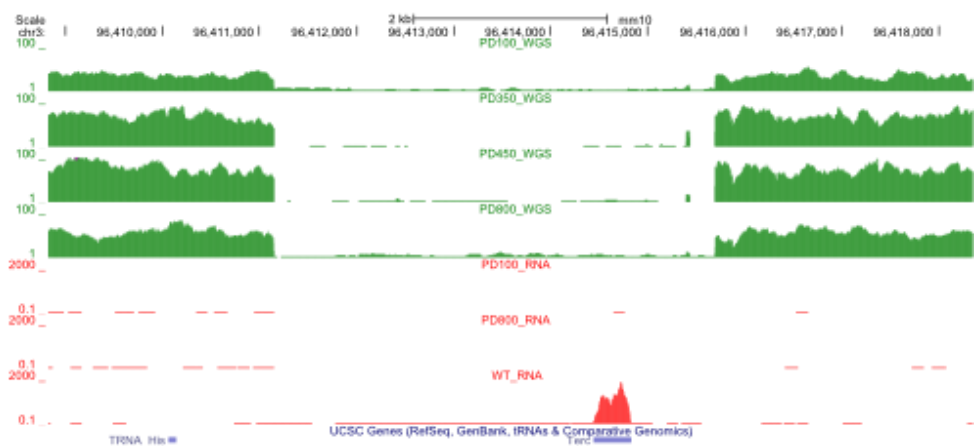


Figure 3. Model of activation process of telomere maintenance mechanisms

a



b

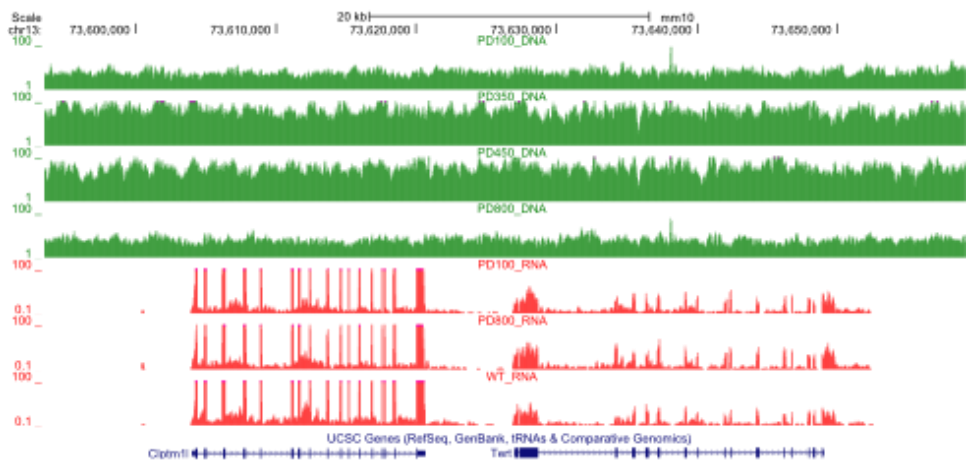
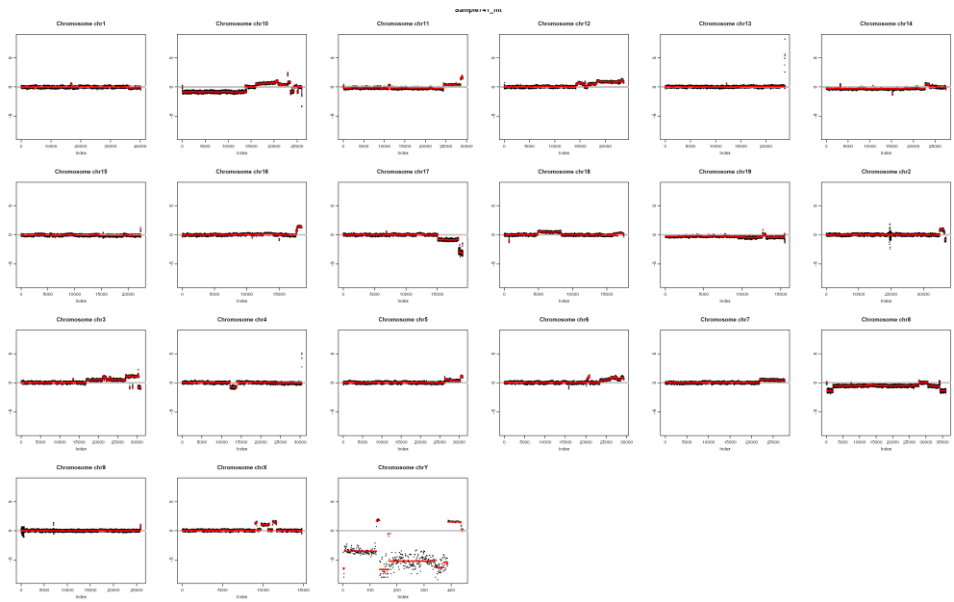


Figure 4. ALT mESCs are telomerase (Terc) knock-out. (a) Genomic snapshot and RNA expression of Terc gene. (b) Genomic snapshot and RNA expression of Tert gene.

a



b

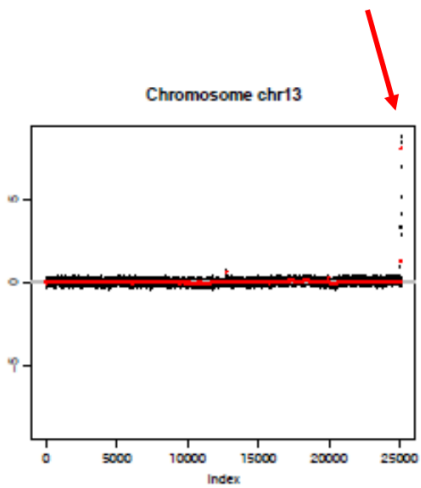


Figure 5. Copy number analysis of ALT mESC genome. (a) Graphical view of copy number changes in chromosome-wide plot. Each chromosome was scaled to

its physical map. (b) Enlarged plot of chromosome13. The subtelomere of right arm of chromosome13 was highly enriched in copy numbers.

a

b

Figure 6. mTALT is located at subtelomere of chromosome13. (a) mTALT locus is at the terminal region of the right arm of chromosome13. (b) Snapshot of mTALT region from WGS data. PD450 and PD800 cells showed much higher read counts of mTALT than PD100 and PD350. mTALT region includes a gene, Tcstv3, with its own promoter region.

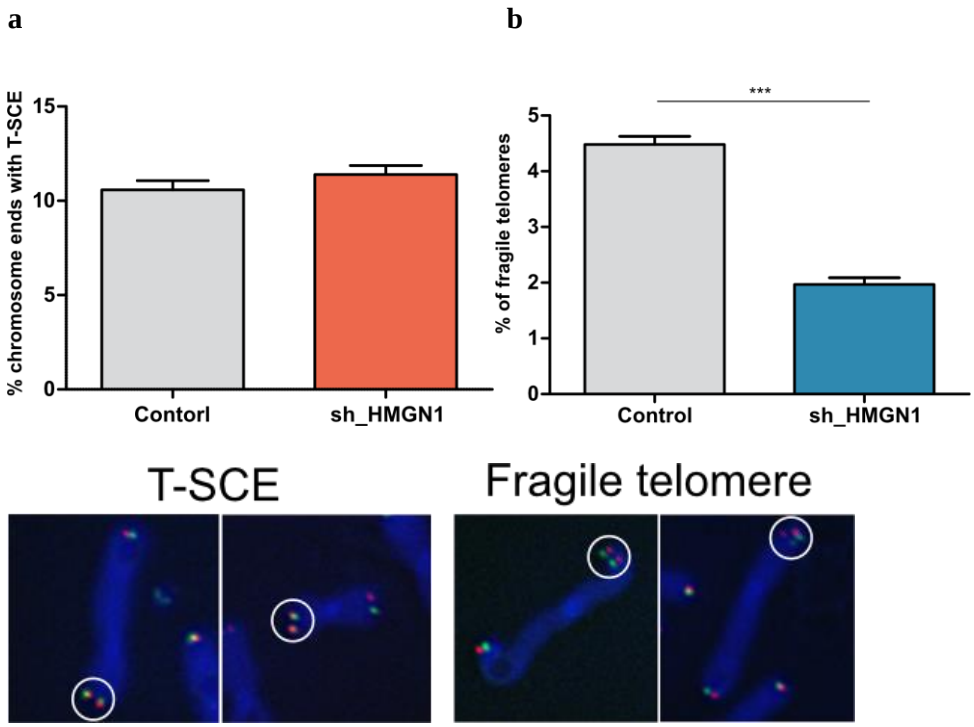


Figure 44. HMGN1 knock-down decreases fragile telomeres. (a) Bar graph showing the ratio of chromosome ends with telomere sister chromatid exchange (T-SCE). (b) Bar graph showing the ratio of fragile telomeres. The bars represent means and SDs from three independent replicates. P value from two-tailed unpaired t-test : *** P<0.001. All infected cells were post-ALT (PD800).

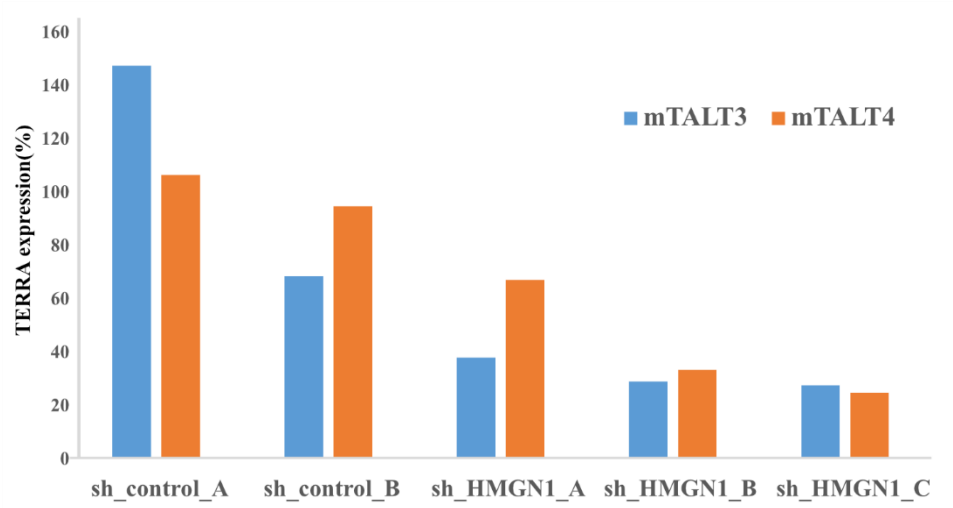


Figure 45. HMGN1 knock-down decreases TERRA expression. Bar graph shows TERRA expression quantified by qPCR with primers specified by colors. qPCR was done with β-actin gene as a control. In each group (control or HMGN1 knock-down) A, B, and C are independently infected clones. All infected cells were post-ALT (PD800).

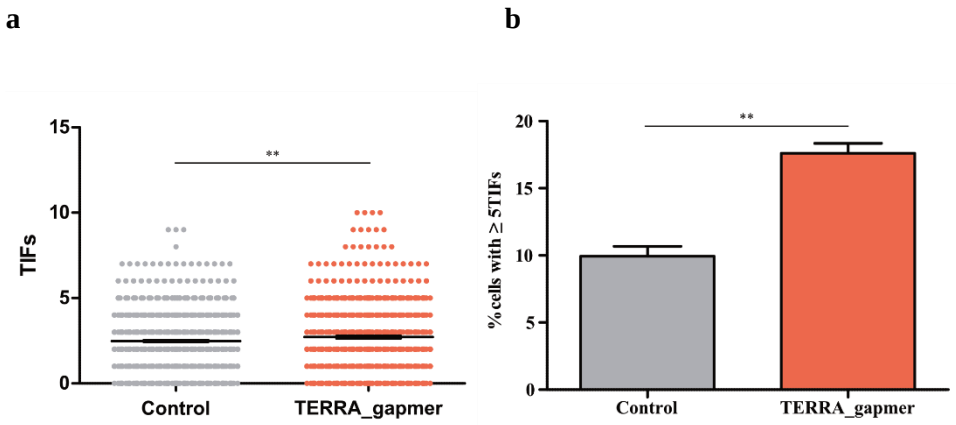


Figure 46. TERRA downregulation induces telomere DNA damages. (a) Dot plot showing the TIFs of cells treated with control gapmer and TERRA-specific gapmer. Each dot represents TIFs in a cell. The bars represent means and SEMs. P values from two-tailed unpaired t-test : ** P<0.005. (b) Quantification of the ratio of cells with more than 5TIFs. The bars represent means and SDs from three independent replicates. P values from two-tailed unpaired t-test : ** P<0.005.

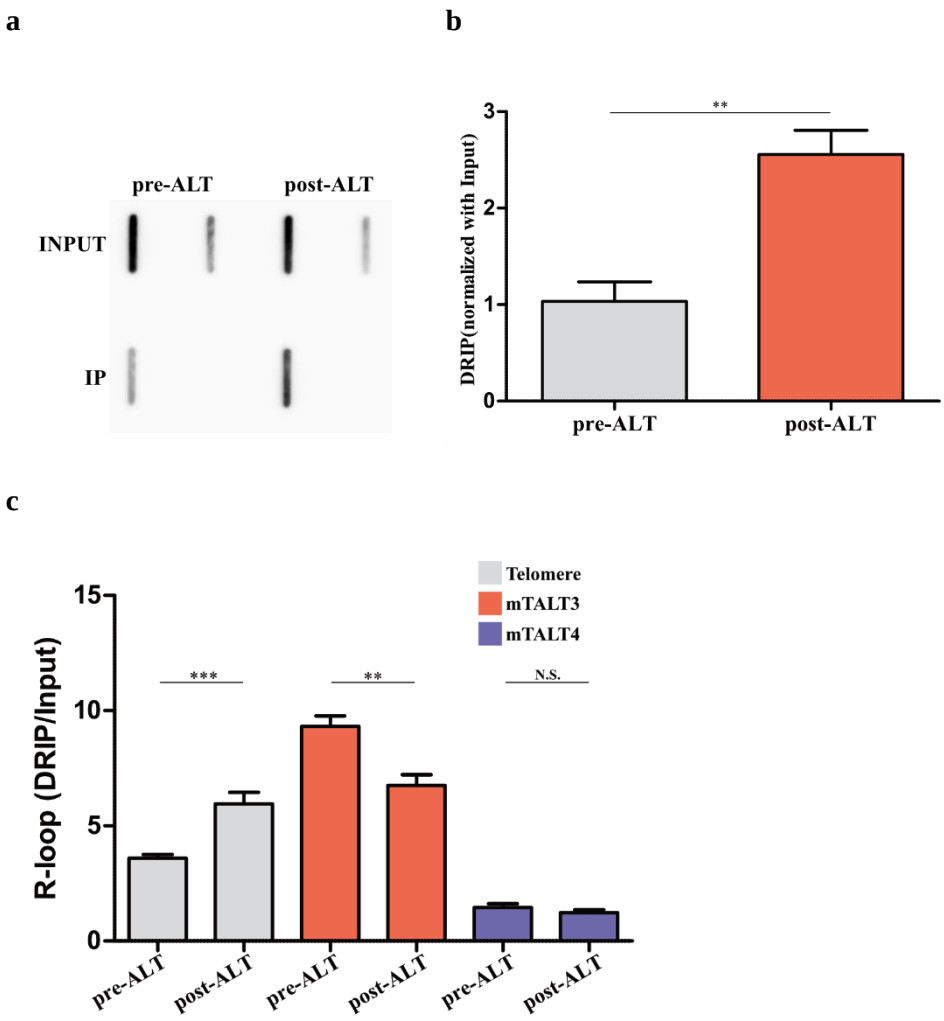


Figure 47. R-loop formation is increased after ALT activation.

(a) Representative figure of the slot-blot with telomere specific probe for DRIP assay. The input DNAs were loaded with 1/10 dilution. (b) Quantification of slot-blot results including (a). The bars represent means and SDs from three independent replicates. P value from two-tailed unpaired t-test : ** P<0.01.

(c) Bar graph shows the DRIP-qPCR result quantified with specific primers indicated as different colors. qPCRs with β-actin primer were taken as controls. The bars represent means and SDs from five independent replicates. pre-ALT cell (PD100) and post-ALT (PD800) cells were used. P values from two-tailed unpaired t-test : *** $P<0.001$, ** $P<0.01$.

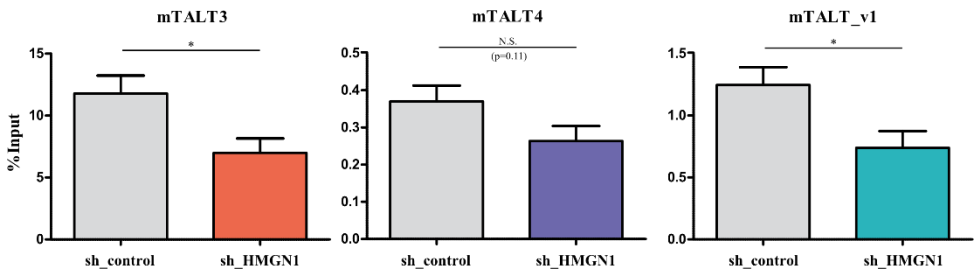
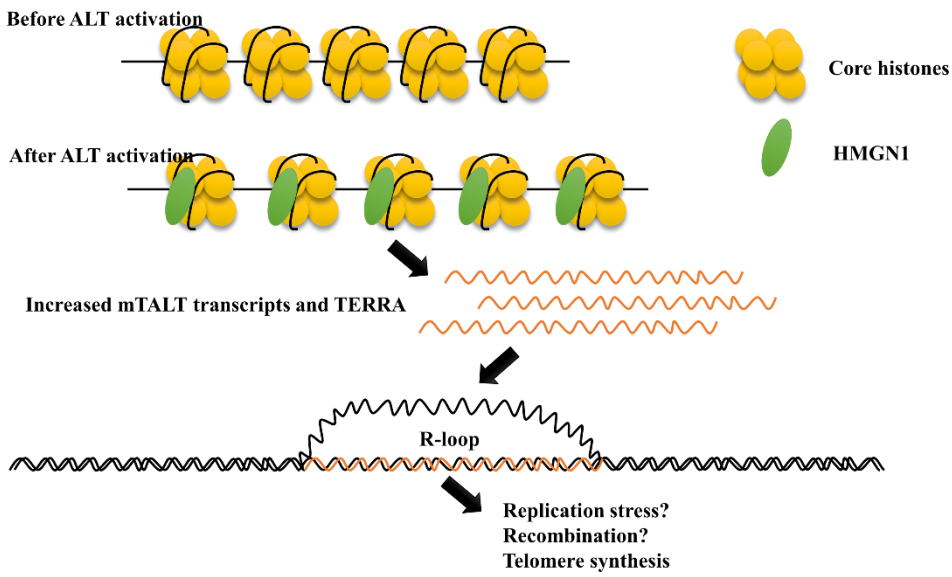


Figure 48. HMGN1 knockdown decreases the formation of R-loop in ALT telomere. Bar graphs shows the DRIP-qPCR result quantified with specific primers indicated at top of each graphs. qPCRs with β-actin primer were taken as controls. The bars represent means and SDs from five independent replicates. All infected cells were post-ALT (PD800). P values from two-tailed unpaired t-test : * $P<0.05$.



Figrue 49. Working model.

Appendix. Sequences of mTALT in mm10 reference genome.

>mm10_dna range=chr13:120311782-120319231 5'pad=0 3'pad=0 strand=+

repeatMasking=none

TTTATCTTCAAATGCAGCTCCATCCATCAGTGTGGCTATGCTGGGGTTCC
CAGGTCCACAGCACCAGGGGTGCTGGAATGAGTGATGTCTCAAGGCCTCA
AAATTATCCATGGCCCACGCCCATGCCTGTTCTCGGGTCATGTCTGGTGG
CATGACTGGGTCTGGTTGGTACATCCTTCCAAGCAGCTCTTCCAAGGGTG
ATGCTCCTGAAGTCATGACTTCTGTTTCCCAAGCAAGTAGCAGAGATTAG
CAGCTGAGGGGCAACTGAGATTGGAAAATCCAAAAGTCTAGTAGTTTCAG
CTGGTCTTCTGTATAAGCTCTAACCCTGAAGAGGTAGATTCAAACAGATG
TGCTGGCAAGAAATGGCAAGCAGGCAGAGTGAGTCCTCCTTCTTCCACTG
TCTTTATATAGGCCTCCAGAAGAAGGTGTGGCTCGGATTAAAGGTGTGTA
CCAGCAGTCCTGGATCTGGAACTTGCTTTATCCCAGGCTGACCTTGGATT
TAGGTCTGATTGACTCAGTTTCTGAGGAATAAAAGCATGAACTACTTTGC
CTAGGGCTAAGGCTTTCATGGCTACTCTGCTTCCGGATATCCATGCCAAG
ATATGTTTCAGAAGCCTGTGTCTTCTGCCCTCAAGATCTGCATCACAGTT
GTGCCCTCAATTTCTGTATTGTACTTCATTCCAGATGTAGTCAAGTTGAC
AACCAGAAATAGCCATCCCAAAAGTACACCTGAAAGCTCTAGAAAAAAAT
GTAGAGGGAAGCACGCTACAGAGGAGTAGAAGGCAGGAAATAATCAAAAT
CAGGGCTGTAAGTGTGTGTAGAGAAATGTAAGCCATATTGGCATGGTTTT
AGCTTCAGGAACCATGCATTGAAGGTCATAACTGAGGCACAAACATCTGA
ATCATGGGTTCTGACAGAATTGAACTGCAGCCTGGCTGGCAGAGTCTCAT
CCTGTCAGCAAAACGAAAGTACTGAGAAAGAAAATCATATACCCGGTTTC
AGCCCGGGCTTGTTAACATCAGAGCAGAATTTTGCTTGGCAGCAAGTAGA
AAGCATTCTCTAACAGGGATGGTCAGCAAGGACTCTGGACATCTGAAAAT
GGCCCAAGGCACACTGCTGGCTCCTGAGCGTCCTCTACTGTGAAAAAGTC
ACAGGCTTGCAACTTGGAGCTAAGAACCCTTTATCTTTCTGTCTAGACTC
ATTTATTGGCCAGGCTCACTTTCTGGGATCCACCAGTGTGGACAAGCAAG
CAGGCTTCTAACTATGGACTTGTCCAGGATCACAAAGCTGGCTCCAACCC

CTGGGTGTGCGAGTGAGAAGCCGTCAGTTAGTGTCTTCTGGCATGTTACA
CGCTGAACAAGTGCACACCAGGGCCATAGCCATGGAGGGACTGCAGAAAG
GAAGGCATGCAGGAGAAAGACAGCTCAGAGAAATACAGCTTAAGCGTCCT
GAGCAATGCTGTCCTGGAAAAACCTCTTCACAACAAAAGGAACGCAGCTT
CAACAGAGATTCCAGCAGACAGCAATTTTGTAGACCTCCATAGGTGGGCA
GCAGTGGGATGGAAAGCGAACCAAGGACAGCTTCGCCATACATTCAGTAA
TGAAAAGAAATTGGACGAATTTTCAGCGTTTAGAAACGAGGAAGAGCGGA
GAGGACACAAGACGATCAGAAACCGCAGGGTCTGCATCCCGTCTCCTGCA
GACTCACACCTGGGCGCCAGGCTTTCTGTCAGCATTTAGGAGCACACCAT
AAGTTCAAAAAAATGGAATCCCACCACATACAAAATCGGATCTGGTTGAC
AAAACGAAAGAACAGAGAAACAAAATCACGTACGCGGTTTGAAATGACCA
TCGGATTTCAGTCAGAGAGAGCATGCTGACAGCCAAGAACTTGGCTCTAT
GAGAAAGTTGCAGTATTTGCTACCTGAGGATCCAAAGGCGGGAAGTTTAC
CATTAATTGGCATTCCTCTCTGGGGGAAGAGCAGACCCACAGCGACCCCT
GCTGTGAAGTCGTGGCCAGACTGGAGCTCAGAGGTAGAGCTTTGAGGCTT
CTGGCTCCTTCATGGGTCCCCACTCCAGCTGTGGCCCCTACACGTTTGGA
AAAGGGGGGGAACAAACTATGGATTTTCCCAGGATCGCACGACTCGCCTA
AAACTCAGGGTGTGCGAGAGAGTCAACCGGGTGGTCCAGTGCTTTCCCTG
TTGGACAAGTGTACTCTTGGGCCATGGCCACGAAGCAACCAGGGAAACTA
CAGTGAGATGGACAGAAAACCTGAGATTGAAAGGCCTTAGGTAAACAGAA
AGGAAAGCAGCTGCGATGGAGAGTTGTGCTTGCACTCCGTTTGTCTTCTT
CAGGGGAAGGGGGCGGTGCTGCAAGTCTTGGCACACTTGGTGGGCTAAGA
AGATGGTATGTAATCTAAGTTTTGAAACGCGGGGAGGTCGAGAGGAGCAG
AAAACAAAACAAATCCGCGGAACCTGAATAGCATCTTCTACATCCCCGCA
CACACGCGCCAGGCATTTTGGTCAGGATTTAAGTGCACACCTTCCCGACC
TGGAGTCAGCACAGAATCCCACCTCCCCCTGCTTGACAGAGTCAAATCCT
GTCAGCACAGTGAATGTACAGAGAAACAAAATCATGAACGACCTTTGAAA
TCACCTTCGGATTTCAGCTCCAGACTTCTGACAGCCCTGTGGAACAAGGG
ATACAAGTTATAGCTACTAGAGAAGGGTGTCCAGAGTGTGAGCACCAAGG
TTGCCACCAGCTGGAGCGCCAGCACCTCAGCGACCCCTGTTGTGAAGAAG
TCTCTGGGCTGCACCTCTGAAGGAGAACTGTGGATCTGCCAATGGAGACT
ATTTGGCTTGTTGCCAGGACTCCTTCCTTGGCTCTGGCTCCTGCATTGCC

CTGGAACCACCAAAACTGTCCTGCATGACAGCGGTTCGTTGGAGGGGACC
AAGCTAACTAAAATGAAATAAAGAACCTACTATTAGTTATACTATATTGG
CGCATGTCTGTTATTTTTGGAGATAACTAACATTTCCCTGGCACTACAAT
TCTCTTCTTACCCCTCGCCTCCTTGGTGTGTTGTTTCTGTTTTATATACT
CTATAAGCAATTCTGCCTTTTTCACTTCTTTGTCTATGTTGTGGATTTGT
AAACACCAGATTTCAGGGTCCTTTGCTTGCTTCCCCCACCTCTTTCTGGA
TTACAAATTCCACCCACAGTGTGACTACAGAAAGTTTATGTGAGCCATCC
CAGAATACACACTGCTATTGTCCTGTGTGCCTGGGCCTAGAGCAGTAGAG
CTGGCTTTTTGGTCTTGTGTGTGAAGAAGCTGTCCTGGGATTCTGCACCA
GTGTCTCCTGGCCTCATCCAGTACTAACATGAGCAGATTTTGAGCCATGG
CTTTCTGCCCTTGACATCTGTACTGGCTAGTTTTGTGTCAACTTGAGTTA
TCTCAAAGAAAGGAGCTTCAGTTGAGGAAATGCCTCCATGAGATCCAACT
GTAAGGCATTTTCTCAATTAGTGATCAAGTGGGAAAGGCCCCTTGTGGGT
GGGACCATCTCTGGGCTGGTAGTCTTGGGATCTATAAGAGAGCAGGCTGA
GCAAGCCAGAGCAAGCAAGCCAGTAAAGAACATCCCTCCATGGCTTCTGC
ATCAGCTCCTGCTTTCTGACCTGCTTGAGTTCCAGTCCTGACTTCCTTTG
GTGATGAACAGCAGTATGGAAGTGTAAGCTGAATAAACCCTTTCCTCCCC
AACTTGCTTCTTGGTCATGATGTTTGTGCAGGAATAGAAACCCTGACTAA
GACAACATTCCTGACTTAGAAGGTCTCTCATTTAATTCTCTTTTCTCATG
TTATGAATCTGCGGGGCTATCTCAGGAAGTCCAAGATCACAAATATATAG
GGAGAGTAGTGAAAACAGGGTAAAATCCTCCATATTTGCTTATTATCTAG
TCCAGGTCCAGGGGAACTGAACTGAAGAGGCTTCTACTTGTTTAAATATA
CACTTTTCAAAATATTACAATGCCCAGAATCCAATCAAGGCTTATATGAG
AAAGTGCTGAGCTATGAGTTTCAAAATAATCCCCATTCAACTTTGATACC
AAATTATTTTGACAGTCCCTTAAAAACAGCTTTCCAGTCTTGCTGCCTGC
TTCCCTGCAAATATGAATCCAGTTCAGGGCAGCCTGGAGCCCAGCTACTC
CAGCAGATCCCTGTCTGCCTTCTGTCCAGACACAAAGACTAGAGGATGGA
AATGTTATGTGCATTCATCTCAAACACATGCATTGAAAACAGGGCTGGGG
GAGTGGGCGGGTGGAGAGGAGATTCCCTGTCCTGAAGCTTCATCTCTACA
CTTAGCTGTTTCCCCTAACCACTTCTCTGGTGGGAATAGAGGATTAGCAT
CTGCTCAGGACTCCAGCTGACCTAAGGACTATTAACAGGATTTCCAAGCA
GGACCTGGTGACTCATGCCTTTAATCTCAACACAAAGGAGACAAGGCAGG

GAGACTTGACTTGGTATCCAGCTTGGTCTACAAAGTGATTTTCATAAGAC
CCAGTGTTACATAGAGAAACACTCTCTGTGTTGGCTATTGTTTGTTATCA
ACTTGGTTATCTCTGCAATTAAGCAAAATCCAAAATGGGAAGATACATCT
GTAAGTAACTTTTGCTTAATTTATAGAAAGTAATCCATACCTGTGAGGTA
GGAAGACTTGCTTTTAATGCAGATCTTTTTGAGCTTGAAAGACAAGGCCT
CCAATTAGGATCTGAAGCCCGAAAGACACACTTTTTTGTTAATGTGATTG
TTGTGATAAGCTAGGGTTTGTCATTAATACCCCTGGCAGTCCTGTAACTC
ACTTTGTAGAGCAGACTACTCACAGAGATCTACTTCCTTGATTCAATTCC
ATAGTCCTTGGATTAAAGGCGTGAGCCACCACACTTAGCATAAGACACTG
AACTTTAATCCAGACGTTGAGGCTGGAACCACACACCTCTAATTCTGGCC
ACACCTTCAGGTGGAGGCCTATATAAGAACAGGGAGGAAGAGAGTTTTCA
CTTGGCCTGCTTGCTCTCTCATAGCTAACAAGTCTATTACTTCGCTGGCA
CTAGAGCCCCCTTGCATGTGATTCCATCTTCTATTGAAGACCAGCTGAAA
CATCCATCCTCAGGAACTGAGAACTTCTGGAATCTTGGACTTTACTTCCT
CTCCAGCTGTTGTGGAATAAGTTCAACTCCAGACTGTAAGTCATTCTAAT
AAATCCCTTTCCAGGAGAGAGAGAGTGAGGTATCATTATATTTAATATTT
TACTCTCTACACCAGGCAGCCCAAGTAGATTTGATGGTCCATTAGCACTT
TCCCCATTGCCTGTTTAGTCAGCTGCCTATAATGTTTAACTGTAGACAAT
TGCATCCAGAGTGATAACATCTTACTTGGAGAAGCTTCACTCCTTTTTGT
TTACTGGTTTTGATTTTTGAGCCTAGGTTTTCCTGTGTAGCACTACCTGA
TCTGAAACTAGGTCTGACAACCAGGCTGGCCTTGAGATCAGAGACCCATC
TGCCTCTGCTTCTGTGTAGGATCTGCTCTTTCCCTTAGGTTCTGGGATGA
AAGGATTTTGCCACCAGATTGGGCAGAATCTTGTCTCCTTAGGCAATGCA
AATAAGATTCTAGGTTTGTAGGTTTCAAACCCAGCTTCCCACAGTGTAAT
GGTTGGGAAGGGTCAAGAAATATAGAGCACAATTAACTTGATTCTTCTCT
GAATCAGGAGGTGGAGGATTTACCTTCAGGGATCCATGGATAAAGCCAAG
AAGATGATGCAGTCCATTCCCAGTTTTGTCAAGGATACATCAGATATTGA
AGAACATGCACTGCCCAGTGCACAGGTCTTGCCAGCCCAGAGTACAAGGT
GTTCCAATTCTGAGACACTTTGTTTCAGCAAAGAGCAAAGCCACTGCTCT
GAGGATGGCTGGATTGCCAATTGGGATCTATACTCCTTTTGTGTATTTGA
GAGTGTGGACTACCTGAAATCCTACCGCAGATTGAATTCTGCCATGAAGA
AGGGCACAGAGGTCTTCCAGAGTGAGAGTCAGAGGGAGCCACAAGTGTCC

CCAGGAGATGTGGAAAACTACAAAGACAAAGATACAGAGGAGCCAGACCA
ACCCTCACTAAGCTTGCTCAGGGAGAAAGGGCTGGAACTTGTGACCTGTG
ATGGTGGAGACTGCCCTGACCAGGATCCTGCATCTTATAGTGCCAGGCAC
CTAGGCTGCTGGGCATGGCTTCAAAGAGCTTTTCGCCAGAAGTGAGAAAG
TCACCCAGAACTGTTTGGATCCCAGATTCCTGCTAAGACTTGCAATTAGG
GGATCTTCTGTCAGCTCCTGCTGGTACAGCAAAGGCACACAAAGGCAGTT
GTGTCTTTTCAGCCATCTGGTTTGTGTTTGTTTGTTTGTTTATTTGTTTG
CAGCTTTCTTAATAAAATTGTTAAAAAGCTGAAAATTTGTGTTCTTCTTA
AAGTGGCCCAAGGGAAAAGATAAAGGGTGGAGTGGACACAGGTGAGGATA
TTCTATTTTTCCTATTGTTGGTTACAGGATAACTGGTGGCTTGTTTTAAA
GGAAGTCCTTAGGGGATGTTAGGCCTCCTTGTTCTCCTTTGTGCTGGGCC
ATAAGATAAACACACAGGGAGCAACTTTGAGCACTTTGGGAAACCTTCCT
GCTGAATCACCATGGATGATGACTGATCTACACAGATGCCTCCTTTCCAA
AAAGATCTCGGGGACACCTGCAACTTCCATTTTTACTCTTTTACCTTATG
CATAGACTAGGCTAGCTCTGAACTCTGATTCACTTTGCCTTGCTTCTCTA
TTATAGGGATTAGAGGCCTGTGCCAGCACATCTGCACAGGAGAATTTTTC
TATTCTCAGCAATTTCCAGAATAAATTTCTACATCATATGTGAACTCACC
CCTAGAAACAGGACTGGCTAAGTTAGAAGGTCAGGGCAGTGCTGGTCCTG
ACAAAGAAAACTCAGCCAACTTGAGAGGGTGTCTTTTGCTTCCAGTCTTG
GAAGAGCTTAACATTTTATCAGACATTAGTTCAGCTGTTGGTCTTCCAGA
AGTCTAGCCTTTCAGTAGTAAATGGATCTGAGGTAGGCATGTTCACTCAG
GCAGGTGGATCCCTGACAGTTTAGGGCAGCCTGATTAACACAGGGAATTC
TAGGACAGACAGCGTTATGTAGTGAGAGGCACAAATAAAAGAAAGAAGGG
AAGGAGGGTGGAAGGCAGAAAGGAAGGAAGTAATAAAGGGAGCGTGTGAA
GGAGGAAGAGAGAAGGAGGGAGAGAGGGAGGAAAGAGGGAGGACTGAGTG
ATGGACAGGTCTGAATGACAGGCCTCACTGCAGTTGTGGAGAATAAATGG
AAGCAGCTAGTAAATCAGGACTCAGGAACTTAGAGAGTCTGTTATGATCT

Table 1. Differentially expressed proteins in quantitative proteomics. Top 100 genes which are enriched in post-ALT cells compared to pre-ALT cells.

	Gene name	log_2(fold change)	q value	Description
1	Tmem35a	-3.09	6.95E-05	Transmembrane protein 35A
2	Hmgn1	-3.01	0.0157	Non-histone chromosomal protein HMG-14
3	Klf4	-2.83	0.000373	Krueppel-like factor 4
4	Prdm16	-2.8	0.00142	PR domain zinc finger protein 16
5	Jdp2	-2.77	0.000264	Jun dimerization protein 2
6	Gli2	-2.66	0.00847	Zinc finger protein GLI2
7	Sall3	-2.51	2.37E-05	Sal-like protein 3
8	Bhmt	-2.42	0.000494	Betaine--homocysteine S-methyltransferase 1
9	Sema5b	-2.3	0.0017	Semaphorin-5B
10	Tfcp2l1	-2.29	0.0013	Transcription factor CP2-like protein 1
11	Esrrb	-2.25	0.000115	Estrogen related receptor, beta, isoform CRA_b
12	Tjp3	-2.23	0.0181	Tight junction protein ZO-3
13	Nanog	-2.19	0.00104	MCG132219
14	Mitf	-2.15	0.00806	Microphthalmia-associated transcription factor
15	Gm14444	-2.15	0.0402	Uncharacterized protein
16	Zfp839	-2.12	0.0105	Zfp839 protein
17	Hmgn2	-2.12	0.0352	High mobility group nucleosomal binding domain 2
18	Tet2	-2.11	0.000312	Tet oncogene 2

19	Rgs18	-2.11	0.00299	Putative uncharacterized protein
20	Atrx	-2.08	0.00718	Transcriptional regulator ATRX (Fragment)
21	Stag3	-2.07	0.00018	Cohesin subunit SA-3
22	Scml2	-2.04	0.00784	Sex comb on midleg-like 2 (Drosophila)
23	Gm14443	-2.03	0.0461	Uncharacterized protein
24	Zfp229	-1.97	0.000333	Zinc finger protein 229
25	Hes1	-1.95	0.0057	Putative uncharacterized protein
26	Oas2	-1.95	0.0228	2'-5'-oligoadenylate synthase 2
27	Zmym3	-1.94	0.000159	Zinc finger MYM-type protein 3
28	Mta1	-1.9	0.0383	Metastasis-associated protein MTA1 isoform 11
29	Rbmxl2	-1.89	0.00592	Putative uncharacterized protein
30	Satb1	-1.81	5.59E-05	DNA-binding protein SATB
31	Enox2	-1.81	0.0013	Putative uncharacterized protein
32	Zeb1	-1.8	0.00307	Zinc finger E-box binding homeobox 1
33	Znf728	-1.8	0.0204	Zinc finger protein 728
34	Brwd1	-1.77	0.00101	Bromodomain and WD repeat-containing protein 1 (Fragment)
35	Pds5b	-1.77	0.0027	Sister chromatid cohesion protein PDS5 homolog B
36	Cmas	-1.77	0.0138	Cytidine monophospho-N-acetylneuraminic acid synthetase
37	Sp4	-1.77	0.0262	Transcription factor Sp4
38	Zfp936	-1.76	0.0037	LOC100045488 protein
39	mCG_120563	-1.76	0.0349	MCG120563, isoform CRA_a
40	H1f0	-1.75	0.0374	Histone H1.0
41	Syce1	-1.74	0.00127	Synaptonemal complex central element protein 1
42	Tbx3	-1.73	6.38E-05	T-box transcription factor TBX3
43	Znf704	-1.73	0.0177	Zinc finger protein 704

44	Zfp990	-1.73	0.0296	Zinc finger protein 990
45	Ap2a1	-1.72	2.99E-05	AP-2 complex subunit alpha-1 (Fragment)
46	Zfp518b	-1.72	0.00853	MCG141317
47	Acsm1	-1.72	0.0449	Acyl-coenzyme A synthetase ACSM1, mitochondrial
48	Gm14325	-1.7	0.0134	Uncharacterized protein
49	A630089N07Rik	-1.69	7.00E-04	MCG4790, isoform CRA_e
50	Nfic	-1.69	0.0068	Nuclear factor 1 C-type
51	Brwd1	-1.67	7.11E-05	Putative uncharacterized protein (Fragment)
52	Arxes1	-1.67	0.00176	Adipocyte-related X-chromosome expressed sequence 1
53	Caap1	-1.67	0.00249	Caspase activity and apoptosis inhibitor 1
54	Nr0b1	-1.67	0.01	Adrenal hypoplasia congenita-like protein
55	Cdca7l	-1.66	2.21E-05	Cell division cycle-associated 7-like protein
56	D1Pas1	-1.65	0.000895	Putative ATP-dependent RNA helicase Pl10
57	Aff1	-1.64	0.00222	AF4/FMR2 family member 1
58	Zfp748	-1.64	0.015	MKIAA4236 protein (Fragment)
59	Olfr639	-1.64	0.03	Olfactory receptor
60	Zfhx2	-1.63	0.000276	ZFH-5
61	Tcl1	-1.63	0.00918	Putative uncharacterized protein
62	Rfx2	-1.63	0.0163	DNA-binding protein RFX2
63	Tmpo	-1.62	0.0389	Putative uncharacterized protein
64	Erf	-1.6	0.000124	ETS domain-containing transcription factor ERF
65	Trerf1	-1.6	0.00212	Transcriptional-regulating factor 1

66	Hnrnph2	-1.59	4.71E-05	Heterogeneous nuclear ribonucleoprotein H2
67	Nol4l	-1.57	5.81E-06	Nucleolar protein 4-like
68	Sap25	-1.57	0.000873	Histone deacetylase complex subunit SAP25
69	Ogfod3	-1.56	0.000661	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 3
70	Nwd1	-1.56	0.00193	NACHT domain- and WD repeat-containing protein 1
71	Zbtb22	-1.55	0.00536	Putative uncharacterized protein
72	Vrk1	-1.55	0.00668	Serine/threonine-protein kinase VRK1
73	Zfp971	-1.55	0.0279	Zinc finger protein 971
74	Pot1b	-1.54	0.00303	Protection of telomeres 1B
75	Tigd2	-1.52	0.00869	Tigger transposable element-derived protein 2
76	Rhox10	-1.52	0.0159	Putative uncharacterized protein
77	Rpl10l	-1.52	0.018	60S ribosomal protein L10-like
78	Trip12	-1.51	0.00484	E3 ubiquitin-protein ligase TRIP12
79	Mki67	-1.51	0.00694	Proliferation marker protein Ki-67
80	Mpnd	-1.51	0.00953	MPN domain-containing protein
81	Hmga1	-1.51	0.0196	High mobility group protein HMG-I/HMG-Y
82	Zdhhc4	-1.5	0.000109	Probable palmitoyltransferase ZDHHC4
83	Rbpj	-1.5	0.0192	Recombining binding protein suppressor of hairless
84	Zscan21	-1.49	0.00268	Zinc finger and SCAN domain-containing protein 21
85	Nova2	-1.49	0.0166	Neuro-oncological ventral antigen 2
86	Chd2	-1.48	0.00141	Chromodomain-helicase-DNA-binding protein 2

87	Acp6	-1.48	0.00769	Lysophosphatidic acid phosphatase type 6
88	N4bp2l2	-1.47	0.00273	NEDD4-binding protein 2-like 2
89	Zic5	-1.46	0.00359	Zinc finger protein ZIC 5
90	Mafg	-1.46	0.00685	Transcription factor MafG
91	Cep290	-1.46	0.0479	Centrosomal protein of 290 kDa
92	Terf2ip	-1.45	0.000249	Telomeric repeat-binding factor 2-interacting protein 1
93	Hsf2bp	-1.45	0.00169	Heat shock factor 2-binding protein
94	Zbtb34	-1.45	0.0194	MKIAA1993 protein (Fragment)
95	AU022751	-1.45	0.0214	Expressed sequence AU022751
96	BC022960	-1.44	0.000661	BC022960 protein (Fragment)
97	Lrif1	-1.43	6.62E-05	Ligand-dependent nuclear receptor-interacting factor 1
98	Chpt1	-1.43	0.000603	Cholinephosphotransferase 1
99	Zfp428	-1.43	0.00523	Zinc finger protein 428
100	Dpf1	-1.42	0.000739	Zinc finger protein neuro-d4

Table 2. Differentially expressed genes in RNA-seq. Top 100 genes which are enriched in post-ALT cells compared to pre-ALT cells.

	Gene name	log_2(fold change)	q value	Locus
1	1700108F19Rik	5.75959	0.000427	chr14:76672962-76687169
2	Prmt8	4.94469	0.000427	chr6:127689010-127769472
3	Tcstv1	4.8976	0.000427	chr13:119893387-119894785
4	Tcstv3	4.75281	0.000427	chr13:120317566-120318076
5	Stra8	4.47472	0.000427	chr6:34920162-34939344
6	Gm14133	4.10168	0.017616	chr2:149743174-149798743
7	Crtac1	4.03185	0.000427	chr19:42283036-42431783
8	Msc	3.99428	0.000427	chr1:14752936-14776931
9	Npb	3.96925	0.006641	chr11:120608476-120609093
10	A2m	3.83397	0.000427	chr6:121635375-121679227
11	Tmem35a	3.81341	0.000427	chrX:134295224-134305969
12	Sohlh2	3.81108	0.000427	chr3:55182027-55209957
13	Sall3	3.71642	0.000427	chr18:80966375-80988575
14	Scn3a	3.70232	0.000427	chr2:65457117-65567627
15	2810410L24Rik	3.69186	0.000427	chr11:120185399-120189852
16	Dock10	3.68328	0.000427	chr1:80501072-80758527
17	AI593442	3.68041	0.017844	chr9:52673043-52679429
18	Bhmt	3.67054	0.000427	chr13:93616890-93637961
19	Pak7	3.62722	0.041662	chr2:136081103-136387967
20	Ntrk3	3.60077	0.000427	chr7:78175958-78738012
21	2210417A02Rik	3.53888	0.018679	chr5:148741839-148743139
22	Gsc	3.53829	0.006641	chr12:104471208-104473330
23	Igsf21	3.52229	0.000427	chr4:140026845-140246784
24	Nlrp9c	3.5054	0.013018	chr7:26322472-26403700

25	Fgf1	3.47978	0.000427	chr18:38601533-39389425
26	St8sia1	3.47039	0.000427	chr6:142814230-142964452
27	Ccdc38	3.4344	0.001831	chr10:93540631-93605245
28	Acot3	3.38582	0.000427	chr12:84038378-84059565
29	Gm20767	3.31182	0.000427	chr13:120154626-120155136
30	Alox8	3.23747	0.000427	chr11:69183931-69197843
31	Scn9a	3.19513	0.000427	chr2:66440879-66634962
32	Ssu2	3.18849	0.037654	chr6:112359323-112388023
33	Sema5b	3.18238	0.000427	chr16:35541144-35664732
34	Syndig1	3.13887	0.000427	chr2:149829210-150004392
35	A530083M17Rik	3.13838	0.000427	chr3:153198265-153725174
36	6330410L21Rik	3.12285	0.001161	chr3:129677564-129763884
37	Grin1	3.11761	0.000427	chr2:25291180-25319187
38	Cfh	3.05083	0.000427	chr1:140084707-140183764
39	Gm37593	3.0494	0.011159	chr17:51842860-51846114
40	Gm12031	3.04939	0.000427	chr11:20145566-20146344
41	Halr1	3.0403	0.000427	chr6:52102948-52113684
42	Fam81a	3.02747	0.000427	chr9:70089309-70142560
43	Hal	3.00324	0.000427	chr10:93488767-93519304
44	Tmtc2	2.99963	0.000427	chr10:105187662-105583874
45	Gm805	2.98576	0.000427	chr12:86169340-86195102
46	Plxna4	2.98245	0.002464	chr6:32144267-32588192
47	Nlrp4a	2.97921	0.000427	chr7:26435112-26476142
48	Lrtm2	2.93745	0.01528	chr6:119236525-119352407
49	Nxf2	2.91892	0.016371	chrX:134944525-134964754
50	Gm44015	2.90543	0.000427	chr7:6651747-6662368
51	Syce1	2.90482	0.000427	chr7:140777228-140787852
52	Chst8	2.89419	0.000427	chr7:34674467-34812711
53	Atp13a4	2.88826	0.014162	chr16:29395852-29544864
54	Klf4	2.87258	0.000427	chr4:55527142-55532466
55	Mybpc1	2.86612	0.002464	chr10:88518278-88605152

56	Sntg1	2.8592	0.000427	chr1:8361474-9299878
57	Notumos	2.84541	0.000427	chr11:120646030-120661815
58	Gm13580	2.82974	0.0015	chr2:60411524-60412579
59	Olfr924	2.7901	0.037654	chr9:38848115-38849042
60	Pabpc6	2.76571	0.000427	chr17:9666496-9669704
61	Spic	2.74223	0.000427	chr10:88674771-88685015
62	2310003L06Rik	2.7342	0.019758	chr5:87908578-88008534
63	Ak7	2.71902	0.000427	chr12:105705981-105782447
64	St6galnac3	2.70686	0.000427	chr3:153198265-153725174
65	Vmn1r53	2.70651	0.006391	chr6:90131440-90232817
66	Sytl4	2.67809	0.000427	chrX:133936384-133981812
67	Gm14410	2.64093	0.000427	chr2:177183452-177206421
68	4933425D22Rik	2.62352	0.045747	chr5:149439659-149470979
69	Syt1	2.61551	0.005895	chr10:108497649-109010982
70	Tmem156	2.60239	0.011583	chr5:65054345-65092132
71	Arxes2	2.60189	0.000427	chrX:135993823-135995355
72	Gm7325	2.60156	0.000427	chr17:45600970-45602102
73	Ldhb	2.58662	0.000427	chr6:142490248-142507957
74	Gpr83	2.5766	0.005624	chr9:14860209-14870789
75	Camk2n1	2.55794	0.000427	chr4:138454313-138460123
76	Ampd1	2.55375	0.039485	chr3:103074013-103099720
77	Lrrn2	2.54153	0.000427	chr1:132880272-132940005
78	Lonrf2	2.53185	0.002152	chr1:38793644-38836711
79	Prdm14	2.52952	0.000427	chr1:13113456-13127163
80	Sfrp1	2.52593	0.000427	chr8:23411501-23449632
81	Atp1a4	2.52371	0.000427	chr1:172223512-172258414
82	Itpka	2.51405	0.000427	chr2:119742336-119751263
83	Gm14406	2.51137	0.000427	chr2:177569203-177578210
84	Platr11	2.51122	0.000808	chr9:118404096-118406099
85	Slc47a1	2.49784	0.000427	chr11:61343400-61378345
86	Fam217a	2.49395	0.005068	chr13:34909963-34919992

87	4930405N21Rik	2.48716	0.000427	chr5:113320453-113321653
88	Gm11766	2.47802	0.000427	chr11:119935802-119942698
89	Slc6a5	2.46124	0.022059	chr7:49910145-49963856
90	Gm53	2.44801	0.000427	chr11:96251104-96264868
91	Pramel6	2.43438	0.000427	chr2:87506564-87510872
92	Olfr153	2.42069	0.02713	chr2:87532034-87532958
93	Gm14393	2.4184	0.000427	chr2:175061548-175067781
94	Notum	2.40653	0.000427	chr11:120646030-120661815
95	Gm281	2.40327	0.002152	chr14:13814617-13914478
96	Myot	2.40108	0.015727	chr18:44334073-44355724
97	Frem1	2.39107	0.005624	chr4:82897919-83052339
98	Filip1	2.38839	0.011362	chr9:79815050-79977804
99	Lrrc34	2.38828	0.000427	chr3:30624266-30672431
100	B230307C23Rik	2.38297	0.000427	chr16:97994600-98010000

Reference

Apte, M.S., and Cooper, J.P. (2017). Life and cancer without telomerase: ALT and other strategies for making sure ends (don't) meet. Critical reviews in biochemistry and molecular biology *52*, 57-73.

Arora, R., and Azzalin, C.M. (2015). Telomere elongation chooses TERRA ALTernatives. RNA biology *12*, 938-941.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T., and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nature communications *5*, 5220.

Benetti, R., Garcia-Cao, M., and Blasco, M.A. (2007). Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. Nature genetics *39*, 243-250.

Berendes, H.D., and Meyer, G.F. (1968). A specific chromosome element, the telomere of Drosophila polytene chromosomes. Chromosoma *25*, 184-197.

Bhatia, V., Barroso, S.I., García-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature *511*, 362.

Birger, Y., Catez, F., Furusawa, T., Lim, J.H., Prymakowska-Bosak, M., West, K.L., Postnikov, Y.V., Haines, D.C., and Bustin, M. (2005). Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMGN1. Cancer research *65*, 6711-6718.

Birger, Y., West, K.L., Postnikov, Y.V., Lim, J.H., Furusawa, T., Wagner, J.P., Laufer, C.S., Kraemer, K.H., and Bustin, M. (2003). Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. The EMBO journal *22*, 1665-1675.

Bisht, K.K., Arora, S., Ahmed, S., and Singh, J. (2008). Role of heterochromatin in suppressing subtelomeric recombination in fission yeast. Yeast (Chichester, England) *25*, 537-548.

Blackburn, E.H. (1991). Structure and function of telomeres. Nature *350*, 569-573.

Bryan, T.M., Englezou, A., Gupta, J., Bacchetti, S., and Reddel, R.R. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. The EMBO journal *14*, 4240-4248.

Casacuberta, E. (2017). Drosophila: Retrotransposons Making up Telomeres. Viruses *9*.

Cawthon, R.M. (2009). Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic acids research *37*, e21.

Cesare, A.J., Hayashi, M.T., Crabbe, L., and Karlseder, J. (2013). The telomere deprotection response is functionally distinct from the genomic DNA damage response. Molecular cell *51*, 141-155.

Cesare, A.J., Heaphy, C.M., and O'Sullivan, R.J. (2015). Visualization of Telomere Integrity and Function In Vitro and In Vivo Using Immunofluorescence Techniques. Current protocols in cytometry *73*, 12.40.11-31.

Chen, Q., Ijpma, A., and Greider, C.W. (2001). Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. Molecular and cellular biology *21*, 1819-1827.

Cheung, A.L., and Deng, W. (2008). Telomere dysfunction, genome instability and cancer. Frontiers in bioscience : a journal and virtual library *13*, 2075-2090.

Cho, N.W., Dilley, R.L., Lampson, M.A., and Greenberg, R.A. (2014). Interchromosomal homology searches drive directional ALT telomere movement and synapsis. Cell *159*, 108-121.

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

Clynes, D., Jelinska, C., Xella, B., Ayyub, H., Scott, C., Mitson, M., Taylor, S., Higgs, D.R., and Gibbons, R.J. (2015). Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. Nature communications *6*, 7538.

Conomos, D., Reddel, R.R., and Pickett, H.A. (2014). NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. Nat Struct Mol Biol *21*, 760-770.

Conomos, D., Stutz, M.D., Hills, M., Neumann, A.A., Bryan, T.M., Reddel, R.R., and Pickett, H.A. (2012). Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. The Journal of cell biology *199*, 893-906.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., *et al.* (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nature methods *14*, 959-962.

Costantino, L., Sotiriou, S.K., Rantala, J.K., Magin, S., Mladenov, E., Helleday, T., Haber, J.E., Iliakis, G., Kallioniemi, O.P., and Halazonetis, T.D. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. Science *343*, 88-91.

Cox, K.E., Marechal, A., and Flynn, R.L. (2016). SMARCAL1 Resolves Replication Stress at ALT Telomeres. Cell reports *14*, 1032-1040.

Cubiles, M.D., Barroso, S., Vaquero-Sedas, M.I., Enguix, A., Aguilera, A., and Vega-Palas, M.A. (2018). Epigenetic features of human telomeres. Nucleic acids research *46*, 2347-2355.

Cusanelli, E., and Chartrand, P. (2015). Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. Frontiers in genetics *6*, 143-143.

de Lange, T. (1998). Telomeres and senescence: ending the debate. Science *279*, 334-335.

de Lange, T. (2004). T-loops and the origin of telomeres. Nature reviews Molecular cell biology *5*, 323-329.

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes & development *19*, 2100-2110.

de Lange, T. (2010). How shelterin solves the telomere end-protection problem. Cold Spring Harb Symp Quant Biol *75*, 167-177.

de Lange, T. (2018). Shelterin-Mediated Telomere Protection. Annual review of genetics.

Dejardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic Loci. Cell *136*, 175-186.

Dilley, R.L., Verma, P., Cho, N.W., Winters, H.D., Wondisford, A.R., and Greenberg, R.A. (2016). Break-induced telomere synthesis underlies alternative telomere maintenance. Nature *539*, 54-58.

Ding, Z., Mangino, M., Aviv, A., Spector, T., Durbin, R., and Consortium, U.K. (2014). Estimating telomere length from whole genome sequence data. Nucleic acids research *42*, e75-e75.

Donnianni, R.A., and Symington, L.S. (2013). Break-induced replication occurs by conservative DNA synthesis. Proceedings of the National Academy of Sciences of the United States of America *110*, 13475-13480.

Draskovic, I., Arnoult, N., Steiner, V., Bacchetti, S., Lomonte, P., and Londono-Vallejo, A. (2009). Probing PML body function in ALT cells reveals spatiotemporal requirements for telomere recombination. Proceedings of the National Academy of Sciences of the United States of America *106*, 15726-15731.

Duan, Y.M., Zhou, B.O., Peng, J., Tong, X.J., Zhang, Q.D., and Zhou, J.Q. (2016). Molecular dynamics of de novo telomere heterochromatin formation in budding yeast. Journal of genetics and genomics = Yi chuan xue bao *43*, 451-465.

Dunham, M.A., Neumann, A.A., Fasching, C.L., and Reddel, R.R. (2000). Telomere maintenance by recombination in human cells. Nature genetics *26*, 447-450.

Episkopou, H., Draskovic, I., Van Beneden, A., Tilman, G., Mattiussi, M., Gobin, M., Arnoult, N., Londono-Vallejo, A., and Decottignies, A. (2014). Alternative Lengthening of Telomeres is characterized by reduced compaction of telomeric chromatin. Nucleic acids research *42*, 4391-4405.

Feuerbach, L., Sieverling, L., Deeg, K., Ginsbach, P., Hutter, B., Buchhalter, I., Northcott, P.A., Lichter, P., Pfister, S.M., Jones, D.T.W., *et al.* (2016). TelomereHunter: telomere content estimation and characterization from whole genome sequencing data. bioRxiv.

Gerlitz, G. (2010). HMGNs, DNA repair and cancer. Biochimica et biophysica acta *1799*, 80-85.

Geronimo, C.L., and Zakian, V.A. (2016). Getting it done at the ends: Pif1 family DNA helicases and telomeres. DNA repair *44*, 151-158.

Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira, M.T., and Luke, B. (2017). Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. Cell *170*, 72-85.e14.

Greenberg, R.A. (2005). Telomeres, crisis and cancer. Current molecular medicine *5*, 213-218.

Greider, C.W. (1993). Telomerase and telomere-length regulation: lessons from small eukaryotes to mammals. Cold Spring Harb Symp Quant Biol *58*, 719-723.

Harley, C.B., and Sherwood, S.W. (1997). Telomerase, checkpoints and cancer. Cancer surveys *29*, 263-284.

Heaphy, C.M., Subhawong, A.P., Hong, S.M., Goggins, M.G., Montgomery, E.A., Gabrielson, E., Netto, G.J., Epstein, J.I., Lotan, T.L., Westra, W.H., *et al.* (2011). Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. The American journal of pathology *179*, 1608-1615.

Henson, J.D., Cao, Y., Huschtscha, L.I., Chang, A.C., Au, A.Y., Pickett, H.A., and Reddel, R.R. (2009). DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. Nature biotechnology *27*, 1181-1185.

Herrera, J.E., West, K.L., Schiltz, R.L., Nakatani, Y., and Bustin, M. (2000). Histone H1 is a specific repressor of core histone acetylation in chromatin. Molecular and cellular biology *20*, 523-529.

Jain, D., Hebden, A.K., Nakamura, T.M., Miller, K.M., and Cooper, J.P. (2010). HAATI survivors replace canonical telomeres with blocks of generic heterochromatin. Nature *467*, 223-227.

Jasin, M., and Rothstein, R. (2013). Repair of strand breaks by homologous recombination. Cold Spring Harbor perspectives in biology *5*, a012740.

Kaul, Z., Cesare, A.J., Huschtscha, L.I., Neumann, A.A., and Reddel, R.R. (2011). Five dysfunctional telomeres predict onset of senescence in human cells. EMBO reports *13*, 52-59.

Kugler, J.E., Deng, T., and Bustin, M. (2012). The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes. Biochimica et biophysica acta *1819*, 652-656.

Kupiec, M. (2014). Biology of telomeres: lessons from budding yeast. FEMS microbiology reviews *38*, 144-171.

Li, X., and Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. Cell Research *18*, 99.

Londono-Vallejo, J.A., Der-Sarkissian, H., Cazes, L., Bacchetti, S., and Reddel, R.R. (2004). Alternative lengthening of telomeres is characterized by high rates of telomeric exchange. Cancer research *64*, 2324-2327.

Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J.H., Sung, P.A., Jasin, M., *et al.* (2012). Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS genetics *8*, e1002772.

Lundblad, V., and Blackburn, E.H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. Cell *73*, 347-360.

McEachern, M.J., and Haber, J.E. (2006). Break-induced replication and recombinational telomere elongation in yeast. Annual review of biochemistry *75*, 111-135.

Muntoni, A., Neumann, A.A., Hills, M., and Reddel, R.R. (2009). Telomere elongation involves intra-molecular DNA replication in cells utilizing alternative lengthening of telomeres. Human Molecular Genetics *18*, 1017-1027.

Nakamura, T.M., Cooper, J.P., and Cech, T.R. (1998). Two modes of survival of fission yeast without telomerase. Science *282*, 493-496.

Nersisyan, L., and Arakelyan, A. (2015). Computel: Computation of Mean Telomere Length from Whole-Genome Next-Generation Sequencing Data. PLOS ONE *10*, e0125201.

Niida, H., Matsumoto, T., Satoh, H., Shiwa, M., Tokutake, Y., Furuichi, Y., and Shinkai, Y. (1998). Severe growth defect in mouse cells lacking the telomerase RNA component. Nature genetics *19*, 203-206.

Niida, H., Shinkai, Y., Hande, M.P., Matsumoto, T., Takehara, S., Tachibana, M., Oshimura, M., Lansdorp, P.M., and Furuichi, Y. (2000). Telomere maintenance in telomerase-deficient mouse embryonic stem cells: characterization of an amplified telomeric DNA. Molecular and cellular biology *20*, 4115-4127.

Oliva-Rico, D., and Herrera, L.A. (2017). Regulated expression of the lncRNA TERRA and its impact on telomere biology. Mechanisms of ageing and development *167*, 16-23.

Pan, X., Drosopoulos, W.C., Sethi, L., Madireddy, A., Schildkraut, C.L., and Zhang, D. (2017). FANCM, BRCA1, and BLM cooperatively resolve the replication stress at the ALT telomeres. Proceedings of the National Academy of Sciences of the United States of America *114*, E5940-e5949.

Postnikov, Y.V., Kurahashi, T., Zhou, M., and Bustin, M. (2012). The nucleosome binding protein

HMGN1 interacts with PCNA and facilitates its binding to chromatin. Molecular and cellular biology *32*, 1844-1854.

Reddel, R.R., Bryan, T.M., and Murnane, J.P. (1997). Immortalized cells with no detectable telomerase activity. A review. Biochemistry Biokhimiia *62*, 1254-1262.

Riethman, H., Ambrosini, A., Castaneda, C., Finklestein, J.M., Hu, X.L., Paul, S., and Wei, J. (2003). Human subtelomeric DNA. Cold Spring Harb Symp Quant Biol *68*, 39-47.

Riethman, H., Ambrosini, A., and Paul, S. (2005). Human subtelomere structure and variation. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *13*, 505-515.

Robin, J.D., Ludlow, A.T., Batten, K., Magdinier, F., Stadler, G., Wagner, K.R., Shay, J.W., and Wright, W.E. (2014). Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. Genes & development *28*, 2464-2476.

Schoeftner, S., and Blasco, M.A. (2009a). A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. The EMBO journal *28*, 2323-2336.

Schoeftner, S., and Blasco, M.A. (2009b). A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. The EMBO journal *28*, 2323-2336.

Sen, S., Dasgupta, A., Roychudhuri, A., Mittra, B., and Majumder, H.K. (1999). Telomere, telomerase, tumorigenesis and therapy: an overview. Indian journal of experimental biology *37*, 839-842.

Seo, B., Kim, C., Hills, M., Sung, S., Kim, H., Kim, E., Lim, D.S., Oh, H.S., Choi, R.M., Chun, J., *et al.* (2015). Telomere maintenance through recruitment of internal genomic regions. Nature communications *6*, 8189.

Sfeir, A., and de Lange, T. (2012). Removal of shelterin reveals the telomere end-protection problem. Science *336*, 593-597.

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

Shay, J.W., and Wright, W.E. (2005). Senescence and immortalization: role of telomeres and telomerase. Carcinogenesis *26*, 867-874.

Silva-Sousa, R., Lopez-Panads, E., and Casacuberta, E. (2012). Drosophila telomeres: an example of co-evolution with transposable elements. Genome dynamics *7*, 46-67.

Sobinoff, A.P., Allen, J.A., Neumann, A.A., Yang, S.F., Walsh, M.E., Henson, J.D., Reddel, R.R.,

and Pickett, H.A. (2017). BLM and SLX4 play opposing roles in recombination-dependent replication at human telomeres. The EMBO journal *36*, 2907-2919.

Sobinoff, A.P., and Pickett, H.A. (2017). Alternative Lengthening of Telomeres: DNA Repair Pathways Converge. Trends in genetics : TIG *33*, 921-932.

Sung, P., and Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nature reviews Molecular cell biology *7*, 739-750.

Tashiro, S., Nishihara, Y., Kugou, K., Ohta, K., and Kanoh, J. (2017). Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis. Nucleic acids research *45*, 10333-10349.

Toubiana, S., and Selig, S. (2018). DNA:RNA hybrids at telomeres - when it is better to be out of the (R) loop. The FEBS journal *285*, 2552-2566.

van Emden, T.S., Forn, M., Forne, I., Sarkadi, Z., Capella, M., Martin Caballero, L., Fischer-Burkart, S., Bronner, C., Simonetta, M., Toczyski, D., *et al.* (2018). Shelterin and subtelomeric DNA sequences control nucleosome maintenance and genome stability. EMBO reports.

Varley, H., Pickett, H.A., Foxon, J.L., Reddel, R.R., and Royle, N.J. (2002). Molecular characterization of inter-telomere and intra-telomere mutations in human ALT cells. Nature genetics *30*, 301-305.

Voon, H.P.J., Collas, P., and Wong, L.H. (2016). Compromised Telomeric Heterochromatin Promotes ALTernative Lengthening of Telomeres. Trends in cancer *2*, 114-116.

Wang, C., Zhao, L., and Lu, S. (2015). Role of TERRA in the regulation of telomere length. International journal of biological sciences *11*, 316-323.

Wang, Y., Xu, Q., Sack, L., Kang, C., and Elledge, S.J. (2016). A gain-of-function senescence bypass screen identifies the homeobox transcription factor DLX2 as a regulator of ATM-p53 signaling. Genes & development *30*, 293-306.

Zalzman, M., Falco, G., Sharova, L.V., Nishiyama, A., Thomas, M., Lee, S.-L., Stagg, C.A., Hoang, H.G., Yang, H.-T., Indig, F.E., *et al.* (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature *464*, 858.

국문 초록

고유의 템플릿을 이용하는 대안적인 텔로미어 유지 기전에 대한 연구

성상현

서울대학교 생명과학부

진핵 생물이 유전체의 정보를 안정적으로 유지하기 위해서는 선형 염색체에만 존재하는 특이적인 문제를 해결해야만 한다. DNA 복제 기구의 생화학적 한계에 기인하는 말단-복제의 문제와 유전체 내부의 DNA 손상과 말단을 구분해야 하는 말단-보호의 문제이다. 두 가지 문제를 제대로 해결하지 못하면, 유전체의 정보는 점차 소실되거나 잘못된 처리에 의해 손상될 수 있다. 두 가지 말단의 문제를 해결하기 위한 특별한 구조체가 텔로미어이다. 텔로미어는 주로 특정한 반복서열과 그에 대한 결합 단백질로 구성되어 있다. 텔로미어의 길이와 구조를 잘 유지하는 것은 세포의 삶과 죽음을 결정하는데 매우 중요하다. 세포가 분열할 때마다 텔로미어는 내부의 유전 정보를 대신하여 소실되고 특정한 한계 길이에 가까워지면 세포의 분열을 정지시키거나 세포를 죽게 하는 타이머의 역할을 한다. 이 한계를 넘어서 세포가 계속 분열하게 되면, 유전체의 불안정성이 과도하게 증가하여 정상적인 조절이 이루어지지 않는 암세포로 발전할 가능성이 높아진다. 다르게 말하면, 암세포가 텔로미어에 의해 부과된 한계를 넘기

위해서는 반드시 텔로미어의 길이를 유지하는 기능을 획득하는 것이 필요한 것이다. 대부분의 암세포는 텔로머레이즈라고 하는 역전사 효소를 활성화시킴으로써 텔로미어의 길이를 유지하고 무한정 증식할 수 있는 능력을 얻는다. 그러나 일부의 암세포에서 텔로머레이즈 없이도 텔로미어 길이를 유지하는 현상이 관찰되었는데, 이를 대안적인 텔로미어 길이 유지 기전 (Alternative lengthening of telomere, ALT) 이라고 한다. 암세포뿐만 아니라 효모나 선충 등의 모델에서 텔로머레이즈 없이도 텔로미어를 유지하는 현상이 관찰되면서, ALT는 자연 상에 널리 존재할 수 있다는 것이 알려졌다. ALT는 상동 재조합에 의존하는 DNA 복제 기전이라는 것이 알려졌지만, 구체적인 기전의 작동 방식에 관해서는 연구되지 않은 바가 많다. 본 연구에서는 쥐 배아줄기세포 (mouse embryonic stem cell, mESC) 모델을 이용해서 ALT의 특징과 분자적 기전을 규명하고자 했다. 전유전체 서열분석을 통해 ALT mESC의 텔로미어가 가지는 독특한 서열과 구조를 밝혔다. 가장 중요한 특징은 서브텔로미어에 존재하는 특이한 주형 서열이 텔로미어를 구성하고 있다는 점이다. 일반적인 텔로미어 반복 서열이 아니라 다른 서열이 많이 끼어든 만큼, ALT mESC의 텔로미어는 독특한 보호 기전을 유지하고 있을 것이라고 추측 가능하다. 이를 확인하기 위해 전사체 분석과 정량적 단백질체 분석을 수행했으며, 그 결과 ALT의 활성화에 뒤따르는 유전자 발현 변화 양상을 관찰할 수 있었다. 대표적으로 전사 조절 네트워크와 염색질 재구성 과정에 변화가 집중되어 있었다. ALT의 기전과 관련되어 있을 것으로 추정되는 후보 유전자 중에서 ALT 특이적으로 가장 많이 증가한, 비일반적인 히스톤인 HMGN1의 기능에 초점을 맞추었다. HMGN1은 뉴클레오좀을 구성하는 연결 히스톤인 H1 히스톤과 경쟁하고 다른 히스톤들의 번역 후 변형에 관여하여 염색질의 구조를 느슨하게

만들 수 있다고 알려져 있다. HMGN1이 ALT mESC의 독특한 텔로미어에 결합한다는 것을 확인했고, HMGN1의 기능을 저해할 경우 텔로미어의 손상 정도가 증가하고 텔로미어에서 생산되는 전사체의 양이 감소했으며 텔로미어 길이가 짧아졌다. 따라서 HMGN1이 만들어내는 변화가 ALT 텔로미어를 안정적으로 유지하는데 필요한 요소임을 확인했다. 본 연구에서는 ALT가 활성화되기 전후의 쥐 배아줄기세포를 이용하여 ALT 특이적인 분자적 특징을 유전체, 전사체, 단백질체의 관점에서 고찰했다. 그 핵심에는 독특한 주형 서열이 만들어내는 특이한 구조의 텔로미어와 후성유전학적 조절자로 작용하는 HMGN1 단백질이 있다. 특이한 주형 서열을 텔로미어 유지에 사용하는 현상은 효모부터 포유류까지 진화적으로 보존된 현상이다. 쥐 배아줄기세포에서 확립된 ALT 모델이 ALT의 분자적 기전뿐만 아니라 유전체의 잠재적인 진화 기전을 이해하는데 기여할 것으로 기대한다.

주요어: 텔로미어, 쥐 배아줄기세포, 대안적인 텔로미어 유지 기작, 유전체 불안정성, DNA 손상, 염색질 재구성, 텔로미어 서열을 포함하는 RNA (TERRA), HMGN1

학번: 2011-20335