이학박사학위논문

예쁜꼬마선충의 새로운 염색체말단 유지기작에 관한 연구

Studies on the Novel Mechanism
of Telomere Maintenance
in *Caenorhabditis elegans*

2015년 8월

서울대학교 대학원

생명과학부

서 범 석

예쁜꼬마선충의 새로운 염색체말단 유지기작에 관한 연구

Studies on the Novel Mechanism
of Telomere Maintenance
in *Caenorhabditis elegans*

지도교수 이 준 호

이 논문을 이학박사학위논문으로 제출함

2014년 12월

서울대학교 대학원
생명과학부
서 범 석

서범석의 이학박사 학위논문을 인준함

2014년 12월

위 원장 성 노현 (인)

부위원장 이 준호 (인)

위 원 이 현숙 (인)

위 원 구 현숙 (인)

위 원 정 인규 (인)

Studies on the Novel Mechanism of Telomere Maintenance in *Caenorhabditis elegans*

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

To the Faculty of

School of Biological Sciences

at

Seoul National University

by

Beomseok Seo

December, 2014

Date approved

December, 2014

ABSTRACT

Study on novel mechanism of telomere mainatenance
in *Caenorhabditis elegans*

Beomseok Seo
Dept. of Biological Sciences
The Graduate School
Seoul National University

When telomeres lose their ability to protect chromosome ends, cells enter crisis and can only escape death by acquiring the ability to maintain telomere length. The survivors of crisis are often tumorigenic and maintain telomere length commonly through the reactivation of telomerase, but occasionally activate an alternative mechanism called ALT (alternative lengthening of telomeres). Here, I established stably-maintained survivor lines of the nematode *C. elegans* that activate an ALT mechanism to escape from the sterility phenotype caused by telomerase deletion. ALT survivors utilized, depending on their genetic backgrounds, either one of two specific internal genomic regions as templates for telomere lengthening. Either one of these experimentally defined DNA fragments 'Template of ALT' (TALT) element 1 and 2, each consisting of a block of genomic

DNA flanked by telomere-like sequences had already been copied to a proximal telomere region of one chromosome before ALT occurred and was incorporated into most chromosome ends by ALT. The TALT elements that I identified in this study represent a novel mechanism of stabilizing linear chromosomes in the absence of telomerase. Our data have implications in cancer biology, as it is possible that at least some ALT tumors maintain telomeres through a TALT-like mechanism.

Keywords : *C. elegans*, germline immortality, end replication problem, genome integrity, telomere maintenance mechanism, alternative lengthening of telomeres(ALT), recombination, cancer, animal model for ALT, template of ALT, TALT

Student Number: 2007-20346

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Introduction

1. Telomere maintenance mechanism

In the evolution of eukaryote, linear chromosome created normal chromosomal ends that is absent in circular genome. This natural ends must be distinguished from DNA breaks that resulted from stresses such as irradiation, free radicals and etc. If this natural ends are repaired inappropriately, the cell division stops and result in chromosome fusions which is harmful to the cell.

The second problem is so-called end-replication problem, as predicted by Russian scientist Alexei Olovnikov (Olovnikov, 1973). Due to the constraint of DNA replication machinery, the RNA primer at the end of chromosome cannot be substituted with DNA. This eventually leads to gradual shortening of chromosomes. However germ cells or stem cells that are highly proliferative compensate this loss by adding telomere repeats with telomerase reverse transcriptase. Telomerase add telomere repeat sequences to the end of the chromosome with its RNA template,

TERC.

Since telomere maintenance mechanism (TMM) is inactivated in most somatic cells, telomeres get shortened in every cell division. Critically short telomeres cannot be distinguished with naturally occurring DNA double strand breaks (DSBs). The cells with critically short telomeres enter senescence, losing its proliferative potential, known as Hayflick limit (Carrel and Ebeling, 1921). However most cancer cells reactivate telomere maintenance mechanism essential for immortalization. However, a subset of cancers lengthen the telomere by recombination mediated Alternative Lengthening of Telomere mechanism, know as ALT (Bryan et al., 1997).

Telomerase negative yeast undergo telomere shortening resulting in catastrophic cell death. However survivors emerges using amplification of subtelomeric Y' for lengthening telomeres. Y' elements are found at some telomeres but not all. This survivor is known as type I survivor and the survivor formation is dependent upon RAD52 and RAD51. As these genes are required for recombination, it is thought that telomere lengthening of yeast type I survivors occur via recombination

(Lundblad and Blackburn, 1993). Due to ease of genetic manipulation of yeast, many genes required for recombinational telomere lenghthening are identified. However yeast is genetically distant to human than *C. elegans* is to human. Establishing multicellur organism of ALT is required for expanding our knowledge.

In pioneering study of human ALT mechanism, Dr. Reddel's group found that DNA 'tag' sequence introduced in between telomere repeat from exogenous source was copied to other telomere in *trans*. In this study copy of 'tag' introduced into subtelomere was not observed (Dunham et al., 2000). Recently it is reported that in human ALT cells variant repeat sequence recruits NurD-ZNF827 complex to telomeres and drives homologous recombination to lengthen the telomeres (Conomos et al., 2014). However the initial molecular event regarding how the variant repeat is dispersed with critically short telomere is not clear. It is also interesting question if there is preferred structure or location for telomere recombination.

Most ALT positive cells contain ALT-associated promyelocytic leukemia nuclear bodies (APBs) (Fasching et al., 2007). The function of PML bodies is not

clear but they are likely to be involved in senescence and DNA-damage response (Cesare and Reddel, 2010). APBs contain large amount of telomere DNA and proteins that is required for recombination such as MRN (MRE11, RAD50 and NBS1) complex, ATM, ATR and RAD51. But because there are also ALT positive cell line, AG11395, that lacks APBs, it is still unclear that APBs are required for ALT activity (Marciniak et al., 2005).

2. *Caenorhabditis elegans* telomere study

Although ALT mechanism has been studied in yeast and human cancer cells, little is known in multicellular organism such as *Caenorhabditis elegans*. *C. elegans* is a good model system to study ALT mechanism as shown in this study. Its moderate telomere length (5-9kb in wild type) allows telomere shortening in fairly short amount time not to mention its short life cycle (Meier et al., 2006). Telomerase deficient *trt-1(ok410)* mutant reaches sterility in 15 to 19 generations.

Due to the lack of sequence homology to telomere protein of other well-studied organism, a few proteins are identified. HRP-1, PLP-1, CEH-37, HMG-5, POT-1

and POT-2 are telomere binding proteins in *C. elegans* (Im and Lee, 2003, 2005; Joeng et al., 2004; Kim et al., 2003; Raices et al., 2008). However telomere repeat (TTAGGC) is highly conserved differing only 1 nucleotide of human telomere repeat (TTAGGG). Telomere RNA component that contains telomere repeat is not yet identified.

C. elegans genome harbors repetitive sequence, which represent ~17% of the genome (Brenner, 1974). *C. elegans* genome is divided into a repeat-poor but gene-rich "cluster" with low recombination rate, and two repeat-rich but gene-poor "arms" with a 7 fold higher recombination rate (Barnes et al., 1995). This repeat include RcS5, Cerep3 and interstitial telomere repeat, thus there is positive correlation between recombination rate and telomere repeat sequences. But their causal relationship is not known (Cangiano and La Volpe, 1993).

Despite the repetitive sequences, the genome size of *C. elegans* is compact enough to use whole genome sequencing approach, which is complicated in organism with large and complex genome such as human and mice. It is reported that *C. elegans* can survive telomere crisis using ALT mechanism (Cheng et al.,

2012; Lackner et al., 2012; Lowden et al., 2011). POT-2 can bind single stranded telomere overhang and is known to repress ALT in *C. elegans*. trt-1;pot-2 double mutant increase formation of survivor that has normal telomere length. Despite the genetic study of *C. elegans* the molecular mechanism is poorly understood.

3. Divergence of wild isolates of *C. elegans*

Due to free living nature of C. *elegans*, wide variety of wild strains are isolated and studied. Among wild isolate, CB4856 strain found in Hawaii shows extensive divergence with N2 laboratory strain found in Bristol resulting in one polymorphism per 840bp in average (Swan et al., 2002). This allows genetic mapping of mutation by crossing out unrelated mutation irresponsible for phenotype of interest leaving closely linked SNPs in the outcrossed mutant. By identifying SNPs mutation mapping can be accelerated compared to classic genetic mapping.

In addition to the ease of SNPs mapping, polymorphism in functional component of genome is directly reponsible for phenotypic variation between N2 and CB4856

strain. Npr-1 gene is the representative example showing single nucletide change can alter various aspect of behavioral response of worms such as O2 sensing, social behavior and susceptibility to pathogens (Rogers et al., 2003). Association mapping can be used to map SNPs with associated phenotype. It is different from linkage mapping in that it is used to map the quantitative traits that naturally exist.

4. Purposes of this study

The purpose of this study is to identify template for telomere recombination in ALT survivors of *C. elegans*. It is not clear whether there is preference for telomere recombination with specific sequence structure or epigenetic state. I found 2 kinds of TALT that share common structure and variant telomere repeat. This information might be used in predicting TALT in various organisms.

Second purpose is to establish multicelluar model organism for ALT. As ALT mechanism were studied mainly in yeast and human cell line, which can miss ALT regulation in multicellular level. It is worth to investigate whether ALT mechanism can support normal development of multicellular organism. To use as

a model organism, it is important whether the model can be maintained stably or not. Although ALT survivor of *C. elegans* were reported already, the survivors could not be stably maintained. In this study we isolated stably maintained ALT survivors. Moreover, it is crucial to identify the hopspot of template for ALT, since it enables researchers to insert positive selection markers such as GFP protein or antibiotics. TALT we found in this study will enable efficient selection assay for ALT induction.

Materials and Methods

Materials and methods

Strains and Culture. Worms were grown at 20℃ using standard conditions (Brenner, 1974). The following strains were used in this study: Bristol N2 wild strain, Hawaiian CB4856 wild isolate, *trt-1(ok410)* I. N2 *trt-1(ok410)* was outcrossed with Hawaiian CB4856 wild isolate to produce CB4856 trt-1(ok410). To maximize the outcrossing effect, SNPs of all chromosome markers were checked. The trt-1(ok410) mutation was confirmed by PCR. The trt-1(ok410) strain was outcrossed with N2 wild type to produce an early generation of N2 *trt-1(ok410)*.

EMS treatment. Synchronized L4 worms were treated with 50 mM EMS in M9 buffer for 4 hr. After 4 hr of recovery, treated P0s were allowed to lay eggs for 12 hrs. F1 worms were isolated by removing P0 worms. Initially ~100 F1 eggs were transferred to fresh plate, then from the F2 generation onwards, 10~15 worms were transferred manually at every generation.

Feeding RNA interference. *Escherichia coli* HT115 expressing dsRNA were grown in LB with 1 mM ampicillin at 37°C overnight and seeded on to NGM plates containing 1 mM IPTG and 1 mM ampicillin. At every generation, 10-15 L1 larva were transferred to fresh RNAi media plates.

Telomere florescent in situ hybridization (FISH). Eggs were isolated by bleaching adult worms. Eggs were fixed in 2% Paraformaldehyde (PFA) for 15 min in room temperature. Eggs were settled on a slide glass coated with poly-lysine for at least 5 min. The slides were covered with clean slide glass and incubated on a aluminum block precooled on the dry ice. Slides were freeze-cracked and incubated in ice-cold methanol and acetone for 5 min each. The slide was rehydrated and washed 3 times with phosphate-buffered saline containing 0.1% Tween-20 (PBST) to remove residual PFA. The slide was incubated in 2 X SSC (0.3 M NaCl, 0.03 M sodium citrate) containing 0.1% Tween-20. The slide was prehybridized for 1 hour with prehybridization solution (3X SSC, 50% formamide, 10% dextran sulfate, 50 µg/ml heparin, 100 µg/ml yeast tRNA, 100 µg/ml salmon sperm DNA) at 37°C. Cy3-(TTAGGC)3 PNA probe and TALT

probe was hybridized for 16 hr in humid chamber at 37°C after 2.5 min of denature in 85°C heat block. Slides were washed twice in wash buffer (2X SSC and 50 % formamide) for 15 min at 37°C. After washing 3 times with PBS-T, slides were blocked with PBS-T containing 5% BSA for 1 hr at room temperature. Slides were stained with fluorscein-conjugated anti-DIG antibody for 3 hour. After washing in PBS-T twice, DNA was counter-stained with DAPI and mounted with anti-bleaching solution Vectashield (Vector Laboratory). The samples were imaged using a confocal microscopy (Zeiss). Telomere signal was quantified with TFL-TELO software (Poon et al., 1999) with its default setting except threshold level 15.

Telomere restriction fragment (TRF) assay. For genomic DNA preparation, worms were harvested and washed 5 times in M9 buffer. Worms were lysed in lysis buffer for 8 hr (100 μg/ml proteinase K, 50 mM KCL, 10 mM Tris pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, 1 % β-mercaptoethanol) DNA was extracted using phenol:chloroform extraction and ethanol precipitaion. DNA

dissolved in TE buffer was treated with RNAse (10 µg/ml) for 2 hr and re-extracted and dissolved in TE buffer.

For Southern hybridization, 5 µg of DNA was treated with 1 unit of restriction enzyme and then separated by gel electrophoresis either using standard equipment or Pulsed Field Gel Electrophoresis equipment. Gels were blotted by capillary transfer on to the Zeta probe membrane (Bio-Rad) overnight. The membranes were cross-linked using a UV cross-linker and hybridized with the Southern probe in DIG Easy Hybridization buffer at 62°C for 16 hr when using oligonucleotide probe or 42°C in case of using TALT probe. The membrane was then washed twice at room temperature in 2x SSC, 0.1 % SDS and twice at 42 °C 0.2x SSC, 0.1 % SDS. The DIG-labeled probe was detected on an ImageQuant LAS-4000 biomolecular imager (GE healthcare) using an anti-DIG-AP antibody chemiluminescence detection kit (Roche). Probes for TALT I and TALT II were labeled with DIG-UTP using primers targeting unique region in TALTs.

Bal 31 exonuclease treatment. 5 μg of DNA was treated with 10 unit of Bal 31 at 30°C in 1x Bal 31 buffer. Reactions were stopped with the addition of EGTA (25 mM final concentration). DNA was collected by ethanol precipitation and digested with *Bam*HI.

Sequence analysis of telomere-subtelomere junctions. To analyze the exact sequence of the junction of telomere and subtelomere, PCR products of junctions were sequenced by Sanger method. For PCR reactions, the forward primer was designed against telomere-adjacent DNA to elongate into the telomere and reverse primer was designed from TALT 1 specific sequence (the first exon of T26H2.5). CS1 genomic DNA and N2 genomic DNA were used as template. 30 PCR cycles were performed with primers annealling at 60°C and elongation for 3 min. After electrophoresis in 1% agarose gel, the CS1-specific amplicon was gel-extracted and sequenced. Primers used for sequencing were the same as those used for PCR reactions.

Variant Discovery

To analyze variant induce by EMS and CB4856 associated variant, WS243 version of C. *elegans* reference genome and annotation data was acquired from Wormbase website (www.wormbase.org). For preprocessing the raw sequencing data, Trimmomatics was applied so that illumina adapter sequence was removed. 3' and 5' end of low quality reads were trimmed. Preprocessed reads were then aligned to *C. elegans* reference genome (WS243) using Burrows-Wheeler Aligner software version 0.7.5a (Raices et al., 2008). Reads were aligned to reference genome with BWA-MEM algorithm using default parameters. Before calling variants, Picard's SortSam and MarkDuplicates and GATK's indel realignment and base quality score recalibration (BQSR) was applied. For the indel realignment, very stringent parameters was used because of the probability of considerable genetic difference between reference genome and survivors. From reference annotation (WS243), SNPs and indels sourced from million mutation project were selected and used in BQSR. According to GATK Best Practices, variant discovery and genotyping were performed across all samples

simultaneously using GATK's haplotype caller. Then standard hard filtering was applied with minimum raw depth of coverage of 4.

De-novo assembly and searching a contig covering T26H2.5

To extract T26H2.5, we assembled WGS data of Hawaiian wild type strain using CLC GenomicsWorkBench 7 (Qiagen) with its default setting. To identify unique contigs that contains T26H2.5 among the contigs with draft genome, BLASTN search was used.

Alignment of whole genome sequencing data. DNA was fragmented by average of 300 bp and sequencing libraries were constructed with an average insert size of 430bp. Libraries were run on Illumina Hiseq2000 platform. Resulting fastq files were aligned to the ce10/WBcel215 reference assembly using bwa, and converted into bam files using Samtools(Li et al., 2009).

Telomere reads analysis. Bam files from different worm strains were analyzed for reads predicted to derive from within telomeres using previously described software (Conomos et al., 2012). Briefly, reads which contained at least 6 canonical telomere repeats (TTAGGC) were extracted and saved into separate files. The number of reads was counted and normalized to total read density to infer telomere length. The saved files were subsequently used to analyze split-pair regions and for variant calling.

TALT bioinformatic analyses. To identify regions that were associated with the telomere, files containing telomeric reads were used to extract their associated paired-ends. Reads in which both pairs were telomeric were excluded, leaving only reads in which one pair mapped to a genomic location. In addition to subtelomeric regions and loci flanking interstitial telomeres, the majority of the reads mapped to TALT1 in the case of CS1 and CS2, and TALT2 in the case of NS1 and NS2. Average genomic coverage across the genome was calculated using genomecov, a program from the BedTools package (Quinlan and Hall,

2010), after the removal of duplicates and low mapping quality (q<20) reads in Samtools. The read depth was assessed at every nucleotide in the genome and normalized by dividing by the average coverage to give a fold-change metric for read depth. The fold-change read depth from the parental strains was subtracted from the ALT survivors to account for strain-specific variations across the genome. Line plots of genome-wide and locus-specific read depth changes were created in R software.

Telomere variant calling. Telomere variants were analyzed from files containing telomere reads. The number of each variant type was counted within this subset of reads from each library as described (Conomos et al., 2012). For each variant repeat, the frequency within the library was calculated and used to generate pie charts.

mRNA sequencing analysis. To align the preprocessed RNA-Seq reads to reference genome and estimate the expression level of transcripts, we followed the Tuxedo pipeline (Tophat2 and Cufflinks pipeline)(Kim et al., 2013). First,

Tophat2 mapped RNA-Seq reads to the *C. elegans* genome using given transcript annotation (WS243). Next, using cuffquant and cuffnorm software in Cufflinks package, we estimate the abundance of known transcripts in normalized RPKM across all samples (Trapnell et al., 2012).

GSEA analysis. The gene ontology (GO) annotation data for *C. elegans* were downloaded from the Wormbase website. Gene sets that significantly changed upon gamma irradiation in *C. elegans* was obtained from (Greiss et al., 2008). For all pairs of ALT samples and a wild type, Gene Set Enrichment Analysis (GSEA) was performed using the gene set obtained above (Subramanian et al., 2005).

Primers. Primers specific to TALT

TALT 1 specific probe

primer089L GAA GGA TGG GAT GGA ACT GA

primer089R AAC AAG CTT CTG GCA CGT CT

TALT 2 specific probe

primer116L gttttcatgctaaattcaaaacg

primer116R aaacacaggagcatataggttgg

Results and Discussion

Isolation of stable ALT survivors in *C. elegans* in two wild isolates.

A deletion allele of the telomerase gene *trt-1*, *trt-1(ok410)* was introduced into wild-type N2 and CB4856 worm backgrounds, one of the strains that show an extensive divergence in single nucleotide variants (SNVs) (Andersen et al., 2012). We then treated these worms with a DNA alkylating agent ethyl methane sulfonate (EMS), and after 8 generations we isolated survivors (**Fig. 1**). We isolated 6 independent survivors (named as 'CS 1' to 'CS 6') from 80 EMS-treated plates of CB4856 *trt-1* worms, while 5 survivors were isolated from 200 EMS-treated plates of N2 *trt-1* worms (named as 'NS 1' to 'NS 5'). All 6 CS lines and 2 of the NS lines have been maintained for at least 300 generations by transferring 10 to 15 individuals every generation, suggesting that these ALT lines are stably maintained except NS3, NS4 and NS5. The other three NS survivors (NS3 to NS5) have been maintained only by massive transfer of worms every generation (**Fig. 2**).

An initial examination of the CB4856 *trt-1* survivors found that telomere lengths were longer than in the starting strains when assessed by both florescent *in*

situ hybridization (FISH) and terminal restriction fragment (TRF) analysis (**Figs. 3, 4**). The telomere signal was abolished after BAL31 exonuclease treatment (**Figs. 5, 6**), suggesting that they were located at chromosomal ends. Surprisingly, unlike the starting strains, all CB4856 *trt-1* survivor lines showed discrete banding patterns on a TRF analysis when *Hin*fI restriction enzyme was used **(Fig. 7**). Since the recognition sequence for this enzyme does not cut the canonical *C. elegans* telomere repeat, these data suggested that additional sequences were interspersed within the telomere. Existence of 3 major bands suggest that additional sequence with at least 2 *Hinf*I restriction sites might exist as units for lethening the telomeres of CS survivors. Therefore, CB4856 *trt-1* survivors were able to maintain genomic integrity by lengthening their telomeres with sequences different from the canonical telomeres.

Chromosome number of intestine in *C. elegans* increases in every round of larval molting by endoreplication (Kipreos, 2005). I expected that large intestinal nuclei and increased numbers of chromosome might facilitate observation of telomeres. As expected bright signal can be observed in intestinal nuclei. Closer

examination of FISH experiment revealed distinct telomere localization (**Fig. 8**). Telomeres of *pot-2(tm1400)* dispersed in the intestinal nucleus of adult. However one of the brightest telomere signal of CS2 survivor was found in the nucleolus consistently. Although ALT-associated PML-nuclear body (APBs) is reported as hallmark of ALT in human (Cesare and Reddel, 2010), I coud not find any similarity with APBs. Localization of telomeres in nucleolus was only observed in postmiotic intestine but not in embryo with dividing cells, suggesting that this localization may not be required for telomere maintenance.

As ALT mechanism is recombination-mediated process, the presenece of single-stranded telomere overhang might be important for ALT by invading double-stranded template for recombination. To find out any increase in overhang, TRF assay was performed in native condition (**Fig. 9**). TRF in denatured condition was done as positive control. There was no significant difference in ALT telomere overhang compared to wild type, suggesting that length and number of overhang was not required for ALT activity.

Association mapping reveals ALT-associated genomic region.

Since we treated EMS as mutagen to induce ALT by inactivating potential ALT repressor genes, we first tried to identify specific mutation(s) responsible for the ALT survivors in the CB4856 *trt-1* background. To do this, we analyzed the whole genome sequences of two typical survivor lines, the CS 1 and CS 2 survivor lines, before and after multiple rounds of backcrosses with N2 *trt-1* worms to exclude unrelated genetic variations. In case of the CS 1 survivor line, none of the variants that were present in the initial isolates were maintained after outcrosses except for a single nucleotide change in an intron in a pseudogene (**Fig. 10A**). In case of CS 2 survivor line, three small genomic regions remained after outcrosses, and there were 3 exonic variants in those regions, none of which exhibited the ALT phenotype when inhibited by RNAi (**Fig. 10B, Fig. 11**). In addition, we were unable to induce ALT by RNAi of the genes identified as in-dels by array CGH (**Fig. 12**). Combined, although we cannot rule out the possibility that there was an ALT-initiating mutation, which was not required for maintenance of ALT in subsequent

generations, our favorite interpretation of these analyses is that survivors did not acquire an ALT-specific mutation, but that potentially a more complex genomic event occurred that was responsible for ALT induction after treatment of EMS.

We then employed an association mapping strategy to identify specific genomic regions that are closely associated with ALT induction from the starting strain of CB4856 *trt-1*. We reasoned that the CB4856 *trt-1* survivors should retain any genomic region derived from CB4856, if any, that conferred the ALT phenotype, even through extensive outcrosses with genetically distinct N2 *trt-1* worms. We therefore looked for CB4856-specific genomic regions that were not erased by successive outcrosses with N2 *trt-1* for both CS 1 and CS 2 survivors. We found that only 2 SNVs on chromosome V derived from CB4856 *trt-1* were retained after multiple outcrosses and therefore tightly associated with the ALT phenotype both in CS 1 and CS 2 survivors (**Fig. 13**). Surprisingly, read depth analysis of these two SNVs revealed that both the wild CB4856 strain and the survivors actually contained both N2 and CB4856-derived SNVs (**Fig. 14**). Furthermore, only the CB4856 type variants were overrepresented in the survivors. For CS1, reads

containing the N2-derived variants were represented at genome-average sequence depth (0.9-fold), while reads containing the CB4856-type variants were increased 76-fold above average depth. Similarly, in CS2 reads containing the N2-derived variants were represented at average levels (1.2-fold), while reads containing the CB4856-derived variants were greatly enriched (122-fold). PCR followed by sequencing revealed that the endogenous sequence of the chromosome V region contained only N2-type SNVs, suggesting that the SNVs that were called as 'CB4856-derived' are actually located somewhere else in the genome. Indeed, we found that the 'CB4856-derived' SNVs are located on the right end of chromosome V only in the CB4856 strains and the survivors, not in the N2 strain (**Fig. 15, 16**). We futher confirmed relative copy number increase in CB4856 compared N2 by quantitative PCR (**Fig. 17**). The replicated units seemed to be multiplicated in tandem repeat in CB4856 (**Fig. 18, Fig. 19**)

Trans-recruitment of an internal genomic region, TALT 1 reservoir, is correlated with ALT in the survivors derived from CB4856 trt-1.

Further analysis of the whole genome sequencing data of CS1 and CS2 survivors showed that a specific genomic region of 1398 bp (V 19,229,887 – 19,231,285) spanning the 2 SNVs associated with the ALT phenotype was over-represented in the CS 1 and CS 2 survivors as high as 100 fold above average sequence depth (**Figs. 20, 21**). This region, which we termed 'Template of ALT 1' (TALT 1) was amplified in all CS survivors (**Fig. 22**), and remained the most amplified sequences even after extensive outcrosses (**Figs. 23**). We confirmed that telomere repeat count from whole genome sequencing data also increased (**Fig. 24**). This consisted of a block of uniquely-mapping sequence flanked by two interstitial telomere-like sequences (ITS), present in the same orientation.

In order to identify the chromosomal locations of the over-represented TALT 1 in the genome of the survivors, we extracted all paired end sequence reads containing parts of the TALT 1, constructed contigs ranging from 200 bp to 1.7kb

long, and then determined where these contigs mapped on the genome by BLAST analysis. We found that, in addition to regions adjacent to TALT 1, five of the contigs also aligned to the ends of chromosomes **(Table 1)**. PCR reactions using primers specific to the TALT 1 and primers specific to these chromosomal ends validated these findings because PCR amplicons were produced only from DNA derived from the survivor lines (**Fig. 25**). We further confirmed this with FISH experiments, where a TALT 1 probe co-localized perfectly with the telomere probe (**Fig. 26**). Finally, we performed TRF analyses using enzymes with no restriction sites in the TALT 1 and enzymes which do cut within the TALT 1 sequence. We recognized telomere smears of CS1 survivor that were longer than CB4856 when the TALT 1 was not cut on a Southern blot using both a telomere probe and a TALT 1 probe (**Fig. 27**), conclusively showing that the TALT 1 is present in multiple copies in the telomeres of the survivors.

N2 survivors utilized another TALT for ALT.

The discovery of the TALT 1 in the CB4856 *trt-1*-derived ALT survivors prompted us to explore whether additional TALT existed in the N2 *trt-1*-derived ALT survivors that we isolated. Similar to the CB4856 *trt-1* survivors, we found that the two independently isolated N2 *trt-1*-derived ALT survivors (NS1 & NS2) contained longer telomeres than the starting strain, and produced discrete bands on TRF blots using restriction enzymes that do not cut canonical telomere sequences **(Fig. 28).** These bands were markedly different from identically digested CB4856-dervied survivor TRFs, suggesting that while the N2 *trt-1* survivors have incorporated specific sequences into their telomeres, these sequences are not TALT 1 elements. WGS analysis of the NS1 & NS2 revealed that both lines had amplification of a specific genomic adjacent to the 5′ telomere of chromosome I (**Fig. 29**). In addition, this was the only region identified as being associated with telomeres by examining the paired ends of reads containing telomeric DNA. We named this region '*T*emplate of *ALT 2*' (TALT 2). A probe specific to the TALT 2

co-localized to telomeres both in TRF and FISH analysis (**Fig. 30, 31**), strongly suggesting that the TALT 2 was present in chromosomal ends in the N2 *trt-1* survivors. PCR reactions using primers specific to the TALT 2 and primers specific to chromosomal ends validated these findings (**Fig. 32**). The genomic structure of the TALT 2 is similar to that of TALT 1, in that the it consists of a unique DNA fragment flanked by telomeric sequences (an ITS at one end and telomere sequences at the other) **(Fig. 33)**. Similar to the TALT 1, which contains a TALT 1 replicated from the endogenous locus to a chromosome end, TALT 2 also is originated from an endogenous locus close to the left arm of chromosome I (**Fig. 34**). Interestingly, the TALT 2 at the left arm of chromosome I is undetectable in the CB4856 genome (**Fig. 35**), suggesting that CB4856 and N2 have evolved different TALT reservoirs for ALT.

The *C. elegans* strain deficient in the telomerase gene was reported to overcome sterility without telomere lengthening only when maintained as large populations, presumably due to stochastic telomere shortening events occurring within the population (Cheng et al., 2012). In stark contrast to these data, most

survivors we present in this study incorporate TALT into their telomeres and can be stably maintained by means of a typical transfer in which a few worms are transferred to a new plates for culture each generation (**Fig. 36**). One reason for this apparent discrepancy is that we treated the worms with EMS; we propose that EMS elevated DNA damage response in some way. Consistent with this, RNA-seq analysis on the survivors showed that among specific gene sets that we compared with, the gamma ray induced-gene set showed a significant correlation with the genes significantly altered in ALT survivors (**Fig. 37**), suggesting that *C. elegans* ALT survivors maintain altered double-strand break (DSB) responses, which is similar to the situation in human ALT cells (Cesare et al., 2009).

In budding yeast, telomeres are transcribed into non-coding telomeric repeat-containing RNA (TERRA). It is recently reported that TERRA can regulate facilitate telomere recombination by forming RNA-DNA hybrids which can be removed by RNase H enzymes(Balk et al., 2013). To seek the possibility that telomeric repeat might be transcribed in ALT survivor, RNA-FISH was performed against TERRA in native condition (**Fig. 38**). TERRA was observed in pachytene

stage of meiotic germline in both N2 wild type and CS2x4 survivor. This signal was abolished after RNase treatment, confirming that this is not single-stranded telomeric DNA. TERRA was increased in CS2 survivor suggesting that increased TERRA might positively regulate telomere recombination.

In this study, we have shown for the first time that internal genomic regions can integrate into the telomeres, resulting in elongation and maintenance of telomeric ends in the absence of functional telomerase. We propose a model of ALT event in *C. elegans* as follows, which may hold true for human ALT (**Fig. 39**): the first prerequisite event was that wild strains randomly incorporated TALT from an original chromosomal region ('TALT Donor') to an end of the same chromosome (cis-recruitment of 'TALT Reservoir'): TALT 1 D from an internal region to TALT 1 R on the right end of chromosome V in CB4856 and TALT 2 D from an internal region to TALT 2 R on the left end of chromosome I. Then when crisis occurs, in rare animals the TALT are recruited to all chromosomal ends (trans-recruitment), stabilizing chromosomes and survive. At present however, the mechanism by which internal TALT s can be recruited and replicated to the telomeric ends by *cis*- or

trans-recruitment still remains unknown. Although circular telomeric DNA formation is a well-known marker of human ALT (Henson et al., 2009), we did not observe an increase in circular telomeric DNA formation in our survivors **(Fig. 40**), suggesting that a different mechanism is involved. Furthermore, we were unable to identify a single factor that suppressed the ALT phenotype in our survivors using a candidate RNAi approach (**Table 2**), although it is still possible that essential processes such as break induced replication (BIR) and recombination may be involved in the genesis of our ALT survivors (Lydeard et al., 2007; McEachern and Haber, 2006).

Recent studies suggest that the intersepersed variant telomere repeat in ALT telomere can create positive feedback loop that can sustain recombinogenic telomeres (Conomos et al., 2013). In this scenario, initial cause that incorporated variant telomere repeat into the telomere can be uncoupled with sustaining ALT activity. As a result telomeres once acquired ALT activity might be able to lengthen itself without initial cause. Although our favorite interpretation is that there was no iniation mutation, it is still possible that initial cause might be lost during

outcrossing to different background due to the same reason. It is consistent with 'positive feedback loop hypothesis' that ALT survivors of *C. elegans* also contained variant telomere repeats. Alternatively it is possible that EMS itself acted as non-specific DNA damaging agent, raising the DNA double strand breaks to increase the substrates for recombination.

Nonetheless, the presence of TALT R sequences near the proximal telomeres is not an accidental event, because the occurrence of the N2 type TALT seem to have evolved several times independently (**Fig. 41**). As human ALT cells often incorporate variant telomere repeats that may originate from proximal telomeres, it is conceivable that human cells also utilize the TALT mechanism for ALT. Therefore, it would be worthwhile to screen human ALT cell lines and tumors to investigate the presence of a TALT-like mechanism facilitating ALT activation in humans. Our results suggest the presence of TALT R in proximal telomeres is an important factor in an ALT activation or initiation.

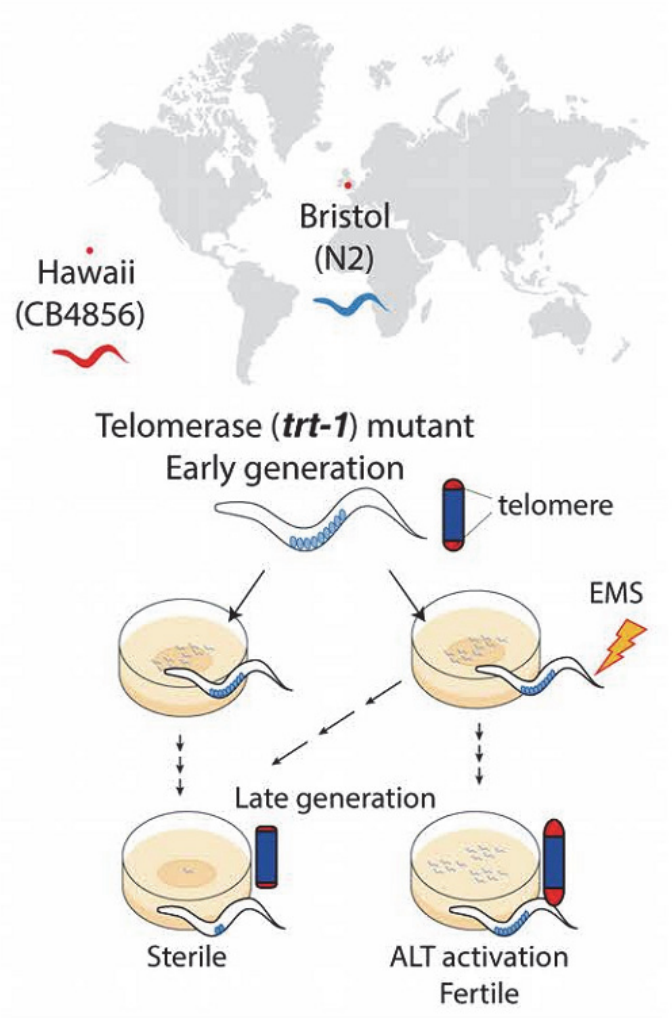


Figure 1. Isolation of stable ALT survivors in C. elegans in two wild isolate. As generation goes, trt-1(ok410) mutant becomes sterile. Most of late generation trt-1 mutant becomes sterile but low portion of survivors arose. In collaboration with C. Kim and D. S. Lim

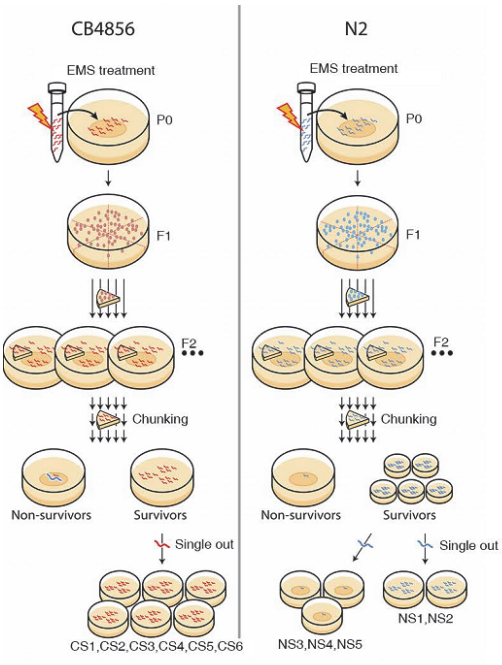


Figure 2. A schematic image of independent ALT survivor isolation. For CB4856 survivors, we separated worms to 80 plates after EMS treatment. After 8 generation, we found 6 independent survivors (CS1-CS6), which were maintained by transfer of small number of worms each generation. For N2 survivors, we separated worms to 200 plates after EMS treatment. We were able to maintain 5 survivors (NS1 – NS5) by large chunking and two of the five survivors (NS1 and NS2) were maintained by transfer of small number of worms each generation. In collaboration with C. Kim, S. Sung and D. S. Lim

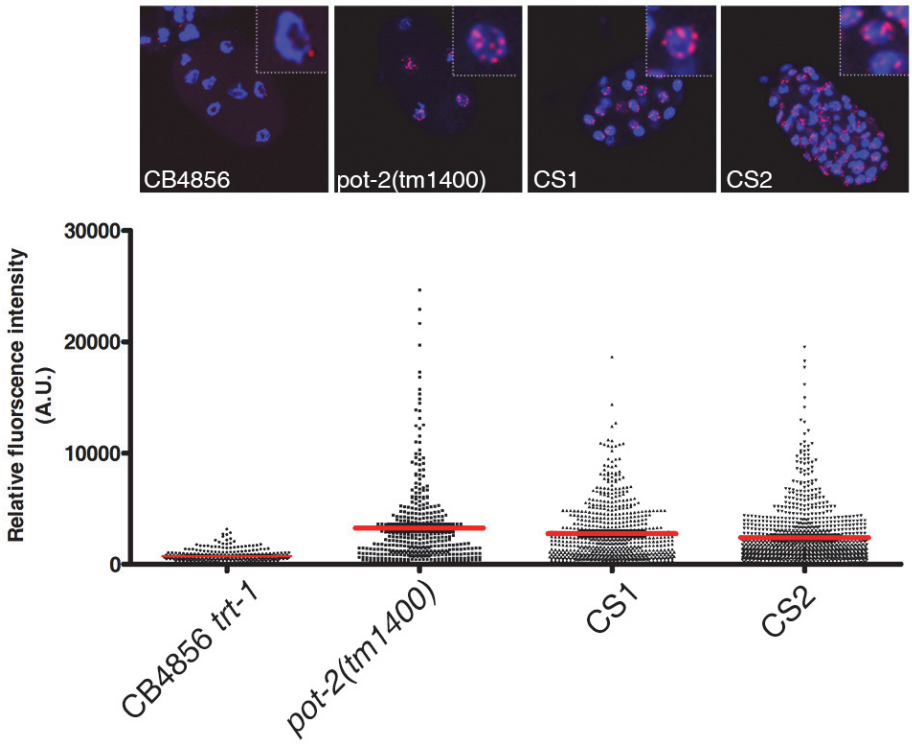


Figure 3. Telomeres become lengthened in CS survivor. FISH analysis using telomere probe in CB4856 trt-1, pot-2(tm1400) CB4856 survivors, CS1 and CS2 at embryonic stage. Upper panel shows representative images. Telomere was detected by Cy-3-TTAGGC*3 PNA probe(red). DNA was counterstained with DAPI, 4',6-diamidino-2-phenylindole (blue). Single nucleus is magnified in the upper right cornor by 300%. Bottom panel shows Quantitative FISH.

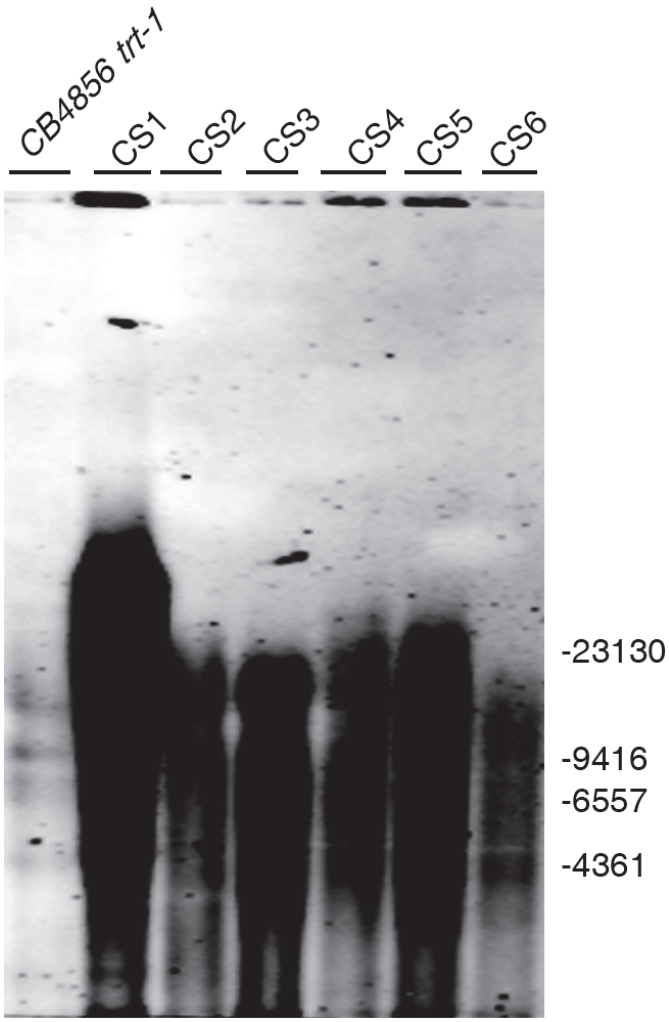


Figure 4. Molecular weight of telomere of CB4856 survivors increased.

TRF analysis of CB4856 *trt-1* and all the CS survivors with *Bam*HI. Southern blot was probed with the digoxigenin-TTAGGC*4.

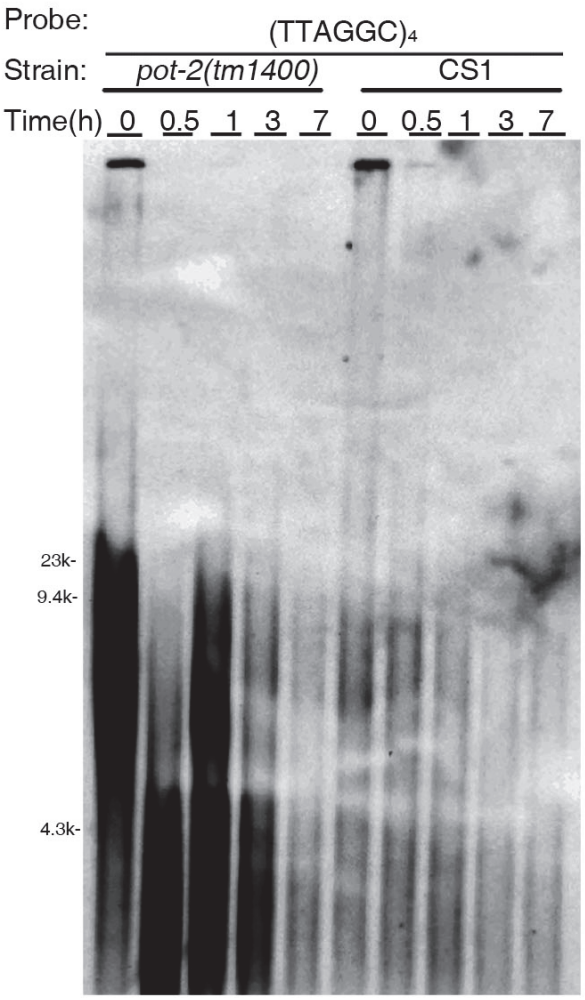


Figure 5. Telomere of CB4856 survivor is located in the terminal of chromosome. Bal 31 exonuclease assay of CS1. The pot-2 mutant line was as a positive control. Genomic DNA was treated by BAL 31 exonuclease prior to digestion with TALT non-cutting restriction enzyme (*Bam*HI). Restricted DNA was then analyzed by Southern blot.

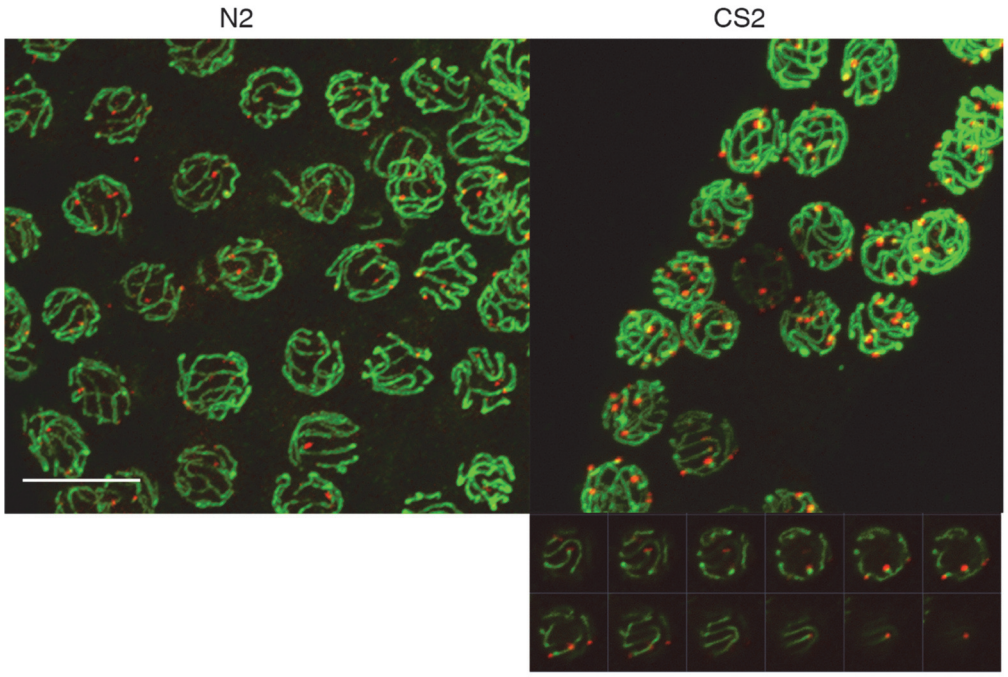


Figure 6. T**elomere of CS2 survivor localized at the end of chromosomal axis(HTP-3).** FISH was performed with TTAGGC*3(red) probe. HTP-3 was stained with Alexa-488 conjugated antibody(green) to mark chromosomal axis in pachytene stage. Lower panel shows Z-stack image of a nucleus of CS2. Scale bar = 10 μm.

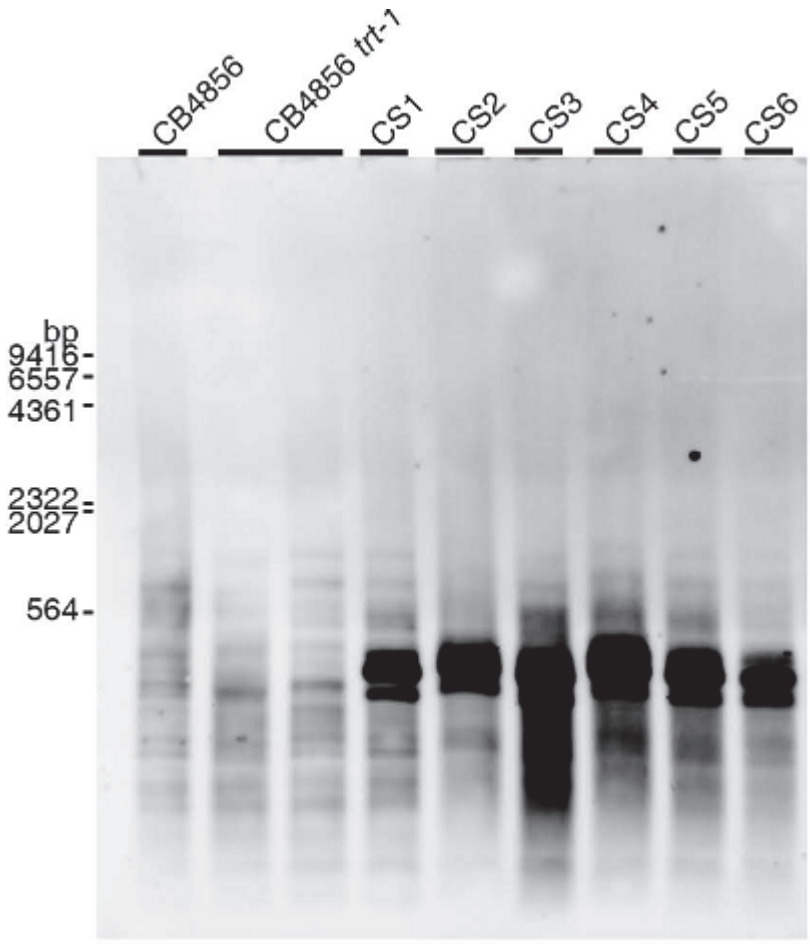


Figure 7. Telomere of CB4856 survivors contained additional sequences which were cut by *Hinf*I. TRF analysis of CB4856, CB4856 trt-1 and all the CS survivors with *Hinf*I enzyme.

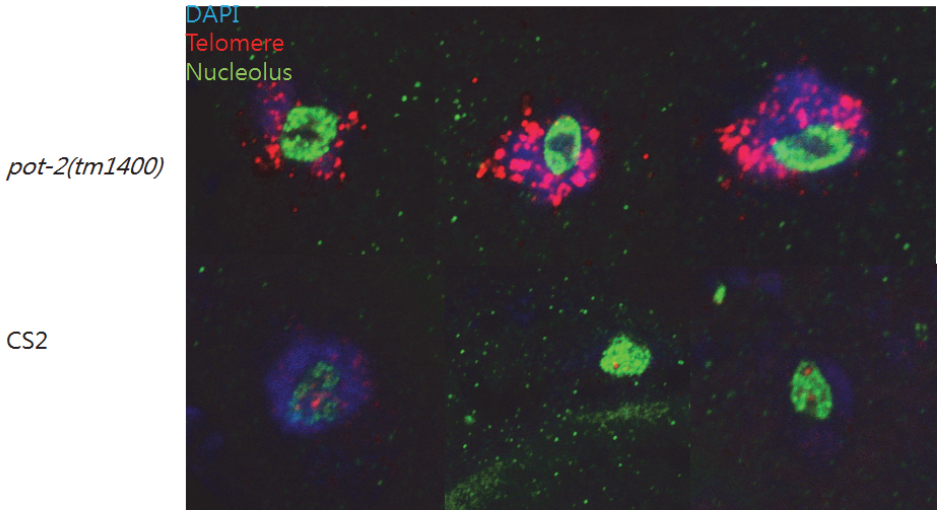


Figure 8. Subcellular localization of telomere in intestinal nuclei. FISH experiment shows one of telomere of CS2 localize within nucleolus(green). *pot-2(tm1400)* did not show this localization.

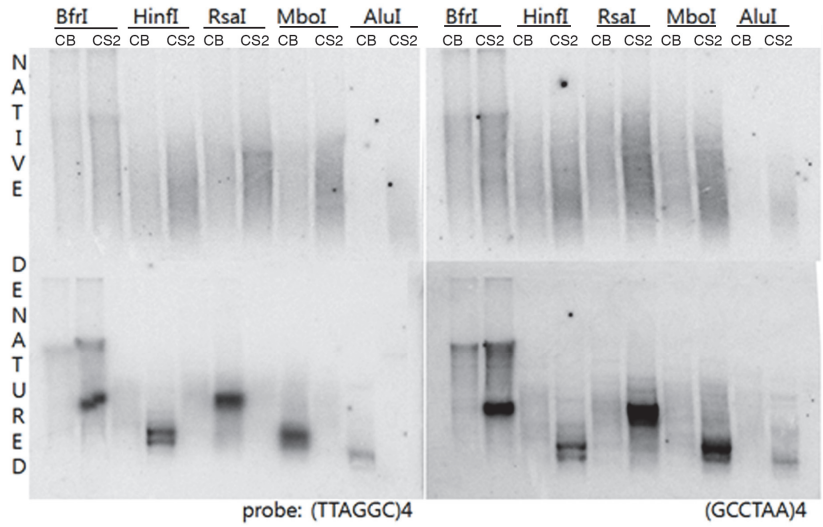


Figure 9. Telomere overhang was not changed in CS2 survivor. TRF assay was done in CB4856 and CS2 survivor either in native or denatured condition.

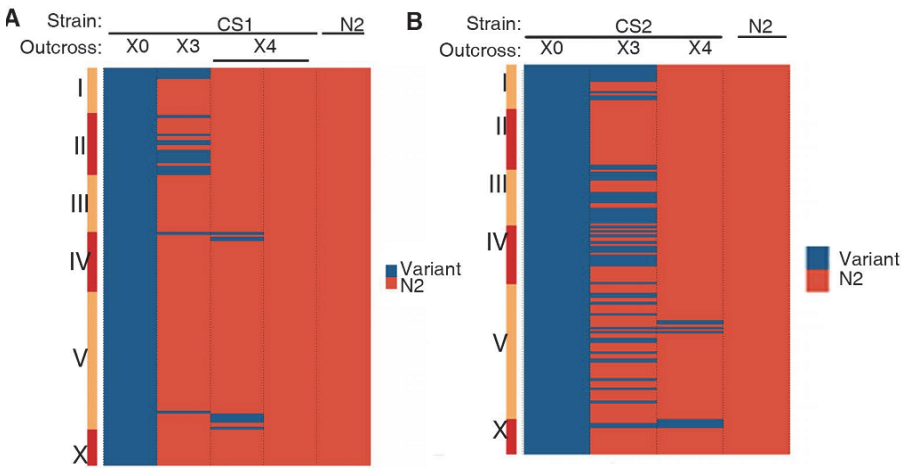
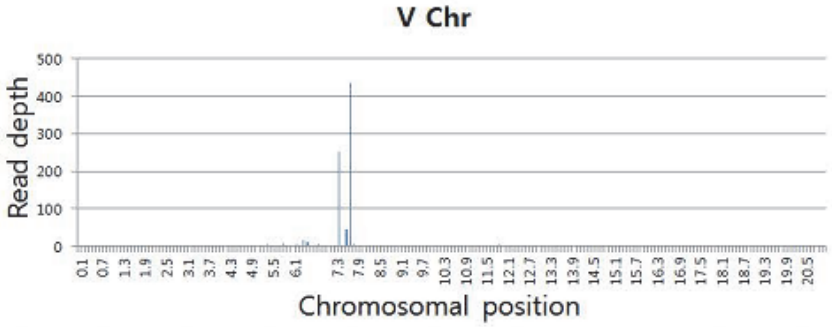


Figure 10. Not a single mutation was responsible for maintaining ALT.

(A) Whole mutation heatmap of CS1. One SNP remained by excessive outcross in CS1. This SNP is located in a pseudogene, sdz-1. (B) Whole mutation heatmap of CS2. After outcross with N2 trt-1, most variant region of CS2 (blue) changed to wild type (red). But three regions remained after three rounds of outcross with N2 trt-1. In collaboration with H. Oh and M. Choi



Chr	Position	reference	Variant	No. of reads	Type	Amino acids	ORF	RNAi result
V	7382318	C	G	4	missense	GGT->GCT[Gly->Ala]	{C03G6.18}	X
V	7620983	C	T	52	missense	CGA->CAA[Arg->Gln]	{C50E3.12}	X
V	7752372	T	G	52	missense	ACT->CCT[Thr->Pro]	{T05B11.4}	X

Figure 11. Exonic variant induced by EMS treatment did not induce ALT phenotype by RNAi in N2 *trt-1(ok410)* mutant. The list of candidate mutations in CS2. Variant was specified by counting EMS specific mutation(C to T, G to A). Feeding RNAi in CB4856 trt-1 did not increased transgenerational lifespan. In collaboration with J. Shim

A

Strain	ORF name	Gene name	Stauts	Function	Southern Lane
CS1X4-1	2L52.1		deletion	ZF TF	1
CS1X4-1	2RSSE.1		deletion	RhoGAP domain	9
CS1X4-1	2RSSE.2		deletion	RhoGAP domain	10
CS1X4-1	Y67D2.6		deletion	RNA helicase	11
CS1X4-1	Y38C1AB.4	frm-5.2	deletion	FERM domain	13
CS1X4-1	Y38C1AB.8	frm-5.1	deletion	FERM domain	13
CS1X4-2	Y66A7A.5	ceh-91	deletion	reproduction 관련	4
CS1X4-2	Y66A7A.4		deletion	unknown	3
CS1X4-2	Y38C1AB.4	frm-5.2	deletion	FERM domain	13
CS1X4-2	Y38C1AB.8	frm-5.1	deletion	FERM domain	13
CS2X4	Y74C9A.3	homt-1	deletion	Hydroxyindole-O-MethylTransferase homolog	7
CS2X4	W02B12.11		deletion	carbohydroxy	2
CS2X4	K04F1.15	alh-2	deletion	dehydrogenase	5
CS2X4	B0213.10	cyp-34A5	deletion	cytochrom P450	8
CS2X4	Y38A10A.7		deletion	unknown	12
CS2X4	cTel55X.1		deletion	unknown	6

B

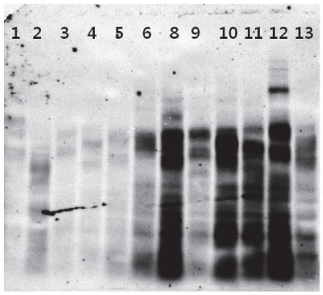


Figure 12. Deletions detected by CGH were not responsible for inducing ALT in N2 *trt-1*. (A) List of candidate genes that have deletions in exon. These lists are from the region that remained after extracting the region of CB4856 trt-1 and N2 trt-1. Yellow indicates common candidates of CS1X4-1 and CS1X4-2. (B) Candidate genes were inhibited by feeding RNAi to N2 trt-1 mutant for 10 generations. They did not increase trans-generational lifespan of N2 trt-1. And these genes did not induce TALT 1 insertion and telomere lengthening. Telomere length was measured by Southern blot. In collaboration with C. Kim

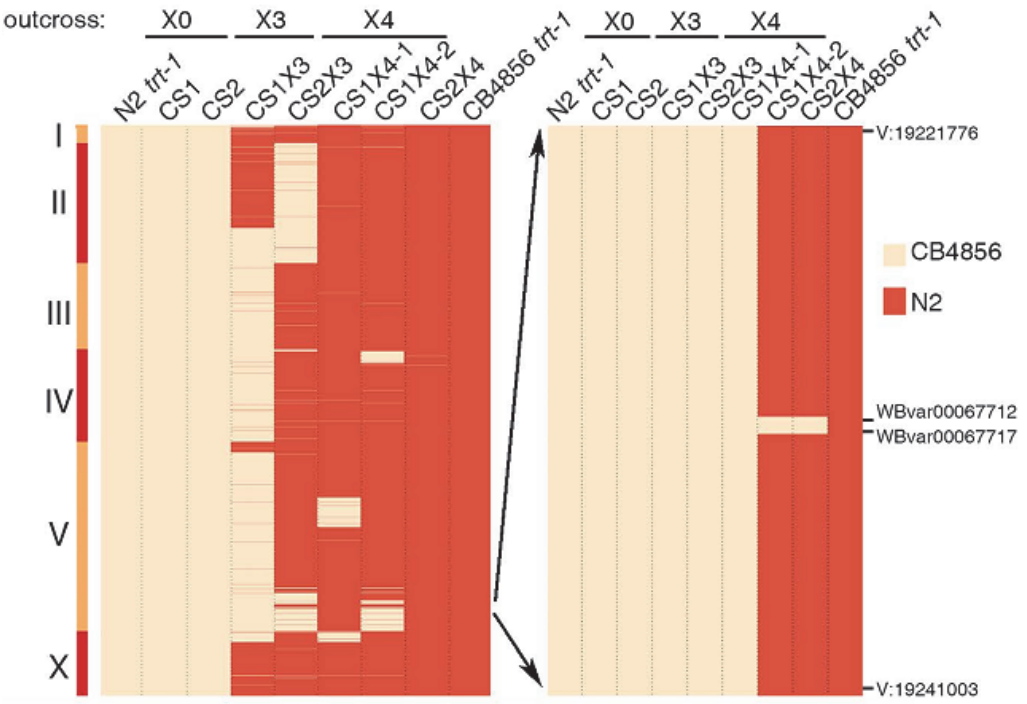


Figure 13. Two CB4856 SNPs that were retained after extensive rounds of outcross. Association heat map with SNPs inherited from parental CB4856 strain. of survivors and survivors outcroseed with N2. The right panel is a blow-up image near the T26H2.5 locus on chromosome V. In collaboration with H. Oh and M. Choi

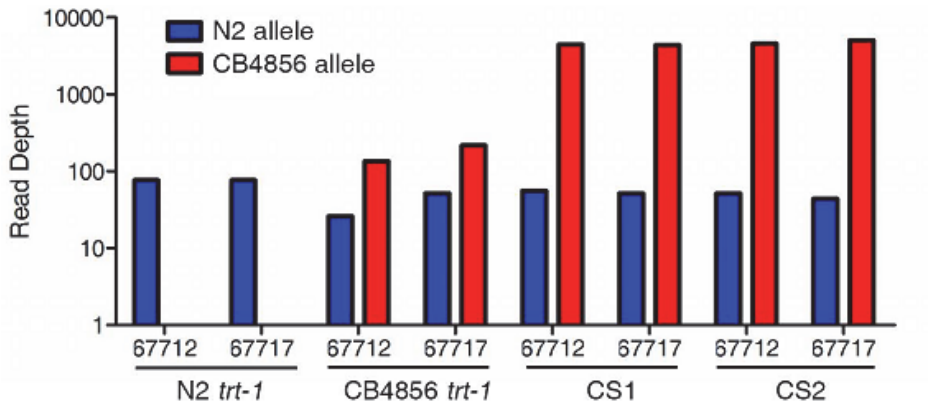


Figure 14. Both the wild CB4856 strain and the survivors actually contained both N2 and CB4856-derived SNVs. Read depth number of SNPs(WBvar00067712 and WBvar00067717) in T26H2.5 locus from WGS data of N2 trt-1, CB4856 trt-1, CS1 and CS2. N2 linked allele(blue) and CB4856 linked allele is plotted separately. Y-axis has a logarithmic scale. In collaboration with M. Hills

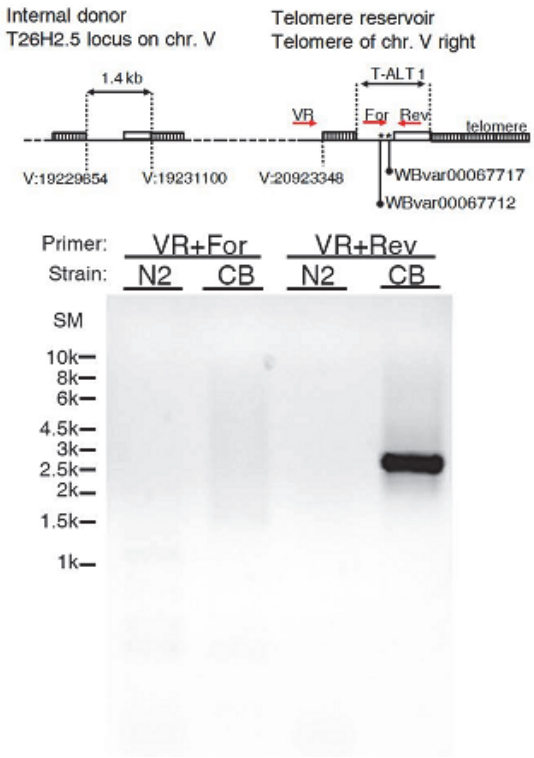


Figure 15. T26H2.5 locus is duplicated at the telomere of chromosome V in CB4856 genome. Breakpoint PCR assay with primer pair(red arrow) designed to anneal to subtelomere region of chromosome V(VR) and T26H2.5 exon I(Rev). To distinguish the orientation of translocation forward primer(For) was used. Upper panel shows T26H2.5 promoter and exon structure. Dashed box represent telomere repeat. Asterisks indicates CB4856 linked SNPs. Dashed line indicates genomic position. N2, N2 wild type. CB, CB4856 wild isolate. In collaboration with C. Kim

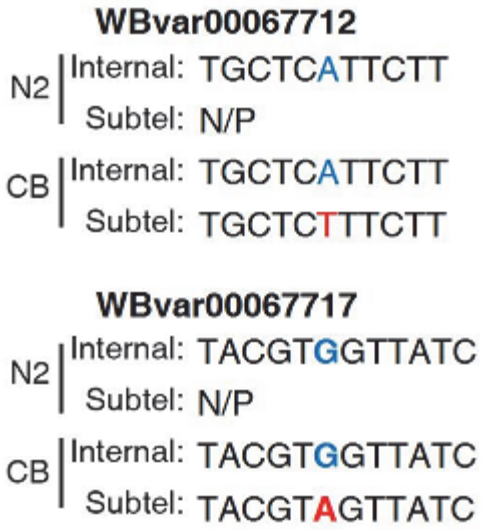


Figure 16. 'CB4856-derived' SNVs are located only on the subtelomere of chromosome V in the CB4856 strain. Nucleotide change of WBvar00067712 and WBvar00067717 in internal and subtelomere. N2 allele is colored in blue while CB4856 allele is colored in red. Nucleotide positions are indicated in panel figure 12. CB, CB4856. N/P, Not Present.

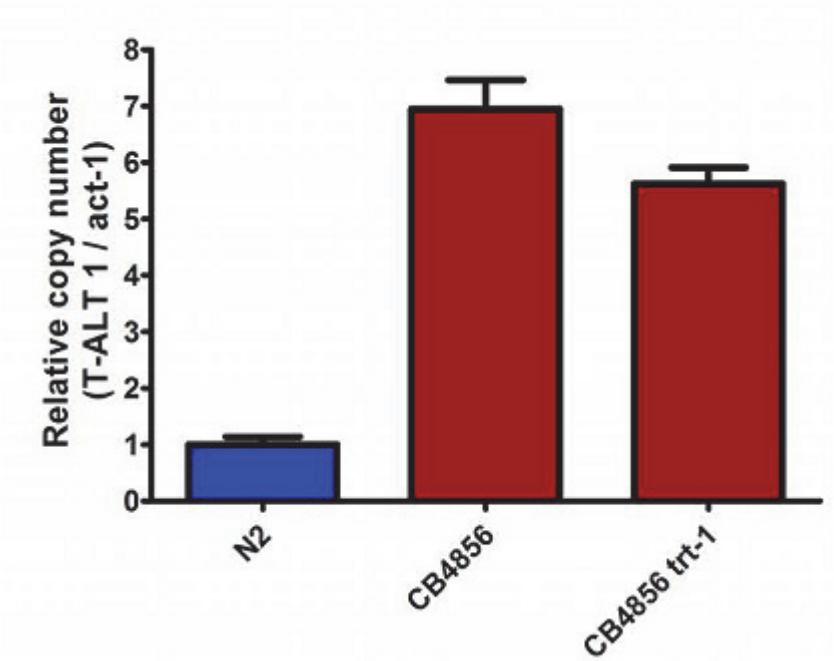


Figure 17. Copy number of T26H2.5 locus is increased in CB4856 compared to N2 by 7 fold. act-1 was used as internal control, relative copy number was normalized with N2. In collaboration with C. Kim.

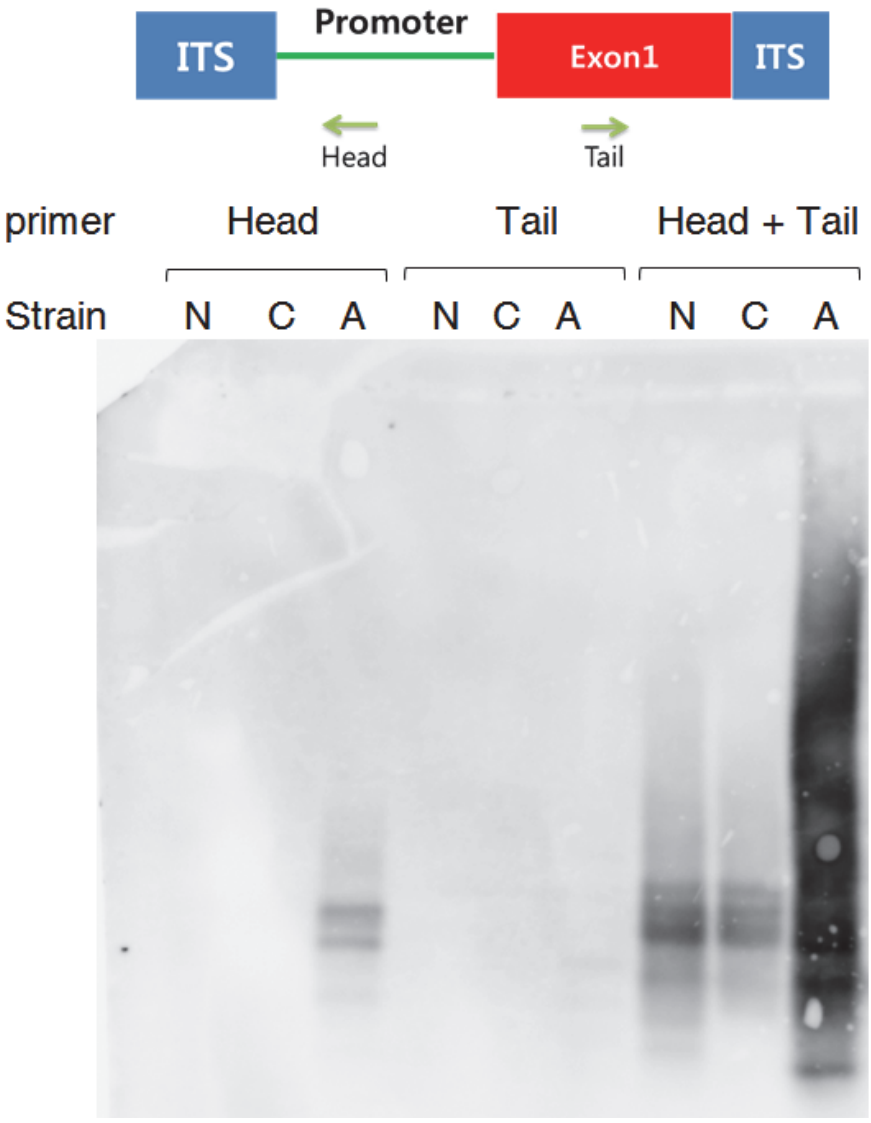


Figure 18. In CB4856, TALT 1 units were multiplicated in tandem repeat.

Head and TAIL primer mix can amplify only DNA from CB4856. The gel was blotted and detected by TALT 1 probe.

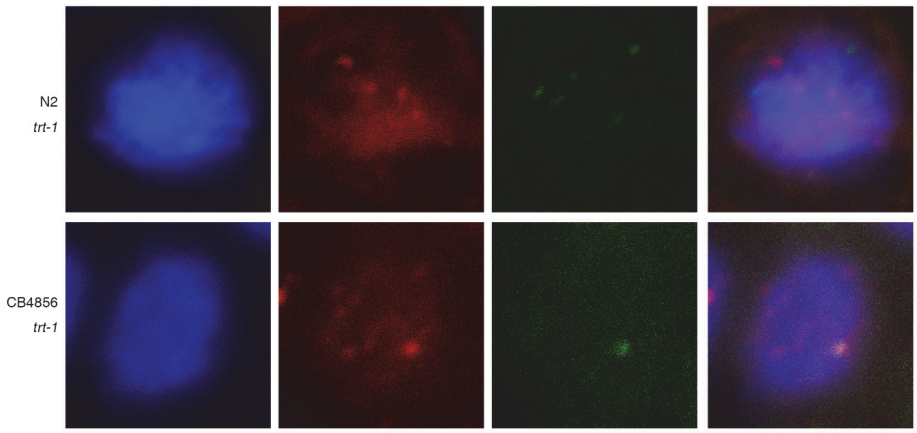


Figure 19. TALT 1 was duplicated at the telomere in CB4856 wild isolate prior to ALT. FISH result show that CB4856 TALT 1 does not colocalize with telomere in N2 *trt-1*. However TALT 1 colocalized with telomere in CB4856 *trt-1*.

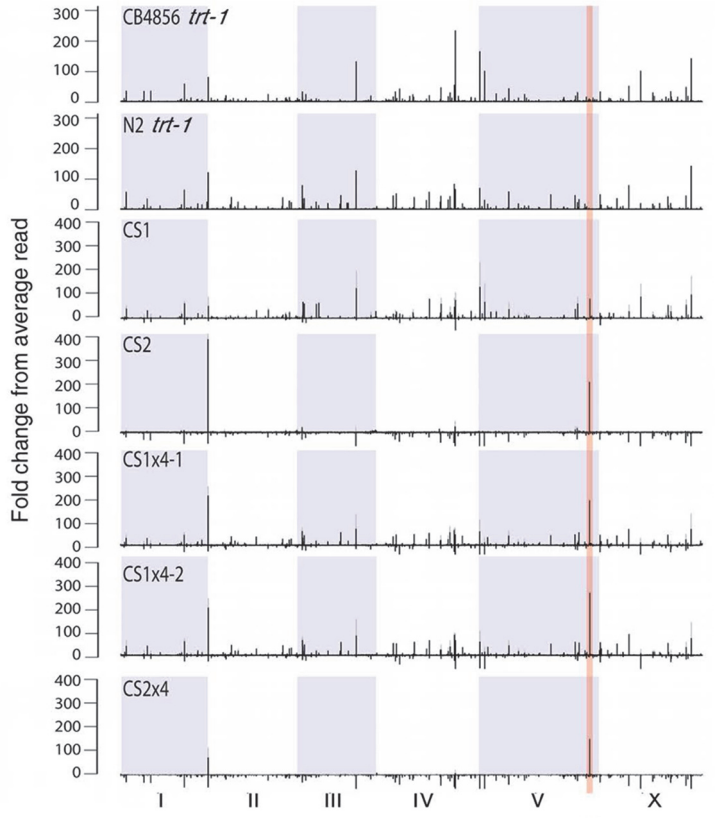


Figure 20. Amplification of T26H2.5 is correlated with ALT in the survivors derived from CB4856 *trt-1*. (A) Fold-change read depth plot of CB4856 *trt-1*, N2 *trt-1*, CS1, CS2 and outcrossed CS suvivors with N2(CS1x4-1, CS1x4-2 and CS2x4) in genome scale. To normalize strain specific variation, fold-change from average depth from parental CB4856 trt-1(grey bar) was subtracted from ALT survivors. In collaboration with M. Hills.

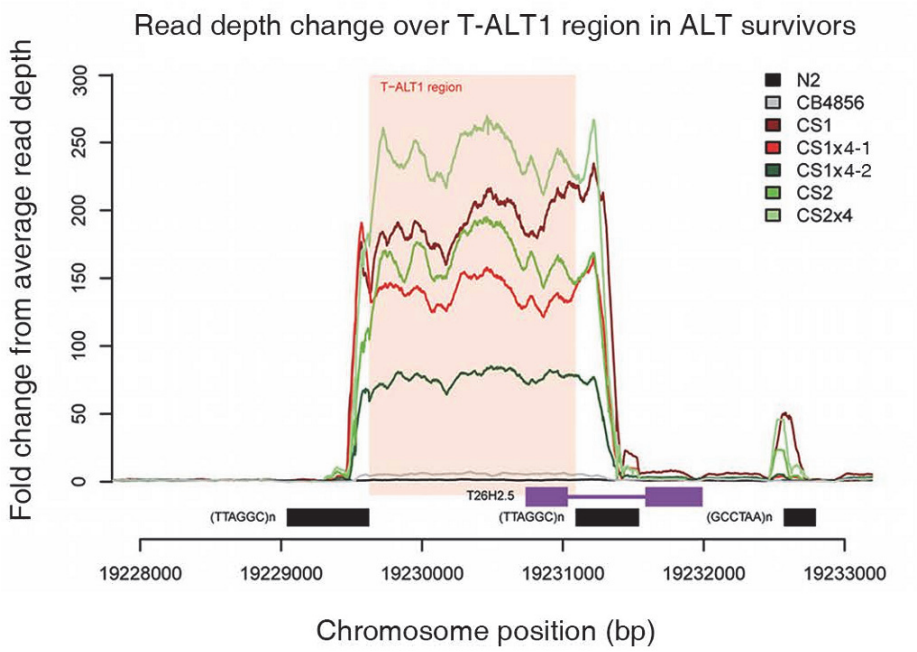


Figure 21. Amplified region in T26H2.5 locus was flanked by telomere repeats on both 3' and 5' ends. Structure of TALT 1 on chromosome V is described. Internal telomere repeat is indicated as black bar. Gene structure of T26H2.5 is indicated as purple bar. In collaboration with M. Hills.

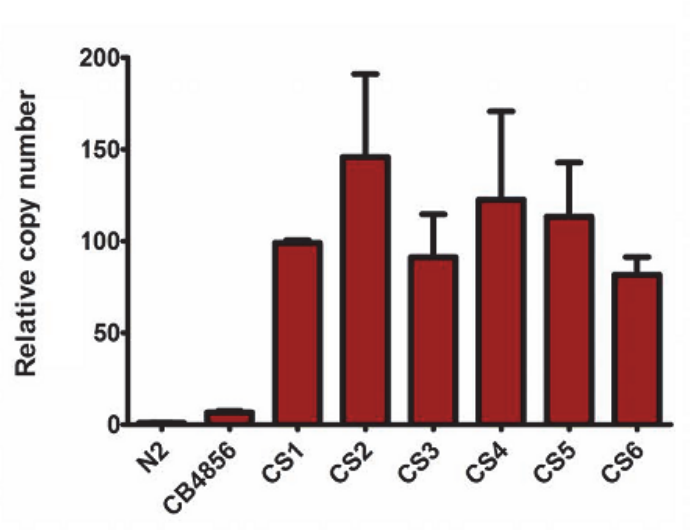


Figure 22. TALT 1 of all CS survivors increased in copy number compared to wild types. qPCR analysis by TALT specific primer. Copy numbers were normalized with act-1 gene. In collaboration with C. Kim.

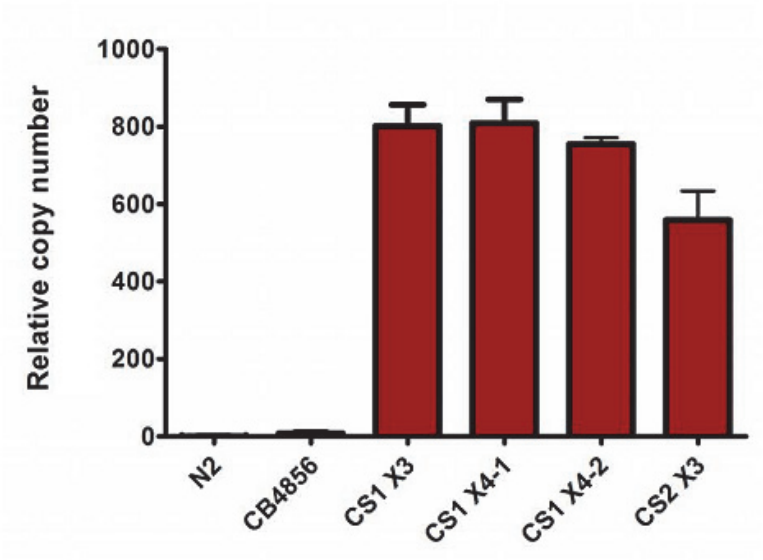


Figure 23. TALT 1 of all CS survivors increased in copy number compared to wild types after outcrosses with N2. Quantitative PCR analysis by TALT specific primer. Copy numbers were normalized with act-1 gene. In collaboration with C. Kim.

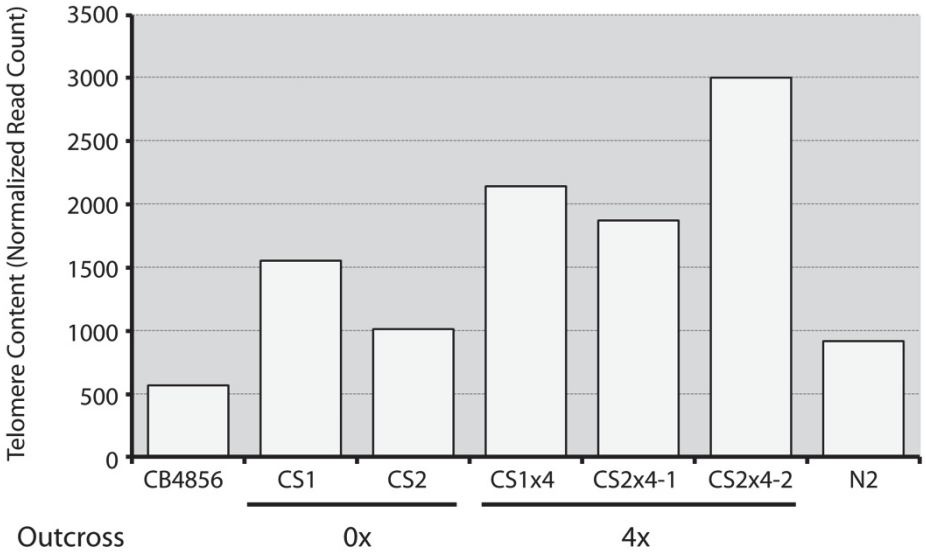


Figure 24. Telomere repeat count from WGS increased in CS survivors compared wild types. Normalized count of reads containing at least 6 telomere repeats in CB4856 trt-1, CS1, CS2 and outcrossed CS1 and CS2 survivors was measured. In collaboration with M. Hills

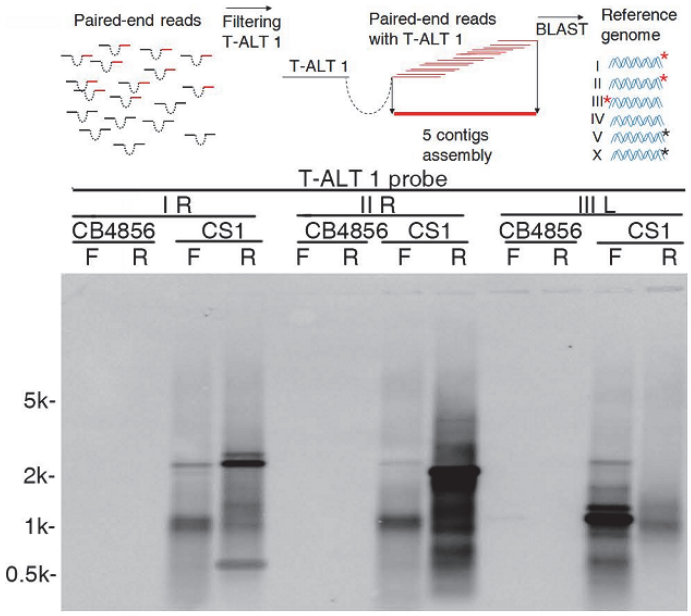


Figure 25. TALT 1 was translocated to at least three of subtelomeres in CS1 survivor. Contigs(red bar) reconstructed from reads in which one pair aligned to TALT 1 were confirmed by breakpoint PCR followed by Southern hybridization with TALT 1 probe. The bottom panel is the breakpoint PCR results of CB4856 and CS1 survivor using primers designed to amplify the trans-localized TALT and subtelomeres. F, forward primer. R, reverse primer on TALT. Black asterisks indicate predicted breakpoint by BLAST while red ones indicate confirmed breakpoint. In collaboration with H. Oh and C. Kim

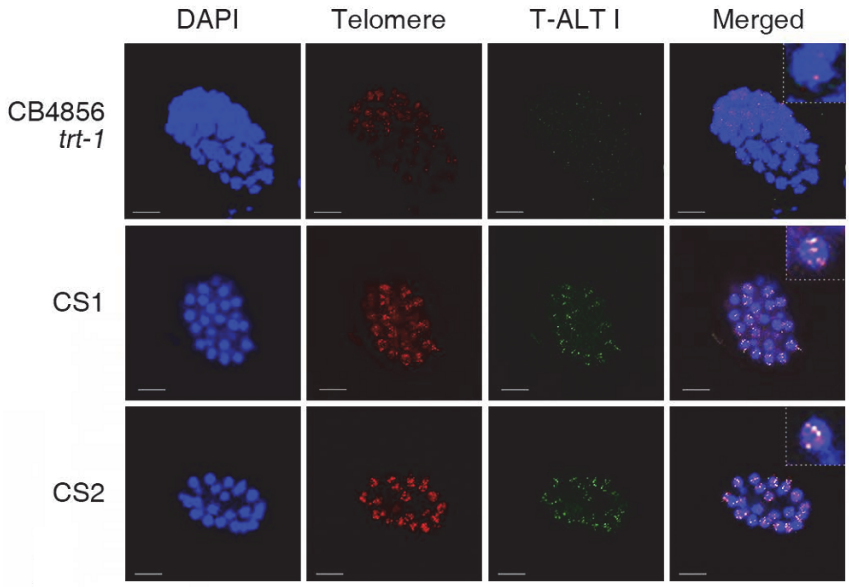


Figure 26. Telomere repeat colocalized with TALT 1 in embryo. Fluorscent *in situ* hybridization using the TALT 1 specific probe(green) and telomere probe(red) in embryonic stage of CB4856 *trt-1*, CS1 and CS2 survivor. DNA was couterstained with DAPI(blue). Single nucleus is magnified by 300%.

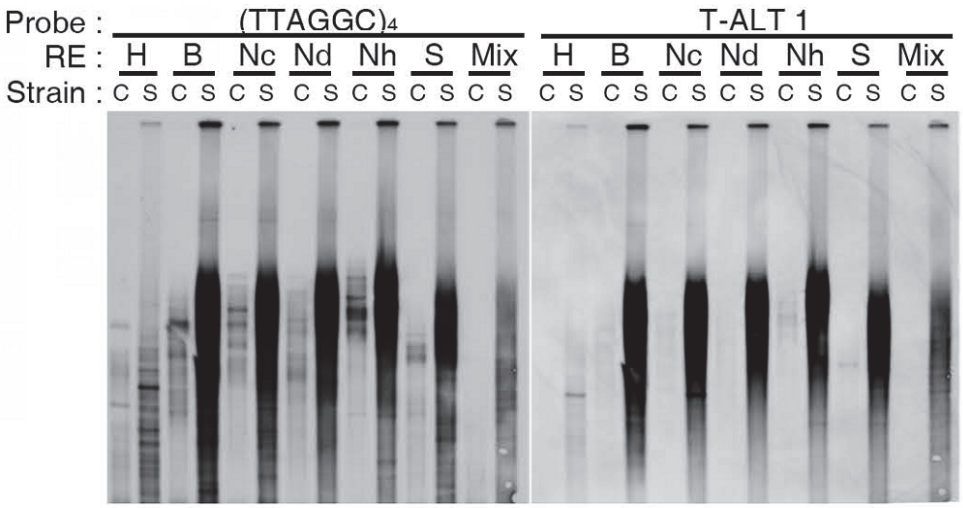


Figure 27. Telomeres of CS1 survivor were longer than CB4856. TRF analysis using a TALT 1 probe and a telomere probe. DNA was digested with indicated enzymes and hybridized with indicated probes. H:HindIII, B:*Bam*HI, Nc:*Nco*I, Nd:*Nde*I, Nh:*Nhe*I, S:*Sac*I, Mix:mixure of *Hind*III, *Bam*HI, *Nco*I, *Nde*I, *Nhe*I and *Sac*I.

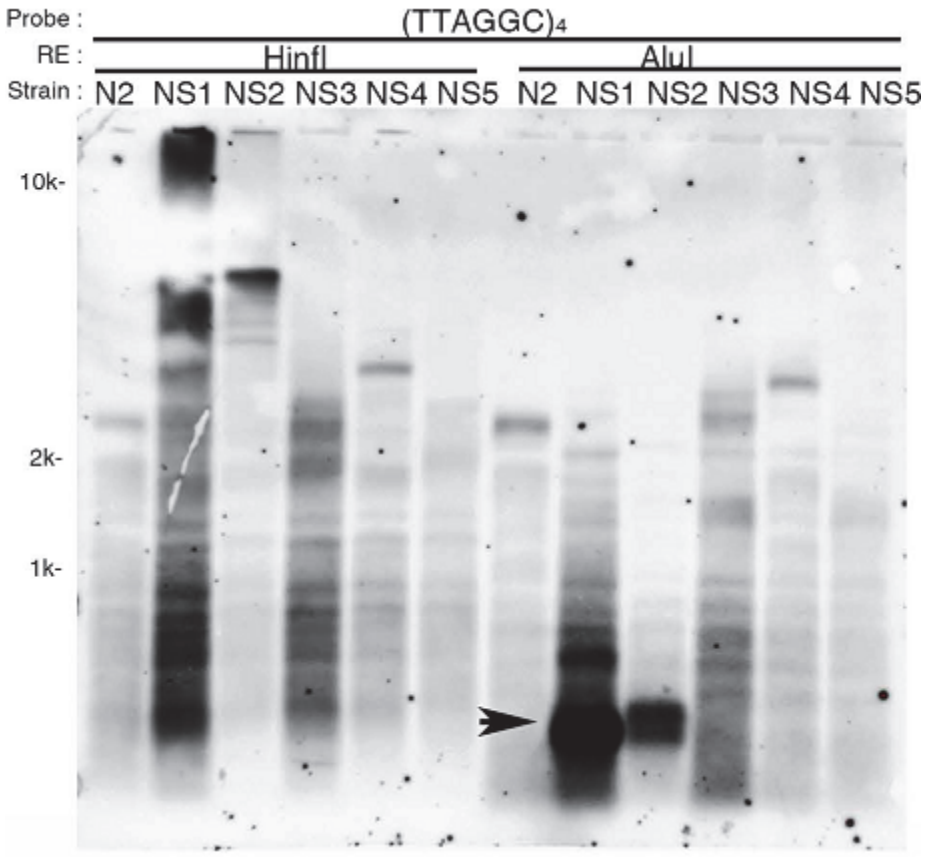


Figure 28. N2 survivors utilized another TALT for ALT. TRF analysis for telomere of N2 survivors. Genomic DNA was digested with indicated enzymes. The blot was hybridized with (TTAGGC)*4 repeat. Arrowhead indicates cleaved TALT 2. In collaboration with C. Kim.

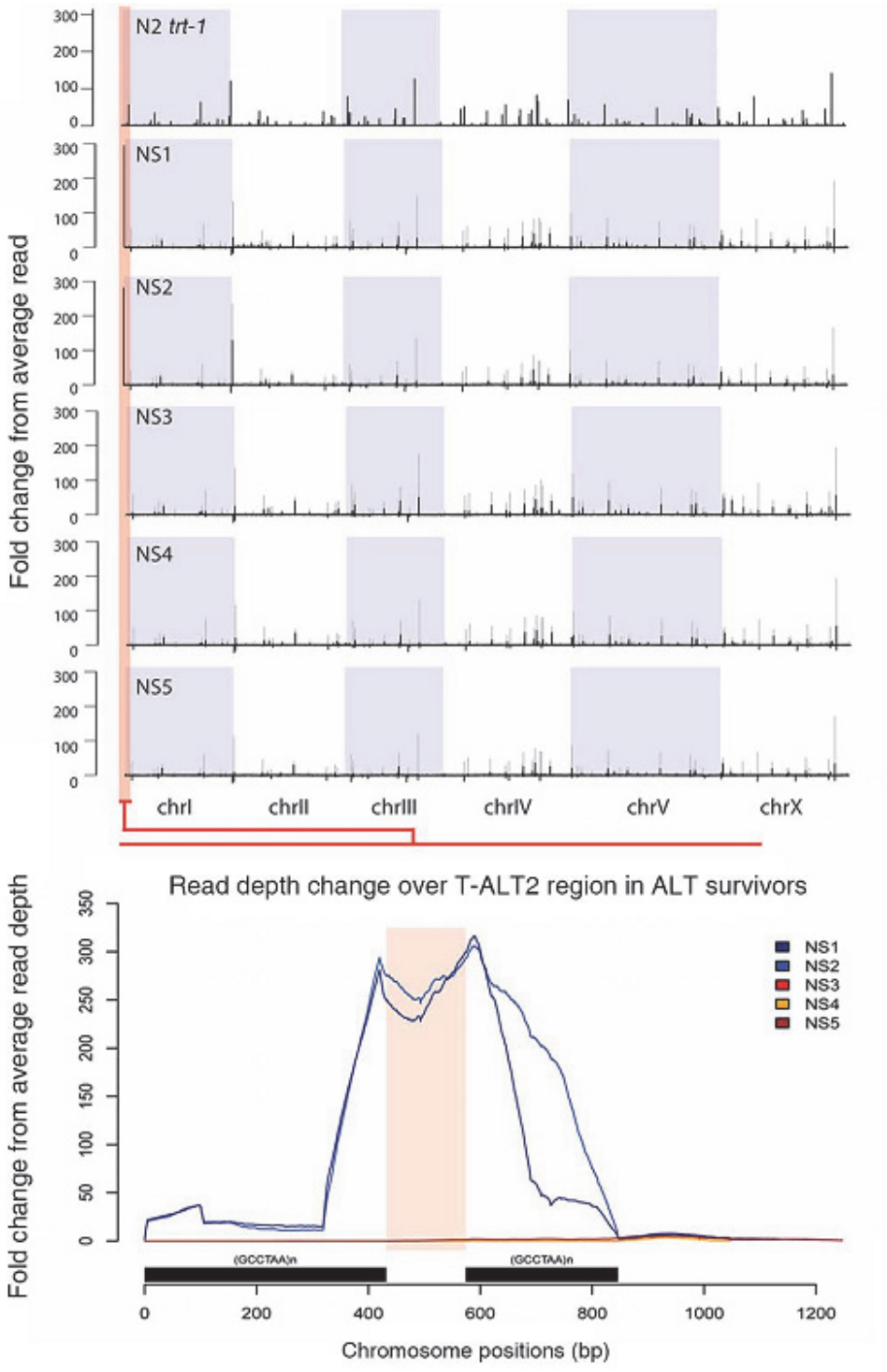

Fold change from average read
N2 trt-1
NS1
NS2
NS3
NS4
NS5
300
200
100
0
chrI
chrII
chrIII
chrIV
chrV
chrX
Read depth change over T-ALT2 region in ALT survivors
Fold change from average read depth
NS1
NS2
NS3
NS4
NS5
(GCCTAA)n
(GCCTAA)n
0
50
100
150
200
250
300
350
0
200
400
600
800
1000
1200
Chromosome positions (bp)

Figure 29. Fold-change read depth plot of N2 trt-1 and N2 survivors (NS1 to NS5) in genome scale. To normalize strain specific variation, fold-change from average depth from parental N2 trt-1(grey bar) was subtracted from ALT survivors. The bottom panel is a blow-up image near the TALT on chromosome I. Internal telomere repeat is indicated as black bar. In collaboration with M. Hills.

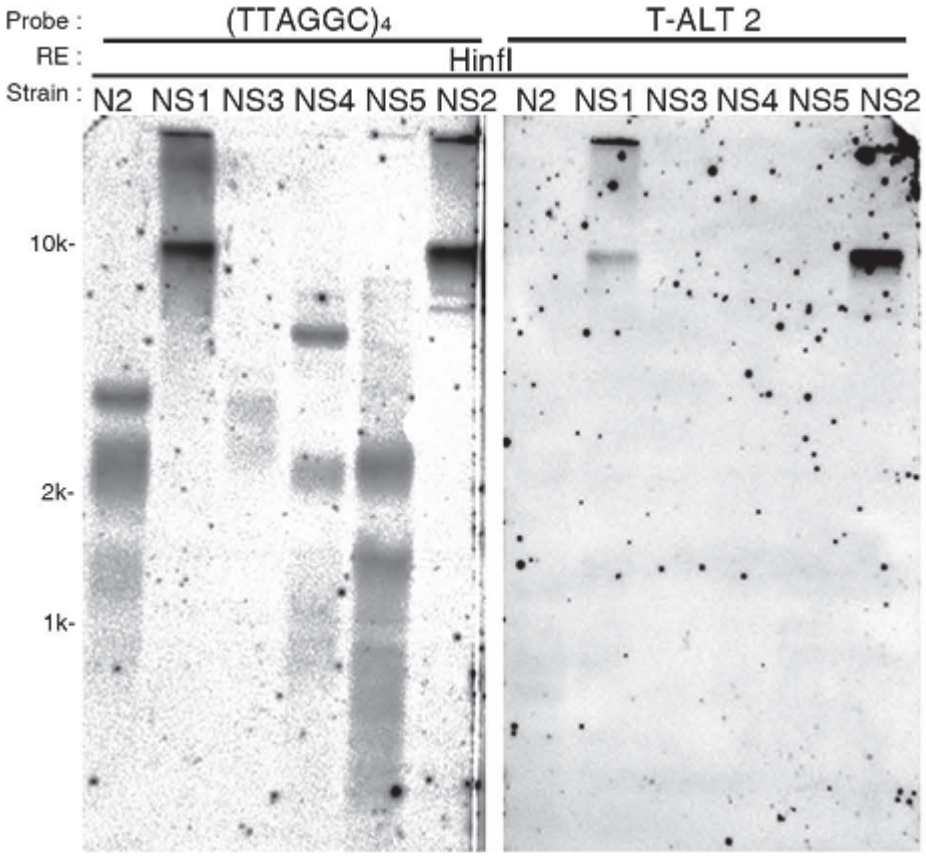


Figure 30. TRF analysis for telomere of N2 survivors. DNA was digested with *Hinf*I and hybridized with (TTAGGC)*4 probe. The same blot was stripped and reprobed with TALT 2 specific probe. In collaboration with C. Kim.

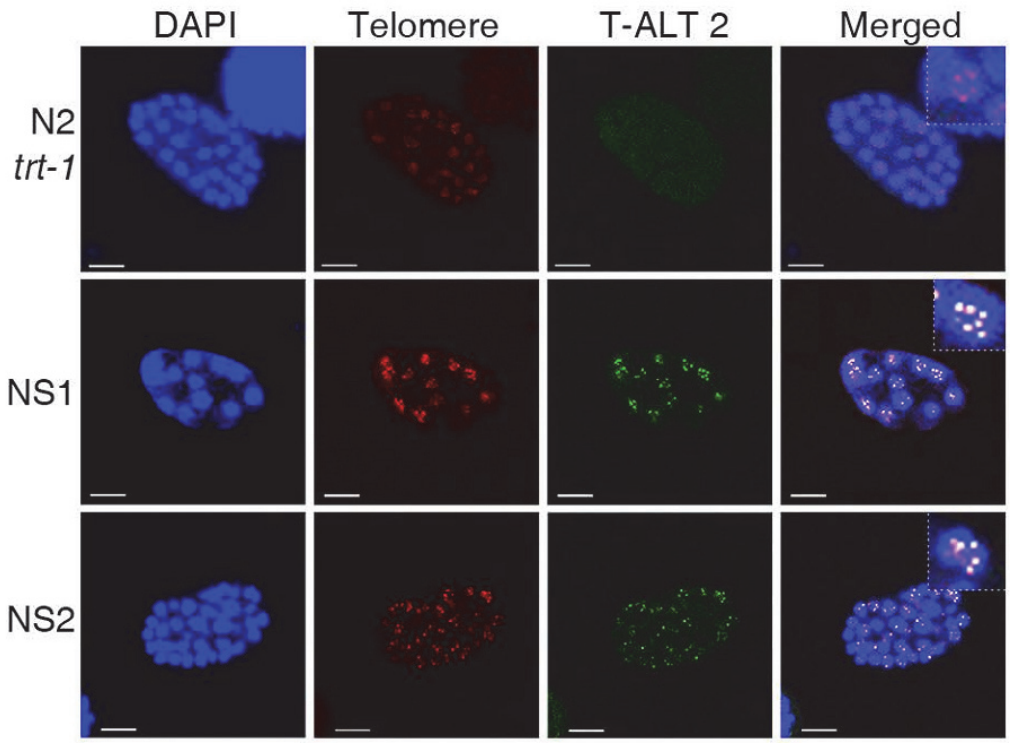


Figure 31. Amplified TALT 2 colocalized to telomere. FISH using the TALT 1 probe and telomere probe in embryonic stage. blue, DAPI; green, TALT 2; red, telomere**.**

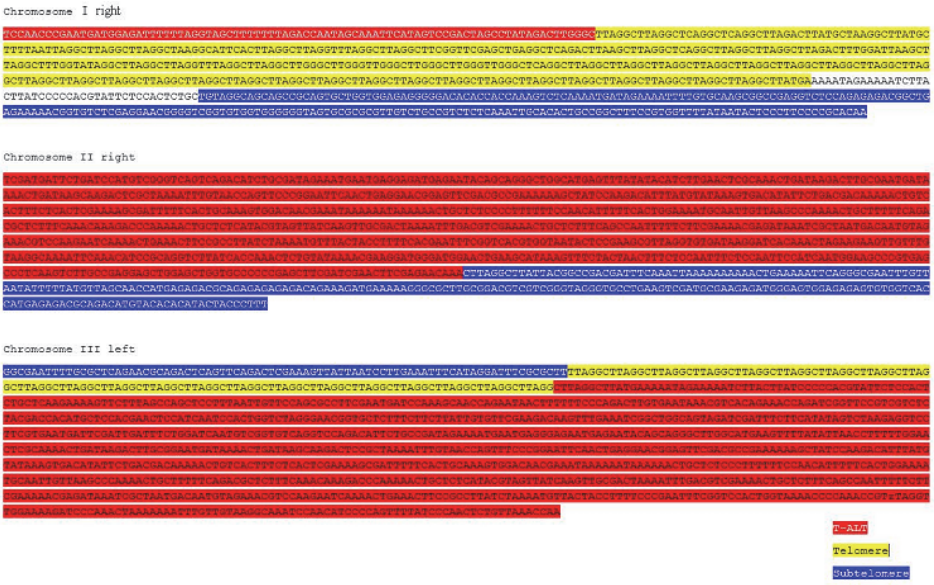


Figure 32. TALT 2 locus was replicated to chromosome ends in NS1 survivor. Sequence analysis of chromosome end and TALT 2 junction by PCR. Upper panel show right arm of chromosome V. Middle panel show right arm of chromosome X. Lower panel show left arm of chromosome II. Red: TALT 2, Yellow: telomere repeats, Blue: chromosome end. In collaboration with S. Sung and E. Kim.

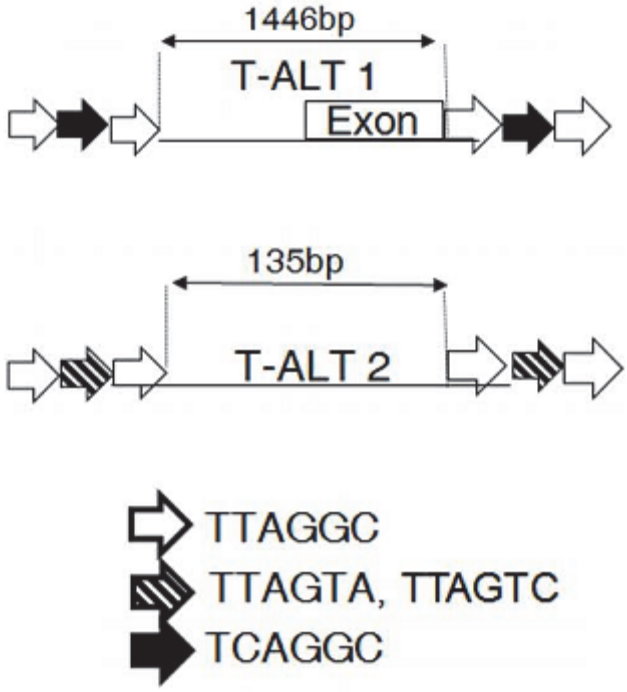


Figure 33. TALT are surrounded by telomere and variant telomere repeat.

Diagram that shows TALT structure.

Figure 34. TALT 2 locus was replicated to chromosome ends in NS1 survivor. Sequence analysis of chromosome end and TALT 2 by BLAST search.

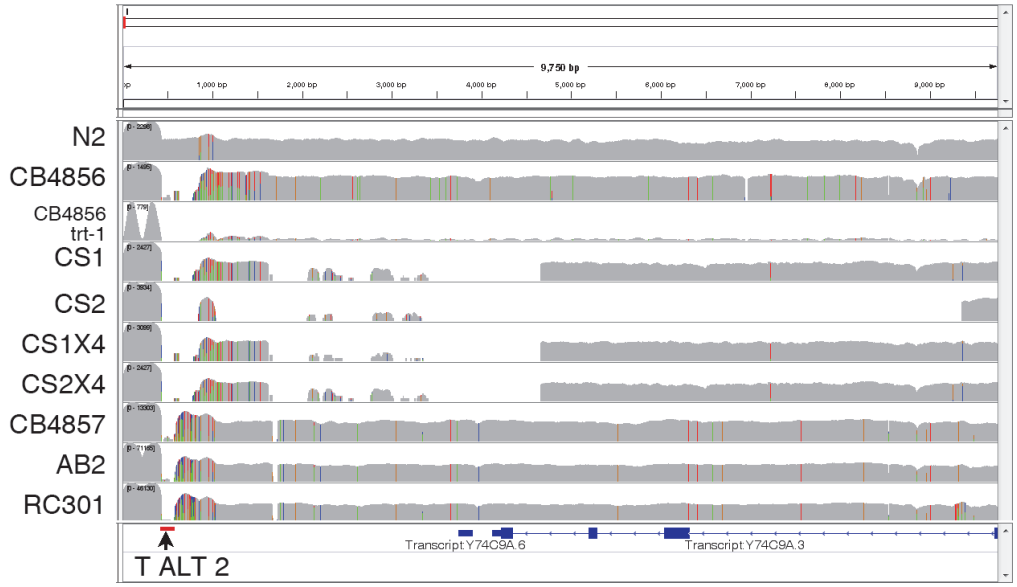


Figure 35. CB4856 has no TALT 2 reservoir in ends of chromosome I. Whole genome sequencing coverage plot in Chr I(1~10kb) indicates TALT 2 is deleted in CB4856, CB4856 *trt-1*, CS survivors, CB4857, AB2 and RC301 strain.

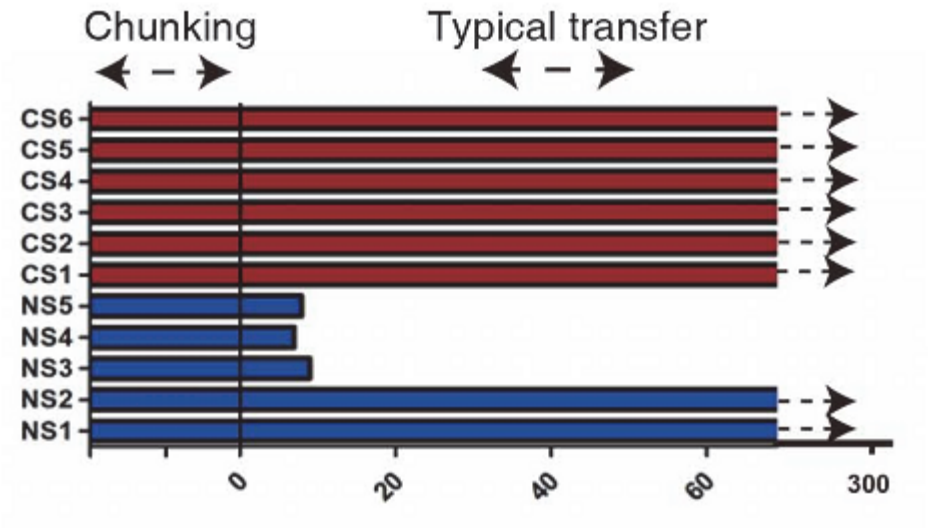


Figure 36. Survivors using either TALT 1 or TALT 2 can be stably maintained. In collaboration with C. Kim.

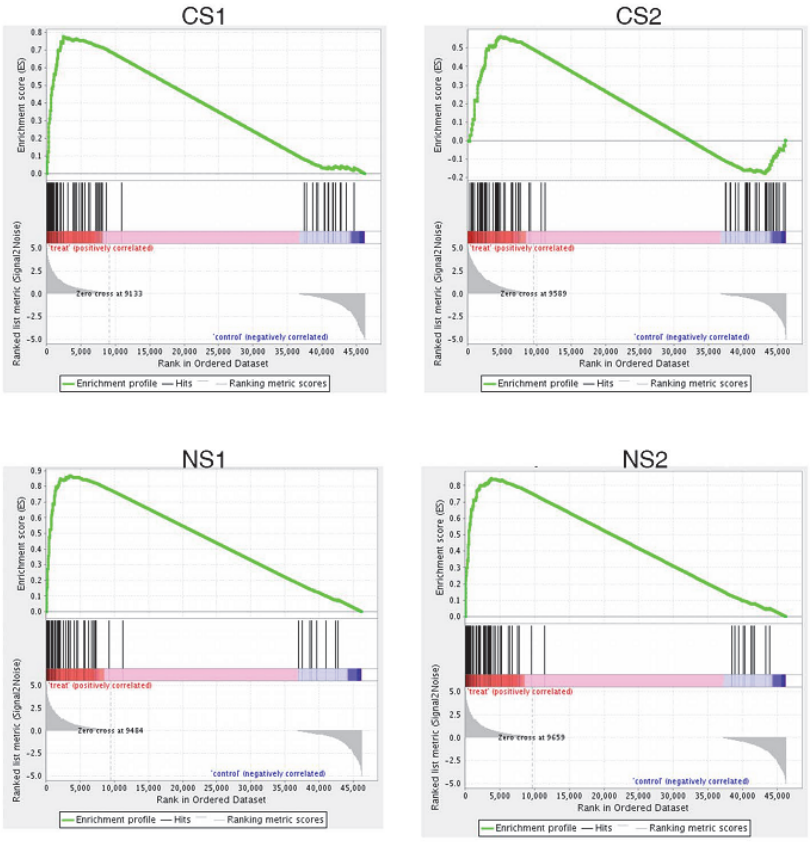


Figure 37. Gamma-ray responsive gene set was enriched in all alt survivors. GSEA results of CB4856 survivors(upper) and N2 survivors (bottom). In collaboration with H. Oh and M. Choi.

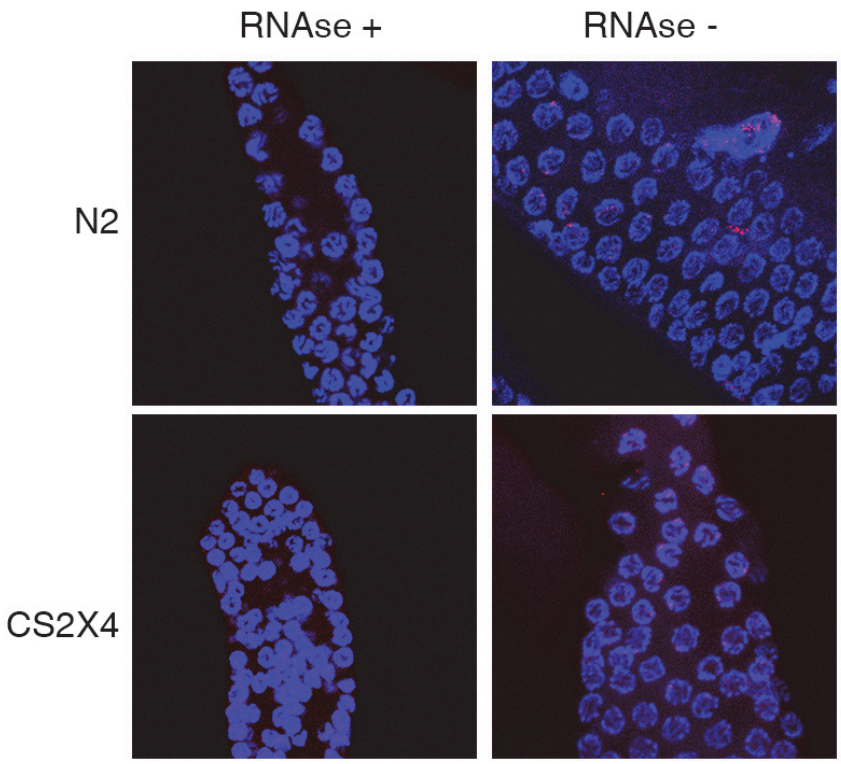


Figure 38. TERRA was transcribed in both wild type and CS2 survivor.

FISH was performed in native condition to detect telomeric RNA. The telomere RNA signal was observed in pachytene stage of meiotic germline and abolished after RNAse treatment.

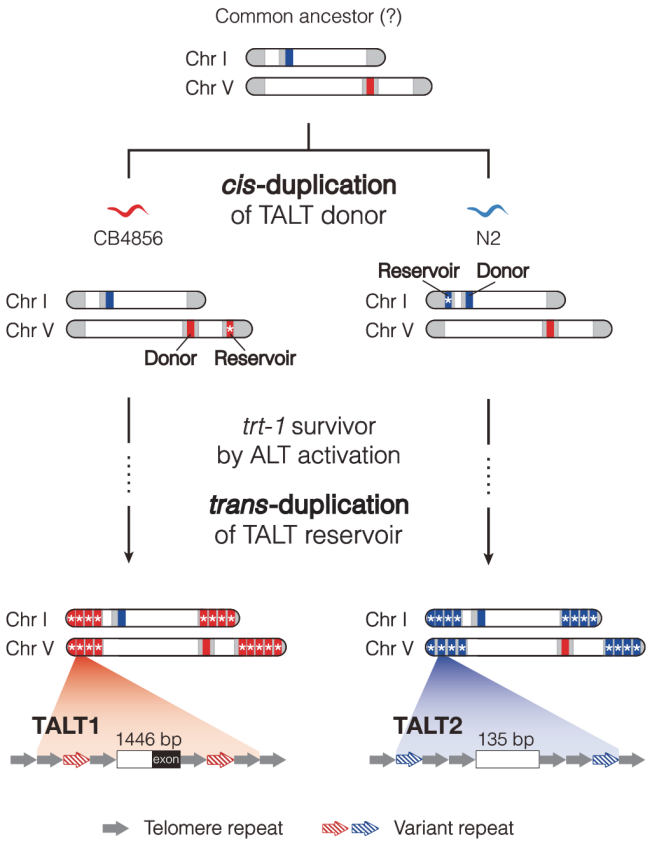


Figure 39. A working model. After divergence from the hypothetical common ancestor, cis-replication of TALT donor into telomere already occurred independently in nature. After telomeric crisis, TALT reservoir in subtelomere could be used as template for telomere maintenance by trans-replication. In collaboration with C. Kim and D. S. Lim

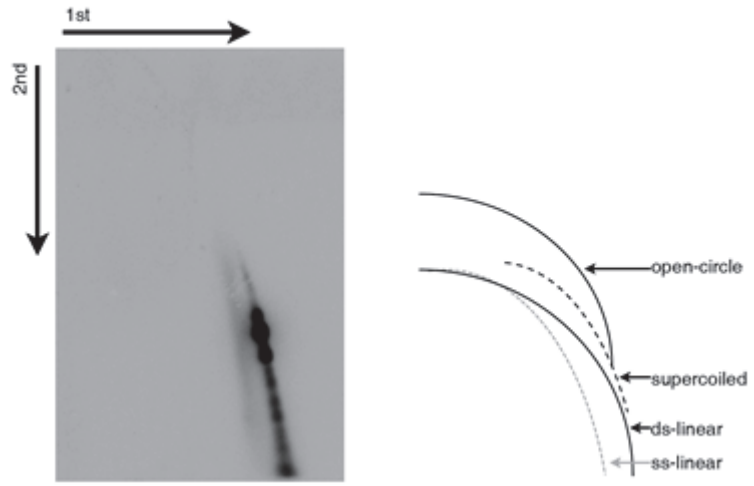


Figure 40. Double strand and single strand telomere DNA was detected in CS3 survivor. Left panel shows 2D gel electrophoresis for detecting circular DNA. First, the *Hinf*I digested genomic DNA of CS3 survivor was fractunated by molecular weight. In second dimension, the DNA was electrophoresed in the presenece of ETBR. The membrane was hybridized with (TTAGGC)*4 probe.

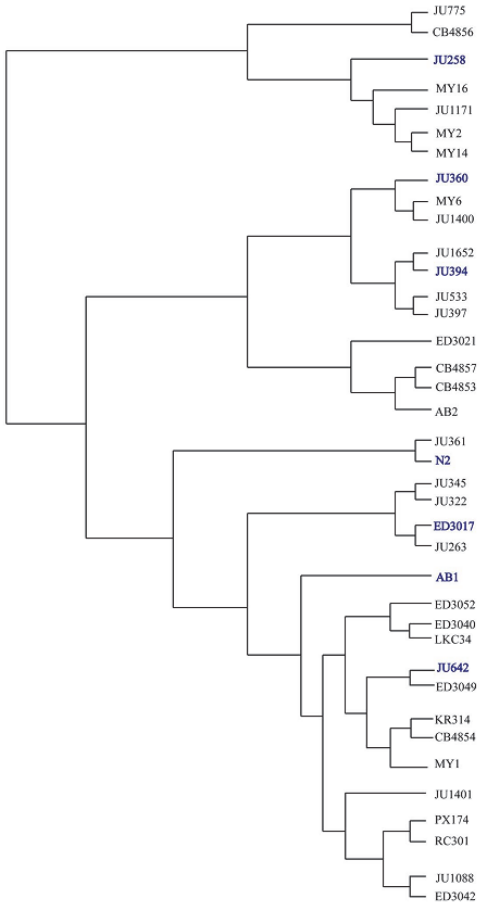


Figure 41. TALT donors are not an extraordinary example. The figure shows the phylogenic tree of 38 wild strains whose genome were fully sequenced. The blue color-coded species contain N2 type TALT on chromosome I. In collaboration with C. Kim.

Table 1. The list of contigs constructed using telomere read-containing reads.

Chromosome	Location	Sequence	Confirmed by PCR?
IR	15071743 ~ 15072072	CGGCACTAGTACTATGCGCCGCGAGACA CACACTACCATCACCAACACAGCCCCCT GAGACATCACGTCTCAGACCAACTGTCA CCCCCCTGACACATATAGGAGTGTGCGG GGAGGGAGTATTATAAAAACCACGGAA AGCCGGCAGTGTGCAATTTGAGAGACGG CAGACAACGCGCGCACTACCCCCCACCA CACCGACCCCCTTCCTCGAGACACCGTT TTTCTCAGCCGTCTCTCTGGAGACCTCGG CCGCTTGCACAAAATTTTCTATCATTTTG AGACTTTGGTGGTGTGTCCCCCTCTCCAC CAGCACTGCGGCTGCTGCCTACAGCAGA GTGGAGAATACGTGGGGGATAAGTAAG ATTTTTCTATTTTTCATAAGCCTA	O
IIR	15255618 ~ 15256023	CGAACTTCGAGAACAAACTTAGGCTTAT TACGGCCGACGATTTCAAATTAAAAAAA AAACTGAAAAATTCAGGGCGAATTTGTT AATATTTTTATGTTAGCAACCATGAGAG ACGCAGAGAGAGAGACAGAAAGATGAA AAAGGGCGCTTGCGGACGTCGTCGGGTA GGGTGCCTGAAGTCGATGCGAAGAGATG GGAGTGGAGAGAGTGTGGTCACCATGA GAGACGCAGACATGTACACACATACTCT ACGTCTCTCTGAGAAGAGCCGCTGTGCT ACTACTATGGGTGCAATAGTATGGAGAA	O

		GATGAGAGATGAACCTTTTTAGATTAGA ATATTTTGAAAACGTGGGATTCGTTTTTA AATAGATTTTACTGAAATATTGCTGCGA AATTTCCATCAATTTTTTTTCTTTTTTGGA AAAACACAAAAATCAAACTACGGTATA AGGTTGCTAACATAAAAATATTAACAAA TTCGCCCTGAATTTTTCAGTTTTTTTTTTA ATTTGAAATCGTCGGCCGTAATAAGCCT AAG	
IIIL	119 ~ 332	ATTTATCGATTCTTATCGATTTTCTTCTCT TTCCGAACCTTTTCGGAATCAAAAAGCC GTGAATCCATAGATTCCGTGCTTTCTCAG ACTTTTCGAGGCCTAATTTTGGTCGAAA AGCCCGATTTTATTTTTATCTGATTTCCG ACTCTTTTCGGACTCAAAAAGTTATAAA TCCTTGAAATTTCATAGGATTTCGCGCTT TTAGGCTTAGGCT	O
VR	20923658 ~ 20923995	AACGGACAAAAACTTTAAAAACCTCCTG TAAAAGAACCTGGGCTGAGCAAAAGCA CATGAAAATTTTTTGAATAAAATATCAT ACTTCATACACCATATACCCATAAAGTA ATAAGAATCTAATGGCAACAGAGGATAC TGGTAAGAATCTAATGGCGACAGAGGAT ACTGGGAATTTTTATCTGTAACTTATTGG CATCCTTAATAGCATAAAGGTTTATTGA ATGTTTTCAAGCTATTTTGAAAATTCTTT AATTCCGCGAAGTTTCTTGATTTCTAAAT TACTGAACAAAAACCAGCTACTTGCGTA AGGTACTGAACCACTTTTCTGGGGATTT	X

		GCTTAGGCTTAGGCTTAGGCTTAGGCTT AGGCTTAGGCTTAGGCTTAGGCTTAGGC TAAGGCTTAGGCACAGGCTCAGGCTTAG	
XR	17718698 ~ 17718815	CCAAAGAACAAAAAAAGAAATTAAAAT ATTTATTTTGCTGTGGTTTTTGATGTGTG TTTTTTATAATGATTTTTGATGTGACCAA TTGTACTTTTCCTTTAAATGAAATGTAAT CTTAAATGTATTTCCGACGAATTCGAGG CCTGAAAAGTGTGACGCCATTCGTATTT GATTTGGGTTTACTATCGAATAATGAGA ATTTTCAGGCTTAGGCTTAG	X

* Blue, Subtelomere sequences; Red, TALT sequence, and Yellow, Telomere repeat

Table 2. The list of RNAi-subjected genes, predicted to have a role in ALT

Gene	Predicted roles
atl-1	DSB response
atm-1	DSB response
brc-1	Recombination
brc-2	fanconi anemia pathway
ceh-37	Telomere binding protein
cep-1	P53 homolog
chk-2	Meiotic recombination
com-1	Meiotic recombination
D1081.9	Meiotic recombination
dna-2	helicase
dog-1	helicase
drh-3	fanconi anemia pathway
exo-3	AP endonuclease
F55A12.10	Meiotic recombination
fanci-1	fanconi anemia pathway
fcd-2	fanconi anemia pathway
hel-308	helicase
him-17	DSB formation
him-18	Homologous recombination
him-3	Meiotic recombination
him-6	BLM helicase
him-8	Meiotic recombination
hmg-5	Telomere binding protein
hpl-2	Heterochromatin protein 1

hrp-1	Telomere binding protein
ku-80	Non-homologous end joining
lin-35	RB homolog, tumor suppressor
mlh-1	mismatched repair
mre-11	DSB response
msh-2	mismatched repair
msh-6	mismatched repair
mus-81	endonuclease

Gene	Predicted roles
plp-1	Telomere binding protein
pms-2	mismatched repair
pot-1	Telomere binding protein
prom-1	Meiotic recombination
rad-50	Homologous recombination
rad-51	Homologus recombination
rad-54	Homologous recombination
rfs-1	Homologous recombination
rpa-1	RPA, replication
rpn-1	protease
rtel-1	helicase
set-11	Histone methyltransferase
set-25	Histone methyltransferase
spo-11	DSB formation
top-3	Topoisomerase IIIa
vhp-1	MAPK
wrn-1	WRN helicase

xpf-1	ERCC1/XPF endonuclease
Y39B6A.16	Meiotic recombination
zhp-3	Meiotic recombination

References

References

Andersen, E.C., Gerke, J.P., Shapiro, J.A., Crissman, J.R., Ghosh, R., Bloom, J.S., Félix, M.-A., and Kruglyak, L. (2012). Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nature genetics *44*, 285-290.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nature structural & molecular biology *20*, 1199-1205.

Barnes, T., Kohara, Y., Coulson, A., and Hekimi, S. (1995). Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. Genetics *141*, 159-179.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. Genetics *77*, 71-94.

Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A., and Reddel, R.R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nature medicine *3*, 1271-1274.

Cangiano, G., and La Volpe, A. (1993). Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. Nucleic acids research *21*, 1133-1139.

Carrel, A., and Ebeling, A.H. (1921). Age and multiplication of fibroblasts. The Journal of experimental medicine *34*, 599-623.

Cesare, A.J., Kaul, Z., Cohen, S.B., Napier, C.E., Pickett, H.A., Neumann, A.A., and Reddel, R.R. (2009). Spontaneous occurrence of telomeric DNA damage response in the absence of chromosome fusions. Nature structural & molecular biology *16*, 1244-1251.

Cesare, A.J., and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. Nature reviews genetics *11*, 319-330.

Cheng, C., Shtessel, L., Brady, M.M., and Ahmed, S. (2012). *Caenorhabditis elegans* POT-2 telomere protein represses a mode of alternative lengthening of telomeres with normal telomere lengths. Proceedings of the National Academy of Sciences *109*, 7805-7810.

Conomos, D., Pickett, H.A., and Reddel, R.R. (2013). Alternative lengthening of telomeres: remodeling the telomere architecture. Frontiers in oncology *3*.

Conomos, D., Reddel, R.R., and Pickett, H.A. (2014). NuRD–ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. Nature structural & molecular biology *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.

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.

Fasching, C.L., Neumann, A.A., Muntoni, A., Yeager, T.R., and Reddel, R.R. (2007). DNA Damage Induces Alternative Lengthening of Telomeres (ALT)–Associated Promyelocytic Leukemia Bodies that Preferentially Associate with Linear Telomeric DNA. Cancer research *67*, 7072-7077.

Greiss, S., Schumacher, B., Grandien, K., Rothblatt, J., and Gartner, A. (2008). Transcriptional profiling in C. elegans suggests DNA damage dependent apoptosis as an ancient function of the p53 family. BMC genomics *9*, 334.

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.

Im, S.H., and Lee, J. (2003). Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode *Caenorhabditis elegans*. FEBS letters *554*, 455-461.

Im, S.H., and Lee, J. (2005). PLP-1 binds nematode double-stranded telomeric DNA. Molecules and cells *20*, 297-302.

Joeng, K.S., Song, E.J., Lee, K.-J., and Lee, J. (2004). Long lifespan in worms with long telomeric DNA. Nature genetics *36*, 607-611.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Kim, S.H., Hwang, S.B., Chung, I.K., and Lee, J. (2003). Sequence-specific binding to telomeric DNA by CEH-37, a homeodomain protein in the nematode *Caenorhabditis elegans*. Journal of Biological Chemistry *278*, 28038-28044.

Kipreos, E.T. (2005). C. elegans cell cycles: invariance and stem cell divisions. Nature Reviews Molecular Cell Biology *6*, 766-776.

Lackner, D.H., Raices, M., Maruyama, H., Haggblom, C., and Karlseder, J. (2012). Organismal propagation in the absence of a functional telomerase pathway in *Caenorhabditis elegans*. The EMBO journal *31*, 2024-2033.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-

2079.

Lowden, M.R., Flibotte, S., Moerman, D.G., and Ahmed, S. (2011). DNA synthesis generates terminal duplications that seal end-to-end chromosome fusions. Science *332*, 468-471.

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

Lydeard, J.R., Jain, S., Yamaguchi, M., and Haber, J.E. (2007). Break-induced replication and telomerase-independent telomere maintenance require Pol32. Nature *448*, 820-823.

Marciniak, R.A., Cavazos, D., Montellano, R., Chen, Q., Guarente, L., and Johnson, F.B. (2005). A novel telomere structure in a human alternative lengthening of telomeres cell line. Cancer research *65*, 2730-2737.

McEachern, M.J., and Haber, J.E. (2006). Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem *75*, 111-135.

Meier, B., Clejan, I., Liu, Y., Lowden, M., Gartner, A., Hodgkin, J., and Ahmed, S. (2006). trt-1 is the *Caenorhabditis elegans* catalytic subunit of telomerase. PLoS genetics *2*, e18.

Olovnikov, A.M. (1973). A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. Journal of theoretical biology *41*, 181-190.

Poon, S.S., Martens, U.M., Ward, R.K., and Lansdorp, P.M. (1999). Telomere length measurements using digital fluorescence microscopy. Cytometry *36*, 267-278.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Raices, M., Verdun, R.E., Compton, S.A., Haggblom, C.I., Griffith, J.D., Dillin, A., and Karlseder, J. (2008). *C. elegans* Telomeres Contain G-Strand and C-Strand Overhangs that Are Bound by Distinct Proteins. Cell *132*, 745-757.

Rogers, C., Reale, V., Kim, K., Chatwin, H., Li, C., Evans, P., and de Bono, M. (2003). Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1. Nature neuroscience *6*, 1178-1185.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L.,

Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

Appendix I

Telomere maintenance through recruitment of internal genomic regions

(Accepted for publication in Nature Communications)

Telomere maintenance through recruitment of internal genomic regions

Beomseok Seo[1,#], Chuna Kim[1,#], Mark Hills[2,#], Sanghyun Sung[1], Hyesook Kim[1], Eunkyeong Kim[1], Daisy S. Lim[1], Hyun-Seok Oh[3], Rachael Mi Jung Choi[3], Jongsik Chun[3], Jaegal Shim[4] and Junho Lee[1,5*]

[1]Department of Biological Sciences, Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Korea, [2]Terry Fox Laboratory, BC Cancer Agency, Vancouver V5Z 1L3, Canada, [3]Department of Biological Sciences, Bioinformatics Institute, BIO-MAX, Seoul National University, Seoul, Korea, [4]Research Institute, National Cancer Center, Goyang, Gyeonggi, Korea, [5]Department of Biophysics and Chemical Biology, Seoul National University, Seoul, Korea.

# These authors equally contributed to this work.

* To whom correspondence should be addressed:
Junho Lee
Dept. Biological Sciences and Dept. Biophysics and Chemical Biology
105-319, Institute of Molecular Biology and Genetics (IMBG)
Seoul National University
Gwanak-ro 1, Seoul, Korea 151-742
82-2-880-6701
elegans@snu.ac.kr

Abstract

Cells surviving crisis are often tumorigenic and their telomeres are commonly maintained through the reactivation of telomerase, but surviving cells occasionally activate a recombination-based mechanism called alternative lengthening of telomeres (ALT). Here we establish stably maintained survivors in telomerase-deleted *Caenorhabditis elegans* that escape from sterility by activating ALT. ALT survivors cis-duplicate internal genomic regions to chromosome ends and subsequently trans-duplicate these regions across the telomeres of all chromosomes. These 'Template for ALT' (TALT) regions consist of a block of genomic DNA flanked by telomere-like sequences, and are different between two genetic background. We establish a model that an ancestral duplication of a donor TALT region to a proximal telomere region forms a genomic reservoir ready to be incorporated into telomeres upon ALT activation.

Introduction

The end-replication problem imposes a replicative limit for linear chromosomes[1], and dis-regulated replication beyond this limit leads to crisis and catastrophic cell death. Rare cell survivors are able to overcome crisis by regaining a telomere maintenance mechanism but often show chromosomal instability and are ultimately tumorigenic. These cells maintain telomere length most commonly through the reactivation of telomerase[2] but occasionally activate the recombination-based mechanism alternative lengthening of telomeres (ALT)[3]. Although specific genes are reportedly important for the maintenance of ALT[4-6], it is largely unknown how specific tumors activate ALT rather than telomerase in maintaining telomere length.

To investigate the mechanism underlying ALT activation, we used a genetic model of the nematode *C. elegans* in which the telomerase gene *trt-1* is deleted[7] such that telomere-mediated sterility occurs after 14-18 generations. In this model system, survivors that are maintained beyond the expected number of generations can arise only through a telomerase-independent ALT mechanism, enabling us to investigate the ALT mechanism by genetically preventing telomerase activation.

We show that the stably maintained survivor lines utilize an internal genomic region as a template for telomere lengthening. We define two such 'Template for

ALT' (TALT) regions, both of which share a common sequence structure consisting of a block of genomic DNA flanked by telomere-like sequences. We also show these regions have *cis*-duplicated to telomeric ends, forming reservoirs which are incorporated into telomeres upon ALT activation. The TALTs we report here represent a novel feature of ALT, whereby internal genomic regions are co-opted to protect chromosome ends in the absence of telomerase. We also discuss the evolutionary implications of our findings.

Results

Isolation of stable ALT survivors of *C. elegans*

A deletion allele of the telomerase gene *trt-1*, *trt-1(ok410)*, was introduced into wild-type N2 and CB4856 worm backgrounds, strains that show an extensive divergence in single nucleotide variants (SNVs)[8]. We treated these worms with a DNA alkylating agent ethyl methane sulfonate (EMS), and after 8 generations, isolated survivors (Fig. 1a and Supplementary Fig. 1a, c and d). We isolated 6 independent survivors (named as 'CS1' to 'CS6') from 80 EMS-treated plates of CB4856 *trt-1* worms, while 5 survivors were isolated from 200 EMS-treated plates of N2 *trt-1* worms (named as 'NS1' to 'NS5'). All 6 CS lines and 2 of the NS lines (NS1 and NS2) were maintained with no gross phenotype for at least 300 generations by transferring 10 to 15 individuals every generation, suggesting that these ALT lines are stably maintained. The remaining three NS survivors (NS3 to

NS5) were maintained only by massive transfer of worms every generation (Supplementary Fig. 1b). The chromosome number of all the ALT survivors decreased due to chromosomal fusions, suggesting that telomeres were critically shortened before ALT was established (Supplementary Fig. 1e).

Similar to ALT in humans, the telomere lengths of CS survivors were longer than in the starting strains when assessed by both florescent *in situ* hybridization (FISH) and terminal restriction fragment (TRF) analysis as well as whole genome sequencing (WGS) analysis (Fig. 1b, c and Supplementary Fig. 2a, 5a). The telomere signal was abolished after BAL 31 exonuclease treatment (Supplementary Fig. 2b), suggesting that they were located at chromosomal ends. Surprisingly, unlike the starting CB4856 *trt-1* strain, all CS survivor lines showed discrete banding patterns on a TRF analysis when the *Hinf*I restriction enzyme was used (Fig. 1d and Supplementary Fig. 5b). Since the recognition sequence for this enzyme does not cut the canonical *C. elegans* telomere repeat, these data suggested that additional sequences were interspersed within the telomere. We therefore hypothesized that CS survivors were able to maintain genomic integrity by lengthening their telomeres with sequences different from the canonical telomeres.

Telomere maintenance without mutations in CB4856 survivors

Since we used EMS, a known potent mutagen, to induce ALT survivors, we identified all specific mutations to find those responsible for the ALT phenotype in the CB4856 *trt-1* background. To do this, we analyzed the whole genome sequences of the CS1 and CS2 survivor lines, before and after multiple rounds of outcrosses with N2 *trt-1* worms to exclude genetic variations unrelated to the ALT phenotype. None of the variants that had been present in the initial isolates were maintained after outcrosses in the CS1 and CS2 survivor lines, suggesting that no point mutation was responsible for ALT activation (Supplementary Fig. 3a, b). In addition, we were unable to induce ALT by RNAi of the genes identified as small indels by array comparative genome hybridization (CGH) (Supplementary Fig. 3c, d). Combined, while we cannot rule out the possibility that there was an ALT-initiating mutation that was subsequently not required for maintenance of ALT, it is more likely that survivors acquired potentially more complex genomic events, which were responsible for ALT induction after EMS treatment.

Telomere maintenance by TALT1 in CB4856 *trt-1* survivors

To identify specific genomic regions that are closely associated with ALT induction from the starting strain of CB4856 *trt-1*, we employed an association mapping strategy. We reasoned that if any specific genomic region conferred the ALT phenotype, the CB4856 *trt-1* survivors should retain this region, even after extensive outcrosses with genetically distinct N2 *trt-1* worms. We therefore

identified all CB4856-specific regions that were not erased by successive outcrosses with N2 *trt-1* in the genomes of both CS1 and CS2 survivors. We found a single region on chromosome V that contained the only two CB4856 *trt-1*-derived SNVs retained after multiple outcrosses, suggesting this region is tightly associated with the ALT phenotype both in CS1 and CS2 survivors (Fig. 2a).

Further analysis of the WGS data of CS1 and CS2 survivors showed that a specific genomic region of 1,464 bp (Chr V:19,229,626-19,231,090) was over-represented in the CS1 and CS2 survivors as high as 200 fold above the average sequence depth (Fig. 2b). This element, which we termed 'Template for ALT 1 (TALT1)', remained the most amplified region even after extensive outcrosses (Fig. 2b and Supplementary Fig. 4a). In addition, the relative copy number of TALT1 increased after outcrosses compared to that of initial survivors, suggesting that TALT1-containing telomeres are actively replicated in the survivors. The genomic structure of TALT1 consisted of a block of uniquely-mapping sequence containing the promoter and the first exon of a coding gene, T26H2.5, flanked by two interstitial telomere-like sequences (ITS), present in the same orientation (Fig. 2c). TALT1 was found to be amplified in all 6 CS survivors (Supplementary Fig. 4b). In order to determine the locations of the multiple copies of TALT1 in the genome of the survivors, we extracted all paired-end sequence reads that mapped to TALT1, and used overlapping mate pairs to construct contiguous fragments ranging from 200 bp to 1.7 kb long.

We generated five contigs that were spanned both TALT1 and specific chromosome ends (Fig. 3a and Supplementary Table 1). PCR reactions using primers specific to TALT1 with primers specific to these chromosomal ends showed that amplicons were produced only from the survivor lines (Fig. 3b and Supplementary Table 2). We further confirmed these data with FISH experiments, where a TALT1 probe co-localized perfectly with a telomere probe only in CS survivors (Fig. 3c). Finally, we performed TRF analyses using enzymes lacking restriction sites in TALT1 and compared them to enzymes which cut within the TALT1 sequence. Using both a telomere probe and a TALT1 probe, we assessed the telomere lengths from the TRFs and found they were longer when TALT1 was uncut (Fig. 3d and Supplementary Fig. 5c), conclusively showing that TALT1 is present in multiple copies in the telomeres of the survivors. We further validated that TALT1 was tandemly integrated into the telomeres of CS survivors by PCR amplifying between TALT1 elements (Supplementary Fig. 4c).

Telomere maintenance by TALT2 in N2 *trt-1* survivors

The discovery of TALT1 in the CS survivors prompted us to explore whether additional TALT regions existed in the N2 *trt-1*-derived ALT survivors that we isolated. Similar to the CS survivors, we found that NS1 and NS2, but not NS3 to NS5, contained longer telomeres than the starting strain, and produced discrete bands on TRF blots using restriction enzymes that do not cut canonical telomere

sequences (Fig. 4a and Supplementary Fig. 5d). These bands were markedly different from identically digested CB4856-dervied survivor TRFs, suggesting that while the NS survivors had incorporated specific sequences into their telomeres, these sequences were not TALT1 (Supplementary Fig. 6a, b). WGS analysis of NS1 and NS2 revealed that both lines had amplification of a specific genomic region adjacent to the chromosome I L telomere (Fig. 4b). We named this region 'Template of ALT 2' (TALT2). A probe specific to TALT2 co-localized to telomeres only in NS1 and NS2 survivors, assessed both by TRF and FISH analyses (Fig. 4a, c), strongly suggesting that TALT2 was enriched at chromosomal ends in the NS1 and NS2 survivors. PCR experiments using primers specific to TALT2 and chromosomal ends validated these findings (Supplementary Table 2). PCR using just TALT2 primers amplified between TALT2 regions and demonstrated that they were arranged tandemly in the telomeres of NS survivors (Supplementary Fig. 4d). The genomic structure of TALT2 is similar to that of TALT1, in that the it consists of a unique DNA fragment flanked by telomeric sequences (an ITS at one end and telomere sequences at the other) (Supplementary Fig. 6c).

Altered double-strand break responses in ALT survivors

The *C. elegans* strain deficient in the telomerase gene was reported to overcome sterility without telomere lengthening only when maintained as large populations, presumably due to stochastic telomere shortening events occurring

within the population[9]. In stark contrast to these data, most survivors in this study incorporated TALT sequences into their telomeres and were stably maintained by means of a typical transfer in which a few worms were transferred to a new plate for culture each generation (Supplementary Fig. 1b). Only the NS3 to NS5 lines did not show TALT amplification (Supplementary Fig. 4e), and these were maintained only by massive transfer, became sterile within 10 generations (Supplementary Fig. 1b) and had short telomeres (Fig. 4a). One reason for the apparent discrepancy is that we treated the worms with EMS, and propose that this elevated the DNA damage response in some way. Consistent with this, by comparing the RNA-seq analysis of the survivors across specific gene expression datasets showed that genes significantly altered in ALT survivors correlated with the expression profile of genes induced by gamma radiation (Supplementary Fig. 7)[10]. This suggests that *C. elegans* ALT survivors maintain altered double-strand break responses, similar to human ALT cells[11].

ALT mechanism involves TALT duplication in *cis* and in *trans*

The mechanism by which TALT regions are utilized at ALT telomeres, was elucidated when we analyzed the SNVs in this region. We identified two heterozygous SNVs within TALT1 common to both the starting CB4856 strain and the CS survivors. The SNVs were inherited together as two haplotypes, and interestingly, they were unequally represented, with the haplotype normally

associated with CB4856 present 5-fold higher than the other haplotype, which is commonly associated with the N2 strain (Fig. 5a, Supplementary Fig. 8a-c). This skewing was more pronounced in the CS survivor lines, with a haplotype shift such that haplotype common to CB4856 was highly overrepresented compared to the other (Fig. 5a). For CS1, reads containing the N2-associated haplotype were represented at genome-average sequence depth (0.9-fold), while reads containing the CB4856-associated were increased 76-fold above average depth. Similarly, reads from CS2 represented the N2-associated haplotype at average levels (1.2-fold), while reads containing the other were greatly enriched (122-fold) (Fig. 5a). Since TALT1 had been amplified at least 100-fold in these lines (Fig. 2c), we reasoned that the SNVs we identified were actually sequence variation between two discrete loci, with the overrepresented CB4856-associated alleles present on the amplified TALT1. To test this we performed PCR across the internal TALT1 on chromosome V, which revealed only the N2-associated alleles represented at genome-average level. Further analysis revealed that the overrepresented CB4856-associated allele were located at the right telomere of chromosome V in only the CB4856 strains and the survivors, but not in the N2 strain (Fig. 5b-d, Supplementary Fig. 5e). We used these variants to distinguish the two locations of TALT1: the internal TALT1 donor (TALT1 D) and the TALT1 reservoir (TALT1 R) at the proximal telomere region.

Similar to the observed TALT1 duplication, we determined that TALT2 also originated from an endogenous locus (TALT2 D) close to the left arm of chromosome I and had duplicated adjacent to the telomere of chromosome I (TALT2 R) (Fig. 6a, Supplementary Fig. 8d). Interestingly, the TALT2 R locus is undetectable in the CB4856 genome (Fig. 6b and Supplementary Fig. 5f, 8d), suggesting that CB4856 and N2 have evolved different TALT reservoirs for ALT. Using variant repeats of TALT2 R and TALT2 D, we were able to identify both locations and found that only TALT2 R sequences were overrepresented in NS survivors, suggesting that the TALT2 R was used as template for ALT (Fig. 6c and Supplementary Fig. 9a, b).

Discussion

Based on these observations, we propose a two-step model for TALT utilization in ALT (Fig. 7). The first step involves a *cis*-duplication of a TALT donor to a chromosomal end to create a TALT reservoir prior to ALT, and the second involves a *trans*-duplication of TALT R to most chromosomes via the ALT pathway. We have demonstrated that this first step event has occurred in wild strains, with TALT from a specific chromosomal 'donor' region being duplicated in *cis* to the end of the same chromosome. This chromosome-adjacent copy (TALT R) acts as a reservoir, so when crisis occurs in individuals, rare survivors recruit

the TALT R sequences to all chromosomal ends (*trans*-duplication), stabilizing their chromosomes.

ALT survivors in this study are similar to type I survivors of *Saccharomyces cerevisiae*, in which Y´elements located to subtelomere are used for telomere lengthening[12]. In humans, it has been shown that a plasmid tag (Tel) integrated within a single telomere duplicated to other telomeres, while a plasmid tag (Subtel) integrated immediately proximal to telomeric DNA did not[13]. While TALT regions are proximal to the telomeric DNA, since they are flanked by an ITS and the telomere, they are more analogous to the plasmid 'Tel' sequences integrated within the telomeres, and behave in a similar way. As such, it is possible that the duplication of regions adjacent to the telomere is an evolutionarily conserved mechanism. Consistent with this, subtelomeric regions in humans rapidly obtained unique sequences during primate evolution[14,15]. Likewise, we observed in *C. elegans* that a sequence near the proximal telomere, TALT2, appears to be rapidly changing, as it has been independently duplicated in multiple wild isolates (Supplementary Fig. 10a). We therefore propose that a rapidly evolving proximal telomere region retains highly replicative potential so that it can be preferentially selected as a template for ALT at telomere crisis. Bioinformatic analysis identified more TALT-like sequences in the *C. elegans* reference genome, and it would be of interest to examine the wild isolates that lack TALT1 and TALT2 to assess whether

they utilize other TALT sequences for ALT (Supplementary Fig. 10b and Supplementary Table 3).

Although we have shown for the first time that internal genomic regions can integrate into the telomeres, resulting in elongation and maintenance of telomeric ends in the absence of functional telomerase, the mechanism by which internal TALT regions are recruited and replicated to the telomeric ends by cis- or trans-duplication still remains unknown. Unfortunately, we were unable to identify a single factor that suppressed the ALT phenotype in our survivors using a candidate RNAi approach (Supplementary Table 4). Furthermore, while C-circle formation is a well-known marker of human ALT[16], we did not observe any increase in C-circle formation in our survivors (Supplementary Fig. 11), suggesting that a different mechanism is involved. It is possible that TALT-mediated recombination and unequal sister chromatid exchange plays a role in the duplication to telomeres on different chromosomes[13] and expansions within telomere tracts from the same chromosome[17]. Without knowing TALT-binding proteins at this moment, it is also difficult to explain how TALT sequences are involved in the capping of chromosomal ends. It is possible that telomere repeats in the TALTs normally function for telomere capping and that telomere variants or non-telomeric sequences found in TALTs may recruit trans-acting proteins that do not interfere telomere capping function completely while simultaneously make the telomeres recombinogenic by their protein interaction. This situation could be similar to the

proposed 'intermediate state telomeres' of ALT cell lines in which the ALT telomeres block chromosome fusions, but not the DNA damage response that can induce recombination[11]. It is also possible that TALT sequences recruit proteins involved in heterochromatinization.

Our discovery of TALT may also help understanding of the mechanism of human ALT. As human ALT cells often incorporate variant telomere repeats that likely originate from proximal telomeres[18,19] and AG11395 ALT cell line has a repeat unit of SV40 DNA with a structure similar to TALT[20,21], it is conceivable that human ALT cells also incorporate TALT elements into their telomeres. Therefore, it would be worthwhile to screen human ALT cell lines and tumors to investigate the presence of a TALT-like mechanism involved in telomere elongation in humans. Our results suggest the presence of TALT R in proximal telomeres is an important DNA signature of ALT activation or initiation.

Methods

C. elegans strains and culture

Worms were cultured at 20°C under standard culture conditions[21]. The following strains were used in this study: Bristol N2 wild strain, Hawaiian CB4856 wild isolate, *trt-1(ok410)* I[7]. N2 *trt-1(ok410)* was outcrossed with Hawaiian CB4856 wild isolate to produce CB4856 *trt-1(ok410)*. To ensure the ALT was

activated in outcrossed progeny, F2 worms were grown at least 20 generations and worms with low fecundity were excluded. To maximize the outcrossing effect, SNVs of all chromosome markers were checked. The *trt-1(ok410)* mutation was confirmed by PCR and WGS (Supplementary Fig. 1c, d). The *trt-1(ok410)* strain was outcrossed with N2 wild type to produce an early generation of N2 *trt-1(ok410)*.

EMS treatment

Synchronized L4 worms were treated with 50 mM EMS in M9 buffer for 4 hours. After 4 hours of recovery, treated P0s were allowed to lay eggs for 12 hours. F1 worms were isolated by removing P0 worms. Initially ~100 F1 eggs were transferred to fresh plate, then from the F2 generation onwards, 10~15 worms were transferred manually at every generation.

Feeding RNA interference

E. coli HT115 expressing dsRNA were grown in LB with 1 mM ampicillin at 37°C overnight and seeded on to NGM plates containing 1 mM IPTG and 1 mM ampicillin. At every generation, 10-15 L1 larvae were transferred to fresh RNAi media plates.

Telomere florescent *in situ* hybridization (FISH)

As cells are highly dividing in embryo stage in *C. elegans*, it is most plausible condition that ALT is activated. Eggs were isolated by bleaching adult worms. Eggs were fixed in 2% paraformaldehyde (PFA). Tubes containing eggs were frozen in liquid nitrogen and thawed in warm water twice in order to crack the eggs.

Eggs were settled on a slide coated with poly-lysine. The slide was washed 3 times with phosphate-buffered saline containing 0.1% Tween-20 (PBST) to remove residual PFA. The slide was incubated in acetone and methanol for 5 minutes each at -20°C and was then rehydrated in 2X SSC (0.3 M NaCl, 0.03 M sodium citrate) containing 0.1% Tween-20. The slide was blocked for 1 hour with prehybridization solution (3X SSC, 50% formamide, 10% dextran sulfate, 50 µg/mL heparin, 100 µg/mL yeast tRNA, 100 µg/mL salmon sperm DNA) at 37°C. PNA-$(TTAGGC)_3$ probe was hybridized for 16 hours in humid chamber at 37°C. Slides were washed twice in wash buffer (2X SSC and 50% formamide) for 15 minutes at 37°C. After washing 3 times with PBS-T, slides were counter-stained with DAPI and mounted with anti-bleaching solution Vectashield (Vector Laboratory). The samples were imaged using a confocal microscope (LSM200, Zeiss).

TALT1 probes were labeled with digoxigenin (DIG) further visualized by FISH. After FISH, slides were blocked with PBS-T containing 5% BSA for 1 hour at room temperature. Slides were stained with rhodamine conjugated anti-DIG antibody for 3 hours. After washing in PBS-T twice, slides were mounted and observed as described above. The telomere signal was quantified using TFL-TELO software (Dr. Peter Lansdorp, Terry Fox Laboratory, Vancouver).

Telomere southern blot

For genomic DNA preparation, worms were harvested and washed 5 times in M9 buffer. Worms were lysed in lysis buffer for 8 hours (100 µg/mL proteinase K,

50 mM KCl, 10 mM Tris pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween-20, 1% beta-mercaptoethanol) DNA was extracted using phenol:chloroform extraction and ethanol precipitaion. DNA in TE buffer was treated with RNAse (10 μg/mL) for 2 hours and re-extracted, before being dissolved in TE buffer.

For Southern hybridization, 5 μg of DNA was treated with 1 unit of restriction enzyme and then separated by gel electrophoresis either using standard equipment or Pulsed Field Gel Electrophoresis equipment. Gels were blotted by capillary transfer on to the Zeta probe membrane (Bio-Rad) overnight. The membranes were cross-linked using a UV cross-linker and hybridized with the Southern probe in DIG Easy Hybridization buffer at 42°C for 16 hours. The membrane was then washed twice at room temperature in 2X SSC, 0.1% SDS and twice at 42°C 0.2X SSC, 0.1% SDS. The DIG-labeled probe was detected on an ImageQuant LAS-4000 biomolecular imager (GE healthcare) using an anti-DIG-AP antibody chemiluminescence detection kit (Roche). The $(TTAGGC)_{30}$ probe was labeled with DIG-UTP by PCR-amplifying telomere sequences cloned in a T-easy vector. Probes for TALT1 and TALT2 were labeled with DIG-UTP using primers targeting unique region in TALTs. Uncropped scans of blots were supplied in Supplementary Fig. 5.

C-circle assay

Worm genomic DNA was digested with TALT1 non-cutting restriction enzyme mix (*Nhe*I, *Bam*HI, *Dra*I, *Apa*I, *Nde*I, *Xho*I, *Nco*I, *Sac*I 4 units/μg each).

Restricted DNA was purified by Phase Lock Gel (5 PRIME) and ethanol precipitation. 10 µL of sample was mixed with 10 µL 0.2 mg/mL BSA, 0.1% Tween-20, 1 mM each dATP, dGTP, dTTP, dCTP, 1X phi29 buffer with or without 7.5 units phi polymerase (NEB) and incubated at 30°C for 8 hours then at 65°C (polymerase inactivation) for 20 minutes. For dot blotting, sample were diluted with 60 µL 2X SSC and blotted onto nylon membrane. DNA was crosslinked with UV onto the membrane and hybridized at the 62°C with $(GCCTAA)_4$-digoxigenin probe in DIG easy hybridization buffer (Roche). C-circle amplified signal was detected by DIG detection kit (Roche) according to the manufacturers' instructions. *pot-1(tm1620)* and *pot-2(tm1400)* were used as positive controls as they are reported to contain elevated levels of C-circles compared to N2 wild type[23,24]. For sample loading confirmation, we stripped by 0.2 M NaOH, 2% SDS and re-hybridization with $(GCCTAA)_4$-digoxigenin probe (denatured blot).

BAL 31 exonuclease treatment

5 µg of DNA was treated with 10 units of BAL 31 at 30°C in 1X BAL 31 buffer. Reactions were stopped with the addition of EGTA (25 mM final concentration). DNA was collected by ethanol precipitation and digested with non-cutting restriction enzyme mix (*Nhe*I, *Bam*HI, *Dra*I, *Apa*I, *Nde*I, *Xho*I, *Nco*I, *Sac*I).

Sequence analysis of telomere-proximal telomere region junctions

To analyze the exact sequence of the junction of telomere and proximal telomere region, PCR products of junctions were sequenced. For PCR reactions,

the forward primer was designed against telomere-adjacent DNA to elongate into the telomere and reverse primer was designed from TALT1 specific sequence (the first exon of T26H2.5). CS1 genomic DNA and N2 genomic DNA were used as template. 30 PCR cycles were performed with primers annealing at 60°C and elongation progressing for 3 minutes. After electrophoresis in 1% agarose gel, the CS1-specific amplicon was gel-extracted and sequenced. Primers used for sequencing were the same as those used for PCR reactions.

Whole genome sequencing

DNA was fragmented to 300 bp and sequencing libraries were constructed with an average insert size of 430 bp using standard Illumina protocols. Libraries were run on an Illumina Hiseq2000 sequencing platform.

Variant Discovery

To analyze variant induce by EMS and CB4856 associated variant, WS243 version of *C. elegans* reference genome and annotation data were acquired from Wormbase website (www.wormbase.org). For preprocessing the raw sequencing data, Trimmomatics was applied so that illumina adapter sequences were removed[25]. 3′ and 5′ ends of low quality reads were trimmed. Preprocessed reads were then aligned to *C. elegans* reference genome (WS243) with the Burrows-Wheeler Aligner software using default parameters (version 0.7.5a)[26]. Before calling variants, Picard's SortSam and MarkDuplicates and GATK's indel realignment[27] and base quality score recalibration (BQSR) was applied. For the

indel realignment, very stringent parameters were used due to the probability of considerable genetic difference between the reference assembly and survivor lines. From reference annotation (WS243), SNVs and indels sourced from the million mutation project were selected and used in BQSR. Variant discovery and genotyping were performed across all samples simultaneously using GATK's haplotype caller as per GATK Best Practices, then standard hard filtering was applied with minimum raw depth of coverage of 4.

TALT1 contig generation

Since TALT1 is amplified from CB4856 allele (TALT1 R) that is absent from genome assembly (based on the N2 genome), we reconstructed TALT1 R region through *de novo* assembly of CB4856 *trt-1* sequencing data using CLCworkbench7 (Qiagen) with default setting. Using BWA, reads of CS1 were collected if only one of the pair was mapped on the reconstructed TALT1 reservoir. Collected reads were used to generate contig sequences, which were then aligned to N2 genome using BLAST.

Whole genome sequencing analysis for telomere reads

WGS data was aligned to the ce10/WBcel215 reference assembly using bwa[26] and converted into bam files using Samtools[28]. Bam files from different worm strains were analyzed for reads predicted to derive from within telomeres using previously described software (http://sourceforge.net/projects/motifcounter/)[18]. Briefly, reads which contained at least 6 canonical telomere repeats (TTAGGC)

were extracted and saved into separate files. The number of reads was counted and normalized to total read density to infer telomere length. The saved files were subsequently used to analyze split-pair regions and for variant calling.

TALT bioinformatic analyses

To identify regions that were associated with the telomere, files containing telomeric reads were used to extract their associated paired-ends. Reads in which both pairs were telomeric were excluded, leaving only reads in which one pair mapped to a genomic location. In addition to subtelomeric regions and loci flanking interstitial telomeres, the majority of the reads mapped to TALT1 in the case of CS1 and CS2, and TALT2 in the case of NS1 and NS2.

Average genomic coverage across the genome was calculated using genomecov from the BedTools package[29], after the removal of duplicates and low mapping quality ($q<20$) reads in Samtools. The read depth was assessed at every nucleotide in the genome and normalized by dividing by the average coverage to give a fold-change metric for read depth. The fold-change read depth from the parental strains was subtracted from the ALT survivors to account for strain-specific variations across the genome. Line plots of genome-wide and locus-specific read depth changes were created in R. Putative additional TALT regions were identified by analyzing the *C. elegans* assembly with the Homer scanMotifGenomeWide function[30] for $(TTAGGC)_7$ with up to 3 mismatches to find all genomic ITS elements. The distance between each ITS was calculated, and data

was extracted if this distance was between 100-2,000 bp. All candidate regions were manually confirmed, the 5´ and 3´ ITS length recorded, and any genes noted.

Telomere variant calling

Telomere variants were analyzed from files containing telomere reads. The number of each variant type was counted within this subset of reads from each library using the program motif_counter[18]. The presence of each variant repeat was assessed within this subset of reads, and their frequency was calculated and used to generate pie charts.

Measurement of TALT copy numbers

TALT copy numbers were measured from worm total genomic DNA by using quantitative real time PCR (qPCR). We calculated TALT copy number normalized by a single copy gene, *act-1*. Each reaction performed on the iCycler iQ5 (Bio-Rad) using following thermal profile: 95°C for 3 minutes; 40 cycles of 95°C for 20 seconds, 60°C anneal for 20 seconds, 72°C extend for 30 seconds along with 81 cycles of melting curve from 60°C to 95°C. The reaction components are as follows : 10 µL 2X SYBR Green mix (Bio-Rad), 1 µL each of 10 µM forward and reverse primers (Supplementary Table 5), 7 µL sterile water and 1 µL total genomic DNA (100 ng/µL) in 20 µL reaction

mRNA sequencing analysis

To align the preprocessed RNA-seq reads to reference genome and estimate the expression level of transcripts, we followed the Tuxedo pipeline (Tophat2 and

Cufflinks pipeline)[31]. First, Tophat2 mapped RNA-seq reads to the *C. elegans* genome using given transcript annotation (WS243)[32]. Next, using cuffquant and cuffnorm software in Cufflinks package[33], we estimate the abundance of known transcripts in normalized RPKM across all samples.

Gene set enrichment analysis (GSEA)[34]

The gene ontology (GO) annotation data for *C. elegans* were downloaded from the Wormbase website (www.wormbase.org). Gene sets that change upon irradiation was obtained from ref. 10. For all pairs of ALT samples and a wild type, Gene Set Enrichment Analysis (GSEA) was performed using the gene set data obtained above.

References

1 Olovnikov, A. M. A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. Journal of theoretical biology 41, 181-190 (1973).

2 Shay, J. & Bacchetti, S. A survey of telomerase activity in human cancer. European journal of cancer 33, 787-791 (1997).

3 Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. Nature reviews genetics 11, 319-330 (2010).

4 Heaphy, C. M. et al. Altered telomeres in tumors with ATRX and DAXX mutations. Science 333, 425-425 (2011).

5 Lovejoy, C. A. et al. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. PLoS genetics 8, e1002772 (2012).

6 Conomos, D., Reddel, R. R. & Pickett, H. A. NuRD–ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. Nature structural & molecular biology 21, 760-770 (2014).

7 Meier, B. et al. trt-1 is the Caenorhabditis elegans catalytic subunit of telomerase. PLoS genetics 2, e18 (2006).

8 Andersen, E. C. et al. Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nature genetics 44, 285-290 (2012).

9 Cheng, C., Shtessel, L., Brady, M. M. & Ahmed, S. Caenorhabditis elegans POT-2 telomere protein represses a mode of alternative lengthening of telomeres with normal telomere lengths. Proceedings of the National Academy of Sciences 109, 7805-7810 (2012).

10 Greiss, S., Schumacher, B., Grandien, K., Rothblatt, J. & Gartner, A. Transcriptional profiling in C. elegans suggests DNA damage dependent apoptosis as an ancient function of the p53 family. BMC genomics 9, 334 (2008).

11 Cesare, A. J. et al. Spontaneous occurrence of telomeric DNA damage response in the absence of chromosome fusions. Nature structural & molecular biology 16, 1244-1251 (2009).

12 Teng, S.-C. & Zakian, V. A. Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in Saccharomyces cerevisiae. Molecular and cellular biology 19, 8083-8093 (1999).

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

14 Royle, N. J., Baird, D. M. & Jeffreys, A. J. A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. Nature genetics 6, 52-56 (1994).

15 Baird, D. & Royle, N. Sequences from higher primates orthologous to the human Xp/Yp telomere junction region reveal gross rearrangements and high levels of divergence. Human molecular genetics 6, 2291-2299 (1997).

16 Henson, J. D. et al. DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. Nature biotechnology 27, 1181-1185 (2009).

17 Bailey, S. M., Brenneman, M. A. & Goodwin, E. H. Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. Nucleic acids research 32, 3743-3751 (2004).

18 Conomos, D. et al. Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. The Journal of cell biology 199, 893-906 (2012).

19 Lee, M. et al. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. Nucleic acids research 42, 1733-1746 (2014).

20 Fasching, C. L., Bower, K. & Reddel, R. R. Telomerase-independent telomere length maintenance in the absence of alternative lengthening of telomeres–associated promyelocytic leukemia bodies. Cancer Research 65, 2722-2729 (2005).

21 Marciniak, R. A. et al. A novel telomere structure in a human alternative lengthening of telomeres cell line. Cancer research 65, 2730-2737 (2005).

22. Brenner, S. The genetics of Caenorhabditis elegans. Genetics 77, 71-94 (1974).

23. Shtessel, L. et al. Caenorhabditis elegans POT-1 and POT-2 repress telomere maintenance pathways. G3: Genes| Genomes| Genetics 3, 305-313 (2013).

24. Lackner, D. H., Raices, M., Maruyama, H., Haggblom, C. & Karlseder, J. Organismal propagation in the absence of a functional telomerase pathway in Caenorhabditis elegans. The EMBO journal 31, 2024-2033 (2012).

25. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, btu170 (2014).

26. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754-1760 (2009).

27. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, The Genome Analysis

Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20, 1297-1303 (2010).

28. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).

29. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842 (2010).

30. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell 38, 576-589 (2010).

31. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7, 562-578 (2012).

32. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36 (2013).

33. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28, 511-515 (2010).

34. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545-15550 (2005).

END NOTES

Acknowledgments

Mutant worm strains were kindly provided by the *Caenorhabditis* Genetics Center and Dr. Shohei Mitani. The authors thank Dr. H. Lee (Seoul National University) for discussion. The authors thank Dr. D. Moerman for CGH analysis. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number : HI14C1277), and a grant of the Korean Research Foundation (grant number NRF-2009-0094019).

Author Contributions

B.S., C.K., M.H. and J.L. designed and analyzed experiments. B.S., C.K., S.S., H.K., E.K. and J.L. performed the molecular biological experiments. M.H., H.O.,

R.M.C., J.C. and J.S. performed the bioinformatic analyses. D.S.L. drew the schematic images. B.S., C.K., M.H. and J.L. wrote the manuscript.

Accession codes

RNA-seq data of ALT survivors are available in the public database. The accession code is SRA278191.

Competing financial interests

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1. **Isolation of stable ALT survivors in *C. elegans* with distinct telomeric sequences** (**a**) A schematic diagram showing the experimental procedures to isolate stable ALT survivors in two wild isolates. (**b**) Quantitative FISH analysis for telomere length of the survivors CB4856 *trt-1, pot-2(tm1400),* CS1 and CS2. Telomere was detected by Cy-3-TTAGGC*3 PNA probe (red) in the embryo. DNA was counterstained with DAPI (4',6-diamidino-2-phenylindole) (blue). The upper panel shows representative images for each strain. T-test was used for statistical

analysis for quantification (*P-value < 0.0001, n = 388/each). Mean value is represented with red bar. Scale bar, 10 μm. (**c**) TRF analysis of N2, CB4856, *pot-1(tm1620), pot-2(tm1400)* and all the CS survivors digested with a combination of 6-cutter restriction endonucleases (*Nhe*I, *Dra*I, *Apa*I, *Nde*I, *Xho*I, *Nco*I and *Sac*I) and probed with DIG-TTAGGC*4. *pot-1(tm1620)* and *pot-2(tm1400)* were used as positive controls as they have long telomeres. (**d**) TRF analysis of CB4856, CB4856 *trt-1* and all the CS survivors probed with DIG-TTAGGC*4 after *Hinf*I digestion.

Figure 2. Identification of template of ALT in CB4856 backgrounds (**a**) Association heat map of CB4856 SNVs that were retained after extensive rounds of outcrosses. The right panel is an enlargement image near the TALT1 locus on chromosome V. (**b**) Fold-change from average read depth plot (black bar) of CB4856 *trt-1*, N2 *trt-1*, CS1, CS2 and outcrossed CS survivors with N2 (CS1x4-1, CS1x4-2 and CS2x4). To normalize strain specific variation, fold-change from average depth of parental CB4856 *trt-1* was subtracted from total fold-change (grey bar) of ALT survivors. While positive value indicates over-representation of the sequence, negative value indicates under-representation of the sequence compared to control. (**c**) An enlargement of TALT1 peak on chromosome V from **b** (red shade). Internal telomere repeats are indicated by black bars. Gene structure flanking TALT1 locus is denoted by purple bar.

Figure 3. TALT1 is replicated in the telomeres of CB4856 survivors (**a**) Diagram illustrating how the contigs (red bar) were reconstructed from reads in which one pair aligned to TALT1. Asterisks indicate predicted breakpoint by BLAST (black) or confirmed breakpoints (red) (**b**) Confirmation of *trans*-duplication by breakpoint PCR followed by Southern hybridization with TALT1 probe. Breakpoint PCR primers (red arrows) were designed to amplify the junction of TALT1 and each of proximal telomere regions. F, forward primer; R, reverse primer; open arrow, telomere repeat (**c**) FISH using the TALT1 probe and telomere probe in embryonic stage. Scale bar, 10 μm; blue, DAPI; green, TALT1; red, telomere. (**d)** TRF analysis using a TALT1 probe and a telomere probe. DNA was digested with indicated enzymes and hybridized with indicated probes. RE, restriction endonuclease; C, CB4856 *trt-1*; S, CS1 survivor; H, *Hin*dIII; B, *Bam*HI; Nc, *Nco*I; Nd, *Nde*I; Nh, *Nhe*I; S, *Sac*I; Mix, mixure of *Hin*dIII, *Bam*HI, *Nco*I, *Nde*I, *Nhe*I and *Sac*I.

Figure 4. N2 *trt-1* survivors amplify and incorporate TALT2 in their telomeres (**a**) TRF analysis for telomere and TALT2 of N2 survivors. Genomic DNA of N2 *trt-1* and NS survivors (NS1 to NS5) was digested with *Hin*fI or *Alu*I. The blot was hybridized with (TTAGGC)*4 repeat (upper panel). The same blot was stripped and reprobed with a TALT2 specific probe (bottom panel). Arrowheads indicate

cleaved TALT2. (**b**) Fold-change from average read depth plot (black bar) of N2 *trt-1* and N2 survivors. To normalize strain specific variation, fold-change from average depth from parental N2 *trt-1* was subtracted from total fold-change (grey bar) of ALT survivors. The bottom panel is an enlargement of TALT on chromosome I. Internal telomere repeats are indicated as black bars. (**c**) FISH using the TALT and telomere probe in embryo of CB4856 *trt-1*, N2 *trt-1* and all the survivors. Scale bar, 10 μm; blue, DAPI; green, TALT; red, telomere.

Figure 5. *trans*-duplication of TALT1 reservoir to other chromosomal ends in CB4856 backgrounds (**a**) Fold change from average coverage of SNVs (WBvar00067712 and WBvar00067717) from WGS data of N2 *trt-1*, CB4856 *trt-1*, CS1 and CS2. N2-associated haplotype (blue) and CB4856-associated haplotype (red) are plotted separately. Due to large difference of value, Y-axis is segmented. (**b**) 'CB4856-derived' SNVs were located at the right end of chromosome V in CB4856, but not in N2. Asterisks indicate CB4856 SNVs. Dashed line indicates genomic position. (**c**) Breakpoint PCR was performed with primer pairs (red arrows in **b**) designed to anneal to the proximal telomere region of chromosome V (VR) and TALT1 in forward (For in **b**) or reverse (Rev in **b**) orientation. CB, CB4856. (**d**) Nucleotide change of WBvar00067712 and WBvar00067717 in internal and telomeric TALT1. N2-associated alleles are blue while CB-associated alleles are red. Nucleotide positions are indicated in **d**.

Figure 6. *trans*-duplication of TALT2 reservoir to other chromosomal ends in N2 backgrounds (**a**) Diagram of TALT2 in N2 genome. Primers used in **b** are indicated as red arrows. Specific variant repeats (asterisks) of TALT donor and reservoir are indicated. (**b**) TALT2 reservoir is absent in CB4856 wild isolate. PCR amplicons from indicated primers were fractionated by gel electrophoresis and stained with ethidium bromide. N2, N2; CB, CB4856. (**c**) Normalized number of variant read count from WGS data of CB4856 *trt-1*, N2 *trt-1* and all the NS survivors. Variant read count was normalized to total sequencing coverage of each sample.

Figure 7. A **model for TALT-mediated ALT** Internal genomic regions with ITS are used as template of ALT in *C. elegans* ALT survivors. The model illustrates that after divergence from the hypothetical common ancestor, *cis*-duplication of TALT donor into telomere already occurred independently in nature. After telomeric crisis, the TALT reservoir at a proximal telomere region is used as a template for telomere maintenance.

국문초록

예쁜꼬마선충의 새로운 염색체말단 유지기작에 관한 연구

서범석

서울대학교 생명과학부

원핵생물과 달리 진핵 생물의 염색체는 선형의 염색체로 유지된다. 염색체의 끝은 텔로미어라는 리보핵산-단백질 복합체라는 구조물로 보호된다. 텔로미어는 텔로미어 반복서열과 텔로미어 결합 단백질로 이루어져 있다. 텔로미어가 짧아져 텔로미어 결합 단백질이 염색체를 보호할 수 없게 되면 세포는 세포분열이 중단될 위기에 처하게 된다. 하지만 줄기세포나 생식세포와 같이 텔로미어 길이를 유지할 수 있는 능력을 획득하게 된다면 세포사를 피할 수 있게 된다. 세포 분열 중단위기를 극복한 세포는 높은 비율로 종양 형성 능력을 갖고 있으며 텔로미어 합성효소를 재활성화하여 텔로미어 길이를 유지한다. 하지만 일부의 암세포는 유전자 재조합 메카니즘을 이용하는 대안적 텔로미어 유지기작인 Alternative Lengthning of telomeres(ALT)를 활성화하여 텔로미어를 유지한다. 이 현상은

효모에서도 발견되어 유전학적으로 유전자 재조합 메커니즘이 ALT 에 관여한다는 것이 알려져 있다. 본 연구에서 나는 예쁜꼬마선충이 텔로미어 합성효소가 결손되면 불임이 되는 점을 이용하여 텔로미어 합성효소가 결손되었지만 ALT 를 작동시켜 불임이 되는 형질을 극복한 개체를 발견하였다. 대안적인 텔로미어 유지개체들은 유전자 배경에 따라 텔로미어 합성연장을 위해 유전체의 두 곳 중 한 곳을 주형삼아 텔로미어를 합성연장하고 있었다. 본 연구에서 이 특별한 유전자좌를 "대안적 텔로미어 유지기작을 위한 주형(TALT) 요소"라 명명하고 이들의 특징을 서술하였다. TALT 는 텔로미어 반복서열과 변형 반복서열에 둘러 싸여 있으며, 이미 야생형에서 염색체 말단에 인접한 곳에 적은 숫자로 복제되어 있었다. 우리가 실험적으로 일으킨 조건하에서 –즉 텔로미어 반복서열 합성효소를 결손시키고 유전적 병목현상을 유도한 조건 –적은 비율로 살아남은 개체들은 텔로미어에 인접한 TALT 를 주형으로 텔로미어를 유지하였다. 본 연구에서 발견한 TALT 요소는 텔로미어가 결손된 상황에서도 텔로미어를 유지하는 새로운 기작을 제시한다. 또한 본 연구는 대안적 텔로미어 유지기작을 이용하는 인간의 암세포의 연구에도 적용될 수 있는 가능성을 지닌다.

주요어 : 예쁜꼬마선충, 텔로미어, 대안적 텔로미어 유지기작, 야생형, 유전자 배경