Conserved and Unique Features of Terminal Telomeric Sequences in ALT-Positive Cancer Cells

Benura Azeroglu[1], Wei Wu[1, #], Raphael Pavani[1], Ranjodh Sandhu[1], Tadahiko Matsumoto[1, ##], André Nussenzweig[1], and Eros Lazzerini Denchi[1*]

1. Laboratory of Genome Integrity, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, USA.

Current address: Key Laboratory of Multi-Cell Systems, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai, China

Current address: Kyoto University, Kyoto, Japan.

* **Correspondence to**:

Eros Lazzerini Denchi (eros.lazzerinidenchi@nih.gov)

Abstract

A significant portion of human cancers utilize a recombination-based pathway, Alternative Lengthening of Telomeres (ALT), to maintain telomere length. Targeting the ALT is of growing interest as a cancer therapy, yet a substantial knowledge gap remains regarding the basic features of telomeres in ALT-positive cells. To address this, we adopted END-seq, an unbiased sequencing-based approach, to define the identity and regulation of the terminal sequences of chromosomes in ALT cells. Our data reveal that the terminal portions of chromosomes in ALT cells contain canonical telomeric sequences with the same 5' terminus bias (-ATC) observed in non-ALT cells. Furthermore, as reported for non-ALT cells, POT1 is required to preserve the precise regulation of the 5' end in cells that maintain telomere length using the ALT pathway. Thus, the regulation of the terminal 5' of chromosomes occurs independently of the mechanism of telomere elongation. Additionally, we employed an S1 endonuclease-based sequencing method to determine the presence and origin of single-stranded regions within ALT telomeres. These data shed light on conserved and unique features of ALT telomeres.

Keywords

Telomere, ALT, ssDNA, END-seq

Introduction

Telomeres are the natural ends of eukaryotic chromosomes and are critical for maintaining chromosome integrity [1]. Human telomeres consist of 5-15 kilobases of repetitive DNA sequences (5'-TTAGGG-3'), which serve as a binding platform for an essential protective protein complex named shelterin [1]. The terminal portion of chromosomes is composed of a single-stranded 3' overhang of approximately 200 nucleotides, generated by the interplay between DNA replication and end-processing, and regulated by the telomere-associated protein POT1 [2, 3]. Noteworthy, the generation of the 3' overhang differs between leading and lagging strand telomeres. At leading strand telomeres, the newly synthesized DNA terminates in a blunt end and requires post-replicative resection by nucleases such as Apollo and Exo1 and a fill-in action promoted by CST complex to create the 3' overhang [4]. In contrast, lagging strand synthesis inherently generates a 3' overhang due to Okazaki fragment processing, where removal of the terminal RNA primer leaves a single-stranded gap that is partially filled in by DNA polymerase α-primase with the aid of the CST complex [5] (for reviews, see [2, 3]). Progressive loss of telomeric repeats during DNA replication leads to telomere shortening and represents a barrier to unlimited proliferation [6-9]. To overcome this tumor suppressive barrier, ~85% of cancers activate telomerase, a reverse transcriptase, to synthesize new telomeric repeats [10, 11]. The remaining ~15% of cancers engage the Alternative Lengthening of Telomeres (ALT) pathway, a recombination-based mechanism to elongate telomeres using existing telomeric DNA as a template [12, 13]. Major advancements in our understanding of the ALT pathway at the molecular level have been made over recent years, shedding light on the molecular pathways involved in telomere elongation as well as the genetic vulnerabilities of ALT-positive cells [14-24] (for reviews, see [25-28]).

Due to the repetitive nature of telomeres, some fundamental questions about the composition of ALT telomeres remain unanswered. Specifically, it is unclear if 5' telomere ends in ALT cells differ from telomerase-positive cells, whether they follow the same rules for end protection, and whether ALT telomeres contain any unique secondary structures. It is foreseeable that telomeres in telomerase-positive cells and ALT cells differ, as ALT can generate complex and non-canonical telomeric sequences, and secondary structures through recombination-based mechanisms. Interestingly, positive staining for the single-strand binding protein RPA and native FISH staining (ssTelo) of telomeres suggested the presence of single-stranded DNA regions in ALT-positive cells [19, 29]. However, the source or extent of ssDNA in ALT telomeres remains

unclear. Intriguingly, EM data and rolling-circle amplification-based assays suggest that ALT frequently contain circular extrachromosomal telomeric DNA termed C-circles [30, 31].

Here, we report a novel approach to studying ALT telomeres using the high-resolution and unbiased method known as END-seq [32]. Using this method, we find that the terminal sequence of ALT chromosomes consists of canonical telomeric repeats, shares the same bias in the 5' terminus as telomeres maintained by telomerase, and terminates with ATC-5'. Furthermore, we demonstrated that this bias is influenced by the shelterin component, POT1.

Secondly, to assess the presence of ssDNA at ALT telomeres, we used S1 nuclease, an enzyme that specifically degrades single-stranded DNA including the single-stranded regions in duplex DNA. Using S1 nuclease treatment coupled with END-seq (S1-END-seq) [33], we show that ALT telomeres have tracts of ssDNA and the presence of ssDNA in telomeres can be used to distinguish ALT-positive from ALT-negative cells.

Results

END-seq efficiently captures 5' chromosome termini

To study the terminal portion of chromosome, we employed an unbiased and high-resolution approach developed to map and investigate double-strand breaks (DSBs): END-seq [32]. In this method, biotin-labeled adaptors are ligated to DSBs, which enable their purification and subsequent sequencing by next-generation sequencing (NGS). Based on the directionality of DNA sequencing (5' to 3'), DSBs are identified as unidirectional reads that originate from the lesion site on either side of the break (Figure 1A). Chromosome ends resemble one-ended DSBs, which by END-seq should be detected as unidirectional reads corresponding to the 5' telomeric strand, the C-rich strand. To test this hypothesis, we performed END-seq on two telomerase-positive human cell lines, HeLa and telomerase expressing RPE1. Sequencing reads containing four consecutive TTAGGG repeats (G-rich) or CCCTAA repeats (C-rich) were defined as "telomeric reads" and were normalized to the total number of reads obtained in each individual library. Strikingly, virtually all (99.9%) telomeric END-seq reads correspond to the C-rich strand, in agreement with the fact that chromosome ends are one-ended DSBs (Figure 1B). To exclude possible artifacts derived from the repetitive nature of telomeric reads, we performed a similar analysis of reads corresponding to pan-centromeric repeats; this analysis shows that although the fraction of their presence in the human genome was similar (Figure S1A), the pan-centromeric sequences are not enriched to the same level as telomeres nor show any strand bias when detected (Figure S1B).

Next, we performed a motif analysis on the C-rich telomeric END-seq reads to determine the prevalent terminal chromosomal sequence. This analysis shows that the vast majority (78% and 65% for HeLa and RPE1, respectively) of chromosomes end with ATC-5' (Figures 1C and 1D). Strikingly, this is in agreement with previous data obtained using a PCR-based approach termed STELA on defined chromosome ends in HeLa cells [34]. A major advantage of END-seq over STELA is that it can be applied to telomeres of any length and chromosome location without the need of specific sub-telomeric primers. This excludes possible biases and allows the analysis of telomeres from any cell type and origin. As a proof of principle, we expanded the analysis to a collection of telomerase-positive human cell lines, as well as mouse and dog cell lines, and mouse tissues. In these experiments, we found that chromosome ends show a strong bias for ATC-5' as a terminal sequencing (Figure 1D), in line with previous findings [35, 36], suggesting a conserved mechanism for its generation.

To further validate END-seq as a method to identify terminal telomeric repeats, we tested whether 5'-3' T7 exonuclease treatment prior to END-seq would be sufficient to change the bias of the ATC-5' sequence (Figure 1E). Since T7 exonuclease preferentially degrades 5' ends of double-stranded DNA while leaving 3' single-stranded overhangs intact, we expected that its treatment would randomize the 5' telomeric ends by progressively resecting them, thereby disrupting the original sequence bias. If END-seq was biased or not sufficiently sensitive, T7-treatment would not significantly alter the detected sequence distribution. As shown in Figures 1F and S1C, T7-treated samples display an altered distribution of telomeric permutations. Quantitative measurement of the difference in distribution of possible telomeric permutations in the presence or absence of T7 treatment was performed using the Kullback-Leibler divergence (KL divergence) score. The result of this analysis indicated that the samples are extremely divergent (●● indicates a KL divergence value between 0.375 and 0.5). To further evaluate the distribution changes between these conditions, we compared the reads obtained from these samples with *in silico*-generated control reads representing various degrees of randomization from the ATC-5' sequences (see Figures S1D and S1E). Using the KL divergence score, we determined that T7-randomized samples were most similar to a population where 100% of the reads were equally randomized. In comparison, in the untreated samples, approximately 70% of the reads correspond to the perfect "canonical" CCAATC-5' sequence. These data indicate END-seq as a method to sensitively detect terminal telomeric repeats while excluding the potential artifacts.

Further, we investigated whether we could detect *in vivo* alterations of the terminal repeat, such as those reported upon depletion of the shelterin component, POT1 [37]. To deplete POT1, we generated a population of HeLa cells harboring a doxycycline-inducible POT1 short hairpin RNA (shPOT1-1) construct. Doxycycline treatment resulted in reduced POT1 expression, confirming efficient knockdown (Figure S1G). In agreement with previously published data, END-seq analysis showed that POT1-depleted cells have a significant reduction in terminal repeats ending with the canonical ATC-5' sequence (Figure 1F). Comparison with control reads generated *in silico* reveals that approximately 45% of reads from POT1-depleted cells are derived from random permutations of the telomeric repeat (Figure S1F).

Lastly, we tested how depletion of either or both POT1 paralogs affected the 5' termini of mouse telomeres. To this end, we engineered mouse embryonic stem cells (mESCs) expressing a FKBP-tagged POT1a protein (POT1a$^{dTAG/dTAG}$) that is degraded when cells are treated with dTAG13 compound (Figures S2A-C). Using this cell line, we then generated cell lines lacking POT1b through CRISPR knockout (POT1a$^{dTAG/dTAG}$ POT1b$^{-/-}$) (Figures S2D-S2F). This strategy allowed us to examine the 5' termini of POT1-proficient ESCs, as well as systematically analyze the effect of either single or double POT1 depletion in a genetically defined background. END-seq analysis revealed that while depletion of POT1a altered 5' termini, depletion of POT1b had no effect (Figures 2A and 2B). Using the KL divergence score, we determined that in the POT1-profient and POT1b-depleted samples, approximately 70% and 85% of reads, respectively, corresponded to the perfect canonical ends. In contrast, in POT1a-depleted samples, this proportion was around 25%, regardless of POT1b status. This striking result highlights a fundamental difference between POT1a and POT1b, demonstrating that POT1a plays a dominant role in maintaining proper 5' end processing, consistent with recently published work [35]. The stark contrast between their effects underscores the distinct functions of these paralogs and reveals an exciting new layer of telomere regulation. Collectively, these data demonstrate the effectiveness of END-seq to study terminal repeats of chromosomes.

The terminal 5' sequence in ALT-positive cells is canonical

Having established that END-seq efficiently captures the terminal repeat of chromosome ends, we sought to investigate whether ALT-positive cells have the same tight regulation of terminal repeats as telomerase-positive cells. We performed END-seq on three ALT-positive cell line: G292, SAOS2 and U2OS. Our analysis shows that in ALT cells, the vast majority of telomeric reads (82-98% across 3 cell lines and 11 biological repeats) correspond to the C-rich strand (Figures 3A-3C), consistent with non-ALT cells and the expected pattern for a one-ended-DSB.

Motif analysis revealed that the majority of telomeric reads in all three tested cell lines start with the canonical ATC-5' sequence (Figures 3D and S3A-S3C). These data suggest that regulation of the terminal 5' of ends of chromosomes occurs independently of the mechanism of telomere elongation. To test this hypothesis, we depleted POT1 in U2OS cells using two different doxycycline-inducible shRNAs (Figure S3D). POT1 depletion resulted in formation of 53BP1 foci (Figures 3E and 3F), consistent with POT1's established role in suppressing the DNA damage response at chromosome ends [37, 38]. Next, we performed END-seq to evaluate the effect of POT1-depletion on the terminal sequence. The data show that, similarly to telomerase-positive HeLa cells, depletion of POT1 dramatically altered the 5' natural termini in ALT-positive cell lines (Figure 3G).

Next, we asked whether ALT cells are more likely to contain variant telomeric repeats (VTRs) in the terminal portions of chromosome ends. VTRs have been shown to occur frequently in telomeric reads isolated from ALT-positive cells [39]. To establish whether VTRs in ALT cells are contained within terminal repeats, we assessed the frequency of non-canonical telomeric repeats within the first 30 nucleotides from the 5' of the Illumina reads from U2OS (1.93%) and HeLa (1.93%) cells (Figure S3E). This result shows that in ALT-positive cells, VTRs do not occur more frequently than non-ALT cells within the very terminal portion of chromosome ends. Due to the short read lengths used in this approach, we cannot determine whether VTRs in ALT cells are specifically excluded from the terminal region of chromosome ends.

Frequent occurrence of ssDNA in ALT telomeres

Having demonstrated that the 5' terminal repeats in ALT cells are indistinguishable from non-ALT cells, we next asked whether the presence of single-stranded DNA (ssDNA) could be a unique feature of telomeric DNA in ALT-positive cells. Previous work provided indirect evidence for the presence of single-stranded telomeric DNA in ALT-positive cells, such as staining for RPA at telomeres, and FISH under native conditions (Figures S4A-S4D) [19, 29]. To test whether telomeres in ALT cells contain ssDNA, we employed S1-END-seq as a direct measure for the presence of ssDNA [33]. For this method, the S1 nuclease is used to convert any ssDNA into DSBs which are detected by END-seq. We anticipated that if telomeres do not contain any single-stranded telomeric DNA in addition to the 3' overhang, telomeric S1-END-seq reads will be indistinguishable from the signal obtained using standard END-seq, corresponding to one-ended DNA ends (Figure 4A). However, S1-mediated cleavage of ssDNA within the internal telomeric tract will generate two-ended DSBs within telomeres. A DSB within telomeres would results in telomeric reads corresponding both to the C-rich and the G-rich

strands (Figure 4A). To test this prediction, we performed S1-END-seq on a panel of ALT-positive and ALT-negative (Figures 4B and S4E). S1-END-seq in ALT-negative cells revealed only C-strand telomeric reads excluding the presence of a significant amount of ssDNA within the telomeric tract in these cells (Figures 4B and S4E). By contrast, performing S1-END-seq on a panel of ALT-positive cell lines revealed that G-rich and C-rich telomeric reads were detected at similar frequencies (Figures 4B and S4E). These data show that telomeres in ALT-positive cells contain frequent regions of single-stranded DNA. Furthermore, our data shows that the ratio of C-rich to G-rich reads in S1-END-seq data can distinguish ALT-negative (0.98-0.99) from ALT-positive cell lines (0.4-0.58). Based on this ratio, we estimate that the ALT-positive cells analyzed here contain a minimum of three single-stranded telomeric DNA regions per chromosome end (Figures S4F and S4G).

We also investigated whether reads detected by S1-END-seq are more likely to contain VTRs compared to the terminal portions of telomeres. We compared the frequency of VTRs within the first 30 nucleotides of the S1-END-seq reads from U2OS libraries (3.8%) to their control END-seq U2OS libraries (1.93%). This shows that the frequency of VTRs is higher in the internal portions of chromosome ends compared to the terminal portions (Figure S4H).

To determine whether the presence of ssDNA at telomeres in ALT cells is a feature intrinsic to these cells or a feature of the ALT-mediated telomere elongation, we measured the levels of ssDNA at telomeres in cells lacking BLM. Depletion of BLM is sufficient to suppress ALT-mediated telomere elongation, as shown by the loss of ALT-associated features and progressive telomere shortening [14, 19, 40-42]. As expected, we found similar levels of C-rich and G-rich reads in S1-END-seq data from U2OS cells (C/G strand ratio: 0.55, Figure 4D), indicating a minimum of 5 ssDNA regions per chromosome end. By contrast, BLM-deficient U2OS cells show a bias towards C-rich reads (C/G strand ratio: 0.7, Figure 4D), indicating that these cells have reduced levels of ssDNA at chromosome ends compared to the parental ALT-positive U2OS cell line (at least 1 ssDNA region per chromosome end). Given that BLM helicase contributes to telomere stability by resolving secondary structures and ensures proper telomeres replication regardless of telomere maintenance mechanism [43-45], it is possible that ssDNA detected in the absence of BLM is due to replication fidelity. Contrary to this, most of the reads detected in both BLM-proficient and deficient cells by END-seq were coming from the C-strand (Figure 4C). As a control, we tested whether artificially introducing DSBs at telomeres using the TRF1-FokI construct [46] would abrogate the difference in BLM-proficient and deficient cells. Expression of TRF1-Fok1 results in DSBs at telomeres, which are detected as

the appearance of G-rich reads by END-seq in both BLM proficient and BLM-deficient cells (Figure 4G). Expression of the TRF1-FokI nuclease in BLM-deficient cells was sufficient to induce single-stranded telomeric DNA at telomeres as seen by ssTelo signal (Figures 4E and 4F) as well as a C/G strand ratio close to what was seen in U2OS cells (Figure 4H).

Collectively, these data show that cells undergoing ALT have multiple single-stranded DNA regions that vastly exceed the number of chromosome ends. Accumulation of single-stranded telomeric DNA is tightly linked to the ALT process itself.

Discussion

END-seq as an unbiased tool to study the terminal telomeric repeat

Here, we show that END-seq can efficiently capture 5' chromosome termini, providing a high-resolution and unbiased approach to studying the terminal portion of chromosome ends. We found that by END-seq, chromosome ends are detected as one-ended DSBs corresponding to the 5' C-rich telomeric strand. Our data show that, in agreement with previous work [34], most chromosomes have the sequence ATC as the 5' end. END-seq has several advantages over alternative approaches: it is unbiased since it does not rely on capture oligos, it can be applied to telomeres of any length, and it does not require any prior knowledge of the subtelomeric sequence. While END-seq is designed to capture chromosome ends in an unbiased manner, we cannot exclude the possibility that certain telomeric structures may hinder detection, potentially leading to underrepresentation of a subset of telomeres. Nevertheless, we have been able to define the terminal identity of chromosome ends in cells with long telomeres, such as murine cells and ALT-positive cells, as well as cells with poorly characterized subtelomeric regions, such as canine cells. These data show that the ATC-5' sequence appears to be universal, highlighting a conserved mechanism of end-processing in mammalian cells. We also examined the effect of the depletion of POT1 paralogs on the 5' chromosome termini in mouse cells. While POT1a depletion altered the sequence at 5' ends, POT1b depletion had no effect. This finding adds to the growing evidence that POT1a and POT1b have distinct functions at mouse telomeres [35, 38, 47-50].

Canonical POT1-dependent regulation of the terminal sequences in ALT-positive cells

Our data show that ALT-positive cells, which use recombination-based mechanisms for telomere elongation, have the same regulation of terminal repeats as telomerase-positive cells. END-seq analysis of three different ALT-positive cell lines (G292, SAOS2 and U2OS) revealed

that the majority of telomeric reads correspond to C-rich reads starting with the canonical CCAATC-5' sequence. This indicates that the regulation of the terminal 5' end of chromosome ends occurs independently of the telomere elongation mechanism and that ALT-based recombination does not alter this process. Furthermore, our data show that POT1 is critical, even in U2OS cells, to define the 5' end termini of telomeres.

Single-stranded DNA in ALT telomeres

Using the unbiased S1-END-seq approach, we show that the presence of single-stranded telomeric DNA characterizes ALT-positive cells. This property of ALT is unique in that cells can be readily classified as ALT-positive or ALT-negative based on the occurrence of S1-cuts within the telomeric reads (Figure 4B).The observed single-stranded telomeric DNA may arise from various sources, including R-loops, D-loops, internal DNA loops (i-loops), and extrachromosomal circles, which are known to be prevalent in ALT [20, 30, 51, 52].

Notably, our findings align with previous indirect evidence suggesting that ALT-positive cells exhibit increased levels of single-stranded telomeric DNA, RPA-staining and FISH under native conditions [19, 29, 53, 54]. Based on our data, we can now conclude that these assays are all indeed detecting the telomeric ssDNA. Due to the nature of S1 cleavage and the repetitive nature of telomeric reads, it is impossible to establish how many regions within telomeres have ssDNA or the average length of the single-stranded DNA tract within telomeres. However, based on the relative ratio between G-rich and C-rich strands, we estimate that there are at least 5 distinct regions of single-stranded DNA for each chromosome end in ALT-positive cells (Figure S4G). The origin of the observed single-stranded telomeric DNA remains to be determined. It is possible that, in ALT cells, partially single-stranded extrachromosomal DNA in the form of C-circles is captured by the S1-END-seq approach. Alternatively, we may be detecting intermediates of the ALT process at chromosome ends or single-stranded DNA generated by strand displacement due to the presence of R-loops and/or G-quadruplexes. Future studies will be needed to elucidate the mechanisms underlying the generation of ssDNA at telomeres in ALT cells. Nevertheless, our findings provide a rationale for using ssTelo as a biomarker for ALT-positive cell detection, offering new avenues for studying and targeting ALT-driven cancers [19, 53-55].

Material and Methods

Cell culture

Human, mouse, and dog cell lines were grown in DMEM medium supplemented with 10% fetal bovine serum (FBS) and 0.5 mg/ml penicillin, streptomycin and L-glutamine (Gibco) and cultured at 37°C and 5%CO_2. HeLa cells used in this study were HeLa1.2.11. G292 cells were grown in McCoy medium supplement with 15% FBS and 0.5 mg/ml penicillin, streptomycin and L-glutamine (Gibco). Cell lines were tested for mycoplasma contamination. BLM knock-out cells were previously described in [19], canine cell line KOS was a kind gift of Dr. Amy LeBlanc (NCI).

Generation of mESCs lines:

Knockout and knock-in cell lines were generated via CRISPR/Cas9 gene targeting by transient transfection using Lipofectamine 2000 (ThermoFisher Scientific). Cells were transfected with a plasmid encoding SpCas9 (px330) and a plasmid encoding specific gRNAs (backbone Addgene 41824). To introduce the dTAG13 controllable FKBP-tag POT1a in mESCs via homologous recombination-mediated repair, a synthetic repair template containing homology arms to the *Pot1a* locus, a blasticidin resistance marker, a 3xFLAG tag, and an FKBP tag was used, along with 2 sgRNA targeting the 5' region of *Pot1a* (guide 1: 5'- TTACATTAGTGCTAATCCAA-3', and guide 2: 5'-TTTTCATCAAACAATGTCTT-3'). To generate POT1b$^{-/-}$ cells two gRNAs were used: guide 1 (5'-GTTCCTTTGCTTAAGTACGG-3') and guide 2 (5'-GACATTTGATGGCACCTTAG-3'). Transfected cells were single-cell cloned, and resulting clones were screened by PCR to confirm the desired genomic deletion, which was further validated by Sanger sequencing.

Lentiviral-mediated gene manipulation:

Lentiviral particles were generated by transfection into HEK293T followed by isolation and transduction as previously described [19]. Lentiviral expression was achieved with the following constructs and procedure:

<u>POT1 depletion:</u> sgRNA targeting POT1 (POT1-1: TGTAGCTTGATCAGACACTTA, POT1-2: TATGTATGCTAAATTGGATGG) were cloned into an inducible lentiviral vector (pRSITEP-U6Tet-sh-EF1-TetRep-2A-Puro) (Cellecta). Following infection, cells were treated with 1 µg/ml of doxycycline to induce shRNA expression and harvested 3 days later.

<u>TRF1-Fok1 expression:</u> pLenti-CMV-TRE3G-ER-TRF1-Fok1 expressing plasmids [20] were a kind gift from Agnel Sfeir. Infected cells were treated with 1 µg/ml of doxycycline and 0.6 µM of 4-hydroxytamoxifen (4-OHT) to induce TRF1-Fok1 and harvested 24 hours later.

RNA isolation and qRT-PCR

Total RNA was extracted from cells using the RNeasy Plus Mini Kit (Qiagen), and cDNA was made using SuperScript II Reverse Transcriptase (ThermoFisher). qPCR was performed using Power SYBR Green PCR Master (Fisher Scientific). Samples were run on a BioRad CFX Opus 96 and analyzed on BioRad CFX Maestro. The following primers were used in this study:

hPOT1: 5'-CAGAACCTGACGACAGCTTTCC and 5'-GCACATAGTGGTGTCCTCTCCA.

hGAPDH: 5'- GTCTCCTCTGACTTCAACAGCG and ACCACCCTGTTGCTGTAGCCAA.

mPOT1b: 5'-TTCGGCCCCAGTAGCACCTT and 5'-TCTCTTGCTTAAAGTACGCAG.

mGAPDH: 5'-TGTGTCCGTCGTGGATCTGA and 5'-TTGCTGTTGAAGTCGCAGGAG.

Immunofluorescence (IF) and FISH- based imaging

IF, IF-FISH and ssTelo staining were carried out as described previously [19, 56] .

PNA probes used in this study were TelC-Cy3 (PNABio, Catalog number F1002) and TelG-Cy3 PNA probe (PNABio, Catalog number F1006)**.**

Primary antibodies used in this study were TRF2 (NB110-57130, Novus Biologicals, rabbit), RPA32/RPA2 (2208, Cell Signaling, rat), Myc (2276, Cell Signaling, mouse) and 53BP1 (NB100-304, Novus Biologicals, rabbit).

Imaging and image analysis

Images were acquired using a Zeiss Axio Imager M2 with an Axiocam 702 camera and ZEN 2.6 (blue edition) software; for each experiment and condition, a minimum of 250 cells were imaged. Final figures were assembled using Adobe Photoshop 25.7.0 and Adobe Illustrator 28.5.

END-seq & S1-END-seq

For END-seq and S1-END-seq, 6–20 x10^6 cells were collected, embedded in 1% agarose plugs, and processed as previously reported [32, 33]. For mouse spleen library, mice were euthanized with CO_2 following ethical guidelines, and the spleens were surgically removed. Spleen tissues were mashed through a 70 µm cell strainer to obtain a single-cell suspension. The suspension was then centrifuged for 5 mins at 300x g, the pellet was washed with ice-cold PBS three times before proceeding with library preparation. Briefly, plugs were incubated with proteinase K solution for 1 hr at 50°C and then for 7 hrs at 37 °C, followed by consecutive washes in a wash buffer (10 mM Tris, pH 8.0, and 50 mM EDTA) and TE (10 mM Tris, pH 8.0, and 1 mM EDTA).

Washed plugs were then treated with RNAseA (Puregene, QIAGEN), washed again in Wash Buffer, and stored at 4°C for up to 2 weeks. End blunting, A-tailing, and ligation to biotinylated hairpin adaptor 1 (ENDseq-adaptor-1, 5′-Phos-GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGUU[Biotin-dT]U[Biotin-dT]UUACACTCTTTCCCTACACGACGCTCTTCCGATC∗T-3′ [∗phosphorothioate bond]) were performed in the plug to minimize shearing and *in vitro* damage. Following this, high molecular weight DNA was isolated from agarose plugs and fragmented by sonication. Subsequently, adaptor-ligated fragments were captured with streptavidin-coated beads and sonication-created ends were repaired, A-tailed, and ligated with a distal adaptor (ENDseq-adaptor-2, 5′-Phos-GATCGGAAGAGCACACGTCUUUUUUUUAGACGTGTGCTCTTCCGATC∗T-3′ [∗phosphorothioate bond]). Subsequent PCR amplification resulted in ready-to-sequence libraries where the sequenced first base corresponds precisely to the first base of the blunted DSB on either side of the break. For S1-END-seq, proteinase K and RNAseA treated plugs were treated with 100 U of S1 nuclease (Thermo Fisher) at 37 °C for 30 minutes and the reaction was terminated by the addition of EDTA to a final concentration of 10 mM. S1-treated plugs were then processed through the standard END-seq protocol. For T7 treatment, proteinase K and RNAseA treated plugs were incubated with 50 U of T7 exonuclease (NEB) at 37 °C for 30 minutes before processing through the standard END-seq protocol.

The libraries were sequenced on the Nextseq 550 or Nextseq 2000 platform (Illumina), using 75-bp single-end read kits. For each treatment condition, the control (non-treated) and sample libraries were sequenced in parallel in the same flow cell.

Telomeric Motif Detector and Quantifier

We developed a Python script, "getTelomereENDseq," to analyze gzipped FASTQ files for the presence of telomeric motifs in DNA sequences on a command-line environment using a Linux-based operating system and the statistical environment RStudio. The script identifies sequences containing G-rich and C-rich telomeric repeats, quantifies their occurrences, and calculates their relative frequencies. It takes a gzipped FASTQ file as input, and the minimum number of telomeric repeats is required. It generates two output files: one with sequences containing G-rich motifs and another with sequences containing C-rich motifs. For this study, we used a minimum of 4 consecutive TTAGGG repeats for G-rich reads and a minimum of 4 consecutive CCCTAA repeats for C-rich reads. Following this, in RStudio, using the stringi, ggseqlogo, ggplot2, ggsci, and tidyverse packages, the telomeric reads were trimmed by the last six

nucleotides, reversed, and organized into data frames. These data frames were then used to plot the frequencies of telomeric repeats or motifs.

Statistical analysis and data visualization

We used RStudio and GraphPad Prim v10 for data visualization and statistical analysis. Statistical analyses are specified in the figure legends. For microscopy data, $P < 0.05$ was considered statistically significant, and P values were assessed by unpaired parametric t-test. The sample size was not predetermined. Statistical analyses are specified in the figure legends. We used Kullback-Leibler (KL) divergence to compare the distributions of frequencies of telomeric repeats between different samples. The KL divergence is a measure of the difference between two probability distributions. It quantifies how one probability distribution diverges from a second, expected distribution. To compute KL divergence, we used the philentropy package in RStudio. For this study, KL divergence values smaller than 0.125 are considered non-significant, and the rest assigned as follows: ● indicates KL divergence value between 0.125 and 0.25, ●● indicates a value between 0.25 and 0.375, and ●●● indicates a value greater than 0.375.

Author contributions

B.A. and E.L.D. designed the project; B.A. performed experiments and bioinformatic analyses; R.P. and A.N. helped optimizing END-seq and S1-END-seq and provided critical suggestions; W.W. helped with bioinformatic analyses of END-seq and S1-END-seq data; R.S generated mESCs lines; T.M. isolated mouse tissues used in the study; E.L.D supervised the project; B.A and E.L.D wrote the manuscript.

Acknowledgement

We are grateful to Agnel Sfeir, Amy LeBlanc, and Gianna Tricola for providing reagents. We thank the CCR Genomic Core for assistance with the NGS experiments performed in this study. We thank members of the Lazzerini Denchi laboratory, Sam John, Niek van Wietmarschen and Travis Stracker for critical feedback on this work. Research in the ELD lab is funded by the NIH Intramural Research Program of the National Institutes of Health (NIH), National Cancer Institute (NCI) and Center for Cancer Research, project 1-ZIA-BC011815-03.

Figure Legends

Fig 1: END-seq successfully captures the 5’ termini of human and mouse chromosomes.

A. Schematic representation of the END-seq method applied to DSBs (left) and natural chromosome ends (right). Blunted dsDNA ends are ligated to biotinylated adaptors, purified, and subsequently sequenced by NGS. Reads originating from DSBs align to either side of the break (left), while telomeric reads align only to the 5' C-rich template (red). **B.** Number of reads containing at least 4 consecutive telomeric repeats corresponding to the G-rich (green) and C-rich (red) strands in HeLa and RPE1 cell lines. Telomeric reads are normalized to the total number of reads identified by END-seq and are represented as reads per million (RPM). **C.** Sequence logo representing the conservation (bits) of the last 6 nucleotides at the 5' end of chromosomes in HeLa and RPE1 cells. **D.** Distribution of the three most frequent telomeric repeats across human cell lines (orange), mouse cells (yellow) and a canine cell line (purple). **E.** Schematic representation of T7-mediated randomization of the 5' termini. **F.** Percentage of telomeric reads that have the indicated sequence as a 5' end. The following sequences (CCCAAT-5', TCCCAA-5', and ATCCCA-5') are grouped and labelled as "Rest". Kullback-Leibler divergence (KL divergence) analysis was used to compare the distributions of the individual conditions. KL divergence between 0.25 and 0.375 is represented by 2 dots. Prior to the END-seq protocol cells were either left untreated (NT) or were treated with T7. **G.** Percentage of telomeric reads are displayed as described in F. Cells expressing either a non-targeting shRNA (shScr) or a shRNA targeting POT1 (shPOT1) were harvested 3 days post induction and analyzed by END-seq. KL divergence between 0.125 and 0.25 is represented by 1 dot (●).

Fig 2: POT1a is the key regulator of 5' telomere end processing in mice, distinct from POT1b

A. Comparison of the effect of single and double depletion of POT1a and POT1b on the percentage of telomeric reads with the indicated 5' end sequences. The following sequences (CCCAAT-5', TCCCAA-5', and ATCCCA-5') are grouped and labelled as "Rest". Kullback-Leibler divergence (KL divergence) analysis was used to compare the distributions across conditions. KL divergence between 0.25 and 0.375 is represented by 2 dots. **B.** Sequence logo illustrating the conservation (in bits) of the last 6 nucleotides at the 5' end of chromosomes in POT1-proficient, POT1a-depleted, POT1b-depleted, and POT1a/POT1b double-depleted mESCs.

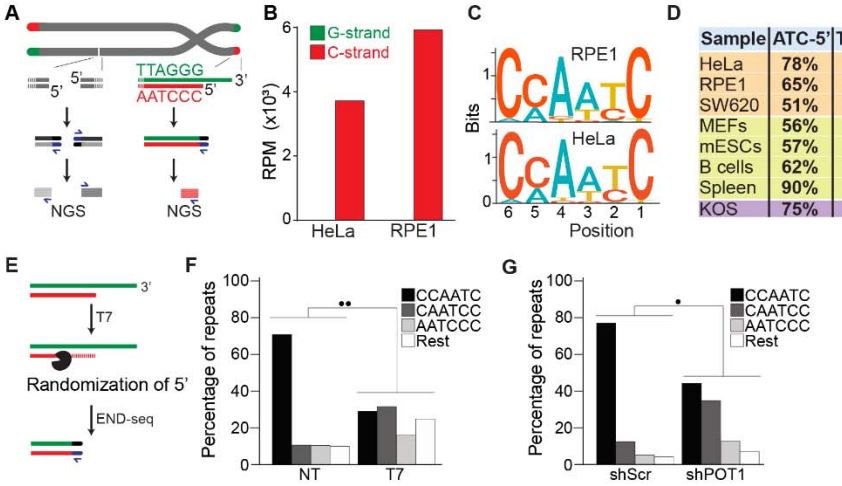
Fig 3: ALT cells have precise 5' termini.

A-C. Number of telomeric reads containing at least 4 consecutive telomeric repeats corresponding to the G-rich (green) and C-rich (red) strands in the ALT-positive G292, SAOS2 and U2OS cell lines. Telomeric reads are normalized to the total number of reads identified by END-seq and are represented as reads per million (RPM). **D.** Sequence logo representing the conservation at chromosome ends. A sequence logo representing the conservation (bits) of the last 6 nucleotides at the 5' end of chromosomes in ALT cells. Fraction of reads with CCAATC as 5' end is displayed. **E.** Cells expressing either a non-targeting shRNA (shScr) or a shRNA targeting POT1 (shPOT1-1) were stained for 53BP1 (red) and telomeric DNA (TTAGGG, green). The scale bar represents 10 µm. **F.** Quantification of the data shown in panel E. Graphs indicated the percentage of cells that have at least 5 telomere dysfunctional foci (TIF) with 53BP1 co-localizing at telomeres. **G.** Percentage of telomeric reads that have the indicated sequence as a 5' sequence that have as a 5' end. The following sequences (CCCAAT-5', TCCCAA-5', and ATCCCA-5') are grouped and labelled as "Rest". Cells expressing either a non-targeting shRNA (shScr) or a shRNA targeting POT1 (shPOT1-1 and shPOT1-2) were harvested 3 days post induction and analyzed by END-seq. Kullback-Leibler divergence (KL divergence) analysis was used to compare the distributions of the individual conditions. KL divergence between 0.25 and 0.375 is represented by 2 dots (●●), KL divergence greater than 0.375 is represented by 3 dots (●●●).

Fig 4: ALT telomeres are readily distinguished due to presence of telomeric ssDNA by sequencing.

A. Schematic Representation of the S1-END-seq method applied to telomeres containing either an internal ssDNA region (left) or only the natural G-rich overhang (right). S1 nuclease treatment cleaves ssDNA regions generating two ended DSB (left) or one ended DSB (right). DNA ends are ligated to biotinylated adaptors, purified, and subsequently sequenced by NGS. Reads originating from the double ended DSB align to either side of the break (left), resulting in C-rich (red) and G-rich (green) reads. Reads originating from the single-ended chromosome ends align only to the C-rich template (red). **B**. Proportion of telomeric reads (containing at least 4 consecutive telomeric reads) corresponding to G-rich reads (green) or C-rich reads (red). **C-D.** Proportion of telomeric reads (containing at least 4 consecutive telomeric reads) corresponding to G-rich reads (green) or C-rich reads (red) in BLM proficient (U2OS) or deficient ($BML^{-/-}$) cells by END-seq (**C**) or S1-seq (**D**). **E.** Native telomeric FISH (ssTelo) in BLM proficient (U2OS) or deficient ($BML^{-/-}$) cells either left untreated (control) or expressing TRF1-Fok1. The scale bar represents 10 µm. **F.** Quantification of the data shown in panel E, cells with 5 ssTelo signal (or

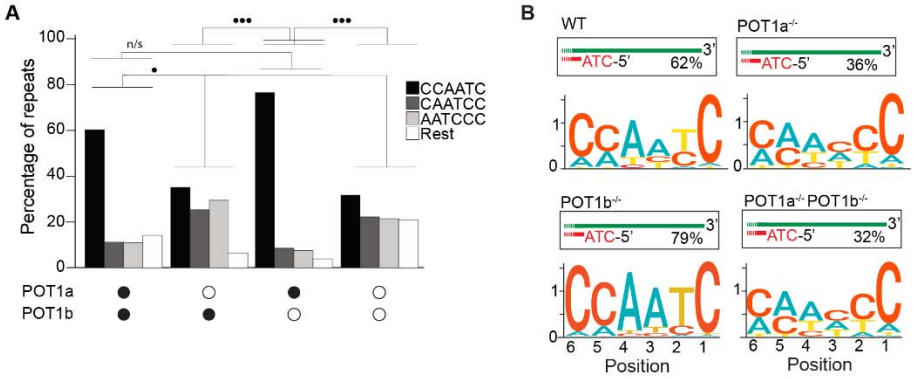
grater) were scored as positive. When indicated, cells were induced to express the TRF1-Fok1 nuclease for 24 hours prior to harvesting. **G-H.** Proportion of telomeric reads (containing at least 4 consecutive telomeric reads) corresponding to G-rich reads (green) or C-rich reads (red) in BLM proficient (U2OS) or deficient ($BML^{-/-}$) cells expressing TRF1-Fok1 by END-seq (**G**) or S1-seq (**H**).

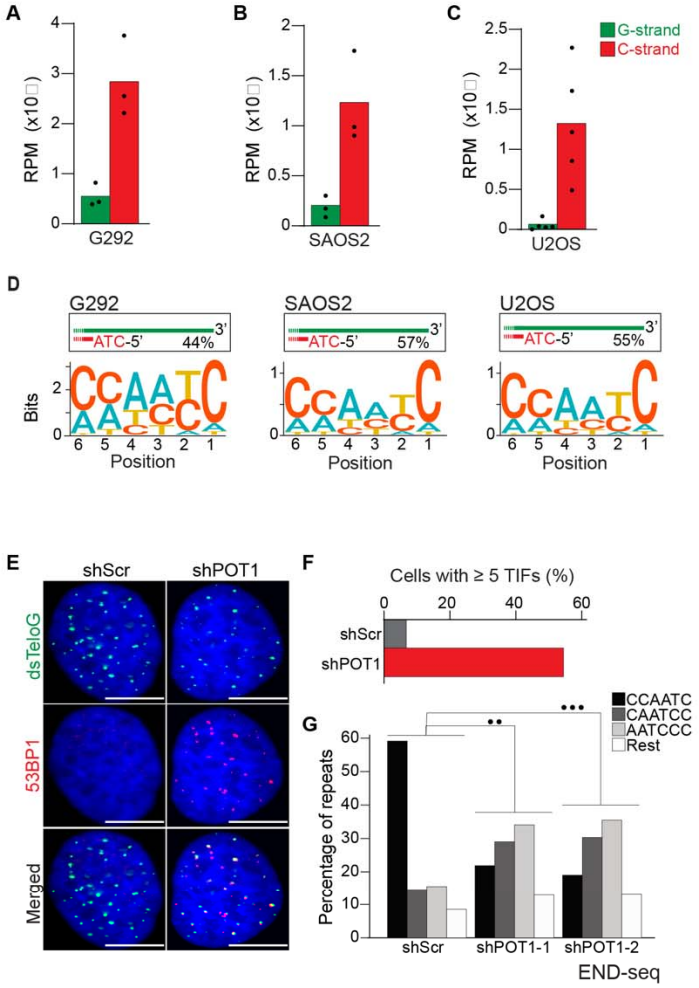


D

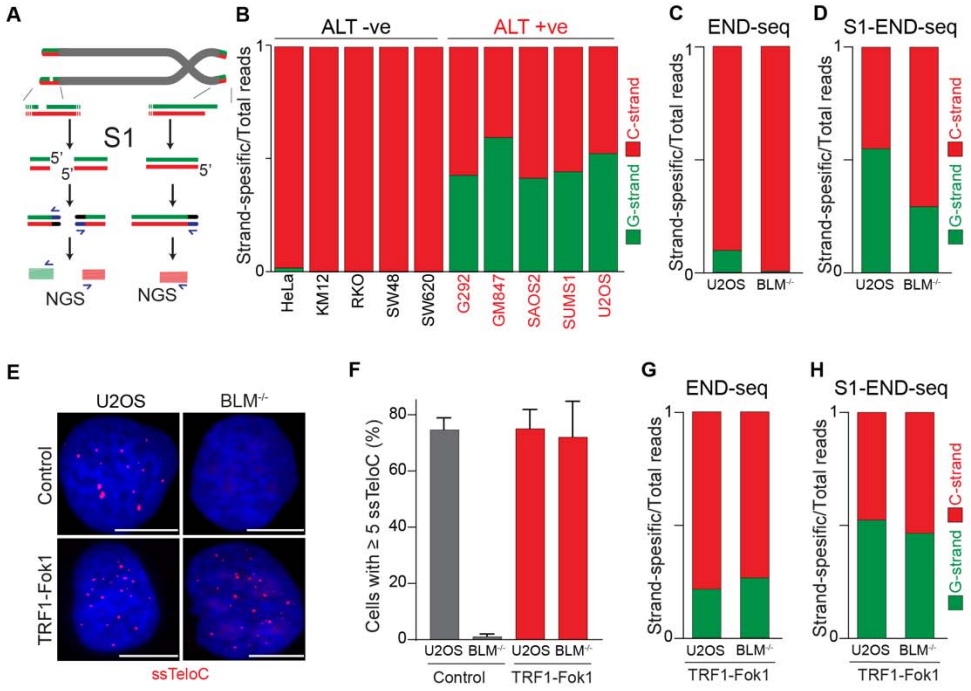
Sample	ATC-5'	TCC-5'	CCC-5'
HeLa	**78%**	12%	5%
RPE1	**65%**	26%	4%
SW620	**51%**	31%	16%
MEFs	**56%**	23%	17%
mESCs	**57%**	20%	16%
B cells	**62%**	28%	5%
Spleen	**90%**	2%	1.4%
KOS	**75%**	12%	0.8%

Azeroglu et al. Figure 2




A
B
C
RPM (x10□)
G292
SAOS2
U2OS
G-strand
C-strand
D
G292
SAOS2
U2OS
ATC-5'
44%
57%
55%
3'
Bits
Position
E
shScr
shPOT1
dsTeloG
53BP1
Merged
F
Cells with ≥ 5 TIFs (%)
shScr
shPOT1
G
CCAATC
CAATCC
AATCCC
Rest
Percentage of repeats
shScr
shPOT1-1
shPOT1-2
END-seq

Azeroglu et al. Figure 4



References

1. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres.* Annu Rev Genet, 2008. **42**: p. 301-34.
2. Cai, S.W. and T. de Lange, *CST-Polalpha/Primase: the second telomere maintenance machine.* Genes Dev, 2023. **37**(13-14): p. 555-569.
3. Lim, C.J. and T.R. Cech, *Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization.* Nat Rev Mol Cell Biol, 2021. **22**(4): p. 283-298.
4. Wu, P., H. Takai, and T. de Lange, *Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST.* Cell, 2012. **150**(1): p. 39-52.
5. Takai, H., et al., *CST-polymerase alpha-primase solves a second telomere end-replication problem.* Nature, 2024. **627**(8004): p. 664-670.
6. Feldser, D.M. and C.W. Greider, *Short telomeres limit tumor progression in vivo by inducing senescence.* Cancer Cell, 2007. **11**(5): p. 461-9.
7. Greenberg, R.A., et al., *Expression of mouse telomerase reverse transcriptase during development, differentiation and proliferation.* Oncogene, 1998. **16**(13): p. 1723-30.
8. Greider, C.W., *Telomere length regulation.* Annu Rev Biochem, 1996. **65**: p. 337-65.
9. Perera, S.A., et al., *Telomere dysfunction promotes genome instability and metastatic potential in a K-ras p53 mouse model of lung cancer.* Carcinogenesis, 2008. **29**(4): p. 747-53.
10. Shay, J.W. and S. Bacchetti, *A survey of telomerase activity in human cancer.* Eur J Cancer, 1997. **33**(5): p. 787-91.
11. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. **266**(5193): p. 2011-5.
12. Dunham, M.A., et al., *Telomere maintenance by recombination in human cells.* Nat Genet, 2000. **26**(4): p. 447-50.
13. Bryan, T.M., et al., *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines.* Nat Med, 1997. **3**(11): p. 1271-4.
14. Jiang, H., et al., *BLM helicase unwinds lagging strand substrates to assemble the ALT telomere damage response.* Mol Cell, 2024. **84**(9): p. 1684-1698 e9.
15. Lu, R., et al., *The FANCM-BLM-TOP3A-RMI complex suppresses alternative lengthening of telomeres (ALT).* Nat Commun, 2019. **10**(1): p. 2252.
16. Pan, X., et al., *FANCM suppresses DNA replication stress at ALT telomeres by disrupting TERRA R-loops.* Sci Rep, 2019. **9**(1): p. 19110.
17. Panier, S., et al., *SLX4IP Antagonizes Promiscuous BLM Activity during ALT Maintenance.* Mol Cell, 2019. **76**(1): p. 27-43 e11.
18. Silva, B., et al., *FANCM limits ALT activity by restricting telomeric replication stress induced by deregulated BLM and R-loops.* Nat Commun, 2019. **10**(1): p. 2253.
19. Loe, T.K., et al., *Telomere length heterogeneity in ALT cells is maintained by PML-dependent localization of the BTR complex to telomeres.* Genes Dev, 2020. **34**(9-10): p. 650-662.
20. Dilley, R.L., et al., *Break-induced telomere synthesis underlies alternative telomere maintenance.* Nature, 2016. **539**(7627): p. 54-58.
21. Barroso-Gonzalez, J., et al., *RAD51AP1 Is an Essential Mediator of Alternative Lengthening of Telomeres.* Mol Cell, 2020. **79**(2): p. 359.
22. Kaminski, N., et al., *RAD51AP1 regulates ALT-HDR through chromatin-directed homeostasis of TERRA.* Mol Cell, 2022. **82**(21): p. 4001-4017 e7.
23. Lynskey, M.L., et al., *HIRA protects telomeres against R-loop-induced instability in ALT cancer cells.* Cell Rep, 2024. **43**(11): p. 114964.

24. Azeroglu, B., et al., *Identification of modulators of the ALT pathway through a native FISH-based optical screen.* Cell Rep, 2025. **44**(1): p. 115114.
25. Bhargava, R., M.L. Lynskey, and R.J. O'Sullivan, *New twists to the ALTernative endings at telomeres.* DNA Repair (Amst), 2022. **115**: p. 103342.
26. Claude, E. and A. Decottignies, *Telomere maintenance mechanisms in cancer: telomerase, ALT or lack thereof.* Curr Opin Genet Dev, 2020. **60**: p. 1-8.
27. Loe, T.K., et al., *ALTercations at telomeres: stress, recombination and extrachromosomal affairs.* Biochem Soc Trans, 2023. **51**(5): p. 1935-1946.
28. O'Sullivan, R.J. and R.A. Greenberg, *Mechanisms of Alternative Lengthening of Telomeres.* Cold Spring Harb Perspect Biol, 2025. **17**(1).
29. Grudic, A., et al., *Replication protein A prevents accumulation of single-stranded telomeric DNA in cells that use alternative lengthening of telomeres.* Nucleic Acids Res, 2007. **35**(21): p. 7267-78.
30. Henson, J.D., et al., *DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity.* Nat Biotechnol, 2009. **27**(12): p. 1181-5.
31. Cesare, A.J. and J.D. Griffith, *Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops.* Mol Cell Biol, 2004. **24**(22): p. 9948-57.
32. Canela, A., et al., *DNA Breaks and End Resection Measured Genome-wide by End Sequencing.* Mol Cell, 2016. **63**(5): p. 898-911.
33. Matos-Rodrigues, G., et al., *S1-END-seq reveals DNA secondary structures in human cells.* Mol Cell, 2022. **82**(19): p. 3538-3552 e5.
34. Sfeir, A.J., et al., *Telomere-end processing the terminal nucleotides of human chromosomes.* Mol Cell, 2005. **18**(1): p. 131-8.
35. Tesmer, V.M., K.A. Brenner, and J. Nandakumar, *Human POT1 protects the telomeric ds-ss DNA junction by capping the 5' end of the chromosome.* Science, 2023. **381**(6659): p. 771-778.
36. Smoom, R., et al., *Telomouse-a mouse model with human-length telomeres generated by a single amino acid change in RTEL1.* Nat Commun, 2023. **14**(1): p. 6708.
37. Hockemeyer, D., et al., *POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end.* EMBO J, 2005. **24**(14): p. 2667-78.
38. Denchi, E.L. and T. de Lange, *Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1.* Nature, 2007. **448**(7157): p. 1068-71.
39. Conomos, D., et al., *Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells.* J Cell Biol, 2012. **199**(6): p. 893-906.
40. O'Sullivan, R.J., et al., *Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1.* Nat Struct Mol Biol, 2014. **21**(2): p. 167-74.
41. Min, J., W.E. Wright, and J.W. Shay, *Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52.* Genes Dev, 2019. **33**(13-14): p. 814-827.
42. Sobinoff, A.P., et al., *BLM and SLX4 play opposing roles in recombination-dependent replication at human telomeres.* EMBO J, 2017. **36**(19): p. 2907-2919.
43. Barefield, C. and J. Karlseder, *The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures.* Nucleic Acids Res, 2012. **40**(15): p. 7358-67.
44. Drosopoulos, W.C., S.T. Kosiyatrakul, and C.L. Schildkraut, *BLM helicase facilitates telomere replication during leading strand synthesis of telomeres.* J Cell Biol, 2015. **210**(2): p. 191-208.
45. Sfeir, A., et al., *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication.* Cell, 2009. **138**(1): p. 90-103.

46. Cho, N.W., et al., *Interchromosomal homology searches drive directional ALT telomere movement and synapsis.* Cell, 2014. **159**(1): p. 108-121.
47. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. **126**(1): p. 49-62.
48. Nurk, S., et al., *The complete sequence of a human genome.* Science, 2022. **376**(6588): p. 44-53.
49. Hockemeyer, D., et al., *Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres.* Cell, 2006. **126**(1): p. 63-77.
50. Gu, P., et al., *Distinct functions of POT1 proteins contribute to the regulation of telomerase recruitment to telomeres.* Nat Commun, 2021. **12**(1): p. 5514.
51. Mazzucco, G., et al., *Telomere damage induces internal loops that generate telomeric circles.* Nat Commun, 2020. **11**(1): p. 5297.
52. Arora, R., et al., *RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells.* Nat Commun, 2014. **5**: p. 5220.
53. Claude, E., et al., *Detection of alternative lengthening of telomeres mechanism on tumor sections.* Mol Biomed, 2021. **2**(1): p. 32.
54. Frank, L., et al., *ALT-FISH quantifies alternative lengthening of telomeres activity by imaging of single-stranded repeats.* Nucleic Acids Res, 2022. **50**(11): p. e61.
55. Quintanilla, I., et al., *Optical pooled screening for the discovery of regulators of the alternative lengthening of telomeres pathway.* Methods, 2025.
56. Azeroglu, B., et al., *Native FISH: A low- and high-throughput assay to analyze the alternative lengthening of telomere (ALT) pathway.* Methods Cell Biol, 2024. **182**: p. 265-284.