

ARTICLE

Open Access

# Critically short telomeres derepress retrotransposons to promote genome instability in embryonic stem cells

Nannan Zhao<sup>1,2</sup>, Guoxing Yin<sup>1,2</sup>, Chun Liu<sup>1,2</sup>, Weiyu Zhang<sup>1,3</sup>, Yang Shen<sup>4</sup>, Dan Wang<sup>5</sup>, Zhenzhen Lin<sup>5</sup>, Jiao Yang<sup>1,2</sup>, Jian Mao<sup>1,2</sup>, Renpeng Guo<sup>1,2</sup>, Yongwang Zhang<sup>1,3</sup>, Feng Wang<sup>6</sup>, Zhe Liu<sup>5</sup>✉, Xinyi Lu<sup>1,3</sup>✉ and Lin Liu<sup>1,2,7,8</sup>✉

## Abstract

Telomeres, at the ends of chromosomes, protect chromosomes from fusion and preserve genomic stability. However, the molecular mechanisms underlying telomere attrition-induced genome instability remain to be understood. We systematically analyzed the expression of retrotransposons and performed genomic sequencing of different cell and tissue types with telomeres of varying lengths due to telomerase deficiency. We found that critically short telomeres altered retrotransposon activity to promote genomic instability in mouse embryonic stem cells, as evidenced by elevated numbers of single nucleotide variants, indels and copy number variations (CNVs). Transpositions of retrotransposons such as LINE1 resulting from the short telomeres can also be found in these genomes with elevated number of mutations and CNVs. Retrotransposon activation is linked to increased chromatin accessibility, and reduced heterochromatin abundance correlates with short telomeres. Re-elongation of telomeres upon recovery of telomerase partly represses retrotransposons and heterochromatin accumulation. Together, our findings suggest a potential mechanism by which telomeres maintain genomic stability by suppressing chromatin accessibility and retrotransposon activity.

## Introduction

Telomeres protect chromosome ends to prevent chromosome fusion and instability<sup>1</sup>. It has long been recognized that telomere dysfunction and rearrangement in sub-telomeric and telomeric regions can lead to genome instability and eventually cause cancer and other diseases in human<sup>2,3</sup>. Telomeres shorten progressively with age, and this shortening is accompanied by the accumulation of somatic mutations and genome instability<sup>4,5</sup>. Telomeres are also generally shorter in tumor cells than in normal somatic cells<sup>6</sup>. Recently, the mapping of somatic mutations

revealed variable mutation rates in different tissues<sup>7</sup>. However, the relationships and mechanisms underlying short or dysfunctional telomere-induced mutations and genomic instability remain elusive.

Endogenous retroviruses (ERVs) constitute ~8%–10% of mammalian genomes<sup>8</sup>. Repression of retrotransposons is critical for the maintenance of genome stability in both somatic cells and embryonic stem cells (ESCs)<sup>9</sup>. Most retrotransposons are silenced to prevent genome instability<sup>10</sup>, but in some cases, retrotransposons are activated. For instance, long terminal repeat (LTR) retrotransposons are highly expressed in the mammalian placenta and trophoctoderm<sup>11</sup>. Intriguingly, retrotransposons are activated in a small number of cases during early embryonic development, contributing to pluripotent identity and networks in mouse and human pluripotent stem cells<sup>12–16</sup>. Moreover, retrotransposon MERV1 and genes in the 2-cell embryo are sporadically

Correspondence: Zhe Liu (zheliu@tmu.edu.cn) or Xinyi Lu (luxy@nankai.edu.cn) or Lin Liu (liulin@nankai.edu.cn)

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China

<sup>2</sup>Frontiers Science Center for Cell Responses, College of Life Sciences, Nankai University, Tianjin, China

Full list of author information is available at the end of the article

These authors contributed equally: Nannan Zhao, Guoxing Yin, Chun Liu

© The Author(s) 2023



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

activated in only a small proportion (1%–5%) of ESC cultures<sup>17,18</sup>. In somatic cells, the activation of retrotransposons can drive cancer gene expression and lead to cancer development<sup>19</sup>. They can also be activated during replicative senescence in mouse and human cells<sup>20</sup>. The accumulation of retrotransposon RNAs can induce inflammation and immune responses<sup>21</sup> and result in cellular senescence<sup>22,23</sup>. Short telomeres can also lead to cell replicative senescence<sup>24–26</sup>. Furthermore, the shortest telomeres, not the average telomere length, is implicated in cell senescence<sup>27</sup>. It is noted that human telomeres are much shorter than mouse telomeres and that mouse somatic cells continue expressing telomerase in adulthood<sup>28</sup>, such that only late-generation telomerase knockout (*Terc*<sup>-/-</sup>) mice exhibit short telomere-associated phenotypes, including accelerated aging, reduced injury repair and predisposition to tumorigenesis<sup>29</sup>. Such mice have been extensively used as study models of telomere biology and function in humans. *Terc* is an essential RNA component of the telomerase complex whose absence affects telomerase activity and causes telomere shortening in diseases<sup>30</sup> and is important for pluripotent stem cells<sup>31</sup>. Short telomeres in mouse ESCs without telomerase component *Terc* lead to reduced cell proliferation<sup>32</sup>, similar to replicative senescence in human cells. It remains unknown whether telomere length regulates retrotransposons to influence genome stability in mammalian cells.

Herein, by systematically examining the effects of telomere length on the transcriptome and on genome integrity in various cell types following telomere shortening, we investigated the role of telomere length in the maintenance of genome stability by controlling the expression of retrotransposons. Critically short telomeres are associated with retrotransposon activation and their subsequent retrotransposition to promote genomic instability.

## Results

### Telomere attrition in various mouse cell types and tissues lacking telomerase

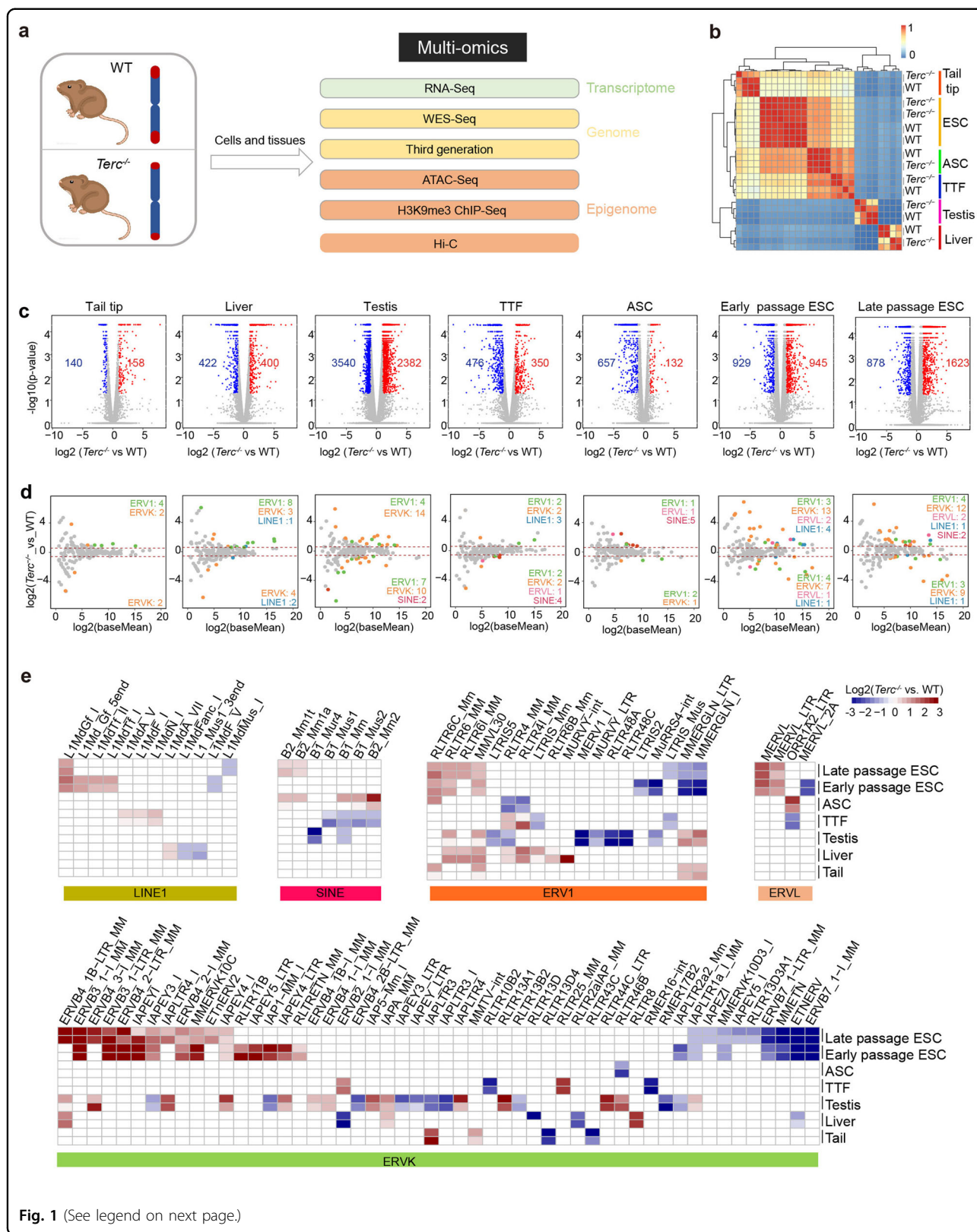
We first compared telomere lengths by telomere restriction fragment (TRF) analysis of various somatic tissues or cells obtained from the third-generation (G3) *Terc*<sup>-/-</sup> mice and ESCs (G4 *Terc*<sup>-/-</sup>) derived from blastocysts of intercrosses of G3 *Terc*<sup>-/-</sup> mice. G1 *Terc*<sup>-/-</sup> ESCs were derived from blastocysts of intercrosses of *Terc*<sup>+/-</sup> mice while G3 *Terc*<sup>-/-</sup> ESCs were derived from blastocysts of intercrosses of G2 *Terc*<sup>-/-</sup> mice<sup>32</sup>. Telomeres were shortened in all somatic tissues including the tail, testis and liver, as well as in cells in culture such as adipose-derived mesenchymal cells (ASCs), tail tip fibroblasts (TTFs) and ESCs obtained from G3 *Terc*<sup>-/-</sup> mice, compared with wild-type (WT)

mice (Supplementary Fig. S1a, b). Moreover, telomeres were shortened in *Terc*<sup>-/-</sup> ESCs with an increasing number of generations (G) from G1 to G4 and with increasing cell culture passages (P) from P18 to P42. P18 was considered as an early passage, and P42 was considered as late passage. G3 *Terc*<sup>-/-</sup> ESCs had shorter telomeres than G1 *Terc*<sup>-/-</sup> ESCs, but longer telomeres than G4 *Terc*<sup>-/-</sup> ESCs. Telomere lengths were measured in somatic cells (TTFs and ASCs) at an earlier passage (P3) because our preliminary observations indicated that these cells would undergo senescence in the absence of telomerase. G4 *Terc*<sup>-/-</sup> ESCs had the shortest telomeres among all tissues, whereas WT ESCs had the longest telomeres (Supplementary Fig. S1a, b).

Telomere quantitative fluorescence in situ hybridization (Q-FISH) of chromosome spreads confirmed the occurrence of telomere shortening in *Terc*<sup>-/-</sup> cells (Supplementary Fig. 1c–e). Notably, G4 *Terc*<sup>-/-</sup> ESCs at both early and late passages exhibited critically short telomeres and telomere loss, as evidenced by telomere signal-free chromosome ends and end-to-end chromosome fusions (Supplementary Fig. S1c–e). G4 *Terc*<sup>-/-</sup> ESCs at both early and late passages exhibited chromosome fusions (Supplementary Fig. S1f). In contrast, somatic TTFs derived from G3 *Terc*<sup>-/-</sup> mice showed no obvious telomere loss or chromosome fusion (Supplementary Fig. S1c–e). G4 *Terc*<sup>-/-</sup> ESCs had the shortest telomeres according to Q-FISH assays (Supplementary Fig. S1c–e), while telomeres were further shortened by passaging of the ESCs (Supplementary Fig. S1c–e). These results also substantiate the notion that telomerase is critical for telomere maintenance in both somatic cells and ESCs. However, mouse ESCs proliferate very rapidly, with cell doubling approximately every 7 h, unlike somatic cells, which may explain the shorter telomeres found in ESCs than in somatic cells lacking telomerase obtained from mice of the same generation.

### Short telomeres alter retrotransposons

To assess the impact of short telomeres on transcriptome and genomic stability, we performed multi-omics analyses to compare various tissues and cell types between G3 *Terc*<sup>-/-</sup> mice and WT mice. These included RNA-seq to interrogate the transcriptome, whole exome-seq and third-generation genome sequencing to examine genome stability, high-throughput/resolution chromosome conformation capture (Hi-C) to observe 3D chromatin organization, and ATAC-seq and chromatin immunoprecipitation (ChIP) coupled with high-throughput sequencing (ChIP-seq) of H3K9me3 to identify epigenomic changes (Fig. 1a). We systematically compared the expression of genes and retrotransposons in somatic tissues and cultured cells. The global gene expression profile of ESCs differed from those of other



(see figure on previous page)

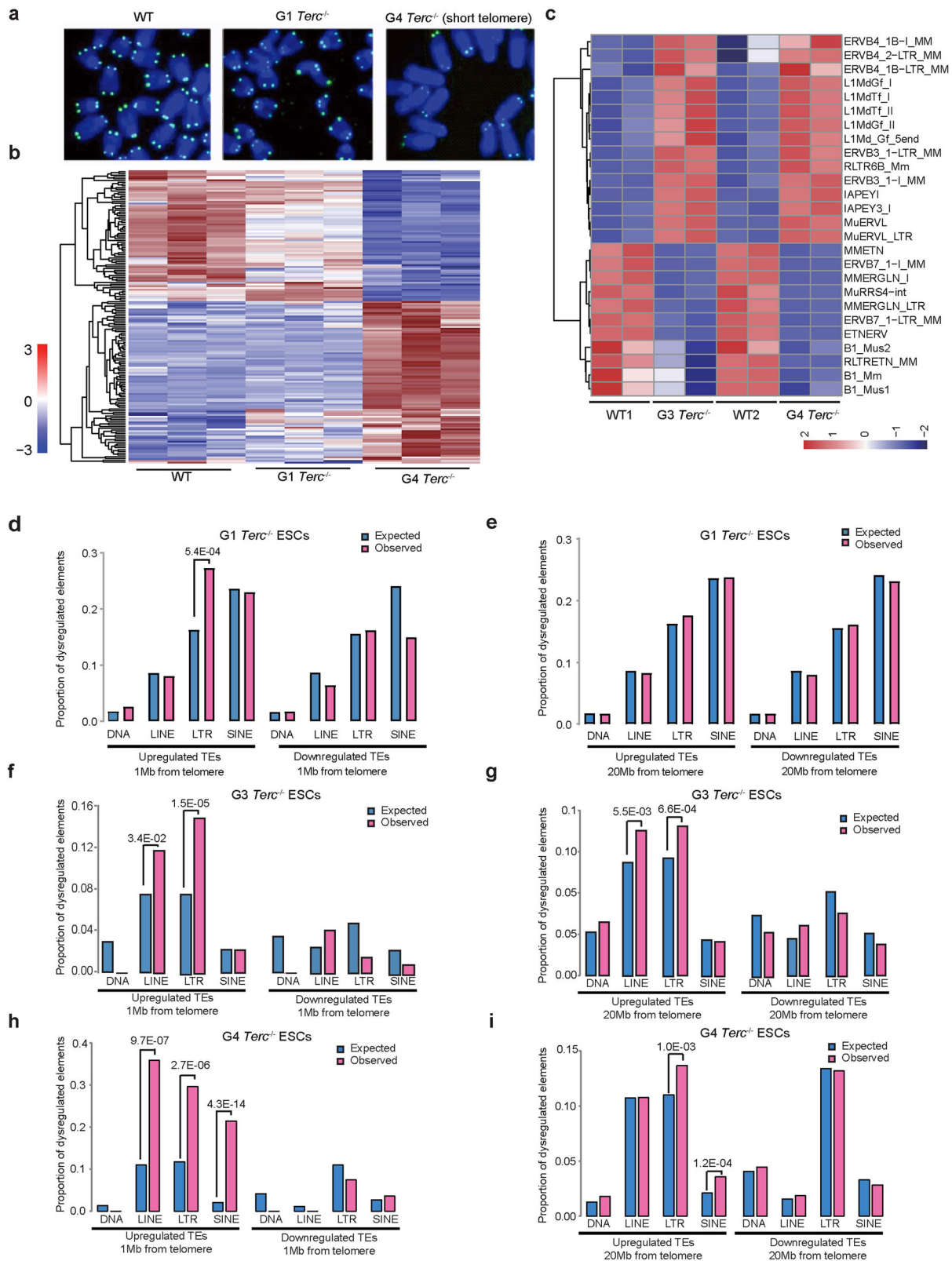
**Fig. 1 Transcriptome of differentially expressed genes and retrotransposons of various cell types from *Terc*<sup>-/-</sup> and WT mice.** **a** Overall experimental design. Several types of cultured cells and tissues from WT and G3 *Terc*<sup>-/-</sup> mice were used for multi-omics analysis, including the transcriptome analysis by RNA-seq, whole-exome sequencing (WES), third-generation whole-genome sequencing, and epigenome analysis using ATAC-seq, ChIP-seq of H3K9me3, and Hi-C sequencing. **b** Heatmap showing the correlation of the expressed genes among the cell types and tissues. ESC embryonic stem cell (P18 for early passage and P42 for late passage), TTF tail-tip fibroblast, ASC adipose-derived mesenchymal cell. The passage number for somatic tissues is P3. The bar color represents the correlation coefficient of the samples. **c** Volcano plot of differentially expressed genes (DEGs) in *Terc*<sup>-/-</sup> and WT cells of various cell types. Red dots, upregulated genes with fold change > 2 and *P* value < 0.05; blue dots, downregulated genes with fold change > 2 and *P* value < 0.05; grey dots, genes without significant change. Early-passage G4 *Terc*<sup>-/-</sup> ESC (P18), late-passage G4 *Terc*<sup>-/-</sup> ESC (P42), G3 *Terc*<sup>-/-</sup> TTF (P3), G3 *Terc*<sup>-/-</sup> ASC (P3), and fresh tissues were collected. **d** MA plot illustrating differentially expressed retrotransposons of the LINE1, ERV1, ERVL and ERVK classes in *Terc*<sup>-/-</sup> and WT cells or tissues. Significantly changed retrotransposons are shown in different colors according to class. **e** Heatmap of all differentially expressed TEs by fold change in short telomere tissues or cells compared with WT tissues or cells. The bar color represents the fold change of the expression of TEs in *Terc*<sup>-/-</sup> vs WT cells; red indicates retrotransposons upregulated in *Terc*<sup>-/-</sup> vs WT cells; blue indicates retrotransposons downregulated in *Terc*<sup>-/-</sup> vs WT cells.

cell types (Fig. 1b). The strongest changes in gene expression among the somatic tissues and cells assessed occurred in *Terc*<sup>-/-</sup> testes, most likely because the testis contains a variety of cell types (Fig. 1c). *Terc*<sup>-/-</sup> ESCs displayed more changes in gene expression than somatic cells such as TTFs and ASCs (Fig. 1c). This might be because ESC cultures undergo more cell passages and rapid cell proliferation, which resulted in short telomeres in the absence of telomerase, whereas only three passages were performed for the somatic cells, as explained above.

The expression of retrotransposons was minimally affected in somatic cell TTFs, ASCs and tail tip samples isolated from G3 *Terc*<sup>-/-</sup> mice compared to WT cells (Fig. 1d). More retrotransposons were activated in liver and testis tissues with short telomeres, which consist of more diverse cell types than other somatic cell types (Fig. 1d). In contrast, more retrotransposons were altered and activated in G4 *Terc*<sup>-/-</sup> ESCs, derived from blastocysts of G3 *Terc*<sup>-/-</sup> mice, than in somatic cell types of G3 *Terc*<sup>-/-</sup> mice (Fig. 1d). Notably, retrotransposons with transposition activity including LINE1 family members were activated most in ESCs (Fig. 1d, e). Among the different cell types, testes and ESCs demonstrated the most changes in retrotransposon expression (Fig. 1d, e). This is probably due to the high proliferation rate of ESCs and spermatogonia stem cells in addition to the presence of more cell types within the testis, which also showed more gene expression changes after telomere shortening (Fig. 1c). Moreover, retrotransposon activation was mainly conserved across passages of ESCs (Fig. 1d). These results suggest that telomere length is a key factor in repressing retrotransposons across different cell types. Furthermore, *Terc*<sup>-/-</sup> ESCs, with the shortest telomeres, showed activation of retrotransposons from all transposable element (TE) families, whereas only a minority of retrotransposon classes were activated in other tissues or cell types where telomeres were not critically shortened (Fig. 1e). Cells with higher proliferation activity and

without telomerase exhibited a more rapid shortening of telomeres. Critically short telomeres in G4 *Terc*<sup>-/-</sup> ESCs after in vitro culture are associated with retrotransposon derepression (Fig. 1e). In contrast, retrotransposons, which were downregulated in cells with short telomeres, were fewer in number and were expressed at overall low levels in WT ESCs (Fig. 1e and Supplementary Fig. S2a). In addition, unlike the upregulated retrotransposons, the downregulated retrotransposons, such as short interspersed nuclear elements (SINEs), are nonautonomous. We speculate that the downregulation of these retrotransposons could be an indirect effect caused by other alterations in G4 *Terc*<sup>-/-</sup> ESCs, which remain to be understood. Therefore, our study mainly focused on the upregulated retrotransposons such as LINE1s, which have been known to retrotranspose<sup>33,34</sup>.

Given that G4 *Terc*<sup>-/-</sup> ESCs manifested the most retrotransposon expression changes, we mainly focused on the dysregulation of retrotransposons by short telomeres. To understand whether the activation of retrotransposons resulted from the loss of telomerase *Terc* or from the critically short telomeres themselves, we examined the expression profile of G1 *Terc*<sup>-/-</sup> ESCs compared with WT ESCs at relatively early passages, in which *Terc* was deleted and telomeres were also shortened but still maintained sufficient length (Fig. 2a and Supplementary Fig. S1a, b). The TE expression profile of G1 *Terc*<sup>-/-</sup> ESCs resembled that of WT ESCs, but the TE profile of G4 *Terc*<sup>-/-</sup> ESCs with the shortest telomeres differed from that of G1 *Terc*<sup>-/-</sup> ESCs as well as that of WT ESCs (Fig. 2b). Similar to the case in WT ESCs, retrotransposons were effectively silenced in G1 *Terc*<sup>-/-</sup> ESCs, which had slightly shorter telomeres than WT ESCs (Fig. 2b), in contrast to retrotransposon activation in G4 *Terc*<sup>-/-</sup> ESCs (Fig. 2b, c). Activation of retrotransposons was also found in G3 *Terc*<sup>-/-</sup> ESCs with comparatively shorter telomeres (Fig. 2c). Retrotransposons belonging to the ERVK retrovirus family (ERV3, ERV4, IAPEY, and IAPEY3\_I) were upregulated in both G3 and G4



**Fig. 2** (See legend on next page.)

(see figure on previous page)

**Fig. 2 Telomere length influences the expression of subtelomeric TEs in ESCs.** **a** Representative telomere Q-FISH images of WT, G1 *Terc*<sup>-/-</sup> and G4 *Terc*<sup>-/-</sup> ESCs. Blue, chromosomes stained by DAPI; green dots, telomeres. **b** Expression of individual TE loci in G1 and G4 *Terc*<sup>-/-</sup> ESCs, compared to WT ESCs. The color bar represents the normalized Z-score value of the TEs. The *P* value cutoff for expression loci was < 0.05, and the fold change cutoff was > 2. Each row represents a transposon locus. **c** Common retrotransposon classes dysregulated in both G3 and G4 *Terc*<sup>-/-</sup> ESCs, compared to WT ESCs. The color bar represents the normalized Z-score value of the TEs. Each row represents a class of TEs. **d, e** Enrichment of TEs in G1 *Terc*<sup>-/-</sup> ESCs located within 1 Mb (**d**) or 20 Mb (**e**) from the telomere. **f, g** Enrichment of TEs in G3 *Terc*<sup>-/-</sup> ESCs located within 1 Mb (**f**) or 20 Mb (**g**) from the telomere. **h, i** Enrichment of TEs in G4 *Terc*<sup>-/-</sup> ESCs located within 1 Mb (**h**) or 20 Mb (**i**) from the telomere. The expected value was calculated by using the number of upregulated TEs expected to be located within the subtelomeric region by chance; the actual genomic density or distributions of each family of retrotransposons were used as the baseline. As the number of the dysregulated TEs was distinct between different generations of ESCs, the expected values are not the same in each case. The calculation method is described in detail in Materials and methods. Fisher's exact test was used to measure the significance of enrichment.

*Terc*<sup>-/-</sup> ESCs (Fig. 2c). Similar to ERVs, LINE1s (L1MdGf and L1MdTf) were upregulated in ESCs with shorter telomeres, compared with WT ESCs (Fig. 2c).

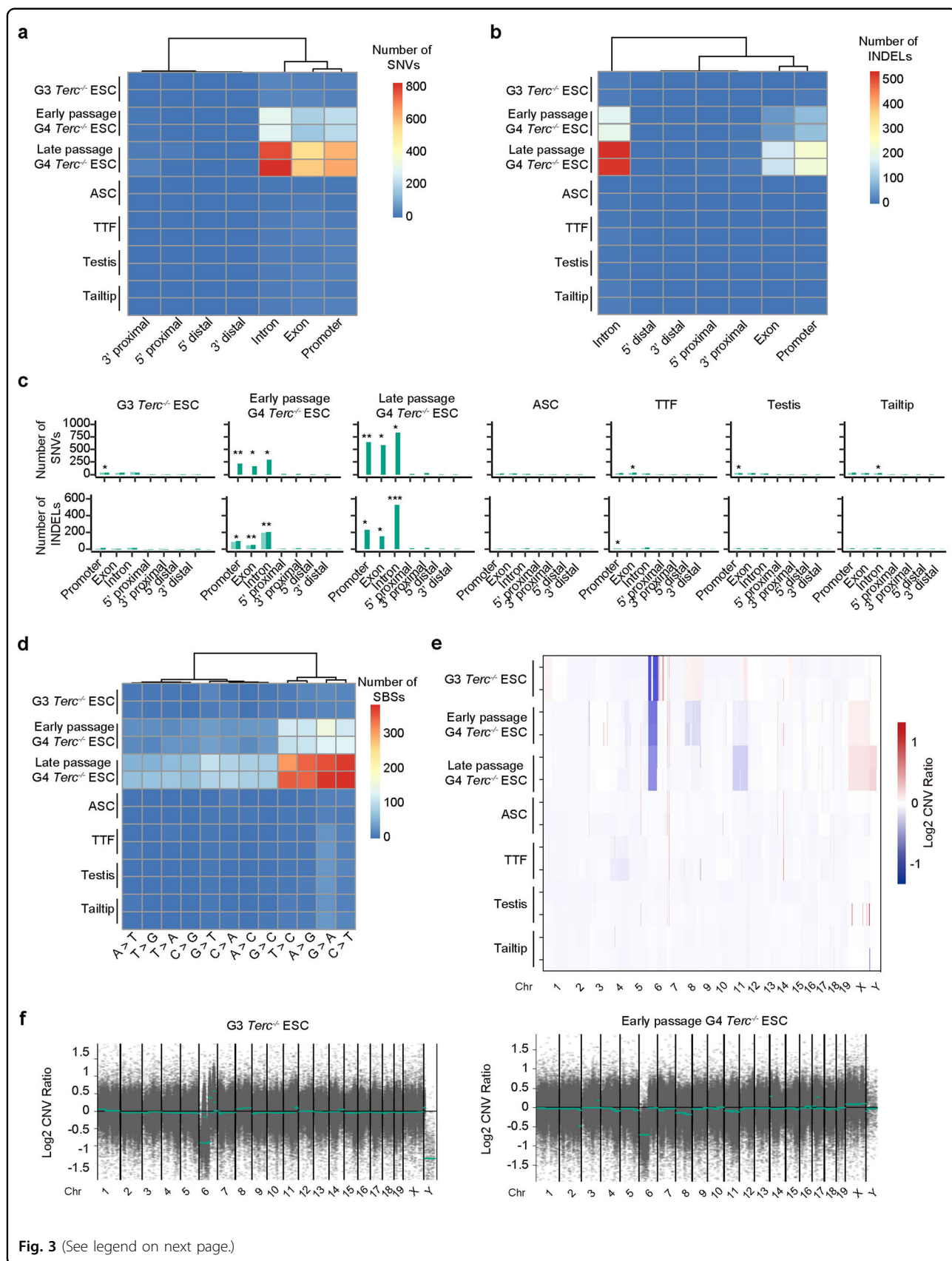
Next, we investigated the role of telomere length in regulating TEs by examining the genomic localization of dysregulated retrotransposons correlated with short telomeres in *Terc*<sup>-/-</sup> ESCs with increasing generations from G1 and G3 to G4 (Fig. 2d–i). In contrast to the non-enrichment of downregulated TEs, upregulated long interspersed nuclear elements (LINEs) and LTRs were both enriched in the 1 Mb subtelomeric region in G3 and G4 *Terc*<sup>-/-</sup> ESCs (Fig. 2f–i), suggesting that telomere length may regulate retrotransposons (Fig. 2f–i). Upregulated LINEs and LTR retrotransposons remained significantly enriched when the subtelomeric region was extended to 20 Mb in G3 *Terc*<sup>-/-</sup> ESCs but not in G1 *Terc*<sup>-/-</sup> ESCs (Fisher's exact test, adjusted *P* < 0.05; Fig. 2d–g). However, when we further extended the regions to 50 Mb from telomeres, retrotransposons were no longer enriched (Supplementary Fig. S2b). These results indicate that retrotransposon activation is linked to telomere position and affected by telomere length. The downregulation of TEs may not be a direct consequence of telomere shortening. Hence, their expression was not enriched at subtelomeric regions (Fig. 2d–i). In addition, recent reports suggest that telomeres can affect distant as well as nearby chromatins<sup>35,36</sup>. This may explain the change in expression of retrotransposons distant from telomeres. Telomeres regulate a subset of retrotransposons instead of all retrotransposons, consistent with the concentration of upregulated elements in subtelomeric regions. Together, these results demonstrate that critically short telomeres alter the transcription of retrotransposons, regardless of cell types.

### Short telomeres increase genomic instability

Next, we asked whether the genomic integrity or stability was altered by telomere attrition. We carried out exome-seq of WT and *Terc*<sup>-/-</sup> cells of various types in biological replicates. To minimize the impact of background single nucleotide variants (SNVs), we normalized the exome-seq

data of *Terc*<sup>-/-</sup> cells to that of corresponding WT cells. The number of SNVs was generally low in the WT tissues or somatic cells examined and did not change substantially in *Terc*<sup>-/-</sup> cells with relatively short telomeres (Fig. 3a). Surprisingly, the number of SNVs markedly increased in G4 *Terc*<sup>-/-</sup> ESCs at both early (P18) and late passages (P42), but only slightly increased in G3 *Terc*<sup>-/-</sup> ESCs (P22), TTFs, testis and tail tip samples (Fig. 3a–c). Similarly, the number of indels did not increase in *Terc*<sup>-/-</sup> somatic cells except for TTFs; rather, indels mainly increased in G4 *Terc*<sup>-/-</sup> ESCs at both early and late passages (Fig. 3b). G4 *Terc*<sup>-/-</sup> ESCs at late passages exhibited more SNVs and indels than those at early passages (Fig. 3a–c). Both the increased SNVs and indels were essentially associated with intron, exon and promoter regions (Fig. 3c), indicating a possible association of genomic instability with gene expression. However, genes with SNVs or indels did not exhibit expression changes at the transcription level (Supplementary Fig. S3a). We further examined the signature of SNVs. Unexpectedly, SNVs did not occur proportionately among different nucleotides. The majority of emerged SNVs are A/G or C/T SNVs (Fig. 3d), implying that the chemical structure of nucleotides may affect SNV occurrence. Increases in copy number variations (CNVs) were found in G4 *Terc*<sup>-/-</sup> ESCs at both early and late passages, but were minimal in other cell types (Fig. 3e). Interestingly, a portion of chromosome 6 was deleted in both G3 and G4 *Terc*<sup>-/-</sup> ESCs (Fig. 3e, f). These data suggest that mutations and genome instability are coincident with retrotransposon activation in both G3 and G4 *Terc*<sup>-/-</sup> ESCs.

Additionally, we analyzed the potential effects (if any) of short telomeres on retrotransposons and genomic stability in human cancer cells. We took advantage of published data that included simultaneous measurements of telomere length and transcriptome analysis from the same single cell<sup>5</sup> and analyzed selected cells with the longest and shortest telomeres. Different groups of retrotransposons were activated by shorter telomeres in human colorectal tumor cells than in tumor cells with long telomeres (Supplementary Fig. S3b, c). By using the transcriptome data<sup>5</sup>, we inferred that CNVs were enriched



**Fig. 3** (See legend on next page.)

(see figure on previous page)

**Fig. 3 Genomic changes in various cell types or tissues with short telomeres.** **a, b** Heatmap showing the number of SNVs (**a**) and indels (**b**) occurring in the genomes of two biological replicates according to exome-seq data. Promoter, 2 kb around transcriptional start site (TSS); 5' proximal, 2–10 kb upstream of gene; 5' distal, 10–100 kb upstream of gene; 3' proximal, 0–10 kb downstream of gene; 3' distal, 10–100 kb downstream of gene. The data were normalized to that of corresponding WT cells. **c** Bar plot of the number of SNVs and indels in the genome. Promoter, 2 kb around TSS; 5' proximal, 2–10 kb upstream of gene; 5' distal, 10–100 kb upstream of gene; 3' proximal, 0–10 kb downstream of gene; 3' distal, 10–100 kb downstream of gene. The data were normalized to that of WT cells. Green bars represent *Terc*<sup>-/-</sup> cells or tissues from G3 *Terc*<sup>-/-</sup> mice. *Terc*<sup>-/-</sup> ASCs, which had the lowest number of total SNVs and indels, were used as the control to determine whether SNVs and indels in other types of cells or tissues were significantly increased at different genomic locations. Statistical analysis was determined by Student's *t*-test, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001. **d** Heatmap showing the number of single-base substitutions (SBSs). Red indicates a larger quantity; blue indicates a smaller quantity. **e** CNV profiles of various cell types or tissues from G3 *Terc*<sup>-/-</sup> mice compared to WT controls. Blue represents copy number deletion, and red represents copy number amplification. **f** CNVs in different chromosomes of G3 *Terc*<sup>-/-</sup> ESCs or early-passage G4 *Terc*<sup>-/-</sup> ESCs. The CNV ratio indicates the fold change in copy number relative to that in the WT sample. CNV regions were filtered with the parameter 'depth > 10' to remove the low-confidence regions.

in chromosomes 1, 6, 11, or 17 of human cancer cells with short telomeres in comparison with those cells with long telomeres (Supplementary Fig. S3d). These data suggest that human colorectal cancer cells are more sensitive to telomere attrition-related retrotransposon activation and genomic instability.

#### Telomere attrition alters transcription and impacts genome stability

Given that G4 *Terc*<sup>-/-</sup> ESCs were derived from blastocysts developed from intercrosses of G3 *Terc*<sup>-/-</sup> mice and exhibited the shortest telomeres as well as the most mutations, we focused on these cells in our analyses of genomic stability. We performed an integrated functional analysis and showed that activated genes in G4 *Terc*<sup>-/-</sup> cells were enriched with similar functional terms (Supplementary Fig. S4a). The enriched pathways in G4 *Terc*<sup>-/-</sup> cells were related to cell death and negative regulation of cell proliferation (Supplementary Fig. S4a). Genes related to pathways in cancer and the inflammatory response were also upregulated in G4 *Terc*<sup>-/-</sup> cells but not in G3 *Terc*<sup>-/-</sup> cells (Supplementary Fig. S4a–d). Interestingly, the testis, which contains spermatogonia stem cells, and ESCs exhibited enrichment of similarly upregulated pathways resulting from short telomeres. Genes that were downregulated in G4 *Terc*<sup>-/-</sup> cells, but not in G3 *Terc*<sup>-/-</sup> cells, were enriched in various pathways, including cellular response to DNA damage stimulus and DNA repair (Supplementary Fig. S4b, e), suggesting impairment of these biological processes by critically short telomeres in G4 *Terc*<sup>-/-</sup> cells. Moreover, 2-cell genes were found to be enriched among the genes upregulated, but not among the genes downregulated, in telomere-shortened ESCs (Supplementary Fig. S4f). Genes related to the piRNA pathway, which is known to repress retrotransposons, were also repressed in *Terc*<sup>-/-</sup> testis cells and late-passage G4 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S4f, g). This may explain the change in retrotransposon expression in *Terc*<sup>-/-</sup> testes (Fig. 1c, d and Supplementary Fig. S4f, g). These data indicate that retrotransposon activation is accompanied by compromised

DNA damage repair pathways and activated cancer-related pathways following telomere attrition.

#### Telomere shortening reduces cell proliferation and increases apoptotic cell death

Furthermore, we evaluated the potential impact of increased DNA mutation on cell proliferation in G4 *Terc*<sup>-/-</sup> ESCs. Cell proliferation slowed after the loss of *Terc* in G1, G3 and G4 ESCs (Supplementary Fig. S5a). G4 *Terc*<sup>-/-</sup> ESCs showed even slower proliferation than G1 and G3 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5a). TUNEL assays and Annexin V immunostaining revealed an increase in apoptotic cells in the G4 *Terc*<sup>-/-</sup> population (Supplementary Fig. S5b–e). G1 and G3 *Terc*<sup>-/-</sup> ESCs had similar rates of apoptosis, lower than that of G4 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5b–e). Consistently, only G4 *Terc*<sup>-/-</sup> ESCs, but not G1 or G3 *Terc*<sup>-/-</sup> ESCs, exhibited elevated DNA damage, as evidenced by  $\gamma$ H2AX immunofluorescence staining patterns (Supplementary Fig. S5f, g). The number of  $\gamma$ H2AX foci was higher in G4 *Terc*<sup>-/-</sup> ESCs than in G1 and G3 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5g). These data show that telomere shortening leads to increased DNA damage and apoptosis. Given that DNA damage repair is required to maintain genome stability<sup>37</sup>, the increased incidences of DNA damage (Supplementary Fig. S5f, g) and suppression of the DNA repair pathway (Supplementary Fig. S4e) also explain why mutations occurred in G4 *Terc*<sup>-/-</sup> ESCs.

To explore the role of LINE1s in the proliferation and telomere length maintenance, we treated G4 *Terc*<sup>-/-</sup> ESCs with the reverse transcriptase inhibitor 3'-azido-3'-deoxythymidine (AZT). Indeed, inhibition of LINE1 reverse transcriptase caused cell apoptosis in both WT and G4 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5h, i), implying that LINE1 elements do not contribute to the selective growth difference.

#### Retrotransposition links critically short telomeres to genome instability

We further performed third-generation sequencing of the mouse genome to identify whether genome stability



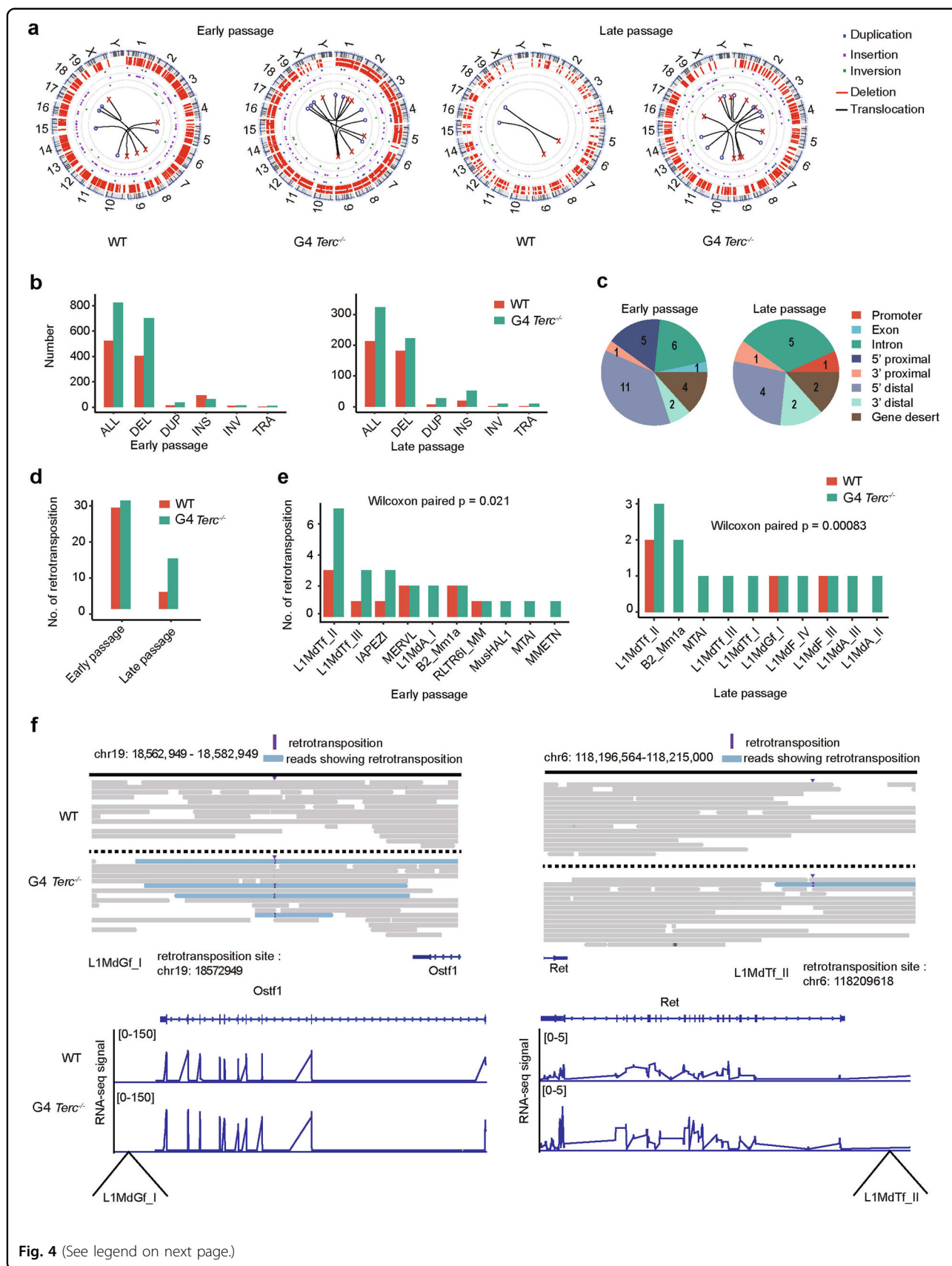
was affected by telomere attrition. In previous work, third-generation sequencing effectively detected structural variations in the genome but exhibited only limited accuracy in the identification of CNVs and SNVs<sup>38</sup>. Hence, our analysis focused on the structural variations in the genome after telomere shortening. Incidences of DNA fragment deletion, duplication, insertion, inversion, and chromosomal translocation all increased in both early (P17)- and late (P42)-passage G4 *Terc*<sup>-/-</sup> ESCs in comparison with respective WT ESCs (P17 or P42) (Fig. 4a). Deletions appeared to be the most frequent genomic instability events that were associated with short telomeres, whereas translocation was the least frequent event in both early- and late-passage G4 *Terc*<sup>-/-</sup> ESCs (Fig. 4b). Deletion, duplication, insertion and inversion events increased across different genomic regions except for exons (Supplementary Fig. S6a), presumably because exon disruption is detrimental to gene and cell functions. Interestingly, there was a decrease in the number of chromatin structure variations in late-passage ESCs compared with early-passage ESCs (Fig. 4a, b). This could be because ESCs without mutations proliferate faster than those with more genomic defects during passaging, such that ESCs with mutations are eventually outcompeted by WT ESCs without mutations.

Furthermore, since TEs were aberrantly activated after telomere attrition, we examined whether any retrotransposition events took place. Most retrotransposons were inserted into non-exon regions such as 5' distal and intron regions (Fig. 4c). The frequency of retrotransposition was slightly higher in G4 *Terc*<sup>-/-</sup> cells than in WT cells at early passages, and the difference notably increased in G4 *Terc*<sup>-/-</sup> cells at late passages. We speculate that sufficient replication time with increasing passages is required for retrotransposition to take place. The number of retrotransposition events in G4 *Terc*<sup>-/-</sup> ESCs at late passages was increased compared with that of WT ESCs at the same passage, but lower than that of early-passage G4 *Terc*<sup>-/-</sup> ESCs (Fig. 4d). This could be due to survival selection during the passages of ESCs. ESCs with severe genome instability (Fig. 4b–e), may undergo apoptotic cell death, as observed in G4 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5a–e), while only cells with more stable genomes survived passaging. This was supported by the fact that telomere shortening can activate an ATM-driven DNA damage response (DDR), which may eventually lead to aneuploidy and mitotic catastrophe<sup>39</sup>. It is possible that the retrotransposition events observed in cells may change after passaging. Therefore, we did not consistently observe retrotransposition events occurring at the same genomic loci. Although the absolute number of retrotransposition events was lower in late-passage than early-passage G4 *Terc*<sup>-/-</sup> ESCs (Fig. 4d), we noticed that retrotransposition events were much more significant

in late-passage than early-passage G4 *Terc*<sup>-/-</sup> ESCs compared to the corresponding WT cells ( $P = 0.00083$  vs  $P = 0.021$ ) (Fig. 4e). More active retrotransposition was associated with more SNVs and indels. Retrotransposition was prominent for young LINE1 family members (L1MdTf, L1MdA, L1MdG) (Fig. 4e and Supplementary Fig. S6b, c), but less frequent for other retrotransposons in telomere-shortened cells (Fig. 4e). Activation of LINE1 is a prerequisite for LINE1 retrotransposition. Additionally, the insertion of retrotransposons into non-gene body regions may affect gene expression. This was exemplified by L1MdGf insertion near *Ostf1* and L1MdTf insertion near *Ret* (Fig. 4f). The retrotransposition event was accompanied by the upregulation of the nearby oncogenes *Ret* and *Ostf1* (Fig. 4f), probably caused by the enhancer function of LINE1. These data suggest that increased expression of retrotransposons influences genome stability and gene expression by retrotransposition.

### Genomic 3D structure is altered in cells with critically short telomeres

Retrotransposons have also been shown to act as enhancers that can interact with and promote the transcription of other genomic loci through chromatin looping<sup>40,41</sup>. It remains unclear whether telomere length influences chromatin 3D structure. Therefore, we analyzed the 3D chromatin structures of G4 *Terc*<sup>-/-</sup> compared with WT ESCs by Hi-C. The deletion of chromosome 6 was again observed, as shown by the reduction in chromosome interactions for a region of chromosome 6 in comparison with other chromosomes (Fig. 5a–c), supporting the CNV change observed in the analysis of exome-seq data (Fig. 3e, f). We calculated the number of interactions at 60 Mb subtelomeric regions (Fig. 5d). Global chromatin interactions appeared to be increased by telomere shortening (Fig. 5a, b). Interestingly, the number of chromatin looping interactions obviously increased at both the 5' and 3' subtelomeric regions (Fig. 5e, f). This was exemplified by the cases of chromosomes 2 and 17 telomeres of G4 *Terc*<sup>-/-</sup> ESCs (Fig. 5g, h). The number of subtelomeric chromosomal interactions on chromosomes 2 and 17 increased (Fig. 5g, h). To test whether the increased contacts are specific to subtelomeres, we compared the interactions of chromosome centers with those of subtelomeres. By normalizing the percentage of interactions against the total number of interactions, we found that the interactions at chromosome centers decreased after telomere shortening (Supplementary Fig. S6d, e), whereas the percentage of interactions specifically increased at subtelomeric regions (Supplementary Fig. S6d, e). Previous studies suggested a positive correlation between the number of chromosomal interactions and gene activation<sup>42,43</sup>. These results indicate that telomere attrition promotes chromatin interactions at subtelomeric regions, consistent with the finding that



(see figure on previous page)

**Fig. 4 Retrotransposition driven by short telomeres.** **a** Circos plot of chromatin structure variations in WT or G4 *Terc*<sup>-/-</sup> ESCs. The chromosome number is indicated on the outside of the Circos plot. **b** Numbers of five different chromatin structure variations in WT cells or G4 *Terc*<sup>-/-</sup> ESCs. DEL deletion, DUP duplication, INS insertion, INV inversion, TRA translocation. **c** Pie chart showing the proportion of retrotransposition events occurring in the genome. Promoter, 2 kb around TSS; 5' proximal, 2–10 kb upstream of gene; 5' distal, 10–100 kb upstream of gene; 3' proximal, 0–10 kb downstream of gene; 3' distal, 10–100 kb downstream of gene; gene desert, > 100 kb away from the nearest gene in WT cells and *Terc*<sup>-/-</sup> ESCs at early or late passage. The number of the retrotransposition events was included on the pie chart. **d** Numbers of TE insertions in WT or G4 *Terc*<sup>-/-</sup> ESCs. **e** Number of TE insertions within specific TE classes in WT or G4 *Terc*<sup>-/-</sup> ESCs. **f** Examples of the locations of L1MdGf\_I and L1MdTf\_II insertion sequencing tags in WT and G4 *Terc*<sup>-/-</sup> ESCs at late passage. The reads supporting the insertion are marked in blue, and the insertion sites are marked with a capital "I". RNA-seq signals of *Ostf1* around the L1MdGf\_I insertion site and *Ret* around the L1MdTf\_II in WT and late-passage *Terc*<sup>-/-</sup> ESCs are shown below the respective genomic regions.

retrotransposons were upregulated at subtelomeric regions because retrotransposons can act as enhancers mediating chromosomal interactions<sup>40,41</sup>.

#### Chromatin accessibility regulates telomere-mediated retrotransposon silencing

The expression of genes and TEs is changed by telomere attrition. We asked whether chromatin accessibility was altered by short telomeres. We performed ATAC-seq to investigate how chromatin structure regulates retrotransposons after telomere attrition. Gene TSSs became more open in G4 *Terc*<sup>-/-</sup> than in WT ESCs (Fig. 6a). Moreover, G4 *Terc*<sup>-/-</sup> ESCs had more ATAC-seq peaks than WT ESCs (Supplementary Fig. S7a, b), indicating that chromatin is more relaxed in ESCs with shorter telomeres. Chromatin accessibility also increased on MT2/MERVL, L1Md\_T, and L1Md\_Gf (Fig. 6b, c), which were upregulated in G4 *Terc*<sup>-/-</sup> ESCs. Furthermore, chromatin accessibility at *Ret* and *Ostf1* increased following LINE1 retrotranspositions in G4 *Terc*<sup>-/-</sup> ESCs (Fig. 6d, e), consistent with the RNA-seq data (Fig. 4f). ESCs sporadically express 2-cell genes and ERVs such as MERVL and Zscan4 in 1%–5% of the ESC population<sup>17,18</sup>. In agreement with increased expression levels of 2-cell genes and MERVL, promoter regions of 2-cell genes exhibited increased chromatin accessibility resulting from short telomeres (Fig. 6f). Additionally, in support of the observed gene expression changes (Supplementary Fig. S4a–d), DNA motifs associated with cancer-related transcription factors were found on the ATAC-seq peaks that were enriched in ESCs with critically short telomeres (Fig. 6g). These data indicate that elevated chromatin accessibility associated with shortest telomeres may account for the upregulation of specific genes and retrotransposons.

#### Retrotransposon activation is accompanied by a change in H3K9me3 enrichment

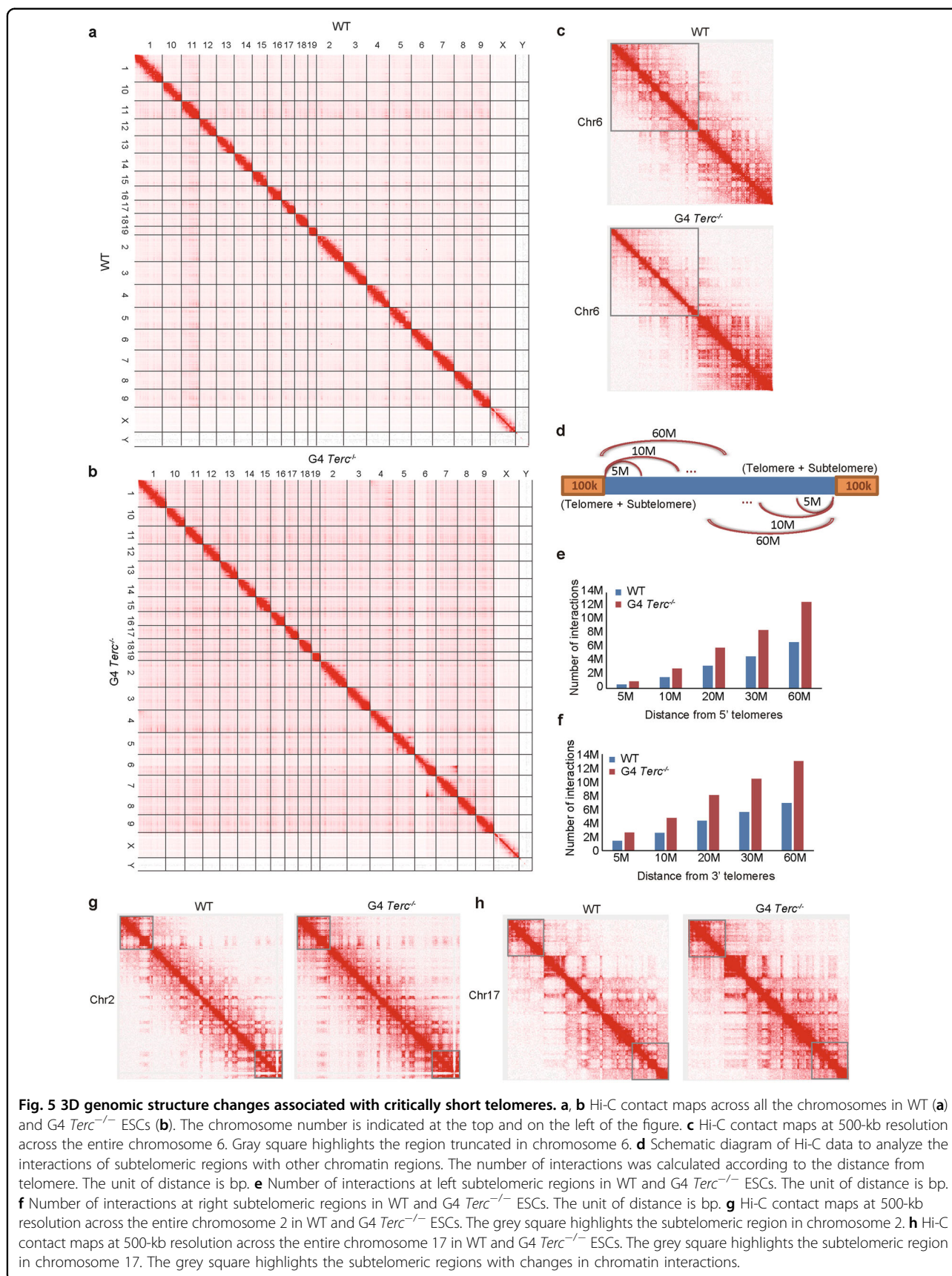
To identify whether chromatin openness is associated with epigenetic changes, we performed ChIP-seq analysis of H3K9me3 and H3K9me2. We identified 512 H3K9me3 peaks that were reduced or lost in G4 *Terc*<sup>-/-</sup> ESCs compared with WT ESCs, and none of these regions were

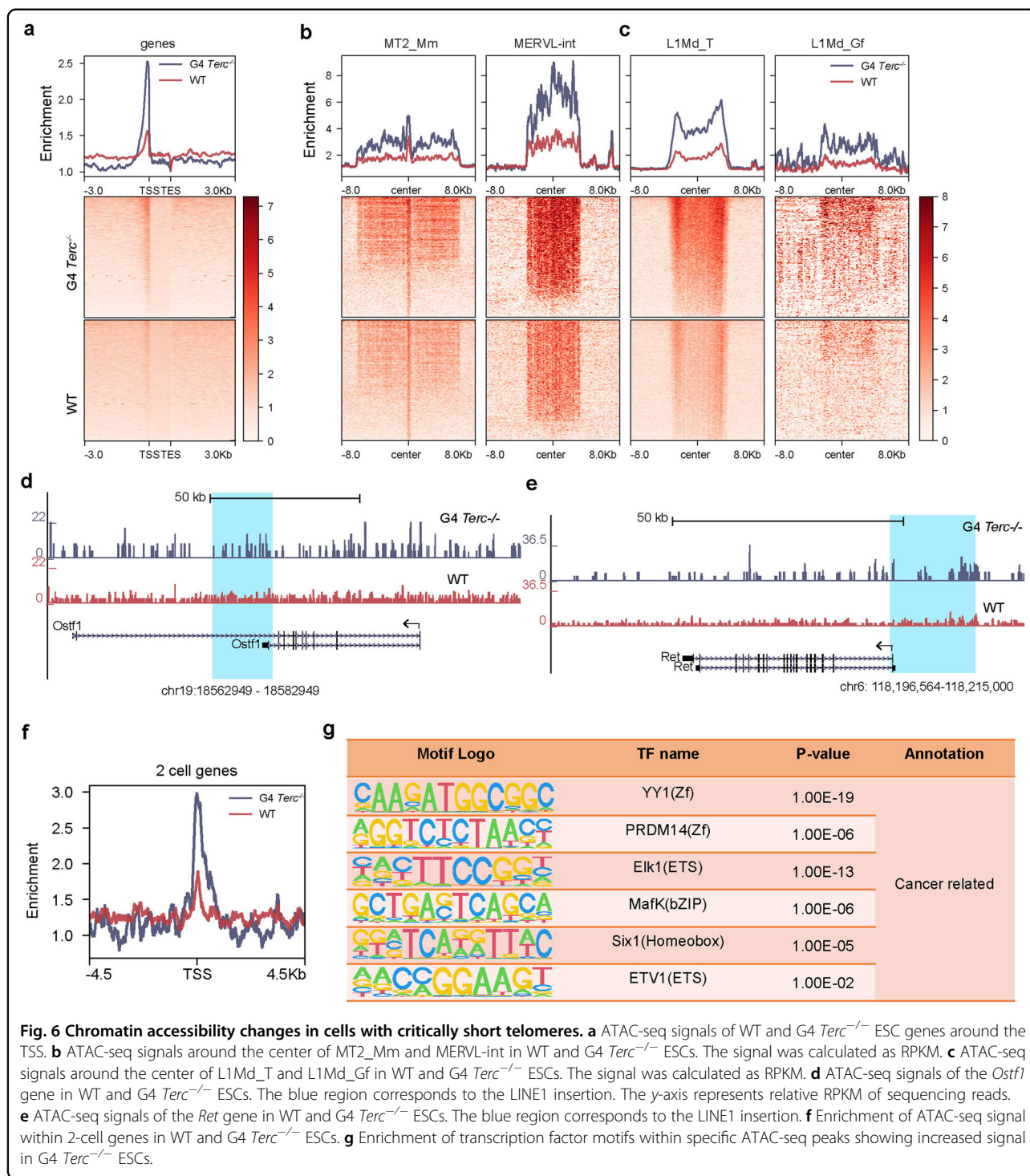
marked by H3K9me2 (Fig. 7a). Moreover, 3160 H3K9me3 reads were mapped to telomeres in WT ESCs in contrast to only 460 H3K9me3 reads on the telomeres of G4 *Terc*<sup>-/-</sup> ESCs (Fig. 7b). Among the 512 peaks with decreased H3K9me3 signals in G4 *Terc*<sup>-/-</sup> ESCs, 200 peaks (39%) were located at subtelomeric regions, and 123 out of these 200 peaks were at retrotransposons (Fig. 7c). These ChIP-seq data further confirmed that H3K9me3 was mainly located at telomeres and subtelomeres and could mediate the silencing of a subset of genes and retrotransposons, suggesting that telomere heterochromatin may regulate the expression of retrotransposons by interacting with neighboring and distant loci.

Genes close to the peaks with decreased H3K9me3 signal tended to be upregulated in G4 *Terc*<sup>-/-</sup> ESCs (Fig. 7d). Upregulation of the LINE1 family member L1MdGf, as revealed by RNA-seq analysis, was associated with reduced enrichment of H3K9me3 (Fig. 7e). To address whether a specific class of retrotransposons is regulated by telomere length, we mapped RNA-seq and ChIP-seq reads to the retrotransposon consensus sequences. Specific classes of retrotransposons, such as L1MdTf (LINE1) and L1MdGf (LINE1), were upregulated by telomere loss, accompanied by a reduction in H3K9me3 modification (Fig. 7f, g). The regions with decreased H3K9me3 peaks were also associated with a slightly increased ATAC-seq signal (Supplementary Fig. S7c). These results indicate the specificity of telomeres in repressing retrotransposons. Together, these findings suggest that H3K9me3 enrichment contributes at least partly to the telomeric repression of specific retrotransposons.

#### Re-elongation of telomeres by the introduction of *Terc* restores telomere-mediated retrotransposable element silencing

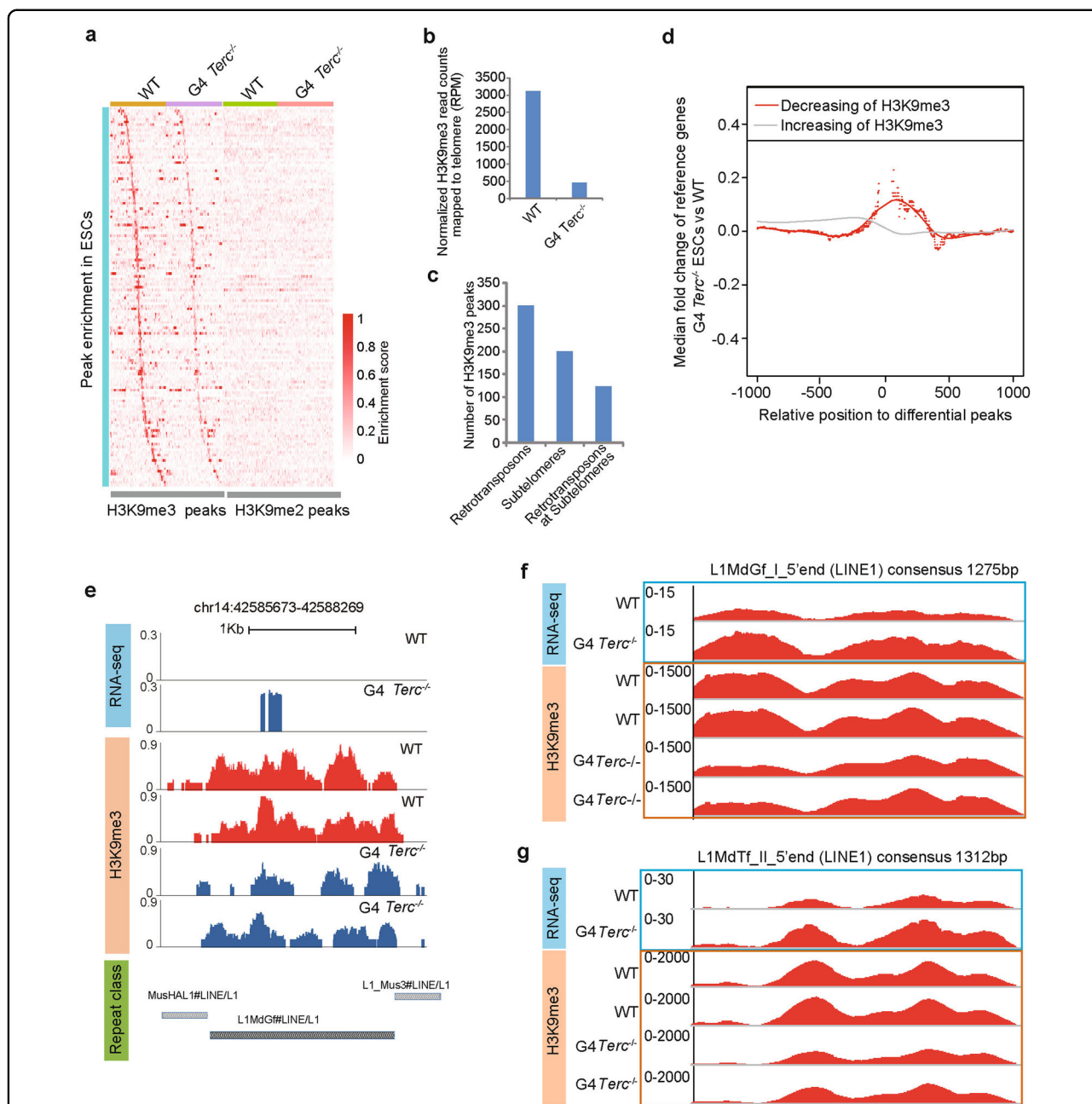
To further substantiate the role of telomeres in silencing retrotransposons, we performed rescue experiments by knock-in of *Terc* via CRISPR/Cas9 into G4 *Terc*<sup>-/-</sup> ESCs (P20) to rejuvenate telomeres (Fig. 8a). We validated the knock-in of *Terc* by PCR (Supplementary Fig. S7d). A sufficient number of passages following *Terc* knock-in were necessary for elongation of telomeres and recovery





of telomerase. Hence, we continued culturing the cells for 22 more passages to obtain P42 G4 *Terc*<sup>-/-</sup> ESCs. Real-time PCR showed that the reduced expression of *Terc* in G4 *Terc*<sup>-/-</sup> ESCs was restored in *Terc*-rescued ESCs (P42) (Fig. 8b). Telomere length was successfully rejuvenated by the introduction of *Terc* (Fig. 8c, d). A higher level of telomere signals suggested longer telomeres

(green) similar to those in WT cells in contrast to the shortened telomeres and telomere signal-free ends and chromosome fusion in G4 *Terc*<sup>-/-</sup> cells, indicating that the telomeres are restored in the G4 *Terc*<sup>-/-</sup> rescue cells following targeted knock-in of *Terc* and recovery of telomerase (Fig. 8d). Southern blot data from TRF analysis also validated the elongation of telomeres after the



**Fig. 7** H3K9me3 regulates the retrotransposon activity induced by critically short telomeres. **a** ChIP-seq read density heatmaps for differentially enriched H3K9me3 peaks around the TSS ( $\pm$  50 kb) and the corresponding H3K9me2 signal in WT and G4 *Terc*<sup>-/-</sup> ESCs. The relative density of H3K9me3 enrichment was presented in red, as indicated by the scale bar. The heatmap was generated using NGSPLLOT. **b** H3K9me3 enrichment on telomeres. The number of reads was mapped to telomere (TTAGGG) repeats and normalized to the total number of reads in each sample. The number of reads was used to indicate H3K9me3 abundance at telomeres in two biological replicates. **c** The number of downregulated H3K9me3 peaks on retrotransposons, subtelomeres and subtelomeric retrotransposons in G4 *Terc*<sup>-/-</sup> ESCs. H3K9me3 peaks indicate their genome-wide enrichment, excluding telomere regions. **d** Metagene expression analyses of genes flanked by H3K9me3 peaks. For both the 5' and 3' genomic directions of H3K9me3 peaks, the median log<sub>2</sub> fold changes of genes around peaks with decreasing (in red) or increasing H3K9me3 (in grey) in different window sizes are plotted. **e** L1MdGf as an example of expression and H3K9me3 change by short telomeres. **f, g** Coverage plot of ChIP-seq and RNA-seq reads on the consensus sequence of LINE1 family member L1MdTf (**f**) and L1MdGf (**g**). The y-axis indicates relative RPKM of sequencing reads. The x-axis indicates the full length of the corresponding retrotransposons.

introduction of *Terc* into *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S1a, b). However, the chromosome fusions present in the original G4 *Terc*<sup>-/-</sup> ESC line were still retained in the telomere-rejuvenated G4 *Terc*<sup>-/-</sup> ESCs even though telomerase was recovered (Fig. 8d). RNA-seq analysis also demonstrated the re-suppression of the expression of retrotransposon, such as IAP members and LINE1 members, by elongated telomeres (Fig. 8e). For example, the expression of LINE1 (L1Md\_T) was rescued by elongated telomeres, accompanied by a partial reduction in H3K9me3 (Fig. 8e, f). Gene expression near the H3K9me3-marked ERVB4\_2-LTR was also restored after telomere elongation (Supplementary Fig. S7e). Hence, re-elongation of telomeres partially restores retrotransposon suppression.

## Discussion

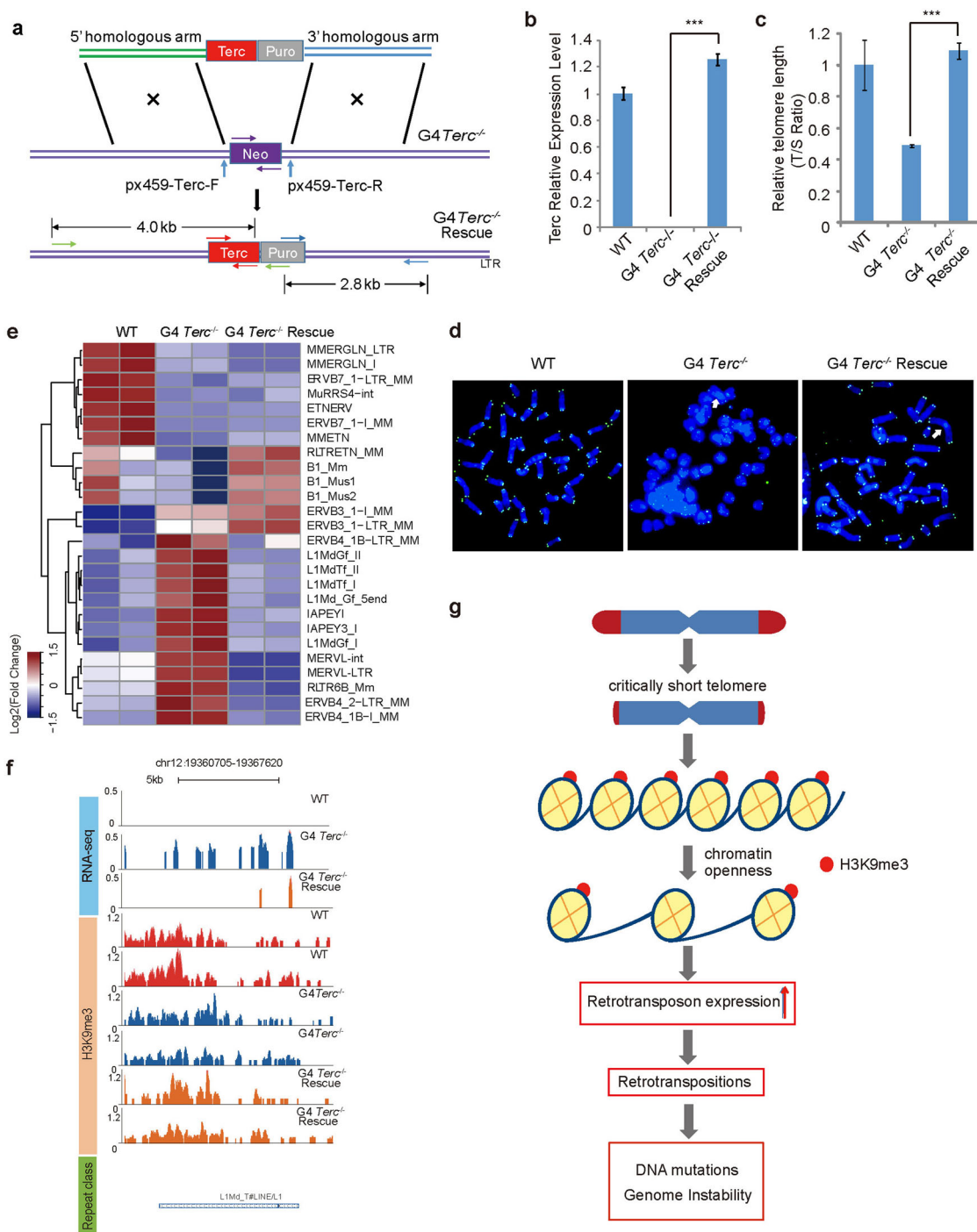
We show that retrotransposons can be activated in cells with very short telomeres and that this activity is linked with genomic instability and promotion of retrotransposition-induced genomic mutations and deletion or amplification; these changes are mediated by the relaxation of chromatin structure. On the basis of our data, we propose a model whereby cells with long telomeres silence a subset of retrotransposons and maintain genome stability (Fig. 8g). After telomere attrition, chromatin accessibility is increased, and retrotransposition occurs following the activation of transposons in a subset of chromatin regions (Fig. 8g). The retrotransposition events destabilize the genome, leading to mutations through the dysregulation of DDR, inflammatory responses and cancer-related pathways (Fig. 8g).

Although telomere attrition can be found in both G3 *Terc*<sup>-/-</sup> somatic cells and G1 *Terc*<sup>-/-</sup> ESCs, G4 *Terc*<sup>-/-</sup> ESCs exhibit the shortest telomeres as evidenced by substantial telomere signal-free ends, presumably because of the highly proliferative nature of ESCs during passaging. Nevertheless, ESCs with telomere loss rarely show chromosome fusions (Supplementary Fig. S1c), which is consistent with recent findings that telomere protection is achieved by distinct mechanisms in pluripotent cells<sup>44,45</sup>. When telomeres become critically short, they activate the DDR in both somatic cells and ESCs<sup>39,46,47</sup>. TRF2, a component of the shelterin complex, prevents telomeric DNA from being recognized as DNA damage, thus preventing the activation of DDR<sup>48</sup>. ESCs devoid of TRF2 at telomeres activate an attenuated telomeric DDR that is not accompanied by telomere fusions<sup>44,45</sup>. Although the specific mechanisms remain to be investigated, critically short telomeres also impair DNA repair pathways (Supplementary Fig. S4e). The finding that G3 *Terc*<sup>-/-</sup> ESCs did not exhibit increased SNVs could be explained by the fact that DNA damage did not occur frequently in G3 *Terc*<sup>-/-</sup> ESCs as in G4 *Terc*<sup>-/-</sup> ESCs

(Fig. 3c and Supplementary Fig. S5f, g). Furthermore, the genes downregulated in G4 *Terc*<sup>-/-</sup> ESCs, but not G3 *Terc*<sup>-/-</sup> ESCs, were enriched for the DNA repair pathway (Supplementary Fig. S4e). DNA repair factors can restrict LINE1 retrotransposition<sup>49</sup>. The undisrupted expression of DNA repair genes may be another reason that fewer mutations are observed in G3 *Terc*<sup>-/-</sup> ESCs than in G4 *Terc*<sup>-/-</sup> ESCs.

Moreover, short telomeres activated retrotransposons and chromosome 6 truncation occurred in both G3 and G4 *Terc*<sup>-/-</sup> ESCs. However, strong DNA damage, which is indicated by high levels of  $\gamma$ H2AX staining, occurred only in G4 *Terc*<sup>-/-</sup> ESCs in which telomeres became critically short and chromosomal fusions emerged (Supplementary Figs. 1a–c and S5f, g). Given that both G3 and G4 *Terc*<sup>-/-</sup> ESCs displayed chromosomal rearrangements, the increase in genomic mutations in G4 *Terc*<sup>-/-</sup> ESCs is unlikely to be caused by clonal expansion of G4 cell lines. Mutations in G4 *Terc*<sup>-/-</sup> ESCs were associated with increased DNA damage. Consistent with this observation, RNA-seq analysis also revealed specific downregulation of DNA damage-related genes in G4 *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S4e). These results imply that the extensive DNA damage may be the reason for the emergence of DNA mutations in G4 *Terc*<sup>-/-</sup> ESCs. The degree of telomere attrition affects retrotransposon activities, with critically short telomeres or telomere loss associated with increased retrotransposon activities. WT ESCs had the longest telomeres, and the shortest length was observed in ESCs after *Terc* deletion (Supplementary Fig. S1a, b). The telomeres of other tissues were also shortened but still maintained at a modest length after *Terc* deletion (Supplementary Fig. S1a, b).

Notably, the shortest telomeres were found in late-generation ESCs (G4 *Terc*<sup>-/-</sup>) without telomerase RNA, which provides an appropriate model for studying the impact of telomere length. Mouse cells generally have very long telomeres<sup>50</sup>; therefore presumably the shortest telomeres are not easily detectable in general mouse tissues or cells without telomerase. Mouse cells have much longer telomeres (50–100 kb) than human cells (5–10 kb)<sup>50</sup>. The telomeres in telomerase-knockout (*Terc*<sup>-/-</sup>) mice are shortened with increasing generations, and these mice can be bred for 6–7 generations<sup>29</sup>. Moreover, the length of the shortest telomeres, as evidenced by telomere signal-free ends or telomere loss, rather than the average telomere length, leads to chromosome instability<sup>27</sup>. In this study, telomere signal-free ends were not found in G3 *Terc*<sup>-/-</sup> MEFs but were notably observed in G6 *Terc*<sup>-/-</sup> MEFs, which exhibit chromosome fusion. Our study also suggested that the essential role of telomerase might not be to maintain average telomere length but rather telomere function at



**Fig. 8 Rejuvenation of telomere length suppresses retrotransposons.** **a** Illustration of the method used to knock in *Terc* into *G4 Terc<sup>-/-</sup>* ESCs using CRISPR/Cas9 in combination with homologous recombination. *Terc*-puro was knocked into the *Terc<sup>-/-</sup>* locus. **b** Relative expression levels of *Terc* in WT, *G4 Terc<sup>-/-</sup>* and *G4 Terc<sup>-/-</sup>* rescue ESCs. **c** Rescue of telomere length shown as the T/S ratio by real-time PCR. **d** Representative telomere Q-FISH images of WT, *G4 Terc<sup>-/-</sup>* and *G4 Terc<sup>-/-</sup>* rescue ESCs. Blue, chromosomes stained by DAPI; green dots, telomeres. White arrows indicate chromosome fusion events. **e** Expression of TEs after restoration of telomere length in *G4 Terc<sup>-/-</sup>* ESCs. The bar color indicates the Z-score value. **f** The UCSC profile displaying the signal of the RNA-seq and H3K9me3 ChIP-seq signals around L1Md\_T in WT, *G4 Terc<sup>-/-</sup>* and *G4 Terc* rescue ESCs. The y-axis represents the relative RPKM of sequencing reads. Repeat classes are indicated at the bottom of the figure in green-edged boxes. **g** A simplified model of telomere-mediated silencing of retrotransposons. Critically short telomeres are correlated with a loss of heterochromatin, which subsequently leads to retrotransposon activation, promoting genome instability.



critically short telomeres<sup>27</sup>. Breeding of the knockout mice in the hybrid genetic background requires significant amount of time, extending one or two years to achieve late generations of *Terc*<sup>-/-</sup> mice. Therefore, *Terc*<sup>-/-</sup> mice at inbred C57BL/6 background were derived and are preferred for use. Hence, in our study, G3 *Terc*<sup>-/-</sup> mice of the inbred C57BL/6 background were used for comparison with WT mice. Indeed, various somatic cells and ESCs of G3 *Terc*<sup>-/-</sup> mice exhibited shorter telomeres than WT somatic cells and ESCs (Supplementary Fig. S1a–c). However, we did not observe telomere loss or critically short telomeres in somatic cell types such as TTFs readily available from G3 *Terc*<sup>-/-</sup> mice (Supplementary Fig. S1a–d). We designated ESCs derived from G3 *Terc*<sup>-/-</sup> blastocysts as G4 *Terc*<sup>-/-</sup> ESCs which had much shorter telomeres than G3 *Terc*<sup>-/-</sup> somatic cells. Moreover, the shortest telomeres as evidenced by telomere signal-free ends or telomere loss were observed in G4 *Terc*<sup>-/-</sup> ESCs (Fig. 2a and Supplementary Fig. 1 c–e). Comparatively, G1 *Terc*<sup>-/-</sup> ESCs also showed shorter telomeres than WT ESCs but no telomere signal-free ends (Fig. 2a), yet the retrotransposon transcription profile appeared similar to that of WT ESCs and distinct from that of G4 *Terc*<sup>-/-</sup> ESCs (Fig. 2b). Similarly, low frequencies of SNVs and indels were found in both G3 *Terc*<sup>-/-</sup> somatic cell types and ESCs (Fig. 3a–c). These data show that responses to telomere shortening-induced transcriptomic alteration and genomic instability can be similar in ESCs and somatic cells.

Telomeres in somatic cells or tissues in mice are generally long, and telomerase knockout shortens somatic telomeres in live mice<sup>29</sup>. The relatively short telomeres in somatic cells and tissues do not show evident changes in the transcription of retrotransposons and genomic stability (Figs. 1d and 3). In contrast, the shortest telomeres, despite being in ESCs, exhibit aberrant upregulation of retrotransposons, retrotranspositions and genomic instability (Figs. 1d, 3 and 4e). However, the relatively short telomeres in *Terc*<sup>-/-</sup> ESCs do not induce these phenotypes (Fig. 2b, c), similar to those of *Terc*<sup>-/-</sup> somatic cells. The differences in telomere shortening between somatic cells and ESCs likely arise from the rapid proliferation of ESCs, which have a doubling time of ~7–10 h, unlike somatic cells which take ~24 h to complete cell cycle. The high proliferation rate of ESCs allows them to undergo more rounds of replication within a fixed period, hence telomeres shorten more rapidly in ESCs than somatic cells without telomerase. The mutation rate was also the lowest in cells with a low rate of cell division such as spermatogonia in testes<sup>7</sup>. Among tested somatic tissues or cells, spermatogonia in testes rank second (behind ESCs) in terms of proliferation rate, but proliferate much faster than other somatic cell types.

G4 *Terc*<sup>-/-</sup> ESCs without telomerase showed elevation of more young LINE1 elements (L1MdTf, L1MDa, L1MdG)<sup>51</sup> (Figs. 2c and 4e), which can actively retrotranspose in genome, leading to genome instability, whereas other elevated TEs, such as MERVL, may not retrotranspose and initiate genome instability. However, when *Terc* was deleted, the speed of cell proliferation was reduced, and replicative senescence occurred, accompanied by telomere attrition with passaging. Sufficient telomere length is critical for cell proliferation<sup>52</sup>. It has also been demonstrated in somatic cells that the shortest telomere length but not the average telomere length limits cell survival<sup>27</sup>. Our data indicated that the shortest telomeres might also be linked to elevated retrotransposon activity, genomic instability and cellular senescence. Retrotransposon activation associated with telomere attrition also contributes to cellular senescence<sup>23</sup>. Hence, cell proliferation cycles affect telomere attrition and retrotransposon activation. Telomere attrition and dysfunction due to telomerase deficiency leads to aging and age-associated diseases resulting from failed protection of chromosome ends, aberrant DDR at telomeres, and genomic instability<sup>53–56</sup>. Short telomeres are also frequently observed during cancer development<sup>57</sup>. Telomere attrition can suppress tumorigenesis by promoting cellular senescence and thus causing the eventual apoptosis of these cells<sup>58,59</sup>. However, cancer cells, in which apoptosis pathways may be dysregulated, can survive cellular senescence despite shortened telomeres. Some cancer cells eventually lose telomere protection at chromosome ends and reach a state of genome instability, which leads to further cancer progression<sup>60</sup>. Our study demonstrates that in addition to the known critical role of telomeres in cancer cells, critically short telomeres may also promote cancer progression indirectly by causing aberrant upregulation of retrotransposons.

Transformed cancer cells exhibit non-telomeric reverse transcriptase activity from retrotransposons. Pharmacological inhibition or RNAi depletion of retrotransposons suppresses proliferation of cancer cells<sup>61</sup>. LINE1 is important for telomere maintenance in telomerase-positive cancer cells<sup>62</sup>. Both LINE1 and another retrotransposon, HERVK, are implicated in the in vivo tumorigenic potential of melanoma cells<sup>61</sup>. In addition, inhibition of LINE1 retrotransposition in mouse 2-cell embryos prevents the activation of 2-cell embryo genes<sup>63</sup>. We also observed the maintenance of telomeres at critically short lengths and activation of LINE1 and 2-cell embryo genes after telomerase loss (Fig. 1d and Supplementary Figs. S1a, b and S4f). Hence, it is possible that cells activate LINE1 after telomere attrition to facilitate the maintenance of telomere length. LINE1 ribonucleoprotein particles (L1-RNPs) are elevated in ALT cells<sup>64</sup>. L1-RNPs interact with *Terra* and L1-RNP depletion

disrupts nuclear localization of *Terra*, which causes telomeric DNA damage and reduces the presence of C-circles<sup>64</sup>. In agreement with the findings in cancer cells, AZT-mediated LINE1 inhibition caused apoptosis in both WT and *Terc*<sup>-/-</sup> ESCs (Supplementary Fig. S5h, i). Accumulation of cytoplasmic retrotransposon DNA activates the innate immunity-associated inflammatory response<sup>22,65,66</sup>, which may cause subsequent DNA damage. Retrotransposon activation can also cause DNA damage and vice versa<sup>49,67,68</sup>. It is worth investigating the mechanism by which the mutation rate is increased by critically short telomeres in the future.

Among the active retrotransposons, both ERVs and LINE1 are activated by short telomeres in ESCs (Fig. 1d, e). The activation of MERVL can be attributed to the upregulation of *Zscan4* in G4 *Terc*<sup>-/-</sup> ESCs<sup>69</sup>. Activation of ERVs and LINE1s is frequently found in cancer cells as well<sup>70–72</sup>. These retrotransposons can lead to the aberrant gene expression, transposition, and genomic instability. Activated ERVs may serve as enhancers/promoters to drive the expression of oncogenes in cancer cells and the expression of ERV protein products may transform cells<sup>73</sup>. LINE1 was reported to drive the interferon response during cell senescence and cause inflammation<sup>22</sup>. This may also explain the activation of genes related to the inflammatory response by short telomeres in ESCs and testes (Supplementary Fig. S4a–c). Epigenetic activation of LINE1s and their retrotransposition were also found in various human diseases, related to genome instability<sup>74,75</sup>. Activation of LINE1 retrotransposition may further contribute to the alteration of gene expression, genome instability, mutations<sup>76–78</sup> and cancer development<sup>72</sup>. For example, the proto-oncogene *Ret* was activated after LINE1 retrotransposition into a nearby region (Figs. 4f and 6e). As the upregulated genes seem to be enriched in cancer-related pathways (Supplementary Fig. S4a, d), the genome of *Terc*<sup>-/-</sup> cells is further destabilized after telomere attrition. This is supported by the observed deletion of a large fraction of chromosome 6 in *Terc*<sup>-/-</sup> ESCs (Fig. 3e, f). The additional role for telomeres in silencing retrotransposons revealed here is fascinating. In addition to protecting chromatin ends from deterioration or fusion with other chromatin ends, telomeres maintain an inactive chromatin state, which is critical for suppressing retrotransposons to maintain genome stability.

Telomeres may also regulate retrotransposons through chromatin looping, which renders telomeres the ability to regulate genes at longer distances, e.g., 10 Mb away from telomeres<sup>79</sup>. Previously, Robin et al. directly mapped the interactions of specific telomere ends by using a Hi-C technique to enrich for specific genomic regions, and described how telomere length regulates gene expression long before telomeres become short enough to produce a

DDR<sup>79</sup>. We compared the chromosomal interactions of critically short telomeres with those of long telomeres in ESCs (Fig. 5a, b). Cells with critically short telomeres exhibit a dramatic reorganization of nearby chromatin structures (Fig. 5d–g), which may create active chromatin enabling transcriptional activation of relevant retrotransposons. The increase in the number of overall chromosomal interactions suggests an active chromatin state (Fig. 5a, b). Previous findings suggest a positive correlation between the number of chromosomal interactions and gene activation<sup>42,43</sup>. Consistent with this notion, chromatin accessibility at the retrotransposons was elevated in ESCs with critically short telomeres (Fig. 6b, c), and heterochromatin deposition at the related foci was reduced (Fig. 7f, g). Hence, future experiments can test the hypothesis that chromatin looping may allow telomeres to control distant retrotransposons.

## Materials and methods

### Mice and tissues

The use of mice for this research was approved by the Nankai University Animal Care and Use Committee. All mice used in this study were taken care of and operated according to the relevant regulations. Mice were housed and cared for in individually ventilated cages (IVCs) on a standard 12 h light:12 h dark cycle in the sterile animal facility. Tail tips, testis, adipose and liver were dissected from G3 *Terc*<sup>-/-</sup> mice and WT mice following humane euthanization.

### Cell cultures

Mouse adipose tissues were obtained from groin of G3 *Terc*<sup>-/-</sup> mice and paired WT mice. They were washed with PBS, aseptically isolated and digested with 0.1% collagenase I, then placed onto culture dishes, and incubated for 4 days in ASC culture medium (knockout-DMEM containing 10% FBS, 0.1 mM non-essential amino acids, 1 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin). Mouse ASCs were digested with 0.25% trypsin-ETDA (Gibco) and transferred to new plates and maintained in ASC culture medium.

Isolation and culture of adult mouse TTFs were carried out as previously described<sup>80</sup>. The tails from adult G3 *Terc*<sup>-/-</sup> mice and paired WT mice were peeled, minced into 1 × 1 mm<sup>3</sup> pieces and cultured for 7 days in MEF medium (DMEM containing 10% FBS, 1 mM L-glutamine, and 100 U/mL penicillin and 100 µg/mL streptomycin). Cells that migrated out of the grafted pieces were mostly TTF cells. TTFs were digested with 0.25% trypsin-ETDA (Gibco) and transferred to fresh plates and maintained in MEF culture medium.

G4 *Terc*<sup>-/-</sup> ESCs at the C57BL/6 genetic background were generated from blastocysts of *Terc*-deficient mice (G3 *Terc*<sup>-/-</sup> mice), G3 *Terc*<sup>-/-</sup> ESCs from G2 *Terc*<sup>-/-</sup>

mice and G1 *Terc*<sup>-/-</sup> ESCs from *Terc*<sup>+/-</sup> mice<sup>32</sup>. WT ESCs of the same C57BL/6 genetic background served as controls. ESCs were cultured as previously described<sup>32</sup>. The ESC culture medium consisted of knock-out DMEM (Invitrogen) containing 20% FBS (HyClone), 1000 U/mL mouse leukemia inhibitory factor (LIF; ESG1107, Millipore), 0.1 mM non-essential amino acids, 0.1 mM  $\beta$ -mecaptoethanol, 1 mM L-glutamine, penicillin (100 U/mL) and streptomycin (100  $\mu$ g/mL). For the culture of ESCs, the medium was changed daily, and cells were routinely passaged every two days using 0.25% trypsin-EDTA (Invitrogen, 25200-072).

#### Generation of *Terc* knock-in ESC lines

For *Terc* knock-in, the *Terc* repair donor sequences were based on the mouse genomic sequence and information<sup>29</sup>. The donor vector contained the *Terc* gene flanked by 5' (left) and 3' (right) homology arms. In addition, there was a puromycin-resistance cassette for drug selection between the *Terc* gene and the 3' homology arm. The DNA fragments were individually amplified using the proper primers and then cloned into one vector with the proper enzymes. G4 *Terc*<sup>-/-</sup> ESCs were transfected with two PX459 plasmids and *Terc* repair donor plasmids using Lipofectamine 2000 transfection reagent (Invitrogen). Twenty-four hours later, 2  $\mu$ g/mL puromycin was applied for 7 days to select positive clones. Single-cell clones were picked, and then the genomic DNA was extracted. To detect whether *Terc* was properly repaired, PCR was performed with primers as described (Supplementary Table S1).

#### Telomere Q-FISH

Telomere length and function (telomere integrity and chromosome stability) were estimated by Q-FISH as previously described<sup>53,81</sup>. Cells were incubated with 0.3  $\mu$ g/mL nocodazole for 3 h to enrich cells at metaphase. Chromosome spreads were made using a routine method. Metaphase-enriched cells were exposed to hypotonic treatment with 0.075 M KCl solution, fixed with methanol: glacial acetic acid (3:1), and spread onto clean slides. The slides were washed in PBS for 15 min, and the chromosomes were fixed in 4% formaldehyde in PBS for 2 min and washed three times in PBS. The slides were then dehydrated in 70%, 90%, and 100% ethanol for 5 min each time and subsequently air-dried. Telomeres were denatured at 80 °C for 3 min and hybridized with an FITC- or Cy3-labeled (CCCTAA)<sub>3</sub> peptide nucleic acid (PNA) probe at 0.5  $\mu$ g/mL (Panagene, Korea) for 2.5 h at 37 °C in a humidified chamber. Chromosomes were counter-stained with 0.5  $\mu$ g/mL DAPI. Fluorescence from chromosomes and telomeres was digitally imaged on a Zeiss Imager Z2 microscope with FITC/DAPI or Rhodamine/DAPI filters using AxioCam and

AxioVision software 4.6. For quantitative measurements of telomere length, telomere fluorescence intensity was integrated using the TFL-TELO program (kindly provided by P. Lansdorp).

#### TRF analysis by Southern blot

The average TRF length was determined according to the commercial kit (TeloTAGGG Telomere Length Assay, 12209136001; Roche Life Science). Genomic DNA was extracted by traditional phenol:chloroform:isoamyl alcohol method. 3  $\mu$ g DNA was digested with *Mbo*I (NEB) overnight, and the DNA fragments were separated by 0.8% agarose gel for 16 h at 6 V/cm in 0.5 $\times$  TBE buffer using CHEF DR-III pulse-field system (Bio-Rad). Gels were denatured, neutralized, and transferred to nylon membranes (RPN2020B; GE Healthcare) for 48 h. The membrane was hybridized with digoxigenin (DIG)-labeled telomere probe at 42 °C overnight and incubated with the anti-DIG-alkaline phosphatase antibody. Telomere signals were detected by chemiluminescence after the addition of the substrate solution on membrane. Telomere length was measured by TeloTool software.

#### Measurement of telomere length by quantitative real-time PCR

Quantitative real-time PCR was used to measure relative telomere lengths (RTLs) as previously described<sup>82</sup>. Genomic DNA was extracted from cells using the DNeasy Blood & Tissue Kit (Qiagen). Average telomere length was measured from total genomic DNA using a real-time PCR assay that was modified to measure mouse telomeres. For each sample, 20 ng of genomic DNA was used in each reaction. PCR reactions were performed on an iCycler MyiQ2 Detection System (Bio-Rad) using telomeric primers and primers for the reference control gene (mouse 36B4 single-copy gene). For each PCR reaction, a standard curve was made by serial dilutions of known amounts of mouse genomic DNA. The telomere signal was normalized to the signal from the single-copy gene to generate a T/S ratio, which was indicative of relative telomere length. The primers used were listed in Supplementary Table S1.

#### Treatment of ESCs with LINE1 reverse transcriptase inhibitor

Zidovudine (AZT) is a nucleoside reverse transcriptase inhibitor. The drug was obtained from Sigma (A2169) and prepared on the day of use. ESC colonies were treated with 10  $\mu$ M AZT.

#### Apoptosis by TUNEL assay

Apoptosis was revealed by catalytically incorporating fluorescein-12-dUTP at 3'-OH DNA ends with the terminal deoxynucleotidyl transferase recombinant enzyme (rTdT) using a commercial kit (G3250, Promega).

Briefly, after being fixed in 4% paraformaldehyde, washed in PBS, and permeabilized in 0.2% Triton X-100, cells were incubated with rTdT buffer at 37 °C for 1 h, immersed in 2× SSC for 15 min to stop the reaction, stained with DAPI for 15 min, washed in PBS, mounted in Vectashield, and immediately analyzed under a fluorescence microscope using FITC/DAPI filters.

#### Annexin V immunostaining

Annexin V staining was performed using the Annexin V-FITC Apoptosis Detection Kit (C1062, Beyotime). Briefly, cells were cultured on cover slips. Cell culture medium was aspirated and washed once with PBS. Annexin V-FITC binding solution was added and mixed gently. Then propidium iodide (PI) staining solution was added and mixed gently. Slides were incubated at room temperature in the dark for 20 min. Samples were loaded in Vectashield medium (Vector) to prevent photo bleaching and imaged under Zeiss Axio Imager Z2 microscope.

#### Immunofluorescence microscopy of DNA damage

Immunofluorescence staining was performed as previously described<sup>32</sup>. Cells were washed in PBS, fixed in 3.7% paraformaldehyde, permeabilized with 0.1% Triton X-100 for 30 min at room temperature, blocked with blocking solution (3% goat serum and 0.1% BSA in PBS) for 2 h at room temperature, and incubated overnight at 4 °C with primary antibodies against  $\gamma$ H2AX (sc517348, Santa Cruz). After three washes with blocking solution, the cells were incubated with a secondary antibody 488 goat anti-mouse IgM (A21042, Invitrogen) for 2 h followed by washing three times with PBS. Nuclei were stained with DAPI (Sigma) and samples were loaded in Vectashield medium (Vector) to prevent photo bleaching. Fluorescence images were captured using a Zeiss fluorescence microscope (LSM 800 with Airyscan) and ZEN software. Positive control for DNA damage was set as treatment of WT ESC with H<sub>2</sub>O<sub>2</sub> for 1 h.

#### Cell proliferation assay

ESCs ( $1 \times 10^5$ ) were plated in 12-well plates and counted every 24 h. For count by passages,  $1 \times 10^5$  ESCs were plated in 12-well plates and passaged into new 12-well plates every 2 days.

#### RNA-seq analysis

Total RNA extracted from WT and *Terc*<sup>-/-</sup> cells and tissues was used for RNA-seq library preparation. Sequencing libraries were constructed by Novogene and sequenced following standard protocols. Paired-end mRNA-seq data consisted of read lengths of 150 bp for each sample, with two biological replicates for each line. The raw reads were first aligned to the mouse reference

genome (mm10) using the STAR software<sup>83</sup> with the default parameters. DEGs were calculated by the Cufflinks software<sup>84</sup> and the expression levels were represented by the RPKM (Reads Per Kilobase per Million mapped reads). DEGs were defined based on the fold change in expression levels and *P* value. If the absolute fold change was > 2 and *P* value < 0.05, genes were considered to be differentially expressed. For the comparison of the function of the DEGs in the seven cell types, we used the Metascape website (<https://metascape.org/gp/index.html#/main/step1>).

For analysis of retrotransposons, the raw reads were first aligned to the consensus sequence of TEs which were downloaded from the REPBASE using the STAR software<sup>83</sup>. Only repeat elements that exist in the mouse genome were used. The number of the retrotransposons was counted based on the BAM files. Differentially expressed retrotransposons were calculated using the software DESeq2<sup>85</sup>. The cutoff for the differentially expressed retrotransposons was fold change > 1.5 and *P* value < 0.05. The volcano plot of the DEGs and the MA plot of the differentially expressed retrotransposons were represented by the R package 'ggplot2'. The PCA (principal component analysis) of the retrotransposons were plotted by the R package 'gmodels'. The heatmap of the differentially expressed genes and retrotransposons was represented by the R package 'gplots'.

For calculating TE enrichment in the subtelomeric region, RNA-seq reads were first aligned to the mouse genome (mm10). A significant number of upregulated TE instances were identified by analyzing uniquely mapped reads of TEs between WT and *Terc*<sup>-/-</sup> samples using DESeq2. A genomic instance of TEs would be excluded if they were only covered by < 5 uniquely mapped reads. For each transposon superfamily, the upregulated instances on each chromosome were classified into subtelomere adjacent and subtelomere non-adjacent instances according to a distance threshold (1 Mb and 20 Mb). The expected value is calculated using the number of up-/downregulated retrotransposons located at telomere region by chance. The observed value was calculated using the number of up-/downregulated retrotransposons at 1 Mb/20 Mb. Fisher's exact tests were used to evaluate the significant enrichment of up-/downregulated instances associated with the telomeric region.

In the analysis of RNA-seq from colorectal cancer cells, the single-cell sequencing data were obtained from the public database under accession number SRP113436<sup>5</sup>. These cells were derived from the same tumor, and the analysis of telomere length and transcriptome was performed on the same single cell with previously described method<sup>5</sup>. We downloaded the sequencing data of 23 cells with the shortest telomeres and 50 cells with the longest telomeres. The raw data were mapped to the human consensus sequence from

the Refbase database with STAR software. Only repeat elements that exist in the human genome were used. The number of the retrotransposons was counted by the self-written script based on the BAM files. Differentially expressed retrotransposons were calculated using the software DESeq2. The cutoff for the differentially expressed retrotransposons was fold change > 1.5 and *P* value < 0.05. For the colorectal cancer cells, we used InferCNV software to estimate the CNVs.

### Exome sequencing and analysis

Sample genomic DNA was extracted, and the exome-seq library construction and sequencing were all carried out by Beijing Novogene Company as follows. Paired-end DNA libraries were prepared according to the manufacturer's instructions (Agilent) for two biological replicates. Genomic DNA from cell samples was sheared into 180–280 bp fragments by Covaris S220 sonicator. Ends of genomic DNA fragments were repaired and adenylated. Both ends of genomic DNA fragments were ligated with paired-end adaptors (Illumina) and a single ‘T’ base overhang, and purified using AMPure SPRI beads from Agencourt. The adaptor-modified genomic DNA fragments were enriched by six cycles of PCR using SureSelect Primer and SureSelect ILM Indexing PreCapture PCR Reverse Primer. The concentration and size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip. Whole-exome capture was carried out using SureSelect Mouse All Exon V1 Agilent 5190-4642. An amount of 0.5 µg prepared genomic DNA library was hybridized with a capture library for 5 min at 95 °C followed by 24 h at 65 °C. The captured DNA\_RNA hybrids were recovered using DynabeadsMyOne Streptavidin T1. Capture products were eluted from the beads and desalted using QIAGEN MinElute PCR purification columns. The purified capture products were then amplified using the SureSelect ILM Indexing Post Capture Forward PCR Primer and PCR Primer Index (Agilent) for 12 cycles. After DNA quality assessment, the captured DNA library was sequenced on the Illumina HiSeq2000 sequencing platform (Illumina) according to the manufacturer's instructions for paired-end 150 bp reads (Novogene). Libraries were loaded onto paired-end flow cells at concentrations of 14–15 pM to generate cluster densities of 800,000–900,000/mm<sup>2</sup> using Illumina cBot and HiSeq paired-end cluster kit.

Analysis of genomic instability of telomere-shortened cells was performed by normalization to corresponding WT cells, according to the protocols specified for mouse cells<sup>86</sup>. Briefly, BWA was used to map clean reads to the mouse mm10 reference genome to obtain the Bam file. Picard and SAMtools were used to sort and process the content of the file, and to generate a final Bam file for calculating sequencing quality (sequencing depth and

coverage). GATK4.1.9 was used to recalibrate the base quality score<sup>87</sup>. Mutect2 in GATK4.1.9, which is recommended in this protocol, was used for filtering SNVs and indels<sup>86</sup>. Refseq was used to further annotate the file and obtain the various locations and related information. CNVkit v0.9.7<sup>88</sup> was used to detect the sequencing information of CNVs and WT cells served as control.

### Third-generation genome sequencing

High molecular weight genomic DNA of each individual was extracted from population cells using HiPure Tissue & Blood DNA Kit (D3018-03, Angen). We performed Nanopore genome sequencing, which represents a robust technology in the DNA sequencing field, producing long-read sequencing data<sup>89</sup>. DNA repair, end repair, and adapter ligation were conducted during library preparation, and 2 µg DNA was fragmented by g-TUBE (Covaris). DNA repair was performed using NEBNext FFPE DNA Repair Mix (M6630L, NEB). End repair was performed using NEBNext Ultra II End Repair/dA-Tailing Module (E7546L, NEB). Adapter ligation was performed using the NEBNext Blunt/TA Ligase Master Mix (M0367L, NEB) and Ligation Sequencing Kit 1D (SQK-LSK109, Oxford Nanopore Technologies). DNA was purified at the end of each step using Agencourt AMPure XP beads (A63882, Beckman Coulter). DNA was quantified via a Qubit Fluorometer 2.0 (Thermo Fisher Scientific, Waltham, MA, USA). Long-read sequencing was carried out using a PromethION sequencer and 1D flow cell with protein pore R9.4.1 1D chemistry according to the manufacturer's instructions. Reads were base-called in batches by guppy v3.2.8 using the default parameters during sequencing.

Reads with an average 8 kb length were obtained and a total of 60 Gb data were analyzed for WT ESC and G4 *Terc*<sup>-/-</sup> ESC genomes. Reads from whole-genome sequencing were mapped with NGLMR according to published method<sup>90</sup>. Genomic structural variants were identified by Sniffles<sup>90</sup>. Structural variants were calculated after normalization to that of MEF for both WT and G4 *Terc*<sup>-/-</sup> ESCs.

### ATAC-seq analysis

The sequencing reads were trimmed with the Cutadapt software with the parameter “-a CTGTCTCTTATA”. Then the reads were mapped to the mouse genome mm10 with the Bowtie2 software<sup>91</sup>, and then the reads, which were mapped to the mitochondria were removed. The PCR duplications were also removed using the Picard tool (<https://broadinstitute.github.io/picard/>). The read counts were normalized with RPKM for the following analysis. The profiles of the ATAC signal were plotted with deeptools<sup>92</sup>. The ATAC-seq signals were visualized using the UCSC genome browser. The peaks were called by the software MACS2 (<https://pypi.org/project/MACS2/>) with

the parameter “-f BAMPE”, and then the peaks were annotated with the R package “ChIPSeeker”<sup>93</sup>.

### ChIP-seq

ChIP-seq was performed as previously described<sup>94</sup>. Briefly, cells were first fixed with 1% paraformaldehyde, and then lysed and sonicated to achieve the majority of DNA fragments at 100–1000 bp. DNA fragments were then enriched by immunoprecipitation with H3K9me3 (ab8898, Abcam), and H3K9me2 (ab1220, Abcam), and Dynabeads M280 (Life Technologies). The immunoprecipitated material was eluted from the beads by heating for 30 min at 68 °C. Then, samples were incubated with Proteinase K at 42 °C for 2 h followed by 67 °C for 6 h to reverse crosslinking. The samples were extracted with phenol:chloroform:isoamyl alcohol, followed by chloroform, ethanol precipitation, and resuspension in TE buffer. The sequencing libraries were prepared using reagents from the Novogene Corporation. Each library was subsequently sequenced, resulting in ~20 million reads (125 bp). The sequencing results were mapped to mouse genome assembly mm10. For reads mapped to repetitive sequences, uniquely mapped reads were used.

### ChIP-seq data analysis

For ChIP-seq data analysis, the raw ChIP-seq reads were aligned to mouse genome assembly (mm10) by BOWTIE2<sup>91</sup> with the parameter “--local -k 1”. This setting would ensure that if more than one equivalent best alignment was found, only one of those hits would be randomly reported. Therefore, instead of excluding reads from repetitive elements, multiple-hit reads are evenly distributed over highly similar repeat elements across the genome. The peaks of H3K9me3 were called using DFilter (v1.5)<sup>95</sup> with the “ks = 20 -lpval = 3 -nonzero” setting. DESeq2 was used to identify differential peaks. The density plot of ChIP-seq reads was plotted with NGSPLOT<sup>96</sup>.

For the generation of H3K9me2 and H3K9me3 enrichment heatmaps, H3K9me3 and H3K9me2 peaks were mapped to the repeat consensus sequence from Repbase, and the relative enrichment of the H3K9me3 signals as the number of reads of H3K9me3 or H3K9me2 ChIP-seq divided by the number of reads of ChIP input. The DESeq2 was used to determine the log<sub>2</sub> fold change of enrichment and measure the significance of enrichment. We mapped reads directly to TTAGGG repeats, the reads that match (TTAGGG)<sub>n</sub> with the parameter -D 15 -R 2 -N 0 -L 20 -i S,1,1.15 using bowtie2 software were defined as telomeric.

### In situ Hi-C library preparation

Cells were fixed in 2% formaldehyde for 10 min, and the reaction was quenched by the addition of glycine (0.125 M final concentration). After centrifugation

(300× g, 10 min), cells were washed with ice-cold PBS, precipitated, and the supernatant was removed. Cell pellets were incubated in 1 mL ice-cold lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630, 1× protease inhibitor (Roche) for 30 min on ice. Nuclei were collected by centrifugation at 2500× g for 5 min at 4 °C. Cell pellets were resuspended in 50 μL of 0.5% SDS, followed by incubation at 65 °C for 10 min. Then, 145 μL H<sub>2</sub>O and 25 μL of 10% Triton X-100 were added to quench the SDS and incubated at 37 °C for 15 min. To digest chromatin, 400 U of *DpnII* (NEB) were added, and samples were incubated at 37 °C overnight with continuous agitation (900 rpm). *DpnII* was then inactivated at 65 °C for 20 min. To fill in the biotin to the DNA, 1.5 μL 10 mM dCTP, 1.5 μL 10 mM dGTP, 1.5 μL 10 mM dTTP, 37.5 μL 0.4 mM biotin-14-dATP (Invitrogen) and 40 U Klenow (NEB) were added to the solution and the reaction was carried out at 37 °C for 1.5 h with rotation. Blunt-end ligation was performed by adding 900 μL of ligation mix (663 μL water, 120 μL of 10× NEB T4 DNA ligase buffer, 100 μL of 10% Triton X-100, 12 μL of 10 mg/mL BSA and 5 μL of 400 U/μL T4 DNA ligase), followed by overnight incubation at 16 °C. The reversal of crosslinks (65 °C, 4 h in the presence of Proteinase K), was followed by RNase A treatment and two sequential phenol chloroform extractions of DNA. DNA was sheared to 200–400 bp with Diagenode Bioruptor pico and double size selected on AMPure XP beads (Beckman Coulter). Biotin was removed from unligated restriction fragment ends by incubating DNA with T4 DNA polymerase (NEB) for 4 h at 20 °C in the presence of dATP and dGTP. The biotin-labeled DNA was then pulled down with 150 μL DynabeadsMyOne Streptavidin T1 (Life Technology) in binding buffer (5 mM Tris, pH 8.0, 0.5 mM EDTA, 1 M NaCl) for 30 min at room temperature. Sequencing library preparation was performed on beads, including end repair, dATP tailing and adaptor ligation, using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) according to the manufacturer’s instructions. 8–12 cycles of PCR amplification were performed with NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB). Finally, size selection was done with AMPure XP beads and fragments ranging from 200 to 600 bp were selected. All libraries were sequenced on Illumina HiSeq X Ten.

### Hi-C data analysis

Paired-end FASTQ files were controlled for quality, mapped to the reference genome (mm10) and converted into interaction matrices using Hi-C-Pro (V2.11.1)<sup>97</sup> using pipeline code available at <https://github.com/nservant/HiC-Pro>. We used 200 kb bin sizes to generate the raw and ICE (iterative correction and eigenvector decomposition) normalized matrix<sup>97</sup>. The valid pair files, were

transformed to Hi-C format files which were ready for juicer (V1.9.8) for exploratory visualization.

For detection of differential chromatin interactions in subtelomeric regions, statistical analysis was performed. The 100 kb at the end of each chromosome, which is located on the heterochromatin region, was excluded for its rarely detected reads. Then we counted the number of chromatin interaction reads mapped at the end of each chromosome (5 M, 10 M, 20 M, 30 M and 60 M, respectively), including chromatin interactions both within the region and with other regions. Left and right subtelomeric regions were calculated, respectively.

### Statistical analysis

Statistical analysis was performed using Excel. Data are presented as mean  $\pm$  SEM from biological triplicate data unless noted otherwise. A two-tailed paired Student's *t*-test was used to estimate the difference between the two groups. The data is expected to be normally distributed, and the variance is expected to be similar between the groups that are being statistically compared.

### Acknowledgements

This study was supported by funding from the National Key R&D Program of China (2018YFA0107000), the National Natural Science Foundation of China (32030033 to L.L., 32070858 and 31871488 to X.L.), and Tianjin Science and Technology Plan Key Project (20JCZDJC00550). We thank Lingyi Chen for helpful discussion. We acknowledge Annoroad Gene Technology and Tianjin Novogene Technology Co., Ltd. for their high-throughput sequencing services.

### Author details

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China. <sup>2</sup>Frontiers Science Center for Cell Responses, College of Life Sciences, Nankai University, Tianjin, China. <sup>3</sup>College of Pharmacy, Nankai University, Tianjin, China. <sup>4</sup>Genome Institute of Singapore, Singapore, Singapore. <sup>5</sup>2011 Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Medical Epigenetics, Department of Immunology, Biochemistry and Molecular Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. <sup>6</sup>Department of Genetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. <sup>7</sup>Haihe Laboratory of Cell Ecosystem, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China. <sup>8</sup>Institute of Translational Medicine, Tianjin Union Medical Center, Nankai University, Tianjin, China

### Author contributions

C.L., R.G. and J.M. performed most of the experiments; N.Z., G.Y., W.Z., Y.Z. and Y.S. performed bioinformatics analyses; C.L. harvested the RNA-seq, exome-seq and third-generation sequencing data; J.Y. helped telomere Q-FISH experiments; Z.Z.L., D.W. and Z.L. performed Hi-C experiments and data analysis; F.W. helped C-circle experiments; L.L. and X.L. planned the project, designed the experiments, and wrote the manuscript.

### Data availability

The whole exome-seq data for *Terc*<sup>-/-</sup> tissues and G3/G4 *Terc*<sup>-/-</sup> ESCs have been uploaded to the SRA with accession number PRJNA764419. The long read data have been uploaded to the SRA with accession number PRJNA764424. The RNA-seq data of the seven tissues/cell types have been uploaded to the GEO under accession number GSE184444. The ATAC-seq data have been uploaded to the GEO with accession number GSE184442. H3K9me3 ChIP-seq data were deposited in the GEO under GSE100168 and H3K9me2 ChIP-seq data in the SRA under PRJNA794026, and G3/G4 *Terc*<sup>-/-</sup> ESCs RNA-seq data were deposited in the GEO with accession numbers GSE118027 and GSE100168. The WT, G1, G4 *Terc*<sup>-/-</sup> RNA-seq data

accession number in the GEO is GSE192876. The Hi-C data accession number in the GEO is GSE210871.

### Conflict of interest

The authors declare no competing interests.

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41421-023-00538-y>.

Received: 21 October 2022 Accepted: 8 March 2023

Published online: 02 May 2023

### References

- Blackburn, E. H. Telomere states and cell fates. *Nature* **408**, 53–56 (2000).
- Mefford, H. C. & Trask, B. J. The complex structure and dynamic evolution of human subtelomeres. *Nat. Rev. Genet.* **3**, 91–102 (2002).
- O'Sullivan, R. J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11**, 171–181 (2010).
- Berrens, R. V. et al. Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat. Biotechnol.* **40**, 546–554 (2022).
- Wang, H. et al. Colorectal cancer stem cell states uncovered by simultaneous single-cell analysis of transcriptome and telomeres. *Adv. Sci.* **8**, 2004320 (2021).
- Barthel, F. P. et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.* **49**, 349–357 (2017).
- Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
- Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–133 (2012).
- Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
- Matsui, T. et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
- Rote, N. S., Chakrabarti, S. & Stetzer, B. P. The role of human endogenous retroviruses in trophoblast differentiation and placental development. *Placenta* **25**, 673–683 (2004).
- Wang, J. et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014).
- Lu, X. et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **21**, 423–425 (2014).
- Goke, J. et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 (2015).
- Fort, A. et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **46**, 558–566 (2014).
- Gao, L. et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* **173**, 248–259 (2018).
- Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Zalzman, M. et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
- Jang, H. S. et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51**, 611–617 (2019).
- De Cecco, M. et al. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247–256 (2013).
- Chiappinelli, K. B. et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
- De Cecco, M. et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73–78 (2019).
- Gorbunova, V. et al. The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43–53 (2021).

24. Espejel, S. & Blasco, M. A. Identification of telomere-dependent "senescence-like" arrest in mouse embryonic fibroblasts. *Exp. Cell Res.* **276**, 242–248 (2002).
25. Stewart, S. A. et al. Erosion of the telomeric single-strand overhang at replicative senescence. *Nat. Genet.* **33**, 492–496 (2003).
26. Shay, J. W. & Wright, W. E. Hallmarks of telomeres in ageing research. *J. Pathol.* **211**, 114–123 (2007).
27. Hemann, M. T., Strong, M. A., Hao, L. Y. & Greider, C. W. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67–77 (2001).
28. Gomes, N. M. et al. Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell* **10**, 761–768 (2011).
29. Blasco, M. A. et al. Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
30. Nagpal, N. et al. Small-molecule PAPD5 inhibitors restore telomerase activity in patient stem cells. *Cell Stem Cell* **26**, 896–909 (2020).
31. Agarwal, S. et al. Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients. *Nature* **464**, 292–296 (2010).
32. Huang, J. et al. Association of telomere length with authentic pluripotency of ES/iPS cells. *Cell Res.* **21**, 779–792 (2011).
33. Wang, Z. et al. Insertion of a chimeric retrotransposon sequence in mouse *Axin1* locus causes metastable kinky tail phenotype. *Mob. DNA* **10**, 17 (2019).
34. Richardson, S. R. et al. The influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol. Spectr.* **3**, MDNA3-0061-2014 (2015).
35. Mukherjee, A. K. et al. Telomere length-dependent transcription and epigenetic modifications in promoters remote from telomere ends. *PLoS Genet.* **14**, e1007782 (2018).
36. Vinayagamurthy, S., Bagri, S., Mergny, J. L. & Chowdhury, S. Telomeres expand sphere of influence: emerging molecular impact of telomeres in non-telomeric functions. *Trends Genet.* **39**, 59–73 (2023).
37. Ulrich, H. D. Ubiquitin and SUMO in DNA repair at a glance. *J. Cell Sci.* **125**, 249–254 (2012).
38. Fan, X. et al. SMOOTH-seq: single-cell genome sequencing of human cells on a third-generation sequencing platform. *Genome Biol.* **22**, 195 (2021).
39. Vessoni, A. T. et al. Telomere erosion in human pluripotent stem cells leads to ATR-mediated mitotic catastrophe. *J. Cell Biol.* **220**, e202011014 (2021).
40. Zhang, Y. et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388 (2019).
41. Asimi, V. et al. Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* **54**, 1238–1247 (2022).
42. Ma, X. et al. Pre-existing chromatin accessibility of switchable repressive compartment delineates cell plasticity. *Natl. Sci. Rev.* **9**, nwab230 (2022).
43. Downen, J. M. et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387 (2014).
44. Ruis, P. et al. TRF2-independent chromosome end protection during pluripotency. *Nature* **589**, 103–109 (2021).
45. Markiewicz-Potoczny, M. et al. TRF2-mediated telomere protection is dispensable in pluripotent stem cells. *Nature* **589**, 110–115 (2021).
46. d'Adda di Fagagna, F. et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
47. Herbig, U., Jobling, W. A., Chen, B. P., Chen, D. J. & Sedivy, J. M. Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol. Cell* **14**, 501–513 (2004).
48. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–334 (2008).
49. Mita, P. et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat. Struct. Mol. Biol.* **27**, 179–191 (2020).
50. Calado, R. T. & Dumitriu, B. Telomere dynamics in mice and humans. *Semin. Hematol.* **50**, 165–174 (2013).
51. Dai, Q. et al. Striking a balance: regulation of transposable elements by Zfp281 and Mll2 in mouse embryonic stem cells. *Nucleic Acids Res.* **45**, 12301–12310 (2017).
52. Allsopp, R. C. et al. Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci. USA* **89**, 10114–10118 (1992).
53. Herrera, E. et al. Disease states associated with telomerase deficiency appear earlier in mice with short telomeres. *EMBO J.* **18**, 2950–2960 (1999).
54. Shay, J. W. Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **6**, 584–593 (2016).
55. Stewart, S. A. & Weinberg, R. A. Telomeres: cancer to human aging. *Annu. Rev. Cell Dev. Biol.* **22**, 531–557 (2006).
56. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
57. Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175–186 (2017).
58. Davoli, T. & de Lange, T. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer Cell* **21**, 765–776 (2012).
59. Bodnar, A. G. et al. Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352 (1998).
60. Zhang, C. Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
61. Oricchio, E. et al. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene* **26**, 4226–4233 (2007).
62. Aschacher, T. et al. LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene* **35**, 94–104 (2016).
63. Wang, F. et al. Inhibition of LINE-1 retrotransposition represses telomere reprogramming during mouse 2-cell embryo development. *J. Assist. Reprod. Genet.* **38**, 3145–3153 (2021).
64. Aschacher, T. et al. Long interspersed element-1 ribonucleoprotein particles protect telomeric ends in alternative lengthening of telomeres dependent cells. *Neoplasia* **22**, 61–75 (2020).
65. Simon, M. et al. LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation. *Cell Metab.* **29**, 871–885.e5 (2019).
66. Thomas, C. A. et al. Modeling of TREG1-dependent autoimmune disease using human stem cells highlights L1 accumulation as a source of neuroinflammation. *Cell Stem Cell* **21**, 319–331.e8 (2017).
67. Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* **357**, 1383–1393 (2006).
68. Ardeljan, D. et al. Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat. Struct. Mol. Biol.* **27**, 168–178 (2020).
69. Dan, J. et al. Zscan4 contributes to telomere maintenance in telomerase-deficient late generation mouse ESCs and human ALT cancer cells. *Cells* **11**, 456 (2022).
70. Shalgi, R., Pilpel, Y. & Oren, M. Repression of transposable elements—a microRNA anti-cancer defense mechanism? *Trends Genet.* **26**, 253–259 (2010).
71. Moulier, T., Nielsen, L. P., Hansen, A. J. & Willerslev, E. Transposable elements in cancer as a by-product of stress-induced evolvability. *Front. Genet.* **5**, 156 (2014).
72. Burns, K. H. Transposable elements in cancer. *Nat. Rev. Cancer* **17**, 415–424 (2017).
73. Kassiotis, G. Endogenous retroviruses and the development of cancer. *J. Immunol.* **192**, 1343–1349 (2014).
74. Hancks, D. C. & Kazazian, H. H. Jr Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22**, 191–203 (2012).
75. Muers, M. Genomic variation: Human retrotransposons keep it active. *Nat. Rev. Genet.* **11**, 527 (2010).
76. Fu, H. et al. Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. *Genome Biol.* **22**, 201 (2021).
77. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
78. Liu, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232 (2018).
79. Robin, J. D. et al. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* **28**, 2464–2476 (2014).
80. Wakayama, T. & Yanagimachi, R. Cloning of male mice from adult tail-tip cells. *Nat. Genet.* **22**, 127–128 (1999).
81. Poon, S. S., Martens, U. M., Ward, R. K. & Lansdorp, P. M. Telomere length measurements using digital fluorescence microscopy. *Cytometry* **36**, 267–278 (1999).
82. Callicott, R. J. & Womack, J. E. Real-time PCR assay for measurement of mouse telomeres. *Comp. Med.* **56**, 17–22 (2006).
83. Dobin, A. & Gingeras, T. R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinform.* **51**, 11.14.1–11.14.19 (2015).
84. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).



85. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
86. Lange, S. et al. Analysis pipelines for cancer genome sequencing in mice. *Nat. Protoc.* **15**, 266–315 (2020).
87. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
88. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
89. Bowden, R. et al. Sequencing of human genomes with nanopore technology. *Nat. Commun.* **10**, 1869 (2019).
90. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
91. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
92. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
93. Yu, G., Wang, L. G. & He, Q. Y. CHIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
94. Yang, B. X. et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
95. Kumar, V. et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 615–622 (2013).
96. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
97. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).