



METHODOLOGY

Open Access



# Telomere-specific chromatin capture using a pyrrole–imidazole polyamide probe for the identification of proteins and non-coding RNAs

Satoru Ide<sup>1,2\*</sup> , Asuka Sasaki<sup>1,2†</sup>, Yusuke Kawamoto<sup>3</sup>, Toshikazu Bando<sup>3</sup>, Hiroshi Sugiyama<sup>3</sup> and Kazuhiro Maeshima<sup>1,2</sup> 

## Abstract

**Background:** Knowing chromatin components at a DNA regulatory element at any given time is essential for understanding how the element works during cellular proliferation, differentiation and development. A region-specific chromatin purification is an invaluable approach to dissecting the comprehensive chromatin composition at a particular region. Several methods (e.g., PICH, enChIP, CAPTURE and CLASP) have been developed for isolating and analyzing chromatin components. However, all of them have some shortcomings in identifying non-coding RNA associated with DNA regulatory elements.

**Results:** We have developed a new approach for affinity purification of specific chromatin segments employing an *N*-methyl pyrrole (P)-*N*-methylimidazole (I) (PI) polyamide probe, which binds to a specific sequence in double-stranded DNA via Watson–Crick base pairing as a minor groove binder. This new technique is called proteomics and RNA-omics of isolated chromatin segments (PI-PRiCh). Using PI-PRiCh to isolate mouse and human telomeric components, we found enrichments of shelterin proteins, the well-known telomerase RNA component (TERC) and telomeric repeat-containing RNA (TERRA). When PI-PRiCh was performed for alternative lengthening of telomere (ALT) cells with highly recombinogenic telomeres, in addition to the conventional telomeric chromatin, we obtained chromatin regions containing telomeric repeat insertions scattered in the genome and their associated RNAs.

**Conclusion:** PI-PRiCh reproducibly identified both the protein and RNA components of telomeric chromatin when targeting telomere repeats. PI polyamide is a promising alternative to simultaneously isolate associated proteins and RNAs of sequence-specific chromatin regions under native conditions, allowing better understanding of chromatin organization and functions within the cell.

**Keywords:** Chromatin, Chromatin purification, Pyrrole–imidazole (PI) polyamide, Non-coding RNA, Telomere, ALT (alternative lengthening of telomeres)

## Background

Genome activities, such as transcription and replication, are often achieved by DNA regulatory elements interacting with protein complexes and modified histone proteins to shape numerous and unique chromatin landscapes [1, 2]. Recently, non-coding RNAs

\*Correspondence: [satide@nig.ac.jp](mailto:satide@nig.ac.jp)

†Satoru Ide and Asuka Sasaki contributed equally to this study

<sup>1</sup> Genome Dynamics Laboratory, National Institute of Genetics, ROIS, Mishima, Shizuoka 411-8540, Japan

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



(ncRNAs), which are defined as RNAs that are not translated into functional proteins, have emerged as key regulators of chromatin states and play roles in diverse biological processes, including X-chromosome dosage compensation, developmental gene expression, and chromosome stability through telomere elongation [3–8]. While chromatin immunoprecipitation (ChIP) is used to map protein–DNA interactions, genome-wide analyses of ncRNA-binding sites, such as CHART, ChIRP and RAP, have been developed to map chromatin-associated RNAs [9–11]. These latter assays utilize antisense oligonucleotide probes to retrieve specific ncRNAs with bound DNA sequences. However, these approaches require prior knowledge of the target ncRNA to assay. Methods for comprehensive and unbiased identification of ncRNAs associated with specific DNA regulatory elements are still in need.

Knowing all the players acting at DNA regulatory elements is a major step toward understanding how the element works. Several approaches are available for isolating and analyzing chromatin at specific regions [12]. We focus on chromatin isolation techniques targeting telomeres consisting of a long array of repetitive sequences (TTAGGG) in mammals, because the chromatin composition of the telomere is well known. One technique is a region-specific chromatin purification called PICh [13, 14]. In PICh, a locked nucleic acid (LNA) probe specifically targeted telomeric chromatin via Watson–Crick base pairing and was used to identify telomere-bound proteins in combination with mass spectrometry [13]. The quantitative telomeric chromatin isolation protocol (QTIP) is an antibody-based method employed to detect proteins bound to telomeres, as it can compare telomere-bound proteins isolated from cells in a variety of states using isotope labeling by amino acids in cell culture (SILAC) [15]. Both techniques efficiently enriched telomeric chromatin but were limited to only analyzing the protein components of the telomere. Apart from the oligonucleotide and antibody probes, the recent development of engineered DNA-binding molecules provides alternative methods for purifying telomeric chromatin. A transcription activator-like effector (TALE)-based strategy or the CRISPR system, which contains nuclease-deficient Cas9 protein and sequence-specific guide RNA, have been used in combination with enChIP [16], CAPTURE [17] and CLASP [18] to identify both telomere-specific protein complexes and also ncRNAs. However, these techniques had less than ten-fold enrichment of telomeric proteins, such as the shelterin complex. Accordingly, the methods still seem primitive for accurate quantification and de novo discovery of chromatin-associated RNAs.

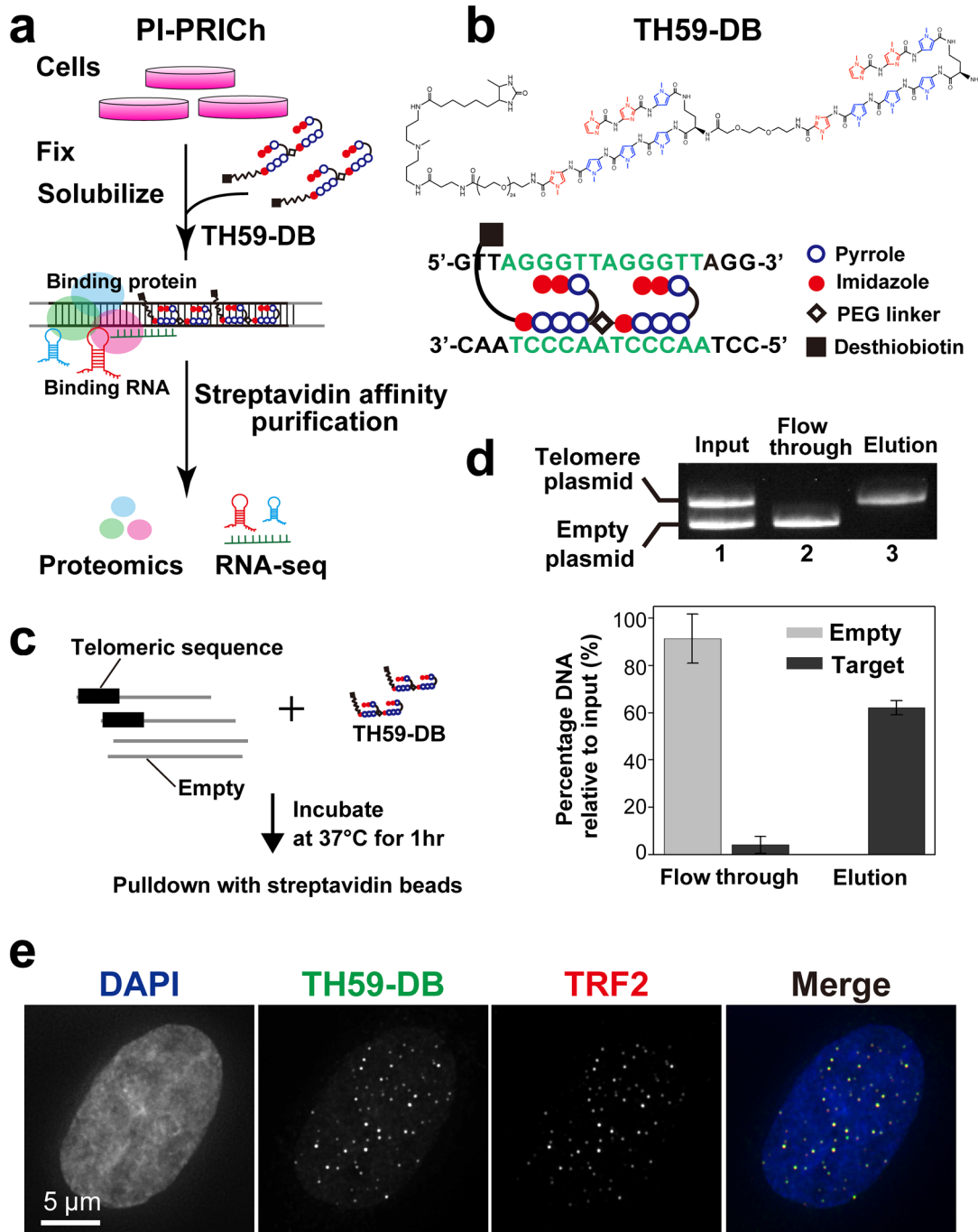
We have previously developed *N*-methyl pyrrole (P)-*N*-methylimidazole (I) (PI) polyamides for the visualization of telomeres and the assessment of telomere length [19–26]. PI polyamide binds to the minor groove of double-stranded DNA without denaturation and can recognize Watson–Crick base pairs [27–34]. We have demonstrated that fluorescently labeled tandem hairpin PI polyamide (TH-59) that targets human telomere sequences (TTAGGG)<sub>n</sub> can stain telomeres in cultured cells and tissue sections [19, 20, 35]. PI polyamide may be a more advantageous alternative to the nucleic acid probes used in any genomic analyses, including region-specific chromatin purification, because PI polyamide can target and bind a specific DNA sequence under mild conditions without non-specific binding to single-stranded RNA.

Here, we show a new region-specific chromatin purification method, using a PI polyamide probe, named Proteomics and RNA-omics of Isolated Chromatin segments (PI-PRICH) (Fig. 1a). We identified telomere-bound proteins from mouse erythrocyte leukemia (MEL) cells, such as the shelterin complex (TRF1, TRF2, POT1a, and TIN2), using PI-PRICH in combination with mass spectrometric analysis. At the same time, we also extracted the RNA fraction associated with telomeric chromatin and used next-generation sequencing (NGS) to comprehensively identify telomeric chromatin-associated RNAs. PI-PRICH mainly identified three types of ncRNAs involved in telomere maintenance: the telomerase RNA component (TERC) [36, 37], the telomeric repeat-containing RNA called TERRA (reviewed in [38, 39]) and ncRNAs transcribed from subtelomeric regions. Lastly, we discovered ncRNAs associated with the de novo telomeric sequence inserted into several intron regions when using PI-PRICH in alternative lengthening of telomere (ALT) cells, whose telomeres are highly recombinogenic. PI polyamides are thus a promising alternative for identifying proteins and ncRNAs to characterize chromatin at specific regions.

## Results

### PI polyamide probe for affinity purification of telomeric repeats DNA

We developed a method, PI-PRICH, which allows unbiased high-throughput identifications of telomeric chromatin-bound proteins and RNAs (Fig. 1a). For this purpose, we used a telomere-targeting tandem hairpin PI polyamide (TH59-DB), which has a biotin analog for affinity purification (Fig. 1b, Additional file 1: Fig. S1a and b). We first confirmed TH59-DB efficiently pulled down a plasmid containing a 750-bp telomeric repeats fragment, whereas the empty plasmid was not retrieved (Fig. 1c, d). Moreover, using streptavidin covalently attached to a



**Fig. 1** Affinity purification procedure of telomeric chromatin by a PI polyamide probe (PI-PRiCh). **a** Scheme for telomeric chromatin isolation using a telomere-targeting PI polyamide probe (TH59-DB). The crude chromatin fraction is mixed and incubated with TH59-DB, and the probe-chromatin complexes are isolated by streptavidin affinity purification. The isolated chromatin fractions are analyzed by mass spectrometry for the study of proteins and by next-generation sequencing for the study of ncRNAs. **b** Chemical structure of TH59-DB. The base recognition profile of TH59-DB is shown in the lower part. **c** Outline of the plasmid pull-down assay. Linearized plasmids (gray line) with or without the telomeric repeat (thick black line) were mixed with TH59-DB. The mixture was incubated at 37 °C for binding of TH59-DB. Plasmid-TH59-DB hybrids were captured using MyOne C1 streptavidin beads. **d** Purification of the telomeric repeat-containing plasmid with TH59-DB. (top) Each fraction (Input, Flow-through, and Elution) was analyzed by agarose gel electrophoresis and EtBr staining. The positions of the telomeric repeat-containing plasmid and the empty vector are indicated. (bottom) Bar graph quantifying telomeric DNA capture. Error bars represent standard deviations. **e** Telomere labeling with TH59-DB in HeLa1.3 cells. Cells were stained with DAPI (first column), TH59-DB (second column) and anti-TRF2 antibody (third column). The merged images are in the fourth column

fluorescent dye, TH59-DB targeted the telomere regions in the cells crosslinked with formaldehyde, which were co-immunostained by the TRF2 antibody (Fig. 1e). After verifying the targeting and binding specificity of TH59-DB, we isolated telomeric chromatin from crosslinked cells. Briefly, cultured cells were crosslinked, and their chromatin was extracted and homogenized. TH59-DB was bound to chromatin containing telomeric repeats. TH59-DB-bound chromatin was isolated using magnetic streptavidin beads. The co-purified proteins and RNAs were eluted and then subjected to downstream assays for identification and quantitation (Fig. 1a).

#### Affinity purification of telomeric chromatin by the PI polyamide probe

We isolated telomeric chromatin from MEL cells using TH59-DB. As a negative control probe, we premixed TH59-DB with telomeric oligo DNA prior to incubation with the chromatin extract (masked TH59-DB, Fig. 2a). The masked TH59-DB failed to capture telomeric DNA in the plasmid pull-down assay (Fig. 2b). TH59-DB and masked TH59-DB were incubated with solubilized MEL-derived chromatin and subsequently washed as described in [Methods](#). Pull-down fractions were electrophoresed and analyzed by silver staining for protein analysis. Several specific bands were detected in the TH59-DB pull-down fraction, but not in the masked TH59-DB pull-down fractions (Fig. 2c, d). The banding pattern resembled that of telomeric proteins purified by PICh, based on LNA (Fig. 2c, d). To verify the specific enrichment of telomeric proteins, we monitored the presence of the known telomere-associating protein TRF1 by immunoblotting. TRF1 was highly enriched in the TH59-DB pull-down fraction but not in the masked TH59-DB pull-down (negative control) or input fractions (Fig. 2e). Importantly, we identified known telomeric proteins

using mass spectrometry, such as shelterin components and chromosome passenger complexes (two independent results in Fig. 2f and Additional file 7: Table S1). Out of 32 proteins reproducibly identified by PI-PRiCh, 30 proteins were found in the published PICh telomere data for mouse embryonic stem cells [40] (Additional file 7: Table S1). These data demonstrated that TH59-DB specifically bound to the telomeric sequence and efficiently purified telomeric chromatin.

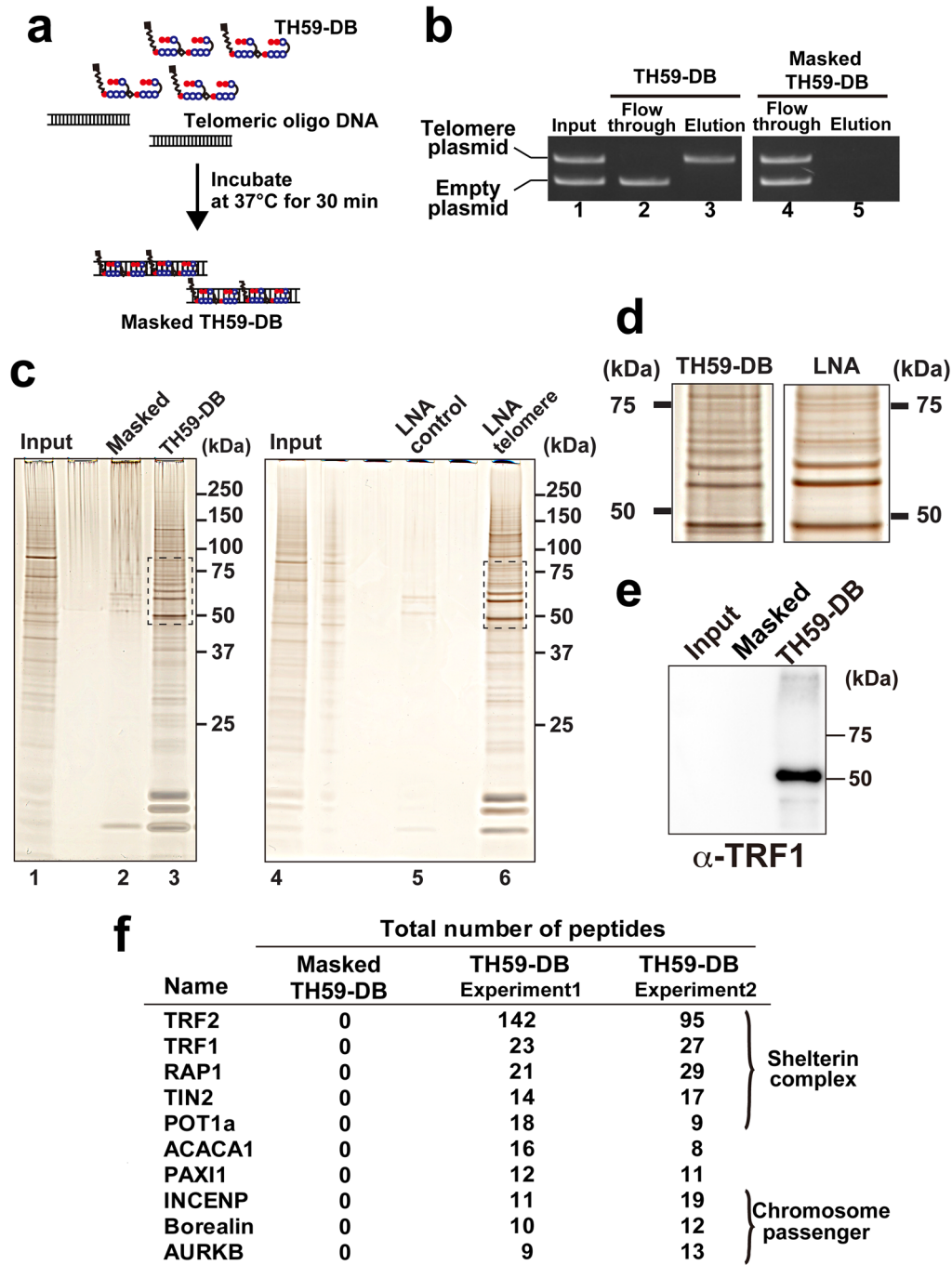
#### Comprehensive identification of telomeric chromatin-associated RNAs

RNA was extracted from the TH59-DB pull-down fraction containing telomeric chromatin and subjected to NGS analysis, herein called RNA-Seq. Sequencing and mapping of these libraries yielded approximately 5.5 million (total) and 2 million (mappable) read pairs in the TH59-DB pull-down fraction and 7.8 million (total) and 3.5 million (mappable) read pairs in input. In the TH59-DB pull-down fraction, the substantial part of RNAs in our dataset (742249 out of 5479729, 13.5%) had > 5 times-repeated telomeric sequences, which corresponded to the telomeric repeat-containing RNAs such as TERRA and ARRET (Fig. 3a). The reads containing (TTAGGG)<sub>5</sub> in the TH59-DB pull-down fraction were enriched 1000-fold over that of the input fraction (0.01%), suggesting PI-PRiCh has a much higher purity of telomeric materials than that previously reported [16]. About 66% of the telomeric repeat-containing RNAs are TERRA and the remaining are ARRET, which is consistent with the previous reports that TERRA is more abundant than ARRET [8, 41].

To identify RNA highly enriched in the telomeric chromatin, we plotted RNA levels in the pull-down fraction (after purification) against that of input (before purification). As shown in Fig. 3b, a cluster of RNAs was

(See figure on next page.)

**Fig. 2** Affinity purification of the telomeric chromatin from mouse erythrocyte leukemia (MEL) cells by PI-PRiCh. **a** Preparation of a negative control probe, masked TH59-DB, for telomeric chromatin isolation. TH59-DB was mixed and incubated with 50 times excess double-stranded oligonucleotide (TTAGGG)<sub>7</sub>/(CCCTAA)<sub>4</sub> (telomeric oligonucleotides). **b** Purification validation of the telomeric repeat DNA with TH59-DB and masked TH59-DB. Telomeric repeat-containing plasmids were purified with TH59-DB (lane 2, 3) but not masked TH59-DB (lane 4, 5). Each fraction (Input, Flow-through, Elution) was analyzed by agarose gel electrophoresis and EtBr staining. The positions of the telomeric repeat-containing plasmid and the empty vector are indicated. **c** Silver staining of proteins obtained from telomeric chromatin purification with TH59-DB or LNA probes for the telomere repeat. (left) Chromatin isolation was performed with masked TH59-DB and TH59-DB. Input representing 0.001% of the starting material (10<sup>4</sup> cells equivalent, lane 1), 8% of the masked TH59-DB pull-down fraction (lane 2) and of the TH59 pull-down fraction (lane 3). (right) Silver staining of proteins obtained from PICh with scrambled (LNA control, lane 5) or telomere LNA probes (LNA telomere, lane 6). **d** Enlarged images of the boxed region between 50 and 75 kDa in **c** to show their similar band patterns. **e** Western blot analysis for TRF1 in each fraction of PI-PRiCh. Input representing 0.0005% of the chromatin extracts, 4% of the materials of masked TH59-DB pull-down fraction or TH59-DB pull-down fraction. **f** List of proteins detected by mass spectrometry analysis of the material purified by TH59-DB from MEL cells. The results from two independent experiments are shown. The top ten proteins are sorted by the total number of peptides

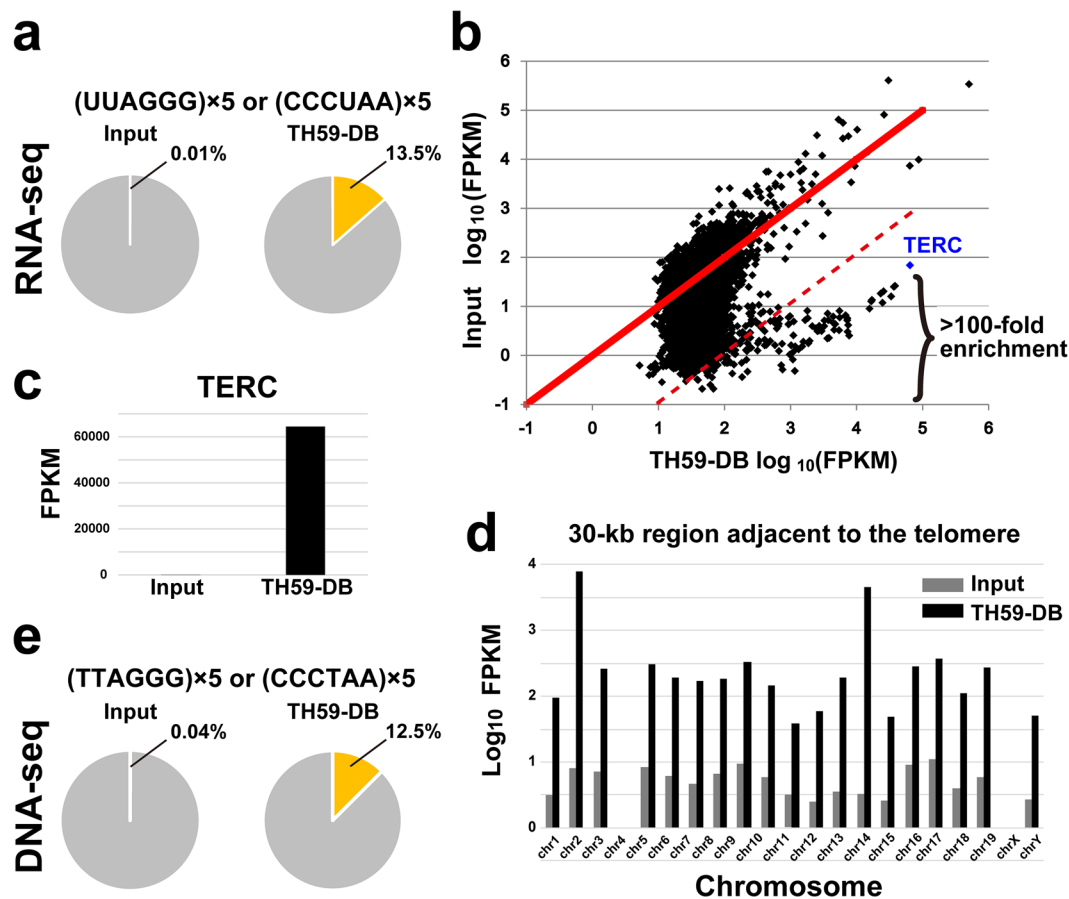


**Fig. 2** (See legend on previous page.)

enriched 100-fold over that of input. The most enriched RNA was the telomerase RNA component TERC (blue diamond in Fig. 3b, c).

Notably, almost all other RNAs also included (TTA GGG)<sub>n</sub> repeat sequences (Additional file 8: Table S2). Such highly enriched RNAs might be derived from

subtelomere regions, because telomeric repeats often exist in these regions [42]. Indeed, some sequence reads in the TH59-DB pull-down fraction were specifically mapped within the last 30-kb region adjacent to the telomere sequence of each chromosome (Fig. 3d and Additional file 1: Fig. S1c), which are considered to be



**Fig. 3** Comprehensive identification of telomeric chromatin-associated RNAs in MEL cells. **a** The percentage (yellow) of telomeric repeat reads, including TERRA transcripts, in the input and TH59-DB pull-down fractions from MEL cells. The number of the single-end reads, including (TTAGGG)<sub>5</sub> or (CCCTAA)<sub>5</sub>, extracted from input and TH59-DB pull-down fractions were divided by the total number of reads. The gray color shows the other reads. **b** Scatter plot of fragments per kilobase of per million mapped reads (FPKM) of the TH59-DB pull-down fraction versus that of the input sample for each RNA. The telomerase RNA component, TERC (most enriched in the TH59-DB pull-down fraction), is highlighted in blue. RNAs that enriched more than 100-fold in TH59-DB pull-down fraction are plotted below a red dotted line. **c** Bar graph of FPKM of telomerase RNA-component (TERC) in input and TH59-DB pull-down fractions. **d** Bar graph of FPKM value of the transcripts at a 30-kb region adjacent to the telomere on the q arm of each chromosome in input and TH59-DB pull-down fractions. **e** Sequence analysis of DNA fragments captured by TH59-DB. The percentage (yellow) of telomeric repeat reads including (TTAGGG)<sub>5</sub> or (CCCTAA)<sub>5</sub>, in the input and TH59-DB pull-down fractions from MEL cells. Note that the results of the second experiment (Experiment 2) are shown in Additional file 2: Figure S2

subtelomeric regions on the q arm of each chromosome [43]. A part of the sequence reads was mapped to a validated TERRA transcript from the subtelomeric regions [44] (Fig. 3d and Additional file 1: Fig. S1c).

To assess the reproducibility of the method, we repeated the telomeric chromatin purification from a different batch of MEL cells using TH59-DB, and obtained similar TH59-DB pull-down fraction results (>10% (39.1%) of RNAs had >5 times-repeated telomeric sequences, and TERC was enriched 100-fold over that of input) (Additional file 2: Fig. S2a, b, c). Moreover, we analyzed DNA fragments by DNA-Seq from two

independent TH59-DB pull-down fractions and estimated the enrichment of the target chromatin. 12.5% (Experiment 1) and 16.3% (Experiment 2) of the reads in each of the TH59-DB pull-down fractions had >5 times-repeated telomeric sequences, indicating about a 300–350-fold enrichment of telomeric TTAGGG repeats (Fig. 3e and Additional file 2: Fig. S2d).

Altogether, we concluded that the PI polyamide probe based telomeric chromatin purification was superior to RNA-omics for telomeric chromatin-associated ncRNA, which has never been achieved by PICh using LNA probes.

### ncRNAs associated with telomeric repeats in ALT

We then focused on ALT cells, which are telomerase-negative and have highly recombinogenic telomeres. ALT telomeres have a different chromatin state from telomerase-driven telomeres [45]. Telomeric RNA, particularly TERRA, is highly expressed and might be a key player to promote homologous recombination in ALT cells [46, 47]. Additionally, telomere elongation by ALT is associated with genomic alternation via the telomeric repeats insertion into non-telomeric genome regions [48–50]. Based on these findings, we expected to pull-down chromatin regions containing the telomeric repeat insertions distributed in the genome as well as in telomeric chromatin.

We applied PI-PRiCh to U2-OS cells (ALT telomeres) and to HeLa1.3 cells (long telomeres with active telomerase). PI-PRiCh data detected TERC in HeLa1.3 cells, but very little was found in U2-OS cells (Fig. 4a). We compared the amounts of TERRA in the pull-down fractions between the two cell lines and found little difference in their percentages of the reads, including  $>(TTAGGG)_5$  (Fig. 4b). Furthermore, the TERRA amount calculated from the mapped reads in U2-OS cells was comparable to that of HeLa1.3 cells (Fig. 4c, d) when we mapped the reads on a reference sequence of the subtelomeric region called TelBam3.4 [51, 52]. Others have suggested that TERRA can trigger recombination through the hybridization with telomeric DNA as TERRA appears to be expressed in ALT cells more than telomerase-positive cells [46, 47]. However, our results imply that not all TERRA RNAs bind to telomere repeats in ALT cells.

To search novel ncRNAs enriched in the pull-down chromatin from ALT cells, we used three filtering criteria: (1) the fragments per kilobase of per million mapped reads (FPKM) value, (2) the fold enrichment over total RNA of input sample before the telomeric purification, and (3) the relative ratio of ncRNAs between U2-OS cells and HeLa1.3 cells. Then, we identified some intronic ncRNAs (SPOCK3 and USP16 gene) with more than 100 FPKM enriched more than 25-fold, which were detected in U2-OS cells 50 times more than in HeLa1.3 cells (Fig. 5a and Additional file 9: Table S3 and Additional file 10: Table S4, Additional file 11: Table S5, Additional file 12: Table S6, Additional file 13: Table S7). On the other hand, TERC was the only ncRNAs enriched in HeLa1.3 cells based on these criteria.

### Identification of DNA fragments associated with telomeric repeats in ALT

As described above, telomere elongation by ALT involves homologous recombination between telomeres and also between non-telomeric regions and telomeres [48–50]. The latter can induce telomeric repeat insertions over

the genome [49, 50]. Therefore, we wondered whether the identified intronic RNAs in ALT cells might be generated by transcription in the intron regions with inserted telomeric repeats. We found that the intron regions in ALT cells have insertions of telomeric repeats using genomic polymerase chain reaction (PCR) with various primer sets: one primer annealing to the intron regions and the other to  $(CCCTAA)_4$  or  $(TTAGGG)_4$  (Additional file 3: Fig. S3). The products were validated by Sanger sequencing.

To further confirm insertions of these telomeric repeats in intron regions, we analyzed DNA fragments captured by PI-PRiCh. These intron regions were highly enriched in the TH59-DB pull-down fraction from U2-OS cells (Fig. 5b). We found chimeric sequence reads of the intron DNA sequence with  $(TTAGGG)_n$  repeats, indicating that telomeric repeats were inserted into the intron regions (Fig. 5c).

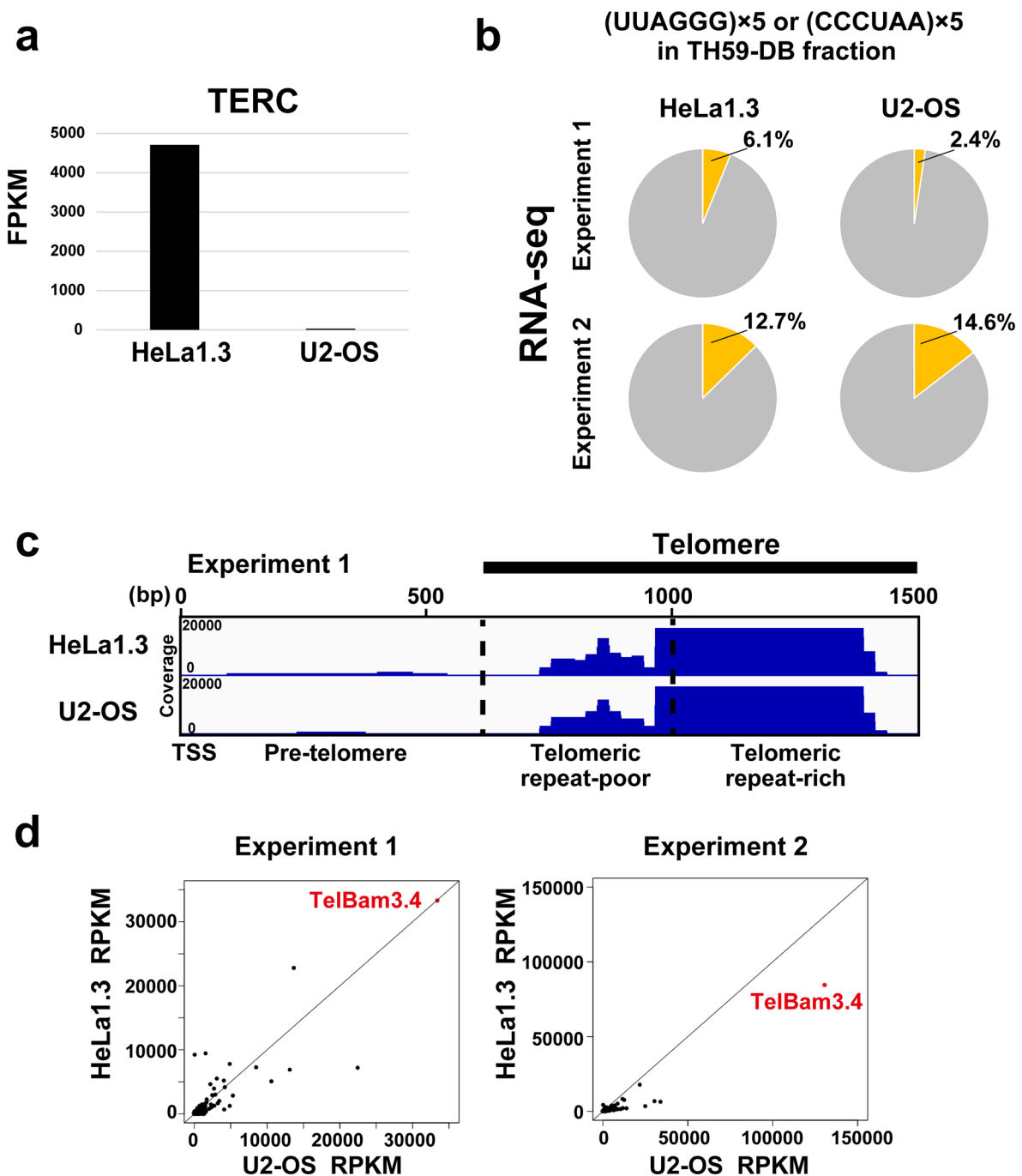
In addition to ncRNAs from SPOCK3 and USP16 genes, we identified four other intronic ncRNAs corresponding to CKS1B, NRDC, PAM, and KIAA1671 genes (Additional file 4: Fig. S4 and Additional file 9: Table S3). Using similar analyses, we found that they also have flanking telomeric repeats (Additional file 3: Fig. S3, Additional file 5: Figure S5).

### Co-localization of ncRNAs at telomeres

To better understand the biological implications of the intronic ncRNAs identified, we examined localizations of the four intronic ncRNAs (USP16, SPOCK3, CKS1B, NRDC genes) in the U2-OS cell line by RNA-FISH (fluorescence in situ hybridization) analysis. Telomeres were simultaneously stained using the fluorescent TH59 [19, 20, 25]. As shown in Additional file 6: Fig. S6a and b, clear signals of USP16-intronic RNA were detected. Interestingly, some of them were co-localized with the telomere signals: 23% of USP16-positive cells, 50% of SPOCK3, 25% of CKS1B and 17% of NRDC-positive cells (Additional file 6: Fig. S6c). These results indicate that the intronic RNAs form foci and can localize near telomeres in the cell nuclei.

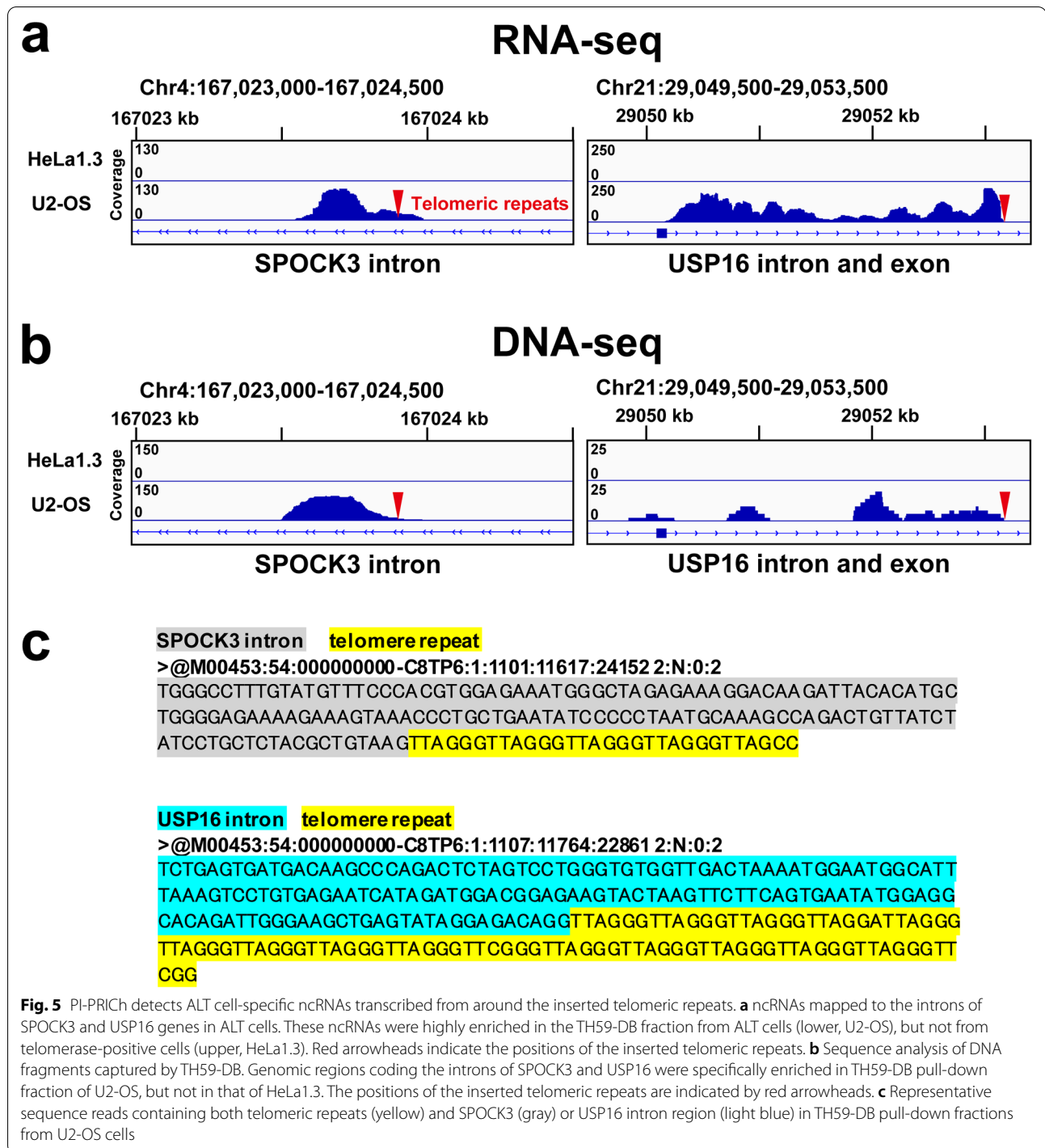
### Discussion

We demonstrated that PI-PRiCh was reproducibly able to purify both proteins and RNAs associating with telomeric chromatin, whereas previous methods could only identify protein composition [13, 15]. Using PI-PRiCh, we succeeded in identifying the known three classes of ncRNAs associated with telomeric chromatin with  $>100$ -fold enrichment from MEL cells. The first class is a trans-acting ncRNA bound as a part of a ribonucleoprotein telomerase complex (i.e., TERC). The second class is a cis-acting ncRNA that functions as a major



**Fig. 4** Comparison of the telomeric chromatin-associated TERRA between HeLa1.3 and ALT U2-OS cell lines. **a** Bar graph of FPKM of the telomerase RNA component (TERC) in telomerase-maintained telomeres (HeLa1.3) and ALT telomeres (U2-OS). **b** Pie charts show the percentage of telomere repeat sequence reads including TERRA/ARRET in the input and TH59-DB pull-down fractions from HeLa1.3 and U2-OS. Upper part and lower part show experiment 1 and experiment 2, respectively. **c** A browser view of chromatin-associated TERRA transcripts at a subtelomeric region downstream transcription start site (TSS) coding TelBam3.4 sequence using Integrative Genomic Viewer (IGV) in HeLa1.3 or U2-OS. The numbers on the left side in tracks indicate the read counts in RNA-seq. Telomere region in TelBam3.4 that was defined previously (Nergadze 2009, Usui 2015). Borderlines (vertical dotted black lines) are shown between pre-telomere, telomeric repeat-poor, and -rich regions based on TTAGGG repeat density. **d** Scatter plots of read per kilobase with normalized count per million (RPKM) in U2-OS cells versus that of HeLa1.3 cells for each RNA from two independent PI-PRiCh experiments. RNA that is transcribed from the subtelomeric region of TelBam3.4 is highlighted in red





modulator of telomere maintenance (i.e., TERRA). Lastly, we identified ncRNAs associated as nascent transcripts from subtelomeric regions and interstitial telomeric sequences (third class). The comprehensive identification of ncRNAs with high enrichment relied on the unique binding mode of PI polyamide to

target sequences of double-stranded DNA through minor grooves (Additional file 1: Fig. S1b) so that PI polyamide has almost no affinity to single-stranded and double-stranded RNA [53]. This property reduces contamination of abundant messenger RNAs and ribosomal RNAs during telomeric chromatin isolation by

PI-PRiCh, whereas chromatin isolation using other nucleic acid probes can potentially have abundant contaminant RNAs, because the nucleic acid probes often nonspecifically hybridize RNAs by mismatch pairing.

PI-PRiCh is distinct from recently published methods for studying a genome-wide RNA–DNA contact, including MARGI, GRID-seq, ChAR-seq and RADICL-seq [54–57]. Chromatin-associated RNAs are immobilized to DNA by proximity ligation after restriction digestion, forming RNA–DNA chimeric sequences for sequencing with these techniques. However, obtained global RNA–chromatin interactomes must exclude the repetitive sequence regions due to a lack of restriction recognition sites. PI-PRiCh is thus an excellent approach for the comprehensive identification of chromatin-associated ncRNAs found on repetitive sequences, which occupy approximately 45% of the human genome [58]. PI-PRiCh can also be applied to other repetitive sequences, such as found in centromeres, to identify new chromatin-associated RNAs.

We also identified novel ncRNAs derived from the introns of gene loci in ALT cells. The intron regions (SPOCK3, USP16, CKS1B, NRDC, KIAA1671 and PAM) commonly included short telomeric repeats. DNA-seq showed that intronic DNA fragments were also enriched in the pull-down fraction from U2-OS cells. TH59-DB bound to inserted telomeric repeats in the PI-PRiCh experiment and pulled down those ncRNAs nascently transcribed by RNA polymerase II. These data suggest that PI-PRiCh can isolate DNA regulatory sequences with their associated RNAs from not only abundant repetitive sequence regions, but also single loci in the future.

Lastly, it would be intriguing to discuss possible functions of identified novel ncRNAs from the introns of several genes in the ALT cells. These intron regions (SPOCK3, USP16, CKS1B, NRDC, PAM, and KIAA1671) have short telomeric repeats. Interestingly, some of them seem to be associated with telomeres (Additional file 6: Fig. S6). This finding suggests that certain genome regions, including inserted telomeric repeats, are somehow tethered to telomeres in ALT cells. Such clustering might be facilitated by ncRNA and/or other proteins such as orphan nuclear receptors [49] (Additional file 6: Fig. S6d) and involve their high recombinogenic property in ALT cells.

With the evolution of new technologies, including cutting-edge imaging, advanced genomics and computational modeling, our knowledge of chromatin organization and dynamics is drastically increasing [1, 59, 60]. Our sensitive telomere labeling and capturing method can be performed under mild conditions, which opens the gate for other intriguing applications

like super-resolution imaging of telomeres and the structural analysis of telomeric chromatin by cryo-electron microscopy without harsh treatments. Combining these techniques could help to elucidate how telomeric chromatin is organized in interphase cell nuclei and mitotic chromosomes. Therefore, telomere visualization and isolation using PI polyamide-based approaches discussed here would expand telomere biology and related medical science.

## Conclusion

PI-PRiCh can simultaneously identify both the protein and RNA components of telomeric chromatin when directed against telomeric sequences. PI polyamide is thus a promising alternative for sequence-specific isolation of native chromatin regions with their associated proteins and RNAs, which will promote an increased understanding of chromatin organization and functions within the cell.

## Methods

### TH59-DB probe synthesis

Telomere PI polyamide TH59 containing a long spacer arm (24 PEG) (Additional file 1: Fig. S1a) was synthesized using the solid-phase synthetic method as described in the references [19, 20, 22]. ESI–TOF–MS  $m/z$  calculated for  $C_{153}H_{232}N_{42}O_{45}^{4+}$   $[M+4H]^{4+}$  844.4284 found 844.4276. For chromatin affinity purification, tandem hairpin TH59 was conjugated with desthiobiotin using EZ-Link™ NHS-desthiobiotin (16129, Thermo) to obtain the PI polyamide probe TH59-DB (Fig. 1b and Additional file 1: Fig. S1a) as described in the reference [23]. ESI–TOF–MS  $m/z$  calculated for  $C_{163}H_{249}N_{44}O_{47}^{5+}$   $[M+5H]^{5+}$  714.9684 found 714.9686. TH59-DB was resuspended in *N,N*-dimethylformamide (DMF) (043-32361, Wako) and kept in the freezer until use.

### Telomere staining with TH59-DB

HeLa1.3 cells were maintained at 37 °C (5% CO<sub>2</sub>) in DMEM (D5796, Sigma) containing 10% FBS. For polyamide staining, cells were grown on coverslips coated with poly-lysine. The cell coverslips were washed twice in phosphate-buffered saline (PBS) and fixed with 1.85% formaldehyde in PBS for 15 min at room temperature. For blocking, cells were treated with 10% normal goat serum (NGS) (S26-100, Millipore) in TE buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA) for 30 min at room temperature. After a brief rinse with TE buffer, the cells were incubated with 10% NGS, 100 nM TH59-DB in DMF, mouse anti-TRF2 (ab13579, Abcam), and 0.5 µg/mL DAPI (4',6-diamidino-2-phenylindole) (10236276001, Roche) in TE buffer at 37 °C for 1 h. After washing with TEN200 buffer (10 mM Tris–HCl, pH7.5, 1 mM EDTA,

and 200 mM NaCl), cells were incubated with 10% NGS, streptavidin Alexa488 (s11223, Invitrogen) and anti-mouse Alexa594 (AA11032, Invitrogen) in TE for 1 h at room temperature. The cells on the coverslips were washed with TEN200 buffer (five times for 3 min), then mounted. Cell images were recorded with a DeltaVision microscope and deconvolved to eliminate out-of-focus blur to obtain clearer pictures. The deconvolved images were projected ('Quick Projection' tool) to obtain the maximum intensity of telomere signals.

#### Plasmid pull-down assay

This assay was performed as described [14]. For each assay, 200 ng linearized plasmids (telomeric repeats containing plasmid and empty vector) were resuspended in LB3JD buffer (10 mM HEPES–NaOH, pH 7.7, 100 mM NaCl, 2 mM EDTA, 1 mM EGTA, 0.2% SDS, 0.1% SLS) containing 0.5  $\mu$ M TH59-DB. The mixture was incubated at 37 °C for 30 min. During incubation, 15  $\mu$ l of MyOne C1 beads (DB6502, VERITAS) was added. After incubation at room temperature for 30 min with shaking, beads were immobilized on a magnetic stand. The supernatant was collected as flow-through. Beads were washed five times with 1 mL of LB3JD at room temperature, resuspended in LB3JD containing 12.5 mM D-biotin (B-20656, Invitrogen) and incubated at 65 °C for 15 min for elution. One-tenth volumes of the input, flow-through, and elution fraction were analyzed by agarose gel electrophoresis, followed by ethidium bromide (EtBr) staining. Band intensities were quantified relative to input by Fiji software in three independent experiments [61]. As a negative control for telomeric chromatin isolation, the masked TH59-DB was prepared as follows: TH59-DB was incubated with a double-stranded DNA oligonucleotide (TTAGGG/CCCTAA)<sub>4</sub> for 60 min at 37 °C.

#### Chromatin purification from mouse (MEL) cells by TH59-DB

MEL cells were grown in DMEM with 10% FBS and 2 mM L-glutamine. The telomere length of MEL cells is not available but could be between >23 and <100 kb based on telomere lengths described in mouse embryonic fibroblasts [62]. 2 L of MEL cells (~2–4  $\times$  10<sup>6</sup> cells/ mL) grown in a roller bottle were crosslinked with 3.7% formaldehyde in PBS for 30 min at room temperature. After washing four times with PBS, cells were transferred in sucrose buffer (0.3 M sucrose, 10 mM HEPES–NaOH, pH 7.7, 1% Triton X-100, 2 mM MgOAc) and lysed with 20 strokes of a Dounce homogenizer with a tight pestle. Chromatin was pelleted by centrifugation at 3200g for 10 min at 4 °C. The pellet was resuspended in the same volume of glycerol buffer (25% glycerol, 10 mM HEPES–NaOH, pH 7.9, 0.1 mM

EDTA, 0.1 mM EGTA, 5 mM MgOAc), and then frozen in liquid nitrogen and stored at –80 °C or used immediately for telomeric-chromatin-isolation. The chromatin pellet was washed with PBS five times and with LB3JD containing 1 mM phenylmethylsulphonyl fluoride (LB3JD-PMSF) (P7626, Sigma). After centrifugation at 3000g for 8 min, the pellet was resuspended in a 1.5  $\times$  volume of LB3JD-PMSF buffer and then passed three times through a French press (FA-078A, Thermo) at 25,000 p.s.i. at room temperature. The solubilized chromatin sample was collected at 20,000g for 15 min at 4 °C and was heated at 58 °C for 5 min. LB3JD-PMSF-pre-equilibrated streptavidin agarose beads (20361, Thermo) were added, and the sample was incubated at 4 °C overnight. The mixture was applied to Sephacryl S-400-HR spin columns (17-0609-10, Roche). The sample was centrifuged at 20,000g for 15 min and SDS was added to the supernatant to a final concentration of 0.2%. About 30 mg chromatin was incubated with 600 pmol of TH59-DB or masked TH59-DB for 2 h at 37 °C. The sample was then centrifuged at 20,000g for 15 min and then the supernatant was added to streptavidin-FG beads (TAS8848 N1170, Tamagawa-Seiki) pre-equilibrated with LB3JD. The sample was incubated at room temperature for 2 h on a nutator. Beads were washed with 10 mL of LB3JD five times at room temperature. Beads were collected in a 1.5 mL tube and additionally washed with shaking in LB3JD containing 30 mM NaCl at 42 °C for 10 min and then in LB3JD containing 10 mM NaCl at 42 °C for 10 min. Beads were resuspended in LB3JD containing 12.5 mM D-biotin and incubated at 37 °C for 1 h for elution with shaking. Cleared supernatant was collected as eluate and kept at –80 °C until further analysis. PICH protocol was performed as previously described with the LNA probe hybridizing telomere sequence [13]. Briefly, the chromatin sample preparation from MEL cells was done as described for PI-PRCh. About 30 mg chromatin was hybridized to 1500 pmol of LNA probe for the telomeric repeat and the scramble LNA probe: desthiobiotin-PEG24- 5'-TtAgGgTtAgGgTtAgGgTtAgGgt-3' and desthiobiotin-PEG24- 5'-GaTgTgTgGaTgTggAtGtGgAtgTgg-3', respectively, where capitalized letters are LNA residues and small letters are DNA residues. Hybridization was performed by sequential incubations at 25 °C for 3 min, 72 °C for 7 min, and 37 °C for 3 h. The sample was then centrifuged at 20,000g for 15 min, and the supernatant was added to MyOne C1 streptavidin beads (DB65002, Thermo) pre-equilibrated with LB3JD. The subsequent procedures, such as wash and elution steps, were performed as described for chromatin purification by TH59-DB.

### Chromatin sample preparation from adherent cells for purification with TH59-DB

HeLa1.3 and U2-OS cells ( $\sim 10^9$  cells) were grown in DMEM with 10% FBS. Telomere lengths of U2-OS cells and HeLa1.3 cells were reported to be  $>30$  kb and  $>20$  kb, respectively [63, 64]. The cells were washed with PBS and crosslinked with 3.7% formaldehyde in PBS for 30 min at room temperature. After being washed twice with cold PBS, cells were scraped with a scraping buffer (0.05% Tween-20 in PBS). The following procedures are the same as chromatin preparation of MEL cells described above.

### Protein extraction for mass spectrometry

For protein analysis of the eluate, trichloroacetic acid (18% final) precipitation was performed, and the pellet was incubated with 80 mL crosslinking reversal solution (250 mM Tris-HCl (pH 8.8), 2% SDS, 1 M 2-mercaptoethanol) at 99°C for 30 min. Proteins from the eluate from PI-PRiCh with masked TH59-DB or TH59-DB probes were separated using a 12% Bis-Tris acrylamide pre-cast gel (NP0343BOX, Invitrogen). The gels were stained with colloidal blue for MS analysis or silver-stained (Silver Quest kit) (LC6070, Invitrogen) for detecting unique bands purified by PI-PRiCh using the TH59-DB probe or subjected to western blot analyses with an anti-TRF1 antibody (kind gift from Y. Shinkai). For comprehensive protein identifications, gel lanes were cut into regions according to the banding pattern and subjected to MS analysis (NIG Mass Spectrometry Facility).

### RNA extraction and preparation of cDNA libraries for next-generation sequencing

To extract total RNA,  $>2.5$  mg isolated chromatin was incubated in de-crosslinking buffer (32 mM Tris-HCl pH 8.0, 320  $\mu$ g/mL Protease K, 0.8% SDS) for 6 h at 65 °C. An equal volume of acidic phenol was added, then vortexed, and incubated for 5 min at room temperature. After adding an equal volume of chloroform: isoamyl alcohol (24:1, v/v) and centrifuging at 16,000g for 5 min at 4 °C, the upper aqueous phase was transferred to a new tube. This step was repeated. One-tenth volume of 3 M sodium acetate, glycogen and double volume of ethanol were added and incubated overnight at  $-80$  °C. After centrifugation at 20,000g for 30 min at 4 °C, the pellet was retained, washed with 70% ethanol and dried up. The dried pellet was dissolved in diethylpyrocarbonate (DEPC)-treated water (36415-54, Nakarai) and kept at  $-80$  °C until use. Total RNA content of each sample was measured using Qubit (Q32851, Thermo), and the quality of RNA samples was assessed by an Agilent 2100 Bioanalyzer using Agilent RNA 6000 pico kit (5067-1513, Agilent). cDNA libraries were synthesized by the SMARTer

Stranded Total RNA-Seq Kit v2 (634411, Takara). The size distributions of the libraries were checked by an Agilent 2100 Bioanalyzer using an Agilent High Sensitivity DNA kit (5067-4626, Agilent). Pooled amplicon library was sequenced with paired-end  $2 \times 150$  bp reads on the Illumina MiSeq platform.

### Detection of TERRA/ARRET in the RNA-Seq data and DNA-Seq data

To estimate the number of putative TERRAs, reads containing (TTAGGG)<sub>5</sub> or (CCCTAA)<sub>5</sub> repeats were extracted from each Read 2 (R2) fastq file using the grep command in the UNIX system.

### Data processing and software for RNA-seq

First, adapter sequences were removed. The reads were trimmed for low quality with CUTADAPT and PRINSEQ using a composite set of Illumina adapters, a minimum quality score of 20 and a minimum length of 25 [65, 66]. Filtered mouse and human sequence data were aligned to the mouse mm10 genome and human hg38 genome, respectively, using HISAT2 [67]. To find telomere-enriched ncRNAs, the mapped fragments in TH59-DB pulled-down fractions were assembled into RNA transcripts and annotated using Cufflinks software [68]. The amounts of RNA transcripts were compared based on fragments per kilobase of per million mapped reads (FPKM) between samples using featureCounts and R [69, 70]. RNA-seq data were also visualized by igvtools [71].

### Mapping of TERRA in MEL cells

For mapping of mouse TERRA, we focused on a 30-kb region adjacent to the telomere repeat of each long (q) arm of the chromosome [43]. Centromere-adjacent telomeres on short (p) arms were not analyzed, because they were not sequenced. Using featureCounts, FPKM values of the subtelomeric regions (Additional file 14: Table S8) were compared between input and TH59-DB pull-down fractions.

### Mapping of TERRA in HeLa1.3 and U2-OS

For mapping human TERRA on the X/Y chromosome, custom gene annotation containing the sequence of TelBam3.4 [51, 52] was generated for quantification of telomere-associated TERRA. Individual reads of RNA-Seq data were mapped to the custom gene and human genome hg38 by Bowtie2 with default settings except for extraction of uniquely mapped reads, respectively [72]. Mapped reads were assembled into transcripts and annotated by Cufflinks. The number of reads overlapped to a region downstream of the transcription start site in TelBam3.4 were counted by featureCounts and normalized

as reads per kilobase of exon model per million mapped reads (RPKM) on TelBam3.4 and human genome hg38.

#### **DNA extraction and preparation of cDNA libraries for next-generation sequencing**

To extract DNA, the pull-down chromatin fractions were incubated in de-crosslinking buffer (50 mM Tris-HCl pH 8.0, 200 µg/mL Protease K, 2% SDS) overnight at 65 °C. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1,v/v) (311-90151, Wako) was added, vortexed and then centrifuge at 16,000g for 5 min at room temperature. The upper aqueous phase was transferred to a new tube. This step was repeated. One-tenth volume of 3 M sodium acetate, glycogen, and the double volume of ethanol were added and incubated for 2 h at -80 °C. After centrifugation at 20,000g for 30 min at 4 °C, the pellet was retained, washed with 70% ethanol and dried. The dried pellet was dissolved in 10 µg/mL RNase A in DNA dilution buffer attached to DNA SMART ChIP-Seq Kit (634865, Takara). After incubating for 1 h at 37 °C, DNA was purified by phenol-chloroform extraction and ethanol precipitation to remove RNase A. DNA samples were kept at -30 °C until use. DNA concentrations of each sample were measured by Qubit. cDNA libraries were synthesized by DNA SMART ChIP-Seq Kit (634865, Takara). The size distributions of the libraries were checked by an Agilent 2100 Bioanalyzer using Agilent High Sensitivity DNA kit. Pooled amplicon library was sequenced on the Illumina MiSeq platform with paired-end 2 × 250 bp reads for HeLa1.3 and U2-OS cells, and with single-end 75 bp reads for MEL cells.

#### **Detection of ncRNA-associated telomeric repeats in the DNA-seq**

To confirm telomeric repeats were inserted near the genomic region, where ncRNA was transcribed, sequence reads in the DNA-seq containing ncRNA sequences were extracted from each fastq file using the grep command in the UNIX system. The sequences for the grep command are shown in Additional file 15: Table S9.

#### **Data processing and software for DNA-seq**

First, adapter sequences were removed. The reads were trimmed for low quality with CUTADAPT and PRINSEQ, using a composite set of Illumina adapters, a minimum quality score of 20, and a minimum length of 25. Filtered human sequence data were aligned to the human hg38 genome using Bowtie2. DNA-seq data were also visualized by igvtools.

#### **Genomic PCR to test telomeric sequence insertion**

To purify genomic DNA, cells grown on φ10 cm dishes were collected and incubated in lysis buffer (10 mM

Tris-HCl pH 8.0, 0.1 M EDTA, 0.5% SDS, 20 µg/mL RNase A) for 1 h at 37 °C. After adding 100 µg/mL Proteinase K (10432, Wako), the cell lysate was incubated for 3 h at 50 °C. DNA was isolated using phenol-chloroform extraction and ethanol precipitation. To test if the telomeric repeat sequence located to the flanking region of the introns, PCR reactions with primer sets shown in Additional file 15: Table S9 were performed on genomic DNA from HeLa1.3 and U2-OS cells using the KOD-FX (KFX-101, Toyobo). PCR products were analyzed by agarose electrophoresis, followed by ethidium bromide staining.

#### **RNA fluorescence in situ hybridization**

Telomere staining with RNA-FISH was performed following an RNA-FISH method combined with immunofluorescence as described in [73]. Briefly, cells were seeded on poly-lysine coated coverslips and incubated overnight at 37 °C in 5% CO<sub>2</sub>. Coverslips were washed with PBS and fixed with 3.7% formaldehyde in PBS at room temperature for 10 min. After 5 min washing with PBS, cells were permeabilized with 0.1% Triton X-100 in PBS for 10 min and then washed with PBS for 5 min. The coverslips were incubated with prehybridization solution [2 × SSC, 1 × Denhards solution, 50% (v/v) formamide, 10 mM EDTA pH 8.0, 100 µg/mL yeast tRNA and 0.01% Tween-20] at 55 °C for 1 h. The prehybridized coverslips were then incubated with hybridization solution [2 × SSC, 1 × Denhards solution, 50% (v/v) formamide, 10 mM EDTA pH 8.0, 100 µg/mL yeast tRNA, 0.01% Tween-20, 5% (w/v), dextran sulfate and DIG-labeled RNA probes, which were generated using DIG RNA Labeling Kit (11277073910, Roche)] at 55 °C overnight. After hybridization, the coverslips were washed twice with wash buffer [2 × SSC, 50% (v/v) formamide and 0.01% Tween-20] at 55 °C for 30 min twice. To remove non-specific binding of RNA probe, cells were incubated in RNase A buffer (1 µg/mL RNase A, 0.5 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) for 1 h at 37 °C. After RNase A treatment, cells were washed with wash buffer2 (2 × SSC, 0.01% Tween-20) and wash buffer3 (0.2 × SSC, 0.01% Tween-20) for 30 min at 55 °C. For detection, coverslips were washed in TBST (Tris-buffered saline solution containing 0.01% Tween-20) for 5 min at room temperature, incubated with blocking solution [1 × Blocking reagent (11096176001, Roche) in TBST] for 5 min at room temperature, and then incubated with anti-DIG Rhodamine (11207750910, Roche) and 15 nM Silicon Rhodamine (SiR)-TH59 for 90 min at room temperature. After staining with 0.5 µg/mL DAPI in TBST for 5 min at room temperature, unbound antibodies were removed by washing three times with TBST for 5 min. Coverslips were then mounted and image acquisition was performed with a DeltaVision microscope. Signal

intensities were measured by Plot Profile tools in Fiji and line plots were created using R software [70].

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-021-00421-8>.

**Additional file 1: Figure S1.** Structure of TH59-DB and enrichment of ncRNAs by PI-PRiCh within subtelomeric regions of MEL cells. **a** Chemical structure of TH59-DB. Py, N-methyl pyrrole (blue) and Im, N-methyl imidazole (red). **b** A structural model of TH59-DB binding to DNA. **c**, ncRNAs mapped to the region near the end of each chromosome (chromosomes 13, 14, 15, 16, 17, 18) from Experiment 1. A genomic view of the terminal regions of the q arm of each chromosome based on the mouse reference genome, GRCm38 (mm10). RNAs were mapped to the terminal region of each chromosome. Results with the input, masked TH59-DB and TH59-DB are shown.

**Additional file 2: Figure S2.** Comprehensive identification of telomeric chromatin-associated RNAs in MEL cells (Experiment 2). **a** The percentage (yellow) of telomeric repeat reads, including TERRA transcripts, in the input and TH59-DB pull-down fractions from MEL cells. The number of single reads, including (TTAGGG)<sub>n</sub> or (CCCTAA)<sub>n</sub>, extracted from input and TH59-DB pull-down fractions were divided by the total number of reads. The gray color depicts other reads. **b** Scatter plot of fragments per kilobase of per million mapped reads (FPKM) of the TH59-DB pull-down fraction versus that of the input sample for each RNA. The telomerase RNA component, TERC (most enriched in the TH59-DB pull-down fraction), is highlighted in blue. RNAs enriched more than 100-fold in the TH59-DB pull-down fraction are plotted below a red dotted line. **c** Bar graph of FPKM of telomerase RNA-component (TERC) in input and TH59-DB pull-down fractions. **d** Sequence analysis of DNA fragments captured by TH59-DB telomeric repeat reads. The percentage (yellow) of telomeric repeat reads including TTAGGG or CCCTAA, in the input and TH59-DB pull-down fractions from MEL cells. Note that results are similar to those shown in Fig. 3.

**Additional file 3: Figure S3.** Genomic PCR to test telomeric sequence insertion. **a** A scheme of genomic PCR to test telomeric repeat insertion into the flanking regions of SPOCK3, USP16, CKS1B, NRDC, PAM and KIAA1671 introns. Primer sets and expected PCR products are shown. **b** PCR products of each intron amplified by the three primer sets with HeLa1.3 and ALT U2-OS genomic DNA. Lanes 1–3 and lanes 5–7 for SPOCK3, lanes 8–10, and lanes 12–14 for USP16, lanes 15–17 and lanes 19–21 for CKS1B, lanes 22–24 and lanes 26–28 for NRDC, lanes 29–31 and lanes 33–35 for PAM, lanes 36–38 and lanes 40–42 for KIAA1671. 100 bp or 1 kb ladder was loaded at the center lane of the gel as a size marker (M).

**Additional file 4: Figure S4.** PI-PRiCh identifies ALT cell-specific ncRNAs transcribed from around the inserted telomeric repeats. Telomeric repeat-associated ncRNAs mapped to intron regions of CKS1B, NRDC, PAM and KIAA1671 genes in ALT U2-OS cells. These ncRNAs were highly enriched in the TH59-DB pull-down fraction of U2-OS cells, but not in HeLa1.3. The positions of the inserted telomeric repeats are indicated by red arrowheads.

**Additional file 5: Figure S5.** DNA sequence analysis of TH59-DB binding sites. Genomic regions coding the introns of CKS1B, NRDC, PAM, and KIAA1671 were specifically enriched in the TH59-DB pull-down fraction from ALT U2-OS cells. These intron regions were enriched in the TH59-DB fraction from U2-OS, but not in HeLa1.3 cells. The positions of the inserted telomeric repeats are indicated by red arrowheads.

**Additional file 6: Figure S6.** RNA-FISH for intronic ncRNAs identified by PI-PRiCh in ALT cells and a model for chromatin tethering to telomeres with ncRNAs. **a** RNA-FISH for the USP16 intron and the simultaneous telomere labeling with the fluorescent TH59 probe. First column, DAPI signal; second column, ncRNA signal of TERRA; third column, fluorescent TH59 signal (telomere); fourth column, merged image of DAPI (blue), ncRNA (green) and telomere (red). **b** Enlarged images of the boxed region in **a**. Line plot of the USP16-intronic RNA signal and the telomeric signal on the

white dotted line in the left merged image. **c** Bar graph of percentages of cells in which USP16, SPOCK3, CKS1B and NRDC intronic RNA signals were co-localized with telomere signal. Three independent experiments were performed and the numbers of ncRNA signals measured for U2-OS cells were from 100 cells for each experiment. Error bars show standard deviation. **d** A model of how the genome region with inserted telomeric repeats tethers to telomere in ALT cells.

**Additional file 7: Table S1.** Mass spectrometry data of PI-PRiCh experiments from MEL cells.

**Additional file 8: Table S2.** Mapped and counted RNA fragments from MEL cells. Sheet 1: RNAs with over 100-fold enrichment in TH59-DB fraction by PRiCh in MEL experiment 1. Sheet 2: A non-filtered result of counted RNA fragments in MEL experiment 1. Sheet 3: RNAs with over 100-fold enrichment in TH59-DB fraction by PRiCh in MEL experiment 2. Sheet 4: A non-filtered result of counted RNA fragments in MEL experiment 2.

**Additional file 9: Table S3.** RNAs enriched in the U2-OS or HeLa1.3 cell line by filtering criteria. Sheet 1: U2-OS enriched RNAs by filtering criteria. Sheet 2: HeLa1.3 enriched RNAs by filtering criteria.

**Additional file 10: Tables S4.** Filtration procedure to identify RNAs enriched in the U2-OS or HeLa1.3 cell line. RNAs in U2-OS experiment 1. Sheet 1: Non-filtrated count data. Sheet 2: Enriched more than 25-fold. Sheet 3: 50 times more than another cell line.

**Additional file 11: Tables S5.** Filtration procedure to identify RNAs enriched in the U2-OS or HeLa1.3 cell line. RNAs in HeLa1.3 experiment 1. Sheet 1: Non-filtrated count data. Sheet 2: Enriched more than 25-fold. Sheet 3: 50 times more than another cell line.

**Additional file 12: Table S6.** Filtration procedure to identify RNAs enriched in the U2-OS or HeLa1.3 cell line. RNAs in U2-OS experiment 2. Sheet 1: Non-filtrated count data. Sheet 2: Enriched more than 25-fold. Sheet 3: 50 times more than another cell line.

**Additional file 13: Table S7.** Filtration procedure to identify RNAs enriched in the U2-OS or HeLa1.3 cell line. RNAs in HeLa1.3 experiment 2. Sheet 1: Non-filtrated count data. Sheet 2: Enriched more than 25-fold. Sheet 3: 50 times more than another cell line.

**Additional file 14: Table S8.** Mouse subtelomeric and telomeric regions for FPKM calculation in Fig. 3d.

**Additional file 15: Table S9.** Primers for telomeric insertion test and for RNA FISH and DNA sequences to search targets for the grep command in the UNIX system.

## Acknowledgements

We are grateful to Ms. H. Ochi for technical assistance and Dr. K.M. Marshall for critical reading and editing of this manuscript. We thank Dr. T. de Lange for the kind gift of HeLa1.3 cells, Dr. Y. Shinkai for anti-TRF1 antibody and S. Sakamoto for mass spectrometry analysis.

## Authors' contributions

SI and KM designed the project; YK, TB and HS synthesized the PI polyamide conjugated with a biotin analog (desthiobiotin) (TH59-DB); AS and SI performed most of the experiments; SI, AS and KM wrote the manuscript with input from all other authors. All authors read and approved the final manuscript.

## Funding

A.S. was a JSPS Fellow (DC2). This work was supported by an NIG collaborative Grant (2015-B6), JSPS Grants (JP17J10836 to A.S.; 15H01361 and 21H02535 to S.I.; 20H05936 and 21H02453 to K.M.), the Takeda Science Foundation to K.M. and the Uehara Memorial Foundation to K.M.

## Availability of data and materials

PI-PRiCh data are available in NCBI GEO database as GSE181609.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

No personal data are included.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Genome Dynamics Laboratory, National Institute of Genetics, ROIS, Mishima, Shizuoka 411-8540, Japan. <sup>2</sup>Department of Genetics, School of Life Science, SOKENDAI, Mishima, Shizuoka 411-8540, Japan. <sup>3</sup>Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606-8502, Japan.

Received: 11 August 2021 Accepted: 23 September 2021

Published online: 09 October 2021

## References

- Misteli T. The self-organizing genome: principles of genome architecture and function. *Cell*. 2020;183(1):28–45.
- Maeshima K, Iida S, Tamura S. Physical nature of chromatin in the nucleus. *Cold Spring Harb Perspect Biol*. 2021;13(5):a040675.
- Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell*. 1999;98(4):513–22.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129(7):1311–23.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120–4.
- Morin GB. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*. 1989;59(3):521–9.
- Shippen-Lentz D, Blackburn EH. Functional evidence for an RNA template in telomerase. *Science*. 1990;247(4942):546–52.
- Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*. 2007;318(5851):798–801.
- Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, et al. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA*. 2011;108(51):20497–502.
- Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011;44(4):667–78.
- Engreitz JM, Pandya-Jones A, McDonell P, Shishkin A, Sirokman K, Surka C, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*. 2013;341(6147):1237973.
- Gauchier M, van Mierlo G, Vermeulen M, Dejardin J. Purification and enrichment of specific chromatin loci. *Nat Methods*. 2020;17(4):380–9.
- Dejardin J, Kingston RE. Purification of proteins associated with specific genomic loci. *Cell*. 2009;136(1):175–86.
- Ide S, Dejardin J. End-targeting proteomics of isolated chromatin segments of a mammalian ribosomal RNA gene promoter. *Nat Commun*. 2015;6:6674.
- Grolimund L, Aeby E, Hamelin R, Armand F, Chiappe D, Moniatte M, et al. A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat Commun*. 2013;4:2848.
- Fujita T, Asano Y, Ohtsuka J, Takada Y, Saito K, Ohki R, et al. Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Sci Rep*. 2013;3:3171.
- Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In situ capture of chromatin interactions by biotinylated dCas9. *Cell*. 2017;170(5):1028–43.
- Tsui C, Inouye C, Levy M, Lu A, Florens L, Washburn MP, et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc Natl Acad Sci USA*. 2018;115(12):E2734–41.
- Maeshima K, Janssen S, Laemmli UK. Specific targeting of insect and vertebrate telomeres with pyrrole and imidazole polyamides. *EMBO J*. 2001;20(12):3218–28.
- Kawamoto Y, Bando T, Kamada F, Li Y, Hashiya K, Maeshima K, et al. Development of a new method for synthesis of tandem hairpin pyrrole-imidazole polyamide probes targeting human telomeres. *J Am Chem Soc*. 2013;135(44):16468–77.
- Hirata A, Nokihara K, Kawamoto Y, Bando T, Sasaki A, Ide S, et al. Structural evaluation of tandem hairpin pyrrole-imidazole polyamides recognizing human telomeres. *J Am Chem Soc*. 2014;136(32):11546–54.
- Kawamoto Y, Sasaki A, Hashiya K, Ide S, Bando T, Maeshima K, et al. Tandem trimer pyrrole-imidazole polyamide probes targeting 18 base pairs in human telomere sequences. *Chem Sci*. 2015;6:2307–12.
- Kawamoto Y, Sasaki A, Chandran A, Hashiya K, Ide S, Bando T, et al. Targeting 24 bp within telomere repeat sequences with tandem tetramer pyrrole-imidazole polyamide probes. *J Am Chem Soc*. 2016;138(42):14100–7.
- Kawamoto Y, Bando T, Sugiyama H. Sequence-specific DNA binding pyrrole-imidazole polyamides and their applications. *Bioorg Med Chem*. 2018;26(8):1393–411.
- Tsubono Y, Kawamoto Y, Hidaka T, Pandian GN, Hashiya K, Bando T, et al. A near-infrared fluorogenic pyrrole-imidazole polyamide probe for live-cell imaging of telomeres. *J Am Chem Soc*. 2020;142(41):17356–63.
- Bando T, Sugiyama H. Sequence-specific PI polyamides make it possible to regulate DNA structure and function. *Bull Chem Soc Jpn*. 2020;93(2):205–15.
- Trauger JW, Baird EE, Dervan PB. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*. 1996;382(6591):559–61.
- White S, Szweczyk JW, Turner JM, Baird EE, Dervan PB. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*. 1998;391(6666):468–71.
- Chenoweth DM, Dervan PB. Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci USA*. 2009;106(32):13175–9.
- Dervan PB. Molecular recognition of DNA by small molecules. *Bioorg Med Chem*. 2001;9(9):2215–35.
- Dervan PB, Edelson BS. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol*. 2003;13(3):284–99.
- Dervan PB, Doss RM, Marques MA. Programmable DNA binding oligomers for control of transcription. *Curr Med Chem Anticancer Agents*. 2005;5(4):373–87.
- Bando T, Sugiyama H. Synthesis and biological properties of sequence-specific DNA-alkylating pyrrole-imidazole polyamides. *Acc Chem Res*. 2006;39(12):935–44.
- Blackledge MS, Melander C. Programmable DNA-binding small molecules. *Bioorg Med Chem*. 2013;21(20):6101–14.
- Sasaki A, Ide S, Kawamoto Y, Bando T, Murata Y, Shimura M, et al. Telomere visualization in tissue sections using pyrrole-imidazole polyamide probes. *Sci Rep*. 2016;6:29261.
- Greider CW, Blackburn EH. A telomeric sequence in the RNA of tetrahymena telomerase required for telomere repeat synthesis. *Nature*. 1989;337(6205):331–7.
- Blackburn EH. Telomeres and telomerase: the means to the end (Nobel lecture). *Angew Chem*. 2010;49(41):7405–21.
- Azzalin CM, Lingner J. Telomere functions grounding on TERRA firma. *Trends Cell Biol*. 2015;25(1):29–36.
- Barral A, Dejardin J. Telomeric chromatin and TERRA. *J Mol Biol*. 2020;432(15):4244–56.
- Gauchier M, Kan S, Barral A, Sauzet S, Agirre E, Bonnell E, et al. SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Sci Adv*. 2019;5(5):eaav3673.
- Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*. 2008;10(2):228–36.
- Riethman H, Ambrosini A, Castaneda C, Finklestein J, Hu XL, Mudunuri U, et al. Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res*. 2004;14(1):18–28.
- Lopez de Silanes I, Grana O, De Bonis ML, Dominguez O, Pisano DG, Blasco MA. Identification of TERRA locus unveils a telomere protection role through association to nearly all chromosomes. *Nat Commun*. 2014;5:4723.

44. Viceconte N, Lorient A, Lona Abreu P, Scheibe M, Fradera Sola A, Butcher F, et al. PAR-TERRA is the main contributor to telomeric repeat-containing RNA transcripts in normal and cancer mouse cells. *RNA*. 2021;27(1):106–21.
45. Tardat M, Dejardin J. Telomere chromatin establishment and its maintenance during mammalian development. *Chromosoma*. 2018;127(1):3–18.
46. Arora R, Azzalin CM. Telomere elongation chooses TERRA ALTERNatives. *RNA Biol*. 2015;12(9):938–41.
47. Graf M, Bonetti D, Lockhart A, Serhal K, Kellner V, Maicher A, et al. Telomere length determines TERRA and R-Loop regulation through the cell cycle. *Cell*. 2017;170(1):72–85.
48. Sakellariou D, Chiourea M, Raftopoulou C, Gagos S. Alternative lengthening of telomeres: recurrent cytogenetic aberrations and chromosome stability under extreme telomere dysfunction. *Neoplasia*. 2013;15(11):1301–13.
49. Marzec P, Armenise C, Perot G, Roumelioti FM, Basyuk E, Gagos S, et al. Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell*. 2015;160(5):913–27.
50. Sieverling L, Hong C, Koser SD, Ginsbach P, Kleinheinz K, Hutter B, et al. Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat Commun*. 2020;11(1):733.
51. Negishi Y, Kawaji H, Minoda A, Usui K. Identification of chromatin marks at TERRA promoter and encoding region. *Biochem Biophys Res Commun*. 2015;467(4):1052–7.
52. Nergadze SG, Farnung BO, Wischnewski H, Khoraiuli L, Vitelli V, Chawla R, et al. CpG-island promoters drive transcription of human telomeres. *RNA*. 2009;15(12):2186–94.
53. Iguchi A, Fukuda N, Takahashi T, Watanabe T, Matsuda H, Nagase H, et al. RNA binding properties of novel gene silencing pyrrole-imidazole polyamides. *Biol Pharm Bull*. 2013;36(7):1152–8.
54. Sridhar B, Rivas-Astroza M, Nguyen TC, Chen W, Yan Z, Cao X, et al. Systematic mapping of RNA-chromatin interactions in vivo. *Curr Biol*. 2017;27(4):602–9.
55. Li X, Zhou B, Chen L, Gou LT, Li H, Fu XD. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol*. 2017;35(10):940–50.
56. Bell JC, Jukam D, Teran NA, Risca VI, Smith OK, Johnson WL, et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife*. 2018;7:e27024.
57. Bonetti A, Agostini F, Suzuki AM, Hashimoto K, Pascarella G, Gimenez J, et al. RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun*. 2020;11(1):1018.
58. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
59. Maeshima K, Tamura S, Hansen JC, Itoh Y. Fluid-like chromatin: toward understanding the real chromatin organization present in the cell. *Curr Opin Cell Biol*. 2020;64:77–89.
60. Itoh Y, Woods EJ, Minami K, Maeshima K, Collepardo-Guevara R. Liquid-like chromatin in the cell: what can we learn from imaging and computational modeling? *Curr Opin Struct Biol*. 2021;71:123–35.
61. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–82.
62. Benetti R, Garcia-Cao M, Blasco MA. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat Genet*. 2007;39(2):243–50.
63. Lee M, Hills M, Conomos D, Stutz MD, Dagg RA, Lau LM, et al. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Res*. 2014;42(3):1733–46.
64. Takai KK, Hooper S, Blackwood S, Gandhi R, de Lange T. In vivo stoichiometry of shelterin components. *J Biol Chem*. 2010;285(2):1457–67.
65. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–2.
66. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863–4.
67. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60.
68. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
69. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
70. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.
71. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.
72. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
73. Kawaguchi T, Tanigawa A, Naganuma T, Ohkawa Y, Souquere S, Pierron G, et al. SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies. *Proc Natl Acad Sci USA*. 2015;112(14):4304–9.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

