

## REPORT

# TERRA ONTseq: a long-read-based sequencing pipeline to study the human telomeric transcriptome

JOANA RODRIGUES,<sup>1</sup> ROBERTA ALFIERI,<sup>2,3</sup> SILVIA BIONE,<sup>2</sup> and CLAUS M. AZZALIN<sup>1</sup>

<sup>1</sup>Instituto de Medicina Molecular João Lobo Antunes (iMM), Faculdade de Medicina da Universidade de Lisboa, Lisbon 1649-028, Portugal

<sup>2</sup>Istituto di Genetica Molecolare Luigi Luca Cavalli-Sforza, Consiglio Nazionale delle Ricerche, Pavia 27100, Italy

<sup>3</sup>Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Segrate (MI) 20054, Italy

## ABSTRACT

The long noncoding RNA TERRA is transcribed from telomeres in virtually all eukaryotes with linear chromosomes. In humans, TERRA transcription is driven in part by promoters comprising CpG dinucleotide-rich repeats of 29 bp repeats, believed to be present in half of the subtelomeres. Thus far, TERRA expression has been analyzed mainly using molecular biology-based approaches that only generate partial and somehow biased results. Here, we present a novel experimental pipeline to study human TERRA based on long-read sequencing (TERRA ONTseq). By applying TERRA ONTseq to different cell lines, we show that the vast majority of human telomeres produce TERRA and that the cellular levels of TERRA transcripts vary according to their chromosomes of origin. Using TERRA ONTseq, we also identified regions containing TERRA transcription start sites (TSSs) in more than half of human subtelomeres. TERRA TSS regions are generally found immediately downstream from 29 bp repeat-related sequences, which appear to be more widespread than previously estimated. Finally, we isolated a novel TERRA promoter from the highly expressed subtelomere of the long arm of Chromosome 7. With the development of TERRA ONTseq, we provide a refined picture of human TERRA biogenesis and expression and we equip the scientific community with an invaluable tool for future studies.

**Keywords:** telomeres; TERRA; transcription; CpG dinucleotide promoters; long-read nanopore sequencing

## INTRODUCTION

Telomeres are protective protein/nucleic acid structures located at the ends of linear eukaryotic chromosomes. In mammalian cells, telomeric DNA is made up of arrays of the (TTAGGG)<sub>n</sub> sequence, varying in length roughly from 10 to 15 kb among different cell types and different chromosome ends within the same cell (Srinivas et al. 2020). Telomeres are transcribed by DNA-dependent RNA polymerase II (RNAPII) into the evolutionarily conserved long noncoding RNA TERRA (Azzalin et al. 2007; Azzalin and Lingner 2008; Schoeftner and Blasco 2008). TERRA transcription starts within subtelomeres and proceeds toward the ends of the chromosomes with the C-rich telomeric strand acting as an RNAPII template. Hence, individual TERRA molecules comprise a sequence specific to the subtelomere of origin followed by stretches of UUAGGG repeats (Azzalin et al. 2007; Schoeftner and Blasco 2008; Porro et al. 2010; Farnung et al. 2012).

In humans, transcription of several TERRA molecules initiates at CpG dinucleotide-rich tandem repeats of 29 base pairs, which were reported to be present in about half of the subtelomeres (Nergadze et al. 2009; Feretzaki et al. 2019). CpG dinucleotides within 29 bp repeats are heavily methylated, in particular in telomerase-positive cells, which normally contain relatively low TERRA levels. Simultaneous inactivation of the DNA methyltransferases DNMT1 and DNMT3b in HCT116 cancer cells (DKO cells) fully eliminates CpG methylation at 29 bp repeats, leading to a substantial increase in TERRA levels and more avid RNAPII binding to 29 bp and TTAGGG repeats (Nergadze et al. 2009). Based on these data, we proposed that 29 bp repeat sequences are bona fide TERRA promoters whose activity is epigenetically regulated. Importantly, chromosome ends devoid of 29 bp repeats are also highly transcribed in DKO cells (Feretzaki et al. 2019), suggesting the existence of CpG-containing promoters other than the canonical 29 bp repeats. However, only with the recent Telomere-to-Telomere (T2T) sequence of the human

**Corresponding author:** [cmazzalin@medicina.ulisboa.pt](mailto:cmazzalin@medicina.ulisboa.pt)

Handling editor: Ling-Ling Chen

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.079906.123>. Freely available online through the RNA Open Access option.

© 2024 Rodrigues et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

genome (Nurk et al. 2022; Rhie et al. 2023) can the exact structure of those 29 bp-less promoters be systematically investigated.

Numerous functions have been assigned to TERRA, including telomerase recruitment to telomeres and activity regulation, heterochromatin establishment, and overall telomere integrity maintenance (Rivosecchi and Cusanelli 2023; Zeinoun et al. 2023). TERRA and telomere transcription are also essential to sustain telomere elongation in telomerase-negative cancers that maintain telomeres through the alternative lengthening of telomeres (ALT) mechanism (Silva et al. 2021, 2022; Bhargava et al. 2022). In agreement, in ALT cells, 29 bp repeat sequences are hypomethylated and TERRA levels are consistently higher compared to those in telomerase-positive cells (Nergadze et al. 2009; Lovejoy et al. 2012; Arora et al. 2014).

The analysis of TERRA transcripts remains technically challenging because of their low abundance and highly repetitive nature. Up to this date, TERRA quantifications have been mainly done using reverse transcription-quantitative PCR (RT-qPCR) with subtelomeric oligonucleotides or northern blotting with probes detecting (UUAGGG)<sub>n</sub> sequences (Azzalin et al. 2007; Porro et al. 2010; Arnoult et al. 2012; Deng et al. 2012; Feretzaki et al. 2019). Although both methods provide robust data, they also generate partial results. With RT-qPCR, only a few chosen telomeres are analyzed in each experiment; on the other side, when performing northern blotting to detect UUAGGG repeats, total TERRA from all telomeres is simultaneously visualized but the amount of TERRA produced from each chromosome end cannot be determined. Illumina sequencing of purified TERRA was also performed (Porro et al. 2014). However, human subtelomeres comprise heterogeneously repeated sequences shared by multiple chromosome ends (Kwapisz and Morillon 2020), impeding to unequivocally assign small reads from Illumina sequencing to the genome. Indeed, different results were obtained by independent groups starting from the same sets of TERRA Illumina reads (Porro et al. 2014; Montero et al. 2016). Overall, systematic studies in which all TERRA molecules are unambiguously analyzed in a quantitative and qualitative way are lacking.

We have established an Oxford Nanopore Technology (ONT)-based long-read sequencing pipeline to sequence and map the subtelomeric tracts of TERRA molecules purified from human cultured cells (TERRA ONTseq). By applying TERRA ONTseq to different human cell lines and using the latest T2T genome as a reference, we show that TERRA transcripts originate from most chromosome ends, although transcript levels vary according to the end of origin. We also show that 29 bp repeat-like sequences are present in the majority of subtelomeres and that TERRA transcription often starts immediately after those repeats. Finally, we isolated a polymorphic CpG-rich TERRA promoter se-

quence that is located within the highly expressed subtelomere of the long arm of Chromosome 7 (7q subtelomere). TERRA ONTseq constitutes a powerful tool generating qualitative and quantitative insights into TERRA biogenesis and regulation.

## RESULTS AND DISCUSSION

### TERRA ONTseq detects transcripts deriving from a multitude of human chromosome ends

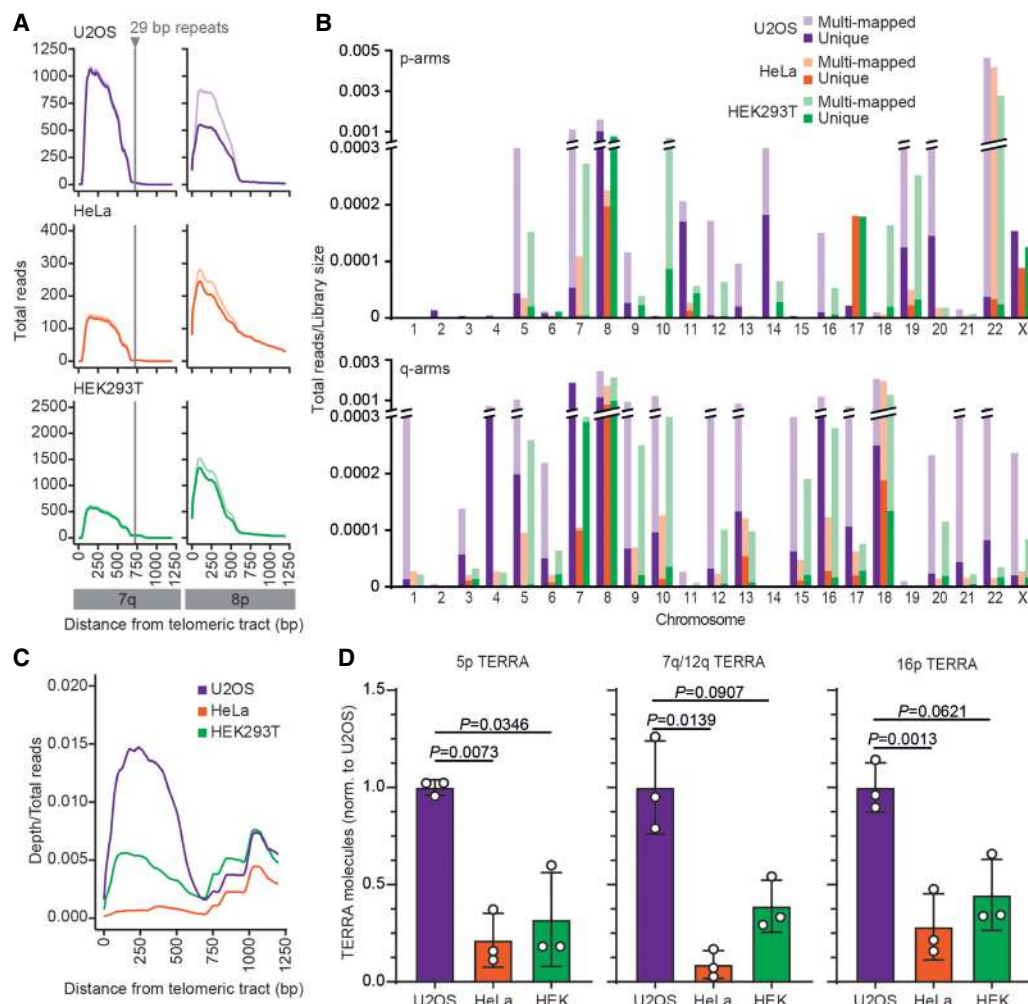
The ONT sequencing technology allows the sequencing of full-length DNA molecules of up to 100 kb, thus bypassing the need to fragment molecules for library preparation and being ideally suited for the sequencing of repetitive nucleic acids (Kwapisz and Morillon 2020; Warburton and Sebra 2023). We affinity-purified TERRA transcripts from nuclear RNA prepared from two telomerase-positive (HeLa and HEK293T) and two ALT (U2OS and SAOS2) cell lines using a biotinylated DNA oligonucleotide comprising five CCCTAA repeats, which are complementary to the UUAGGG tracts within TERRA molecules (Supplemental Fig. S1A; Supplemental Table S1). Biotinylated oligonucleotides were 3' end blocked with dideoxycytosine to avoid interference with the subsequent step of reverse transcription. As shown by northern blotting, purified RNA fractions were largely devoid of the U6 small nuclear RNA, while containing telomeric UUAGGG repeats (Supplemental Fig. S1B). As quantified by RT-qPCR, the purification efficiencies of TERRA transcripts differed depending on the chromosome end and cell line (Supplemental Fig. S1C), possibly because of different conservations of the (UUAGGG)<sub>n</sub> sequence. With the exception of SAOS2, TERRA purification consistently provided sufficient material for ONT sequencing across all tested cell lines (Supplemental Fig. S1C).

TERRA-enriched RNA was reverse transcribed with an oligonucleotide complementary to the UUAGGG repeats (Supplemental Fig. S1A; Supplemental Table S1). Because the C-rich oligonucleotide can hybridize randomly along the (UUAGGG)<sub>n</sub> stretch, including in close proximity of the subtelomeric part of the transcript, this step assures that the majority of reads contain subtelomeric sequences and can be mapped. However, our protocol cannot provide information on the length of the UUAGGG tract. The obtained cDNA was Nanopore sequenced and quality score analysis identified reads up to 11,375, 12,310, and 23,869 bp long for U2OS, HEK293T, and HeLa, respectively (mean read length: 1064, 866, and 964 bp, respectively; median read length: 887, 684, and 779 bp, respectively). Preliminary alignments of the reads to the build T2T CHM13v2.0/hs1 of the human genome (Nurk et al. 2022) did not detect any splicing signature (data not shown), which is consistent with previous reports (Porro et al. 2014). Hence, we refined the alignment

strategy by generating a subtelomeric reference sequence from the T2T genome comprising 25 kb sequences terminating at the fourth telomeric TTAGGG repeats on each chromosome end. Because all cell lines used in this study are from female donors, the recently completed sequence of Chromosome Y was not included in the reference (Rhie et al. 2023).

For each cell line, we calculated the total number of both unique and multimapped reads (Supplemental Table S2), and when all reads from the three cell lines were pooled, at least 10 reads were assigned to 40 out of 46 chromosome ends (19 short arms and 21 long arms) (Fig. 1A,B; Supplemental Figs. S2–S4; Supplemental Table S2). In

the majority of the cases, read density peaked in the vicinity of the telomeric repeats and declined when moving toward the inside of the subtelomere, with reads terminating ~500–750 bp from the telomeric repeats (Fig. 1A,C; Supplemental Figs. S2–S4). A striking different behavior was observed for the 18q and 22p (short arm) subtelomeres, for which reads started 750–1000 bp away from the first telomeric repeat and covered a region of ~500 bp (Supplemental Figs. S2–S4). This discrepancy may reflect differently structured 18q and 22p subtelomeres in the T2T genome and the cell lines analyzed in this study. Altogether these data reveal that (1) TERRA is transcribed from a large number of telomeres, as previously proposed



**FIGURE 1.** TERRA expression in different human cell lines. (A) TERRA ONTseq read coverage for 7q and 8p subtelomeres in U2OS, HeLa, and HEK293T. Mapping was performed against the T2T CHM13v2.0/hs1 subtelomeric reference genome. Alignments span 1250 bp regions starting from the first telomeric repeat. The darker and lighter shade lines represent unique and multimapped reads, respectively. The vertical gray lines indicate the position of the most telomere-proximal 29 bp–like repeats. (B) Quantification of the unique and multimapped reads (darker and lighter shades, respectively) within 1250 bp regions as in A for each human subtelomere. Reads are normalized to the corresponding library size. (C) Metaprofiles of uniquely mapped reads for U2OS, HeLa, and HEK293T cells. Note that the read density profiles between 750 and 1250 bp largely derive from Chromosomes 8q, 18q, and 22p only. (D) RT-qPCR quantifications of TERRA transcripts from 5p, 7q/12q, and 16p subtelomeres in the indicated cell lines. Values are normalized to the number of molecules in U2OS cells. Bars and error bars are means and SDs from three independent experiments. *P* values (Student’s *t*-test) are indicated.

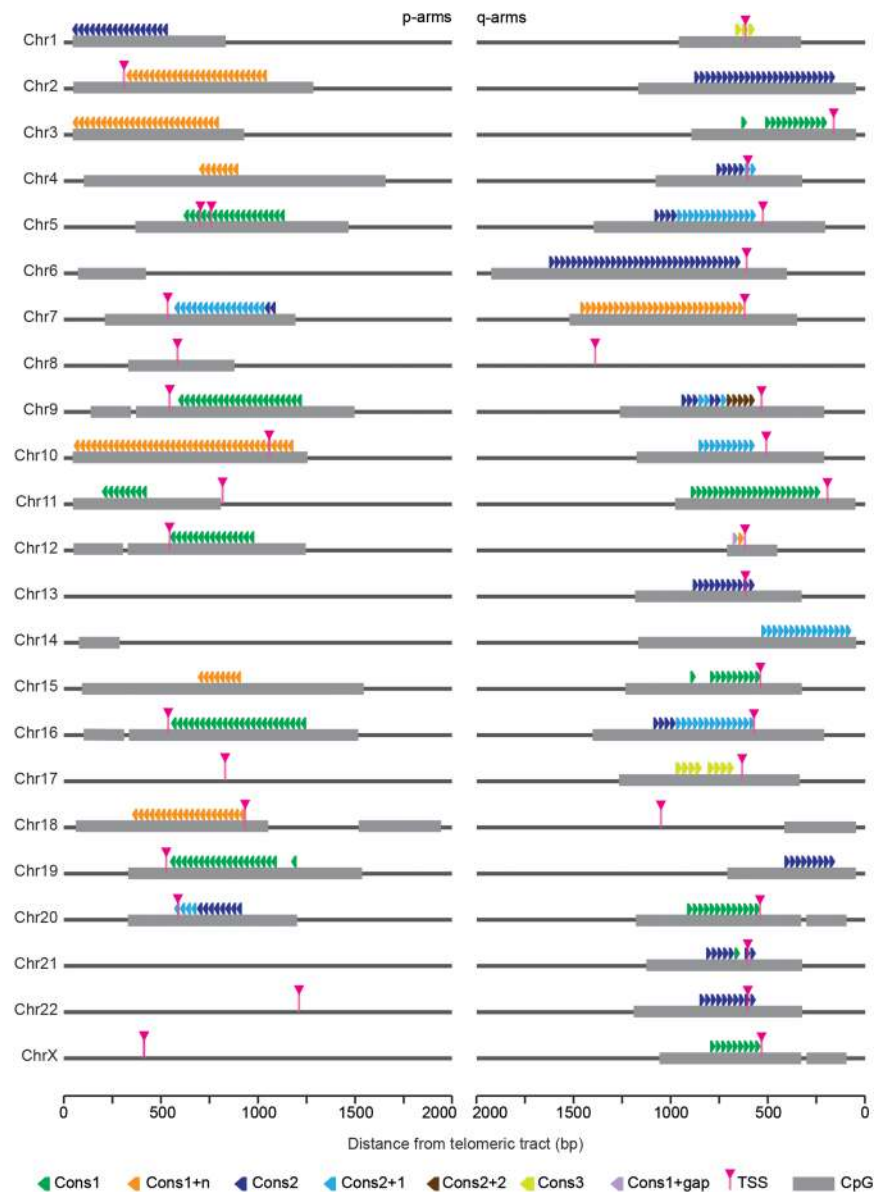
(Azzalin et al. 2007; Nergadze et al. 2009; Arora et al. 2014; Episkopou et al. 2014; Porro et al. 2014; Feretzaki et al. 2019); and (2) most of TERRA molecules originate from regions very close to the first telomeric repeats and not from distant subtelomeric loci; this confirms that the heterogeneity of TERRA transcripts largely stems from the UUAGGG tracts (Azzalin et al. 2007; Farnung et al. 2012).

Read mapping across different chromosome ends was not uniform. For example, the subtelomeres 8p, 7q, and 8q were highly expressed in all cell lines (Fig. 1A,B; Supplemental Figs. S2–S4; Supplemental Table S2). Conversely, some chromosome ends were underrepresented in our alignments, including the ones of the short arms of Chromosomes 1, 2, 3, 4, 6, 15, and 21 and the long arms of Chromosomes 2, 11, 14, and 19 (Fig. 1B; Supplemental Figs. S2–S4; Supplemental Table S2). We conclude that TERRA transcripts from different ends are synthesized and/or degraded at different rates, as previously suggested based on the analysis of a few chromosome ends (Feretzaki et al. 2019; Savoca et al. 2023). Technical limitations, including differences in the efficiencies of purification (Supplemental Fig. S1C) and/or reverse transcription for different TERRA species, may also account for the observed variabilities. Finally, we generated TERRA ONTseq metaprofiles and found that U2OS, HeLa, and HEK293T showed similar profile shapes; however, U2OS cells had the highest overall read count, followed by HEK293T and HeLa (Fig. 1C). These data were validated using RT-qPCRs detecting a few selected TERRA species (Fig. 1D). Because ALT cells are characterized by elevated TERRA levels (Lovejoy et al. 2012; Arora et al. 2014; Episkopou et al. 2014), TERRA ONTseq appears to be able to quantitatively measure TERRA transcripts not only across different chromosome ends but also different samples.

### TERRA ONTseq identifies TERRA transcription start site regions

To position TERRA promoters and transcription start sites (TSSs), we first reanalyzed the distribution of tandem

repeats related to the canonical 29 bp repeats in our T2T-derived subtelomeric reference and identified them in almost all subtelomeres, with the exception of 6p, 8p, 13p, 14p, 17p, 21p, 22p, Xp, 8q, and 18q (Fig. 2; Supplemental Table S4). More specifically, we found 12 subtelomeres containing the canonical 29 bp sequence (Cons1) (Fig. 2; Supplemental Table S4); 13 subtelomeres containing the canonical 29 bp sequence with a single mismatched nucleotide (Cons2); eight subtelomeres containing a 29 bp



**FIGURE 2.** Twenty-nine#bp-like repeats, CpG-rich islands, and TERRA TSSs across all human chromosome ends. Schematic representation of the positions of the different variants of 29 bp-like repeats (triangles) and subtelomeric CpG islands (gray boxes) in the T2T CHM13v2.0/hs1 subtelomeric reference genome. TERRA TSSs identified by TERRA ONTseq are also indicated (pink arrowheads). For each chromosome arm, the most telomere-proximal 2000 bp are shown. Note that Chromosome 5p has two TERRA TSSs. The one on the left was identified in HeLa cells while the one on the right was identified in U2OS and HEK293T cells.



sequence with two mismatches compared to Cons1 and 1 to Cons 2 (Cons2 + 1); and eight subtelomeres containing a 29 bp sequence with a variable number of mismatches with respect to Cons1 (Cons1 +  $n$ ). In addition, the 9q subtelomere contains a 29 bp repeat with two mismatches with respect to Cons2 (Cons2 + 2) (Fig. 2; Supplemental Table S4); the 12q subtelomere contains a 30 bp sequence with 1 nt insertion and two mismatches relative to Cons1 (Cons1 + gap); and finally, the 1q and 17q subtelomeres contain a 34 bp repeat sequence with two insertions of 2 and 3 nt and one mismatch with respect to Cons2 (Cons2). We also predicted telomere-adjacent CpG islands using the T2T reference sequence and identified at least one of them within 2000 bp from the first telomeric repeat on all subtelomeres with the exception of the ones of 13p, 17p, 21p, 22p, Xp, and 8q arms. The sizes of the identified CpG islands are comprised between 259 and 1553 bp (in 12q and 4p, respectively) and often contain multiple repetitions of one or more 29 bp consensus sequences (Fig. 2).

We then used TERRA ONTseq reads to identify TSS-containing regions by defining a threshold of at least three subtelomeric reads with 5' ends mapping to the same 10 bp window. This strategy successfully identified 10 bp regions containing TERRA TSSs in 35 subtelomeres (Fig. 2; Supplemental Table S3). Notably, TERRA TSS regions are predominantly positioned within the CpG islands, just downstream from the 29 bp repeats (Fig. 2). This validates the ability of our pipeline to define TERRA TSSs and confirms that TERRA promoters are comprised of a variety of tandemly repeated 29 bp-like sequences. Importantly, some of the identified 29 bp-like repeats are only a few nucleotides away from the first telomeric repeats, as it is the case for the 1p, 3p, 10p, and 14q subtelomeres (Fig. 2), for which virtually no unique TERRA reads were detected (Fig. 1B; Supplemental Figs. S2–S4; Supplemental Table S2). It is plausible that those chromosome ends are also transcribed into TERRA molecules, but those transcripts escaped our analysis because the subtelomeric tract is too short to be mapped.

### TERRA ONTseq detects quantitative differences in TERRA molecules

As mentioned, TERRA ONTseq appears to be able to quantitate differences among different samples. To corroborate this assumption using isogenic cell systems, we took advantage of a previously described U2OS-derived cell line (SID4 cells), in which TERRA expression from subtelomeres containing a 20 bp sequence within the 29 bp repeat consensus can be suppressed using activator-like effectors (T-TALEs) fused to the transcription repressor domain SID4X and cloned under the control of a doxycycline (dox) inducible promoter (Silva et al. 2021). Control cells expressing unfused T-TALEs (NLS3 cells) (Silva et al.

2021) were also used. SID4 and NLS3 cells were treated with dox for 24 h or left untreated and subjected to TERRA ONTseq (Supplemental Fig. S5A). T-TALE expression and suppression of TERRA transcription from selected chromosome ends containing 20 bp target sequences were validated by western blot and RT-qPCR, respectively (Supplemental Fig. S5B,C).

SID4 cells generated lower TERRA read counts (normalized to library size) than NLS3 cells already in the absence of dox treatment, possibly because of leaky expression of the T-TALE constructs (Supplemental Fig. S5A,B). The dox treatment decreased the number of reads mapped to subtelomeres containing the T-TALE target in the SID4 samples (average 3.4-fold decrease), whereas a milder effect was observed at subtelomeres devoid of the 20 bp T-TALE target (average 1.4-fold decrease) (Supplemental Fig. S5A; Supplemental Table S2). In NLS3 cells, dox treatment did not cause noticeable changes in TERRA reads at any chromosome ends (Supplemental Fig. S5A; Supplemental Table S2). Overall, these data confirm that TERRA ONTseq detects quantitative differences in TERRA levels between independent samples. Moreover, suppression of TERRA transcription from a fraction of chromosome ends does not trigger a compensatory mechanism increasing transcription from the other ends, at least in U2OS cells.

While analyzing the T-TALE data, we noticed that the 6q and 7q subtelomeres did not respond to dox treatment, despite being annotated in the T2T genome as containing numerous copies of the 20 bp target sequence (Supplemental Fig. S6A; Supplemental Table S4). This raised the possibility that the 6q and 7q subtelomeres contain 29 bp repeat-like sequences in the CHM13 cell line that was used for T2T sequencing (Nurk et al. 2022) but not in other cell lines. We thus focused on the highly expressed 7q subtelomere, which is annotated in the T2T genome assembly as having 29 copies of the Cons1 +  $n$  29 bp repeat variant (Supplemental Fig. S6A; Supplemental Table S4). The same locus is devoid of 29 bp repeats in the GRCh38/hg38 human genome assembly (Supplemental Fig. S6A, B; Supplemental Table S4). Additionally, a previous study on the architecture of 7q subtelomere did not detect 29 bp repeats in 10 DNA samples from unrelated individuals (Baird et al. 2000). We PCR-amplified HeLa and U2OS genomic DNA with oligonucleotides flanking the 29 bp repeats in the T2T 7q subtelomere and obtained only amplicons of sizes consistent with the lack of the 29 bp repeats tract (1537 and 697 bp for fragments with and without insertion, respectively; data not shown). We also cloned the amplified fragments into a plasmid and Sanger sequenced independent inserts for each cell line, and did not detect 29 bp repeat-like sequences in any of the inserts (Supplemental Fig. S6A,B). Hence, the 29 bp repeats present in the T2T genome at the 7q subtelomere might correspond to an insertion polymorphism that is not

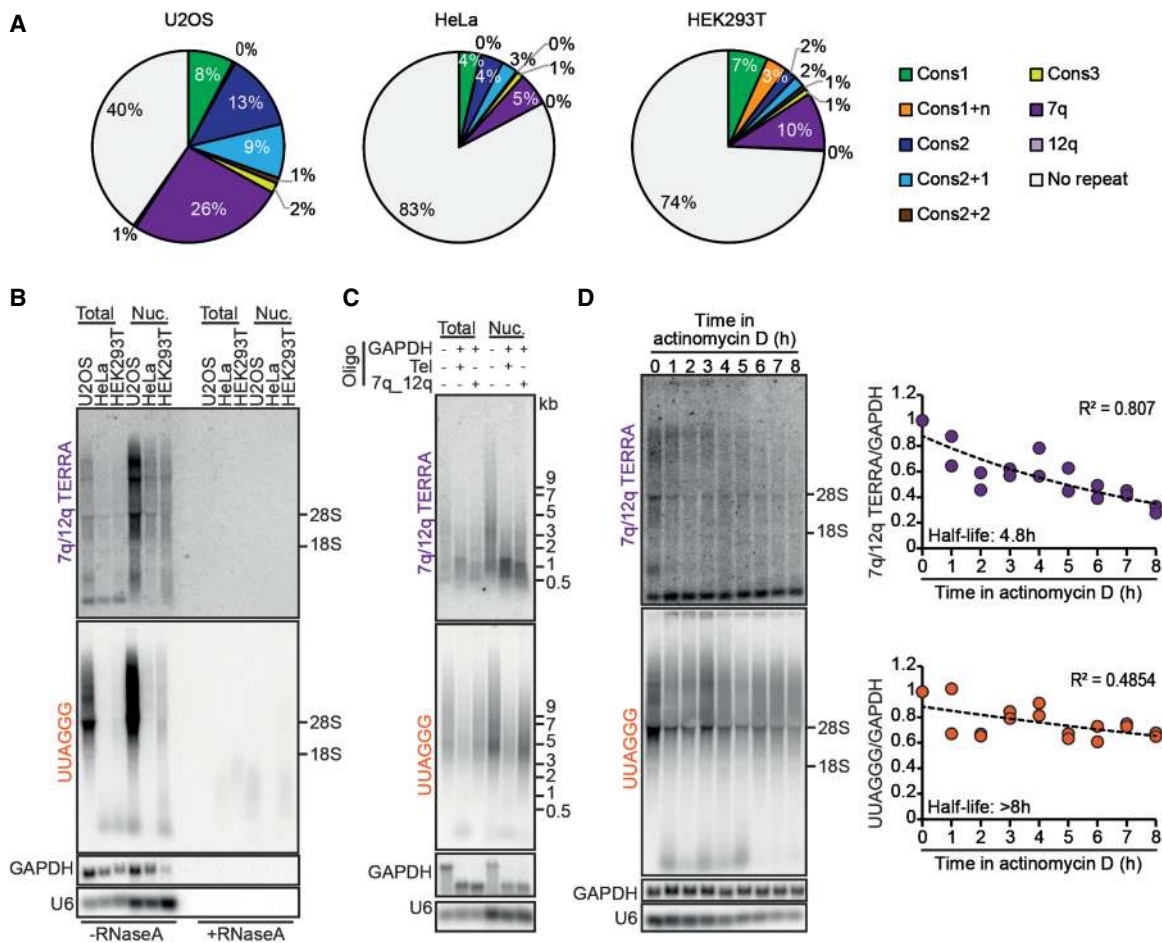
shared by all individuals or by HeLa and U2OS cells. This explains why our T-TALEs did not affect the expression of 7q TERRA and suggests a similar explanation also for the 6q subtelomere. Given these findings, we used the GRCh38/hg38 sequence for the 7q alignments performed later during this study.

### Analysis of 7q TERRA expression

The highly expressed 7q TERRA accounts for 26%, 5%, and 10% of total unique TERRA reads in U2OS, HeLa, and HEK293T cells, respectively (Fig. 3A). To better characterize this transcript, we first performed northern blotting using a probe corresponding to a 441 bp fragment from the 7q subtelomere (7q/12q probe) (Supplemental Fig. S6C).

Because of the high identity between 7q and 12q subtelomeres (Supplemental Fig. S6C), our probe is expected to detect TERRA species from both loci; nonetheless, based on our TERRA ONTseq data, 74% of the signal generated by this probe should derive only from 7q TERRA (Fig. 1B; Supplemental Table S2). We isolated total and nuclear RNA from U2OS, HeLa, and HEK293T cells and subjected it to northern blotting using 7q/12q and total (UUAGGG) TERRA probes. The signals generated by the two probes were similar and appeared as smears ranging from <0.5 to >9 kb (Fig. 3B,C).

We then used a northern blot-based approach previously developed to analyze TERRA transcripts from specific subtelomeres (Nergadze et al. 2009). We incubated total and nuclear RNA from U2OS cells with an oligonucleotide



**FIGURE 3.** Expression of 7q TERRA. (A) Distribution of TERRA reads from 7q, 12q, 29 bp-like-containing subtelomeres and from repeat-devoid subtelomeres in the indicated cell lines. Only unique reads were considered for the analysis. (B) Northern blot hybridizations of total and nuclear RNA from U2OS, HeLa, and HEK293T cells using probes for the 7q/12q TERRA or UUAGGG repeats (total TERRA). The positions of 28S and 18S ribosomal RNAs are indicated on the right. GAPDH and U6 were used as loading controls. RNaseA treatments were used to assure that signals derived from RNA and not genomic DNA contaminations. (C) Total and nuclear RNA from U2OS cells was incubated with RNaseH in the presence of oligonucleotides complementary to a sequence from 7q/12q TERRA or UUAGGG repeats (Tel). Oligonucleotides complementary to a GAPDH sequence were also included as indicated and served as the control for the RNaseH treatment. Products were analyzed by northern blotting as in B. (D) U2OS cells were treated with actinomycin D for the indicated times. Total RNA was extracted and analyzed by northern blotting as in C. The long-lived GAPDH mRNA was used as a loading control. Half-lives and  $R^2$  values were calculated using a scatter plot analysis after the 7q/12q or the TERRA signals were normalized to the GAPDH signals. Data points are from two independent experiments and the values for time 0 are set to 1.

corresponding to a unique sequence shared by 7q and 12q subtelomeres and located 85 bp from the telomeric tract in the 7q subtelomere or with a (CCCTAA)<sub>5</sub> oligonucleotide (7q<sub>12q</sub> or Tel oligonucleotide, respectively) (Supplemental Fig. S6C). In both reactions, we also included a control oligonucleotide annealing to GAPDH mRNA. We treated the RNA with RNaseH, which digests specifically RNA molecules engaged in DNA/RNA hybrids, and analyzed it by northern blotting (Fig. 3C). When compared to the undigested controls, the 7q/12q TERRA signal in RNaseH-treated sample incubated with the telomeric C-rich oligonucleotide appeared as a smear between ~1000 and 500 bases; this smear should correspond to the 7q/12q TERRA subtelomeric sequences largely devoid of telomeric repeats. RNaseH digestion in the presence of the 7q<sub>12q</sub> oligonucleotide produced a slightly lower major band compared to the digestion with the telomeric oligonucleotide, confirming that the 7q/12q probe mainly detects 7q/12q TERRA and that these transcripts contain long telomeric UUAGGG tracts (Fig. 3C). Hybridization with a telomeric probe showed, as expected, a significantly reduced signal from the RNA incubated with the Tel oligonucleotide, whereas the incubation with the 7q<sub>12q</sub> oligonucleotide led to a substantial disappearance of longer TERRA species above 9 kb (Fig. 3C). Attesting the efficiency and specificity of the RNaseH treatment, all GAPDH transcripts were shorter in samples treated with RNaseH, whereas the length of the unrelated U6 transcript remained unaffected (Fig. 3C). Altogether, these experiments confirm that 7q TERRA constitutes a substantial fraction of total TERRA and reveals that individual 7q TERRA species contain long and heterogeneous UUAGGG repeat tracts in all tested cell lines.

To determine whether the increased levels of 7q TERRA derive from increased transcription or RNA stability, we measured 7q/12q TERRA stability by treating U2OS cells with actinomycin D to halt transcription and collecting total RNA every hour over 8 h. Northern blot analysis showed that 7q/12q TERRA has a half-life of 4.8 h, whereas total TERRA has a half-life of >8 h (Fig. 3D), which is consistent with previously reported values (Azzalin and Lingner 2008; Porro et al. 2010; Savoca et al. 2023). Hence the abundance of 7q TERRA cannot be ascribed to overall increased RNA stability, but rather it should derive from higher transcription rates.

### Analysis of 7q TERRA promoter regulation

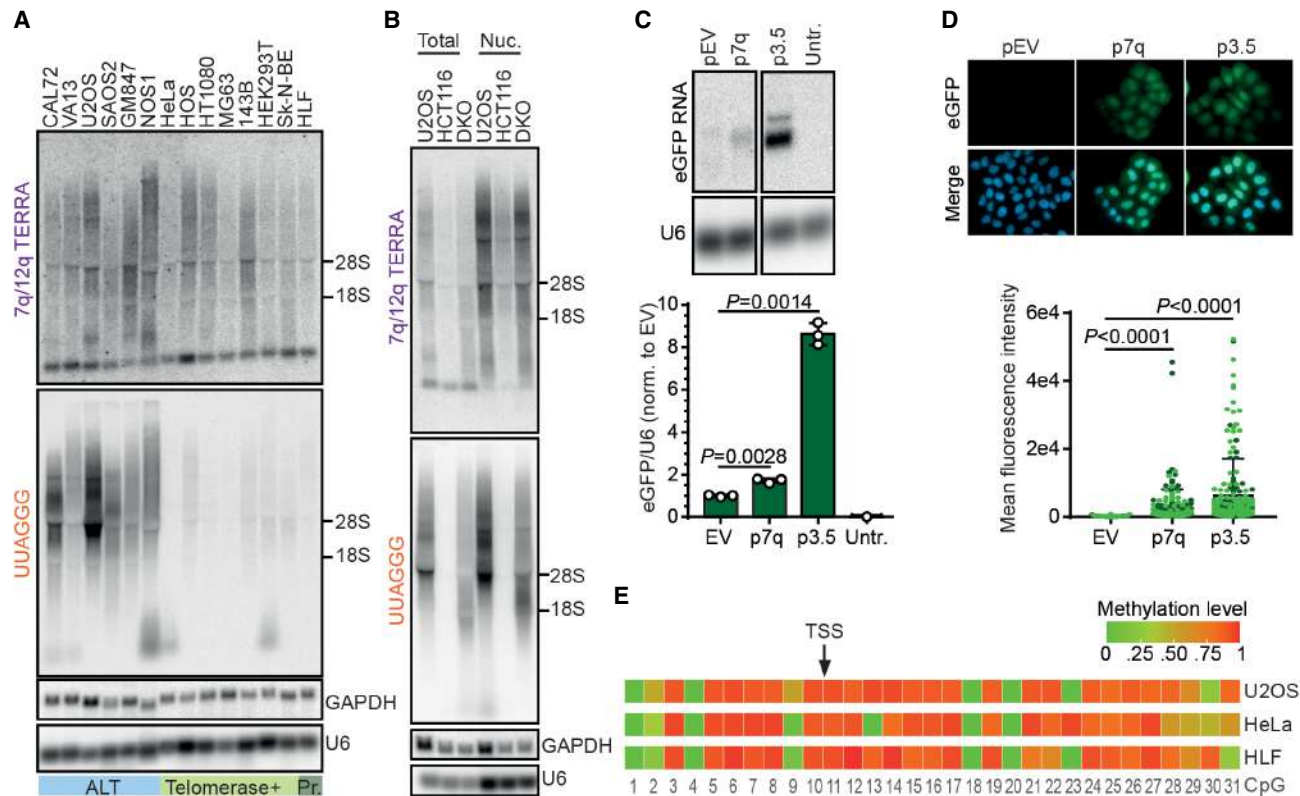
7q/12q TERRA levels are higher in U2OS ALT cells than in HeLa and HEK293T telomerase-positive cells; however, those differences are not as pronounced as the ones detected for total UUAGGG repeats (Fig. 3B). We expanded our analysis to a larger panel of cell lines and again observed that 7q/12q TERRA was, on average, 1.5 times higher in ALT cells compared to telomerase-positive and

primary lung fibroblasts, whereas the total UUAGGG signal was roughly 5.5 times higher in ALT cells compared to the others (Fig. 4A). We conclude that 7q/12q TERRA expression is not regulated identically to the majority of TERRA species.

Because TERRA 29 bp repeat promoters are silenced by CpG methylation mediated by DNMT1 and DNMT3b (Nergadze et al. 2009; Feretzaki et al. 2019), it is possible that CpG methylation is not substantially affecting 7q/12q TERRA expression. We performed northern blotting to analyze 7q/12q and total TERRA expression in human DKO cells and parental HCT116 cells. As expected, we detected a very faint signal for total TERRA in HCT116 parental cell lines when compared to the very strong signal detected for DKO cells (16%–20% of the DKO signal) (Fig. 4B). On the contrary, the 7q/12q TERRA signal was easily detected in HCT116 cells and its intensity was roughly half of the one in DKO cells (44%–49% of the DKO signal) (Fig. 4B). These data suggest that the 7q/12q promoters might be overall hypomethylated across different cell lines regardless of their telomere elongation mechanism, thus explaining why 7q/12q TERRA transcripts are more homogeneously expressed among cell lines and constitute a large fraction of total TERRA transcripts.

To test the hypothesis above, we first set up to isolate the 7q TERRA promoter. We generated a promoter reporter plasmid (p7q) by cloning a 695 bp fragment from the 7q subtelomere, overlapping a predicted CpG island of 333 bp (Fig. 2; Supplemental Fig. S6C), in front of an eGFP coding sequence. We transfected the plasmid into HeLa cells and analyzed eGFP expression by northern blotting and fluorescence microscopy. As a control, we transfected cells with a plasmid in which the eGFP cDNA is preceded by a 29 bp repeat promoter from the XqYq subtelomere (p3.5 plasmid) (Nergadze et al. 2009). We found that the selected 7q subtelomeric sequence was able to induce the expression of eGFP (Fig. 4C,D), confirming that we isolated a novel TERRA promoter sequence that does not contain 29 bp repeats. The p7q plasmid produced eGFP transcripts ~5× less abundant, and functional eGFP proteins ~2.4× less abundant than the ones produced from the p3.5 plasmid (Fig. 4C,D). This indicates that the 29 bp repeat promoters are stronger than the 7q promoter; however, they are likely to be silenced more efficiently in cells.

Having defined the 7q promoter, we performed bisulfite sequencing of that region using genomic DNA from U2OS, HeLa, HEK293T, HLF, HCT116, and DKO cells and primers flanking the identified CpG island (Supplemental Fig. S6C). Unfortunately, we were not able to isolate amplification products for the 7q subtelomere from HEK293T, HCT116, and DKO cells, possibly because of polymorphisms within the regions annealing with the chosen oligonucleotides. Nonetheless, we could analyze CpG methylation in the other three cell lines. 7q



**FIGURE 4.** Transcriptional regulation of 7q TERRA. (A) Total RNA from ALT, telomerase-positive (Telomerase+), or primary (Pr.) cells was analyzed by northern blotting using probes for the 7q/12q subtelomere or UUAGGG repeats (total TERRA). GAPDH and U6 RNAs were used as loading controls. (B) Total and nuclear RNA from HCT116 and DKO cells was analyzed by northern blotting as in A. U2OS RNA was included in the same blots for comparison with an ALT cell line. (C) Northern blot analysis of eGFP expression in HeLa cells transfected with the p7q promoter reporter. An empty vector plasmid devoid of promoter sequences (pEV) and the p3.5 plasmid containing a 29 bp repeat promoter were used as negative and positive controls, respectively. Twenty-four hours after transfection, cells were selected using puromycin and 48 h later harvested for RNA preparation. The bar graph shows the quantification of eGFP signals normalized through the corresponding U6 signals. Bars and error bars are the means and SDs from three independent experiments. *P* values (Student's *t*-test) are indicated. (D) Representative images of eGFP fluorescence in HeLa cells treated as in C. eGFP is shown in green, DAPI-stained DNA in blue. The bar graph shows the total eGFP fluorescence per cell. Bars and error bars are the means and SDs from two independent experiments. *P* values (Student's *t*-test) are indicated. (E) Heatmap representation of 7q CpG island methylation for the indicated cell lines. The CpG island coordinates are chr7:159,335,203–159,335,443 in the GRCh38/hg38 reference genome. The position of 7q TERRA TSS gauged from TERRA ONTseq is indicated.

methylation patterns from U2OS, HeLa, and HLF did not show any striking differences (Fig. 4D; Supplemental Fig. S6C), indicating that CpG methylation has only a secondary role in regulating 7q TERRA expression. On the other hand, it was previously reported that the 29 bp repeat promoters are more methylated in HLF and HeLa cells (low TERRA levels) than in U2OS cells (high TERRA levels), pointing to methylation as a primary mechanism silencing the expression of 29 bp repeat promoters (Nergadze et al. 2009).

## Conclusions

TERRA ONTseq allows users to simultaneously analyze TERRA transcripts from nearly all telomeres and provides robust and reproducible data. Although confirming that 29 bp repeat-like sequences are the main TERRA promot-

ers, other promoter regions devoid of such repeats, such as the one of the 7q subtelomere identified here, are present in humans and substantially contribute to TERRA cellular levels. Similarly, TERRA ONTseq should allow us to identify and validate all TERRA promoter sequences. TERRA ONTseq also showed that the most transcribed chromosome ends are consistent between cell lines, likely pointing to the existence of conserved regulatory elements maintained across different cell types. Given the qualitative and quantitative nature of TERRA ONTseq, we are convinced that it will become an extremely useful tool for a rapid and precise analysis of the human TERRA transcriptome in different laboratory and clinical settings. Further methodological improvements will broaden the reach of this tool—for example, by developing protocols to also sequence the entire UUAGGG tract within TERRA or by adapting the TERRA ONTseq pipeline to organisms other than humans.



## MATERIALS AND METHODS

### Oligonucleotides and plasmid construction

Oligonucleotides were purchased from Integrated DNA Technologies and their sequences can be found in Supplemental Table S1. To generate the promoter reporter plasmid p7q, a 695 bp fragment from the 7q subtelomere was PCR-amplified from HeLa genomic DNA using oligonucleotides #2384 and #2385 (Supplemental Table S1). The PCR product was cloned into the pJET1.2 plasmid using the CloneJET PCR Cloning Kit (Thermo Scientific), according to the manufacturer's instructions. The resulting plasmid was digested with XhoI and NcoI and the insert was cloned in front of an eGFP cDNA. The p3.5 plasmid was a kind gift from Elena Giulotto (University of Pavia, Italy).

### Cell lines and tissue culture procedures

HeLa cervical cancer, HT1080 fibrosarcoma, MG63 osteosarcoma, and HEK293T embryonic kidney cells were purchased from ATCC. CAL72 osteosarcoma cells were purchased from DSMZ. U2OS osteosarcoma cells were a kind gift from Massimo Lopes (IMCR, Zurich, Switzerland). HOS, HOS 143B, a derivative of the HOS cell line, and NOS1 cells were a kind gift from Bruno Fuchs (Balgrist University Hospital, Zurich, Switzerland). Sk-N-BE neuroblastoma cells were a kind gift from Olga Shakhova (University Hospital Zurich, Zurich, Switzerland). SAOS2 osteosarcoma, VA13 lung and GM847 skin SV40-transformed fibroblasts were a kind gift from Arturo Londoño-Vallejo (CNRS, Paris, France). HLF primary lung fibroblasts were a kind gift from Joachim Lingner (EPFL, Lausanne, Switzerland). HCT116 human colon carcinoma cells and DKO cells were a kind gift from Bert Vogelstein (Johns Hopkins University, Baltimore, USA). The LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich) was used to assure that all cell lines were mycoplasma-free. HeLa, HT1080, HEK293T, U2OS, HCT116, and DKO cells were cultured in high-glucose DMEM, GlutaMAX (Thermo Fisher Scientific) supplemented with 5% tetracycline-free fetal bovine serum (FBS; Pan BioTech) and 100 U/mL penicillin–streptomycin (Thermo Fisher Scientific). WI-38 VA13 and HLF cells were cultured in high-glucose DMEM, GlutaMAX supplemented with 10% tetracycline-free FBS and 100 U/mL penicillin–streptomycin. GM847 cells were cultured in high-glucose DMEM, GlutaMAX supplemented with 10% tetracycline-free FBS, 100 U/mL penicillin–streptomycin, and 1% nonessential amino acids (Thermo Fisher Scientific). NOS1 cells were cultured in high-glucose RPMI 1640, GlutaMAX (Thermo Fisher Scientific), 5% tetracycline-free FBS, 1% sodium pyruvate (Thermo Fisher Scientific), and 100 U/mL penicillin–streptomycin. SAOS2, HOS, CAL72, MG63, and Sk-N-BE cells were cultured in high-glucose DMEM/F12, GlutaMAX (Thermo Fisher Scientific), supplemented with 10% tetracycline-free FBS, 100 U/mL penicillin–streptomycin, and 1% nonessential amino acids; 143B cells were cultured in high-glucose DMEM/F12, GlutaMAX, supplemented with 10% tetracycline-free FBS and 100 U/mL penicillin–streptomycin. U2OS T-TALE cells were previously described (Silva et al. 2021) and were maintained in the same conditions as U2OS cells. For T-TALE expression, 100 ng/mL doxycycline (Sigma-Aldrich) was added to the culture medium for 24 h. For promoter reporter as-

says, HeLa cells were transfected using the Lipofectamine 2000 reagent (Invitrogen). Twenty-four hours after transfection, cells were incubated in medium containing 1 µg/mL puromycin (VWR International) and selected for 2 days. eGFP expressing cells were fixed with 4% formaldehyde (Sigma-Aldrich) in 1× PBS for 10 min, washed twice with 1× PBS, permeabilized with 1× PBS, 0.5% Triton X-100, and washed again with 1× PBS. DNA was counterstained with 100 ng/mL DAPI (Sigma) in 1× PBS and coverslips were mounted on slides in Vectashield (Vectorlabs). Images were acquired with an inverted widefield fluorescence microscope (ZEISS Axio Observer) and eGFP signal was quantified using ImageJ software. For TERRA half-life measurements, cells were incubated with 5 µg/mL actinomycin D (Sigma-Aldrich).

### RNA preparation and analysis

RNA from whole cells or nuclei-enriched cellular fractions was prepared using the TRIzol reagent (Invitrogen) as previously described (Azzalin et al. 2007). RNA was treated twice with RNase-free DNase I (New England Biolabs) for northern blotting analysis and TERRA purification or three times for RT-qPCR analysis. For RNaseH experiments, 10 µg of nuclear RNA or 15 µg of total RNA were mixed with 625 nM GAPDH oligonucleotide (#134; Supplemental Table S1) and 625 nM telomeric oligonucleotide (#18; Supplemental Table S1) or 7q<sub>12q</sub> oligonucleotide (#2336; Supplemental Table S1). The mix was incubated at 65°C for 4 min and then at room temperature for 20 min. Ten units of RNaseH (New England Biolabs) were added to the mix and allowed to digest at 37°C for 1 h. For northern blots, 10–15 µg of RNA were electrophoresed in 1% formaldehyde agarose gels and blotted onto nylon membranes. Membranes were then hybridized for ~18 h in Church buffer containing <sup>32</sup>P-labeled probes at 50°C–64°C. The strand-specific telomeric probe used to detect total TERRA was described previously (Azzalin et al. 2007). The 7q/12q probe was obtained by performing PCR on the plasmid p7q with the oligonucleotides #2384 and #1438 (Supplemental Table S1). The PCR product was strand-specifically radiolabeled by primer extension using the #2384 oligonucleotide. The eGFP probe was a DNA fragment obtained by digestion of the p3.5 plasmid with BamHI and XbaI and labeled by primer extension using the oligonucleotide #123 (Supplemental Table S1). The probes detecting β-actin, GAPDH, and U6 RNAs were 5' end-labeled DNA oligonucleotides (#5, #134, and #217, respectively; Supplemental Table S1). After hybridization, membranes were washed in 0.2×–2× SSC, 0.2% SDS at the same temperatures used for hybridizations. Signals were detected using a Typhoon IP imager (Amersham) and quantified using ImageJ software. For quantitative RT-PCR experiments, 5 µg of total RNA or 1.5 µg of nuclear RNA were reverse transcribed with 0.5 µM TeloR and 0.05 µM U6 oligonucleotides (#18 and #217, respectively; Supplemental Table S1) using the SUPERScript IV reverse transcriptase (Invitrogen). Quantitative PCRs were performed with specific oligonucleotide pairs as indicated in Supplemental Table S1 using the iTaq Universal SYBR Green Supermix (Bio-Rad) and a Rotor-Gene 6000 instrument (Corbett). Absolute TERRA quantifications were performed as previously described (Feretaki et al. 2019) using plasmids containing subtelomeric fragments from specific chromosome ends to generate standard curves.

## DNA methylation analysis

Genomic DNA was extracted using the Promega Wizard Genomic DNA Purification Kit according to the manufacturer's instructions. Bisulfite treatment was performed on 500 ng of DNA using the EZ DNA Methylation-Lightning Kit (Zymo Research). Bisulfite-converted DNA was amplified using the oligonucleotides #2479 and #2480 (Supplemental Table S1) and the Q5U Hot Start High-Fidelity DNA Polymerase (New England Biolabs). PCR products were sequenced at Eurofins Scientific using the NGSelect Amplicon option (two-step amplicon generation workflow).

## Western blotting

Cells pellets were resuspended in lysis buffer (4% SDS, 20% Glycerol, 120 mM Tris-HCl pH 6.8), boiled at 95°C for 5 min, and centrifuged at 1600g at 4°C for 10 min. Supernatants were recovered and protein concentrations were determined using a NanoDrop 2000 (Thermo Fisher Scientific). Thirty-five micrograms of proteins were supplemented with 0.004% Bromophenol blue and 1%  $\beta$ -mercaptoethanol (Sigma-Aldrich), incubated at 95°C for 5 min, separated in a 10% polyacrylamide gel, and transferred to a nitrocellulose membrane (Maine Manufacturing, LLC) using a Trans-Blot SD Semi-Dry Transfer Cell apparatus (Bio-Rad). Primary antibodies were a rabbit monoclonal anti-HA (Cell Signaling 3724; 1:1000 dilution) and a mouse monoclonal anti- $\beta$ -actin (Abcam ab8224; 1:5000 dilution). Secondary antibodies were HRP-conjugated goat anti-mouse and goat anti-rabbit IgGs (Bethyl Laboratories A90-116P and A120-101P; 1:2000 dilution). Signal detection was performed using the ECL detection reagents (GE HealthCare) and an Amersham 680 blot and gel Imager.

## TERRA purification and ONT sequencing

For each TERRA purification reaction, 25  $\mu$ g of nuclear RNA were incubated at 75°C for 4 min in 0.5 $\times$  TE (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA). Twenty-five pmol of biotinylated, 3' dideoxycytosine (ddC)-blocked telomeric oligonucleotides (#1; Supplemental Table S1) were added to the RNA and the mixture incubated at 50°C for 50 min in 0.5 $\times$  TE supplemented with 1 M of NaCl. Fifty micrograms of streptavidin magnetic beads (New England Biolabs) were added and the mixture was incubated in a thermomixer at 50°C with 300 rpm agitation for 3 h. Beads were washed once in 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 3 times in 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.5 M NaCl, and twice in 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.15 M NaCl. RNA was eluted in 10 mM Tris-HCl pH 8.0 by heating the mixture for 3 min at 80°C. RNA was then treated with DNase I at 37°C for 30 min and precipitated in the presence of glycogen. TERRA sequencing was performed using the Direct cDNA Sequencing Kit from Oxford Nanopore Technologies (SQK-DCS109) according to the manufacturer's instructions with minor modifications during the cDNA synthesis. Specifically, 400–800 ng of purified TERRA, 10 nM of VNP-TERRA oligonucleotide (#2161; Supplemental Table S1), and 10  $\mu$ M of dNTPs were mixed together and heated to 85°C for 3 min. The cDNA synthesis reaction was performed with SUPERSRIPT IV reverse transcriptase (Invitrogen) at 55°C for 70 min followed by 20 min at 42°C. HeLa, HEK293T, and

T-TALE cDNAs were barcoded using the Native Barcoding Expansion kit from Oxford Nanopore Technologies (EXP-NBD104). Sequencing was performed using the MiniON device with the flow cells MK 1 R9 version (Oxford Nanopore Technologies). For U2OS, a total of 225  $\mu$ g of nuclear RNA was used for two sequencing runs; for HeLa, 727  $\mu$ g of nuclear RNA was used for six sequencing runs (three of which barcoded); for HEK293T 1 mg of nuclear RNA was used for seven sequencing runs (four of which barcoded); for T-TALES, 225  $\mu$ g of nuclear RNA was used for three sequencing barcoded runs.

## Nanopore sequence analysis

Nanopore reads were debarcoded using the Porechop tool (<https://github.com/rwick/Porechop>). The alignment to the T2T CHM13v2.0/hs1 genome reference or a derived subtelomeric reference was performed using the minimap2 aligner (Li 2018). The options used for the alignment step were `-k15 -w6` (minimizers on the query), `-ax map-ont` (for Oxford nanopore reads). The alignment results were statistically analyzed through the functions `stats` and `idxstats` from SAMtools (Danecsek et al. 2021). The mapping quality is given by the MAPQ value, which describes the uniqueness of the alignment. MAPQ = 0 indicates the reads that can possibly map to many locations and have the lowest quality, whereas MAPQ > 0 are primary reads, or uniquely mapped. The coverage of the aligned reads was calculated using the function `genomeCoverageBed` from BEDTools suite (Quinlan and Hall 2010). Multimapped reads were distributed equally along the reference sequence. Custom scripts in R language were set up for coverage visualization, for comparison among the different cell lines, and for the position of 29 and 20 bp repeats. Metaprofiles were generated with a custom R script from the global coverage of the uniquely mapped reads normalized through the total number of reads for each cell line.

## DNA and bisulfite sequencing analysis

7q and 12q subtelomere sequences were aligned using the EMBOSS Needle tool at the EMBL's European Bioinformatics Institute ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)). CpG dinucleotide contents were analyzed using the CpGPlot software ([https://www.ebi.ac.uk/Tools/seqstats/emboss\\_cpplot/](https://www.ebi.ac.uk/Tools/seqstats/emboss_cpplot/)). The distribution of 29 bp repeat-like sequences was determined using the simple repeat annotation and BLAT at UCSC Genome Browser (<https://genome.ucsc.edu/>). For methylation analysis, the forward and reverse reads from bisulfite sequencing were merged using the `fastq-join` function (<https://manpages.ubuntu.com/manpages/bionic/man1/fastq-join.1.html#name>) and the `filter-fastq` function (<https://github.com/Floor-Lab/filter-fastq>) was used to select reads shorter than 400 bp to perform the mapping to 7q and reads longer than 400 bp to 12q. Read mapping was performed using the software `bwameth` (<https://github.com/brentp/bwa-meth>) using a FASTA reference containing the 7q and 12q subtelomeric sequences. The computation of methylation metrics from alignment files was performed with `MethylDackel` software (<https://github.com/dpryan79/MethylDackel>). A custom script in R language was set up to generate methylation state heatmaps.

## Statistical analysis

For direct comparison of two groups, a paired two-tailed Student's *t*-test was performed using Microsoft Excel. *P* values are shown in the figures.

## DATA DEPOSITION

TERRA ONTseq data were deposited at the Sequence Read Archive (SRA) with the code PRJNA997813. Sanger and bisulfite sequencing data can be provided upon request.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## COMPETING INTEREST STATEMENT

C.M.A. is a cofounder and shareholder of TessellateBIO.

## ACKNOWLEDGMENTS

We thank Miguel Alves for help with isolating and sequencing the 7q subtelomere, Filipe Cadete for help with the ONTseq analysis pipeline, and Patricia Abreu for critical reading of this manuscript. We thank the Bioimaging and Information System units of iMM for services. We thank Elena Giulotto, Massimo Lopes, Bruno Fuchs, Olga Shakhova, Arturo Londoño-Vallejo, Joachim Lingner, and Bert Vogelstein for sharing the reagents. This work was supported by LaCaixa Foundation (project LCF/PR/HP21/52310016 to C.M.A.), national funds through FCT (Fundação para a Ciência e a Tecnologia), I.P. (projects PTDC/BIA-MOL/6624/2020 and 2021.00143.CEECIND to C.M.A.), TessellateBIO (to C.M.A.) and CNR (Consiglio Nazionale delle Ricerche) (project DBA.AD005.225—NUTRAGE—FOE2021 to S.B.).

*Author contributions:* J.R. and C.M.A. conceptualized the study and designed the experiments. J.R. performed the experiments. R.A. and S.B. performed all bioinformatic analyses. J.R. and C.M.A. wrote the manuscript and all authors reviewed and edited the manuscript. C.M.A. and S.B. acquired funding and provided supervision.

Received November 30, 2023; accepted May 5, 2024.

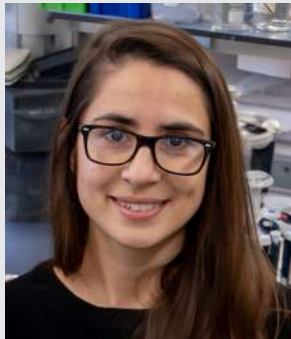
## REFERENCES

- Arnoult N, Van Beneden A, Decottignies A. 2012. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1 $\alpha$ . *Nat Struct Mol Biol* **19**: 948–956. doi:10.1038/nsmb.2364
- Arora R, Lee Y, Wischniewski H, Brun CM, Schwarz T, Azzalin CM. 2014. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun* **5**: 5220. doi:10.1038/ncomms6220
- Azzalin CM, Lingner J. 2008. Telomeres: the silence is broken. *Cell Cycle* **7**: 1161–1165. doi:10.4161/cc.7.9.5836
- Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J. 2007. Telomeric repeat containing RNA and RNA surveillance factors

- at mammalian chromosome ends. *Science* **318**: 798–801. doi:10.1126/science.1147182
- Baird DM, Coleman J, Rosser ZH, Royle NJ. 2000. High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. *Am J Hum Genet* **66**: 235–250. doi:10.1086/302721
- Bhargava R, Lynskey ML, O'Sullivan RJ. 2022. New twists to the ALTerNative endings at telomeres. *DNA Repair (Amst)* **115**: 103342. doi:10.1016/j.dnarep.2022.103342
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: 1–4. doi:10.1093/gigascience/giab008
- Deng Z, Wang Z, Stong N, Plasschaert R, Moczan A, Chen HS, Hu S, Wikramasinghe P, Davuluri RV, Bartolomei MS, et al. 2012. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J* **31**: 4165–4178. doi:10.1038/emboj.2012.266
- Episkopou H, Draskovic I, Van Beneden A, Tilman G, Mattiussi M, Gobin M, Arnoult N, Londoño-Vallejo A, Decottignies A. 2014. Alternative lengthening of telomeres is characterized by reduced compaction of telomeric chromatin. *Nucleic Acids Res* **42**: 4391–4405. doi:10.1093/nar/gku114
- Famung BO, Brun CM, Arora R, Lorenzi LE, Azzalin CM. 2012. Telomerase efficiently elongates highly transcribing telomeres in human cancer cells. *PLoS ONE* **7**: e35714. doi:10.1371/journal.pone.0035714
- Feretzi M, Renck Nunes P, Lingner J. 2019. Expression and differential regulation of human TERRA at several chromosome ends. *RNA* **25**: 1470–1480. doi:10.1261/ma.072322.119
- Kwapisz M, Morillon A. 2020. Subtelomeric transcription and its regulation. *J Mol Biol* **432**: 4199–4219. doi:10.1016/j.jmb.2020.01.026
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, De S, Petrini JH, Sung PA, Jasin M, et al. 2012. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet* **8**: e1002772. doi:10.1371/journal.pgen.1002772
- Montero JJ, López de Silanes I, Graña O, Blasco MA. 2016. Telomeric RNAs are essential to maintain telomeres. *Nat Commun* **7**: 12534. doi:10.1038/ncomms12534
- Nergadze SG, Famung BO, Wischniewski H, Khoriauli L, Vitelli V, Chawla R, Giulotto E, Azzalin CM. 2009. CpG-island promoters drive transcription of human telomeres. *RNA* **15**: 2186–2194. doi:10.1261/ma.1748309
- Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. 2022. The complete sequence of a human genome. *Science* **376**: 44–53. doi:10.1126/science.abj6987
- Porro A, Feuerhahn S, Reichenbach P, Lingner J. 2010. Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol Cell Biol* **30**: 4808–4817. doi:10.1128/MCB.00460-10
- Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. 2014. Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* **5**: 5379. doi:10.1038/ncomms6379
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, Hook PW, Koren S, Rautiainen M, Alexandrov IA, et al. 2023. The complete sequence of a human Y chromosome. *Nature* **621**: 344–354. doi:10.1038/s41586-023-06457-y

- Rivosecchi J, Cusanelli E. 2023. TERRA beyond cancer: the biology of telomeric repeat-containing RNAs in somatic and germ cells. *Front Aging* **4**: 1224225. doi:10.3389/fragi.2023.1224225
- Savoca V, Rivosecchi J, Gaiatto A, Rossi A, Mosca R, Gialdini I, Zubovic L, Tebaldi T, Macchi P, Cusanelli E. 2023. TERRA stability is regulated by RALY and polyadenylation in a telomere-specific manner. *Cell Rep* **42**: 112406. doi:10.1016/j.celrep.2023.112406
- Schoeftner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**: 228–236. doi:10.1038/ncb1685
- Silva B, Arora R, Bione S, Azzalin CM. 2021. TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nat Commun* **12**: 3760. doi:10.1038/s41467-021-24097-6
- Silva B, Arora R, Azzalin CM. 2022. The alternative lengthening of telomeres mechanism jeopardizes telomere integrity if not properly restricted. *Proc Natl Acad Sci* **119**: e2208669119. doi:10.1073/pnas.2208669119
- Srinivas N, Rachakonda S, Kumar R. 2020. Telomeres and telomere length: a general overview. *Cancers (Basel)* **12**: 558. doi:10.3390/cancers12030558
- Warburton PE, Sebra RP. 2023. Long-read DNA sequencing: recent advances and remaining challenges. *Annu Rev Genomics Hum Genet* **24**: 109–132. doi:10.1146/annurev-genom-101722-103045
- Zeinoun B, Teixeira MT, Barascu A. 2023. TERRA and telomere maintenance in the yeast *Saccharomyces cerevisiae*. *Genes (Basel)* **14**: 618. doi:10.3390/genes14030618

## MEET THE FIRST AUTHOR



Joana Rodrigues

**Meet the First Author(s)** is an editorial feature within *RNA*, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of *RNA* and the RNA research community. Joana Rodrigues is the first author of this paper, “TERRA ONTseq: a long-read-based sequencing pipeline to study the human telomeric transcriptome.” Joana is a postdoctoral researcher in Claus Azzalin’s laboratory at the Institute of Molecular Medicine in Lisbon, Portugal.

### What are the major results described in your paper and how do they impact this branch of the field?

In 2007, Claus Azzalin and Joachim Lingner made the groundbreaking discovery that telomeres are transcribed into a low-abundance, repetitive, long noncoding RNA known as TERRA. In our study, we used long-read sequencing to comprehensively analyze TERRA molecules from all telomeres, both qualitatively and quantitatively. Our findings enabled us to identify TERRA transcription start sites for most subtelomeres as well as a previously uncharacterized TERRA promoter, present at telomere 7q, which is responsible for the production of a significant fraction of TERRA molecules present in the cell.

### What led you to study RNA or this aspect of RNA science?

The central role of telomeres in both cancer and aging sparked my interest when I was still an undergraduate. In school, I was taught

that telomeres were transcriptionally silent. However, the revelation that telomeres are transcribed into an RNA, which plays crucial roles in various aspects of cell biology, has kept my curiosity and interest piqued ever since I stumbled upon it while studying the consequences of deleting the PAF1 complex in budding yeast.

### During the course of these experiments, were there any surprising results or particular difficulties that altered your thinking and subsequent focus?

There are still many open questions about TERRA biology that can only be answered with long-read sequencing, such as length of the repetitive tract, polyadenylation status, etc. When we started this project, we wanted to answer all those questions simultaneously. However, because of the limited abundance of TERRA molecules, we prioritized answering fundamental questions first: identifying which telomeres are transcribed and their transcription start sites. With this groundwork laid, we are now better equipped to tackle the more intricate aspects of TERRA biology.

### What are some of the landmark moments that provoked your interest in science or your development as a scientist?

The moment that ignited my passion for science was when I realized that everything we learn in biology or genetics class was initially discovered by someone in a lab. The idea of being the first to uncover new knowledge is incredibly exciting. There is nothing quite like the thrill of seeing the data that proves your hypothesis right.

### Are there specific individuals or groups who have influenced your philosophy or approach to science?

Throughout my career, I have been fortunate to encounter numerous individuals who have influenced my scientific journey. Among them, David Lydall and Claus Azzalin stand out as pivotal figures. They both incited in me the love for telomeres and the telomere community as well as the importance of independence, rigor, and critical thinking. I am particularly grateful for the opportunity to share Claus’ positive mindset and unwavering motivation. His resilience and optimism have been a constant source of inspiration in my scientific journey.





# RNA

A PUBLICATION OF THE RNA SOCIETY

## TERRA ONTseq: a long-read-based sequencing pipeline to study the human telomeric transcriptome

Joana Rodrigues, Roberta Alfieri, Silvia Bione, et al.

*RNA* 2024 30: 955-966 originally published online May 22, 2024  
Access the most recent version at doi:[10.1261/ma.079906.123](https://doi.org/10.1261/ma.079906.123)

---

**Supplemental Material** <http://rnajournal.cshlp.org/content/suppl/2024/05/22/rna.079906.123.DC1>

**References** This article cites 29 articles, 5 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/30/8/955.full.html#ref-list-1>

**Open Access** Freely available online through the *RNA* Open Access option.

**Creative Commons License** This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>