TERRA transcripts and promoters from telomeric and interstitial sites

Marco Santagostino*[1], Lorenzo Sola*[1], Eleonora Cappelletti*[1], Nicolò Gennari[1], Francesca M. Piras[2], Marialaura Biundo[1], Solomon G. Nergadze[1], Elena Giulotto[#1]

[1] Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, Via Adolfo Ferrata 9, 27100 Pavia, Italy

[2] Unit of Anatomic Pathology, IRCCS San Matteo Hospital Foundation, Viale Camillo Golgi 19, 27100 Pavia, Italy

* Equal contribution

[#] Corresponding author

Short title: Telomere and ITS transcription

Key words: telomere, interstitial telomeres, transcription, TERRA, promoter, human genome

ABSTRACT

The transcription of human telomeres gives rise to a family of long noncoding RNAs, named TERRA. We previously showed that TERRA transcription is driven by CpG island promoters that are composed by stretches of three types of repeats. Using the human genome assembly that was available at that time, putative promoter sequences were localized at several subtelomeres. In this work, using the T2T-CHM13v2.0 human reference genome, we found that 39 out of 46 subtelomeres contain TERRA promoters and grouped them in classes depending on their organization. The sequence of subtelomeric promoters was variable in different reference genomes suggesting that they are polymorphic in the human population. We then discovered 106 intrachromosomal TERRA-like promoters, adjacent to interstitial telomeric sequences (ITSs) or far away from them. Fortyseven of these promoters are flanked and may regulate the transcription of coding genes, ncRNAs or pseudogenes. Comparative sequence analysis showed that interstitial and subtelomeric promoters belong to a previously undescribed family of segmental duplications deriving from common ancestral sequences. RT-PCR experiments in seven cell lines demonstrated that TERRA transcripts can be synthesized from ITSs. TERRA expression was always low in primary fibroblasts and HeLa cells while highly variable in the other two telomerase positive (HT1080 and HEK293) and in the three telomerase negative ALT cell lines (GM847, U2OS and VA13). Given the large number of ITSs in the human genome, we propose that these loci contribute significantly to the production of the TERRA pool.

INTRODUCTION

Telomeres, the terminal regions of eukaryotic linear chromosomes, are chromatin structures essential for chromosome integrity. In vertebrates, the telomeric DNA is composed of extended tandem arrays of the hexanucleotide TTAGGG and is associated with the protein complex called Shelterin and with other accessory proteins (de Lange 2018). Maintenance of telomere length and protection of chromosome ends from damage is ensured by the specialized enzyme telomerase (Shay and Wright 2019), a nucleoprotein composed by a reverse transcriptase (TERT) and an RNA moiety (TERC) containing the template for telomere synthesis. The telomerase enzyme is active in the germline, in stem cells and in most tumors while, in a fraction of human cancers, a telomerase independent mechanism of telomere maintenance (Alternative Lengthening of Telomeres or ALT) is present (Azzalin 2025).

Telomeres are transcribed by RNA polymerase II into a family of long non-coding RNAs called Telomeric Repeat-containing RNA (TERRA) that, in human cells, are highly heterogeneous in length, ranging from 100 to 9000 bp (Azzalin et al. 2007). Telomeric transcripts have been identified also in mouse (Schoeftner and Blasco 2008), fish (Park et al. 2019), plants (Vrbsky et al. 2010), nematodes (Manzato et al. 2023), protozoa (Rudenko and Van der Ploeg 1989) and yeast (Luke et al. 2008). Thus, telomere transcription can be considered a conserved feature in eukaryotes.

In a previous work, we found that, in humans, TERRA is transcribed from CpG island promoters that are embedded within the subtelomeric sequence immediately upstream the telomeric repeats (Nergadze et al. 2009). The pro-terminal subtelomeric regions from the long-arm of human chromosomes X/Y and 10 were isolated and named TelBam3.4 and TelSau2.0, respectively (Brown et al. 1990). In that early study, it was shown that these regions comprise three minisatellites composed of stretches of 61, 29 and 37 bp repeats in centromere to telomere orientation. The 61 bp and the 29 bp repeats are separated by a sequence related to the 29 bp repeat that lies outside the array and was defined as 29like. Using a gene reporter assay, we then demonstrated that these sequences function as TERRA promoters and that the 29 bp repeat is sufficient to drive transcription (Nergadze et al. 2009). Through sequence analysis, using the GRCh37/hg19 reference genome, we identified TERRA promoters at 14 subtelomeres (1p, 2p, 3q, 4p, 5p, 8p, 9p, 11p, 15q, 16p, 17p, 19p, 21q, Xq) and on both sides of the telomere fusion at 2q14 (Nergadze et al. 2009). In the same work, we showed that the telomeric tract and the last repeat of the 37 bp unit are separated by a highly conserved

sequence named pre-telomeric region or pre-tel. Through a 5'-RACE protocol, we identified a TERRA transcription start site (TSS), located at about 250 bp from the telomeric repeats. The regulation of TERRA transcription involves methylation of the CpG island (Nergadze et al. 2009; Le Berre et al. 2019), interaction with DNA binding proteins (Deng et al. 2012; Feretzaki et al. 2019; Vohhodina et al. 2021) and histone 3/lysine 4 methylation (Caslini et al. 2009). Telomere transcription can increase in response to alterations of the metabolism (Diman et al. 2016) and to radiation-induced DNA damage (Al-Turki et al. 2024).

The results of RT-qPCR experiments demonstrated that TERRA is transcribed from several chromosome ends and that its expression varies not only between different cell lines but also between different telomeres in the same cell line (Vitelli et al. 2013; Episkopou et al. 2014; Feretzaki et al. 2019; Silva et al. 2021; Savoca et al. 2023). A comparison of subclones isolated from four telomerase positive and one ALT cell line revealed that TERRA levels from specific subtelomeres are variable (Smirnova et al. 2013).

TERRA is a key player in telomere function and regulates several aspects of telomere biology, including chromatin structure (Deng et al. 2009; Arnoult et al. 2012), telomerase activity (Redon et al. 2010; Cusanelli et al. 2013) recruitment of telomeric binding proteins (Deng et al. 2009; López de Silanes et al. 2010; Scheibe et al. 2013; Abreu et al. 2022), telomere replication (Maicher et al. 2012; Beishline et al. 2017) and activation of the DNA damage response (Porro et al. 2014; Tutton et al. 2016; Graf et al. 2017). The observation that deregulation of TERRA synthesis causes genomic instability and telomere loss suggested that inappropriate telomere transcription may be involved in tumorigenesis (Azzalin et al. 2007). Indeed, low levels of TERRA expression in tumors has been associated to tumor aggressiveness (Schoeftner and Blasco 2008; Deng et al. 2012; Sampl et al. 2012; Kienzl et al. 2024). A comparison between TERRA levels in the tumor and matched normal tissue from head and neck squamous cell carcinoma patients showed that lower TERRA expression in the tumor was associated with worse prognosis (Vitelli et al. 2018). Interestingly, TERRA binds not only chromosome ends but also extratelomeric genomic locations (Chu et al. 2017; Avogaro et al. 2018) and might carry out functions unrelated to telomere maintenance (Reig-Viader et al. 2014; Wang et al. 2015; Wang and Lieberman 2016; Marión et al. 2019; Nassour et al. 2023).

Besides chromosome ends, stretches of TTAGGG repeats are present also at internal genomic sites, where they constitute the so called interstitial telomeric sequences or ITSs. ITSs have been described in

many species (Meyne et al. 1990), including humans (Azzalin et al. 1997, 2001) and other primates (Ruiz-Herrera et al. 2002; Nergadze et al. 2004; Mazzoleni et al. 2017), rodents (Faravelli et al. 2002; Nergadze et al. 2007), equids (Santagostino et al. 2020) and birds (Nanda et al. 2002).

According to chromosomal localization, sequence organization and mechanism of origin, ITSs can be classified in heterochromatic, fusion and short ITSs (Ruiz-Herrera et al. 2008). Heterochromatic ITSs, extended blocks of centromeric and pericentromeric repeats, have been described in several species (Faravelli et al. 2002; Colomina et al. 2017; Clemente et al. 2020) but not in the human genome where one fusion (IJdo et al. 1991) and numerous short ITSs are present (Nergadze et al. 2007). Although short ITSs can be considered a particular type of microsatellite, they were not generated by the expansion of pre-existing units through DNA polymerase slippage (Messier et al. 1996) but were inserted in one step during evolution preferentially at sites where microhomologies with the telomeric repeat insertion were present in the ancestral DNA sequence (Nergadze et al. 2004, 2007). These observations, along with the identification of a subclass of ITS loci flanked by a sequence retrotranscribed from the Telomerase RNA component (TERC), suggested that the insertion of ITSs into the genome occurred, in the course of evolution, during the repair of DNA double strand breaks through peculiar mechanisms involving the telomerase enzyme (Nergadze et al. 2004, 2007; Santagostino et al. 2020; Sola et al. 2021). ITS loci are characterized by two types of polymorphism: insertion and copy number variation. To our knowledge, insertion polymorphism was never detected in the human population (Santagostino et al. 2020) but in some species, such as the horse, this is an important source of genetic variation. Similar to other microsatellites, copy number variation of TTAGGG repeats, was detected at several loci in all species analyzed (Mondello et al. 2000; Aksenova and Mirkin 2019; Santagostino et al. 2020).

In the present work we characterized the organization of 39 putative subtelomeric TERRA promoters and of a family of TERRA promoter-like sequences that are positioned at interstitial sites. We then demonstrated that TERRA molecules synthesized from ITSs constitute an important component of the TERRA pool.

RESULTS AND DISCUSSION

Organization of subtelomeric TERRA promoters in the T2T-CHM13v2.0 reference genome

Taking advantage of the T2T-CHM13v2.0 genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/) (Nurk et al. 2022) we updated the analysis of human subtelomeres, completed the list of TERRA promoters and characterized their sequence organization. Using UCSC Genome Browser (https://genome.ucsc.edu), we downloaded the 10 kb sequence immediately adjacent to each telomere. We then used the 1311 bp sequence of the previously described Xq TERRA promoter (Nergadze et al. 2009) (T2T-CHM13v2.0, chrX:154255311-154256622) as query on each subtelomere. Sequences were manually analyzed using MultAlin to identify 61bp, 29like, 29bp, 37bp units and the pre-tel sequence. The subtelomeres of the Y chromosome were not included in the analysis because the cell line used to generate the original T2T-CHM13v2.0 genome derived from a complete hydatidiform mole with a 46XX karyotype (Nurk et al. 2022). Supplemental Table S1 reports the genomic coordinates and sequence organization of all subtelomeres, including the ones previously described.

With this analysis we confirmed the presence of TERRA promoters at the 14 previously described subtelomeres (Nergadze et al. 2009) and identified 25 new ones. Altogether, these results suggest that, at 39 subtelomeres, TERRA promoters may drive telomere transcription.

We observed seven different sequence organizations of subtelomeric TERRA promoters (Fig. 1). Figure 1A shows a sketch of the 25 TERRA promoters characterized by the previously described sequence organization (Brown et al. 1990; Nergadze et al. 2009) with the 61-29like-29-37 repeat tracts and the pre-tel region immediately adjacent to the telomere. Fifteen of these sequences were identified in the present work (1q, 4q, 5q, 6p, 7p, 9q, 11q, 12p, 13q, 15p, 16q, 17q, 20p, 20q, 22q). At two loci (5p, 6q) we detected an additional element between the pre-tel sequence and the telomeric repeats, which we named spacer sequence (Fig. 1B). At six loci (2q, 3q, 11p, 14q, 18q and 19q) the 61-29like-29-37 repeat tracts were present while the pre-tel region was absent (Fig. 1C). At two subtelomeres (7q, 12q) the 61-29like tracts were absent and the promoter sequence comprised only the 29 and 37 repeats, the pre-tel region and the spacer sequence (Fig. 1D). In addition to these categories, three other classes were identified (Fig. 1E-G). The 29 bp repeat, which was previously shown by us to possess promoter activity in a reporter gene assay in HeLa cells (Nergadze et

al. 2009), is conserved in all TERRA promoters. In a similar *in vitro* assay it was shown that the 61 bp element, which contains CTCF binding sites, favors the recruitment of RNAPII at telomeres (Deng et al. 2012). Although no data were obtained so far on the other elements found at TERRA promoters (29like, 37, pre-tel and spacer), they may function as regulators of TERRA transcription *in vivo* possibly contributing to differential expression of subtelomeres.

At seven subtelomeres (8q, 13p, 14p, 17p, 21p, 22p, Xp) the sequence adjacent to the telomere did not show homology to TERRA promoters (Supplemental Table S1). At these loci, the analysis of the 10 kb preceding the telomere, with a combination of different promoter predicting software (EMBOSS Cpgplot, YAPP and Promoter 2.0), did not reveal any putative promoter element with the exception of CpG islands located 4136 bp and 64 bp away from the telomeric repeats in the Xp and 14p, respectively. Sequence comparison of these subtelomeric regions extending for 5 kb from the telomere revealed that the 21p and 22p share 99% identity suggesting a common origin. The remaining five promoter-less subtelomeres did not seem to derive from a common ancestral sequence. In conclusion, in the T2T-CHM13v2.0 assembly, a TERRA promoter was identified at 39 subtelomeres. At seven subtelomeres, no TERRA promoters were found suggesting that they may not be the only sequences driving telomere transcription since, as described in a following paragraph, the Xp telomere, which is not adjacent to a TERRA promoter, is transcribed in the cell lines tested.

We compared the sequences of the subtelomeres in the T2T-CHM13v2.0 and in the GRCh38 reference genomes. In Supplemental Table S1, the results of the analysis of two genomes are reported.

Twentyfive out of the 39 TERRA promoters were identified in both genome assemblies. At 5 of these informative promoters (1p, 2q, 7p, 10p and 12p) the sequence organization varied in the two assemblies. A peculiar situation was observed at the 7p subtelomere where the promoter identified in the T2T-CHM13v2.0 genome corresponded, in the GRCh38 reference, to a sequence unrelated to TERRA promoters. Interestingly, the length of the 29 bp repeat stretch, that is the minimum promoter (Nergadze et al. 2009), varied at different subtelomeres being comprised, in the T2T-CHM13v2.0 genome, between 25 and 1302 bp and corresponding to a number of repeat units comprised between about 1 and 45. The length of the other repeats varied at different loci but, at each locus, tended to be conserved in the two genomes. We cannot exclude that some of the differences between the two genomes are due to mis-assembly of the

GRCh38 genome. Overall, these results suggest that the sequence of at least some TERRA promoters is polymorphic in the human population. We do not know whether this sequence variability causes TERRA expression variation among individuals, however, it may affect the efficiency of primer pairs used in RT-PCR experiments and should be taken into account when TERRA expression levels from different loci and cell lines are compared. While our search of the entire TERRA promoter allowed us to identify 39 subtelomeric loci, the search carried out by Rodrigues and colleagues (Rodrigues et al. 2024), based on a custom script to detect the 29 bp repeat only, revealed 36 such loci.

Interstitial Telomeric Sequences (ITSs) in the T2T-CHM13v2.0 reference genome

In our previous work (Nergadze et al. 2007; Santagostino et al. 2020), we described Interstitial Telomeric Sequences in the human genome assemblies that were available at that time. Through comparative analysis with orthologous loci in non-human primates, we analyzed their origin and proposed molecular mechanisms for their insertion into genomes during DNA repair events that occurred in the course of evolution (Nergadze et al. 2004, 2007; Santagostino et al. 2020). In the present work, we updated the list of human ITSs performing a BLAST search using as query the $(TTAGGG)_4$ sequence against T2T-CHM13v2.0. The results are reported in Supplemental Table S2 where, for each locus, the localization, length and number of mismatches per repeat is reported. While in our previous work (Santagostino et al. 2020) ITS loci were searched using as query the telomeric hexamer repeated four times with a number of mismatch/repeat ≤1, in the present work we used less stringent conditions considering as ITSs telomeric repeat loci with a minimum length of 16 bp and a number of mismatch/repeat ≤2. With this approach also degenerated loci were found. We identified 522 loci ranging in length from 16 to 1840 bp. At the majority of the ITS loci (421) the conservation of the telomeric repeats is relatively high (less than 1.4 mismatch per repeat). As previously described (IJdo et al. 1991), one of these loci, that is composed by two head-to-head stretches (HSA2q14, chr2:114027333-114028210), derives from the telomere-telomere fusion of two ancestral acrocentric chromosomes that generated human chromosome 2 (number 57 in Supplemental Table S2).

Intrachromosomal TERRA promoters

During the search of ITS loci we discovered that 66 of them were adjacent to sequences homologous to subtelomeric TERRA promoters. At the 2q14 fusion ITS, two promoters were localized at both sites of the fusion point. Therefore, 67 TERRA-like promoters, adjacent to ITSs, were localized at intrachromosomal sites. We then detected 39 additional loci homologous to the TERRA promoter but not adjacent to telomeric repeats. Altogether, 106 interstitial TERRA promoters were found in the T2T-CHM13v2.0 genome. In Figure 1, the sequence organization (N-S) of all interstitial TERRA promoters is sketched. In Supplemental Table S3, the coordinates and the number assigned to each locus are reported.

Figure 1H shows the organization of three interstitial promoters containing the 61-29like-29-37 bp repeats and the pre-tel fragment. This is the same organization found in the majority of subtelomeres (Fig. 1A). Figures 1J and I show the organization of two promoters flanking the two ITSs, at chromosome 2q14, oriented head-to-head and deriving from the fusion of two ancestral telomeres (Ijdo et al 1991). One of these promoters is composed by the 61-29-37 bp units and the pre-tel while the other one shows the same organization of the subtelomeres shown in Figure 1C. In 55 interstitial promoters flanking an ITS (Fig. 1K) the 61 bp repeat is missing. Finally, four loci (Fig. 1M) do not contain the 29 bp core promoter.

Although at most (34 out of 39) subtelomeric promoters the 61 bp repeat is present, at most (97 out of 106) interstitial promoters this sequence is missing. It was suggested that, at telomeres, this sequence, which contains CTCF binding sites, may limit transcription proceeding towards the centromere (Deng et al. 2012). Since, as described in the next paragraph, some interstitial promoters are flanked by genes, it is tempting to speculate that the absence of the 61 bp repeat may allow transcription in both directions.

The organization of interstitial TERRA-like promoters in the T2T-CHM13v2.0 and in the GRCh38 reference genome was then compared. Only at one intrachromosomal locus (10.1 in Supplemental Table 3) the sequence organization was different in the two genomes. At the majority of loci also the length of the repeat arrays was the same in the two genomes suggesting that their sequence is conserved in the human population.

Comparative analysis and chromosomal localization of TERRA promoters

To test whether the sequences containing TERRA promoters are related, we downloaded the 5 kb flanking the 5' end of each subtelomeric and of each interstitial promoter. We then aligned them using the Kalign software and produced a heat map. In Supplemental Figure S1, the promoter organization, as in Figure 1, is reported. In cluster 1, 22 loci containing type K interstitial promoters flanking an ITS were included. Sixteen of these promoters were located on the acrocentric chromosomes. The majority of subtelomeric sequences were grouped in two clusters (2 and 3 in Supplemental Fig. S1) comprising 15 and 9 regions, respectively. The 5 kb regions belonging to cluster 2 were composed by a single copy sequence shared among all the 15 subtelomeres while a different single copy sequence was shared among the 9 subtelomeres in the second cluster. These single copy sequences were interrupted by a few transposons. Interestingly, four regions flanking interstitial promoters adjacent to ITSs (2.14, 2.15, 19.1 and 20.1 in Supplemental Table S3) were included in these subtelomeric clusters. The presence of subtelomeric and interstitial sites was also observed in cluster number 9. In each one of the remaining clusters (4-8 and 10) promoters with the same sequence organization were included.

A possible interpretation of these results is that the subtelomeric regions belonging to clusters 2 and 3, together with the interstitial regions included in the same clusters, derived from a common ancestral sequence. Successive expansions of variants that were generated through mutation may have given rise to the other clusters. In this scenario, interstitial promoters may have originated from telomeric promoters with the adjacent ITS deriving from telomeric repeats. The interstitial promoters lacking the ITS may have been generated following loss of telomeric repeats. Alternatively, these ITS-less loci may derive from duplication events that did not involve the telomeric repeats. The second hypothesis is supported by the observation that promoters with and without ITSs are separated in different clusters (Supplemental Fig. S1). In our previous work (Nergadze et al. 2004, 2007; Santagostino et al. 2020), we showed that short ITSs were inserted in the genomes during the repair of double-strand break. Here, we propose that the new class of ITSs, adjacent to TERRA promoter-like sequences, derived from recombination events involving subtelomeric and interstitial sites.

We then localized all TERRA promoters on an ideogram of human G-banded chromosomes (Fig. 2). The intrachromosomal promoters are not randomly distributed but most of them are grouped within single chromosome bands. As shown in Supplemental Table S3, about 50% of the loci are located at less than 1 Mb from adjacent loci and, in several cases at less than 100 kb. The regions containing TERRA promoters are mainly localized near pericentromeric or subterminal regions (Fig. 2) and 29 out the 106 loci are on the short arm of the acrocentric chromosomes (13, 14, 15, 21 and 22). Some chromosomal bands on metacentric chromosomes are also particularly enriched in these loci, such as 1q21.1, 2p11.2, 9p11.2, 9q21.11, 16p11.2. In the Figure the telomeres containing TERRA promoters are also marked (red arrows). It is well known that pericentromeres, subtelomeres and acrocentric short arms are highly enriched in segmental duplications (SDs). The SD content of the human genome is about 7% and SDs make up two-thirds of acrocentric short arms (Vollger et al. 2022; Jeong et al. 2025). Given their chromosomal localization (Fig. 2) and similarity (Fig. 1 and Supplemental Fig. S1), we propose that the TERRA promoter containing regions belong to a large family of both intra- and inter-chromosomal segmental duplications that were generated during evolution via recombination events involving subtelomeric and interstitial regions. Using the SEDEF segmental duplication track from genome browser, we confirmed that 105 out of the 106 interstitial promoters are located in regions marked as duplications.

Sequence analysis of the regions containing intrachromosomal TERRA promoters was carried out using UCSC genome browser. At 47 loci, pseudogenes or genes for non-coding RNAs, whose transcription starts at less than 100 bp from a TERRA promoter, were found (Supplemental Table S4). At six loci, predicted protein coding genes were found. Genes were in the forward or in the reverse orientation relative to the promoter. At sixteen loci, genes were found on both sides of the promoter. Although localized on different chromosomes, several of these genes share high sequence identity with forward and reverse genes being included in different clusters (Supplemental Fig. S2). This is not surprising given that they are contained in duplicated regions. It is tempting to speculate that specific transcription factors may be responsible for a concerted regulation of the genes sharing common promoter organizations and that the regulation of TERRA transcription may share some of these factors.

Transcription of TERRA molecules from interstitial telomeric sites

TERRA can be transcribed from several telomeres and transcription levels can vary in different cell lines (Azzalin et al. 2007; Nergadze et al. 2009; Schoeftner and Blasco 2008; Smirnova et al. 2013; Vitelli et al. 2018; Feretzaki et al. 2019; Rodrigues et al. 2024). In this work, we tested whether TERRA-like molecules are also transcribed from ITS loci producing interstitial TERRA molecules that, from now on, we will call iTERRA. The results are reported in Figure 3. We analyzed the transcription of some ITS loci adjacent to interstitial TERRA promoters by constructing primer pairs to amplify fragments from the pre-tel regions (Supplemental Table S5). We then tested the specificity of these primers using BLAT and *in silico* PCR in the T2T-CHM13v2.0 assembly. Five primer pairs were specific for a single locus (1.1, 2.18, 9.10, 14.3, 22.1). Given the high identity among TERRA promoter containing regions, the remaining two primer pairs amplified several loci. We also analyzed the transcription of three ITSs not flanked by TERRA promoters by constructing primer pairs on the sequence upstream the telomeric repeats, that is between 6 and 278 bp from the (TTAGGG)n array (Supplemental Table S5). These primer pairs are locus specific. As positive controls for TERRA expression, we used a primer pair specific for the 20p subtelomere, a primer pair specific for the Xp subtelomere and a primer pair that amplifies six subtelomeres (5q, 7p, 9q, 10q,13q and 16q). We measured the levels of TERRA transcription in a primary fibroblast (HPF) cell line, in three telomerase-positive (HeLa, HT1080 and HEK 293-F) and three ALT (GM847, U2OS and VA13) cancer cell lines by carrying out RT-qPCR experiments on cDNA synthesized using the ACCCTAACCCTAACC oligonucleotide. The expression levels of subtelomeres characterized by TERRA promoters was variable in the different cell lines (Fig. 3A left and middle) and were particularly high in U2OS cells but lower in the other two ALT cell lines. In primary fibroblasts, HeLa and HT1080 cells, expression levels were particularly low. These results are in agreement with previous reports were TERRA levels in primary fibroblasts, HeLa and U2OS cells were compared (Azzalin et al. 2007; Feretzaki et al. 2019; Rodrigues et al. 2024). Variable expression levels were also found at the Xp subtelomere that is characterized by the absence of a TERRA promoter. Interestingly, transcription of iTERRA molecules was detected from all the interstitial loci flanked by TERRA promoters analyzed here (Fig. 3B). As for the subtelomeric loci, great variability among cell lines was observed. The expression levels were always low in primary fibroblasts and HeLa cells while extremely variable in the other telomerase positive and in the ALT cell lines. At the ITS loci shown in the

first row of Figure 3B, expression levels were higher in U2OS cells compared to HeLa and fibroblasts. However, with three primer pairs (second row in Fig. 3B), we detected higher expression levels in two telomerase positive HT1080 and HEK 293 cell lines compared to the three ALT cell lines (GM847, U2OS and VA13). The three ITS loci not flanked by a TERRA promoter were also expressed, although at relatively low levels (Fig. 3C).

Although only a subset of interstitial telomeric loci was tested, the results demonstrated that, in all cell lines, the pool of TERRA transcripts is composed by molecules synthesized from both telomeres and ITS loci. The length of the telomeric repeat tract in the 66 ITS loci flanked by a TERRA promoter is comprised between 201 and 878 bp (Supplemental Table S3), therefore, the length of iTERRA telomeric repeats produced by such loci is expected to be of a few hundred bp. The loci analyzed here (Fig. 3B) are transcribed at levels comparable to the telomeres (Fig. 3A). It was reported that the bulk of the TERRA telomeric repeats has a size ranging between 100 and 400 bp with a median length of 200 bp (Porro et al. 2010), which is compatible with the expected size of iTERRA molecules produced by ITSs adjacent to TERRA promoters. In addition, we showed that the size of promoter-less ITSs is highly variable (16-1840 bp) and their transcription levels may be relatively low (Fig. 3C). However, since hundreds of such loci are present in the human genome, their contribution to the production of TERRA transcripts may also be relevant. All together, these observations suggest that iTERRA molecules constitute an important component of the total TERRA pool.

CONCLUSIONS

In this work we carried out a comprehensive search and sequence organization analysis of TERRA promoters in the T2T-CHM13v2.0 human reference genome. We found that 39 out of 46 subtelomeres contained TERRA promoters. We grouped them in classes depending on their sequence organization and on the presence or absence of the various repeats. The comparison of these loci in two reference genomes revealed that the presence and extension of the repeats can vary suggesting that TERRA promoters are polymorphic in the human population. Although it was previously shown that the 29 bp repeat is sufficient to drive transcription (Nergadze et al. 2009), we can speculate that the other repeats may have a functional role in the regulation of TERRA transcription and, therefore, the sequence organization of the promoter may

influence transcription efficiency of telomeres in different cell lines, tissues and individuals. No TERRA promoters were found at seven subtelomeres. The Xp telomere, which is not adjacent to a TERRA promoter, is transcribed suggesting that alternative promoters can drive TERRA transcription.

We then updated the list of ITS loci in the T2T genome and discovered that 66 of them are flanked by TERRA-like promoters which are, therefore, localized at intrachromosomal sites. Additional intrachromosomal promoters, not adjacent to telomeric repeats, were found at 39 loci. Sequence analysis with UCSC revealed that 47 interstitial promoters are flanked, on one or on both sides, by putative or annotated genes suggesting that common factors may regulate the transcription of TERRA and of these genes.

Comparative sequence analysis showed that interstitial and subtelomeric promoters derive from a common ancestral sequence that gave rise to a previously undescribed family of segmental duplications.

To test whether TERRA molecules can be transcribed from ITS loci, we carried out RT-PCR experiments in seven cell lines and demonstrated that TERRA transcripts were synthesized from all the interstitial loci. TERRA expression was always low in primary fibroblasts and HeLa cells while highly variable in the other two telomerase positive and in the three ALT cell lines. Given the large number of ITSs (522) that are present in the human genome, we propose that these loci contribute significantly to the production of the TERRA pool.

MATERIALS AND METHODS

Identification of subtelomeric and interstitial TERRA promoters

Subtelomeric TERRA promoters were identified in the T2T-CHM13v2.0 genome assembly (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/) (Nurk et al. 2022). We downloaded the 10 kb sequences immediately adjacent to each telomere using UCSC Genome Browser (https://genome.ucsc.edu). Using Multalin, we searched each subtelomere for sequences related to the previously described Xq TERRA promoter (chrX:154255311-154256622, T2T-CHM13v2.0 assembly) (Nergadze et al. 2009). Sequences were manually analyzed to define the organization of the promoters. Using Clustal Omega (Sievers et al. 2011) sequence identity with the Xq reference was determined and varied between 80 and 100% for the 61 bp, between 64 and 99 % for the 29like, between 55 and 100% for

the 29 bp, between 71 and 100% for the 37 bp and between 65 and 100% for the pre-tel. Subtelomeres that did not contain sequences related to the TERRA promoter were analyzed for the presence of predicted promoters using CpGplot (https://www.ebi.ac.uk/Tools/seqstats/ emboss_cpgplot/), YAPP (http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi/), and Promoter 2.0 (https://services.healthtech.dtu.dk/services/Promoter-2.0/).

Interstitial TERRA promoters were BLAST-searched (https://blast.ncbi.nlm.nih.gov/Blast.cgi) in the T2T-CHM13v2.0 genome assembly using the sequence chrX:154255311-154256622 as query. The search was performed using this setup: blastn algorithm, max target sequences = 100; automatically adjust parameters for short input sequences = disabled; expect threshold = 10; word size = 11; max matches in a query range = 0; match/mismatch scores = 2, -3; gap costs = existence: 5, extension: 2, and filters for low complexity regions and species-specific repeats disabled. The automatic adjustment of search parameters for short sequences was disabled.

Further manipulations of the hit list were carried out using tools available on the Galaxy platform (https://usegalaxy.org/) (Afgan et al. 2016). Sequences were analyzed manually and through Multalin alignment against the chrX:154255311-154256622 sequence to analyze promoter sequences.

Comparative analysis of TERRA promoters and flanking genes

To perform sequence comparison, we downloaded the 5 kb flanking the 5' end of each subtelomeric and of each interstitial promoter. Sequences were aligned using the Kalign software (Lassmann and Sonnhammer 2005) (https://www.ebi.ac.uk/jdispatcher/msa/kalign?stype=dna) to produce a percentage identity matrix that was converted into a heatmap using the R package "pheatmap" (Kolde 2018) (https://rdrr.io/cran/pheatmap/). Using the UCSC genome browser, we searched in the T2T-CHM13v2.0 genome for the presence of putative or annotated genes whose transcription starts within 100 bp from each interstitial promoter. The genomic sequences corresponding to these gene were downloaded and aligned using the Kalign software to produce a percentage identity matrix that was converted into a heatmap.

Localization of TERRA promoters on the human chromosomes

The kariogram of human G-banded chromosomes showing the localization of the subtelomeric and interstitial TERRA promoters was prepared using the R package RIdeogram (Hao et al. 2020) and cytoband information from the T2T-CHM13v2.0 genome assembly available at the UCSC Genome Browser.

To test whether regions containing intrachromosomal TERRA promoters belong to a family of Segmental Duplications, we used the Table Browser tool to find the promoters that have any overlap with regions in the SEDEF Segmental Duplications (https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=2516619039_nMpaYeBSA3J2vYcU3JP5EXXUa27D&db=hub_3671779_hs1&c=chr9&g=hub_3671779 sedefSegDups).

Search for ITS loci in the human T2T-CHM13v2.0 genome assembly

To update the list of human ITSs, the sequence $(TTAGGG)_4$ was used as query for a BLAST search (https://blast.ncbi.nlm.nih.gov/Blast.cgi/)(Altschul et al. 1990) against the human genome reference sequence *Homo sapiens* T2T-CHM13v2.0 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009914755.1/)(Nurk et al. 2022). The search was performed with the "blastn" algorithm and the following parameters: Max target sequences = 100, Expected threshold = 10, Word sizes =11, the automatic adjustment parameters for short sequences, low complexity regions and species-specific repeats regions were disabled. Further manipulations of the hit list were carried out using tools available on the Galaxy platform (https://usegalaxy.org/) (Afgan et al. 2016).

The exact position and organization of each ITS in the human genome was manually determined using the UCSC Genome Browser (https://genome.ucsc.edu/) (Kent 2002).

Cell culture

Seven human cell lines were used to test TERRA transcription at interstitial loci: human normal primary fibroblasts (HPFs), three telomerase-positive (HeLa, HEK293-F adenovirus-transformed human embryonic kidney and HT1080 fibrosarcoma cells) and three ALT cancer cell lines (U2OS osteosarcoma cells, VA13 SV40-immortalized lung fibroblast and GM847 SV40-immortalized skin fibroblasts cells). All cell lines were routinely cultured in 5% CO_2 at 37°C. Human primary fibroblasts were propagated in

Dulbecco's Modified Eagle's Media (DMEM, Euroclone), supplemented with 20% Fetal bovine serum (Euroclone), 1% L-glutamine (Sigma), 2% non-essential amino acids (Euroclone) and 1% Penicillin-Streptomycin (Sigma). Cancer cell lines were cultured in DMEM supplemented with 10% Fetal bovine serum (Euroclone), 1% L-glutamine (Sigma), 1% non-essential amino acids (Euroclone) and 1% Penicillin-Streptomycin (Sigma).

RNA preparation, cDNA synthesis and RT-qPCR

Total RNA was extracted from 10^7 cells using QIAzol Lysis Reagent (QIagen) according to the manufacturer's protocol. To remove genomic DNA, we repeated the following procedure twice: RNase-free DNase I (Thermo Scientific) (1 U per µg of RNA) at 37°C for 30 min; "on column" DNase I digestion and purification using the Zymo Research RNA Clean & Concentrator kit (Zymo Research) according to the manufacturer's instructions.

To synthesize the cDNA for quantitative Real Time PCR experiments, we set up reverse transcription reactions in a 20 µl volume containing 5 µg of total RNA, 10 pmoles of the ACCCTAACCCTAACC oligonucleotide to reverse transcribe TERRA, 5 pmoles of the AGTTGCCAGCCCTCCTAGA oligonucleotide to reverse transcribe the reference gene β2-microglobulin. Reverse transcriptions were carried out at 58°C using the Maxima H- Reverse Transcriptase Transcriptase (Thermo Scientific) according to the manufacturer's protocol for GC-rich transcripts. Finally, to prevent the formation of RNA:DNA hybrids, the reaction was incubated with 1 U of RNase H (Promega) at 37°C for 30 min. RT- reactions were prepared using the same protocol and substituting the Maxima H- Reverse Transcriptase Transcriptase with an equivalent volume of water.

Forward and reverse primers for RT-qPCR reactions were designed on the DNA sequence upstream the TTAGGG array of telomeres and of ITS loci, within 400 bp from of the TTAGGG repeats (Supplemental Table S5). Quantitative Real Time PCRs were performed using the 2X GoTaq qPCR master mix (Promega) in a 96-well reaction plate (Thermo Scientific). For each sample, 3 cDNAs and 2 RT- replicates were analyzed. Mixes were prepared according to the manufacturer's protocol in a total volume of 20 µl containing 2 µl of cDNA, and the forward and reverse primer (5 pmoles each) (Supplemental Table S5). For each primer pair a no-template control (NTC) containing 2 µl of water instead of cDNA was tested to detect

the formation of primer dimers. Reactions were carried out using the PikoReal Real-Time PCR System (Thermo Scientific). Cycling parameters included an initial denaturation at 95°C for 7 min, followed by 45 cycles composed by the following steps: 95°C for 10 sec, annealing coupled to extension at 62°C for 30 sec with fluorescence detection. A final fluorescence acquisition for melting curve analysis (60°C to 95°C) was done.

The relative TERRA expression was calculated according to the standard "$2^{-\Delta\Delta Ct}$ method" (Livak and Schmittgen 2001). Data were normalized against the hB2M gene using the human primary fibroblast cell line as reference. When no TERRA expression was detectable in fibroblasts, we used the cell line with the lowest expression as reference. For each cell line, the minimum and maximum relative TERRA expression detected for each primer pair in each triplet of reactions was calculated using the error propagation.

ACKNOWLEDGMENTS

We thank Francesco Lescai for helpful suggestions on bioinformatic analyses.

REFERENCES

Abreu PL, Lee YW, Azzalin CM. 2022. In Vitro Characterization of the Physical Interactions between the Long Noncoding RNA TERRA and the Telomeric Proteins TRF1 and TRF2. *Int J Mol Sci* **23**: 10463.

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3–W10.

Aksenova AY, Mirkin SM. 2019. At the Beginning of the End and in the Middle of the Beginning: Structure and Maintenance of Telomeric DNA Repeats and Interstitial Telomeric Sequences. *Genes (Basel)* **10**.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Al-Turki TM, Maranon DG, Nelson CB, Lewis AM, Luxton JJ, Taylor LE, Altina N, Wu F, Du H, Kim J, et al. 2024. Telomeric RNA (TERRA) increases in response to spaceflight and high-altitude climbing. *Commun Biol* **7**: 698.

Arnoult N, Van Beneden A, Decottignies A. 2012. Telomere length regulates TERRA levels through

increased trimethylation of telomeric H3K9 and HP1α. *Nat Struct Mol Biol* **19**: 948–956.

Avogaro L, Querido E, Dalachi M, Jantsch MF, Chartrand P, Cusanelli E. 2018. Live-cell imaging reveals the dynamics and function of single-telomere TERRA molecules in cancer cells. *RNA Biol* **15**: 787–796.

Azzalin CM. 2025. TERRA and the alternative lengthening of telomeres: a dangerous affair. *FEBS Lett* **599**: 157–165.

Azzalin CM, Mucciolo E, Bertoni L, Giulotto E. 1997. Fluorescence in situ hybridization with a synthetic (T2AG3)n polynucleotide detects several intrachromosomal telomere-like repeats on human chromosomes. *Cytogenet Cell Genet* **78**: 112–115.

Azzalin CM, Nergadze SG, Giulotto E. 2001. Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma* **110**: 75–82.

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.

Beishline K, Vladimirova O, Tutton S, Wang Z, Deng Z, Lieberman PM. 2017. CTCF driven TERRA transcription facilitates completion of telomere DNA replication. *Nat Commun* **8**: 2114.

Brown WR, MacKinnon PJ, Villasanté A, Spurr N, Buckle VJ, Dobson MJ. 1990. Structure and polymorphism of human telomere-associated DNA. *Cell* **63**: 119–132.

Caslini C, Connelly JA, Serna A, Broccoli D, Hess JL. 2009. MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. *Mol Cell Biol* **29**: 4519–4526.

Chu H-P, Cifuentes-Rojas C, Kesner B, Aeby E, Lee H-G, Wei C, Oh HJ, Boukhali M, Haas W, Lee JT. 2017. TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell* **170**: 86-101.e16.

Clemente L, Mazzoleni S, Pensabene Bellavia E, Augstenová B, Auer M, Praschag P, Protiva T, Velenský P, Wagner P, Fritz U, et al. 2020. Interstitial Telomeric Repeats Are Rare in Turtles. *Genes (Basel)* **11**: 657.

Colomina V, Catalan J, Britton-Davidian J, Veyrunes F. 2017. Extensive Amplification of Telomeric Repeats in the Karyotypically Highly Diverse African Pygmy Mice. *Cytogenet Genome Res* **152**: 55–64.

Cusanelli E, Romero CAP, Chartrand P. 2013. Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell* **51**: 780–791.

de Lange T. 2018. Shelterin-Mediated Telomere Protection. *Annu Rev Genet* **52**: 223–247.

Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. 2009. TERRA RNA binding to TRF2

facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* **35**: 403–413.

Deng Z, Wang Z, Stong N, Plasschaert R, Moczan A, Chen H-S, 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.

Diman A, Boros J, Poulain F, Rodriguez J, Purnelle M, Episkopou H, Bertrand L, Francaux M, Deldicque L, Decottignies A. 2016. Nuclear respiratory factor 1 and endurance exercise promote human telomere transcription. *Sci Adv* **2**: e1600031.

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.

Faravelli M, Azzalin CM, Bertoni L, Chernova O, Attolini C, Mondello C, Giulotto E. 2002. Molecular organization of internal telomeric sequences in Chinese hamster chromosomes. *Gene* **283**: 11–16.

Feretzaki M, Renck Nunes P, Lingner J. 2019. Expression and differential regulation of human TERRA at several chromosome ends. *RNA* **25**: 1470–1480.

Graf M, Bonetti D, Lockhart A, Serhal K, Kellner V, Maicher A, Jolivet P, Teixeira MT, Luke B. 2017. Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell* **170**: 72-85.e14.

Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**: e251.

IJdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci USA* **88**: 9051–9055.

Jeong H, Dishuck PC, Yoo D, Harvey WT, Munson KM, Lewis AP, Kordosky J, Garcia GH, Human Genome Structural Variation Consortium (HGSVC), Yilmaz F, et al. 2025. Structural polymorphism and diversity of human segmental duplications. *Nat Genet* **57**: 390–401.

Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome Res* **12**: 656–664.

Kienzl P, Deloria AJ, Hunjadi M, Hadolt JM, Haering M-F, Bothien A, Mejri D, Korkut-Demirbaş M, Sampl S, Weber G, et al. 2024. Telomere transcripts act as tumor suppressor and are associated with favorable prognosis in colorectal cancer with low proliferating cell nuclear antigen expression. *Cell Oncol (Dordr)*.

Kolde R. 2018. pheatmap: Pretty Heatmaps. https://github.com/raivokolde/pheatmap.

Lassmann T, Sonnhammer ELL. 2005. Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* **6**: 298.

Le Berre G, Hossard V, Riou J-F, Guieysse-Peugeot A-L. 2019. Repression of TERRA Expression by Subtelomeric DNA Methylation Is Dependent on NRF1 Binding. *Int J Mol Sci* **20**: 2791.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402–408.

López de Silanes I, Stagno d'Alcontres M, Blasco MA. 2010. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nat Commun* **1**: 33.

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. 2008. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Mol Cell* **32**: 465–477.

Maicher A, Kastner L, Dees M, Luke B. 2012. Deregulated telomere transcription causes replication-dependent telomere shortening and promotes cellular senescence. *Nucleic Acids Res* **40**: 6649–6659.

Manzato C, Larini L, Oss Pegorar C, Dello Stritto MR, Jurikova K, Jantsch V, Cusanelli E. 2023. TERRA expression is regulated by the telomere-binding proteins POT-1 and POT-2 in Caenorhabditis elegans. *Nucleic Acids Res* **51**: 10681–10699.

Marión RM, Montero JJ, López de Silanes I, Graña-Castro O, Martínez P, Schoeftner S, Palacios-Fábrega JA, Blasco MA. 2019. TERRA regulate the transcriptional landscape of pluripotent cells through TRF1-dependent recruitment of PRC2. *Elife* **8**: e44656.

Mazzoleni S, Schillaci O, Sineo L, Dumas F. 2017. Distribution of Interstitial Telomeric Sequences in Primates and the Pygmy Tree Shrew (Scandentia). *Cytogenet Genome Res* **151**: 141–150.

Messier W, Li SH, Stewart CB. 1996. The birth of microsatellites. *Nature* **381**: 483.

Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG, Wiley JE, Wurster-Hill DH, Yates TL, Moyzis RK. 1990. Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. *Chromosoma* **99**: 3–10.

Mondello C, Pirzio L, Azzalin CM, Giulotto E. 2000. Instability of interstitial telomeric sequences in the human genome. *Genomics* **68**: 111–117.

Nanda I, Schrama D, Feichtinger W, Haaf T, Schartl M, Schmid M. 2002. Distribution of telomeric (TTAGGG)(n) sequences in avian chromosomes. *Chromosoma* **111**: 215–227.

Nassour J, Aguiar LG, Correia A, Schmidt TT, Mainz L, Przetocka S, Haggblom C, Tadepalle N, Williams A, Shokhirev MN, et al. 2023. Telomere-to-mitochondria signalling by ZBP1 mediates replicative crisis. *Nature* **614**: 767–773.

Nergadze SG, Farnung BO, Wischnewski H, Khoriauli L, Vitelli V, Chawla R, Giulotto E, Azzalin CM. 2009. CpG-island promoters drive transcription of human telomeres. *RNA* **15**: 2186–2194.

Nergadze SG, Rocchi M, Azzalin CM, Mondello C, Giulotto E. 2004. Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res* **14**: 1704–1710.

Nergadze SG, Santagostino MA, Salzano A, Mondello C, Giulotto E. 2007. Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol* **8**: R260.

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.

Park J, Zhu Q, Mirek E, Na L, Raduwan H, Anthony TG, Belden WJ. 2019. BMAL1 associates with chromosome ends to control rhythms in TERRA and telomeric heterochromatin. *PLoS One* **14**: e0223803.

Porro A, Feuerhahn S, Lingner J. 2014. TERRA-reinforced association of LSD1 with MRE11 promotes processing of uncapped telomeres. *Cell Rep* **6**: 765–776.

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.

Redon S, Reichenbach P, Lingner J. 2010. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res* **38**: 5797–5806.

Reig-Viader R, Vila-Cejudo M, Vitelli V, Buscà R, Sabaté M, Giulotto E, Caldés MG, Ruiz-Herrera A. 2014. Telomeric repeat-containing RNA (TERRA) and telomerase are components of telomeres during mammalian gametogenesis. *Biol Reprod* **90**: 103.

Rodrigues J, Alfieri R, Bione S, Azzalin CM. 2024. TERRA ONTseq: a long-read-based sequencing pipeline

to study the human telomeric transcriptome. *RNA* **30**: 955–966.

Rudenko G, Van der Ploeg LH. 1989. Transcription of telomere repeats in protozoa. *EMBO J* **8**: 2633–2638.

Ruiz-Herrera A, García F, Azzalin C, Giulotto E, Egozcue J, Ponsà M, Garcia M. 2002. Distribution of intrachromosomal telomeric sequences (ITS) on Macaca fascicularis (Primates) chromosomes and their implication for chromosome evolution. *Hum Genet* **110**: 578–586.

Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E. 2008. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res* **122**: 219–228.

Sampl S, Pramhas S, Stern C, Preusser M, Marosi C, Holzmann K. 2012. Expression of telomeres in astrocytoma WHO grade 2 to 4: TERRA level correlates with telomere length, telomerase activity, and advanced clinical grade. *Transl Oncol* **5**: 56–65.

Santagostino M, Piras FM, Cappelletti E, Del Giudice S, Semino O, Nergadze SG, Giulotto E. 2020. Insertion of Telomeric Repeats in the Human and Horse Genomes: An Evolutionary Perspective. *Int J Mol Sci* **21**: E2838.

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.

Scheibe M, Arnoult N, Kappei D, Buchholz F, Decottignies A, Butter F, Mann M. 2013. Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome Res* **23**: 2149–2157.

Schoeftner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**: 228–236.

Shay JW, Wright WE. 2019. Telomeres and telomerase: three decades of progress. *Nat Rev Genet* **20**: 299–309.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

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.

Smirnova A, Gamba R, Khoriauli L, Vitelli V, Nergadze SG, Giulotto E. 2013. TERRA Expression Levels Do Not Correlate with Telomere Length and Radiation Sensitivity in Human Cancer Cell Lines. *Front Oncol* **3**: 115.

Sola L, Nergadze SG, Cappelletti E, Piras FM, Giulotto E, Santagostino M. 2021. Telomeric-Like Repeats Flanked by Sequences Retrotranscribed from the Telomerase RNA Inserted at DNA Double-Strand Break Sites during Vertebrate Genome Evolution. *Int J Mol Sci* **22**: 11048.

Tutton S, Azzam GA, Stong N, Vladimirova O, Wiedmer A, Monteith JA, Beishline K, Wang Z, Deng Z, Riethman H, et al. 2016. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J* **35**: 193–207.

Vitelli V, Falvo P, G Nergadze S, Santagostino M, Khoriauli L, Pellanda P, Bertino G, Occhini A, Benazzo M, Morbini P, et al. 2018. Telomeric Repeat-Containing RNAs (TERRA) Decrease in Squamous Cell Carcinoma of the Head and Neck Is Associated with Worsened Clinical Outcome. *Int J Mol Sci* **19**.

Vitelli V, Falvo P, Khoriauli L, Smirnova A, Gamba R, Santagostino M, Nergadze SG, Giulotto E. 2013. More on the Lack of Correlation between Terra Expression and Telomere Length. *Front Oncol* **3**: 245.

Vohhodina J, Goehring LJ, Liu B, Kong Q, Botchkarev VV, Huynh M, Liu Z, Abderazzaq FO, Clark AP, Ficarro SB, et al. 2021. BRCA1 binds TERRA RNA and suppresses R-Loop-based telomeric DNA damage. *Nat Commun* **12**: 3542.

Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, Diekhans M, Sulovari A, Munson KM, Lewis AP, et al. 2022. Segmental duplications and their variation in a complete human genome. *Science* **376**: eabj6965.

Vrbsky J, Akimcheva S, Watson JM, Turner TL, Daxinger L, Vyskot B, Aufsatz W, Riha K. 2010. siRNA-mediated methylation of Arabidopsis telomeres. *PLoS Genet* **6**: e1000986.

Wang Z, Deng Z, Dahmane N, Tsai K, Wang P, Williams DR, Kossenkov AV, Showe LC, Zhang R, Huang Q, et al. 2015. Telomeric repeat-containing RNA (TERRA) constitutes a nucleoprotein component of extracellular inflammatory exosomes. *Proc Natl Acad Sci USA* **112**: E6293-6300.

Wang Z, Lieberman PM. 2016. The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol* **13**: 690–695.

FIGURE LEGENDS

Figure 1. Organization of subtelomeric and interstitial TERRA promoters.

Sketches of the organization of TERRA promoters found at subtelomeres (A-G) and interstitial loci flanked (H-M) or not flanked by ITSs (N-S). Telomeric repeats are indicated by red circles. The 61, 29like, 29 and 37 bp repeats are indicated with red, light blue, dark blue and orange rectangles, respectively. The pre-tel and the spacer sequences are indicated with gray and green lines, respectively. Flanking sequences unrelated to promoters are indicated with black lines. The subtelomeric or interstitial loci containing each type of promoter organization are listed below its sketch. Subtelomeres are indicated using the chromosome arm, interstitial loci are indicated using the names from Supplemental Table S3.

Figure 2: Localization of TERRA promoters in the human genome.

Localization of subtelomeric (red triangles) and interstitial TERRA promoters flanked (light blue circles) or not flanked by ITSs (dark blue diamonds) on ideograms of human G-banded chromosomes. Light blue rectangles on chromosomes mark the location of groups of interstitial promoters. Magnifications on the left and right side of the chromosomes show the position of interstitial promoters in each group. Numbers at each end of the magnifications indicate the start and end coordinates of each group.

Figure 3: Transcription levels of iTERRA molecules measured by Real Time Quantitative PCR.

Relative quantification of TERRA transcribed from telomeres (A), from ITSs flanked by TERRA promoters (B) and from ITSs not flanked by TERRA promoters (C) in a primary fibroblast (HPF) cell line, three telomerase-positive (HeLa, HT1080 and HEK 293-F) and three ALT (GM847, U2OS and VA13) cancer cell lines. The names of the loci are reported on the top of each graph. For telomeric loci the chromosome arms (A), for ITSs flanked by interstitial promoters the numbers from Supplemental Table S3 (B) and for ITSs not flanked by interstitial promoters the coordinates of the telomeric repeat starting nucleotides are reported (C). Relative quantification was calculated using the standard "$2^{-\Delta\Delta Ct}$ method" (Livak and Schmittgen 2001). Bars indicate the minimum and maximum relative TERRA expression measured for each sample and calculated using the error propagation method. The primary fibroblasts cell

line was used as reference. When no TERRA expression was detectable in fibroblasts, the cell line with the lowest expression was used as reference (HT1080 in the 22.1 locus and GM847 in the 2.18 locus).

Subtelomeric promoters

Figure 1

A)

1q, 2p, 4p, 4q, 5q, 6p, 7p, 8p, 9p, 9q, 10q, 11q, 12p, 13q, 15p, 15q, 16p, 16q, 17q, 19p, 20p, 20q, 21q, 22q, Xq (25 subtelomeres)

B)

5p and 6q (2 subtelomeres)

C)

2q, 3q, 11p, 14q, 18q, 19q (6 subtelomeres)

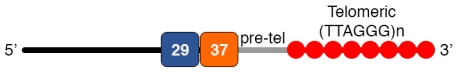
D)

7q and 12q (2 subtelomeres)

E)

1p (1 subtelomere)

F)



18p (1 subtelomere)

G)

3p and 10p (2 subtelomeres)

Intrachromosomal promoters

H)

19.1, 20.1, 20.4 (3 loci)

I)

2.15 (1 locus)

J)

2.14 (1 locus)

K)

1.1, 1.3, 1.5,1.6, 2.2, 2.3, 2.6, 2.7, 2.10, 2.11, 2.16, 2.17, 2.18, 3.1, 4.1, 4.2, 9.3, 9.5, 9.10, 9.12, 9.16, 9.17, 10.1, 13.1, 13.2, 13.4, 14.1, 14.2, 14.3, 14.4, 14,5, 14.8, 15.1, 15.2, 15.3, 15.4, 16.1, 16.4, 18.6, 20.2, 20.3, 21.1, 21.2, 21.3, 21.4, 21.5, 21.7, 22.1, 22.2, 22.3, 22.4, 22.5, 22.6, 22.8, 22.9 (55 loci)

L)

18.1, 18.2, 18.3 (3 loci)

M)

9.7, 9.8, 18.5, 21.8 (4 loci)

N)

9.18 (1 locus)

O)

11.1, 17.1 (2 loci)

P)

1.2, 1.4, 1.7, 2.1, 2.4, 2.5, 2.8, 2.9, 2.12, 2.13, 9.1, 9.2, 9.4, 9.6, 9.9, 9.11, 9.13, 9.14, 9.15, 13.3, 14.6, 14.7, 15.5, 15.6, 15.7, 15.8, 16.2, 16.3, 16.6, 16.7, 21.6, 22.7, 22.10 (33 loci)

Q)

18.4 (1 locus)

R)

16.5 (1 locus)

S)

7.1 (1 locus)

Figure 2

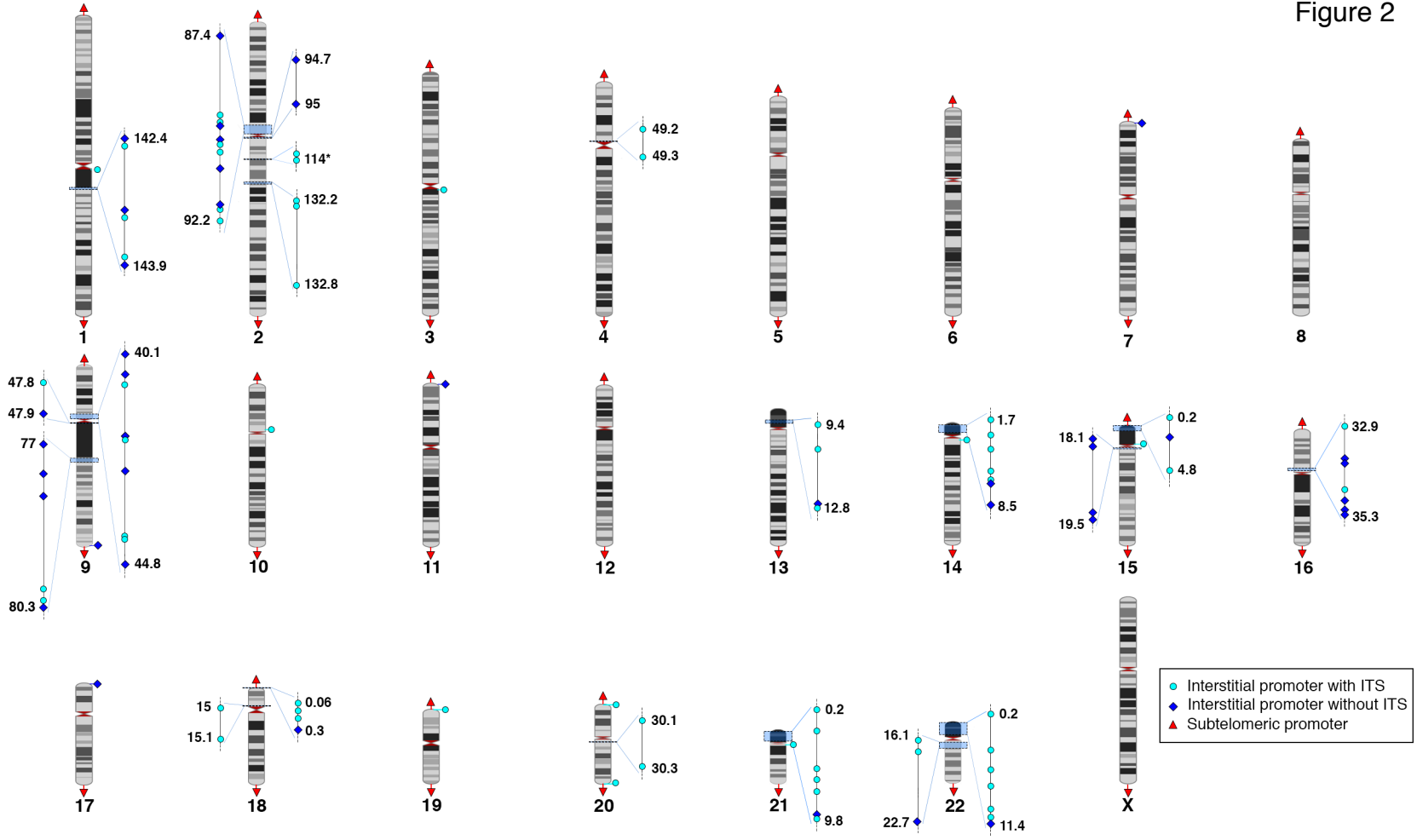
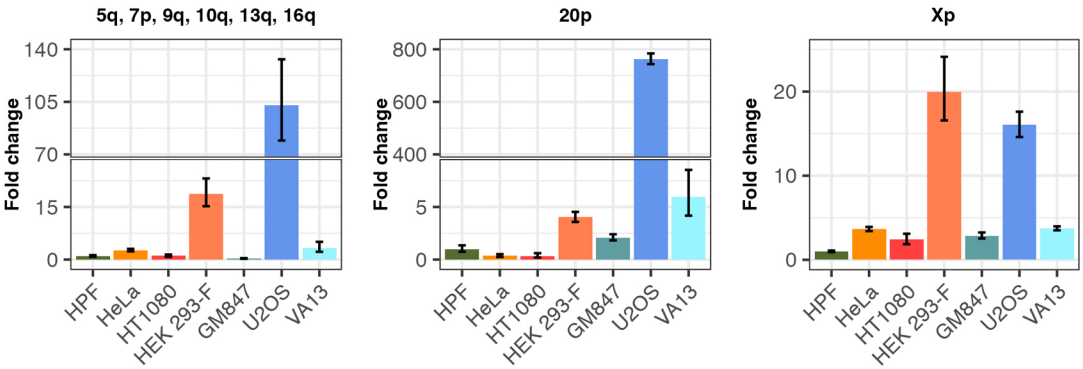
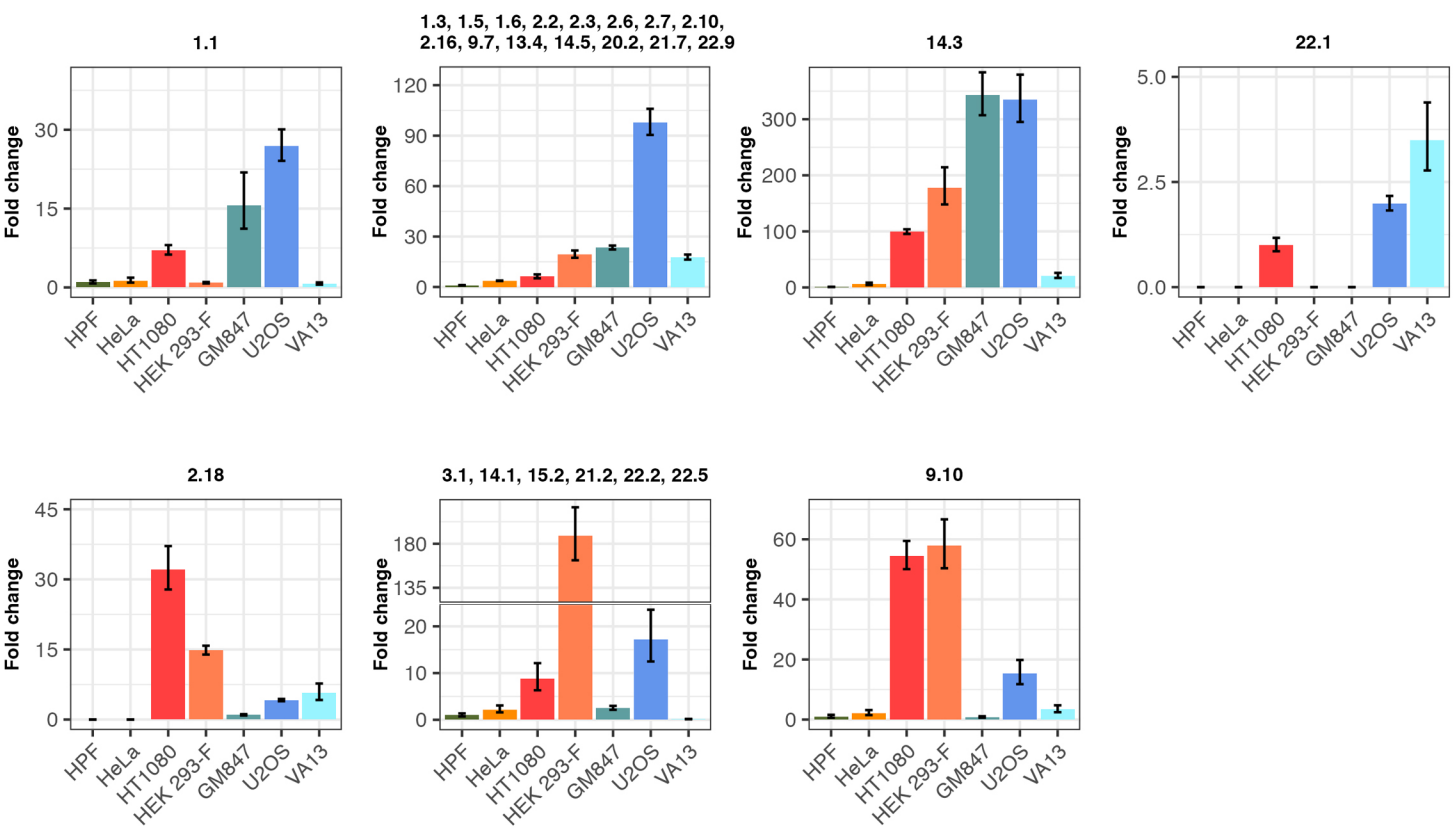
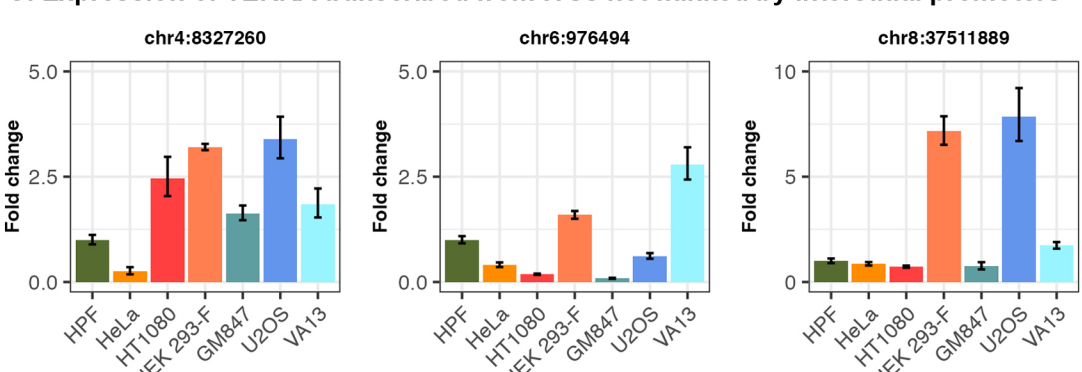

142.4
143.9
1
87.4
94.7
95
114*
132.2
92.2
132.8
2
3
49.2
49.3
4
5
6
7
8
40.1
47.8
47.9
77
44.8
80.3
9
10
11
12
9.4
12.8
13
1.7
8.5
14
0.2
18.1
4.8
19.5
15
32.9
35.3
16
17
15
0.06
15.1
0.3
18
19
30.1
30.3
20
0.2
9.8
21
0.2
16.1
22.7
22
11.4
X
Interstitial promoter with ITS
Interstitial promoter without ITS
Subtelomeric promoter

Figure 3

A. Expression of telomeric TERRA



B. Expression of TERRA transcribed from ITSs flanked by interstitial promoters



C. Expression of TERRA transcribed from ITSs not flanked by interstitial promoters