Telomeric Repeat-Containing RNA Increases in Aged Human Cells

Yu-Hung Hsieh[1], Chin-Hua Tai[1], Meng-Ting Yeh[1], Yu-Chen Chen[1], Po-Cheng Yang[1], Chien-Ping Yen[1], Der-Sheng Han[2,3,]* and Hsueh-Ping Catherine Chu[1,]*

[1] Institute of Molecular and Cellular Biology, National Taiwan University, No. 1 Sec. 4 Roosevelt Road, Taipei, Taiwan.

[2]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Bei-Hu Branch, Taipei 100, Taiwan.

[3]Department of Physical Medicine and Rehabilitation, College of Medicine, National Taiwan University.

Email: angelcemept@gmail.com (Yu-Hung Hsieh); kristoff0109@gmail.com (Po-Chen Yang);

briantai85@gmail.com (Chin-Hua Tai); ns99073100@gmail.com (Meng-Ting Yeh); ycchen99123@gmail.com (Yu-Chen Chen); d06b48002@ntu.edu.tw (Chien-Ping Yen)

* To whom correspondence should be addressed. Email: cchu2017@ntu.edu.tw (Hsueh-Ping Catherine Chu), dshan1121@yahoo.com.tw (Der-Sheng Han)

Abstract

Telomeric repeat-containing RNA (TERRA), transcribed from subtelomeric regions towards telomeric ends, poses challenges in deciphering its complete sequences. Utilizing TERRA-capture RNA-seq and Oxford Nanopore direct RNA sequencing to acquire full-length TERRA, we annotate TERRA transcription regions in the human T2T-CHM13 reference genome. TERRA transcripts encompass hundreds to over a thousand nucleotides of telomeric repeats, predominantly originating from 61-29-37 bp repeat promoters enriched with H3K4me3, RNA pol II, CTCF, and R-loops. We develop a bioinformatics tool, TERRA-QUANT, for quantifying TERRA using RNA-seq datasets and find that TERRA increases with age in blood, brain, and fibroblasts. TERRA upregulation in aged leukocytes is confirmed by RT-qPCR. Single-cell RNA-seq analysis demonstrates TERRA expression across various cell types, with upregulation observed in neurons during human embryonic stem cell differentiation. Additionally, TERRA levels are elevated in brain cells in the early stage of Alzheimer's disease. Our study provides evidence linking TERRA to human aging and diseases.

Introduction

Chromosome ends synthesize a heterogeneous population of long noncoding RNAs called "TERRA" (1), which are composed of the telomeric repeat sequence UUAGGG and sequences unique to the subtelomeric region of each chromosome. TERRA is transcribed by RNA polymerase II (2), and its expression is regulated by CpG-islands in the subtelomeric regions and altered through cell cycle progression (3-5). In human cells, only a fraction of TERRA undergoes polyadenylation (6), while the majority remains non-polyadenylated and is stabilized by RALY (7), a member of the hnRNP family that interacts with both protein-coding and long non-coding RNAs (8). The regulatory regions influencing TERRA expression have been identified by the presence of three repetitive motifs, referred to 61-29-37 base pair (bp) repeat sequences, featuring a high GC content, and common to several human chromosome ends (4). The defect in DNA methyltransferase enzyme DNMT3B results in DNA hypomethylation in the subtelomeric regions and upregulation of TERRA expression (3,9-11). Studies have shown that TERRA expression can be regulated by several transcription regulators including CTCF, ATRX, NRF1, HSF1, p53, Snail1, ZNF148, ZFX, EGR1, and PLAG1 (3,12-17). TERRA can form RNA:DNA hybrids with telomeric DNA, facilitating telomere extension in telomerase-negative cancer cells using a mechanism termed "Alternative Lengthening of Telomeres" (ALT) (18-21).

TERRA exhibits binding capabilities with telomeric DNA binding proteins (22-24) and non-telomeric DNA binding proteins such as chromatin modifiers (24). It binds to both telomeric and non-telomeric chromatin (24), regulating telomere integrity and gene expression *in cis* and *in trans* (23-25). For instance, TERRA interacts with ATRX, a protein involved in histone H3.3 recruitment and H3K9me3 formation (26), and prevents ATRX from binding to chromatin (24,27). TERRA depletion leads to increased ATRX recruitment to repetitive sequences such as rDNA, retrotransposons, subtelomeric sequences, and telomeric repeats, as well as the regions close to transcription start sites (24,27). Furthermore, TERRA plays a crucial role in modulating DNA G-quadruplex structures around the transcription start sites, suggesting a function of TERRA beyond telomere biology (27).

Elevated TERRA levels are accompanied by telomere attrition (5,28-30) and damage (9,31,32). For instance, uncapping telomeres by TRF2 depletion upregulates TERRA expression (32). Cells derived from patients with ICF (Immunodeficiency, Centromeric instability, and Facial anomalies) syndrome display accelerated telomere shortening, premature replicative senescence, and significantly elevated levels of TERRA (9,10). Bidirectional telomeric transcription is induced by DNA damage at telomeres in Hutchinson-Gilford progeria syndrome (HGPS) fibroblasts (33). Reducing telomere length by ablating Bqt4 in fission yeast results in a greatly increased TERRA expression (34). When both telomerase and homologous recombination mechanisms are absent in budding yeast, the accumulation of RNA:DNA hybrids at telomeres results in telomere loss and increased rates of cellular senescence (35). TERRA levels are elevated at senescence in budding yeast cells lacking telomerase, while TERRA ablation by expression of an artificial antisense TERRA (anti-TERRA) delays senescence (36). Furthermore, telomere shortening induced by TNF-α in mice is mediated by TERRA elevation (37). While major progress has been made in TERRA involvement in telomere lengthening and cellular senescence in yeast, TERRA expression during human aging is poorly investigated.

Examining TERRA transcripts presents several challenges, such as unassembled genomic sequences at various chromosome ends and the lack of an annotated TERRA profile in the human genome. To address these questions, we enriched TERRA transcripts using antisense oligos to capture TERRA and conducted Nanopore direct RNA sequencing (RNA-seq). Through sequence alignment to the T2T-CHM13 genome reference, we achieve the delineation of TERRA transcription regions at the end of chromosomes. Utilizing our bioinformatics tool to quantify TERRA levels from publicly available RNA-seq datasets, we demonstrate that TERRA is upregulated during human aging. Additionally, our analyses reveal differential TERRA expression in various tissues and in individuals with human diseases.

MATERIALS AND METHODS

Cell culture

U2OS and HeLa cells were consistently tested and confirmed to be free of mycoplasma contamination. Cells were cultured using Gibco Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum, L-glutamine, and 1% penicillin/streptomycin in a 37°C incubator supplied with 5% CO_2.

TERRA capture for Illumina RNA-seq

To enrich TERRA, total RNA was extracted from 1×10^7 cells using the TRIzol™ Reagent (Invitrogen™, Cat:15596018) following the manufacturer's protocol. The total RNA was treated with DNaseI (Invitrogen™, AM2238) and supplied with Ribonucleoside vanadyl complexes (VRC, New England Biolabs, E7760S) under 37°C for 15 minutes to eliminate DNA (0.4 Unit/µl DNaseI and 10 mM VRC). The enzyme was inactivated by 5 mM of EDTA. The DNaseI-treated RNA sample was then subjected to a second round of purification using TRIzol™ reagent. For one TERRA-capture experiment, a total of 20 µg of the DNaseI-treated RNA was used. The RNA sample was dissolved in 90 µl of 6x SSC buffer (0.9M of NaCl, 0.9M of Sodium citrate). The RNA sample and 10 µl of 1 µM biotinylated TERRA antisense probe (5'-$(CCCATT)_5$-3'BioTEG) were individually denatured at 70°C for 2 minutes, and then were mixed together (final: 20 µg RNA, 2x SSC buffer, 100 nM probe) for an additional 8 minutes at 70°C. The RNA-probe annealing was carried out by gradually decreasing the temperature from 70°C to 44°C, followed by incubation at 44°C for an additional 30 minutes. To capture TERRA RNA, 100 µl of streptavidin beads (MyOneC1, Invitrogen™, Cat#65002) were prepared and mixed with the sample. The mixture was then incubated on a rotating wheel at 37°C incubator for 15 minutes. The supernatant was separated from beads by a magnet and carefully removed. Beads were washed 4 times with 2x SSC buffer containing 0.1% NP40 (0.3M of NaCl, 0.3M of Sodium citrate, 0.1% NP-40) at 37°C, followed by two washes with 1x SSC buffer containing 0.1% NP40 (0.15M of NaCl, 0.15M of Sodium citrate, 0.1% NP-40), once at 37°C and once at room temperature. Finally, beads were washed once with 1x SSC buffer without NP40 (0.15M of NaCl, 0.15M of Sodium citrate) at 37°C. Washed beads were resuspended in 100 µl of nuclease-free water and incubated at 70°C to elute RNA. The

elution step was repeated twice. The eluted RNA sample was concentrated by precipitating in 3× volume of 100% ethanol (300 µl), containing 0.1× volume of 3M NaOAc (40 µl, Thermo Fisher Scientific, AM9740) and 2 µl of GlycoBlue™ (Invitrogen™, AM9515) overnight at -30℃. The RNA was precipitated by centrifuge at 4℃ and 148000 RPM. The RNA pellet was washed twice with 70% EtOH and eluted in 30 µl of nuclease-free water. RNA concentration was measured using Qubit RNA HS kit (Invitrogen™, Q32852). The captured RNA was converted into cDNA, followed by paired-end Illumina sequencing and RT-qPCR.

Illumina RNA seq library preparation

The cDNA library was constructed using the NEBNext® Ultra II Directional RNA Library Prep Kit (New England Biolabs, E7760S) according to the manufactory instruction with some modifications. The first-strand synthesis step was carried out by SuperScript IV (Thermo Fisher Scientific, Cat: 18090010). The TERRA-captured RNA sample (30 µl) was divided into two equal halves for the reverse transcription reaction. Half of the TERRA-captured sample was reverse transcribed using 2 µM random hexamers (Thermo Fisher Scientific, 48190011) and the other half using 1 µM of telomeric C-rich primer 5'-$(CCCTAA)_5$-3'. These two groups are referred to as the random and specific groups, respectively. Followed by reverse transcription, second strand synthesis, end repair, adaptor ligation, and size selection. Libraries were amplified using NEBNext Ultra II Q5 Master Mix (New England Biolabs, #M0544L) with optimal cycles determined by qPCR, and size-selected twice using AMPure XP beads. Finally, the DNA was eluted in 15 µL of 0.1× TE buffer and stored at -20℃. The library size distribution was assessed using the Agilent 2100 Bioanalyzer. Each library was quantified using the NEBNext Library Quant Kit (New England Biolabs, #E7630S). The pooled libraries (fragment size 300~500 bp) were subjected to paired-end 150 bp sequencing on Illumina MiSeq or NovaSeq platforms.

Reverse transcription and Quantitative PCR

TERRA enrichment was assayed by RT-qPCR using cDNA. The reverse transcription was performed using random hexamers (Thermo Fisher Scientific, #48190011) and SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, Cat#18090200)(final: 1µM random hexamer, 400 µM

dNTP, 1×SuperScript IV buffer, 8 Unit/µl SuperScript IV reverse transcriptase, 4 mM DTT, 1.6 Unit/µl RNaseOUT). The qPCR analysis was conducted using a 1:5 dilution of TERRA-enriched cDNA and a serial dilution of control cDNA, ranging from 1× to 1000× dilution (1% to 0.001% of input), to generate a standard curve. Subtelomeric hg38-2q primers: forward: 5'-AAAGCGGGAAACGAAAAGC-3'; reverse: 5'- GCCTTGCCTTGGGAGAATCT-3'. GAPDH primers : forward: 5'-CGTCCCGTAGACAAAATGGT-3'; reverse:5'- TTGATGGCAACAATCTCCAC-3'. The qPCR reaction was performed with IQ SYBR Green Supermix (Bio-Rad, Cat: 170-8882) using the CFX Real-Time PCR Detection System. The thermal cycling conditions were as follows: Initial heat activation of polymerase at 95℃ for 3 minutes, followed by 35 cycles of denaturation at 95℃ for 15 seconds, annealing and extension at 58℃ for 1 minute. Data analysis was carried out using CFX Maestro software v1.1 (Bio-Rad). For Illumina samples, the standard curve was generated from a series of dilutions from the input. TERRA RNA abundance was normalized to GAPDH. TERRA-captured samples were compared to the control sample (no capture) to calculate TERRA enrichment. For Illumina samples, TERRA enrichment ratio was calculated using the ddCt method, where the Ct value of TERRA was normalized to that of GAPDH.

TERRA capture for Nanopore Direct RNA-seq

We harvested cells from 200 of 150 mm culture dishes for RNA extraction. The TERRA-capture procedure followed the method described previously with a few modifications. First, the RNA samples were not treated with DNase I. Second, the input RNA amount for each capture was increased to 500 µg. Third, to achieve the required RNA quantity for Nanopore direct RNA-seq, 10 captures were pooled. RNA was concentrated by precipitation with 3× volume of 100% ethanol, 1/10 × volume of 3M NaOAc (Thermo Fisher Scientific, AM9740), and 2 µl of GlycoBlue™ (Invitrogen™, AM9515) for 3 days at -20℃. 100 captures (a total of 50 mg input RNA) were required to acquire a sufficient amount of TERRA-enriched RNA for sequencing. Subsequently, RNA was precipitated by centrifugation at 4°C at 14800 RPM for 20 minutes. The pellet was washed with 70% ethanol and eluted in 20 µl of nuclease-free water. Finally, a total of 3.4 µg RNA was obtained after capture. Next, 1.5 µg of TERRA-enriched RNA was polyadenylated using E. coli Poly (A) polymerase (New England Biolabs, #M0276S) at a final concentration of 0.25 U/µl in 1X

Poly(A) Polymerase reaction buffer with 1 mM dATP at 37°C for 30 minutes. To prevent RNA degradation, the reaction included 2 Unit/µl RNaseOUT, (Thermo Fisher Scientific, Cat: 10777019). The polyadenylation reaction was terminated by adding 10 mM EDTA. Polyadenylated RNA was purified using 2× sample volume of AMPureXP RNA beads (Beckman Coulter, A66514), following the manufacturer's protocol. RNA was eluted in 20 µl of nuclease-free water. RNA concentration was measured using Qubit RNA High-Sensitivity assay (Thermo Fisher Scientific, Q32852). The fragment size was assessed using Bioanalyzer (Agilent RNA 6000 Pico Kit). TERRA-enriched RNA was subjected to Nanopore direct RNA sequencing (Cat#SQK-RNA002) with R9.4.1 flow cell (FLO-MIN106).

Analysis of TERRA-capture Illumina RNA-seq

The fastq files were subjected to a quality filter (Q > 20), and Illumina adaptor sequences were trimmed using Trimgalore v0.6.7 (trimgalore --illumina --pair Read1.fastq Read2.fastq -o /path/to/output). The reads were aligned to the T2T-CHM13 genome using the STAR aligner v2.7.9a (star --runMode alignReads --genomeDir /path/to/T2T-CHM13 --outSAMtype BAM SortedByCoordinate --outFileNamePrefix output_path/id --readFilesIn trimmed_Read1.fastq trimmed_Read2.fastq). The aligned BAM files were indexed by samtools v1.17 (samtools index -b exp.bam -o exp.bam.bai) and converted to coverage files in bigwig format by deepTools v3.3.1 bamCoverage, with RPKM as the normalization method (bamCoverage -b BAM_files --normalizeUsing RPKM --binSize 30 -of bigwig -o output.bw). Furthermore, a comparison was performed between TERRA-captured groups and no-captured RNA-seq data (control) using log2 ratio analysis by bigwigCompare (deepTools). The coverage files (bigwig) were visualized on the Integrative Genomics Viewer (IGV) v2.16.0. TERRA transcription regions were manually assigned according to TERRA enrichment, nanopore reads, and CAGE tags at chromosome ends of the T2T-CHM13 genome. The BED and GTF files of TERRA transcription regions were used for TERRA quantification. To quantify TERRA read counts, the alignments were extracted from the BAM files using samtools v1.17 (samtools view -b -h -M -L TERRA_region.bed --fetch-pairs input.bam -o TERRA_region_alignment.bam), and filtered with MAPQ $\geq$ 30.

Analysis of TERRA Nanopore direct RNA-seq

Fast5 files were processed using Guppy base caller (guppy_GPUv5.0.7) with a default quality filter of Q > 7 (guppy -i /path/to/fast5 -c rna_r9.4.1_70bps_hac.cfg -s /path/to/output_fastq). After base-calling, reads were aligned to CHM13 reference genome using the minimap2 aligner v2.22 (minimap2 -ax splice -uf -k14 /path/to/CHM13.mmi TERRA_Nanopore.fastq > TERRA_Nanopore_aligned.sam). The alignment was converted to BAM format and sorted using samtools v1.17 (samtools view -b TERRA_Nanopore_aligned.sam -o TERRA_Nanopore_aligned.bam; samtools sort TERRA_Nanopore_aligned.bam -o TERRA_Nanopore_aligned.sorted.bam). A bam index was created by samtools v1.17 (samtools index TERRA_Nanopore_aligned.sorted.bam -o TERRA_Nanopore_aligned.sorted.bam.bai). The aligned reads displayed on the genome browser were manually examined to assist in defining TERRA transcription regions. Reads aligned to TERRA transcription regions were selected using samtools v1.17 with the TERRA transcription region bed file (samtools view -b -h -M -L TERRA_region.bed TERRA_Nanopore_aligned.sorted.bam -o Nanopore_TERRA_regions_aligned.bam), and the counts of each TERRA region were quantified. For the analysis of TERRA from individual chromosome ends (Figure 1D), the unique reads were filtered based on MAPQ (≥ 1), FLAG values (FLAG = 0, or 16 for primary alignments, FLAG = 2048, or 2064 for supplementary alignments) provided by minimap2 (**Supplementary Table 5.1**). The supplementary alignments with the same read IDs were removed (**Supplementary Table 5.2**).

Analysis of ChIP-seq datasets

The ChIP-seq datasets are listed as the following: U2OS H3K4me3 (GSE114703) (38), U2OS RNA polymerase II (GSE87324) (39), U2OS CTCF (GSE87831) (40), U2OS MeDIP-seq (GSE81165) (41), U2OS DRIP-seq (GSE115957) (42), HeLa CAGE-seq (GSE121351). Additionally, the annotated CpG island BED file was downloaded from UCSC (assembly: T2T-CHM13 v2.0, group: Expression and Regulation, track: CpG Islands, table: hub_3671779_cpgIslandExtUnmasked). Reads with quality scores below Q30 were filtered out, and adaptor sequences were trimmed using TrimGalore v0.6.7 (trimgalore --illumina --pair -q 30 Read1.fastq Read2.fastq -o /path/to/output). The processed reads were aligned to the CHM13 genome using bowtie2 with default settings. The

resulting Sequence Alignment Map (SAM) files were converted into Binary Alignment Map (BAM) format, sorted, duplicate reads were removed, and the files were indexed using samtools v1.17 (samtools view -b exp.sam -o exp.bam, samtools sort exp.bam -o exp.sorted.bam, samtools rmdup exp.sorted.bam -o exp.sorted.rmdup.bam, samtools index exp.sorted.rmdup.bam -o exp.sorted.rmdup.bam.bai). ChIP-seq coverage files (bigwig) were generated using deepTools v3.3.1 bamCompare, normalizing using CPM method and subtracting to input signal. (bamCompare -b1 ChIP_BAM_file -b2 Input_BAM_file --normalizeUsing CPM --binSize 30 --operation subtract -of bigwig -o output.bw). Meta-analysis surrounding the 61-29-37 nt repeat regions was performed using deepTools v3.3.1 computeMatrix and plotHeatmap (computeMatrix reference-point --referencePoint center -a 500 -b 500 -S ChIP-seq_coverage.bw -R 61_29_37_region.bed -o Matrix.gz, plotHeatmap -m Matrix.gz --colorMap bwr -o output_meta-analysis.pdf). For visualization on CHM13 genome, the bigwig files were examined in Integrative Genomics Viewer (IGV).

Analysis of TERRA expression by TERRA-QUANT

RNA-seq datasets were downloaded and pre-processed using the previously described methods (SRAToolkit and TrimGalore commands). The processed reads were aligned to the T2T-CHM13 genome using STAR aligner with default settings. Read counts for human genes were calculated using HTseq-count (htseq -f bam -s reverse -t transcript --idattr gene_name -m intersection-nonempty --nonunique all BAM_file CHM13_GTF > output_count.txt). Reads aligned to TERRA regions were identified and quantified using TelomereHunter (43) with the criteria of at least 2 continuous TTAGGG, TCAGGG, TGAGGG, or TTGGGG repeats (telomerehunter -ibt BAM_files -p sample_id -o /path/to/output -b cytoband_with_TERRA_region.bed -pl -rt 2 -rl). This allowed for the generation of read counts for each TERRA region containing telomeric repeats. For quantifying TERRA from individual chromosome ends, reads were extracted with MAPQ $\geq$ 30 for paired-end reads, or with MAPQ = 255 for single-end reads by samtools. For total TERRA quantification, reads were aligned by STAR with default settings, and duplicates were removed and filtered with primary alignments (MAPQ $\geq$ 1) by TelomereHunter. The reads mapped to TERRA regions and all human genes were then analyzed together and normalized using the YARN

package (44) in the R language. The total TERRA levels were accumulated from reads that mapped to all TERRA transcription regions at chromosome ends.

NGS sequencing for TERRA RT-qPCR products

RNA was extracted using TRIzol™ Reagent (Invitrogen™, Cat:15596018). Reverse transcription was performed using the SuperScript™ IV system (Thermo Fisher Scientific, Cat# 18090200) with random hexamers (2.5 nM). The RT-qPCR reaction was conducted by using IQ SYBR Green Supermix (Bio-Rad, Cat: 170-8882). The design of the primer sequences was based on the CHM13 genome (**Supplementary Table 6**). TERRA RT-qPCR products were purified using Gel/PCR Purification Mini Kit (FAVORGEN, Cat# FAGCK 001-1) before DNA library preparation. Eluted DNA was subjected to library construction using NEBNext Ultra II DNA Library Prep kit (NEB, Cat#E7645S). Paired-end 150bp reads were obtained from a Illumina NovaSeq system. The sequencing reads of TERRA RT-qPCR products underwent quality filtration (Q > 30) and adapters pruning using TrimGalore (v0.6.3) (https://github.com/FelixKrueger/ TrimGalore). Trimmed reads were aligned to the CHM13 genome using STAR aligner, filtering with MAPQ $\geq$ 30 by samtools. SAM files were transformed to BAM files, and were deduplicated using samtools (v.1.18). Bigwig files were generated by DeepTools (v.3.3.1) bamCoverage (bamCoverage --normalizeUsing CPM, --centerReads, and --extendReads 150). To assess the proportion of PCR products in the TERRA transcription regions, we calculated the read counts of PCR products on TERRA transcription regions using samtools (v.1.13) view (samtools view --fetch-pairs, -h, -b, -M (--use-index), and -L (--target-file)) with a bed file including TERRA transcription regions.

Research participants for collecting blood cells

Blood samples were collected as previously described (45). The study was approved by the institutional review board (IRB No. 201601091RIND) of the National Taiwan University Hospital. All participants were provided with written informed consent before participation in the study. Trial registration: ClinicalTrials.gov: NCT02779088. Registered 20 May 2016, https://clinicaltrials.gov/ct2/keydates/NCT02779088.

Telomere length measurement in blood cells

Blood samples were collected from the participants as previously described (45). Briefly, genomic DNA was extracted from frozen human buffy coats using the QIAamp® DNA Mini Kit (QIAGEN, Hilden, Germany, #Cat 51306). Quantitative analysis was performed by quantitative polymerase chain reaction (qPCR). The relative telomere length for each participant was calculated using the telomere-to-single copy gene (T/S) ratio, defined as the number of telomeric repeats (T) divided by a standard reference DNA (S). The *36B4* gene was used as the single-copy DNA. A standard curve was generated from a series of dilutions of genomic DNA from HeLa cells, and was used for calibration for each qPCR plate. Primers for telomeric DNA and *36B4*: *36B4* forward primer sequence: 5'-CAGCAAGTGGGAAGGTGTAATCC-3', and *36B4* reverse primer sequence: 5'-CCCATTCTATCATCAACGGGTACAA-3'; telomeric DNA forward primer sequence: 5'-GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT-3' and telomeric DNA reverse primer sequence: 5'-TCCCGACTATCCCTATCCCTATCCCTATCCCTATCCCTA-3'.

Quantification of TERRA by RT-qPCR in blood samples

RNA was extracted from buffy coat samples using TRIzol (Thermo Fisher Scientific, Cat# 15596018) and purified by acid-phenol:chloroform (Thermo Fisher Scientific, Cat# AM9722). RNA was treated with ezDNase™ Enzyme (Thermo Fisher Scientific) to eliminate contaminating genomic DNA. Reverse transcription was carried out using the SuperScript™ IV system (Thermo Fisher Scientific, Cat# 18090200). Random (2.5 nM) hexamers combined with telomeric repeat primers (0.25 nM) to enrich TERRA complementary DNA (cDNA). The protocol for total TERRA qPCR was similar to that for the telomere qPCR using telomeric repeat primers. Beta-2 microglobulin (*B2M*) was used as the reference gene. B2M primers: forward primer sequence, 5'-CTATCCAGCGTACTCCAAAG-3'; reverse primer sequence, 5'-GAAAGACCAGTCCTTGCTGA-3'. TERRA RNA extracted from U2OS cells was used to obtain a standard curve for calibration between samples from different qPCR plates. The relative TERRA expression level was analyzed using the TERRA RNA/B2M RNA ratio. Subtelomeric primers (hg38-2q) for TERRA RT-qPCR in blood cells were listed in **Supplementary Table 6**.

Analysis of TERRA expression across different tissues

Publicly available RNA-seq datasets were gathered from NCBI GEO datasets (32,46,47). Sample sources were documented in **Supplementary Table 7**. To compare TERRA levels across different tissues or cell lines, we employed the YARN R package (48). The whole gene counts derived from all RNA-seq datasets were merged in the same expression table for the following process. The filtration of low-expressed genes was performed based on an expression threshold (CPM < 1). TERRA counts for each sample were determined using TERRA-QUANT as described previously. Normalization of all data was performed using the qsmooth method (48).

Single-cell and Single-nucleus RNA-seq analyses

Neuronal single-cell RNA-seq datasets (GSE186698) and PBMC single-cell RNA-seq datasets (GSE157007) were downloaded as previously described (**Supplementary Table 8**). Alignment, barcode assignment and Unique Molecular Identifier (UMI) counting were performed using CellRanger v7.2.0 (49) with a default setting (MAPQ = 255). Downstream analysis of CellRanger outputs was performed using Seurat v4.4.0 (50). The low-quality cells were filtered out if the number of genes detected or mitochondria gene reads were three Mean Absolute Deviations (MAD) away from the average. Genes expressed in fewer than 3 cells are excluded. 2,000 highly variable genes were identified using the variance stabilizing transformation (VST) method and were used for Principle Component Analysis (PCA). The top 20 principal components were selected for cell clustering. The result was visualized with Uniform Manifold Approximation and Projection (UMAP). Data integration was performed by using FindIntegrationAnchors() and IntegrateData() functions according to the union of the top 2,000 highly variable genes of each data. Plots of TERRA, ATRX, and MAP2 were generated by scCustomize v2.0.1 (Samuel Marsh, Maëlle Salmon, & Paul Hoffman. (2023). https://doi.org/10.5281/zenodo.10161832). In neuronal differentiation samples, cells positive for CD29, CD44, and CD105 but negative for CD14, CD34, and CD45 were classified as mesenchymal stem cells; cells expressing SOX2 and Nestin were identified as neuroepithelial cells, while those positive for MAP2 were categorized as mature neurons. Cells in PBMC were annotated by the expression of representative cell markers: CD4+ T cells (CD4, IL7R and CD3G), NK cells (NCAM1 and NKG7), CD8+ T cells (CD8A and CD8B),

B cells (CD19 and MS4A1), Monocytes (CD14 and CD16), Dendritic cells(IL3RA). Single-nucleus RNA-seq datasets were obtained from The Rush Alzheimer's Disease Center (RADC) Research Resource Sharing Hub at Synapse (https://www.synapse.org/#!Synapse:syn18485175) with the DOI 10.7303/syn18485175. The ROSMAP metadata can be found at https://www.synapse.org/#!Synapse:syn3157322. Details of individual samples, including those with no pathology (13 individuals), AD-early pathology (8 individuals), and AD-late pathology (5 individuals), utilized in the analysis of this study are outlined in **Supplementary Table 9**. In Alzheimer's disease samples, we assigned cell types by marker genes: excitatory neurons (NRGN), inhibitory neurons (GAD1), astrocytes (AQP4), oligodendrocytes (MBP), microglia (CSF1R and CD14) and oligodendrocyte progenitor cells (VCAN).

Results

Nanopore Direct RNA sequencing for TERRA transcripts

To investigate TERRA transcripts in human cells, we captured RNA molecules containing UUAGGG repeats by using biotinylated antisense oligos in U2OS cells (**Figure 1A**). The TERRA-enriched RNA samples were subjected to two different sequencing methods: Illumina paired-end sequencing and Nanopore direct RNA sequencing. For Illumina sequencing, TERRA-enriched RNA was reverse transcribed using either random hexamers (random group) or telomere-specific primers (specific group) for cDNA library construction. For Nanopore Direct RNA sequencing, enriched TERRA molecules were polyadenylated by poly(A) polymerase before library preparation (**Figure 1A**). Approximately, a few hundred to over two-thousand-fold enrichment was observed after TERRA capture (**Supplementary Figure S1A**). The reads from both methods were aligned to the T2T-CHM13 human genome reference, which includes nearly complete coverage of the entire genome including subtelomeric and telomeric regions (51).

TERRA transcription regions at chromosome ends

The criteria for identification of TERRA transcription regions included the observations of Nanopore reads, CAGE (Cap Analysis of Gene Expression) tags, and TERRA enrichment in the random or specific groups from TERRA-capture RNA-seq (**Figure 1B, Supplementary Figure S1B**). Compared to ribosomal depleted RNA-seq (control), TERRA-capture RNA-seq (enriched by antisense oligos) exhibited an increase in read counts derived from chromosome ends, shown by TERRA enrichment in log2 values (**Figure 1C**). In cases where no CAGE-seq tags mapped to the subtelomeric regions, we estimated transcription start sites (TSS) based on subtelomeric regions with Nanopore long reads and TERRA enrichment. A total of 39 TERRA transcription regions from chromosome ends were identified (**Supplementary Tables 1, 2**). Remarkably, Nanopore and Illumina sequencing data displayed the presence of TERRA transcription across a majority of chromosome termini in U2OS cells (**Figure 1D**). The results aligned well with a recent study showing that TERRA can be detected in most chromosome ends in human cell lines (52). To further verify the TERRA transcription regions, we performed TERRA-capture RNA-seq in another

cell line (HeLa). The TERRA enrichment was also observed in defined TERRA transcription regions at 39 chromosome ends in HeLa cells (**Supplementary Figure S2**), showing the consistency of TERRA transcription regions between two different cell lines.

Notably, most CAGE tags were mapped to the 37 bp repeat regions (**Supplementary Figure S3**). In agreement with this, TERRA nanopore reads were aligned with the 37 bp repeats at their 5' termini (**Supplementary Figure S3**). These results suggest that the regions preceding the 37 bp repeats may function as potential TERRA promoters, supporting previous studies that showed the importance of the 29 bp repeat regions in regulating transcription activity (4,21).

We classified three distinct types of TERRA transcription regions according to their locations and promoters (**Figure 1B**). The majority of telomeres (37 out of 46 chromosome ends) were characterized as Type I TERRA transcription regions by the presence of 61-29-37 bp repeats at the promoter (**Supplementary Figure S3**). Only chr10p and chr17p termini were characterized as Type II TERRA transcription region, which lacks 61-29-37 bp repeats (**Figure 1B, Supplementary Figure S3**). TERRA transcription regions with 61-29-37 bp repeats or CAGE tags were summarized in **Supplementary Tables 1 and 2**. Not all Type I TERRA promoters encompass all three repetitive motifs. 16 chromosome ends are composed of all three repetitive motifs, while 21 chromosome ends lack either one or two repetitive motifs (**Supplementary Figure 4A**, **Supplementary Table 3**). To investigate the impact of the 29 bp repeat element on TERRA expression, we compared TERRA levels at chromosome ends with or without the 29 bp repeats. Our analysis revealed no significant difference in TERRA expression between these regions (**Supplementary Figure S4B**), implying the presence of an additional promoter that lacks the 29 bp repeats for TERRA transcription.

TERRA transcription from interstitial telomeric repeats

In addition to TERRA transcription at chromosome ends, TERRA can be transcribed from interstitial telomeric sequences (ITS), classified as Type III TERRA transcription regions (**Figure 1B**, **Supplementary Figure S3, Supplementary Table 4**). In the T2T-CHM13 genome, we identified 79 ITS sites containing telomeric repeats longer than 200 base pairs. Among these, 22 sites showed TERRA enrichment and Nanopore reads containing long telomeric repeats at the 3' ends. Notably,

Type III TERRA transcription regions exhibit CAGE tags, and the promoters usually comprise the 37 bp repeats without 29 bp repeats (**Supplementary Figure S3)**. These results indicate that most TERRA transcripts originated from the 37 bp repeat regions, regardless of whether TERRA derived from chromosome ends or ITS. Genome browser views demonstrated some examples of TERRA transcription regions carrying these features (**Supplementary Figure S3**).

Interestingly, we observed antisense TERRA transcripts, named ARIA (53) in Nanopore direct RNA-seq data. ARIA reads were mapped to chromosome 2, a region containing an ancestral telomere-telomere fusion site (54). The 5' ends of ARIA reads comprise C-rich telomeric repeats near the junction of the telomere-telomere fusion site, following the unique sequences aligned to chromosome 2 (**Supplementary Figure S4D**). TERRA transcripts were also observed near the junction but were not directly connected to ARIA transcripts (**Supplementary Figure S4D**).

The length of telomeric repeats in TERRA transcripts

We sought to elucidate telomeric repeat length within TERRA molecules, and examined the telomeric repeat length in Nanopore reads that aligned with the subtelomeric regions and extended into the telomeric repeats. The length of the telomeric repeats in TERRA transcripts was determined by identifying the longest telomeric sequences from reads mapped to TERRA transcription regions (**Supplementary Figure 5A**). The analysis indicates telomeric repeat length within TERRA raw reads, ranging from several hundred nucleotides to over one thousand nucleotides.

TERRA promoters are associated with H3K4me3, Pol II, CpGs, and R-loops at distinct repeat elements

Next, we searched for common features among the transcription start sites of TERRA. ChIP-seq data for active transcription marks, including RNA polymerase II (39) and H3K4me3 (38), along with CTCF (40) and DNA methylation (41) were analyzed using T2T-CHM13 reference genome. Additionally, DRIP-seq data (42) was investigated to examine the profiles of R-loops surrounding the 61-29-37 bp repeats of TERRA promoters located in the subtelomeric regions (Type I). The meta-analysis revealed epigenetic profiles corresponding to each repeat element (**Figure 2A**).

The 61 bp repeat element exhibited an enrichment of CTCF, RNA polymerase II, and R loops. The 29 bp repeat element showed an enrichment of DNA methylation compared to other repeats. Remarkably, CAGE tags and H3K4me3 were enriched at the 37 bp repeat regions. The distribution of analyzed epigenetic marks along the TERRA promoter was illustrated in **Figure 2B**, and an example of the genomic view of these marks was shown in **Figure 2C**. We conclude that DNA methylation predominantly occurs at 29 bp repeats, while the active marks such as RNA polymerase II and H3K4me3 are enriched at 61 bp and 37 bp repeats respectively.

Our observations suggest that TERRA are transcribed mainly from the 61-29-37 bp repeat regions. Only chr10p and chr17p do not contain the 61-29-37 repeats adjacent to TERRA transcription regions. It's worth noting that while certain TERRA transcription regions lack 29 bp repeats, they still consist of 61-37 bp repeats. Intriguingly, even in regions without 29 bp repeats, DNA methylation was still detected in the CpG islands between the 61 and 37 bp repeat tracts (**Supplementary Figure S5B**).

To investigate the correlation between these epigenetic marks and TERRA levels, we compared TERRA expression with the abundance of these epigenetic marks at TERRA promoters. The promoter regions were selected from 3 kb upstream of the transcription start sites, where the 61-29-37 bp repeats were located. Scatter plots demonstrated that TERRA expression is positively correlated with H3K4me3, R-loops, and CTCF at chromosome ends (**Figure 2D**).

TERRA-QUANT for measuring TERRA expression

To analyze TERRA expression using publicly available short-read RNA-seq datasets, we established a bioinformatics tool, dubbed TERRA-QUANT, for measuring TERRA levels based on the transcription regions defined in this study (**Figure 3A**). TERRA-QUANT integrates reads from HTseq-count and TelomereHunter, with the merged reads subsequently undergoing YARN normalization. This method selects reads aligned to TERRA transcription regions, consisting of at least 2 consecutive telomeric repeats to be classified as TERRA reads. The idea is to quantify TERRA fragments covering the regions between subtelomeric regions and telomeric repeats.

To test the reliability of this tool, we applied TERRA-QUANT to analyze TERRA expression in HeLa cells using Illumina sequencing data obtained from a previous study (32). The previous

study enriched TERRA RNA by using biotinylated antisense oligos to capture TERRA from nuclear RNA. Notably, TERRA read counts were detected at the majority of chromosome ends in the TERRA capture group, whereas they were absent at multiple chromosome ends in the control group without TERRA capture (**Figures 3B-3C**). TRF2 deficiency led to increased TERRA read counts at multiple chromosome ends compared to wildtype cells (**Figures 3B-3C**). Additionally, TERRA read counts in HeLa cells were significantly lower than those in U2OS cells (**Figure 3C**). These results are consistent with earlier studies (4,32), affirming the robustness of the TERRA-QUANT methodology.

To further validate the annotations for TERRA transcription regions, we designed subtelomeric primer sets specific to CHM13-1q, 3q, 8p, and 15q chromosome ends based on the T2T-CHM13 genome sequence (**Supplementary Table 6**). RT-qPCR analysis using these subtelomeric primers revealed increased TERRA levels from various chromosome ends in U2OS compared to HeLa cells (**Figure 3D**). Subsequently, the RT-qPCR products were subjected to Illumina sequencing to validate their sequences. We observed that the reads of RT-qPCR products were predominantly mapped to their target chromosome ends. These results indicate the specificity of these subtelomeric primers (**Supplementary Figure 5C**).

ALT positive osteosarcomas show elevated TERRA levels

We gathered RNA-seq datasets from a previous study, which included three biological replicates of 13 human osteosarcoma cell lines, with 4 cell lines categorized as ALT negative (n=12) and 9 cell lines classified as ALT positive (n=27) (46). The RNA-seq reads of these datasets were aligned to the T2T-CHM13 genome and processed using TERRA-QUANT. Total TERRA levels were then determined by cumulating TERRA reads from all chromosome ends. Notably, ALT-positive cells exhibited a higher level of total TERRA, and a lower RNA level of telomerase reverse transcriptase (TERT) (**Figures 4A-4B**). These results agree with the previous finding showing ALT cells with elevated TERRA expression (2).

TERRA increases along with human aging in blood cells

To compare TERRA expression across different ages, we collected RNA-seq datasets from a

previous study (47), which comprised 222 human blood samples for analysis. These blood samples cover a range of ages from 24 to 73 years old and were separated into three groups: young (<30 years old), mid-age (30-59 years old), and old groups (≥60 years old). Interestingly, TERRA-QUANT analysis demonstrated that total TERRA levels were positively correlated with age (**Figure 4C**). When grouped by different ages, total TERRA levels were significantly increased in old individuals, compared to young and mid-age individuals (**Figure 4D**).

Next, we confirmed the association of TERRA levels with aging by collecting blood samples from individuals of various ages and conducted quantitative PCR to measure telomere length and TERRA levels. The samples were grouped by young (21-65y) or old (>65y) ages. Genomic DNA was extracted from buffy coats, and telomere length (T/S ratio) was analyzed by quantitative PCR. We observed that telomere lengths significantly decreased in elderly people in both males and females (**Figure 4E**). Interestingly, total TERRA RNA levels in blood samples determined by RT-qPCR were significantly elevated in old males and females (**Figure 4F**). Using subtelomeric primers for RT-qPCR also showed a significant increase in TERRA expression in the old group compared to the young group (**Figure 4G**), confirming an elevation in TERRA levels in aged leukocytes.

TERRA is upregulated in the aged brain

To explore the relationship between TERRA and human aging, we analyzed published RNA-seq datasets obtained from various human tissues including brains, hearts, and ovaries (47) by TERRA-QUANT. The scatter plots displayed a significant positive correlation between total TERRA levels and age in brains, but not in ovaries or hearts (**Figure 5A**). Remarkably, total TERRA levels in brain tissues exhibited an upregulation in the old group (≥70 years old), compared to the adult group (30-69 years old) (**Figure 5B**). Considering that ATRX is important for neuronal development and functionality (55-57), we analyzed ATRX expression in human brains. Interestingly, the levels of ATRX, which is a TERRA-interacting protein (24), declined in old individuals (**Figure 5C**) and displayed an anticorrelation with TERRA expression in brain tissues (**Figure 5D**).

Elevated TERRA levels in aged fibroblasts and abnormal TERRA profiles in Hutchinson-Gilford progeria syndrome patients

Next, we quantified TERRA expression in fibroblasts obtained from healthy individuals of various ages and patients diagnosed with Hutchinson-Gilford progeria syndrome (HGPS), a genetic disorder characterized by premature aging features in childhood (58). Total TERRA levels, analyzed by TERRA-QUANT, displayed a significant upregulation in fibroblasts derived from healthy older individuals (**Figures 5E-5F**), with an exponential increase observed in those over 75 years old. Considering that HGPS patients were below 30 years old, we compared TERRA levels with those of their age-matched controls. Notably, HGPS patients displayed an aberrant and diverse TERRA expression pattern compared to healthy individuals (**Figure 5G**).

Single-cell RNA-seq analysis reveals high TERRA levels in neurons

We assessed TERRA levels across various cell types by analyzing published RNA-seq datasets from human tissues, including adipose, blood, bone, brain, heart, lung, ovary, pancreas, and retina (**Supplementary Table 7**). Among these tissues, TERRA levels were particularly high in blood, bone, ovary, and retina (**Figure 6A**). Given the highest level in blood cells, we sought to identify the specific cell types exhibiting high TERRA expression. To achieve this, we analyzed single-cell RNA-seq datasets (59) from peripheral blood mononuclear cells (PBMC) obtained from healthy young individuals. Notably, dendritic cells displayed the highest TERRA level among various immune cells (**Figure 6B**).

To investigate the potential role of TERRA in neurons, we examined TERRA levels using single-cell RNA-seq datasets obtained from human embryonic stem cells (hESCs) undergoing neuronal differentiation (60). The hESCs were converted to neuromesodermal progenitors (NMPs) with distinct HOX profiles along rostrocaudal axis by modulation of Wnt, FGF and retinoic acid signaling, and were further differentiated into ventral neuron phenotypes. Six different NMP cultures from hESCs were generated corresponding to HOX patterning periods: 24, 48, 72, 120, 168, and 216 hours. The cultures exhibited a progressive expression of caudal HOX paralogs, corresponding to cervical (HOX1-8; observed in 24h,48h, and 72h), thoracic (HOX1-9; observed in 120h), lumbar (HOX1-11; observed in 168h), and lumbosacral spinal regions (HOX1-13;

observed in 216h). Cells were clustered based on gene expression into mesenchymal cells, neuroepithelial cells, and neurons. Notably, higher levels of TERRA were observed in neurons compared to neuroepithelial cells and mesenchymal cells (**Figure 6C**). Additionally, TERRA levels displayed a concordant expression pattern with ATRX during neuronal differentiation (**Figures 6D-6F**).

Elevated TERRA levels during the early stage of Alzheimer's disease progression

Next, we sought to analyze TERRA levels using TERRA-QUANT in neuronal disorder diseases. We analyzed published single-nucleus RNA-seq datasets from the prefrontal cortex of patients with Alzheimer's disease (AD) (61). Postmortem human brain samples were obtained from participants enrolled in the Religious Order Study and Rush Memory and Aging Project (ROSMAP) (62). Individuals with AD pathology were categorized into two subgroups according to the previous study (61), reflecting distinct stages of AD pathological progression: 1) early-pathology characterized by amyloid burden, modest neurofibrillary tangles, and cognitive impairment; and 2) late-pathology marked by higher amyloid levels, elevated neurofibrillary tangles, increased global pathology, and cognitive impairment. Notably, increased TERRA levels were observed in various types of brain cells in the early stage of AD pathology compared to no AD pathology, with more profound TERRA levels in excitatory neurons (**Figure 6G, Supplementary Figure 6A**).

Discussion

In this study, we investigated TERRA transcripts in human cells using Nanopore direct RNA-seq and TERRA capture RNA-seq. We identified 39 TERRA transcription regions at chromosome ends and utilized these defined regions to quantify TERRA expression across various tissues. Our bioinformatics pipeline is versatile, accommodating different RNA-seq datasets, including single-cell RNA-seq experiments. Notably, substantial mapped reads per cell were observed in single-cell RNA-seq datasets without YARN normalization (**Supplementary Figures 6A-6C**), indicating that TERRA can be detected at the single-cell level.

Commonly available datasets largely comprise RNA samples enriched with poly-A tails. Nevertheless, a significant portion of TERRA molecules lack poly-A tails, which limits the detection

of TERRA using poly-A enriched RNA-seq data. Here, we accumulated TERRA reads from all chromosome ends to enhance read counts, thereby increasing overall TERRA counts for quantifying total TERRA levels. Due to low read counts mapped to individual chromosome ends, analyzing chromosome-end-specific TERRA levels remains challenging using publicly available RNA-seq datasets. TERRA-enriched sequencing and longer read lengths could enable more precise quantification of TERRA at specific chromosome ends.

Utilizing TERRA-QUANT for quantifying TERRA levels across diverse tissues, we identified the presence of TERRA in most tissues. High TERRA levels were observed in blood and bone tissues, suggesting a potential role of TERRA in immune responses. This aligns with earlier research indicating that TERRA interacts with ZBP1 to activate immune response during replicative crisis (63). Moreover, shorter TERRA can be secreted into the extracellular environment through exosomes to induce inflammatory cytokines (64). Notably, we observed elevated total TERRA levels in fibroblasts, brain, and blood cells as individuals aged. The telomere shortening may contribute to TERRA upregulation during aging. However, neurons are postmitotic cells that are not undergoing telomere shortening. Therefore, the increase of TERRA in the brain is unlikely due to shortened telomeres. The underlying mechanism needs further investigation. In addition, fibroblasts derived from HGPS patients demonstrated abnormal TERRA expression, with some individuals showing high levels while others displayed lower levels. We reason that the diverse TERRA counts observed in HGPS patients could be attributed to variable telomere lengths or defects in telomere integrity. Previous studies have shown that telomere length in HGPS patients is generally shorter, yet exhibits variability (65,66). Low TERRA counts in some HGPS patients could result from either low TERRA expression or the loss of subtelomeric and telomeric DNA at chromosome ends.

Examination of single-nucleus RNA-seq datasets unveiled an upregulation of TERRA levels in neurons of the prefrontal cortex during the early stages of Alzheimer's disease progression. These intriguing findings raise questions about the potential association between elevated TERRA levels and changes in chromatin status near telomeres, increased oxidative stress, chronic inflammation, or the occurrence of damaged telomeres in aged cells.

Upregulation of TERRA and ATRX during neuronal differentiation implies the potential function of TERRA in neuronal cells. Interestingly, the inverse relationship between TERRA and ATRX in the aged brain was observed, suggesting that low levels of ATRX might contribute to the dysfunction of the aged neurons. Previous research indicated that the absence of ATRX leads to cognitive impairment and promotion of cellular senescence (67), further highlighting the significance of exploring the interplay between TERRA, ATRX, and aging-related cellular processes in the brain and neuronal tissues. It is plausible that proper levels of TERRA and ATRX are crucial for maintaining normal gene expression in neuronal cells, and dysregulation of TERRA and ATRX may impact neuron function during aging.

Our study provides the annotation of TERRA transcripts in the T2T-CHM13 genome reference and introduces the bioinformatics tool “TERRA-QUANT” for quantifying TERRA levels. This lays the foundation for studying TERRA in diverse conditions, cell types, and human diseases.

Data availability

TERRA capture RNA-seq datasets including Illumina short-read RNA-seq, Nanopore direct RNA-seq, and the sequencing data of TERRA RT-qPCR products are accessible at GEO for reviewers (GEO accession GSE250303): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250303. Enter token mrwvskqonlutzaf into the box. The published sequencing datasets analyzed in this study are described in **Supplementary Tables 7, 8, and 9**.

Code availability

A copy of the custom code utilized for quantification of TERRA is accessible on GitHub (https://github.com/ls807terra/TERRA_RNA-seq_pipeline). Publicly available packages were used in this study as indicated in the Methods section.

Acknowledgments

We appreciated the insightful discussions with Liuh-Yow Chen, Ching-Shyi Peter Wu, as well as the contributions of our lab members Hong-Jhin Shen, Chia-Yu Guh, and Liv Weichien Chen. This work was supported by National Science and Technology Council (NSTC) grants NSTC 112-2628-B-002-008 (H-P. C. C.), NSTC 112-2320-B-002-058 (H-P. C. C.), and National Taiwan University grants NTU-111L7880 (H-P. C. C.), NTU-AS-112L104312 (H-P. C. C), NTU-CDP-112L7721 (H-P. C. C.), NTU-CDP-113L7705 (H-P. C. C.), and National Health Research Institutes grants NHRI-EX111-11107SI, NHRI-EX112-11107SI (H-P. C. C.). We thank Technology Commons, College of Life Science, National Taiwan University for supporting the usage of equipment, and National Center for High-performance Computing (NCHC) for providing computational and storage resources (MST109178).

Author contributions

H-P. C. C. and Y-H. H. conceived of and designed the study. Y-H. H. conducted TERRA capture, library constructions for sequencing, and data analysis. C-H. T., M-T. Y. P-C. Y, analyzed RNA-seq datasets. Y-C. C. performed qPCR for blood cells. C-P. Y. prepared samples for experiments. D-S H. collected blood samples. H-P. C. C., Y-H. H., C-H. T., P-C. Y, and M-T. Y. wrote the manuscript.

Completing interests

The authors declare that they have no competing interests.

References

1. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798-801.
2. Schoeftner, S. and Blasco, M.A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*, **10**, 228-236.
3. Feretzaki, M., Renck Nunes, P. and Lingner, J. (2019) Expression and differential regulation of human TERRA at several chromosome ends. *RNA*, **25**, 1470-1480.
4. Nergadze, S.G., Farnung, B.O., Wischnewski, H., Khoriauli, L., Vitelli, V., Chawla, R., Giulotto, E. and Azzalin, C.M. (2009) CpG-island promoters drive transcription of human telomeres. *RNA*, **15**, 2186-2194.
5. Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira, M.T. and Luke, B. (2017) Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell*, **170**, 72-85 e14.
6. Azzalin, C.M. and Lingner, J. (2008) Telomeres: the silence is broken. *Cell Cycle*, **7**, 1161-1165.
7. Savoca, V., Rivosecchi, J., Gaiatto, A., Rossi, A., Mosca, R., Gialdini, I., Zubovic, L., Tebaldi, T., Macchi, P. and Cusanelli, E. (2023) TERRA stability is regulated by RALY and polyadenylation in a telomere-specific manner. *Cell Rep*, **42**, 112406.
8. Rossi, A., Moro, A., Tebaldi, T., Cornella, N., Gasperini, L., Lunelli, L., Quattrone, A., Viero, G. and Macchi, P. (2017) Identification and dynamic changes of RNAs isolated from RALY-containing ribonucleoprotein complexes. *Nucleic Acids Res*, **45**, 6775-6792.
9. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K. and Selig, S. (2008) Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum Mol Genet*, **17**, 2776-2789.
10. Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chedin, F. and Selig, S. (2017) Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat Commun*, **8**, 14015.
11. Deng, Z., Campbell, A.E. and Lieberman, P.M. (2010) TERRA, CpG methylation and telomere heterochromatin: lessons from ICF syndrome cells. *Cell Cycle*, **9**, 69-74.
12. Deng, Z., Wang, Z., Stong, N., Plasschaert, R., Moczan, A., Chen, H.S., Hu, S., Wikramasinghe, P., Davuluri, R.V., Bartolomei, M.S. *et al.* (2012) A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J*, **31**, 4165-4178.
13. Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J.H., Sung, P.A., 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.
14. Diman, A., Boros, J., Poulain, F., Rodriguez, J., Purnelle, M., Episkopou, H., Bertrand, L., Francaux, M., Deldicque, L. and Decottignies, A. (2016) Nuclear respiratory factor 1 and endurance exercise promote human telomere transcription. *Sci Adv*, **2**, e1600031.

15. Tutton, S., Azzam, G.A., Stong, N., Vladimirova, O., Wiedmer, A., Monteith, J.A., 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.
16. Koskas, S., Decottignies, A., Dufour, S., Pezet, M., Verdel, A., Vourc'h, C. and Faure, V. (2017) Heat shock factor 1 promotes TERRA transcription and telomere protection upon heat stress. *Nucleic Acids Res*, **45**, 6321-6333.
17. Mazzolini, R., Gonzalez, N., Garcia-Garijo, A., Millanes-Romero, A., Peiro, S., Smith, S., Garcia de Herreros, A. and Canudas, S. (2018) Snail1 transcription factor controls telomere transcription and integrity. *Nucleic Acids Res*, **46**, 146-158.
18. Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T. and Azzalin, C.M. (2014) RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun*, **5**, 5220.
19. Yu, T.Y., Kao, Y.W. and Lin, J.J. (2014) Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc Natl Acad Sci U S A*, **111**, 3377-3382.
20. Guh, C.Y., Shen, H.J., Chen, L.W., Chiu, P.C., Liao, I.H., Lo, C.C., Chen, Y., Hsieh, Y.H., Chang, T.C., Yen, C.P. *et al.* (2022) XPF activates break-induced telomere synthesis. *Nat Commun*, **13**, 5781.
21. Silva, B., Arora, R., Bione, S. and Azzalin, C.M. (2021) TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nature Communications*, **12**.
22. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. and Lieberman, P.M. (2009) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Molecular Cell*, **35**, 403-413.
23. Mei, Y., Deng, Z., Vladimirova, O., Gulve, N., Johnson, F.B., Drosopoulos, W.C., Schildkraut, C.L. and Lieberman, P.M. (2021) TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. *Sci Rep-Uk*, **11**.
24. Chu, H.P., Cifuentes-Rojas, C., Kesner, B., Aeby, E., Lee, H.G., Wei, C., Oh, H.J., Boukhali, M., Haas, W. and Lee, J.T. (2017) TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell*, **170**, 86-101 e116.
25. Flynn, R.L., Centore, R.C., O'Sullivan, R.J., Rai, R., Tse, A., Zhou, S.Y., Chang, S., Karlseder, J. and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature*, **471**, 532-+.
26. Voon, H.P., Hughes, J.R., Rode, C., De La Rosa-Velazquez, I.A., Jenuwein, T., Feil, R., Higgs, D.R. and Gibbons, R.J. (2015) ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Rep*, **11**, 405-418.
27. Tsai, R.X., Fang, K.C., Yang, P.C., Hsieh, Y.H., Chiang, I.T., Chen, Y., Lee, H.G., Lee, J.T. and Chu, H.C. (2022) TERRA regulates DNA G-quadruplex formation and ATRX recruitment to chromatin. *Nucleic Acids Res*, **50**, 12217-12234.
28. Arnoult, N., Van Beneden, A. and Decottignies, A. (2012) Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1alpha. *Nat Struct Mol Biol*, **19**, 948-956.

29. Cusanelli, E., Romero, C.A. and Chartrand, P. (2013) Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell*, **51**, 780-791.
30. Misino, S., Busch, A., Wagner, C.B., Bento, F. and Luke, B. (2022) TERRA increases at short telomeres in yeast survivors and regulates survivor associated senescence (SAS). *Nucleic Acids Res*, **50**, 12829-12843.
31. Caslini, C., Connelly, J.A., Serna, A., Broccoli, D. and Hess, J.L. (2009) MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. *Mol Cell Biol*, **29**, 4519-4526.
32. Porro, A., Feuerhahn, S., Delafontaine, J., Riethman, H., Rougemont, J. and Lingner, J. (2014) Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun*, **5**, 5379.
33. Aguado, J., Sola-Carvajal, A., Cancila, V., Revechon, G., Ong, P.F., Jones-Weinert, C.W., Wallen Arzt, E., Lattanzi, G., Dreesen, O., Tripodo, C. *et al.* (2019) Inhibition of DNA damage response at telomeres improves the detrimental phenotypes of Hutchinson-Gilford Progeria Syndrome. *Nat Commun*, **10**, 4990.
34. Maestroni, L., Reyes, C., Vaurs, M., Gachet, Y., Tournier, S., Geli, V. and Coulon, S. (2020) Nuclear envelope attachment of telomeres limits TERRA and telomeric rearrangements in quiescent fission yeast cells. *Nucleic Acids Res*, **48**, 3029-3041.
35. Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K. and Luke, B. (2013) Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol*, **20**, 1199-1205.
36. Wanat, J.J., Logsdon, G.A., Driskill, J.H., Deng, Z., Lieberman, P.M. and Johnson, F.B. (2018) TERRA and the histone methyltransferase Dot1 cooperate to regulate senescence in budding yeast. *PLoS One*, **13**, e0195698.
37. Liu, B., Maekawa, T., Yoshida, K., Ly, N.H., Inoue, K., Hasegawa, A., Chatton, B., Ogura, A. and Ishii, S. (2019) Telomere shortening by transgenerational transmission of TNF-alpha-induced TERRA via ATF7. *Nucleic Acids Res*, **47**, 283-298.
38. Jang, S.M., Zhang, Y., Utani, K., Fu, H., Redon, C.E., Marks, A.B., Smith, O.K., Redmond, C.J., Baris, A.M., Tulchinsky, D.A. *et al.* (2018) The replication initiation determinant protein (RepID) modulates replication by recruiting CUL4 to chromatin. *Nat Commun*, **9**, 2782.
39. Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B., Bernard, X., Masamsetti, V.P., Roston, S. *et al.* (2018) EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature*, **555**, 387-391.
40. Ibarra, A., Benner, C., Tyagi, S., Cool, J. and Hetzer, M.W. (2016) Nucleoporin-mediated regulation of cell identity genes. *Genes Dev*, **30**, 2253-2258.
41. Wang, Y., Khan, A., Marks, A.B., Smith, O.K., Giri, S., Lin, Y.C., Creager, R., MacAlpine, D.M., Prasanth, K.V., Aladjem, M.I. *et al.* (2017) Temporal association of ORCA/LRWD1 to late-firing origins during G1 dictates heterochromatin replication and organization. *Nucleic Acids Res*, **45**, 2490-2502.
42. De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O. and Capranico, G. (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc Natl Acad Sci U S A*, **116**, 816-825.

43. Feuerbach, L., Sieverling, L., Deeg, K.I., Ginsbach, P., Hutter, B., Buchhalter, I., Northcott, P.A., Mughal, S.S., Chudasama, P., Glimm, H. *et al.* (2019) TelomereHunter - in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics*, **20**, 272.
44. Paulson, J.N., Chen, C.Y., Lopes-Ramos, C.M., Kuijjer, M.L., Platig, J., Sonawane, A.R., Fagny, M., Glass, K. and Quackenbush, J. (2017) Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics*, **18**, 437.
45. Chang, K.V., Chen, Y.C., Wu, W.T., Shen, H.J., Huang, K.C., Chu, H.P. and Han, D.S. (2020) Expression of Telomeric Repeat-Containing RNA Decreases in Sarcopenia and Increases after Exercise and Nutrition Intervention. *Nutrients*, **12**.
46. Mason-Osann, E., Dai, A., Floro, J., Lock, Y.J., Reiss, M., Gali, H., Matschulat, A., Labadorf, A. and Flynn, R.L. (2018) Identification of a novel gene fusion in ALT positive osteosarcoma. *Oncotarget*, **9**, 32868-32880.
47. Shokhirev, M.N. and Johnson, A.A. (2021) Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell*, **20**, e13280.
48. Hicks, S.C., Okrah, K., Paulson, J.N., Quackenbush, J., Irizarry, R.A. and Bravo, H.C. (2018) Smooth quantile normalization. *Biostatistics*, **19**, 185-198.
49. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, **8**, 14049.
50. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C. *et al.* (2023) Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*.
51. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A. *et al.* (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, **585**, 79-84.
52. Rodrigues, J., Alfieri, R., Bione, S. and Azzalin, C.M. (2024) TERRA ONTseq: a long-read-based sequencing pipeline to study the human telomeric transcriptome. *RNA*, **30**, 955-966.
53. Azzalin, C.M. and Lingner, J. (2015) Telomere functions grounding on TERRA firma. *Trends Cell Biol*, **25**, 29-36.
54. JW, I.J., Baldini, A., Ward, D.C., Reeders, S.T. and Wells, R.A. (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci U S A*, **88**, 9051-9055.
55. Gibbons, R.J., Picketts, D.J., Villard, L. and Higgs, D.R. (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell*, **80**, 837-845.
56. Quesnel, K.M., Martin-Kenny, N. and Berube, N.G. (2023) A mouse model of ATRX deficiency with cognitive deficits and autistic traits. *J Neurodev Disord*, **15**, 39.
57. Berube, N.G., Mangelsdorf, M., Jagla, M., Vanderluit, J., Garrick, D., Gibbons, R.J., Higgs, D.R., Slack, R.S. and Picketts, D.J. (2005) The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. *J Clin Invest*, **115**, 258-267.
58. Gilford, H. (1897) On a Condition of Mixed Premature and Immature Development. *Med Chir Trans*, **80**, 17-46 25.

59. Luo, O.J., Lei, W., Zhu, G., Ren, Z., Xu, Y., Xiao, C., Zhang, H., Cai, J., Luo, Z., Gao, L. *et al.* (2022) Multidimensional single-cell analysis of human peripheral blood reveals characteristic features of the immune system landscape in aging and frailty. *Nat Aging*, **2**, 348-364.
60. Iyer, N.R., Shin, J., Cuskey, S., Tian, Y., Nicol, N.R., Doersch, T.E., Seipel, F., McCalla, S.G., Roy, S. and Ashton, R.S. (2022) Modular derivation of diverse, regionally discrete human posterior CNS neurons enables discovery of transcriptomic patterns. *Sci Adv*, **8**, eabn7430.
61. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X. *et al.* (2019) Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, **570**, 332-337.
62. Bennett, D.A., Buchman, A.S., Boyle, P.A., Barnes, L.L., Wilson, R.S. and Schneider, J.A. (2018) Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis*, **64**, S161-S189.
63. Nassour, J., Aguiar, L.G., Correia, A., Schmidt, T.T., Mainz, L., Przetocka, S., Haggblom, C., Tadepalle, N., Williams, A., Shokhirev, M.N. *et al.* (2023) Telomere-to-mitochondria signalling by ZBP1 mediates replicative crisis. *Nature*, **614**, 767-773.
64. Wang, Z., Deng, Z., Dahmane, N., Tsai, K., Wang, P., Williams, D.R., Kossenkov, A.V., Showe, L.C., Zhang, R., Huang, Q. *et al.* (2015) Telomeric repeat-containing RNA (TERRA) constitutes a nucleoprotein component of extracellular inflammatory exosomes. *Proc Natl Acad Sci U S A*, **112**, E6293-6300.
65. Decker, M.L., Chavez, E., Vulto, I. and Lansdorp, P.M. (2009) Telomere length in Hutchinson-Gilford progeria syndrome. *Mech Ageing Dev*, **130**, 377-383.
66. Li, Y., Zhou, G., Bruno, I.G., Zhang, N., Sho, S., Tedone, E., Lai, T.P., Cooke, J.P. and Shay, J.W. (2019) Transient introduction of human telomerase mRNA improves hallmarks of progeria cells. *Aging Cell*, **18**, e12979.
67. Kovatcheva, M., Liao, W., Klein, M.E., Robine, N., Geiger, H., Crago, A.M., Dickson, M.A., Tap, W.D., Singer, S. and Koff, A. (2017) ATRX is a regulator of therapy induced senescence in human cells. *Nat Commun*, **8**, 386.

Figure legends

Figure 1. Identification of TERRA transcription regions by short-read and long-read RNA-seq in human cells.

A. The flowchart of TERRA-capture RNA-seq. TERRA was captured by using biotinylated antisense oligos. Captured TERRA RNAs were subjected to Illumina or Oxford nanopore long-read RNA direct sequencing.
B. Definition of three types of TERRA transcription regions. TERRA transcription start sites were determined by TERRA enrichment from Illumina short-reads, Nanopore long reads and CAGE-seq.
C. Heatmap represents TERRA enrichment (log2 values) at different chromosome ends based on Illumina read counts of TERRA transcripts. Capture: TERRA capture RNA-seq. Control: regular RNA-seq with ribosomal RNA depletion.
D. Heatmap represents TERRA read counts from different chromosome ends. Reads from Illumina or Nanopore sequencing were mapped to T2T-CHM13. CPM: counts per million mapped reads.

Figure 2. Enrichments of H3K4me3, RNA Pol II, CpGs, DNA methylation and R-loop at TERRA promoter

A. Meta-analysis showing the distribution of epigenetic marks on 61-29-37 bp repeats located at subtelomeric regions. Each plot represents the coverage of indicated marks.
B. A schematic model of the epigenetic profiles at TERRA promoter near telomeres.
C. Genome browser view showing the coverage of indicated epigenetic marks at TERRA promoter on chr22q arm.
D. Scatter plots of TERRA v.s. H3K4me3, R-loop (DRIP-seq), DNA methylation (MeDIP-seq), CTCF, or RNAPII near TERRA transcription start sites. Each dot indicates TERRA enrichment and epigenetic marks at individual chromosome ends in U2OS cells. TERRA enrichment (log2 ratio) was calculated by comparing TERRA capture and no capture. P values, by Pearson's correlation.

Figure 3. TERRA-QUANT for measuring TERRA expression.

A. The workflow of bioinformatics pipeline for quantification of TERRA expression. Reads that mapped to TERRA transcription regions were counted. YARN was used for normalization between different tissues.
B. Heatmap showing TERRA expression from individual chromosome ends in HeLa and U2OS cells. TERRA capture, using antisense oligos to enrich TERRA. Control, no TERRA capture.
C. TERRA normalized counts were analyzed by TERRA-QUANT. Each dot indicates TERRA counts at individual chromosome ends. TRF2 deletion (TRF2KO) in HeLa cells increases TERRA expression. P values, by Wilcoxon matched-pairs signed rank test. Bars, median.
D. RT-qPCR to detect TERRA in HeLa and U2OS cells. Subtelomeric primers for TERRA were designed based on the T2T-CHM13 genome sequence. Data from three biological replicates. P values, by two-tailed Student's t-test. Bars, mean±SD.

Figure 4. TERRA levels increase with age in blood cells.

A. TERRA expression in ALT positive (+) or negative (-) cell lines was quantified by TERRA-QUANT. Total TERRA counts at all chromosome ends were accumulated. Each dot indicates the normalized TERRA read counts of each RNA-seq dataset. Bars, mean±SD. P values, by Mann-Whitney U test.
B. TERT expression of each RNA-seq data from ALT+ or ALT- cells. Bars, mean±SD. P values, by Mann-Whitney U test.
C. TERRA expression in human blood cells was analyzed by TERRA-QUANT using RNA-seq datasets. Scatter plots showing the correlation of TERRA v.s. age in blood. Each dot indicates the total TERRA normalized reads from all chromosome ends of each individual. P values, by Pearson's correlation.
D. Normalized TERRA counts in blood cells of different ages. Young (<30 years); adult (30~59 years); old (≥60 years). Bars, median with interquartile. P values, by Mann-Whitney U test.
E. The T/S ratio represents the relative telomere length (T) to the single-copy gene (S, 36B4 gene). Each dot represents an individual T/S ratio. Young (21~65 years); old (≥65 years).
F. RT-qPCR to detect total TERRA levels in blood cells using telomeric repeat primers. Each dot indicates each individual.
G. RT-qPCR to detect TERRA levels using subtelomeric primers (hg38-2q). Each dot indicates each individual. (E-G) P values, by two-tailed Student's t-test. n=sample size. Error bars are SDs.

Figure 5. TERRA expression increases with age in brain tissues and fibroblasts.

A. Scatter plots showing the correlation of TERRA v.s. age in brain, ovary and heart tissues. Each dot indicates total TERRA level from individuals. P values, by Pearson's correlation.
B. TERRA levels in various tissues with different ages. Adult (<70 years); old (≥70 years). Bars, median with interquartile. P values, by Mann-Whitney U test.
C. Boxplots of ATRX expression in brain tissues. Bars, median with interquartile. P values, by Mann-Whitney U test.
D. Scatter plots of TERRA and ATRX expression in brain tissues. P values, by Pearson's correlation.
E. TERRA levels in human fibroblasts derived from healthy individuals of different ages. P values, by Pearson's correlation.
F. Boxplots showing the upregulation of TERRA levels in old human fibroblasts. Bars, median with interquartile. P values, by Mann-Whitney U test.
G. Fibroblasts derived from HGPS patients display abnormal TERRA levels. Each dot represents the TERRA level of each individual. Bars, median with interquartile. P values, by Kolmogorov-Smirnov test.

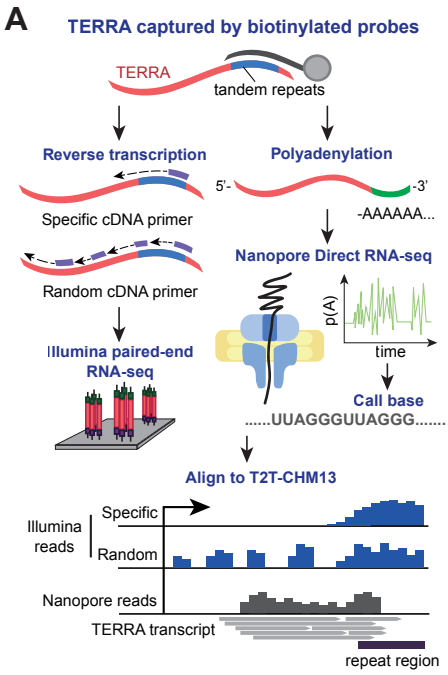
Figure 6. Single-cell analysis of TERRA levels undergoing neuronal differentiation and in Alzheimer's disease.

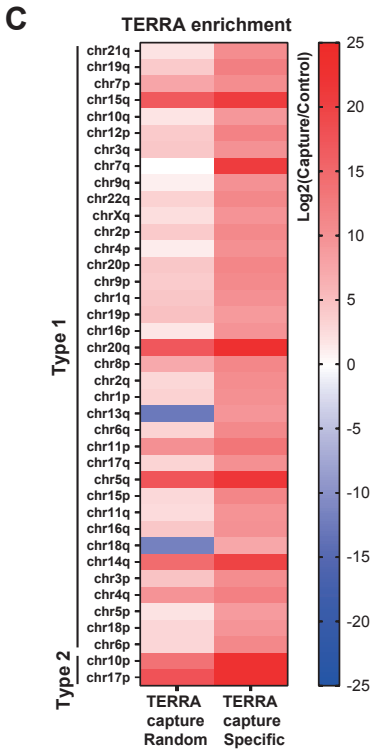
A. Violin plot showing normalized TERRA counts in various human tissues. The solid lines indicate the median and the dashed lines are interquartile range.
B. Single-cell RNA-seq showing relative TERRA levels in various immune cells from PBMC. NK, nature killer cell; DC, dendritic cells. Solid lines, median.

C. Single-cell RNA-seq analysis of relative TERRA levels in human embryonic stem cells (hESCs) undergoing neuronal differentiation with various HOX patterning periods (24h,72h,48h, 120h, 168h, 216h). MSC, mesenchymal stem cells. Solid lines, median.

D-F. Uniform manifold approximation and projection (UMAP) plots of single-cell RNA-seq datasets from hESCs undergoing neuronal differentiation using various HOX patterning periods. Colors according to TERRA, ATRX, or neuron marker MAP2 expression levels.

G. TERRA levels in cells isolated from the prefrontal cortex of individuals with Alzheimer's disease (AD). Single-nucleus RNA-seq datasets were obtained from the ROSMAP project, and grouped into healthy (no AD pathology), early-AD pathology, and late-AD pathology. P values, by Mann-Whitney U test.

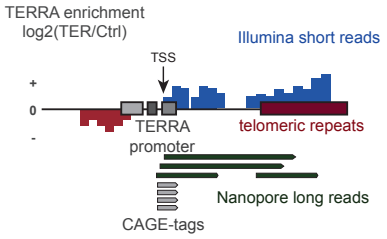




B

Types of TERRA transcription regions

Types	TSS definition
Type I 61-29-37 bp repeat-containg promoter (37/46 telomeres)	Define TSS according to TERRA enrichment signal log2(TERRA capture / Control) Nanopore reads, or CAGE-tag 5'-end.
Type II No 61-29-37 bp repeats on the promoter (chr10p and chr17p)	
Type III TERRA from interstitial telomeric sequences (22/79 ITSs contain TERRA transcripts)	



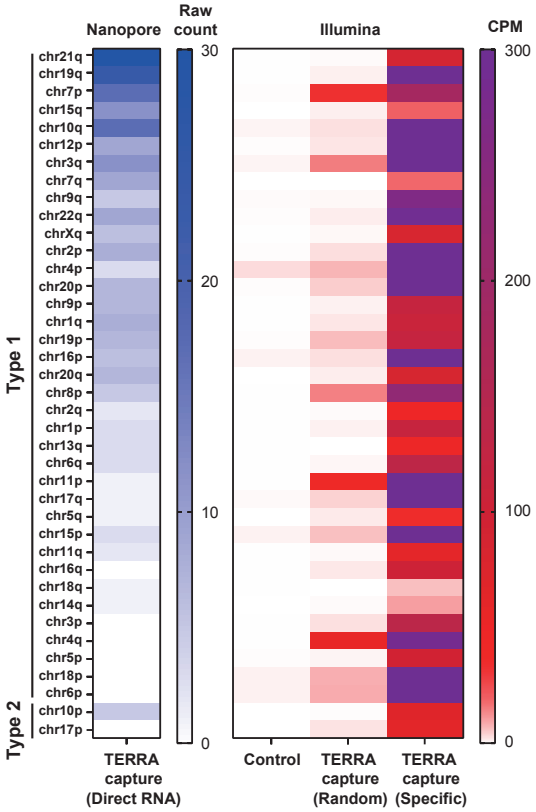


Figure 1. Identification of TERRA transcription regions by short-read and long-read RNA-sequencing in human cells.
(**A**)The flowchart of TERRA-capture RNA-seq. TERRA was captured by using biotinylated antisense oligos. Captured TERRA RNAs were subjected to Illumina or Oxford nanopore long-read RNA direct sequencing. (**B**) Definition of three types of TERRA transcription regions. TERRA transcription start sites were determined by TERRA enrichment from Illumina short-reads, Nanopore long reads and CAGE-seq. (**C**) Heatmap represents TERRA enrichment (log2 values) at different chromosome ends based on Illumina read counts of TERRA transcripts. Capture: TERRA capture RNA-seq. Control: regular RNA-seq with ribosomal RNA depletion. (**D**) Heatmap represents TERRA read counts from different chromosome ends. Reads from Illumina or Nanopore sequencing were mapped to T2T-CHM13. CPM: counts per million mapped reads.

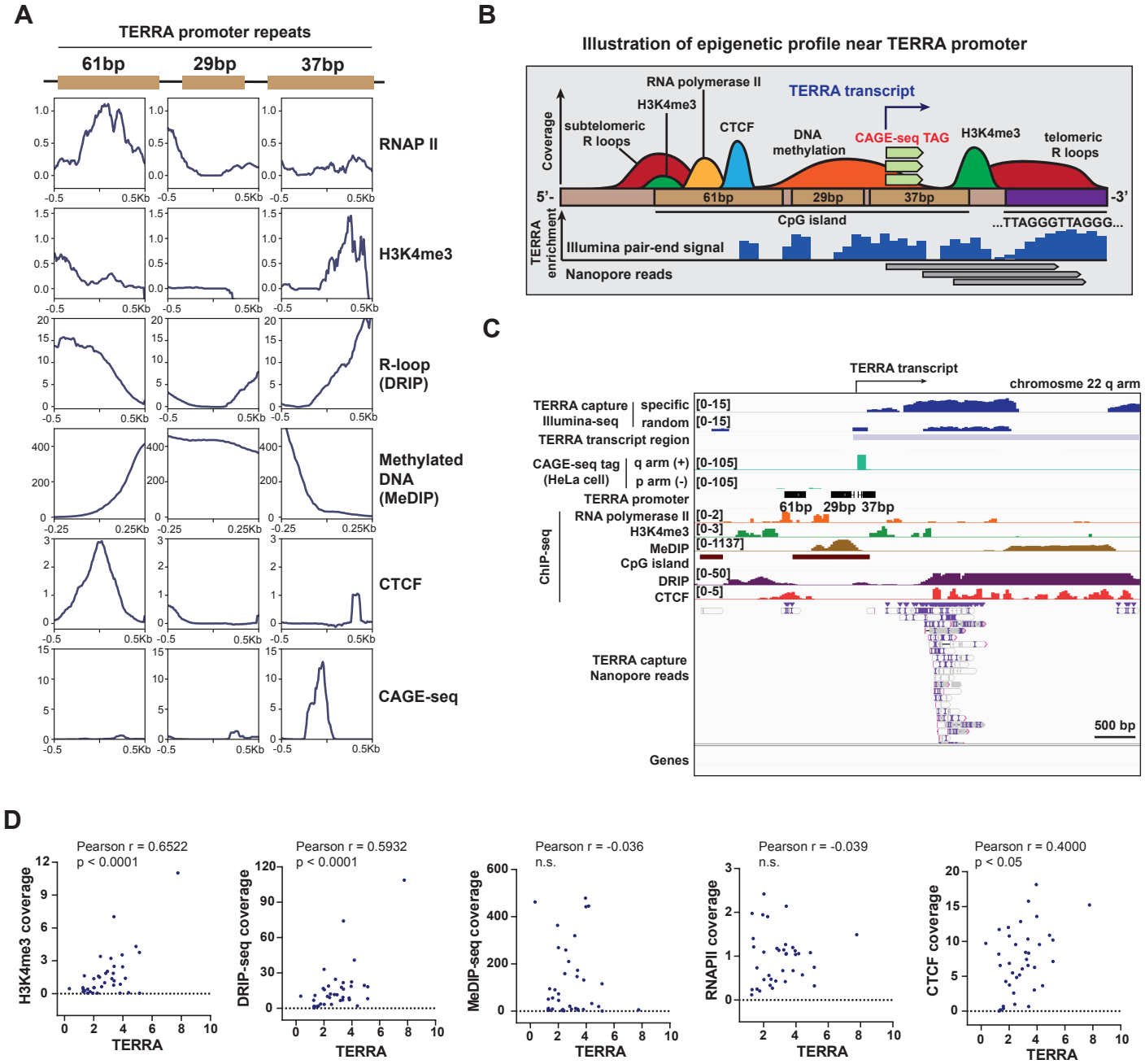


Figure 2. Enrichments of H3K4me3, Pol II, CpGs, DNA methylation and R-loops at TERRA promoter
(**A**) Meta-analysis showing the distribution of epigenetic marks on 61-29-37 bp repeats located at subtelomeric regions. Each plot represents the coverage of indicated marks. (**B**) A schematic model of the epigenetic profiles at TERRA promoter near telomeres. (**C**) Genome browser view showing the coverage of indicated epigenetic marks at TERRA promoter on chr22q arm. (**D**) Scatter plots of TERRA v.s. H3K4me3, R-loop (DRIP-seq), DNA methylation (MeDIP-seq), CTCF, or RNAPII near TERRA transcription start sites. Each dot indicates TERRA enrichment and epigenetic marks at individual chromosome ends in U2OS cells. TERRA enrichment (log2 ratio) was calculated by comparing TERRA capture and no capture. P values, by Pearson's correlation.

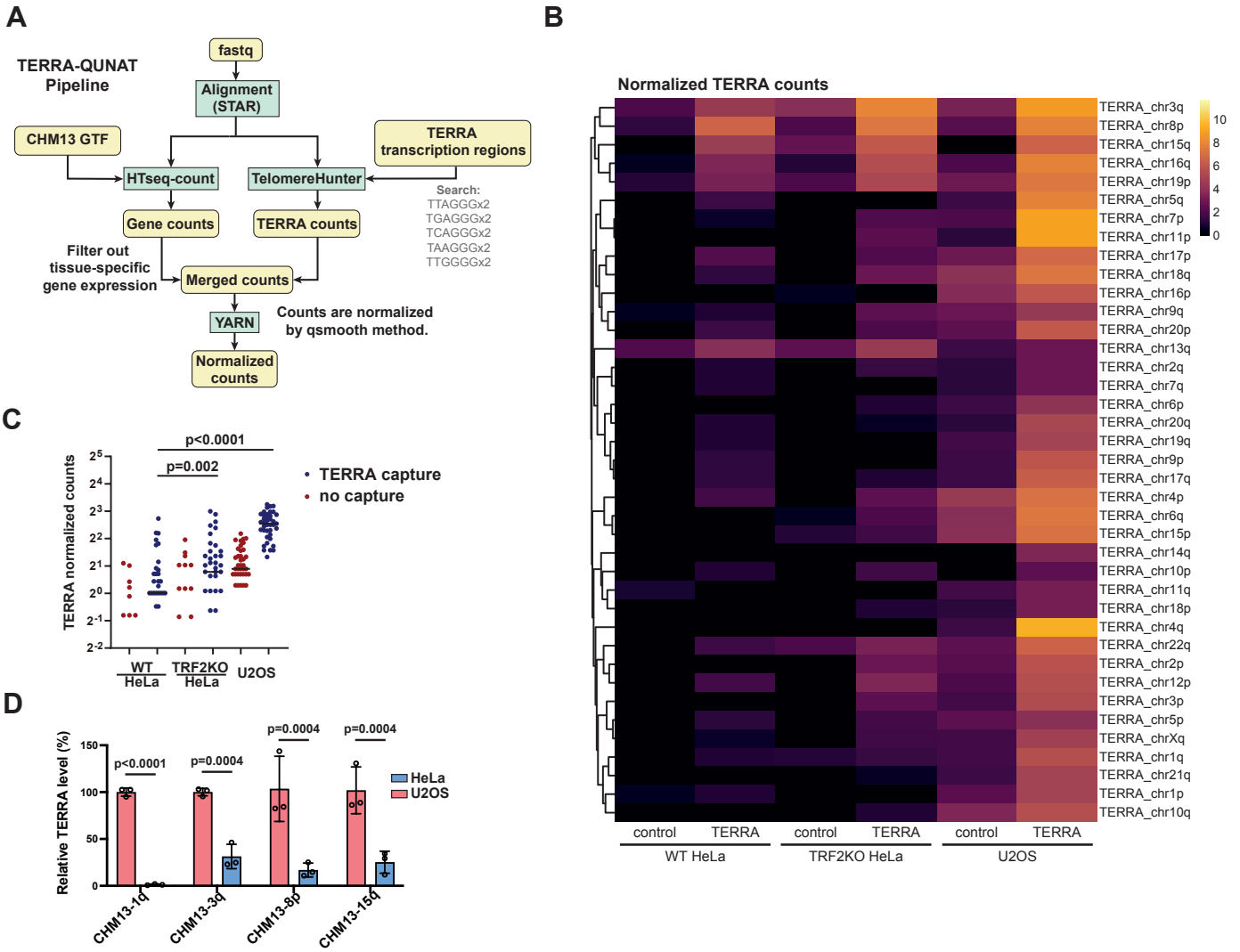


Figure 3. TERRA-QUANT for measuring TERRA expression

(**A**)The workflow of bioinformatics pipeline for quantification of TERRA expression at each chromosome end. Reads that mapped to TERRA transcriptions were counted. YARN was used for normalization between different tissues. (**B**) Heatmap showing TERRA expression from individual chromosome ends in HeLa and U2OS cells. TERRA capture, using antisense oligos to enrich TERRA. Control, no TERRA capture. (**C**) TERRA normalized counts were analyzed by TERRA-QUANT. Each dot indicates TERRA counts at individual chromosome ends. TRF2 deletion (TRF2KO) in HeLa cells increases TERRA expression. *P*-values by Wilcoxon matched-pairs signed rank test. Bars, median with interquartile range. (**D**) RT-qPCR to detect TERRA levels in HeLa and U2OS cells. Subtelomeric primers for TERRA were designed based on the T2T-CHM13 genome sequence. Data from three biological replicates. P-values by two-tailed Student's t test. Bars, +/-SD.

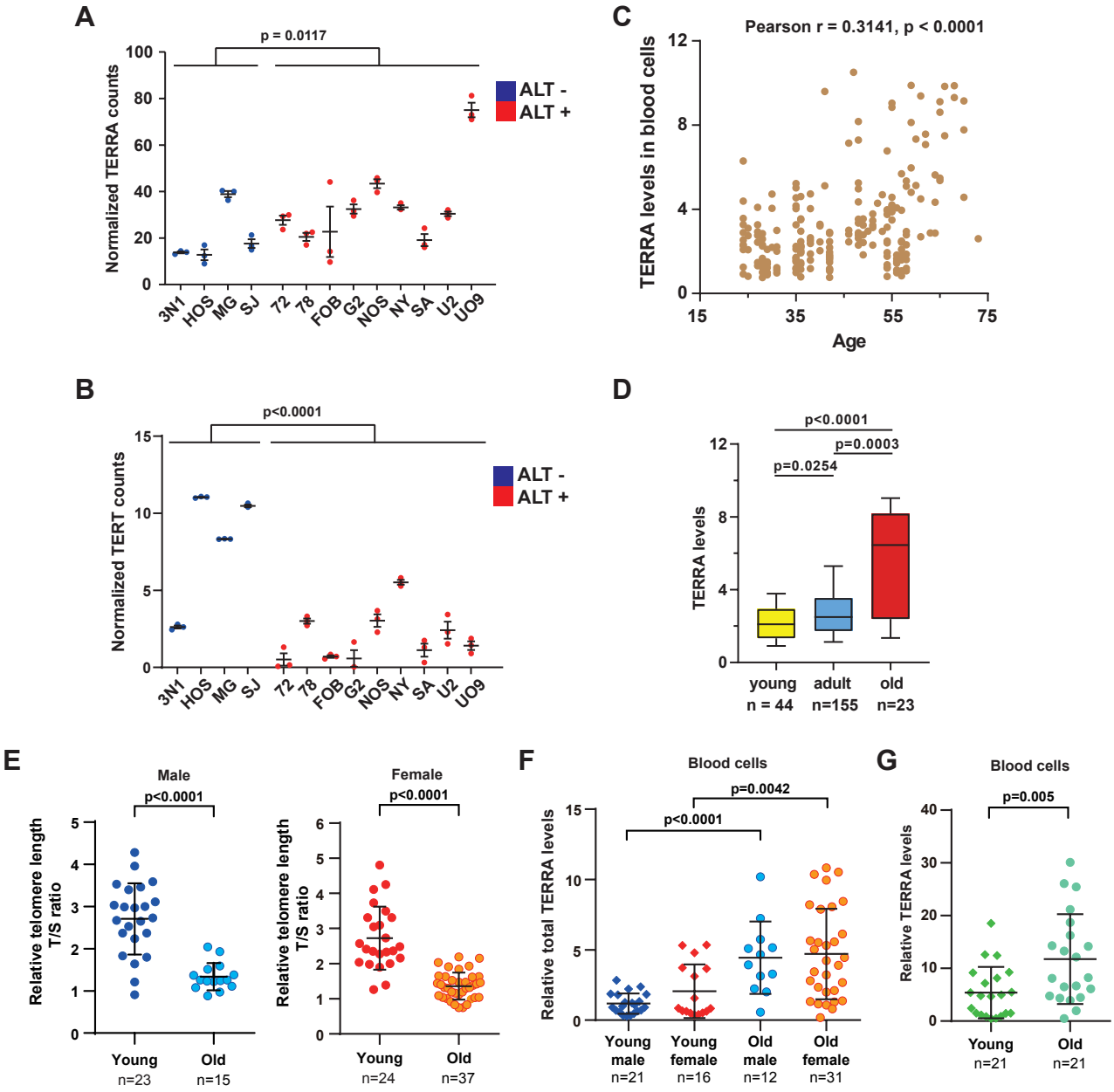


Figure 4. TERRA levels increase with age in blood cells.
(**A**) TERRA expression in ALT positive (+) or negative (-) cell lines was quantified by TERRA-QUANT using RNA-seq datasets. Total TERRA counts at all chromosome ends were accumulated. Each dot indicates the normalized TERRA read counts of each RNA-seq dataset. Bars, mean+/-SD. P values, by Mann-Whitney U test. (**B**) TERT expression of each RNA-seq data from ALT+ or ALT- cells. Bars, mean+/-SD. P values, by Mann-Whitney U test. (**C**) TERRA expression in human blood cells was analyzed by TERRA-QUANT using RNA-seq datasets. Scatter plots showing the correlation of TERRA v.s. age in blood cells. Each dot indicates total TERRA levels from all chromosome ends of each individual. P values, by Pearson's correlation. (**D**) Normalized TERRA counts in blood cells of different ages. Young (<30 years); adult (30~59 years); old (≥60 years). Bars, median with interquartile. P values, by Mann-Whitney U test. (**E**) The T/S ratio represents the relative telomere length (T) to the single-copy gene (S, 36B4 gene). Each dot represents an individual T/S ratio. Young (21~65 years); old (≥65 years). (**F**) RT-qPCR to detect total TERRA levels in blood cells using telomeric repeat primers. Each dot indicates each individual. (**G**)RT-qPCR to detect TERRA levels using subtelomeric primers (hg38-2q). Each dot indicates each individual. P vales, by two-tailed Student's t-test. n=sample size. Error bars are SDs.

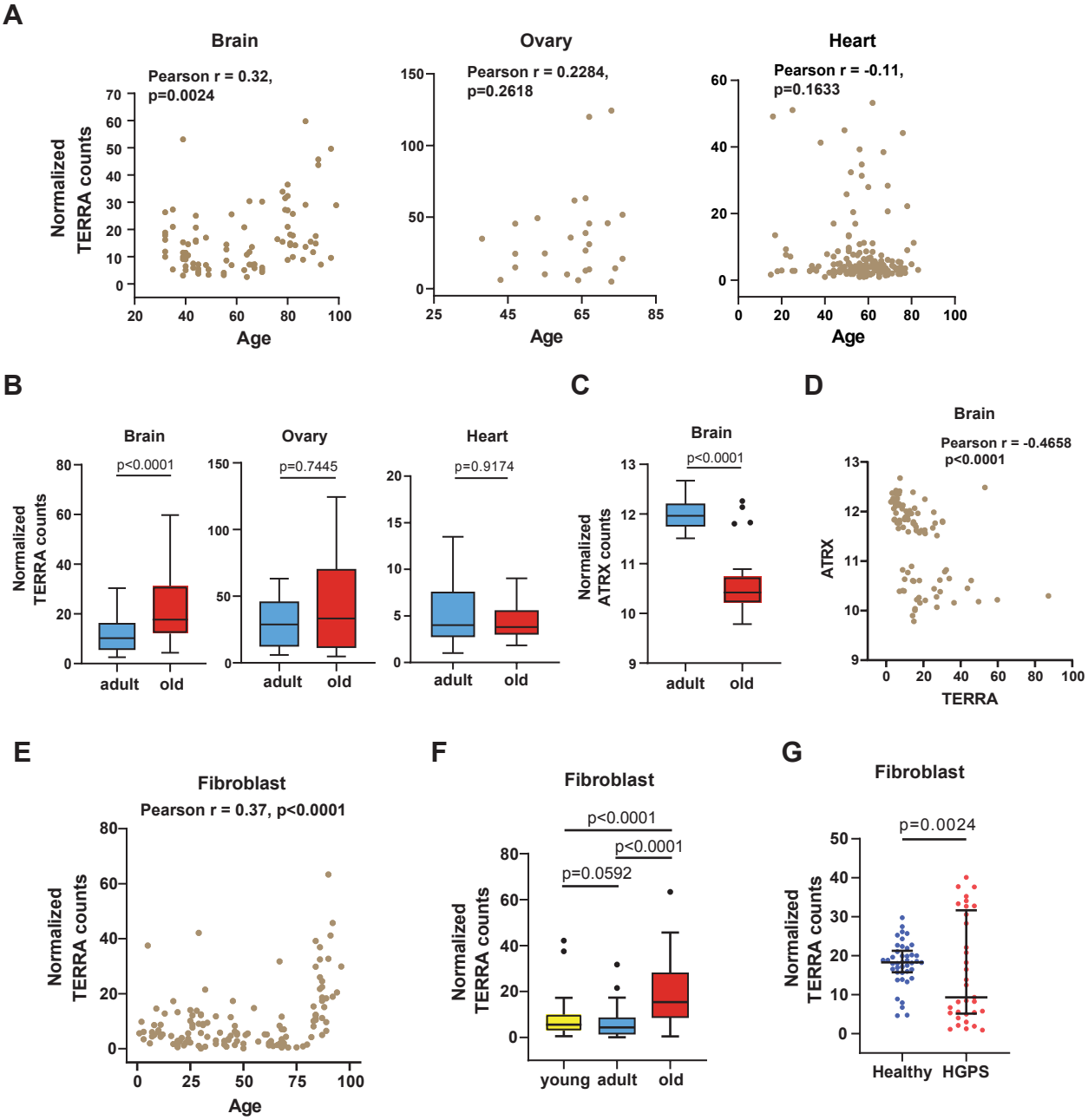


Figure 5. TERRA expression increases with age in brain tissues and fibroblasts.
(**A**) Scatter plots showing the correlation of TERRA v.s. age in brain, ovary and heart tissues. Each dot indicates total TERRA level from individuals. P values, by Pearson's correlation. (**B**)TERRA levels in various tissues with different ages. Adult (<70 years); old (≥70 years). Bars, median with interquartile. P values, by Mann-Whitney U test. (**C**) Boxplots of normalized ATRX counts in brain tissues. Bars, median with interquartile. P values, by Mann-Whitney U test. (**D**) Scatter plots of normalized TERRA and ATRX counts in brain tissues. P values, by Pearson's correlation. (**E**) TERRA levels in human fibroblasts derived from healthy people of various ages. P values, by Pearson's correlation. (**F**) Boxplots showing the upregulation of TERRA levels in old human fibroblasts. Bars, median with interquartile. P values, by Mann-Whitney U test. n.s. no significance. (**G**) Fibroblasts derived from HGPS patients display abnormal TERRA levels. Each dot represents TERRA level of each individual. Bars, median with interquartile. P values, by Kolmogorov-Smirnov test.

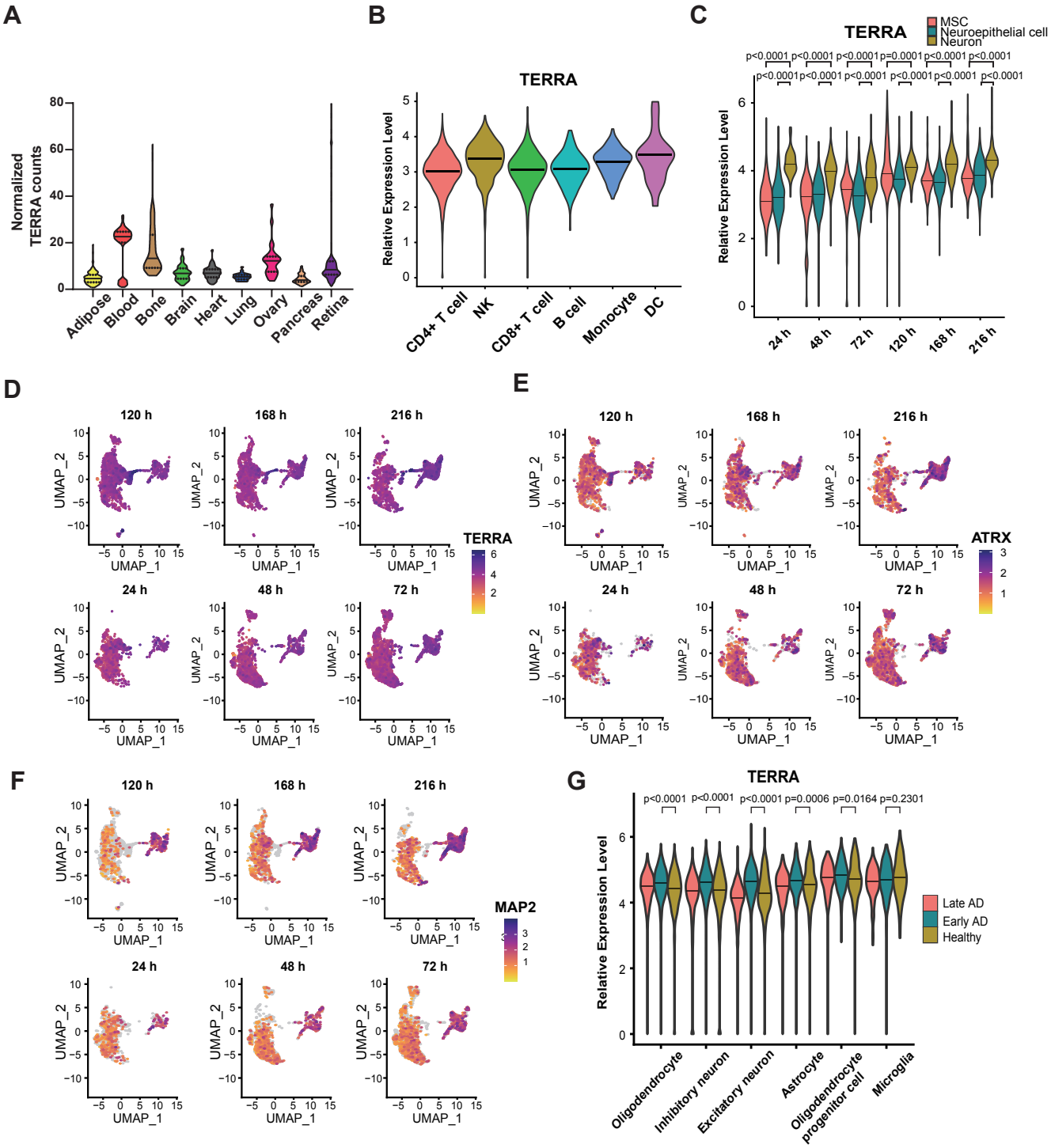


Figure 6. Single-cell analysis of TERRA levels under neuronal differentiation and in Alzheimer's disease.
(**A**) Violin plot showing normalized TERRA counts in various human tissues. The solid lines indicate the median, and the dash lines are interquartile range. (**B**) Single cell RNA-seq showing relative TERRA levels in various immune cells from PBMC. NK, nature killer cell; DC, dendritic cells. Solid lines, median. (**C**) Single cell RNA-seq analysis of relative TERRA levels in human embryonic stem cells (hESCs) undergoing neuronal differentiation with various HOX patterning periods (24h,72h,48h, 120h, 168h, 216h). MSC, mesenchymal stem cells.Solid lines, median. (**D-F**) Uniform manifold approximation and projection (UMAP) plots of single-cell RNA-seq datasets from hESCs undergoing neuronal differentiation using various HOX patterning periods. Colors according to TERRA, ATRX, or neuron marker MAP2 expression levels. (**G**) TERRA levels in cells isolated from prefrontal cortex of individuals with Alzheimer's disease (AD). Single-nucleus RNA-seq datasets were obtained from the ROSMAP project, and grouped into healthy (no AD pathology), early-AD pathology, and late-AD pathology. Solid lines, median. P values, by Mann-Whitney U test.