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], Hong-Jhih Shen[1], Chan-Hsien Yeh[2,3], Hung-Chih Kuo[2,3], Der-Sheng Han[4,5,*] 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]Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, 10051, Taiwan

[3]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan

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

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

[#]These authors contributed equally

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); Owen11533@gmail.com (Hong-Jhih Shen); chanhsien@gate.sinica.edu.tw; kuohuch@gate.sinica.edu.tw; 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)

This manuscript contains 7 main figures, 9 supplementary figures, and 9 supplementary tables.

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.

48 **Introduction**

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

65 TERRA exhibits binding capabilities with telomeric DNA binding proteins (22-24) and non-
66 telomeric DNA binding proteins such as chromatin modifiers (24). It binds to both telomeric and
67 non-telomeric chromatin, regulating telomere integrity and gene expression *in cis* and *in trans*
68 (23-25). For instance, TERRA interacts with ATRX, a protein involved in histone H3.3 recruitment
69 and H3K9me3 formation (26), and prevents ATRX from binding to chromatin (24,27). TERRA
70 depletion leads to increased ATRX recruitment to repetitive sequences such as rDNA,
71 retrotransposons, subtelomeric sequences, and telomeric repeats, as well as the regions close to
72 transcription start sites (24,27). Furthermore, TERRA plays a crucial role in modulating DNA G-
73 quadruplex structures around the transcription start sites, suggesting a function of TERRA beyond
74 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.

101 **Materials & Methods**

102 **Cell culture**

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

107

108 **TERRA capture for Illumina RNA-seq**

109 To enrich TERRA, total RNA was extracted from 1×10^7 cells using the TRIzol™ Reagent
110 (Invitrogen™, Cat:15596018) following the manufacturer's protocol. The total RNA was treated
111 with DNaseI (Invitrogen™, AM2238) and supplied with Ribonucleoside vanadyl complexes (VRC,
112 New England Biolabs, E7760S) under 37°C for 15 minutes to eliminate DNA (0.4 Unit/µl DNaseI
113 and 10 mM VRC). The enzyme was inactivated by 5 mM of EDTA. The DNaseI-treated RNA sample
114 was then subjected to a second round of purification using TRIzol™ reagent. For one TERRA-
115 capture experiment, a total of 20 µg of the DNaseI-treated RNA was used. The RNA sample was
116 dissolved in 90 µl of 6x SSC buffer (0.9M of NaCl, 0.9M of Sodium citrate). The RNA sample and
117 10 µl of 1 µM biotinylated TERRA antisense probe (5'-$(CCCATT)_5$-3'BioTEG) were individually
118 denatured at 70°C for 2 minutes, and then were mixed together (final: 20 µg RNA, 2x SSC buffer,
119 100 nM probe) for an additional 8 minutes at 70°C. The RNA-probe annealing was carried out by
120 gradually decreasing the temperature from 70°C to 44°C, followed by incubation at 44°C for an
121 additional 30 minutes. To capture TERRA RNA, 100 µl of streptavidin beads (MyOneC1,
122 Invitrogen™, Cat#65002) were prepared and mixed with the sample. The mixture was then
123 incubated on a rotating wheel at 37°C incubator for 15 minutes. The supernatant was separated
124 from beads by a magnet and carefully removed. Beads were washed 4 times with 2x SSC buffer
125 containing 0.1% NP40 (0.3M of NaCl, 0.3M of Sodium citrate, 0.1% NP-40) at 37°C, followed by
126 two washes with 1x SSC buffer containing 0.1% NP40 (0.15M of NaCl, 0.15M of Sodium citrate,
127 0.1% NP-40), once at 37°C and once at room temperature. Finally, beads were washed once with
128 1x SSC buffer without NP40 (0.15M of NaCl, 0.15M of Sodium citrate) at 37°C. Washed beads
129 were resuspended in 100 µl of nuclease-free water and incubated at 70°C to elute RNA. The

130 elution step was repeated twice. The eluted RNA sample was concentrated by precipitating in 3×
131 volume of 100% ethanol (300 µl), containing 0.1× volume of 3M NaOAc (40 µl, Thermo Fisher
132 Scientific, AM9740) and 2 µl of GlycoBlue™ (Invitrogen™, AM9515) overnight at -30℃. The RNA
133 was precipitated by centrifuge at 4℃ and 21,100 x g . The RNA pellet was washed twice with 70%
134 EtOH and eluted in 30 µl of nuclease-free water. RNA concentration was measured using Qubit
135 RNA HS kit (Invitrogen™, Q32852). The captured RNA was converted into cDNA, followed by
136 paired-end Illumina sequencing and RT-qPCR.
137
138 **Illumina RNA seq library preparation**
139 The cDNA library was constructed using the NEBNext® Ultra II Directional RNA Library Prep Kit
140 (New England Biolabs, E7760S) according to the manufactory instruction with some
141 modifications. The first-strand synthesis step was carried out by SuperScript IV (Thermo Fisher
142 Scientific, Cat: 18090010). The TERRA-captured RNA sample (30 µl) was divided into two equal
143 halves for the reverse transcription reaction. Half of the TERRA-captured sample was reverse
144 transcribed using 2 µM random hexamers (Thermo Fisher Scientific, 48190011) and the other
145 half using 1 µM of telomeric C-rich primer 5'-$(CCCTAA)_5$-3'. These two groups are referred to as
146 the random and specific groups, respectively. Followed by reverse transcription, second strand
147 synthesis, end repair, adaptor ligation, and size selection. Libraries were amplified using NEBNext
148 Ultra II Q5 Master Mix (New England Biolabs, #M0544L) with optimal cycles determined by qPCR,
149 and size-selected twice using AMPure XP beads. Finally, the DNA was eluted in 15 µL of 0.1× TE
150 buffer and stored at -20℃. The library size distribution was assessed using the Agilent 2100
151 Bioanalyzer. Each library was quantified using the NEBNext Library Quant Kit (New England
152 Biolabs, #E7630S). The pooled libraries (fragment size 300~500 bp) were subjected to paired-end
153 150 bp sequencing on Illumina MiSeq or NovaSeq platforms.
154
155 **Reverse transcription and Quantitative PCR**
156 TERRA enrichment was assayed by RT-qPCR using cDNA. The reverse transcription was performed
157 using random hexamers (Thermo Fisher Scientific, #48190011) and SuperScript™ IV Reverse
158 Transcriptase (Thermo Fisher Scientific, Cat#18090200)(final: 1µM random hexamer, 400 µM

159 dNTP, 1×SuperScript IV buffer, 8 Unit/µl SuperScript IV reverse transcriptase, 4 mM DTT, 1.6
160 Unit/µl RNaseOUT). The qPCR analysis was conducted using a 1:5 dilution of TERRA-enriched
161 cDNA and a serial dilution of control cDNA, ranging from 1× to 1000× dilution (1% to 0.001% of
162 input), to generate a standard curve. Subtelomeric primer sequences (hg38-2q, CHM13-1q,
163 CHM13-3q, CHM13-8p, and CHM13-15q) were listed in **Supplementary Table 6**. The qPCR
164 reaction was performed with IQ SYBR Green Supermix (Bio-Rad, Cat: 170-8882) using the CFX
165 Real-Time PCR Detection System. The thermal cycling conditions were as follows: Initial heat
166 activation of polymerase at 95℃ for 3 minutes, followed by 35 cycles of denaturation at 95℃ for
167 15 seconds, annealing and extension at 58℃ for 1 minute. Data analysis was carried out using
168 CFX Maestro software v1.1 (Bio-Rad). For Illumina samples, the standard curve was generated
169 from a series of dilutions from the input. TERRA RNA abundance was normalized to GAPDH.
170 TERRA-captured samples were compared to the control sample (no capture) to calculate TERRA
171 enrichment. For Illumina samples, TERRA enrichment ratio was calculated using the ddCt method,
172 where the Ct value of TERRA was normalized to that of GAPDH.
173
174 **TERRA capture for Nanopore Direct RNA-seq**
175 We harvested cells from 200 of 150 mm culture dishes for RNA extraction. The TERRA-capture
176 procedure followed the method described previously with a few modifications. First, the RNA
177 samples were not treated with DNase I. Second, the input RNA amount for each capture was
178 increased to 500 µg. Third, to achieve the required RNA quantity for Nanopore direct RNA-seq,
179 10 captures were pooled. RNA was concentrated by precipitation with 3× volume of 100%
180 ethanol, 1/10 × volume of 3M NaOAc (Thermo Fisher Scientific, AM9740), and 2 µl of GlycoBlue™
181 (Invitrogen™, AM9515) for 3 days at -20℃. 100 captures (a total of 50 mg input RNA) were
182 required to acquire a sufficient amount of TERRA-enriched RNA for sequencing. Subsequently,
183 RNA was precipitated by centrifugation at 4°C at 21,100 x g for 20 minutes. The pellet was washed
184 with 70% ethanol and eluted in 20 µl of nuclease-free water. Finally, a total of 3.4 µg RNA was
185 obtained after capture. Next, 1.5 µg of TERRA-enriched RNA was polyadenylated using E. coli Poly
186 (A) polymerase (New England Biolabs, #M0276S) at a final concentration of 0.25 U/µl in 1X Poly(A)
187 Polymerase reaction buffer with 1 mM dATP at 37°C for 30 minutes. To prevent RNA degradation,

188 the reaction included 2 Unit/µl RNaseOUT, (Thermo Fisher Scientific, Cat: 10777019). The
189 polyadenylation reaction was terminated by adding 10 mM EDTA. Polyadenylated RNA was
190 purified using 2× sample volume of AMPureXP RNA beads (Beckman Coulter, A66514), following
191 the manufacturer's protocol. RNA was eluted in 20 µl of nuclease-free water. RNA concentration
192 was measured using Qubit RNA High-Sensitivity assay (Thermo Fisher Scientific, Q32852). The
193 fragment size was assessed using Bioanalyzer (Agilent RNA 6000 Pico Kit). TERRA-enriched RNA
194 was subjected to Nanopore direct RNA sequencing (Cat#SQK-RNA002) with R9.4.1 flow cell (FLO-
195 MIN106).
196
197 **Analysis of TERRA-capture Illumina RNA-seq**
198 The fastq files were subjected to a quality filter (Q > 20), and Illumina adaptor sequences were
199 trimmed using Trimgalore v0.6.7 (trimgalore --illumina --pair Read1.fastq Read2.fastq -o
200 /path/to/output). The reads were aligned to the T2T-CHM13 genome using the STAR aligner
201 v2.7.9a (star --runMode alignReads --genomeDir /path/to/T2T-CHM13 --outSAMtype BAM
202 SortedByCoordinate --outFileNamePrefix output_path/id --readFilesIn trimmed_Read1.fastq
203 trimmed_Read2.fastq). The aligned BAM files were indexed by samtools v1.17 (samtools index -
204 b exp.bam -o exp.bam.bai) and converted to coverage files in bigwig format by deepTools v3.3.1
205 bamCoverage, with RPKM as the normalization method (bamCoverage -b BAM_files --
206 normalizeUsing RPKM --binSize 30 -of bigwig -o output.bw). Furthermore, a comparison was
207 performed between TERRA-captured groups and no-captured RNA-seq data (control) using log2
208 ratio analysis by bigwigCompare (deepTools). The coverage files (bigwig) were visualized on the
209 Integrative Genomics Viewer (IGV) v2.16.0. TERRA transcription regions were manually assigned
210 according to TERRA enrichment, nanopore reads, and CAGE tags at chromosome ends of the T2T-
211 CHM13 genome. The BED and GTF files of TERRA transcription regions were used for TERRA
212 quantification. To quantify TERRA read counts, the alignments were extracted from the BAM files
213 using samtools v1.17 (samtools view -b -h -M -L TERRA_region.bed --fetch-pairs input.bam -o
214 TERRA_region_alignment.bam), and filtered with MAPQ ≥ 30.
215
216 **Analysis of TERRA Nanopore direct RNA-seq**

217 Fast5 files were processed using Guppy base caller (guppy_GPUv5.0.7) with a default quality filter
218 of Q > 7 (guppy -i /path/to/fast5 -c rna_r9.4.1_70bps_hac.cfg -s /path/to/output_fastq). After
219 base-calling, reads were aligned to CHM13 reference genome using the minimap2 aligner v2.22
220 (minimap2 -ax splice -uf -k14 /path/to/CHM13.mmi TERRA_Nanopore.fastq >
221 TERRA_Nanopore_aligned.sam). The alignment was converted to BAM format and sorted using
222 samtools v1.17 (samtools view -b TERRA_Nanopore_aligned.sam -o
223 TERRA_Nanopore_aligned.bam; samtools sort TERRA_Nanopore_aligned.bam -o
224 TERRA_Nanopore_aligned.sorted.bam). A bam index was created by samtools v1.17 (samtools
225 index TERRA_Nanopore_aligned.sorted.bam -o TERRA_Nanopore_aligned.sorted.bam.bai). The
226 aligned reads displayed on the genome browser were manually examined to assist in defining
227 TERRA transcription regions. Reads aligned to TERRA transcription regions were selected using
228 samtools v1.17 with the TERRA transcription region bed file (samtools view -b -h -M -L
229 TERRA_region.bed TERRA_Nanopore_aligned.sorted.bam -o
230 Nanopore_TERRA_regions_aligned.bam), and the counts of each TERRA region were quantified.
231 For the analysis of TERRA from individual chromosome ends (Figure 1D), the unique reads were
232 filtered based on MAPQ (≥ 1), FLAG values (FLAG = 0, or 16 for primary alignments, FLAG = 2048,
233 or 2064 for supplementary alignments) provided by minimap2 (**Supplementary Table 5.1**). The
234 supplementary alignments with the same read IDs were removed (**Supplementary Table 5.2**).
235

236 **Analysis of ChIP-seq datasets**

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

246 format, sorted, duplicate reads were removed, and the files were indexed using samtools v1.17
247 (samtools view -b exp.sam -o exp.bam, samtools sort exp.bam -o exp.sorted.bam, samtools
248 rmdup exp.sorted.bam -o exp.sorted.rmdup.bam, samtools index exp.sorted.rmdup.bam -o
249 exp.sorted.rmdup.bam.bai). ChIP-seq coverage files (bigwig) were generated using deepTools
250 v3.3.1 bamCompare, normalizing using CPM method and subtracting to input signal.
251 (bamCompare -b1 ChIP_BAM_file -b2 Input_BAM_file --normalizeUsing CPM --binSize 30 --
252 operation subtract -of bigwig -o output.bw). Meta-analysis surrounding the 61-29-37 nt repeat
253 regions was performed using deepTools v3.3.1 computeMatrix and plotHeatmap
254 (computeMatrix reference-point --referencePoint center -a 500 -b 500 -S ChIP-seq_coverage.bw
255 -R 61_29_37_region.bed -o Matrix.gz, plotHeatmap -m Matrix.gz --colorMap bwr -o
256 output_meta-analysis.pdf). For visualization on CHM13 genome, the bigwig files were examined
257 in Integrative Genomics Viewer (IGV).
258
259 **Analysis of TERRA expression by TERRA-QUANT**

260 RNA-seq datasets were downloaded and pre-processed using the previously described methods
261 (SRAToolkit and TrimGalore commands). The processed reads were aligned to the T2T-
262 CHM13v1.1 genome using STAR aligner with default settings. Annotated TERRA transcription
263 regions were divided into subtelomeric and telomeric repeat portions and incorporated into the
264 CHM13 GTF file. Read counts for both human genes and TERRA were calculated using HTseq-
265 count (htseq -f bam -s no/yes/reverse -t transcript --idattr gene_name -m intersection-nonempty
266 --nonunique all BAM_file CHM13_GTF_with_TERRA_regions > output_count.txt). For quantifying
267 TERRA from individual chromosome ends, reads were extracted with MAPQ ≥ 30 for paired-end
268 reads, or with MAPQ = 255 for single-end reads by samtools. Only reads mapped to the
269 subtelomeres within defined TERRA transcription regions were calculated for chromosome-arm-
270 specific TERRA. For total TERRA quantification, reads were aligned by STAR with default settings,
271 and duplicates were removed and filtered with primary alignments (MAPQ ≥ 1) by samtools. The
272 reads mapped to TERRA regions and all human genes were then analyzed together and
273 normalized using the YARN package (44) in the R language. The total TERRA levels in human
274 tissues were calculated based on reads mapped to all TERRA transcription regions with proximal

275 TSSs located at chromosome ends (Type I and Type II, but not Type III), including both
276 subtelomeric regions and pure telomeric repeats. "TERRA levels" refer to normalized TERRA read
277 counts, representing the relative expression of TERRA transcripts compared to a control gene set
278 across all samples. Counts for other genes were obtained using HTseq-count, and all counts were
279 merged for downstream normalization. For normalization, gene counts—including TERRA and
280 other genes quantified using HTseq-count—were processed using the YARN R package, which
281 supports multi-sample normalization across diverse tissues and cell types.

282

283 **NGS sequencing for TERRA RT-qPCR products**

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

301

302 **Poly(A)+ and poly(A)- TERRA capture sequencing and analysis**

303 TERRA RNA was first enriched by hybridization with a biotin-labeled C-rich telomeric probe.
304 Following this, polyadenylated TERRA transcripts were selectively isolated using the NEBNext®
305 Poly(A) mRNA Magnetic Isolation Module (E7490S, New England Biolabs), allowing for the
306 separation of poly(A)+ and poly(A)− fractions. Both RNA fractions were then subjected to strand-
307 specific library preparation (NEBNext® Ultra II Directional RNA Library Prep Kit, E7760S) for
308 Illumina sequencing, using 150 bp paired-end reads. To further enhance the representation of
309 TERRA transcripts during cDNA synthesis, C-rich telomeric primers were used, ensuring
310 preferential reverse transcription of TERRA RNA. The fastq files obtained from TERRA-capture
311 Illumina RNA-seq were subjected to a quality filter (Q > 30), and Illumina adaptor sequences were
312 trimmed using Trimgalore v0.6.7. The reads that passed the quality filter were aligned to the T2T-
313 CHM13 genome using the STAR aligner v2.7.9a. (star --runMode alignReads --genomeDir
314 /path/to/T2T-CHM13 --outSAMtype BAM SortedByCoordinate --outFileNamePrefix
315 output_path/id --readFilesIn trimmed_Read1.fastq trimmed_Read2.fastq). To quantify TERRA
316 read counts, the alignments were extracted from the BAM files using samtools v1.17 (samtools
317 view -b -h -q 30). Reads mapped to TERRA transcription regions including Type I, II, and III, were
318 gathered as total TERRA reads. The proportion of TERRA reads in poly(A)+ and poly(A)- samples
319 was calculated as the number of TERRA reads in each region divided by the total TERRA reads.
320 The ratio of poly(A)+ to poly(A)- reads was determined by % of total TERRA reads in poly(A)+ over
321 % of total TERRA reads in poly(A)-. Enrichment of poly(A)+ relative to poly(A)- was represented
322 as the $\log_2$ values of this ratio.
323

324 **Research participants for collecting blood cells**

325 Blood samples were collected as previously described (38). The study was approved by the
326 institutional review board (IRB No. 201601091RIND) of the National Taiwan University Hospital.
327 All participants were provided with written informed consent before participation in the study.
328 Trial registration: ClinicalTrials.gov: NCT02779088. Registered 20 May 2016,
329 https://clinicaltrials.gov/ct2/keydates/NCT02779088.

330

331 **Telomere length measurement in blood cells**

332 Blood samples were collected from the participants as previously described (38). Briefly, genomic
333 DNA was extracted from frozen human buffy coats using the QIAamp® DNA Mini Kit (QIAGEN,
334 Hilden, Germany, #Cat 51306). Quantitative analysis was performed by quantitative polymerase
335 chain reaction (qPCR). The relative telomere length for each participant was calculated using the
336 telomere-to-single copy gene (T/S) ratio, defined as the number of telomeric repeats (T) divided
337 by a standard reference DNA (S). The *36B4* gene was used as the single-copy DNA. A standard
338 curve was generated from a series of dilutions of genomic DNA from HeLa cells, and was used for
339 calibration for each qPCR plate. Primers for telomeric DNA and *36B4*: *36B4* forward primer
340 sequence: 5'-CAGCAAGTGGGAAGGTGTAATCC-3', and *36B4* reverse primer sequence: 5'-
341 CCCATTCTATCATCAACGGGTACAA-3'; telomeric DNA forward primer sequence: 5'-
342 GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT-3' and telomeric DNA reverse primer
343 sequence: 5'-TCCCGACTATCCCTATCCCTATCCCTATCCCTATCCCTA-3'.
344

345 **Quantification of TERRA by RT-qPCR in blood samples**

346 RNA was extracted from buffy coat samples using TRIzol (Thermo Fisher Scientific, Cat# 15596018)
347 and purified by acid-phenol:chloroform (Thermo Fisher Scientific, Cat# AM9722). RNA was
348 treated with ezDNase™ Enzyme (Thermo Fisher Scientific) to eliminate contaminating genomic
349 DNA. Reverse transcription was carried out using the SuperScript™ IV system (Thermo Fisher
350 Scientific, Cat# 18090200). Random (2.5 nM) hexamers combined with telomeric repeat primers
351 (0.25 nM) to enrich TERRA complementary DNA (cDNA). The protocol for total TERRA qPCR was
352 similar to that for the telomere qPCR using telomeric repeat primers. Beta-2 microglobulin (*B2M*)
353 was used as the reference gene. B2M primers: forward primer sequence, 5'-
354 CTATCCAGCGTACTCCAAAG-3'; reverse primer sequence, 5'-GAAAGACCAGTCCTTGCTGA-3'.
355 TERRA RNA extracted from U2OS cells was used to obtain a standard curve for calibration
356 between samples from different qPCR plates. The relative TERRA expression level was analyzed
357 using the TERRA RNA/B2M RNA ratio. Subtelomeric primers (hg38-2q) for TERRA RT-qPCR in
358 blood cells were listed in **Supplementary Table 6**. Data points from samples with poor RNA
359 quality—assessed by RT-qPCR analysis of the B2M control were excluded from the analysis.
360

361

362 **Analysis of TERRA expression across different tissues**

363 Publicly available RNA-seq datasets were gathered from NCBI GEO datasets (32,39,40). Sample
364 sources were documented in **Supplementary Table 7**. To compare TERRA levels across different
365 tissues or cell lines, we employed the YARN R package (41). The whole gene counts derived from
366 all RNA-seq datasets were merged in the same expression table for the following process. The
367 filtration of low-expressed genes was performed based on an expression threshold (CPM < 1).
368 TERRA counts for each sample were determined using TERRA-QUANT as described previously.
369 Normalization of all data was performed using the qsmooth method (41).

370

371 **Single-cell and Single-nucleus RNA-seq analyses**

372 Neuronal single-cell RNA-seq datasets (GSE186698) and PBMC single-cell RNA-seq datasets
373 (GSE157007) were downloaded as previously described (**Supplementary Table 8**). Alignment,
374 barcode assignment and Unique Molecular Identifier (UMI) counting were performed using
375 CellRanger v7.2.0 (42) with a default setting (MAPQ = 255). Downstream analysis of CellRanger
376 outputs was performed using Seurat v4.4.0 (43). The low-quality cells were filtered out if the
377 number of genes detected or mitochondria gene reads were three Mean Absolute Deviations
378 (MAD) away from the average. Genes expressed in fewer than 3 cells are excluded. 2,000 highly
379 variable genes were identified using the variance stabilizing transformation (VST) method and
380 were used for Principle Component Analysis (PCA). The top 20 principal components were
381 selected for cell clustering. The result was visualized with Uniform Manifold Approximation and
382 Projection (UMAP). Data integration was performed by using FindIntegrationAnchors() and
383 IntegrateData() functions according to the union of the top 2,000 highly variable genes of each
384 data. Plots of TERRA, ATRX, and MAP2 were generated by scCustomize v2.0.1 (Samuel Marsh,
385 Maëlle Salmon, & Paul Hoffman. (2023). https://doi.org/10.5281/zenodo.10161832). In
386 neuronal differentiation samples, cells positive for CD29, CD44, and CD105 but negative for CD14,
387 CD34, and CD45 were classified as mesenchymal stem cells; cells expressing SOX2 and Nestin
388 were identified as neuroepithelial cells, while those positive for MAP2 were categorized as
389 mature neurons. Cells in PBMC were annotated by the expression of representative cell markers:

390 CD4+ T cells (CD4, IL7R and CD3G), NK cells (NCAM1 and NKG7), CD8+ T cells (CD8A and CD8B),
391 B cells (CD19 and MS4A1), Monocytes (CD14 and CD16), Dendritic cells(IL3RA). Single-nucleus
392 RNA-seq datasets were obtained from The Rush Alzheimer's Disease Center (RADC) Research
393 Resource Sharing Hub at Synapse (https://www.synapse.org/#!Synapse:syn18485175) with the
394 DOI 10.7303/syn18485175. The ROSMAP metadata can be found at
395 https://www.synapse.org/#!Synapse:syn3157322. Details of individual samples, including those
396 with no pathology (13 individuals), AD-early pathology (8 individuals), and AD-late pathology (5
397 individuals), utilized in the analysis of this study are outlined in **Supplementary Table 9**. In
398 Alzheimer's disease samples, we assigned cell types by marker genes: excitatory neurons (NRGN),
399 inhibitory neurons (GAD1), astrocytes (AQP4), oligodendrocytes (MBP), microglia (CSF1R and
400 CD14) and oligodendrocyte progenitor cells (VCAN).
401
402 **Alzheimer's disease patient-derived iPSC differentiation to neurons**
403 To generate cortical neurons, induced pluripotent stem cells (iPSCs) derived from an Alzheimer's
404 disease (AD) patient carrying the P117L mutation in presenilin-1 (*PS1*)—the catalytic subunit of
405 γ-secretase—and mutation-corrected isogenic controls were used. iPSCs were dissociated with
406 dispase and cultured in suspension using human embryonic stem cell (hESC) medium lacking
407 bFGF, but supplemented with 100 nM LDN193189 and 10 µM SB431542 to initiate neural
408 induction. Cells were then transferred to N2 medium and cultured until neural rosette structures
409 formed. These neural rosettes were manually isolated and further cultured in NI medium to
410 promote neural sphere formation. The resulting neural spheres were plated onto Matrigel-
411 coated 6-well plates and cultured in ND medium to induce cortical neuron differentiation. The
412 differentiated cortical neurons were subsequently collected for RT-qPCR analysis.
413
414
415
416

417 ## Results

418 **Nanopore Direct RNA sequencing for TERRA transcripts**

419 To investigate TERRA transcripts in human cells, we captured RNA molecules containing UUAGGG
420 repeats by using biotinylated antisense oligos in U2OS cells (**Figure 1A**). The TERRA-enriched RNA
421 samples were subjected to two different sequencing methods: Illumina paired-end sequencing
422 and Nanopore direct RNA sequencing. For Illumina sequencing, TERRA-enriched RNA was reverse
423 transcribed using either random hexamers (random group) or telomere-specific primers (specific
424 group) for cDNA library construction. For Nanopore Direct RNA sequencing, enriched TERRA
425 molecules were polyadenylated by poly(A) polymerase before library preparation (**Figure 1A**).
426 Approximately, a few hundred to over two-thousand-fold enrichment was observed after TERRA
427 capture (**Supplementary Figure S1A**). The reads from both methods were aligned to the T2T-
428 CHM13 human genome reference, which includes nearly complete coverage of the entire
429 genome including subtelomeric and telomeric regions (44).

430

431 **TERRA transcription regions at chromosome ends**

432 The criteria for identification of TERRA transcription regions included the observations of
433 Nanopore reads, CAGE (Cap Analysis of Gene Expression) tags, and TERRA enrichment within 100
434 kb of chromosome ends (**Figure 1B, Supplementary Figure S1B**). Compared to ribosomal
435 depleted RNA-seq (control), TERRA-capture RNA-seq (enriched by antisense oligos) exhibited an
436 increase in read counts derived from chromosome ends in U2OS cells, shown by TERRA
437 enrichment in log2 values from Illumina sequencing data (**Figure 1C**). In cases where no CAGE-
438 seq tags mapped to the subtelomeric regions, we estimated transcription start sites (TSSs) based
439 on subtelomeric regions with Nanopore long reads and TERRA enrichment. We identified 39
440 chromosome ends that contain TERRA transcription (**Supplementary Tables 1, 2**). Notably, chr5p,
441 chr18p, and chr20p each potentially contain two TERRA transcription start sites (one proximal
442 and one distal to telomere), while other chromosome ends have a single TERRA transcription
443 start site located near the telomeric repeat tracts (**Supplementary Figure S1C**). The distance of
444 the identified TERRA TSS from the telomeric repeat tract per chromosome end is shown in

445 **Supplementary Figure S2**. These results demonstrate the presence of TERRA transcription across
446 a majority of chromosome termini in U2OS cells (**Figure 1D**), aligning with a recent study showing
447 that TERRA is detectable at most chromosome ends in human cell lines (45).

448 To further verify the TERRA transcription regions, we performed TERRA-capture RNA-seq
449 in another cell line (HeLa). The TERRA enrichment was also observed in defined TERRA
450 transcription regions at chromosome ends in HeLa cells (**Supplementary Figure S3**), showing the
451 consistency of TERRA transcription regions between two different cell lines.

452 Notably, most CAGE tags were mapped to the 37 bp repeat regions (**Supplementary**
453 **Figure S4**). In agreement with this, TERRA nanopore reads were aligned with the 37 bp repeats
454 at their 5' termini (**Supplementary Figure S4**). These results suggest that the regions preceding
455 the 37 bp repeats may function as potential TERRA promoters, supporting previous studies that
456 showed the importance of the 29 bp repeat regions in regulating transcription activity (4,21).

457 We classified three distinct types of TERRA transcription regions according to their
458 locations and promoters (**Figure 1B**). The majority of telomeres (37 out of 46 chromosome ends)
459 were characterized as Type I TERRA transcription regions by the presence of 61-29-37 bp repeats
460 at the promoter (**Supplementary Figure S4**). Only chr10p and chr17p termini were characterized
461 as Type II TERRA transcription region, which lacks 61-29-37 bp repeats (**Figure 1B, Supplementary**
462 **Figure S4**). TERRA transcription regions with 61-29-37 bp repeats or CAGE tags were summarized
463 in **Supplementary Tables 1 and 2**. Not all Type I TERRA promoters encompass all three repetitive
464 motifs. 16 chromosome ends are composed of all three repetitive motifs, while 21 chromosome
465 ends lack either one or two repetitive motifs (**Supplementary Figure S5A**, **Supplementary Table**
466 **3**). To investigate the impact of the 29 bp repeat element on TERRA expression, we compared
467 TERRA levels at chromosome ends with or without the 29 bp repeats. Our analysis revealed no
468 significant difference in TERRA expression between these regions (**Supplementary Figure S5B**),
469 implying the presence of an additional promoter that lacks the 29 bp repeats for TERRA
470 transcription.

471

472 **TERRA transcription from interstitial telomeric repeats**

473 In addition to TERRA transcription at chromosome ends, TERRA can be transcribed from
474 interstitial telomeric sequences (ITS), classified as Type III TERRA transcription regions (**Figures**
475 **1B-1D**, **Supplementary Figure S4, Supplementary Table 4**). We quantified Type I, II, and III TERRA
476 transcription and found that most TERRA transcripts in U2OS cells originate from chromosome
477 ends (Type I + II) (**Figure 1E**, **Supplementary Figure S3B**). Type III accounts for only ~4% of
478 Nanopore and 9.1% of Illumina TERRA reads in U2OS but is more prominent in HeLa cells,
479 comprising 46% of Nanopore and 17.7% of Illumina reads. In the T2T-CHM13 genome, we
480 identified 79 ITS sites containing telomeric repeats longer than 200 base pairs (**Supplementary**
481 **Figure S5C**). Among these, 19 sites showed TERRA enrichment and Nanopore reads containing
482 long telomeric repeats at the 3' ends. Notably, Type III TERRA transcription regions exhibit CAGE
483 tags, and the promoters usually comprise the 37 bp repeats without 29 bp repeats
484 (**Supplementary Figure S4)**. These results indicate that most TERRA transcripts originated from
485 the 37 bp repeat regions, regardless of whether TERRA derived from chromosome ends or ITS.
486 Genome browser views demonstrated some examples of TERRA transcription regions carrying
487 these features (**Supplementary Figure S4**).
488 Interestingly, we observed antisense TERRA transcripts, named ARIA (46) in Nanopore
489 direct RNA-seq data. ARIA reads were mapped to chromosome 2, a region containing an ancestral
490 telomere-telomere fusion site (54). The 5' ends of ARIA reads comprise C-rich telomeric repeats
491 near the junction of the telomere-telomere fusion site, following the unique sequences aligned
492 to chromosome 2 (**Supplementary Figure S5D**). TERRA transcripts were also observed near the
493 junction but were not directly connected to ARIA transcripts (**Supplementary Figure S5D**).
494
495 **The length of telomeric repeats in TERRA transcripts**
496 We sought to elucidate the length of telomeric repeat tracts within TERRA molecules by analyzing
497 Nanopore sequencing reads that aligned to subtelomeric regions and extended into the
498 telomeric repeats. Among TERRA reads mapping to chromosome ends, the maximum lengths are
499 2478 bp in U2OS cells and 4639 bp in HeLa cells (**Figures 2A-2B)**. The mean lengths of bulk TERRA
500 reads were 721 bp in U2OS cells and 982 bp in HeLa cells. For the pure telomeric tracts in TERRA
501 reads, the maximum lengths observed were 1080 bp in U2OS cells and 1487 bp in HeLa cells,

502 while the mean lengths were 222 bp and 284 bp in U2OS and HeLa cells, respectively. Overall,
503 telomeric repeat tracts within TERRA reads ranged from several hundred nucleotides to over
504 one thousand nucleotides. Dot plots of Nanopore TERRA reads per chromosome arm, grouped
505 by types of TERRA transcription regions, revealed considerable variability in length of bulk TERRA
506 reads and telomeric repeats across chromosomes in both U2OS and HeLa cells (**Figure 2C**). The
507 maximum bulk TERRA read lengths were longer in Type I compared to Type II and Type III in both
508 cell cells. Mean telomeric repeat lengths in TERRA reads were slightly shorter in Type III in both
509 cell lines. Given the fragility of RNA molecules, it is possible that some TERRA transcripts were
510 partially degraded during sample preparation, potentially resulting in shorter observed lengths
511 and increased variation in transcript size. Nevertheless, our findings are consistent with previous
512 reports indicating that TERRA transcripts generally exceed 1kb in length and that the average
513 telomeric repeat length within TERRA is approximately 200~300 bp (32,47) .
514

515 **TERRA promoters are associated with H3K4me3, Pol II, CpGs, and R-loops at distinct repeat**
516 **elements in Type I TERRA transcription regions**

517 Next, we searched for common features among the transcription start sites of TERRA. ChIP-seq
518 data for RNA polymerase II (48), H3K4me3 (49), CTCF (50) and DNA methylation (51) were
519 analyzed using T2T-CHM13 reference genome. Additionally, DRIP-seq data (52) was investigated
520 to examine the profiles of R-loops surrounding the 61-29-37 bp repeats of TERRA promoters
521 located in the subtelomeric regions of the Type I TERRA transcription regions. The meta-analysis
522 revealed epigenetic profiles corresponding to each repeat element (**Figure 3A**). The 61 bp repeat
523 element exhibited an enrichment of CTCF, RNA polymerase II, and R loops. The 29 bp repeat
524 element showed an enrichment of DNA methylation compared to other repeats. Remarkably,
525 CAGE tags and H3K4me3 were enriched at the 37 bp repeat regions. The distribution of analyzed
526 epigenetic marks along the TERRA promoter was illustrated in **Figure 3B**, and an example of the
527 genomic view of these marks was shown in **Figure 3C**. We conclude that DNA methylation
528 predominantly occurs at 29 bp repeats, while RNA polymerase II and H3K4me3 are enriched at
529 61 bp and 37 bp repeats respectively in Type I TERRA transcription regions.

530 Our observations suggest that TERRA are transcribed mainly from the 61-29-37 bp repeat
531 regions. Only chr10p and chr17p do not contain the 61-29-37 repeats adjacent to TERRA
532 transcription regions. It's worth noting that while certain TERRA transcription regions lack 29 bp
533 repeats, they still consist of 61-37 bp repeats. Intriguingly, even in regions without 29 bp repeats,
534 DNA methylation was still detected in the CpG islands between the 61 and 37 bp repeat tracts
535 (**Supplementary Figure S6A**).
536 To investigate the correlation between these epigenetic marks and TERRA levels, we
537 compared TERRA expression with the abundance of these epigenetic marks at TERRA promoters.
538 The promoter regions were selected from 3 kb upstream of the transcription start sites, where
539 the 61-29-37 bp repeats were located. Scatter plots demonstrated that TERRA expression is
540 positively correlated with H3K4me3, R-loops, and CTCF at chromosome ends (**Figure 3D**).
541
542 **TERRA-QUANT for measuring TERRA expression**
543 To analyze TERRA expression using publicly available short-read RNA-seq datasets, we
544 established a bioinformatics tool, dubbed TERRA-QUANT, for measuring TERRA levels based on
545 the transcription regions defined in this study (**Figure 4A**). TERRA-QUANT quantifies reads
546 mapped to TERRA transcription regions at each chromosome end, including reads from Type I, II,
547 and III. For total TERRA levels from chromosome ends, only TERRA reads mapped to Type I and II
548 TERRA transcription regions were selected and subsequently underwent YARN normalization. For
549 chromosome-end-specific TERRA analysis, we separated reads from subtelomeric regions and
550 pure telomeric repeat tracts, and only quantified subtelomeric reads and excluded pure
551 telomeric repeat reads.
552 To test the reliability of this tool, we applied TERRA-QUANT to analyze TERRA expression
553 in HeLa cells using Illumina sequencing data obtained from a previous study (32). The previous
554 study enriched TERRA RNA by using biotinylated antisense oligos to capture TERRA from nuclear
555 RNA. Notably, TERRA read counts were detected at the majority of chromosome ends in the
556 TERRA capture group, whereas they were absent at multiple chromosome ends in the control
557 group without TERRA capture (**Figures 4B**). TRF2 deficiency led to increased TERRA read counts
558 at multiple chromosome ends compared to wildtype cells (**Figures 4B**). Additionally, TERRA read

559 counts from chromosome ends in HeLa cells were significantly lower than those in U2OS cells
560 (**Figure 4C**). These results are consistent with earlier studies showing elevated TERRA levels upon
561 TRF2 depletion (4,32), affirming the robustness of the TERRA-QUANT methodology.
562 To further validate the annotations for TERRA transcription regions, we designed
563 subtelomeric primer sets specific to CHM13-1q, 3q, 8p, and 15q chromosome ends based on the
564 T2T-CHM13 genome sequence (**Supplementary Table 6**). We used these subtelomeric primers
565 to detect TERRA following TERRA capture and observed over 1000-fold enrichment
566 (**Supplementary Figure S6B)**, indicating that the primers efficiently amplified TERRA transcripts.
567 Notably, RT-qPCR analysis using these subtelomeric primers revealed increased TERRA levels
568 from various chromosome ends in U2OS compared to HeLa cells (**Figure 4D**). Subsequently, the
569 RT-qPCR products were subjected to Illumina sequencing to validate their sequences. We
570 observed that the reads of RT-qPCR products were predominantly mapped to their target
571 chromosome ends (**Supplementary Figure S6C**). These results indicate the specificity of these
572 subtelomeric primers.
573
574 **Poly(A)+ and non-poly(A) TERRA capture sequencing**
575 Since most publicly available RNA-seq datasets are generated from polyadenylated RNA--and
576 only a portion of TERRA transcripts are polyadenylated--we sought to investigate the genomic
577 origins of TERRA molecules with and without poly(A) tails. To achieve this, we performed TERRA
578 capture followed by an additional poly(A) selection step to separate poly(A)+ and poly(A)-
579 fractions. Both poly(A)+ and poly(A)- TERRA fractions were subjected to Illumina sequencing. We
580 analyzed the ratio of poly(A)+ to poly(A)- TERRA reads in both U2OS and HeLa cells and found
581 that certain chromosome ends, such as chr1p, chr2q, chr3p, and 18q, produced more poly(A)+
582 TERRA transcripts in both cell lines (New **Figure 4E**; **Supplementary Figure S6D**). In contrast, other
583 chromosome ends, including chr8p, chr11p, chr15q, and chr21q, produced more poly(A)- TERRA
584 transcripts. Notably, we observed a strong positive correlation (Pearson $r=0.7$, $p<0.0001$)
585 between U2OS and HeLa cells in the ratio of poly(A)+ to poly(A)- across different chromosome
586 ends (**Figure 4F**), suggesting that polyadenylation patterns in TERRA are largely consistent across
587 cell types.

588

589 **ALT positive osteosarcomas show elevated TERRA levels**

590 We gathered rRNA-depleted RNA-seq datasets from a previous study, which included three
591 biological replicates of 13 human osteosarcoma cell lines, with 4 cell lines categorized as ALT
592 negative (n=12) and 9 cell lines classified as ALT positive (n=27) (39). The RNA-seq reads of these
593 datasets were aligned to the T2T-CHM13 genome and processed using TERRA-QUANT. Total
594 TERRA levels were then determined by cumulating TERRA reads from all chromosome ends.
595 Notably, ALT-positive cells exhibited a higher level of total TERRA, and a lower RNA level of
596 telomerase reverse transcriptase (TERT) (**Figures 5A-5B**). These results agree with the previous
597 finding showing ALT cells with elevated TERRA expression (2,47).

598

599 **TERRA increases along with human aging in blood cells**

600 To compare TERRA expression across different ages, we collected poly(A)-enriched RNA-seq
601 datasets from a previous study (47), which comprised 222 human blood samples for analysis.
602 These blood samples cover a range of ages from 24 to 73 years old and were separated into three
603 groups: young (<30 years old), mid-age (30-59 years old), and old groups (≥60 years old).
604 Interestingly, TERRA-QUANT analysis demonstrated that total TERRA levels accumulated from all
605 chromosome ends were positively correlated with age (**Figure 5C**). When grouped by different
606 ages, total TERRA levels were significantly increased in old individuals, compared to young and
607 mid-age individuals (**Figure 5D**). For chromosome-end-specific analysis, reads mapping to
608 individual chromosome ends were low, preventing statistical analysis (**Supplementary Figure**
609 **S7A**). Notably, TERRA reads were predominately mapped to Type I transcription regions
610 (**Supplementary Figure S7B**).

611 Next, we confirmed the association of TERRA levels with aging by collecting blood samples
612 from individuals of various ages and conducted quantitative PCR to measure telomere length and
613 TERRA levels. The samples were grouped by adult (21-60y) or old (> 60y) ages. Genomic DNA was
614 extracted from buffy coats, and telomere length (T/S ratio) was analyzed by quantitative PCR.
615 We observed that telomere lengths significantly decreased in elderly people in both males and
616 females (**Figure 5E**). Interestingly, total TERRA RNA levels in blood samples determined by RT-

617 qPCR were significantly elevated in old males and females (**Figure 5F**), and were negatively
618 correlated with telomere lengths (**Figure 5G**). Using subtelomeric primers for RT-qPCR also
619 showed a significant increase in TERRA expression in the old group compared to the adult group
620 (**Figure 5H**), confirming an elevation in TERRA levels in aged leukocytes.
621
622 **TERRA is upregulated in the aged brain**

623 To explore the relationship between TERRA and human aging, we analyzed published poly(A)-
624 enriched RNA-seq datasets obtained from various human tissues including brains, hearts, and
625 ovaries (40) by TERRA-QUANT. The scatter plots displayed a significant positive correlation
626 between total TERRA levels and age in brains, but not in ovaries or hearts (**Figure 6A**).
627 Remarkably, total TERRA levels in brain tissues exhibited an upregulation in the old group (≥60
628 years old), compared to the adult group (30-59 years old) (**Figure 6B**). Considering that ATRX is
629 important for neuronal development and functionality (53-55), we analyzed ATRX expression in
630 human brains. Interestingly, the levels of ATRX, which is a TERRA-interacting protein (24),
631 declined in old individuals (**Supplementary Figure S7C**) and displayed an anticorrelation with
632 TERRA expression in brain tissues (**Supplementary Figure S7D**).
633
634 **Elevated TERRA levels in aged fibroblasts and abnormal TERRA profiles in Hutchinson-Gilford**
635 **progeria syndrome patients**
636 Next, we quantified TERRA expression in fibroblasts obtained from healthy individuals of various
637 ages and patients diagnosed with Hutchinson-Gilford progeria syndrome (HGPS), a genetic
638 disorder characterized by premature aging features in childhood (56). Total TERRA levels,
639 analyzed by TERRA-QUANT using poly(A)-enriched RNA-seq datasets, displayed a significant
640 upregulation in fibroblasts derived from healthy older individuals (**Figures 6C-6D**), with an
641 exponential increase observed in those over 75 years old. Chromosome-end-specific analysis
642 showed that TERRA increased at multiple chromosome ends in the old group (**Supplementary**
643 **Figure S8A**). Type I TERRA transcription was dominant in fibroblasts compared to Type II and III
644 (**Supplementary Figure S8B**). Considering that HGPS patients were below 30 years old, we
645 compared TERRA levels with those of their age-matched controls. Notably, HGPS patients

646 displayed an aberrant and diverse TERRA expression pattern compared to healthy individuals
647 (**Figure 6E**).
648
649 **Single-cell RNA-seq analysis reveals high TERRA levels in neurons**
650 We assessed TERRA levels across various cell types by analyzing published RNA-seq
651 datasets from human tissues, including adipose, blood, bone, brain, heart, lung, ovary, pancreas,
652 and retina (**Supplementary Table 7**). Among these tissues, TERRA levels were particularly high
653 in blood, bone, and ovary (**Figure 7A**). Given the highest level in blood cells, we sought to identify
654 the specific cell types exhibiting high TERRA expression. To achieve this, we analyzed single-cell
655 RNA-seq datasets (59) from peripheral blood mononuclear cells (PBMC) obtained from healthy
656 young individuals. Notably, dendritic cells displayed the highest TERRA level among various
657 immune cells (**Figure 7B**).
658 To investigate the potential role of TERRA in neurons, we examined TERRA levels using
659 single-cell RNA-seq datasets obtained from human embryonic stem cells (hESCs) undergoing
660 neuronal differentiation (57). The hESCs were converted to neuromesodermal progenitors
661 (NMPs) with distinct HOX profiles along rostrocaudal axis by modulation of Wnt, FGF and retinoic
662 acid signaling, and were further differentiated into ventral neuron phenotypes. Six different NMP
663 cultures from hESCs were generated corresponding to HOX patterning periods: 24, 48, 72, 120,
664 168, and 216 hours. The cultures exhibited a progressive expression of caudal HOX paralogs,
665 corresponding to cervical (HOX1-8; observed in 24h,48h, and 72h), thoracic (HOX1-9; observed
666 in 120h), lumbar (HOX1-11; observed in 168h), and lumbosacral spinal regions (HOX1-13;
667 observed in 216h). Cells were clustered based on gene expression into mesenchymal cells,
668 neuroepithelial cells, and neurons. Notably, higher levels of TERRA were observed in neurons
669 compared to neuroepithelial cells and mesenchymal cells (**Figure 7C**). Additionally, TERRA levels
670 displayed a concordant expression pattern with ATRX during neuronal differentiation (**Figures**
671 **7D-7F**).
672
673 **Elevated TERRA levels during the early stage of Alzheimer's disease progression**

674 Next, we sought to analyze TERRA levels using TERRA-QUANT in neuronal disorder diseases. We
675 analyzed published single-nucleus RNA-seq datasets from the prefrontal cortex of patients with
676 Alzheimer's disease (AD) (58). Postmortem human brain samples were obtained from
677 participants enrolled in the Religious Order Study and Rush Memory and Aging Project (ROSMAP)
678 (59). Individuals with AD pathology were categorized into two subgroups according to the
679 previous study (58), reflecting distinct stages of AD pathological progression: 1) early-pathology
680 characterized by amyloid burden, modest neurofibrillary tangles, and cognitive impairment; and
681 2) late-pathology marked by higher amyloid levels, elevated neurofibrillary tangles, increased
682 global pathology, and cognitive impairment. Notably, increased TERRA levels were observed in
683 various types of brain cells in the early stage of AD pathology compared to no AD pathology, with
684 more profound TERRA levels in excitatory neurons (**Figure 7G, Supplementary Figure S9A**). To
685 further investigate TERRA in AD neurons, induced pluripotent stem cells (iPSCs) derived from an
686 AD patient, along with mutation-corrected isogenic control cells, were generated. The AD patient
687 carried the P117L mutation in presenilin-1 (PS1mut), the catalytic subunit of γ-secretase. Notably,
688 RT-qPCR demonstrated that AD-PS1mut neuronal cells differentiated from iPSCs exhibited
689 significantly higher TERRA levels compared to control cells (**Figure 7H**), further supporting the
690 upregulation of TERRA expression in AD neurons.
691
692 **Discussion**
693 In this study, we investigated TERRA transcripts in human cells using Nanopore direct RNA-seq
694 and TERRA capture RNA-seq. We identified TERRA transcription regions at 39 chromosome ends
695 and utilized these defined regions to quantify TERRA expression across various tissues. Our
696 bioinformatics pipeline is versatile, accommodating different RNA-seq datasets, including single-
697 cell RNA-seq experiments. Notably, substantial mapped reads per cell were observed in single-
698 cell RNA-seq datasets without YARN normalization (**Supplementary Figures S9A-S9C**), indicating
699 that TERRA can be detected at the single-cell level.
700 Some discrepancies were observed between our findings and those reported by
701 Rodrigues et al. (45) regarding TERRA expression at specific chromosome ends in HeLa and U2OS
702 cells. A comparison of our and their results revealed consistent identification of TERRA expression

703 at 31 chromosome ends, while 3 chromosome ends were consistently found to lack detectable
704 TERRA expression (**Supplementary Figure S9D**). Our analysis revealed 8 additional chromosome
705 ends expressing TERRA that were not reported by Rodrigues et al., while their study identified 4
706 ends absent from our dataset. We attribute these differences primarily to the distinct mapping
707 strategies applied. Rodrigues et al. used a limited reference consisting of 25 kb subtelomeric
708 regions, whereas we utilized the entire genome for read alignment. When we applied 25 kb
709 subtelomeric regions as a reference genome for mapping our data, the results aligned closely
710 with their findings. These observations suggest that the choice of mapping strategy impacts the
711 sensitivity and specificity of TERRA detection at individual chromosome ends.

712 Commonly available datasets largely comprise RNA samples enriched with poly-A tails.
713 Nevertheless, a significant portion of TERRA molecules lack poly-A tails, which may limit the
714 detection of TERRA using poly-A enriched RNA-seq data. To investigate potential biases toward
715 specific chromosome ends in datasets derived from poly(A)-enriched RNA-seq libraries, we
716 performed poly(A)+ and poly(A)- TERRA capture sequencing in U2OS and HeLa cells. Our analysis
717 revealed that poly(A)+ TERRA transcripts may preferentially be produced from certain
718 chromosome ends and vice versa, consistent with the previous study (7). This preference is
719 conserved in both U2OS and HeLa cells. Although there were detectable preference in TERRA
720 detection from specific chromosome ends using poly(A)-enriched RNA-seq datasets, we still
721 observed TERRA originating from multiple chromosome ends in aged fibroblasts, despite this bias
722 (**Supplementary Figure S8A**). For analysis of total TERRA levels in human tissues, we accumulated
723 TERRA reads from all chromosome ends to enhance read counts, thereby increasing overall
724 TERRA counts for quantifying total TERRA levels. Due to low read counts mapped to individual
725 chromosome ends, analyzing chromosome-end-specific TERRA levels remains challenging using
726 publicly available RNA-seq datasets from human tissues. TERRA-enriched sequencing and longer
727 read lengths could enable more precise quantification of TERRA at specific chromosome ends.

728 Utilizing TERRA-QUANT for quantifying TERRA levels across diverse tissues, we detected
729 TERRA expression in most tissues examined. Notably, high TERRA levels were observed in blood
730 and bone tissues, raising the possibility of a role for TERRA in immune-related processes. The
731 earlier research has shown that TERRA can be secreted via extracellular exosomes to stimulate

732 inflammatory cytokines during telomere dysfunction (60,61). TERRA also interacts with ZBP1 to
733 activate immune response under conditions of replicative crisis (62). However, it is important to
734 note that such ZBP1-TERRA complex-mediated immune activation is unlikely to occur under
735 normal physiological conditions in healthy blood cells. Notably, we observed elevated total
736 TERRA levels in fibroblasts, brain, and blood cells as individuals aged. The telomere shortening
737 may contribute to TERRA upregulation during aging. However, neurons are postmitotic cells that
738 are not expected to undergo telomere shortening. Therefore, the increase of TERRA in the brain
739 is unlikely to result from telomere shortening, but may instead be associated with oxidative stress,
740 neuronal damage, and telomere dysfunction in the aging brain. It has been previously reported
741 that telomeric transcription is induced by DNA damage at telomeres in Hutchinson-Gilford
742 progeria syndrome (HGPS) fibroblasts (33). In our analysis, fibroblasts derived from HGPS
743 patients demonstrated abnormal TERRA expression, with some individuals showing high levels
744 while others displayed lower levels. We reason that the diverse TERRA counts observed in HGPS
745 patients could be attributed to variable telomere lengths or defects in telomere integrity.
746 Previous studies have shown that telomere length in HGPS patients is generally shorter, yet
747 exhibits variability (63,64). Low TERRA counts in some HGPS patients could result from either low
748 TERRA expression or the loss of subtelomeric and telomeric DNA at chromosome ends.

749 Examination of single-nucleus RNA-seq datasets and RT-qPCR analysis of TERRA levels
750 unveiled an upregulation of TERRA levels in Alzheimer's disease (AD) neurons. AD is associated
751 with several cellular stress responses—such as oxidative stress and chronic inflammation (65-
752 67)—that are also known to upregulate TERRA expression (37,68). Moreover, the accumulation
753 of DNA G-quadruplex (G4) structures has been reported in AD neurons (69), and TERRA has also
754 been implicated in promoting G4 formation (27). These results indicate elevated TERRA levels
755 correlate with increased DNA G4 formation and oxidative stress, which are aligned with the
756 observations in AD neurons. Taken together, these connections provide a compelling rationale
757 for further investigation into the role of TERRA in AD pathogenesis.

758 Upregulation of TERRA and ATRX during neuronal differentiation implies the potential
759 function of TERRA in neuronal cells. Interestingly, the inverse relationship between TERRA and
760 ATRX in the aged brain was observed, suggesting that low levels of ATRX might contribute to the

761 dysfunction of the aged neurons. Previous research indicated that the absence of ATRX leads to
762 cognitive impairment and promotion of cellular senescence (70), further highlighting the
763 significance of exploring the interplay between TERRA, ATRX, and aging-related cellular processes
764 in the brain and neuronal tissues. It is plausible that proper levels of TERRA and ATRX are crucial
765 for maintaining normal gene expression in neuronal cells, and dysregulation of TERRA and ATRX
766 may impact neuron function during aging.

767 Our study provides the annotation of TERRA transcripts in the T2T-CHM13 genome
768 reference and introduces the bioinformatics tool "TERRA-QUANT" for quantifying TERRA levels.
769 This lays the foundation for studying TERRA in diverse conditions, cell types, and human diseases.
770

771 **Data availability**

772 TERRA capture RNA-seq datasets including Illumina short-read RNA-seq, Nanopore direct RNA-
773 seq, poly(A)+ and poly(A)- TERRA capture RNA-seq, and the sequencing data of TERRA RT-qPCR
774 products are accessible at GEO (GEO accession GSE250303)
775 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250303. The published sequencing
776 datasets analyzed in this study are described in **Supplementary Tables 7, 8, and 9**.

777

778 **Code availability**

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

782

783 **Funding**

784 This work was supported by National Science and Technology Council (NSTC) grants NSTC 112-
785 2628-B-002-008 (H-P. C. C.), NSTC 112-2320-B-002-058 (H-P. C. C.), NSTC 113-2320-B-002-009 (H-
786 P. C. C.), NSTC 113-2628-B-002-010-MY3 (H-P. C. C.) and National Taiwan University grants NTU-
787 111L7880 (H-P. C. C.), NTU-AS-112L104312 (H-P. C. C), NTU-CDP-112L7721 (H-P. C. C.), NTU-CDP-
788 113L7705 (H-P. C. C.), and National Health Research Institutes grants NHRI-EX111-11107SI, NHRI-
789 EX112-11107SI (H-P. C. C.), NHRI-EX113-11107SI (H-P. C. C.), NHRI-EX114-11411SI.

790

791 **Conflict of Interest declaration**

792 The authors declare that they have no competing interests.

793

794 **Acknowledgments**

795 We appreciated the insightful discussions with Liuh-Yow Chen, Ching-Shyi Peter Wu, as well as
796 the contributions of our lab members Hong-Jhin Shen, Chia-Yu Guh, and Liv Weichien Chen. We
797 thank Kevin, Tsai at IBMS, Academia Sinica, for his support with Nanopore Direct RNA-seq. We
798 thank Technology Commons, College of Life Science, National Taiwan University for supporting
799 the usage of equipment, and National Center for High-performance Computing (NCHC) of the

National Institutes of Applied Research of Taiwan (NIAR) for providing computational and storage resources.

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. H-J. S. analyzed TERRA levels in AD neurons. C-H. Y and H-C. K contributed to AD iPSC generation and neuron differentiation. 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.

813 **References**
814

815 1. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric
816 repeat containing RNA and RNA surveillance factors at mammalian chromosome ends.
817 *Science*, **318**, 798-801.
818 2. Schoeftner, S. and Blasco, M.A. (2008) Developmentally regulated transcription of
819 mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*, **10**, 228-236.
820 3. Feretzaki, M., Renck Nunes, P. and Lingner, J. (2019) Expression and differential
821 regulation of human TERRA at several chromosome ends. *RNA*, **25**, 1470-1480.
822 4. Nergadze, S.G., Farnung, B.O., Wischnewski, H., Khoriauli, L., Vitelli, V., Chawla, R.,
823 Giulotto, E. and Azzalin, C.M. (2009) CpG-island promoters drive transcription of human
824 telomeres. *RNA*, **15**, 2186-2194.
825 5. Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., Jolivet, P., Teixeira,
826 M.T. and Luke, B. (2017) Telomere Length Determines TERRA and R-Loop Regulation
827 through the Cell Cycle. *Cell*, **170**, 72-85 e14.
828 6. Azzalin, C.M. and Lingner, J. (2008) Telomeres: the silence is broken. *Cell Cycle*, **7**, 1161-
829 1165.
830 7. Savoca, V., Rivosecchi, J., Gaiatto, A., Rossi, A., Mosca, R., Gialdini, I., Zubovic, L., Tebaldi,
831 T., Macchi, P. and Cusanelli, E. (2023) TERRA stability is regulated by RALY and
832 polyadenylation in a telomere-specific manner. *Cell Rep*, **42**, 112406.
833 8. Rossi, A., Moro, A., Tebaldi, T., Cornella, N., Gasperini, L., Lunelli, L., Quattrone, A., Viero,
834 G. and Macchi, P. (2017) Identification and dynamic changes of RNAs isolated from
835 RALY-containing ribonucleoprotein complexes. *Nucleic Acids Res*, **45**, 6775-6792.
836 9. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K. and Selig, S. (2008)
837 Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally
838 short telomeres and enhanced transcription from telomeric regions. *Hum Mol Genet*,
839 **17**, 2776-2789.
840 10. Sagie, S., Toubiana, S., Hartono, S.R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel,
841 C., Velasco, G., Chedin, F. and Selig, S. (2017) Telomeres in ICF syndrome cells are
842 vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat Commun*, **8**, 14015.
843 11. Deng, Z., Campbell, A.E. and Lieberman, P.M. (2010) TERRA, CpG methylation and
844 telomere heterochromatin: lessons from ICF syndrome cells. *Cell Cycle*, **9**, 69-74.
845 12. Deng, Z., Wang, Z., Stong, N., Plasschaert, R., Moczan, A., Chen, H.S., Hu, S.,
846 Wikramasinghe, P., Davuluri, R.V., Bartolomei, M.S. *et al.* (2012) A role for CTCF and
847 cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end
848 protection. *EMBO J*, **31**, 4165-4178.
849 13. Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S.,
850 Petrini, J.H., Sung, P.A., Jasin, M. *et al.* (2012) Loss of ATRX, genome instability, and an
851 altered DNA damage response are hallmarks of the alternative lengthening of telomeres
852 pathway. *PLoS Genet*, **8**, e1002772.
853 14. Diman, A., Boros, J., Poulain, F., Rodriguez, J., Purnelle, M., Episkopou, H., Bertrand, L.,
854 Francaux, M., Deldicque, L. and Decottignies, A. (2016) Nuclear respiratory factor 1 and
855 endurance exercise promote human telomere transcription. *Sci Adv*, **2**, e1600031.

856 15. Tutton, S., Azzam, G.A., Stong, N., Vladimirova, O., Wiedmer, A., Monteith, J.A.,
857 Beishline, K., Wang, Z., Deng, Z., Riethman, H. *et al.* (2016) Subtelomeric p53 binding
858 prevents accumulation of DNA damage at human telomeres. *EMBO J*, **35**, 193-207.
859 16. Koskas, S., Decottignies, A., Dufour, S., Pezet, M., Verdel, A., Vourc'h, C. and Faure, V.
860 (2017) Heat shock factor 1 promotes TERRA transcription and telomere protection upon
861 heat stress. *Nucleic Acids Res*, **45**, 6321-6333.
862 17. Mazzolini, R., Gonzalez, N., Garcia-Garijo, A., Millanes-Romero, A., Peiro, S., Smith, S.,
863 Garcia de Herreros, A. and Canudas, S. (2018) Snail1 transcription factor controls
864 telomere transcription and integrity. *Nucleic Acids Res*, **46**, 146-158.
865 18. Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T. and Azzalin, C.M. (2014)
866 RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT
867 tumour cells. *Nat Commun*, **5**, 5220.
868 19. Yu, T.Y., Kao, Y.W. and Lin, J.J. (2014) Telomeric transcripts stimulate telomere
869 recombination to suppress senescence in cells lacking telomerase. *Proc Natl Acad Sci U S*
870 *A*, **111**, 3377-3382.
871 20. Guh, C.Y., Shen, H.J., Chen, L.W., Chiu, P.C., Liao, I.H., Lo, C.C., Chen, Y., Hsieh, Y.H.,
872 Chang, T.C., Yen, C.P. *et al.* (2022) XPF activates break-induced telomere synthesis. *Nat*
873 *Commun*, **13**, 5781.
874 21. Silva, B., Arora, R., Bione, S. and Azzalin, C.M. (2021) TERRA transcription destabilizes
875 telomere integrity to initiate break-induced replication in human ALT cells. *Nature*
876 *Communications*, **12**.
877 22. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. and Lieberman, P.M. (2009) TERRA RNA
878 Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at
879 Telomeres. *Molecular Cell*, **35**, 403-413.
880 23. Mei, Y., Deng, Z., Vladimirova, O., Gulve, N., Johnson, F.B., Drosopoulos, W.C.,
881 Schildkraut, C.L. and Lieberman, P.M. (2021) TERRA G-quadruplex RNA interaction with
882 TRF2 GAR domain is required for telomere integrity. *Sci Rep-Uk*, **11**.
883 24. Chu, H.P., Cifuentes-Rojas, C., Kesner, B., Aeby, E., Lee, H.G., Wei, C., Oh, H.J., Boukhali,
884 M., Haas, W. and Lee, J.T. (2017) TERRA RNA Antagonizes ATRX and Protects Telomeres.
885 *Cell*, **170**, 86-101 e116.
886 25. Flynn, R.L., Centore, R.C., O'Sullivan, R.J., Rai, R., Tse, A., Zhou, S.Y., Chang, S., Karlseder,
887 J. and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on
888 telomeric single-stranded DNA. *Nature*, **471**, 532-+.
889 26. Voon, H.P., Hughes, J.R., Rode, C., De La Rosa-Velazquez, I.A., Jenuwein, T., Feil, R.,
890 Higgs, D.R. and Gibbons, R.J. (2015) ATRX Plays a Key Role in Maintaining Silencing at
891 Interstitial Heterochromatic Loci and Imprinted Genes. *Cell Rep*, **11**, 405-418.
892 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
893 Chu, H.C. (2022) TERRA regulates DNA G-quadruplex formation and ATRX recruitment to
894 chromatin. *Nucleic Acids Res*, **50**, 12217-12234.
895 28. Arnoult, N., Van Beneden, A. and Decottignies, A. (2012) Telomere length regulates
896 TERRA levels through increased trimethylation of telomeric H3K9 and HP1alpha. *Nat*
897 *Struct Mol Biol*, **19**, 948-956.

898 29. Cusanelli, E., Romero, C.A. and Chartrand, P. (2013) Telomeric noncoding RNA TERRA is
899 induced by telomere shortening to nucleate telomerase molecules at short telomeres.
900 *Mol Cell*, **51**, 780-791.
901 30. Misino, S., Busch, A., Wagner, C.B., Bento, F. and Luke, B. (2022) TERRA increases at
902 short telomeres in yeast survivors and regulates survivor associated senescence (SAS).
903 *Nucleic Acids Res*, **50**, 12829-12843.
904 31. Caslini, C., Connelly, J.A., Serna, A., Broccoli, D. and Hess, J.L. (2009) MLL associates with
905 telomeres and regulates telomeric repeat-containing RNA transcription. *Mol Cell Biol*,
906 **29**, 4519-4526.
907 32. Porro, A., Feuerhahn, S., Delafontaine, J., Riethman, H., Rougemont, J. and Lingner, J.
908 (2014) Functional characterization of the TERRA transcriptome at damaged telomeres.
909 *Nat Commun*, **5**, 5379.
910 33. Aguado, J., Sola-Carvajal, A., Cancila, V., Revechon, G., Ong, P.F., Jones-Weinert, C.W.,
911 Wallen Arzt, E., Lattanzi, G., Dreesen, O., Tripodo, C. *et al.* (2019) Inhibition of DNA
912 damage response at telomeres improves the detrimental phenotypes of Hutchinson-
913 Gilford Progeria Syndrome. *Nat Commun*, **10**, 4990.
914 34. Maestroni, L., Reyes, C., Vaurs, M., Gachet, Y., Tournier, S., Geli, V. and Coulon, S. (2020)
915 Nuclear envelope attachment of telomeres limits TERRA and telomeric rearrangements
916 in quiescent fission yeast cells. *Nucleic Acids Res*, **48**, 3029-3041.
917 35. Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K. and Luke, B.
918 (2013) Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence.
919 *Nat Struct Mol Biol*, **20**, 1199-1205.
920 36. Wanat, J.J., Logsdon, G.A., Driskill, J.H., Deng, Z., Lieberman, P.M. and Johnson, F.B.
921 (2018) TERRA and the histone methyltransferase Dot1 cooperate to regulate senescence
922 in budding yeast. *PLoS One*, **13**, e0195698.
923 37. Liu, B., Maekawa, T., Yoshida, K., Ly, N.H., Inoue, K., Hasegawa, A., Chatton, B., Ogura, A.
924 and Ishii, S. (2019) Telomere shortening by transgenerational transmission of TNF-alpha-
925 induced TERRA via ATF7. *Nucleic Acids Res*, **47**, 283-298.
926 38. Chang, K.V., Chen, Y.C., Wu, W.T., Shen, H.J., Huang, K.C., Chu, H.P. and Han, D.S. (2020)
927 Expression of Telomeric Repeat-Containing RNA Decreases in Sarcopenia and Increases
928 after Exercise and Nutrition Intervention. *Nutrients*, **12**.
929 39. Mason-Osann, E., Dai, A., Floro, J., Lock, Y.J., Reiss, M., Gali, H., Matschulat, A., Labadorf,
930 A. and Flynn, R.L. (2018) Identification of a novel gene fusion in ALT positive
931 osteosarcoma. *Oncotarget*, **9**, 32868-32880.
932 40. Shokhirev, M.N. and Johnson, A.A. (2021) Modeling the human aging transcriptome
933 across tissues, health status, and sex. *Aging Cell*, **20**, e13280.
934 41. Hicks, S.C., Okrah, K., Paulson, J.N., Quackenbush, J., Irizarry, R.A. and Bravo, H.C. (2018)
935 Smooth quantile normalization. *Biostatistics*, **19**, 185-198.
936 42. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B.,
937 Wheeler, T.D., McDermott, G.P., Zhu, J. *et al.* (2017) Massively parallel digital
938 transcriptional profiling of single cells. *Nat Commun*, **8**, 14049.
939 43. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava,
940 A., Molla, G., Madad, S., Fernandez-Granda, C. *et al.* (2023) Dictionary learning for
941 integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*.

942 44. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S.,
943 Howe, E., Porubsky, D., Logsdon, G.A. *et al.* (2020) Telomere-to-telomere assembly of a
944 complete human X chromosome. *Nature*, **585**, 79-84.
945 45. Rodrigues, J., Alfieri, R., Bione, S. and Azzalin, C.M. (2024) TERRA ONTseq: a long-read-
946 based sequencing pipeline to study the human telomeric transcriptome. *RNA*, **30**, 955-
947 966.
948 46. Azzalin, C.M. and Lingner, J. (2015) Telomere functions grounding on TERRA firma.
949 *Trends Cell Biol*, **25**, 29-36.
950 47. Rosso, I., Jones-Weinert, C., Rossiello, F., Cabrini, M., Brambillasca, S., Munoz-Sagredo,
951 L., Lavagnino, Z., Martini, E., Tedone, E., Garre, M. *et al.* (2023) Alternative lengthening
952 of telomeres (ALT) cells viability is dependent on C-rich telomeric RNAs. *Nat Commun*,
953 **14**, 7086.
954 48. Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B.,
955 Bernard, X., Masamsetti, V.P., Roston, S. *et al.* (2018) EWS-FLI1 increases transcription
956 to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature*, **555**, 387-391.
957 49. Jang, S.M., Zhang, Y., Utani, K., Fu, H., Redon, C.E., Marks, A.B., Smith, O.K., Redmond,
958 C.J., Baris, A.M., Tulchinsky, D.A. *et al.* (2018) The replication initiation determinant
959 protein (RepID) modulates replication by recruiting CUL4 to chromatin. *Nat Commun*, **9**,
960 2782.
961 50. Ibarra, A., Benner, C., Tyagi, S., Cool, J. and Hetzer, M.W. (2016) Nucleoporin-mediated
962 regulation of cell identity genes. *Genes Dev*, **30**, 2253-2258.
963 51. Wang, Y., Khan, A., Marks, A.B., Smith, O.K., Giri, S., Lin, Y.C., Creager, R., MacAlpine,
964 D.M., Prasanth, K.V., Aladjem, M.I. *et al.* (2017) Temporal association of ORCA/LRWD1
965 to late-firing origins during G1 dictates heterochromatin replication and organization.
966 *Nucleic Acids Res*, **45**, 2490-2502.
967 52. De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O. and Capranico,
968 G. (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by
969 R loops in human cancer cells. *Proc Natl Acad Sci U S A*, **116**, 816-825.
970 53. Gibbons, R.J., Picketts, D.J., Villard, L. and Higgs, D.R. (1995) Mutations in a putative
971 global transcriptional regulator cause X-linked mental retardation with alpha-
972 thalassemia (ATR-X syndrome). *Cell*, **80**, 837-845.
973 54. Quesnel, K.M., Martin-Kenny, N. and Berube, N.G. (2023) A mouse model of ATRX
974 deficiency with cognitive deficits and autistic traits. *J Neurodev Disord*, **15**, 39.
975 55. Berube, N.G., Mangelsdorf, M., Jagla, M., Vanderluit, J., Garrick, D., Gibbons, R.J., Higgs,
976 D.R., Slack, R.S. and Picketts, D.J. (2005) The chromatin-remodeling protein ATRX is
977 critical for neuronal survival during corticogenesis. *J Clin Invest*, **115**, 258-267.
978 56. Gilford, H. (1897) On a Condition of Mixed Premature and Immature Development. *Med*
979 *Chir Trans*, **80**, 17-46 25.
980 57. Iyer, N.R., Shin, J., Cuskey, S., Tian, Y., Nicol, N.R., Doersch, T.E., Seipel, F., McCalla, S.G.,
981 Roy, S. and Ashton, R.S. (2022) Modular derivation of diverse, regionally discrete human
982 posterior CNS neurons enables discovery of transcriptomic patterns. *Sci Adv*, **8**,
983 eabn7430.

58. 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.
59. 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.
60. Wang, Z. and Lieberman, P.M. (2016) The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol*, **13**, 690-695.
61. 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.
62. 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.
63. 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.
64. 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.
65. Klein, J.A. and Ackerman, S.L. (2003) Oxidative stress, cell cycle, and neurodegeneration. *J Clin Invest*, **111**, 785-793.
66. Ionescu-Tucker, A. and Cotman, C.W. (2021) Emerging roles of oxidative stress in brain aging and Alzheimer's disease. *Neurobiol Aging*, **107**, 86-95.
67. Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G.M., Cooper, N.R., Eikelenboom, P., Emmerling, M., Fiebich, B.L. *et al.* (2000) Inflammation and Alzheimer's disease. *Neurobiol Aging*, **21**, 383-421.
68. Galigniana, N.M., Charo, N.L., Uranga, R., Cabanillas, A.M. and Piwien-Pilipuk, G. (2020) Oxidative stress induces transcription of telomeric repeat-containing RNA (TERRA) by engaging PKA signaling and cytoskeleton dynamics. *Biochim Biophys Acta Mol Cell Res*, **1867**, 118643.
69. Hanna, R., Flamier, A., Barabino, A. and Bernier, G. (2021) G-quadruplexes originating from evolutionary conserved L1 elements interfere with neuronal gene expression in Alzheimer's disease. *Nat Commun*, **12**, 1828.
70. 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.

1023 **Figure legends**

1024 **Figure 1. Identification of TERRA transcription regions by short-read and long-read RNA-seq in**
1025 **human cells.**

1026 A. The flowchart of TERRA-capture RNA-seq. TERRA was captured by using biotinylated
1027 antisense oligos. Captured TERRA RNAs were subjected to Illumina or Oxford nanopore
1028 long-read RNA direct sequencing.
1029 B. Definition of three types of TERRA transcription regions. TERRA transcription start sites
1030 were determined by TERRA enrichment from Illumina short-reads, Nanopore long reads
1031 and CAGE-seq.
1032 C. Heatmap represents TERRA enrichment (log2 values) at different chromosome ends and
1033 ITSs based on Illumina read counts of TERRA transcripts in U2OS cells. Capture: TERRA
1034 capture RNA-seq. Control: regular RNA-seq with ribosomal RNA depletion.
1035 D. Heatmap represents TERRA read counts from different chromosome ends and ITSs in
1036 U2OS cells. Reads from Illumina or Nanopore sequencing were mapped to T2T-CHM13.
1037 CPM: counts per million mapped reads.
1038 E. Pie charts show the numbers of Type I, II, III TERRA reads from Nanopore or Illumina
1039 sequencing in U2OS cells.
1040

1041 **Figure 2. TERRA length distribution from nanopore reads in U2OS and HeLa cells.**

1042 A. Density plots showing the distribution of TERRA bulk read lengths and telomeric repeat
1043 lengths in TERRA reads in U2OS and HeLa cells. Nanopore reads mapped to T2T-CHM13
1044 TERRA transcription regions were calculated.
1045 B. Statistic analysis of TERRA bulk read lengths and telomeric repeat lengths in TERRA reads.
1046 C. Dotplots show TERRA bulk read lengths and telomeric repeat lengths in TERRA reads at
1047 each chromosome end and ITS. Each dot represents each nanopore read. Solid bar,
1048 median.
1049

1050 **Figure 3. Enrichments of H3K4me3, RNA Pol II, CpGs, DNA methylation and R-loop at TERRA**
1051 **promoters.**

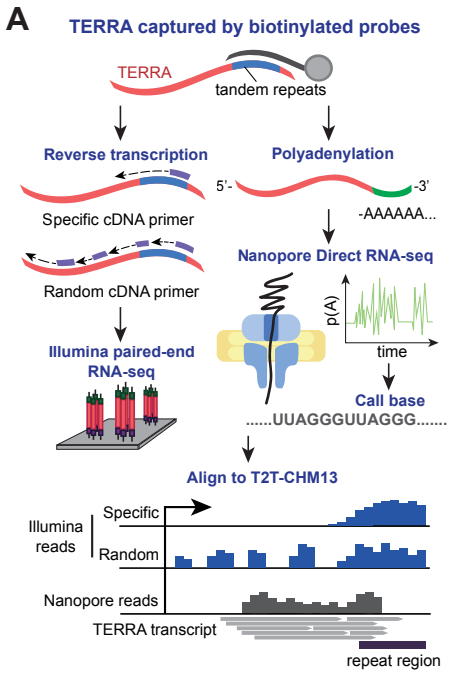
1052 A. Meta-analysis showing the distribution of epigenetic marks on 61-29-37 bp repeats
1053 located at subtelomeric regions (Type I promoter). Each plot represents the coverage of
1054 indicated marks.
1055 B. A schematic model of the epigenetic profiles at Type I TERRA promoter near telomeres.
1056 C. Genome browser view showing the coverage of indicated epigenetic marks at TERRA
1057 promoter on chr22q arm.
1058 D. Scatter plots of TERRA v.s. H3K4me3, R-loop (DRIP-seq), DNA methylation (MeDIP-seq),
1059 CTCF, or RNAPII near TERRA transcription start sites. Each dot indicates TERRA enrichment
1060 and epigenetic marks at individual chromosome ends (Type I +Type II TERRA transcription
1061 regions) in U2OS cells. TERRA enrichment (log2 ratio) was calculated by comparing TERRA
1062 capture and no capture. P values, by Pearson's correlation.
1063

1064 **Figure 4. TERRA-QUANT for measuring TERRA expression.**

1065 A. The workflow of bioinformatics pipeline for quantification of TERRA expression. Reads
1066 that mapped to TERRA transcription regions were counted. YARN was used for
1067 normalization between different tissues.
1068 B. Heatmap showing TERRA expression from individual chromosome ends and ITSs in HeLa
1069 and U2OS cells. TERRA capture, using antisense oligos to enrich TERRA. Control, no TERRA
1070 capture.
1071 C. TERRA normalized counts were analyzed by TERRA-QUANT. Each dot indicates TERRA
1072 counts at individual chromosome ends and ITSs. TRF2 deletion (TRF2KO) in HeLa cells
1073 increases TERRA expression. P values, by Wilcoxon matched-pairs signed rank test. Bars,
1074 median.
1075 D. RT-qPCR to detect TERRA in HeLa and U2OS cells. Subtelomeric primers for TERRA were
1076 designed based on the T2T-CHM13 genome sequence. Data from three biological
1077 replicates. P values, by two-tailed Student's t-test. Bars, mean±SD.
1078 E. Log2 ratios of poly(A)+ to poly(A)- TERRA reads mapped to individual chromosome ends
1079 and ITSs in U2OS and HeLa cells.
1080 F. Scatter plots showing the correlation between U2OS and HeLa cells in the
1081 poly(A)+/poly(A)- TERRA ratios. Each dot represents the ratio (log2) of poly(A)+/poly(A)-
1082 reads at each TERRA transcription region.
1083
1084 **Figure 5. TERRA levels increase with age in blood cells.**
1085 A. TERRA expression in ALT positive (+) or negative (-) cell lines was quantified by TERRA-
1086 QUANT. Total TERRA counts at all chromosome ends were accumulated. Each dot
1087 indicates the normalized TERRA read counts of each RNA-seq dataset. Bars, mean±SD. P
1088 values, by Mann-Whitney U test.
1089 B. TERT expression of each RNA-seq data from ALT+ or ALT- cells. Bars, mean±SD. P values,
1090 by Mann-Whitney U test.
1091 C. TERRA expression in human blood cells was analyzed by TERRA-QUANT using RNA-seq
1092 datasets. Scatter plots showing the correlation of TERRA v.s. age in blood. Each dot
1093 indicates the total TERRA normalized reads from all chromosome ends of each individual.
1094 P values, by Pearson's correlation.
1095 D. Normalized TERRA counts in blood cells of different ages. Young (<30 years); adult (30~59
1096 years); old (≥60 years). Bars, median with interquartile. P values, by Mann-Whitney U test.
1097 E. The T/S ratio represents the relative telomere length (T) to the single-copy gene (S, 36B4
1098 gene). Each dot represents an individual T/S ratio. Adult (21~59 years); old (≥60 years).
1099 F. RT-qPCR to detect total TERRA levels in blood cells using telomeric repeat primers. Each
1100 dot indicates each individual.
1101 G. Scatter plot showing negative correlation between TERRA and telomere length in blood
1102 cells. P values, by Pearson's correlation.
1103 H. RT-qPCR to detect TERRA levels using subtelomeric primers (hg38-2q). Each dot indicates
1104 each individual. (E, F and H) P values, by two-tailed Student's t-test. n=sample size. Error
1105 bar, SD.
1106
1107 **Figure 6. TERRA expression increases with age in brain tissues and fibroblasts.**

1108 A. Scatter plots showing the correlation of TERRA v.s. age in brain, ovary and heart tissues.
1109 Each dot indicates the total TERRA level from all chromosome ends in an individual. P
1110 values, by Pearson's correlation.
1111 B. TERRA levels in various tissues with different ages. Adult (21~59 years); old (≥60 years).
1112 Bars, median with interquartile. P values, by Mann-Whitney U test.
1113 C. TERRA levels in human fibroblasts derived from healthy individuals of different ages. P
1114 values, by Pearson's correlation.
1115 D. Boxplots showing the upregulation of TERRA levels in old human fibroblasts. Young (<30
1116 years); adult (30~59 years); old (≥60 years). Bars, median with interquartile. P values, by
1117 Mann-Whitney U test.
1118 E. Fibroblasts derived from HGPS patients display abnormal TERRA levels. Each dot
1119 represents the TERRA level of each individual. Bars, median with interquartile. P values,
1120 by Kolmogorov-Smirnov test.
1121
1122 **Figure 7. Single-cell analysis of TERRA levels undergoing neuronal differentiation and in**
1123 **Alzheimer's disease.**
1124 A. Violin plot showing normalized TERRA counts in various human tissues. The solid lines
1125 indicate the median and the dashed lines are interquartile range. P values, by Mann-
1126 Whitney U test.
1127 B. Single-cell RNA-seq showing normalized TERRA counts in various immune cells from
1128 PBMC. NK, nature killer cell; DC, dendritic cells. Solid lines, median. Dashed lines are
1129 interquartile range. P values, by Mann-Whitney U test.
1130 C. Single-cell RNA-seq analysis of normalized TERRA counts in human embryonic stem cells
1131 (hESCs) undergoing neuronal differentiation with various HOX patterning periods
1132 (24h,72h,48h, 120h, 168h, 216h). MSC, mesenchymal stem cells. Solid lines, median.
1133 Dashed lines are interquartile range. P values, by Mann-Whitney U test.
1134 D-F. Uniform manifold approximation and projection (UMAP) plots of single-cell RNA-seq
1135 datasets from hESCs undergoing neuronal differentiation using various HOX patterning
1136 periods. Colors according to TERRA, ATRX, or neuron marker MAP2 expression levels.
1137 G. TERRA levels in cells isolated from the prefrontal cortex of individuals with Alzheimer's
1138 disease (AD). Single-nucleus RNA-seq datasets were obtained from the ROSMAP project,
1139 and grouped into healthy (no AD pathology), early-AD pathology, and late-AD pathology.
1140 Solid lines, median. Dashed lines are interquartile range. P values, by Mann-Whitney U
1141 test.
1142 H. RT-qPCR analysis showing elevated TERRA levels in AD neurons. AD-PS1mut or Ctrl
1143 neurons were differentiated from iPSC. AD-PS1mut carrying P117L mutation in *PS1* gene.
1144 Ctrl: mutation-corrected isogenic control. P values, by two-tailed Student's t-test. n=3.
1145 TERRA levels normalized to GAPDH.
1146
1147
1148

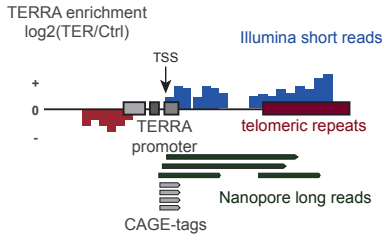
Figure 1

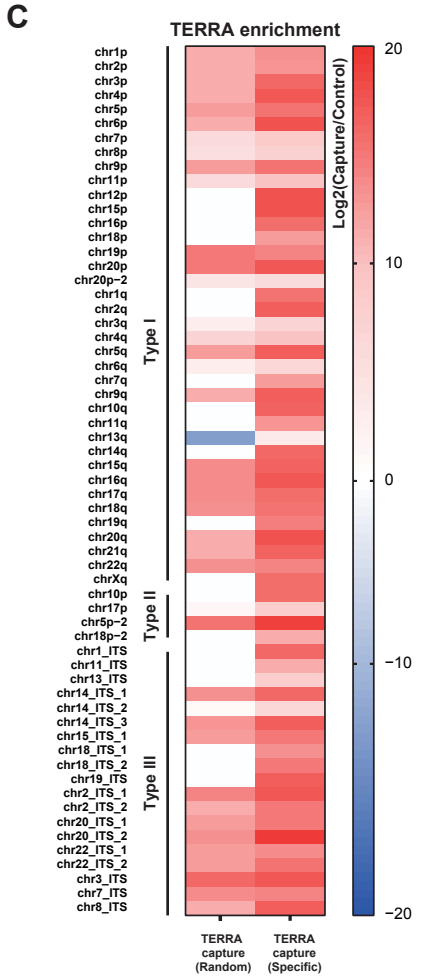


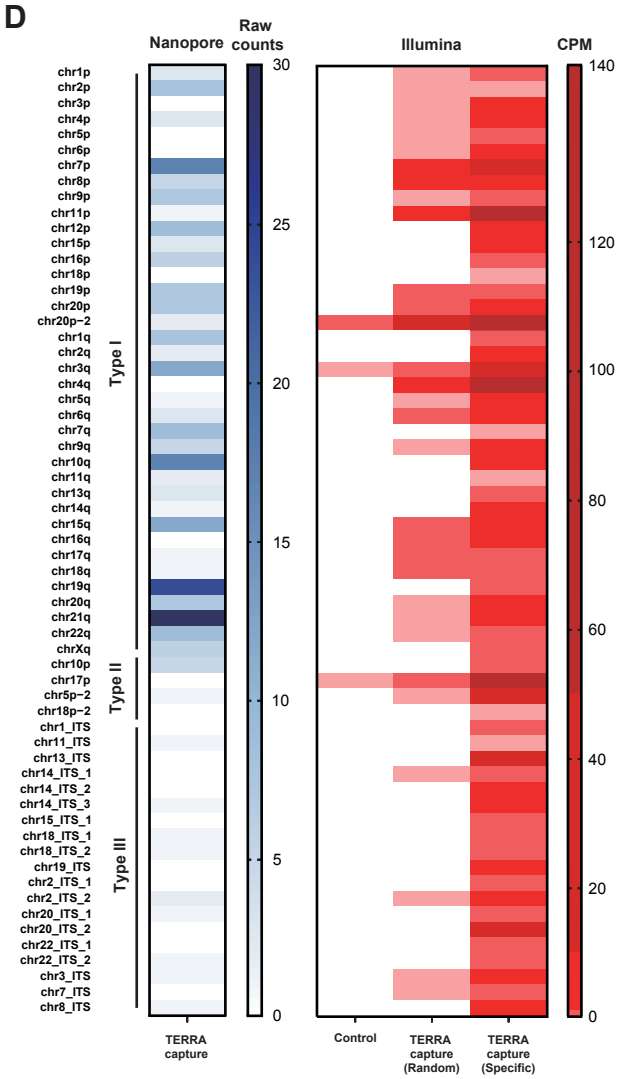
B

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, and CAGE-tags.
Type II No 61-29-37 bp repeats on the promoter	
Type III TERRA from interstitial telomeric sequences (19 ITSs)	







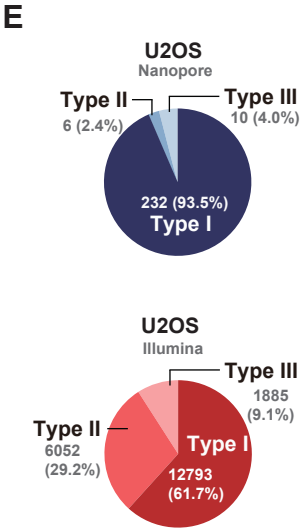
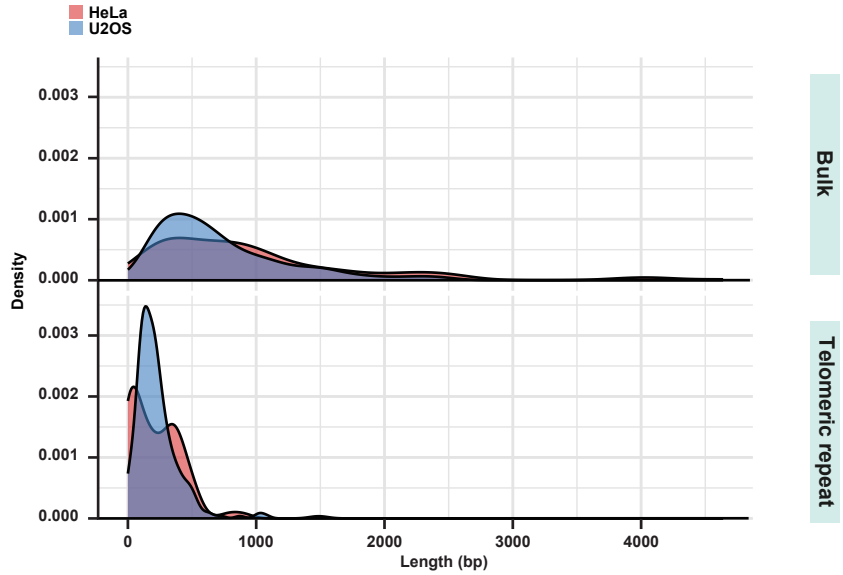


Figure 2

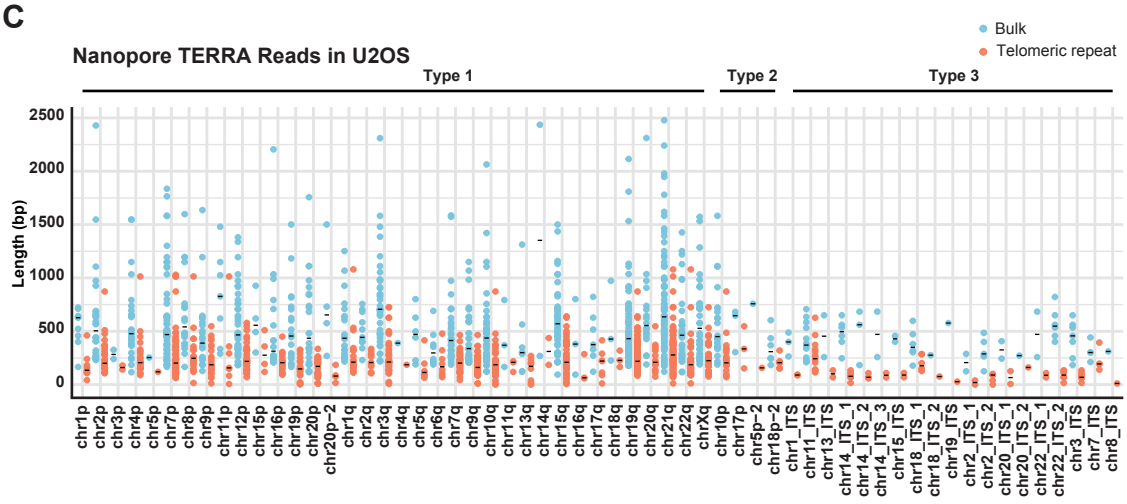
A



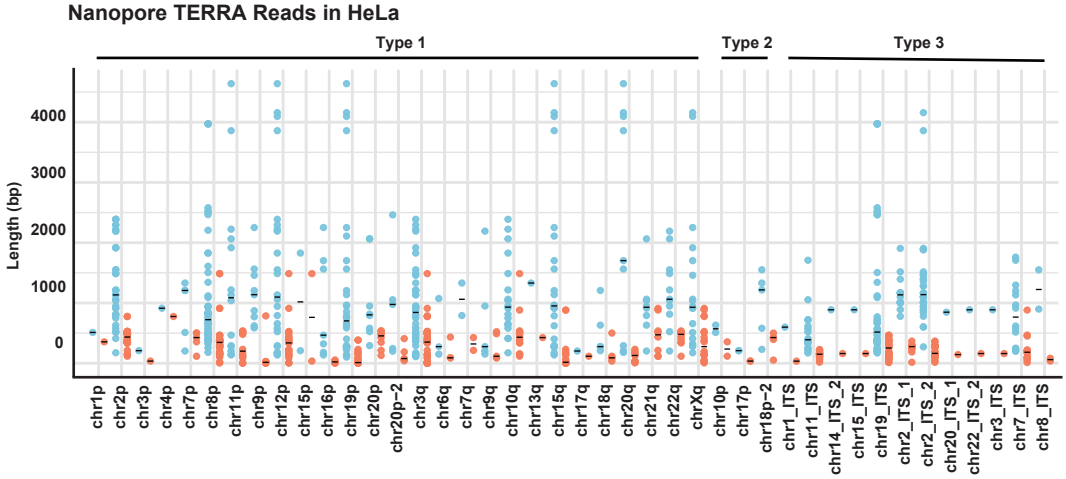
B

Statistic	U2OS	HeLa
No. of reads	281	153
Bulk Read Length		
Max (bp)	2478	4639
Median (bp)	578	779
Mean (bp)	721.3096	982.0523
Min (bp)	94	102
SD (bp)	501.9355	851.7684
Telomeric repeat Length		
Max (bp)	1080	1487
Median (bp)	189	278.5
Mean (bp)	222.0747	284.36
Min (bp)	0	0
SD (bp)	160.9017	219.2933

C



Statistic	U2OS		
Promoter	Type 1	Type 2	Type 3
Bulk Read Length			
Max (bp)	2478	1582	822
Median (bp)	480.5	542.09	402
Mean (bp)	583.26	438.5	416
Min (bp)	94	183	125
SD (bp)	386.69	309.24	165.93
Telomeric repeat Length			
Max (bp)	1080	873	627
Median (bp)	246.13	204	110
Mean (bp)	207.5	250.34	139.48
Min (bp)	0	70	0
SD (bp)	166.63	154.83	114.52



Statistic	HeLa		
Promoter	Type 1	Type 2	Type 3
Bulk Read Length			
Max (bp)	4639	1551	3972
Median (bp)	914	580	694
Mean (bp)	1193.87	718.89	968.55
Min (bp)	102	209	170
SD (bp)	1041.38	517.72	922.29
Telomeric repeat Length			
Max (bp)	1487	501	882
Median (bp)	282	355	185
Mean (bp)	298.59	260.56	214.38
Min (bp)	0	36	12
SD (bp)	280.38	195.23	137.42

Figure 3

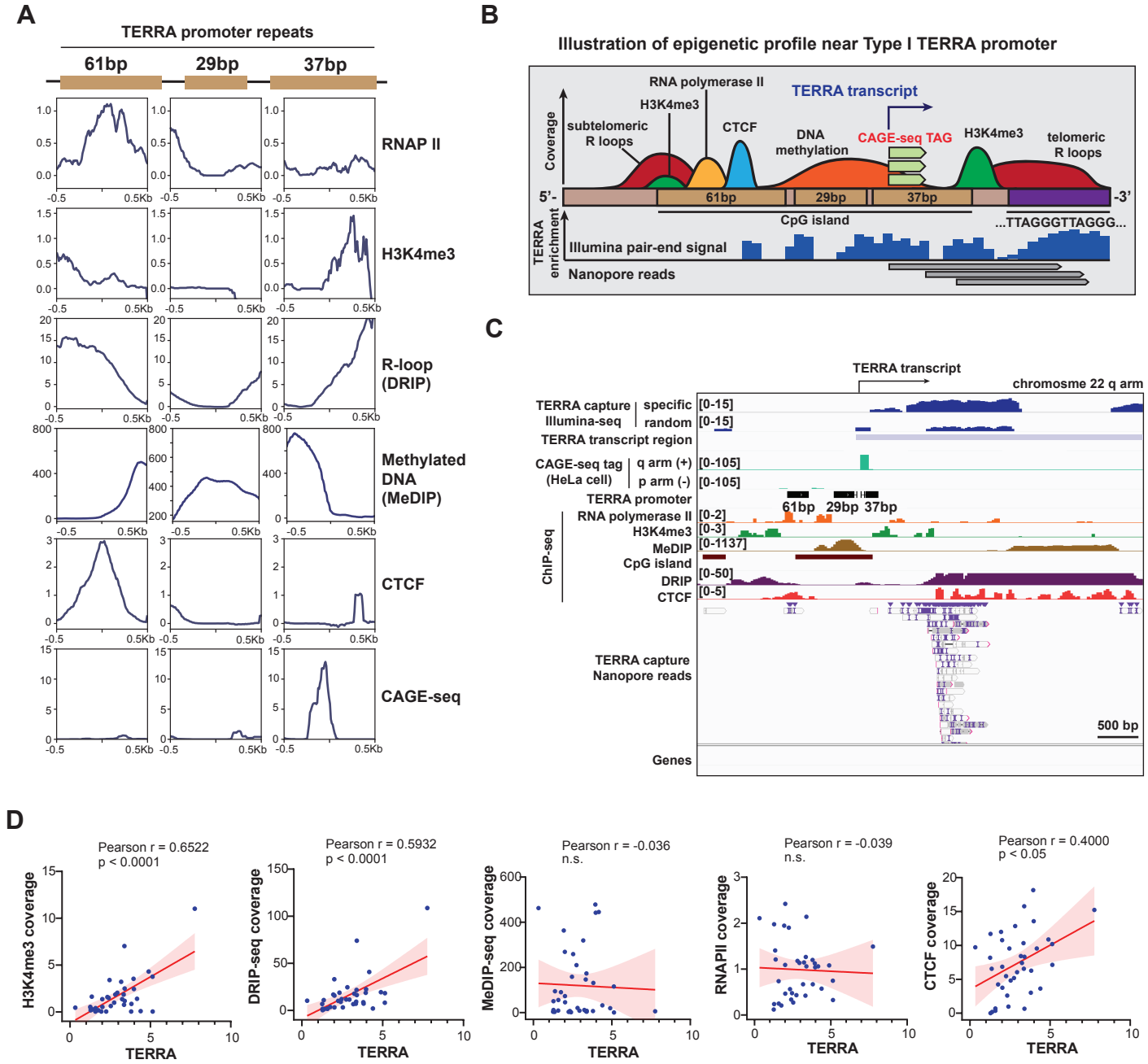

A
TERRA promoter repeats
61bp
29bp
37bp
RNAP II
H3K4me3
R-loop (DRIP)
Methylated DNA (MeDIP)
CTCF
CAGE-seq
-0.5
0.5Kb
B
Illustration of epigenetic profile near Type I TERRA promoter
RNA polymerase II
H3K4me3
TERRA transcript
subtelomeric R loops
CTCF
DNA methylation
CAGE-seq TAG
H3K4me3
telomeric R loops
Coverage
5'-
-3'
61bp
29bp
37bp
CpG island
...TTAGGGTTAGGG...
TERRA enrichment
Illumina pair-end signal
Nanopore reads
C
TERRA transcript
chromosme 22 q arm
TERRA capture Illumina-seq
specific
random
[0-15]
[0-15]
TERRA transcript region
CAGE-seq tag (HeLa cell)
q arm (+)
p arm (-)
[0-105]
[0-105]
TERRA promoter
61bp 29bp 37bp
ChIP-seq
RNA polymerase II
[0-2]
H3K4me3
[0-3]
MeDIP
[0-1137]
CpG island
DRIP
[0-50]
CTCF
[0-5]
TERRA capture Nanopore reads
500 bp
Genes
D
Pearson r = 0.6522
p < 0.0001
H3K4me3 coverage
TERRA
Pearson r = 0.5932
p < 0.0001
DRIP-seq coverage
TERRA
Pearson r = -0.036
n.s.
MeDIP-seq coverage
TERRA
Pearson r = -0.039
n.s.
RNAPII coverage
TERRA
Pearson r = 0.4000
p < 0.05
CTCF coverage
TERRA

Figure 4

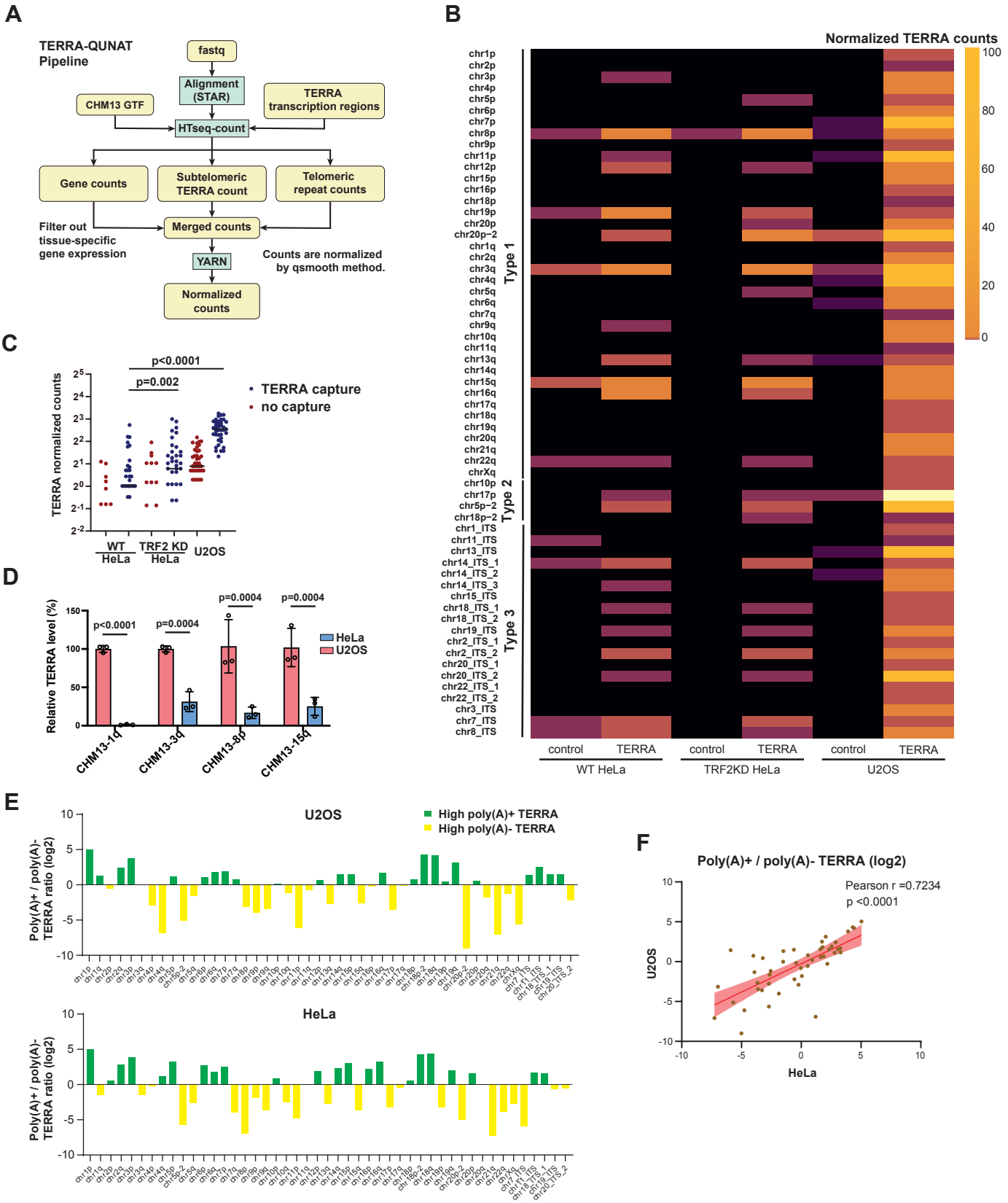

A
TERRA-QUNAT Pipeline
fastq
Alignment (STAR)
CHM13 GTF
TERRA transcription regions
HTseq-count
Gene counts
Subtelomeric TERRA count
Telomeric repeat counts
Filter out tissue-specific gene expression
Merged counts
YARN
Counts are normalized by qsmooth method.
Normalized counts
B
Normalized TERRA counts
Type 1
Type 2
Type 3
control
TERRA
WT HeLa
TRF2KD HeLa
U2OS
C
TERRA normalized counts
p<0.0001
p=0.002
TERRA capture
no capture
WT HeLa
TRF2 KD HeLa
U2OS
D
Relative TERRA level (%)
p<0.0001
p=0.0004
p=0.0004
p=0.0004
HeLa
U2OS
CHM13-1q
CHM13-3q
CHM13-8p
CHM13-15q
E
U2OS
High poly(A)+ TERRA
High poly(A)- TERRA
Poly(A)+ / poly(A)- TERRA ratio (log2)
HeLa
F
Poly(A)+ / poly(A)- TERRA (log2)
Pearson r =0.7234
p <0.0001
U2OS
HeLa

Figure 5

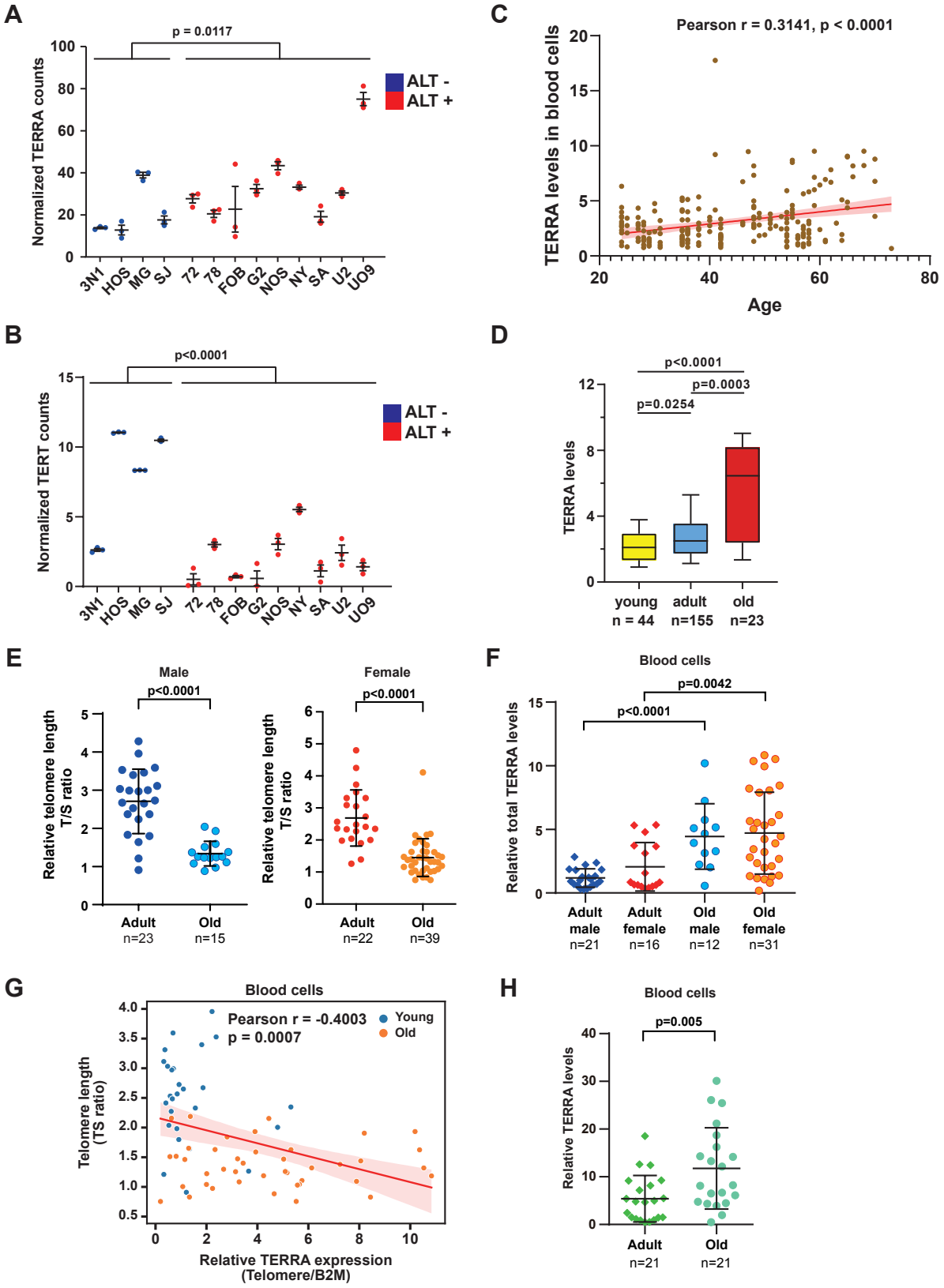

A
p = 0.0117
Normalized TERRA counts
ALT -
ALT +
3N1
HOS
MG
SJ
72
78
FOB
G2
NOS
NY
SA
U2
U09
B
p<0.0001
Normalized TERT counts
C
Pearson r = 0.3141, p < 0.0001
TERRA levels in blood cells
Age
D
p<0.0001
p=0.0003
p=0.0254
TERRA levels
young
n = 44
adult
n=155
old
n=23
E
Male
p<0.0001
Female
p<0.0001
Relative telomere length
T/S ratio
Adult
n=23
Old
n=15
Adult
n=22
Old
n=39
F
Blood cells
p=0.0042
p<0.0001
Relative total TERRA levels
Adult
male
n=21
Adult
female
n=16
Old
male
n=12
Old
female
n=31
G
Blood cells
Pearson r = -0.4003
p = 0.0007
Young
Old
Telomere length
(TS ratio)
Relative TERRA expression
(Telomere/B2M)
H
Blood cells
p=0.005
Relative TERRA levels
Adult
n=21
Old
n=21

Figure 6

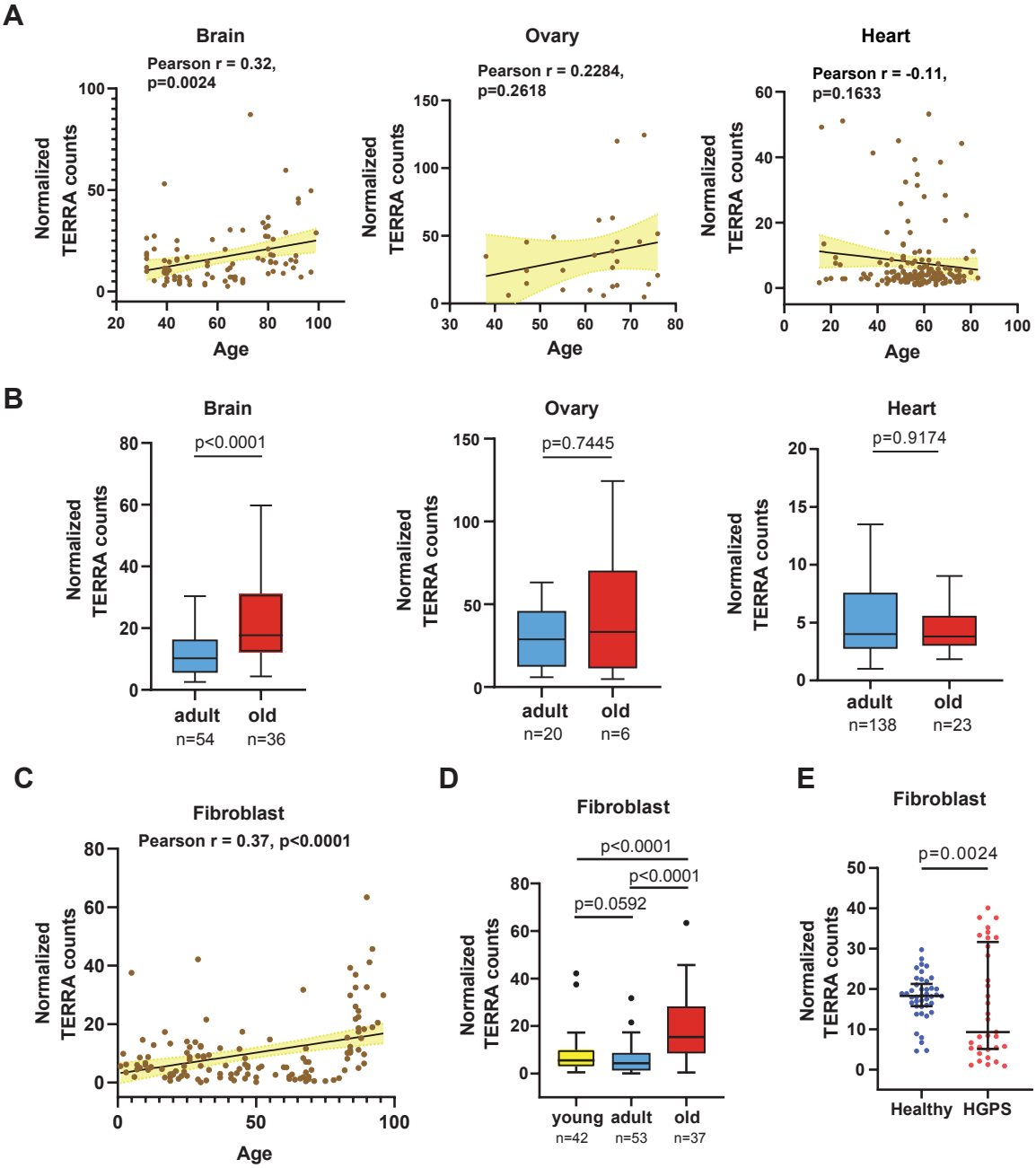
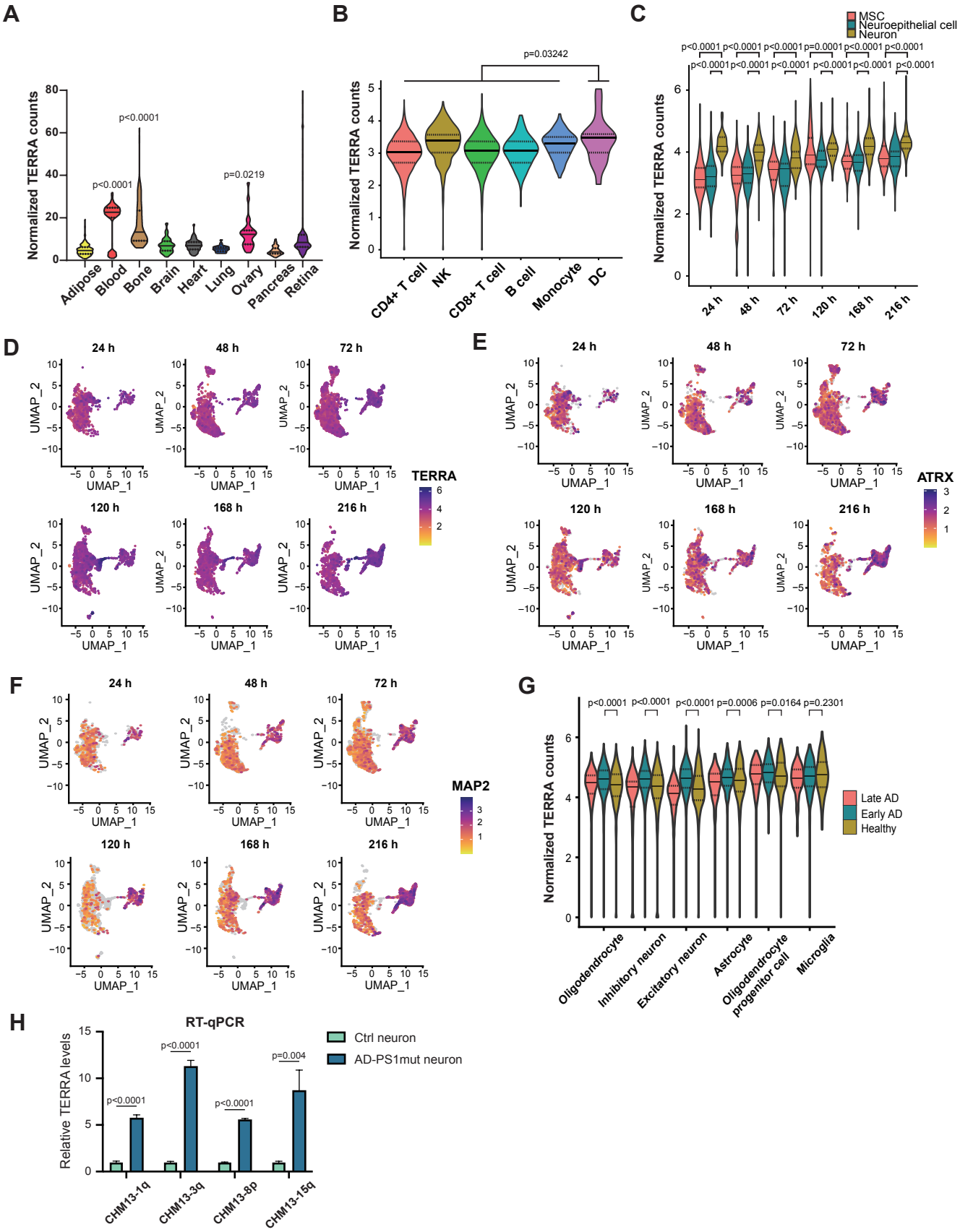

A
Brain
Pearson r = 0.32, p=0.0024
Normalized TERRA counts
Age
Ovary
Pearson r = 0.2284, p=0.2618
Heart
Pearson r = -0.11, p=0.1633
B
Brain
p<0.0001
adult n=54
old n=36
Ovary
p=0.7445
adult n=20
old n=6
Heart
p=0.9174
adult n=138
old n=23
C
Fibroblast
Pearson r = 0.37, p<0.0001
D
Fibroblast
p<0.0001
p<0.0001
p=0.0592
young n=42
adult n=53
old n=37
E
Fibroblast
p=0.0024
Healthy
HGPS

Figure 7


A
Normalized TERRA counts
p<0.0001
p<0.0001
p=0.0219
Adipose
Blood
Bone
Brain
Heart
Lung
Ovary
Pancreas
Retina
B
p=0.03242
Normalized TERRA counts
CD4+ T cell
NK
CD8+ T cell
B cell
Monocyte
DC
C
MSC
Neuroepithelial cell
Neuron
Normalized TERRA counts
24 h
48 h
72 h
120 h
168 h
216 h
D
24 h
48 h
72 h
120 h
168 h
216 h
UMAP_1
UMAP_2
TERRA
E
ATRX
F
MAP2
G
p<0.0001
p<0.0001
p<0.0001
p=0.0006
p=0.0164
p=0.2301
Normalized TERRA counts
Late AD
Early AD
Healthy
Oligodendrocyte
Inhibitory neuron
Excitatory neuron
Astrocyte
Oligodendrocyte progenitor cell
Microglia
H
RT-qPCR
Relative TERRA levels
Ctrl neuron
AD-PS1mut neuron
p<0.0001
p<0.0001
p<0.0001
p=0.004
CHM13-1q
CHM13-3q
CHM13-8p
CHM13-15q