國立臺灣大學生命科學院分子與細胞生物學研究所

碩士論文

Institute of Molecular and Cellular Biology

College of Life Science

National Taiwan University

Master's Thesis

Lurbinectedin 標靶 ALT 癌細胞中的端粒及探討 TERRA 與人類老化的關聯性

Lurbinectedin targets telomeres in ALT cancer and Telomeric Repeat-Containing RNA is associated with human aging

戴進華

Chin-Hua Tai

指導教授：朱雪萍 博士

Advisor: Hsueh-Ping Chu, Ph.D.

中華民國 114 年 7 月

July, 2025

doi:10.6342/NTU202504254

審定書

請至本校教務處/(右方)檔案與表單下載/研究生教務組/審定書下載。

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

（論文中文題目）(Chinese title of Master's thesis)

Lurbinectedin 標靶 ALT 癌細胞中的端粒及探討 TERRA 與人類老化的關聯性

（論文英文題目）(English title of Master's thesis)

Lurbinectedin targets telomeres in ALT cancer and Telomeric Repeat-Containing RNA is associated with human aging.

本論文係 戴進華 (姓名) R12B43029 (學號) 在國立臺灣大學 分子與細胞生物學系/所/學位學程 完成之碩士學位論文，於民國 114 年 7 月 22 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Department / Institute of Molecular and Cellular Biology on 22 (date) 7 (month) 2025 (year) have examined a Master's thesis entitled above presented by DAI, CHIN-HUA (name) R12B43029 (student ID) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

朱雪萍 (指導教授 Advisor)　陳律佑　林敬哲

系主任/所長 Director: 董桂書

doi:10.6342/NTU202504254

誌謝

這段求學旅程中，有辛苦也有收穫。謝謝朱雪萍老師與實驗室夥伴的指導與協助，讓我一步步累積經驗、完成這篇論文。也謝謝家人與朋友在背後默默的支持與鼓勵，讓我能安心專注在研究上。這段經歷雖然告一段落，但這些學習與回憶，將陪著我繼續前行。

doi:10.6342/NTU202504254

摘要

端粒維持機制在癌症進展與細胞老化中扮演關鍵角色。多數癌細胞依賴端粒酶以維持端粒長度，然而約有 10–15%的癌症，特別是肉瘤，則使用另一種機制，稱為非典型端粒延長模式（alternative lengthening of telomeres, ALT）。ALT 癌細胞常對傳統化學治療藥物產生抗藥性，因此尋找新的治療策略至關重要。本研究利用生物素標記的 lurbinectedin 結合 chem-seq 技術，探討其在 ALT 陽性癌細胞中的基因組結合分佈。我們的結果顯示，lurbinectedin 偏好結合於端粒區域及 G-四股螺旋結構（G-quadruplex, G4），此現象經由 motif 分析及 GC 含量富集進一步確認。此外，lurbinectedin 亦顯示在基因啟動子（transcription start sites, TSS）有顯著結合，暗示其可能具備更廣泛的調控功能。本研究揭示 lurbinectedin 於 ALT 癌細胞中標靶端粒與 G4 結構的潛力，提供未來治療策略的新方向。另一部分的研究著重於一種長鏈分編碼 RNA，稱之為（telomeric repeat-containing RNA, TERRA），此 RNA 包含端粒重複序列，源自亞端粒區域，參與端粒功能及基因組穩定性的調控。然而，TERRA 表達與人類老化之間的關係仍未被充分探討。本研究應用 RNA 納米孔定序（Nanopore RNA direct sequencing）及次世代定序（Next generation sequencing ）結合生物資訊分析（TERRA-QUANT），量化 TERRA 表達量，並進一步使用 poly(A)+/non-poly(A) TERRA 捕獲定序以解析染色體端特定的多腺苷酸化偏好。分析結果顯示，TERRA 在血液、腦部及纖維母細胞中隨年齡顯著增加。此外，早老症（Hutchinson-Gilford progeria syndrome, HGPS）患者的纖維母細胞則表現出異常的 TERRA 表達型態。此結果指出，TERRA 可能為老化及相關疾病的重要分子標誌。

關鍵字:替代性端粒延長、癌症、端粒重複序列核醣核酸、老化

doi:10.6342/NTU202504254

Abstract

Telomere maintenance mechanisms play crucial roles in cancer progression and cellular aging. While most cancers rely on telomerase to sustain telomere length, approximately 10-15% of cancers, particularly sarcomas, utilize the alternative lengthening of telomeres (ALT) pathway. ALT cancers often exhibit resistance to conventional chemotherapeutic agents, highlighting the need for alternative treatment strategies. In this study, we employed chem-seq with biotinylated lurbinectedin to investigate its genome-wide binding profile in ALT-positive cancer cells. Our results reveal that lurbinectedin preferentially binds to telomeric regions and G-quadruplex (G4) structures, as confirmed by motif analysis and GC content enrichment. Furthermore, lurbinectedin exhibits notable binding at transcription start sites (TSS), suggesting broader regulatory roles. These findings reveal the mechanism by which lurbinectedin efficiently eliminates ALT cancer cells through targeting telomeres.

Second part of my thesis focuses on telomeric repeat-containing RNA (TERRA), a long non-coding RNA transcribed from subtelomeric regions. TERRA serves as a key regulator of telomere function and genome stability, but its relationship to human aging remains poorly understood. To address this, we combine Nanopore RNA direct sequencing and next-generation sequencing to annotate TERRA transcription regions. We established a customized bioinformatics pipeline (TERRA-QUANT) to quantify

doi:10.6342/NTU202504254

TERRA levels. Coupling this with poly(A)+ and non-poly(A) TERRA capture sequencing allowed us to uncover chromosome-specific polyadenylation preferences conserved across cell types. Our analyses also demonstrated that TERRA expression significantly increases with age in blood, brain, and fibroblasts. Notably, abnormal TERRA expression patterns were detected in Hutchinson-Gilford progeria syndrome (HGPS) fibroblasts. These results suggest that TERRA may serve as a molecular marker for aging and age-associated diseases.

Keywords: alternative lengthening of telomeres, cancer, telomeric repeat-containing RNA, aging.

doi:10.6342/NTU202504254

Content

doi:10.6342/NTU202504254

doi:10.6342/NTU202504254

doi:10.6342/NTU202504254

doi:10.6342/NTU202504254

Content of figures

doi:10.6342/NTU202504254

doi:10.6342/NTU202504254

Content of table

doi:10.6342/NTU202504254

Chapter 1 Lurbinectedin target telomeres in ALT cancer

INTRODUCTION

1 Telomeres and the Need for Maintenance Mechanisms

Telomeres are nucleoprotein structures located at the ends of linear chromosomes, composed of tandem TTAGGG repeats and associated with a specialized protein complex known as shelterin [1, 2]. This complex, which includes TRF1, TRF2, TIN2, POT1, TPP1 and RAP1, plays essential roles in protecting chromosome ends from being misrecognized as DNA double-strand breaks (DSBs), thereby suppressing activation of the DNA damage response (DDR) and preventing end-to-end chromosomal fusions [2]. Through the formation of T-loops and regulation of DNA repair pathway accessibility, shelterin maintains telomere integrity and contributes to genome stability [3]. During each round of DNA replication, telomeres undergo gradual shortening due to the end-replication problem, in which lagging-strand synthesis cannot fully replicate the 3′ end of linear DNA [4]. Additional shortening may result from oxidative stress and nucleolytic processing [5]. As a result, telomeres limit the replicative capacity of somatic cells. Critically short telomeres trigger a persistent DDR and induce cellular senescence or apoptosis [6]. To achieve cellular immortality, one of the hallmarks of cancer, tumor cells

doi:10.6342/NTU202504254

must circumvent telomere attrition by activating a telomere maintenance mechanism (TMM) [7]. While the majority of human cancers (~85–90 %) upregulate telomerase—a reverse transcriptase that elongates telomeres using an RNA template [7, 8]—a significant subset (~10–15 %) utilize an alternative, telomerase-independent strategy to maintain telomere length [9].

2 Alternative Lengthening of Telomeres in Cancer

The alternative lengthening of telomeres (ALT) is a telomerase-independent pathway that sustains chromosome ends in roughly 10–15 % of human tumours—most frequently in mesenchymal cancers such as osteosarcomas, liposarcomas and high-grade gliomas—and operates through homologous-recombination–driven DNA synthesis in which a shortened telomere invades a homologous template (another telomere, a sister chromatid or an extrachromosomal C-circle) and copies sequence via break-induced replication or mitotic DNA synthesis [9, 10]. This recombination activity generates hallmark features: extreme telomere-length heterogeneity; abundant extrachromosomal telomeric DNA—including C-circles that serve as quantitative ALT markers [11]; ALT-associated PML bodies (APBs) containing telomeric DNA, PML protein and repair factors such as RAD51, BLM and RPA; frequent telomere sister-chromatid exchanges; and high transcription of telomeric repeat-containing RNA (TERRA) that forms RNA-

doi:10.6342/NTU202504254

DNA hybrids (R-loops) [12]. Genetically, ALT is strongly linked to loss-of-function mutations or epigenetic silencing of the chromatin remodelers ATRX or DAXX, whose absence destabilizes telomeric heterochromatin, derepresses TERRA and promotes unscheduled recombination [13, 14]. Functionally, ALT telomeres are prone to secondary structures such as G-quadruplexes and persistent R-loops, leading to replication stress and chronic activation of ATM/ATR signaling; although these structures facilitate the recombination necessary for telomere elongation, they simultaneously expose ALT cells to DNA-damage-induced vulnerabilities that distinguish them from telomerase-positive tumors [15, 16].

3 Therapeutic Challenges in ALT-Positive Tumors

Although tumors that rely on the alternative lengthening of telomeres (ALT) account for only about 10–15 % of all cancers, they are over-represented in mesenchymal malignancies—particularly osteosarcoma, liposarcoma and high-grade glioma—and are frequently associated with pronounced genomic instability and poor prognosis [11, 14]. Because these tumors lack telomerase activity, standard telomerase-targeted agents are ineffective, and they often display intrinsic or rapidly acquired resistance to first-line chemotherapies such as doxorubicin, cisplatin and etoposide. Their heavy reliance on homologous-recombination repair suggests that inhibitors of DNA-damage-response kinases (e.g., ATR or CHK1) should be efficacious, yet clinical and pre-clinical studies

doi:10.6342/NTU202504254

have yielded inconsistent responses, pointing to a highly adaptable repair network [17, 18]. The relative rarity and biological heterogeneity of ALT-positive cases further complicate trial design and impede the development of robust companion biomarkers, leaving this aggressive subgroup of cancers therapeutically underserved [15].

4 G-Quadruplexes, R-Loops, and Telomeric Vulnerabilities

G-quadruplexes (G4s) are four-stranded DNA or RNA structures formed by stacked guanine tetrads stabilized by Hoogsteen hydrogen bonding and monovalent cations. They commonly form in guanine-rich regions, including telomeric DNA, gene promoters and replication origins [16, 19]. At telomeres, G4 structures can impede replication fork progression, alter telomere elongation and promote DNA-damage signaling [16]. In ALT cells, where telomeres are actively transcribed into TERRA and recombinogenic, G4s are particularly abundant and may participate in both telomere elongation and genome instability [19]. R-loops are another class of noncanonical nucleic-acid structures consisting of RNA-DNA hybrids and a displaced single-stranded DNA strand. In the telomere context, R-loops are typically generated co-transcriptionally by TERRA and are prevalent in ALT cells [12]. While R-loops may support telomere recombination by exposing single-stranded DNA, excessive R-loop formation leads to replication-transcription conflicts, replication-fork stalling and DNA damage [20]. The concurrent

doi:10.6342/NTU202504254

presence of G4s and R-loops at ALT telomeres creates a structurally complex and fragile chromatin environment—one that may be exploited by DNA-binding agents capable of recognizing or stabilizing such features [21].

5 Lurbinectedin for cancer treatment application in ALT cancer

Lurbinectedin is a synthetic tetrahydroisoquinoline alkaloid structurally related to trabectedin, originally derived from the marine tunicate *Ecteinascidia turbinata*. The drug binds to the DNA minor groove at specific GC-rich sequences, interferes with RNA-polymerase-II-mediated transcription and induces DNA damage, particularly in transcriptionally active genomic regions [22]. Mechanistically, lurbinectedin causes stalling of RNA polymerase II, recruitment of nucleotide-excision-repair factors and formation of single- and double-strand breaks, ultimately leading to apoptosis [22]. Clinically, lurbinectedin has demonstrated efficacy in small-cell lung cancer and is being investigated in other solid tumors [23, 24]. While its genome-wide binding pattern had not been systematically studied, its sequence preference and transcription-blocking activity raise the possibility that lurbinectedin may selectively localize to G4-enriched or transcriptionally vulnerable chromatin regions. Given that ALT telomeres are GC-rich, G4-prone and sites of active transcription, we hypothesized that lurbinectedin may exhibit

doi:10.6342/NTU202504254

preferential binding to ALT-specific telomeric regions and interfere with their function. This model suggests that lurbinectedin might exploit the unique vulnerabilities of ALT cells by exacerbating DNA damage at already fragile telomeric loci, thereby inducing selective cytotoxicity [25].

6 Objectives of This Study

In this study, we sought to characterize the genome-wide DNA binding profile of lurbinectedin in ALT-positive cancer cells. To this end, we performed chem-seq in U2OS cells using biotinylated lurbinectedin (Bio-Lurbinectedin) to identify high-confidence binding sites across the genome. We then examined the sequence and chromatin features associated with Bio-Lurbinectedin peaks, including G-quadruplex content, GC composition, proximity to transcription start sites (TSSs), and subtelomeric enrichment. By integrating Bio-Lurbinectedin chem-seq data with G4 ChIP-seq and epigenomic annotations, we aimed to determine whether lurbinectedin preferentially targets ALT telomeres and G4-rich regulatory regions. These results provide mechanistic insight into the interaction between lurbinectedin and chromatin architecture in ALT cells and support its potential use as a therapeutic agent targeting telomerase-independent cancers

doi:10.6342/NTU202504254

MATERIALS AND METHODS

1 Cell culture

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

2 Biotin-lurbinectedin chromatin precipitation sequencing (ChemIP-seq)

Cells were fixed in PBS containing 1% formaldehyde for 10 min at room temperature (RT). The reaction was quenched by adding 1/10 volume of 1.25 M glycine and incubated at RT for 5 min. Cells were washed with ice-cold PBS and lysed in 0.5 ml Buffer 1 (50 mM HEPES, 150 mM NaCl, 1 mM EDTA (pH 8.0), 0.5% NP-40, 0.25% Triton X-100) supplemented with 2× complete EDTA-free Protease Inhibitor Cocktails (PIC, Roche, Cat#04693132001), and incubated for 10 min with rotation at 4°C. After centrifugation at 21,100 *g* for 10 min at 4°C, cells were resuspended in Buffer 2 (200 mM NaCl, 5 mM EDTA, 2.5 mM EGTA, 10 mM Tris pH 8.0) supplemented with 2× PIC, and rotated at 4°C for 10 min. After centrifugation, cells were resuspended in Buffer 3 (5 mM EDTA, 2.5 mM EGTA, 10 mM Tris pH 8.0, and freshly added 2× PIC), mixing on a

doi:10.6342/NTU202504254

rotator for 10 min at 4°C. 30 µl of 10% *N*-lauroylsarcosine was added to lysates, followed by incubation on a rotator for 10 min at 4°C. 130 µl of samples were loaded into Covaris microtube (6 × 16 mm) for sonication, using the program with 12-height-level for water-bath, at 4–7°C, 5% duty cycle, intensity 4 and 200 bursts per cycle for 10 min to generate fragmented DNA size (200–500 bp) for biotin-lurbinectedin ChemCP-seq. Biotin-lurbinectedin bound chromatin per sample was captured using Dynabeads™ M-280 Streptavidin (20 µl of beads per capture, washed with Buffer 1 twice and resuspended in Buffer 3) on a rotator at 4°C for 2 h. Beads were washed with low salt buffer (20 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1% Triton-X 100, 0.1% SDS, and 150 mM NaCl) 2 times and high salt buffer (20 mM Tris-HCl (pH 7.5), 2 mM EDTA, 1% Triton-X 100, 0.1% SDS, and 500 mM NaCl) 2 times, 5 min each at 4°C with rotation. Beads were resuspended in 50 µl elution buffer (100 mM NaHCO3, 1% SDS, and 50 mM Tris pH 8.0) at 85°C for 15-30 min twice for elution. 1 µl of 20 mg/ml RNase A was added to each sample and incubated at 37°C for 20 min. Samples were further incubated with 5 µl of 20 mg/ml Proteinase K at 55°C overnight. After reverse crosslinking, DNA was purified using a Plus DNA Clean / Extraction Kit. Eluted DNA was subjected to library construction using NEBNext Ultra II DNA Library Prep kit (NEB, Cat#E7645S) for Illumina. Paired-end 150bp reads were obtained by NovaSeq systems. Part of the eluted DNA was analysed by qRT-PCR to assess the amount of the telomeric DNA.

doi:10.6342/NTU202504254

3 Library Preparation for Illumina Sequencing

Chromatin immunoprecipitated (ChIP) DNA samples were subjected to library preparation using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, #E7104S), following the manufacturer's protocol with minor modifications. For each sample, 25 µl of DNA was combined with 3.5 µl of End Prep Reaction Buffer and 1.5 µl of End Prep Enzyme Mix, bringing the total volume to 30 µl. The reaction was incubated in a thermocycler as follows: 20°C for 30 minutes, 65°C for 30 minutes, then held at 4°C. Subsequently, 30 µl of end-repaired DNA was mixed with 15 µl of Ligation Master Mix, 0.5 µl of Ligation Enhancer, and 1.25 µl of diluted adapter (5× dilution). The ligation reaction was performed at 20°C for 15 minutes and held at 4°C. Adapter-ligated DNA was then treated with 3 µl of USER enzyme at 37°C for 30 minutes. Size selection was performed using magnetic beads for fragment size enrichment. The ligated DNA was mixed with 0.3× volume (30 µl) of sample purification beads and incubated at room temperature for 10 minutes. The supernatant was transferred to a fresh tube and mixed with 0.5× volume (50 µl) of beads for 10 minutes. Beads were washed twice with 80% ethanol and air-dried. The DNA was eluted in 20 µl of 0.1× TE buffer. To determine the optimal cycle number for PCR amplification, 2 µl of size-selected DNA was mixed with NEBNext Ultra II Q5 Master Mix and subjected to qPCR. The amplification cycles were based on Ct values obtained from the qPCR assay. Library amplification was carried out

doi:10.6342/NTU202504254

in a 50 µl reaction volume containing 25 µl Q5 Master Mix, 2 µl Universal Primer, 2 µl Index Primer, and size-selected DNA. The PCR program was: 98°C for 30 seconds, followed by variable cycles (based on qPCR evaluation) of 98°C for 10 seconds and 65°C for 75 seconds, with a final extension at 65°C for 5 minutes. Amplified libraries underwent a second round of size selection. Initially, 0.9× volume of beads was added to the amplified DNA, incubated at room temperature, and washed twice with 80% ethanol. DNA was eluted in 52 µl of 0.1× TE buffer. A second 0.9× bead selection was performed to ensure library purity, followed by elution in 15 µl of 0.1× TE buffer. Library fragment size distribution was assessed using a Bioanalyzer (Agilent Technologies) with High Sensitivity DNA kits. Library concentrations were quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs, #E7630S), following serial dilutions and qPCR amplification with standard curves.

4 Sequencing data analysis

Chem-seq raw FASTQ reads were trimmed by Trim-galore (v0.6.3) with parameters: –illumina –fastq -q 30. Trimmed reads were aligned to the CHM13 reference genome using bowtie2 aligner (v2.4.2) with the default setting. BAM files were sorted and deduplicated with SAMTools (v1.18). Enriched peaks were determined using MACS2 (v2.2.9.1)with cutoff values: – p value 0.01. Bedtools intersect (v2.3.0) was used to

doi:10.6342/NTU202504254

determine overlapping regions of two interested data. Bedtools fastaFromBed (v2.3.0) converted a region to fasta information for motif sequence analysis, which was conducted using MEME Suite (v5.3.2). Bigwig files were generated using DeepTools (v3.3.1) bamCoverage. Normalized bigwig files were normalized to inputs by DeepTools (v3.3.1) bamCompare with parameters: –operation subtract, –centerReads, –normalizeUsing CPM, –scaleFactorsMethod None. Metaplots were generated using DeepTools (v3.3.1) computeMatrix and plotProfile. Heatmap were generated using DeepTools (v3.3.1) computeMatrix and plotHeatmap. For GC content, the genome was divided into six GC content groups. Lurbinectedin binding sites from Chem-seq data were intersected with these groups using Bedtools intersect (v2.3.0). Binding site distributions were normalized by genome length and compared to a random background. Percentage of counts in regions were generated by BEDTOOLS coverageBed(v2.31.1) -counts.

doi:10.6342/NTU202504254

RESULTS

1 Bio-lurbinectedin recognizes G-rich DNA

To investigate the genomic binding sites of lurbinectedin, we performed Bio-lurbinectedin Chem-seq in U2OS cells. In this approach, cells were treated with biotin-labelled lurbinectedin (Bio-Lurbinectedin, PM120306, Chapter 1 Figure 1) for 10 minutes, and drug-bound chromatin fragments were captured using streptavidin-coated beads, followed by next-generation sequencing. This method enables unbiased genome-wide identification of lurbinectedin binding sites. Quantitative PCR (qPCR) analysis of the captured DNA revealed that lurbinectedin binds telomeric TTAGGG repeats within ten minutes of treatment (Chapter 1 Figure 2A). Motif analysis using MEME Suite (Motif-based Sequence Analysis Tools) revealed that lurbinectedin has a strong binding preference for G-rich sequences, while base-composition profiling showed that most of the peaks lie in regions whose GC content exceeds 50 % (Chapter 1 Figure 2B, C). Together, these observations confirm a strong preference of lurbinectedin for guanine- and GC-rich sequences.

2 Lurbinectedin preferentially binds ALT telomeres

Given our earlier observation that lurbinectedin targets G-rich regions, we next asked whether this preference extends to ALT telomeres. Metaprofiles aligned at chromosome

doi:10.6342/NTU202504254

endsMetaprofiles aligned at chromosome ends displayed a sharp rise in Bio-lurbinectedin signal within the last 25 kb of every arm in U2OS cells (Chapter 1 Figure 3A, B). To compare the telomeric targeting of lurbinectedin between ALT and non-ALT cells, we performed Bio-lurbinectedin Chem-seq in HeLa cells and analyzed previously published Chem-seq data from A549 cells, both representing non-ALT backgrounds. The results showed that lurbinectedin preferentially binds to telomeres in U2OS cells, whereas such enrichment was not observed in HeLa or A549 cells, indicating distinct telomeric binding patterns between ALT and non-ALT cell lines Chapter 1 (Figure. 3C). Genome-browser snapshots further illustrated dense peak clusters at multiple p- and q-arm telomeres in U2OS (Chapter 1 Figure 4).

3 Lurbinectedin binging sites are associate with G-quadruplex

To further explore whether the preferential binding of lurbinectedin to telomeres is linked to G-quadruplex (G4) DNA structures and specific genomic features, we next examined the overlap between Bio-lurbinectedin Chem-seq peaks and G4 ChIP-seq data, as well as their distribution relative to transcription start sites. Intersecting Bio-lurbinectedin peaks with G4 ChIP-seq data showed that roughly one-third of experimentally mapped DNA G-quadruplex (G4) loci overlap Chem-seq peaks—far above the frequency obtained with randomized peaks (Chapter 1 Figure 5).

In addition to telomeres, metagene plots on TSSs revealed a second binding hotspot.

doi:10.6342/NTU202504254

Metagene profiles revealed an enrichment of lurbinectedin binding near transcription start sites (TSSs) (Chapter 1 Figure 6A), particularly those marked with high levels of active histone mark H3K4me3 (Chapter 1 Figure 6B).

doi:10.6342/NTU202504254

DISUCSSION

In this study, we demonstrated that lurbinectedin preferentially binds to telomeric regions and G-quadruplex (G4) structures specifically in ALT-positive cancer cells. Using Bio-lurbinectedin Chem-seq in U2OS cells, we identified a significant enrichment of lurbinectedin at chromosome ends, while this phenomenon was not observed in non-ALT cells such as HeLa. This ALT-specific accumulation was supported by metagene analysis and genome browser inspections. Motif analysis and GC content profiling further revealed that lurbinectedin favors guanine-rich, GC-rich sequences, consistent with its known DNA-binding properties.

Notably, we observed that Bio-lurbinectedin peaks overlapped substantially with previously defined G4 ChIP-seq peaks, indicating that lurbinectedin targets DNA G-quadruplex structures in addition to telomeres. Furthermore, we found lurbinectedin binding enriched at transcription start sites (TSSs), especially those marked by high H3K4me3 signals, suggesting a preference for transcriptionally active regions. However, telomeric targeting remained the most prominent feature unique to ALT cells.

Together, our results support a model where lurbinectedin exploits ALT-specific telomere structures—characterized by high G4 and R-loop levels—as preferential binding targets. This property may underlie lurbinectedin's selective cytotoxicity against ALT cancer cells observed in previous studies. Our Chem-seq data provide genome-wide

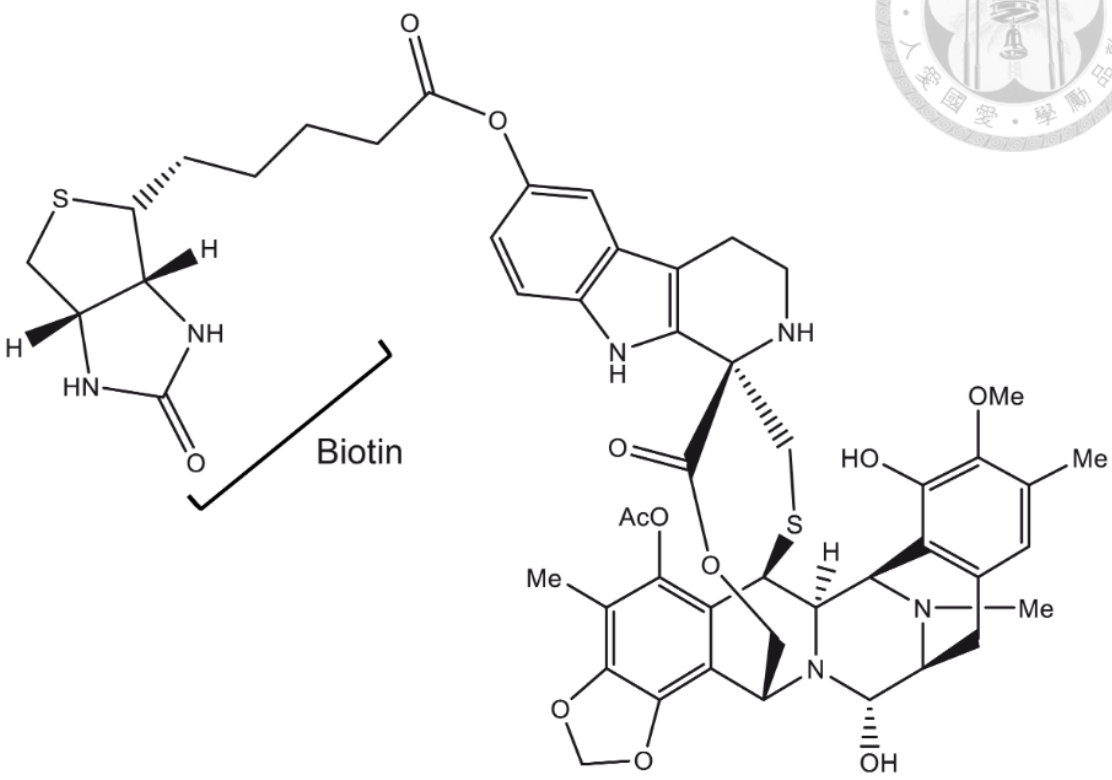
doi:10.6342/NTU202504254

confirmation of lurbinectedin's binding landscape and offer mechanistic insights.

By integrating these results with other studies from our laboratory, which demonstrated that lurbinectedin induces R-loops and recruits XPF endonuclease leading to DNA double-strand breaks at telomeres in ALT cells, we propose a mechanistic model: lurbinectedin binds to G4-rich telomeric DNA, stabilizing G4 structures and promoting R-loop formation. These R-loops then recruit XPF, which introduces DNA breaks, resulting in telomere dysfunction and cell death. While our study specifically focused on mapping lurbinectedin binding sites, these combined findings suggest that the drug's ALT-selective cytotoxicity arises from both its binding specificity and downstream DNA damage processes.

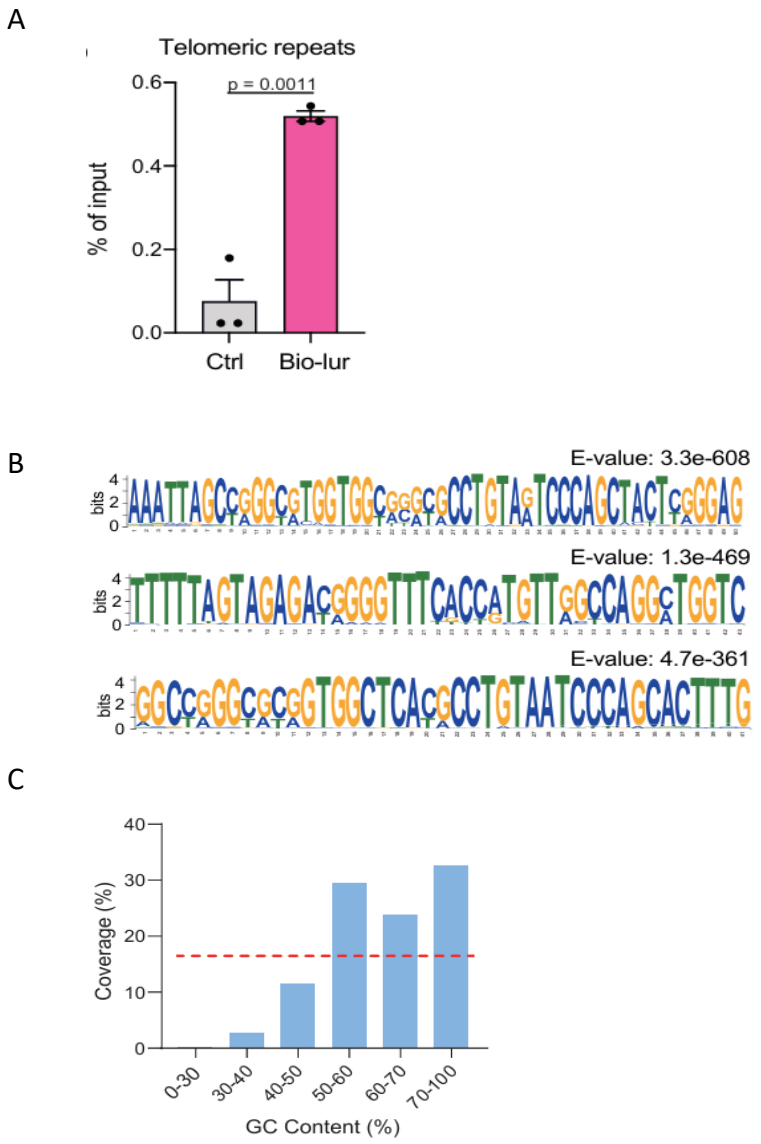
doi:10.6342/NTU202504254

FIGURES AND TABLES



Chapter 1 Figure 1. Structure of Bio-lurbinectedin

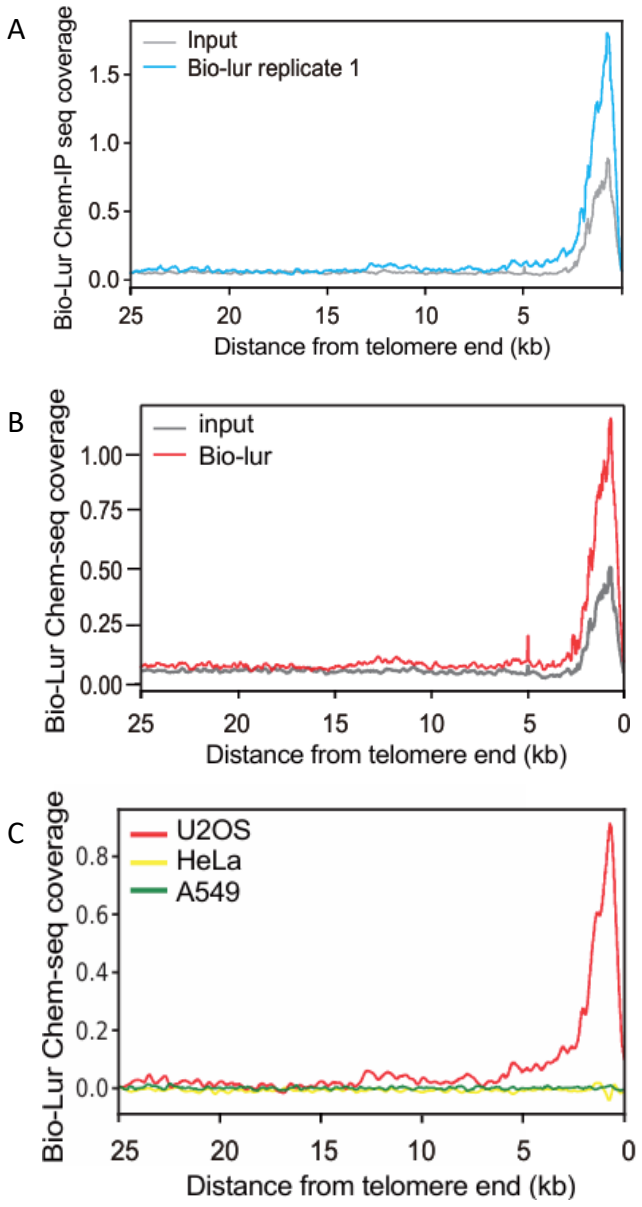
doi:10.6342/NTU202504254



Chapter 1 Figure 2. Bio-lurbinectedin recognizes G-rich DNA

(A) Telomeric DNA enrichment in Bio-lurbinectedin capture, quantified by qPCR. Data represent mean ± SEM from three independent experiments. P-values were calculated using unpaired t-test. **(B)** Motif analysis of Bio-lurbinectedin Chem-seq peaks using MEME, revealing G-rich sequence motifs. **(C)** GC content distribution of Bio-lurbinectedin Chem-seq peaks. The red line indicates the percentage of randomized peaks.

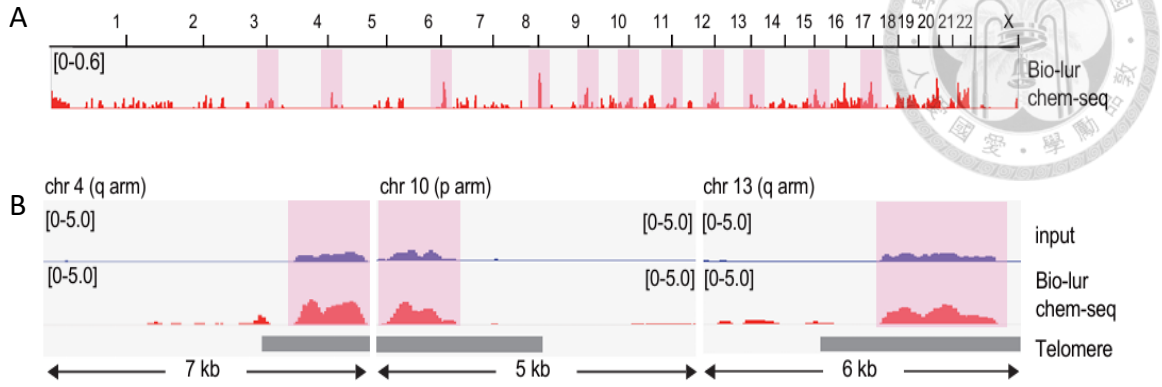
doi:10.6342/NTU202504254



Chapter 1 Figure 3. Bio-lurbinectedin is enriched at chromosome ends

(A-B) Metaplot of Bio-lurbinectedin Chem-seq coverage at chromosome ends within 0–25 kb from telomeric regions from two biological replicates. **(C)** Metaplots of Bio-lurbinectedin Chem-seq coverage at chromosome ends (0–25 kb from telomeres) in ALT (U2OS) and non-ALT (HeLa and A549) cell lines.

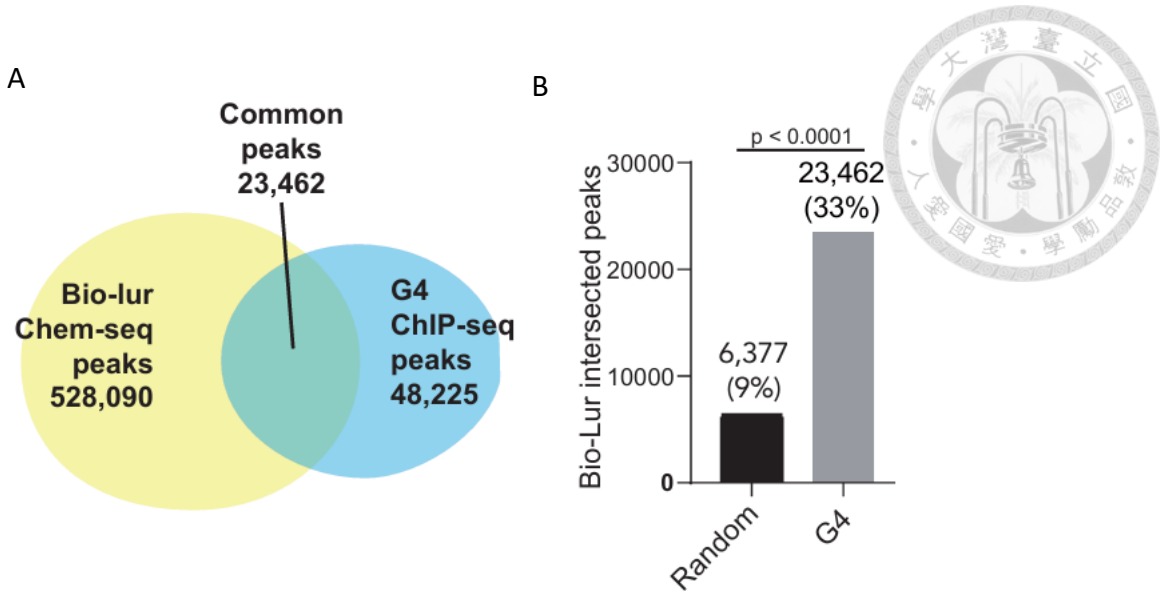
doi:10.6342/NTU202504254



Chapter 1 Figure 4. Enrichment of lurbinectedin Chem-seq at telomeres

(A) Genome browser views showing Bio-lurbinectedin Chem-seq coverage at multiple chromosome ends. **(B)** Genome browser views showing examples of the enrichment of lurbinectedin Chem-seq at telomeres.

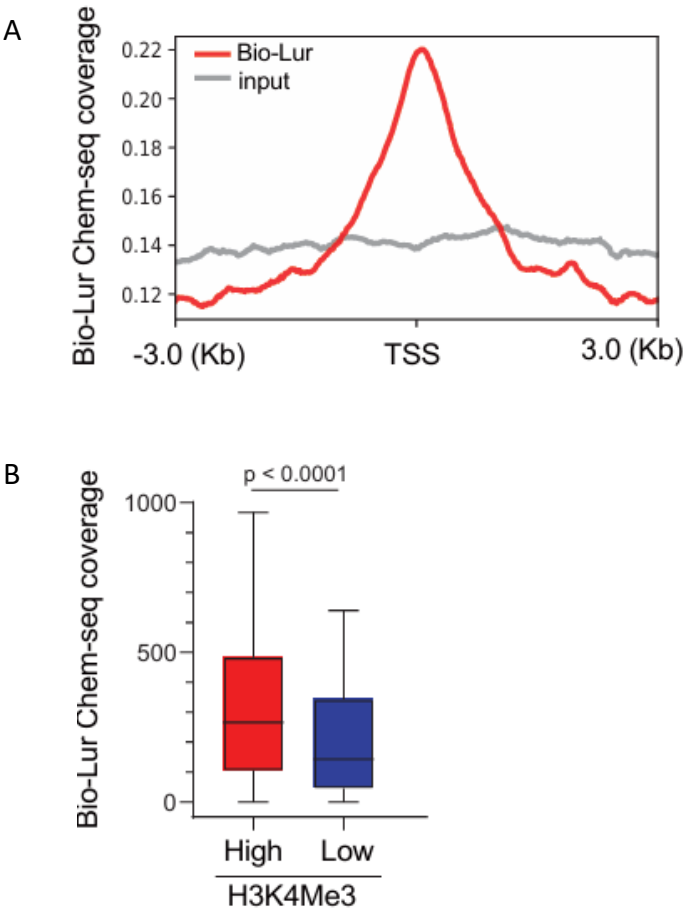
doi:10.6342/NTU202504254



Chapter 1 Figure 5. Lurbinectedin binding sites are associated with DNA G-quadruplex (G4) structures in ALT cancer cells.

(A) Number and percentage of Bio-lurbinectedin Chem-seq peaks in U2OS cells overlapping with G4 ChIP-seq peaks from HaCaT cells. **(B)** Bar graph showing the number of G4 binding sites containing Bio-lurbinectedin Chem-seq peaks. Random peaks were used as a control for overlap comparison. P-values were calculated using Fisher's exact test.

doi:10.6342/NTU202504254



Chapter 1 Figure 6. Lurbinectedin binding at transcription start sites

(A) Metagene analysis of Bio-lurbinectedin Chem-seq coverage around transcription start sites (TSSs). **(B)** Bar graph presenting the coverage of Bio-lur Chem-seq at TSSs with high or low H3K4me3 levels. P-values were calculated using Mann–Whitney test.

doi:10.6342/NTU202504254

REFERENCES

1. Palm W, de Lange T. How shelterin protects mammalian telomeres. Annu Rev Genet. 2008;42:301-334.
2. Nandakumar S, Cech TR. Structural biology of shelterin and telomeric chromatin. Curr Opin Struct Biol. 2024;79:102632.
3. Griffith JD, Comeau L, Rosenfield S, et al. Mammalian telomeres end in a large duplex loop. Cell. 1999;97:503-514.
4. Olovnikov AM. Principle of marginotomy in the synthesis of polynucleotides. J Theor Biol. 1973;41:181-190.
5. Coluzzi E, et al. Factors that influence telomeric oxidative base damage and repair. DNA Repair. 2014;21:60-70.
6. d'Adda di Fagagna F, Reaper PM, Clay-Farrace L, et al. A DNA-damage checkpoint response in telomere-initiated senescence. Nature. 2003;426:194-198.
7. Kim NW, Piatyszek MA, Prowse KR, et al. Specific association of human telomerase activity with immortal cells and cancer. Science. 1994;266:2011-2015.
8. Shay JW, Wright WE. Telomeres and telomerase: three decades of progress. Nat Rev Genet. 2019;20:299-309.
9. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet. 2010;11:319-330.
10. Dilley RL, Verma P, Cho NW, et al. Break-induced telomere synthesis underlies alternative telomere maintenance. Nature. 2016;539:54-58.
11. Henson JD, Neumann AA, Yeager TR, Reddel RR. Assay for monitoring ALT activity in mammalian cells. Nat Protoc. 2009;4:1831-1843.
12. Arora R, Lee YC, Wischnewski H, et al. RNase H1 regulates TERRA-dependent R-loop formation at telomeres. Nat Commun. 2014;5:5220.
13. Heaphy CM, de Wilde RF, Jiao Y, et al. Altered telomeres in tumours with ATRX and DAXX mutations. Nat Genet. 2011;43:748-753.
14. Lovejoy CA, Li W, Reisenweber S, et al. Loss of ATRX, genome instability and an altered DNA damage response are hallmarks of ALT. PLoS Genet. 2012;8:e1002772.

doi:10.6342/NTU202504254

15. Carson LM, Flynn RL. Highlighting vulnerabilities in the alternative lengthening of telomeres pathway. Curr Opin Pharmacol. 2023;70:102380.
16. Paeschke K, Bochman ML, Garcia PD, et al. Pif1 family helicases suppress genome instability at G-quadruplex motifs. Nature. 2013;497:458-462.
17. Flynn RL, Cox KE, Jeitany M, et al. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. Science. 2015;347:273-277.
18. Deeg KI, et al. Cancer cells with ALT do not display general hypersensitivity to ATR inhibition. Front Oncol. 2016;6:186.
19. Hansel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat Rev Mol Cell Biol. 2017;18:279-284.
20. Li Y, Smith P, LaBoissiere S, et al. Elevated ROS drives telomeric R-loop accumulation and sustains ALT activity. Nat Commun. 2024;15:4112.
21. Panda S, Avin B, Gupta R, et al. Promoter G-quadruplex stabilisation selectively kills ALT cancer cells. ACS Chem Biol. 2024;19:1433-1439.
22. Santamaría-Nuñez G, Vara-Ciruelos D, Moneo V, et al. Lurbinectedin triggers degradation of phosphorylated RNA polymerase II. Mol Cancer Ther. 2016;15:2399-2412.
23. Trigo J, et al. Lurbinectedin as second-line treatment for small-cell lung cancer. Lancet Oncol. 2020;21:645-653.
24. Leal JF, Martínez-Díez M, García-Hernández V, et al. PM01183 (lurbinectedin) inhibits growth of soft-tissue sarcoma models. Drugs. 2020;80:2189-2200.
25. Conroy R. Lurbinectedin plus atezolizumab earns FDA priority review for ES-SCLC maintenance (IMforte trial). Cancer Network News. 2025 Jun 10.

doi:10.6342/NTU202504254

Chapter 2 Telomeric Repeat-Containing RNA is associated with human aging

INTRODUCTION

1 Telomeres and Transcription at Chromosome Ends

Telomeres are nucleoprotein structures at eukaryotic chromosome ends composed of tandem TTAGGG repeats bound by the shelterin complex (TRF1, TRF2, TIN2, RAP1, POT1, TPP1); these complexes are essential for preventing DNA-damage signaling and chromosome fusion [1]. Although telomeres were initially deemed transcriptionally silent—the telomere-position effect (TPE) showed telomere-proximal reporters were repressed [2]—subsequent work revealed that telomeres themselves are transcribed, giving rise to the long-non-coding telomeric repeat-containing RNA (TERRA) [3]. Telomeric transcription was first detected in *Trypanosoma brucei* [4], later in mouse and *Arabidopsis* [5], and human HeLa cells [3], underscoring evolutionary conservation. TERRA molecules comprise UUAGGG repeats plus subtelomeric sequence, range from several hundred nucleotides to kilobases, and occur in both polyadenylated and non-polyadenylated forms [6].

2 TERRA Biogenesis, Structure, and Regulatory Mechanisms

TERRA is transcribed by RNA polymerase II from multiple chromosome ends,

doi:10.6342/NTU202504254

frequently initiating in CpG-rich subtelomeric promoters that contain conserved 61-29-37-nt tandem repeats [7]. Epigenetic state—DNA methylation and histone marks such as H3K4me3, H3K27ac and H3K9me3—together with transcription factors CTCF and ATRX shapes TERRA output [8]. Approximately 90 % of transcripts are non-polyadenylated, whereas a minor (~7 %) poly(A)$^+$ fraction is more stable and differentially localized [6]. TERRA readily forms RNA–DNA hybrids (R-loops) at telomeres, and is enriched in ALT cancer cells and broadly linked to genome instability [9].

3 Functional Roles of TERRA

TERRA plays multifaceted roles in maintaining telomere integrity and regulating chromatin structure. TERRA suppress heterochromatin formation in the subtelomeric regions in mouse embryonic stem cells [18], and modulates telomerase activity by acting as a negative regulator of telomerase access to chromosome ends [11]. TERRA is also involved in the formation of RNA:DNA hybrids known as R-loops, particularly at telomeres in ALT cells, where it promotes homologous-recombination-based telomere elongation [9, 12]. While R-loops can facilitate recombination, excessive accumulation leads to replication stress and genome instability [12]. Beyond its telomeric functions, TERRA binds non-telomeric chromatin regions and regulates gene expression both in cis and in trans [13]. It interacts with various chromatin regulators and transcription factors,

doi:10.6342/NTU202504254

contributing to broader genome organization. A notable interaction is with ATRX, a chromatin remodeler involved in histone variant deposition. TERRA limits ATRX recruitment to repetitive chromatin regions, thereby influencing heterochromatin composition without completely displacing ATRX from the genome [14]. Additionally, TERRA contains guanine-rich sequences capable of forming G-quadruplex structures (G4), which influence transcriptional regulation near promoters [10]. TERRA also can regulate gene expression via regulating ATRX and G4 structures near the transcription start sites. It has been reported that TERRA depletion in mouse embryonic stem cells results in decrease of G-quadruplex structures and increase of ATRX occupancy at TSSs [18]. TERRA itself is bound and regulated by multiple RNA-binding proteins, including members of the hnRNP family such as RALY, which stabilizes non-polyadenylated TERRA transcripts [15]. In budding yeast lacking telomerase, TERRA levels increase during replicative senescence, while experimental ablation of TERRA by expressing an artificial antisense TERRA transcript delays the onset of senescence [16]. Similarly, increased TERRA expression has been reported in Hutchinson–Gilford progeria syndrome (HGPS) patient cells, linking TERRA dysregulation to premature aging and altered telomeric chromatin states [17].

doi:10.6342/NTU202504254

4 Technological Challenges and Experimental Strategy

Short-read RNA-seq, Northern blot and RNA-FISH offer limited resolution and cannot reliably assign TERRA transcripts to specific chromosome ends. To overcome these, we employed Oxford Nanopore direct RNA sequencing in U2OS cells, generating long reads that defined genome-wide TERRA-producing loci. The resulting chromosome-end annotations were validated and extended by sequencing native RNA from HeLa cells. We then pursued a multi-pronged strategy: (i) locus-specific RT-qPCR with CHM13-based subtelomeric primers; (ii) separation of total TERRA into poly(A)+/− fractions for abundance profiling; and (iii) expression analysis across multiple human tissues.

doi:10.6342/NTU202504254

MATERIALS AND METHODS

1 Cell culture

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

2 RNA extraction

Total RNA was extracted from cell pellets using QIAzol Lysis Reagent (Qiagen, Cat. No. 79306) according to the manufacturer's protocol with minor modifications. Cells were first collected by centrifugation and washed once with cold phosphate-buffered saline (PBS). The resulting cell pellet was lysed directly in 1 mL of QIAzol Lysis Reagent by thorough pipetting until complete homogenization was achieved. The lysate was incubated at room temperature for 5 minutes to permit the dissociation of nucleoprotein complexes. For phase separation, 0.2 mL of chloroform was added to the homogenate. Tubes were tightly capped, shaken vigorously for 15 seconds, and incubated at room temperature for an additional 2 minutes. The samples were then centrifuged at 12,000 × g for 15 minutes at 4°C to separate the mixture into aqueous and organic phases. The upper aqueous phase, containing RNA, was carefully transferred to a new RNase-free

doi:10.6342/NTU202504254

microcentrifuge tube. To precipitate RNA, an 1.25x volume of isopropanol was added to the aqueous phase, mixed by gentle inversion, and incubated at -30°C for over night. Samples were centrifuged again at 12,000 × g for 10 minutes at 4°C to collect the RNA pellet. The pellet was washed once with 75% ethanol (prepared with RNase-free water) and centrifuged at 12,000 × g for 5 minutes at 4°C. After removing the supernatant, the RNA pellet was briefly air-dried and then resuspended in DEPC-treated water.

3 TERRA capture for Illumina RNA-seq

Total RNA was isolated from 1×10^7 cells with QIAzol Lysis Reagent. To remove genomic DNA, 0.4 U μl^{-1} DNase I (Invitrogen, AM2238) and 10 mM ribonucleoside-vanadyl complexes (New England Biolabs, E7760S) were added and incubated at 37 °C for 15 min; the reaction was quenched with 5 mM EDTA, and the sample was repurified with QIAzol Lysis Reagent. For each TERRA capture reaction, 500 µg of total RNA was mixed with 600 µl of DEPC-treated water, 300 µl of 20× SSC buffer, 10 µl of 10 µM biotinylated TERRA antisense oligonucleotide ($(CCCTAA)_5$-TEG-biotin), and 90 µl of DEPC-treated water. The RNA and probe were separately denatured at 70°C for 2 minutes, then combined and incubated again at 70°C for 8 minutes to promote hybridization. The mixture was subsequently transferred to a 44°C incubator and rotated for 1 hour to allow annealing. During the annealing step, MyOne C1 streptavidin magnetic beads

doi:10.6342/NTU202504254

(Invitrogen™) were prepared by vortexing and washing twice with DEPC-treated water. Beads were then resuspended in 6× SSC buffer and incubated at 37°C for at least 10 minutes for equilibration. After annealing, the RNA-probe mixture was cooled from 44°C to 37°C. Then, 100 µl of prepared beads were added to the annealed RNA-probe mixture (total 1 ml), and the sample was rotated at 37°C for 15 minutes to capture the hybridized TERRA molecules. Beads were subsequently washed to remove nonspecifically bound material. Washing steps included four washes with 2× SSC containing 0.1% NP-40 at 37°C, one wash with 1× SSC containing 0.1% NP-40 at 37°C, one wash with 1× SSC containing 0.1% NP-40 at room temperature, and one wash with 1× SSC without NP-40 at 37°C. For each wash, beads were resuspended in buffer, rotated for 5 minutes, and separated on a magnetic stand. After the final wash, beads were resuspended in 30 µl of DEPC-treated water and heated at 70°C for 5 minutes to elute the captured RNA. The eluate was separated by magnet, collected, and the elution step was repeated once more to maximize RNA recovery. Eluates from multiple captures were pooled and precipitated by adding 0.1 volume of 3 M sodium acetate, 2 µl of GlycoBlue™ coprecipitant, and 3 volumes of 100% ethanol, followed by incubation at –30°C for several days. Precipitated RNA was collected by centrifugation, washed twice with 70% ethanol, air-dried, and resuspended in RNase-free water.

doi:10.6342/NTU202504254

4 TERRA capture for Nanopore Direct RNA-seq

Cells from two hundred 150-mm dishes were harvested and extracted with TRIzol as above, but without DNase treatment. Each capture used 500 µg total RNA; ten captures were pooled, and precipitated for 3 days at −20 °C with 3 vol 100 % EtOH, 1/10 vol 3 M NaOAc and 2 µl GlycoBlue™. Precipitated RNA from 100 captures (≈ 50 mg input) yielded 3.4 µg TERRA-enriched RNA after centrifugation (21 100 × g, 4 °C, 20 min) and 70 % EtOH wash, followed by elution in 20 µl water.

5 Poly(A)+ and Poly(A)− TERRA Isolation

Following TERRA capture, RNA samples were subjected to poly(A)+ and poly(A)− fractionation using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB, #E7490), following the manufacturer's standard protocol with minor modifications. TERRA-enriched RNA was first diluted with nuclease-free water to a final volume of 50 µl and kept on ice. For each sample, 20 µl of NEBNext Oligo $d(T)_{25}$ magnetic beads were transferred to a 1.5 mL RNase-free tube and washed twice with 100 µl of 2× RNA Binding Buffer (provided in the kit) by gentle pipetting. Beads were separated from the wash buffer using a magnetic stand, ensuring minimal disturbance to the bead pellet. After the final wash, beads were resuspended in 50 µl of 2× RNA Binding Buffer and mixed thoroughly. Fifty microliters of the prepared beads were added to the 50 µl RNA sample,

doi:10.6342/NTU202504254

and the mixture was pipetted gently to ensure homogeneity. The samples were immediately placed into a thermal cycler with a heated lid set at ≥75°C and subjected to the following incubation: 65°C for 5 minutes, followed by cooling to 4°C. Upon completion of the thermal program, samples were incubated at room temperature for 10 minutes to enhance binding. Beads were then magnetically separated, and the supernatant containing poly(A)– RNA (non-polyadenylated TERRA fraction) was carefully removed and saved for downstream analysis. The beads bound with poly(A)+ RNA were subjected to two washes with 200 µl of NEBNext Wash Buffer at room temperature to remove residual non-specific RNA. After washing, poly(A)+ RNA was eluted from the beads by adding 50 µl of NEBNext Tris Buffer, followed by heating at 80°C for 2 minutes and immediate cooling to 25°C. Beads were magnetically separated, and the supernatant containing the purified poly(A)+ TERRA RNA was collected.

6 Library Preparation for Illumina Sequencing

After first-strand cDNA synthesis from TERRA-enriched RNA, second-strand synthesis was performed using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Cat# E7765) according to the manufacturer's instructions. Specifically, 20 µl of first-strand product was combined with 8 µl of NEBNext Second Strand Synthesis Reaction Buffer (with dUTP Mix, 10X), 4 µl of

doi:10.6342/NTU202504254

NEBNext Second Strand Synthesis Enzyme Mix, and 48 µl of nuclease-free water, yielding a final volume of 80 µl. The reaction was incubated at 16°C for 1 hour in a thermocycler with the lid set to ≤ 40°C. The resulting double-stranded cDNA was purified using 144 µl (1.8×) of NEBNext Sample Purification Beads. Beads were mixed thoroughly with the cDNA by pipetting, incubated at room temperature for 5 minutes, then separated on a magnetic stand. The beads were washed twice with 200 µl of freshly prepared 80% ethanol, air-dried until no visible liquid remained (3–5 minutes), and eluted in 53 µl of 0.1× TE buffer. After a brief incubation, 50 µl of the eluate was collected for end-repair. End repair and A-tailing were performed in a total volume of 60 µl by combining 50 µl of the purified cDNA with 7 µl of NEBNext Ultra II End Prep Reaction Buffer and 3 µl of NEBNext Ultra II End Prep Enzyme Mix. The mixture was incubated at 20°C for 30 minutes, followed by 65°C for 30 minutes, and held at 4°C. Adaptor ligation was then carried out by adding 2.5 µl of diluted NEBNext Adaptor, 1 µl of NEBNext Ligation Enhancer, and 30 µl of NEBNext Ultra II Ligation Master Mix to the 60 µl end-prepped cDNA, resulting in a 93.5 µl reaction volume. The ligation reaction was incubated at 20°C for 15 minutes. After ligation, 3 µl of USER enzyme was added and incubated at 37°C for 15 minutes to degrade the second strand, ensuring strand specificity. The ligated library was purified and size-selected using 90 µl (0.9×) of NEBNext Sample Purification Beads. After incubation and magnetic separation, the

doi:10.6342/NTU202504254

beads were washed twice with 80% ethanol, briefly air-dried, and the DNA was eluted in 17 µl of 0.1× TE buffer. A total of 15 µl was recovered for PCR amplification. PCR enrichment was performed in a 50 µl reaction containing 15 µl of adaptor-ligated DNA, 25 µl NEBNext Ultra II Q5 Master Mix, and 10 µl NEBNext Index Primer Mix (NEBNext Multiplex Oligos for Illumina). The thermal cycling conditions were: 98°C for 30 seconds, followed by 10–14 cycles of 98°C for 10 seconds and 65°C for 75 seconds, with a final extension at 65°C for 5 minutes. The exact cycle number was determined by prior qPCR-based evaluation to prevent overamplification. After PCR, libraries were cleaned up using 45 µl (0.9×) NEBNext Sample Purification Beads, washed twice with 80% ethanol, and eluted in 23 µl of 0.1× TE buffer. Finally, 20 µl was recovered and assessed for quality and concentration using the Agilent Bioanalyzer (High Sensitivity DNA Kit) and Qubit dsDNA HS Assay Kit, respectively. Qualified libraries were pooled and sequenced on the Illumina NovaSeq or MiSeq platform using paired-end 150 bp reads.

7 Nanopore Direct RNA Sequencing Library Preparation

Following the enrichment of TERRA RNA through hybridization with a biotinylated $(CCCTAA)_5$ antisense probe and streptavidin magnetic bead purification, the eluted RNA was subjected to enzymatic polyadenylation in preparation for Oxford Nanopore direct

doi:10.6342/NTU202504254

RNA sequencing. Up to 1 μg of purified RNA was incubated in a 20 μl reaction containing 1× E. coli Poly(A) Polymerase Reaction Buffer, 1 mM ATP, and 1 μl of E. coli Poly(A) Polymerase (New England Biolabs, Cat# M0276S). The reaction was maintained at 37°C for 30 minutes to ensure the addition of a poly(A) tail compatible with Nanopore library preparation. Upon completion, the RNA was purified using a 1.8× volume of RNAClean XP beads (Beckman Coulter, Cat# A63987), washed twice with 70% ethanol, air-dried briefly, and resuspended in 8 μl of nuclease-free water (Thermo Fisher Scientific, Cat# AM9937). RT adapter annealing and ligation were then performed by adding 3 μl of NEBNext Quick Ligation Reaction Buffer (NEB, Cat# B6058), 1 μl of RT Adapter (RTA; SQK-RNA004, Blue Cap), 1 μl of Murine RNase Inhibitor (NEB, Cat# M0314), and 1.5 μl of T4 DNA Ligase (NEB, Cat# M0202M) to the polyadenylated RNA. This 14.5 μl mixture was incubated at room temperature for 10 minutes to facilitate efficient ligation of the RT adapter. Subsequently, reverse transcription was performed by adding a freshly prepared 23 μl master mix containing nuclease-free water, 10 mM dNTPs (NEB, Cat# N0447), and 5× Induro RT Reaction Buffer (NEB, Cat# M0681) to the ligation reaction, followed by the addition of 2 μl of Induro Reverse Transcriptase (NEB, Cat# M0681), bringing the total volume to 39.5 μl. The reaction was incubated at 60°C for 30 minutes to synthesize cDNA, and the enzyme was inactivated by heating at 70°C for 10 minutes. The resulting RNA–cDNA hybrids were then purified using 1.8× RNAClean XP beads

doi:10.6342/NTU202504254

and eluted in 23 μl of nuclease-free water. For sequencing adapter ligation, 8 μl of NEBNext Quick Ligation Reaction Buffer, 6 μl of RNA Ligation Adapter (RLA; SQK-RNA004, Green Cap), and 3 μl of T4 DNA Ligase were added to the 23 μl RT product, yielding a total reaction volume of 40 μl. This ligation step was carried out at room temperature for 10 minutes. The adapter-ligated library was purified with 0.4× (16 μl) RNAClean XP beads and washed twice with 150 μl of Wash Buffer (WSB; SQK-RNA004, Orange Cap). The final library was eluted with 13 μl of RNA Elution Buffer (REB; SQK-RNA004, Black Cap). To prepare for sequencing, the RNA library was quantified using the Qubit 1× dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat# Q33230), ensuring a yield of at least 30 ng. A final sequencing mix of 75 μl was prepared by combining 12 μl of the library with 37.5 μl of Sequencing Buffer (SB; Red Cap) and 25.5 μl of Library Solution (LIS; White Cap). The mix was gently loaded onto a FLO-MIN004RA flow cell pre-equilibrated with Flush Buffer and Flush Tether. Sequencing was performed on the MinION platform using MinKNOW software with live basecalling enabled.

8 NGS Sequencing for TERRA RT-qPCR Products

Total RNA was first extracted using TRIzol™ Reagent (Invitrogen™, Cat# 15596018) following the manufacturer's instructions. For reverse transcription, 500 ng of total RNA was used as input and incubated with 2.5 nM random hexamer primers and

doi:10.6342/NTU202504254

SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, Cat# 18090200) under the recommended conditions to generate cDNA. Subsequent amplification of telomeric repeat-containing RNA (TERRA) was performed via quantitative PCR using IQ™ SYBR® Green Supermix (Bio-Rad, Cat# 170-8882), with primer sequences designed based on the CHM13 reference genome (see Supplementary Table 6 in the original publication). Following RT-qPCR, the amplified TERRA products were purified using the Gel/PCR DNA Fragment Extraction Kit (FAVORGEN, Cat# FAGCK 001-1). Purified PCR products were then subjected to DNA library construction using the NEBNext® Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Cat# E7645S), following the manufacturer's guidelines for adapter ligation, size selection, and PCR enrichment. Sequencing was conducted using a paired-end 150 bp run on an Illumina NovaSeq platform.

9 Analysis of TERRA Expression by TERRA-QUANT

To quantitatively assess TERRA expression across different human tissues and cell lines, we utilized TERRA-QUANT, a computational pipeline developed based on the TERRA transcription regions identified in the T2T-CHM13 reference genome. Public RNA-seq datasets were first downloaded and pre-processed using SRAToolkit and TrimGalore. The processed reads were aligned to the T2T-CHM13v1.1 genome using the

doi:10.6342/NTU202504254

STAR aligner with default settings. To facilitate quantification, we modified the GTF annotation file to include both telomeric repeat regions and subtelomeric regions corresponding to Type I, II, and III TERRA transcription units. Read counts were calculated using HTseq-count (-m intersection-nonempty --nonunique all) to obtain expression values for both annotated genes and TERRA regions. For chromosome-end-specific analysis, we extracted high-confidence mapped reads (MAPQ ≥ 30 for paired-end or MAPQ = 255 for single-end reads) using SAMtools. These reads were filtered to include only those mapped within subtelomeric TERRA regions, excluding reads mapped exclusively to pure telomeric repeats. For quantifying total TERRA levels, STAR-aligned reads were filtered for primary alignments (MAPQ ≥ 1), duplicates were removed, and TERRA-mapped reads were aggregated. These counts were then normalized together with all other gene expression counts using the YARN R package, which enables cross-sample normalization across diverse tissues and conditions.

10 cDNA Synthesis and Quantitative PCR

All TERRA-enriched RNA obtained after capture and fractionation was used for complementary DNA (cDNA) synthesis. The RNA was reverse transcribed using SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, Cat#18090200) according to the manufacturer's protocol with minor modifications. Specifically, the

doi:10.6342/NTU202504254

entire amount of TERRA RNA was mixed with 1 μM of a telomeric C-rich primer (5'-$(CCCTAA)_5$-3') in a total volume of 13 μl. The mixture was heated at 65°C for 5 minutes to denature RNA secondary structures and immediately chilled on ice for at least 1 minute. Following primer annealing, 7 μl of a reverse transcription master mix containing 4 μl of 5× SuperScript IV buffer, 1 μl of 100 mM DTT, 1 μl of 10 mM dNTPs, and 1 μl of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Thermo Fisher Scientific, Cat#10777019) was added to each reaction tube. Subsequently, 1 μl of SuperScript™ IV Reverse Transcriptase (200 U/μl) was added, and the reactions were incubated at 55°C for 10 minutes, followed by 80°C for 10 minutes for enzyme inactivation.

doi:10.6342/NTU202504254

RESULTS

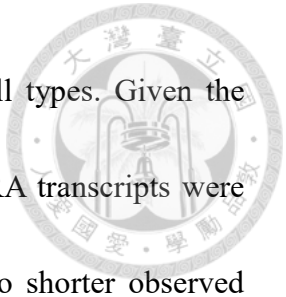
1 HeLa TERRA Profiling by Integrated Long- and Short-Read Sequencing

To verify the TERRA transcription regions, we performed TERRA-capture RNA-seq in telomerase-positive cells (HeLa) using both Nanopore direct RNA sequencing and Illumina capture RNA-seq. First, we performed RT-qPCR to detect TERRA enrichment after TERRA capture and found tens of thousands-fold enrichment was observed after TERRA capture (Chapter 2 Figure 1). TERRA enrichment in HeLa was detected at defined chromosome ends, showing a consistent pattern with U2OS-defined transcription regions. This was confirmed by heatmaps comparing Nanopore and Illumina signal intensity (Chapter 2 Figure 2) and genome browser views illustrating subtelomeric enrichment profiles (Chapter 2 Figure 3).

Nanopore sequencing of captured TERRA from HeLa yielded 153 high-quality reads with a mean length of 982 bp and a maximum length of 4639 bp (Chapter 2 Figure 4). In comparison, U2OS TERRA reads had a mean length of 721 bp and a maximum of 2478 bp. For the pure telomeric repeat tracts within TERRA reads, the mean length was 284 bp in HeLa and 222 bp in U2OS, with maximum lengths of 1487 bp and 1080 bp, respectively.

These observations demonstrate that telomeric repeat tracts within TERRA transcripts

doi:10.6342/NTU202504254



span several hundred to over one thousand nucleotides in both cell types. Given the inherent fragility of RNA molecules, it is possible that some TERRA transcripts were partially degraded during sample preparation, potentially leading to shorter observed lengths and increased variation in transcript size.

2 Validation of subtelomeric primers for chromosome-specific TERRA

We designed seven primer pairs targeting unique subtelomeric regions in the T2T-CHM13 assembly and confirmed exclusive alignment to their intended chromosome ends. RT-qPCR revealed that U2OS cells exhibit significantly higher TERRA levels than HeLa when assayed with these CHM13 primers (Chapter 2 Figure 5A). To assess specificity, we sequenced the RT-PCR amplicons by Illumina and calculated the fraction of reads mapping back to each target end (Chapter 2 Figure 5B). CHM13-1q, -3q, -8p and -15q primer sets yielded high percentage on-target reads, whereas CHM13-4q, -13q and -22q showed partial off-target mapping—predominantly to 8q—consistent with local sequence homology (Chapter 2 Figure 6). In parallel, we tested eight previously published subtelomeric primer sets designed against hg38; none achieved comparable specificity, with each amplifying TERRA across multiple chromosome ends (Chapter 2 Figure 4B). Overall, CHM13-derived primers robustly outperformed hg38-based sets in

doi:10.6342/NTU202504254

chromosome-specific TERRA detection.

3 Polyadenylation profiles of TERRA across chromosome ends

To quantify TERRA expression in short-read RNA-seq datasets, we developed a bioinformatics pipeline termed TERRA-QUANT based on TERRA transcription regions (Chapter 2 Fig. 7A). TERRA-QUANT measures reads mapped to these regions, including all three transcript classes: Type I, II, and III, where Type I denotes canonical subtelomeric promoters (37 of 46 chromosome ends) that carry the characteristic 61-29-37 bp tandem repeat adjacent to telomeric repeats, Type II refers to the atypical subtelomeric promoters on chr10p and chr17p that lack the 61-29-37 bp repeat but still initiate TERRA transcription. Besides the transcription that occurs at chromosome ends, TERRA is also generated from interstitial telomeric sequences (ITSs), which are defined as Type III TERRA transcription regions. For total TERRA levels, only reads mapped to TERRA transcription regions are counted, followed by normalization using the YARN package. For chromosome-end-specific analysis, TERRA-QUANT quantifies reads from subtelomeric regions only, excluding pure telomeric repeat reads to avoid multi-mapping bias. Because conventional RNA-seq predominantly captures polyadenylated transcripts, we next sought to distinguish the genomic origins of poly(A)+ versus non-poly(A)

doi:10.6342/NTU202504254

TERRA. Following TERRA capture, we split each sample into poly(A)+ and poly(A)– fractions using the NEBNext® Poly(A) mRNA Magnetic Isolation Module, then performed deep Illumina sequencing on both. Analysis of chromosome-end-specific read counts revealed that certain loci—chr1p, chr2q, chr3p and chr18q—were enriched for poly(A)+ TERRA in both U2OS and HeLa cells, whereas others—chr8p, chr11p, chr15q and chr21q—were predominantly represented in the poly(A)– fraction (Chapter 2 Figure 7B). Importantly, the ratio of poly(A)+ to poly(A)– reads at each chromosome end was highly correlated between U2OS and HeLa (Pearson's $r = 0.70$, $p < 0.0001$, Chapter 2 Figure 7C), indicating that locus-specific polyadenylation of TERRA is a conserved feature across these cell types.

4 TERRA Expression Increases with Human Aging.

Applying TERRA-QUANT to poly(A)-enriched RNA-seq datasets from blood, brain, and skin fibroblasts, we observed age-associated TERRA upregulation. Across 222 blood samples, TERRA expression increased significantly with age, with older donors (≥ 60 years) exhibiting higher levels than middle-aged and young individuals (Chapter 2 Figure 8A, B). In brain tissues, analysis of 90 RNA-seq datasets similarly revealed a positive correlation between age and TERRA expression (Chapter 2 Figure 9A, B). We further analyzed 132 fibroblast RNA-seq datasets and observed a consistent increase in TERRA levels in samples from aged donors (≥ 75 years) compared to younger individuals

doi:10.6342/NTU202504254

(Chapter 2 Figure 10A, B). Hutchinson–Gilford progeria syndrome (HGPS) is a rare premature aging disorder caused by LMNA gene mutations, leading to nuclear shape abnormalities, genomic instability, and accelerated telomere attrition. Fibroblasts derived from HGPS patients exhibited abnormally heterogeneous TERRA expression profiles compared to age-matched controls (Chapter 2 Figure 10C), likely reflecting altered telomere status and chromatin structure in progeroid cells.

doi:10.6342/NTU202504254

DISUCSSION

In this study, we defined the transcription regions and expression profiles of TERRA across different cell types, tissues, and age groups using a combination of capture RNA-seq, Nanopore direct RNA-seq, and the TERRA-QUANT pipeline. Our work established four key findings: (1) confirmation of TERRA transcription regions in telomerase-positive cells, (2) validation of chromosome-specific detection using CHM13-based subtelomeric primers, (3) characterization of polyadenylation profiles across chromosome ends, and (4) demonstration that TERRA expression increases with human aging.

First, by performing TERRA capture RNA-seq in HeLa cells and comparing it with U2OS, we showed that telomerase-positive cells also express TERRA from defined chromosome ends. Nanopore sequencing revealed that telomeric repeat tracts in TERRA transcripts spanned several hundred nucleotides, consistent across U2OS and HeLa, although with differences in mean length. These observations confirm that TERRA is broadly present beyond ALT contexts and that its structure is conserved.

Second, we validated a set of CHM13-derived subtelomeric primers that outperform hg38-based primers in detecting chromosome-specific TERRA expression. While hg38 primers have been widely used, they often amplify TERRA transcripts arising from multiple chromosome ends due to sequence similarity, reducing their specificity. In

doi:10.6342/NTU202504254

contrast, most CHM13 primers demonstrated significantly improved on-target specificity. However, some CHM13 primer sets—such as those targeting 4q, 13q, and 22q—still captured signals from unintended loci like 8p, reflecting the inherent sequence homology shared among subtelomeric regions. This highlights a key technical challenge: even with improved primer design based on the T2T-CHM13 assembly, complete chromosome-end specificity remains difficult to achieve by RT-qPCR. Nonetheless, CHM13-derived primers represent a substantial advancement over hg38-based sets, providing a more reliable tool for locus-specific TERRA analysis.

Third, we uncovered that TERRA polyadenylation profiles vary systematically across chromosome ends. Commonly available RNA-seq datasets predominantly use poly(A)-enriched RNA libraries, which may overlook the substantial fraction of TERRA molecules lacking poly(A) tails. To address this, we performed both poly(A)+ and poly(A)− TERRA capture sequencing in U2OS and HeLa cells. Our analysis revealed that poly(A)+ TERRA transcripts are preferentially produced from certain chromosome ends—such as chr1p, 2q, 3p, and 18q—while others, like chr8p, 11p, 15q, and 21q, mainly give rise to poly(A)− TERRA. Importantly, this locus-specific polyadenylation pattern was conserved between U2OS and HeLa cells, indicating that the bias is an intrinsic property of TERRA transcription rather than a cell-type-specific phenomenon. These results also underscore the importance of considering both poly(A)+ and poly(A)−

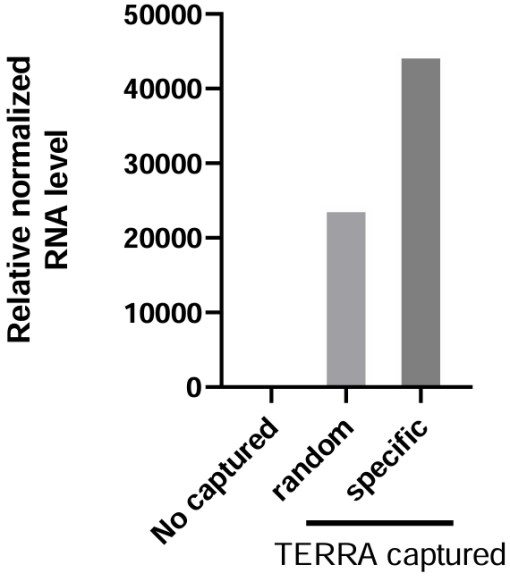
doi:10.6342/NTU202504254

fractions when studying TERRA expression profiles.

Finally, we observed consistent upregulation of TERRA with human aging across blood, brain, and fibroblasts, using both publicly available and laboratory-generated datasets. Our analysis demonstrated that older donors exhibited higher TERRA levels than younger individuals in all tested tissue types. In fibroblasts derived from Hutchinson–Gilford progeria syndrome (HGPS) patients, TERRA expression profiles were abnormally heterogeneous compared to age-matched controls. These findings suggest that increased TERRA expression is a common feature of aging and may reflect changes in telomere transcription regulation associated with age-related cellular states.

doi:10.6342/NTU202504254

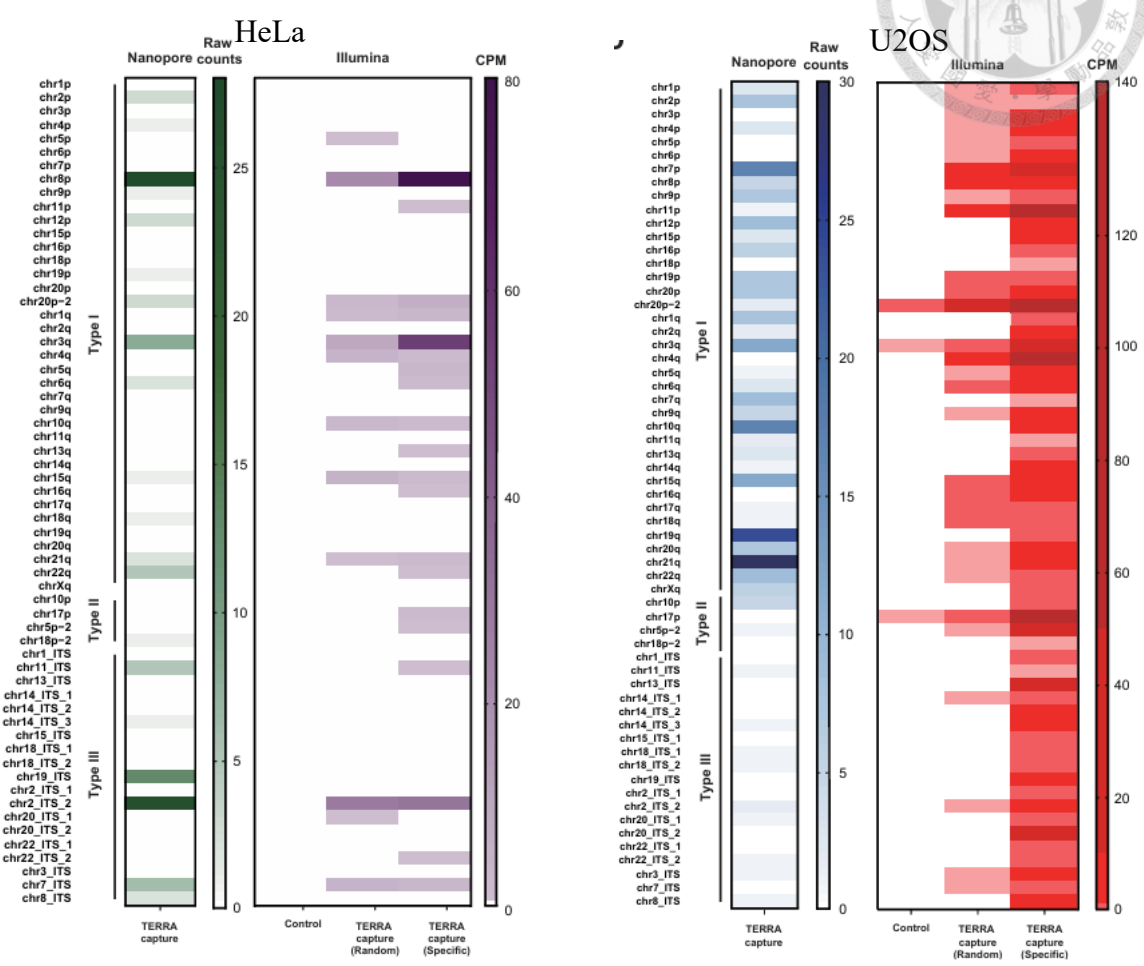
FIGURES AND TABLES



Chapter 2 Figure 1. TERRA is enriched by TERRA-capture

RT-qPCR results show TERRA enrichment in HeLa cells after TERRA capture for Illumina RNA-seq. cDNA synthesis was performed using either random or telomeric-specific primers for Illumina RNA-seq.

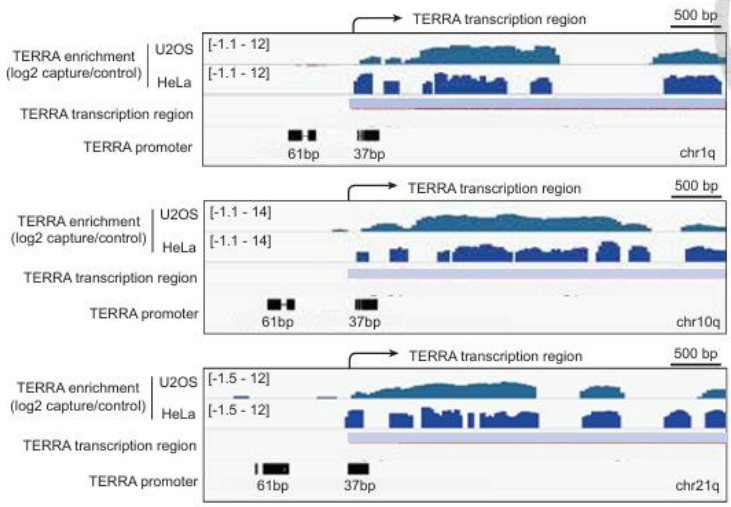
doi:10.6342/NTU202504254



Chapter 2 Figure 2. HeLa TERRA Profiling by Integrated Long- and Short-Read Sequencing

Heatmap showing TERRA read counts from different chromosome ends and interstitial telomeric sequences (ITSs) in HeLa and U2OS cells. Reads from Illumina or Nanopore sequencing were mapped to the T2T-CHM13 genome. CPM, counts per million mapped reads. This data was published in 2025 in NAR [19]

doi:10.6342/NTU202504254

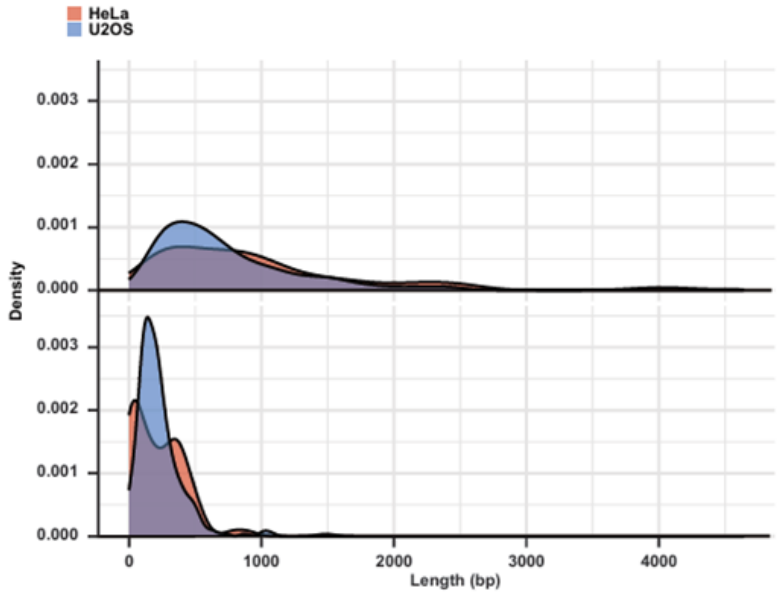


Chapter 2 Figure 3. TERRA enrichment in U2OS and HeLa cells

Genome browser screenshots showing TERRA enrichment in U2OS and HeLa cells, displaying a similar pattern between the two cell lines. This data was published in 2025 in NAR [19]

doi:10.6342/NTU202504254

A



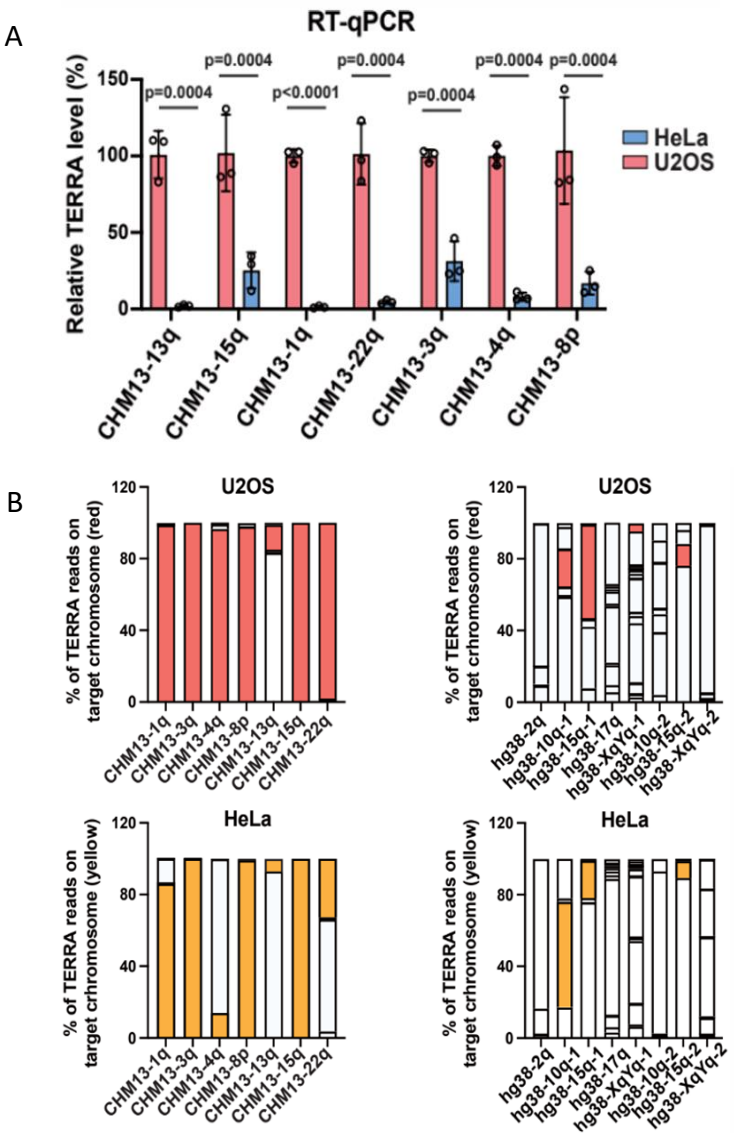
B

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

Chapter 2 Figure 4. TERRA length distribution from nanopore reads in U2OS and HeLa cells.

(A) Density plots showing the distribution of TERRA bulk read lengths and telomeric repeat lengths in TERRA reads from U2OS and HeLa cells. Nanopore reads mapped to T2T-CHM13 TERRA transcription regions were used for analysis. **(B)** Statistical analysis of TERRA bulk read lengths and telomeric repeat lengths in TERRA reads. This data was published in 2025 in NAR [19]

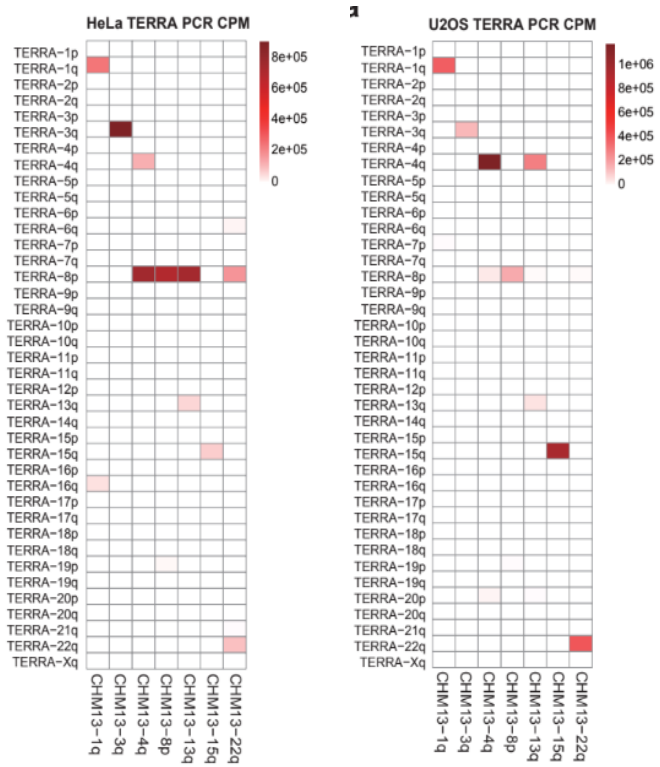
doi:10.6342/NTU202504254



Chapter 2 Figure 5. Validation of TERRA RT-PCR products

(A) RT-qPCR analysis of TERRA levels in HeLa and U2OS cells. Subtelomeric primers specific to individual chromosome ends were designed based on the T2T-CHM13 genome. Data represent mean ± SD from three biological replicates. P-values were calculated using two-tailed Student's t-test. This data was published in 2025 in NAR [19] **(B)** Proportion of TERRA reads mapped to target chromosome ends relative to all subtelomeric reads within TERRA transcription regions. Colored bars indicate the percentage of TERRA reads on each target chromosome end, while empty bars represent reads mapped to non-target chromosome ends.

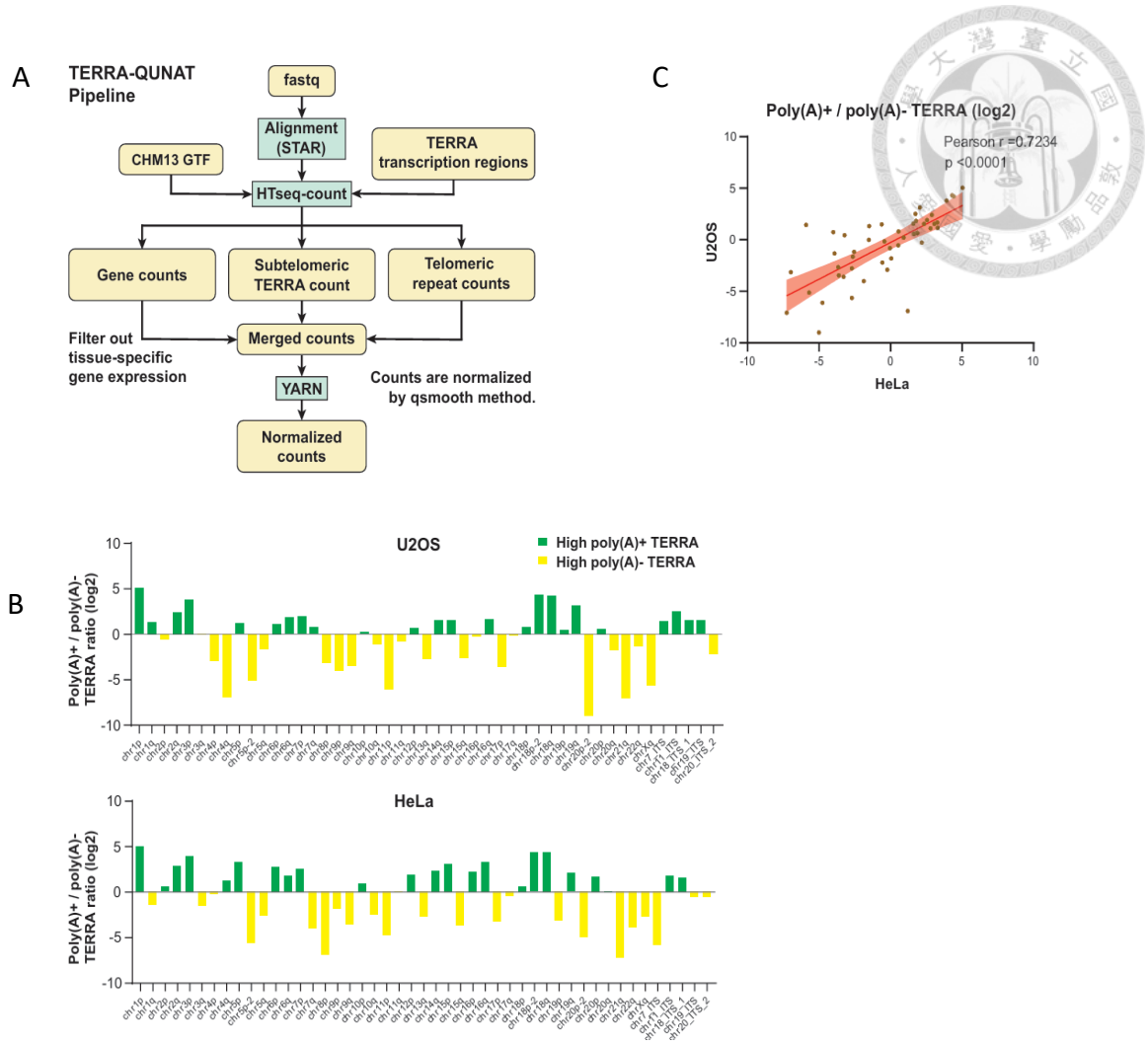
doi:10.6342/NTU202504254



Chapter 2 Figure 6. NGS sequencing to validate TERRA RT-PCR products

Heatmaps showing TERRA counts per million (CPM) mapped to TERRA transcription regions at specific chromosome ends in U2OS and HeLa cells.

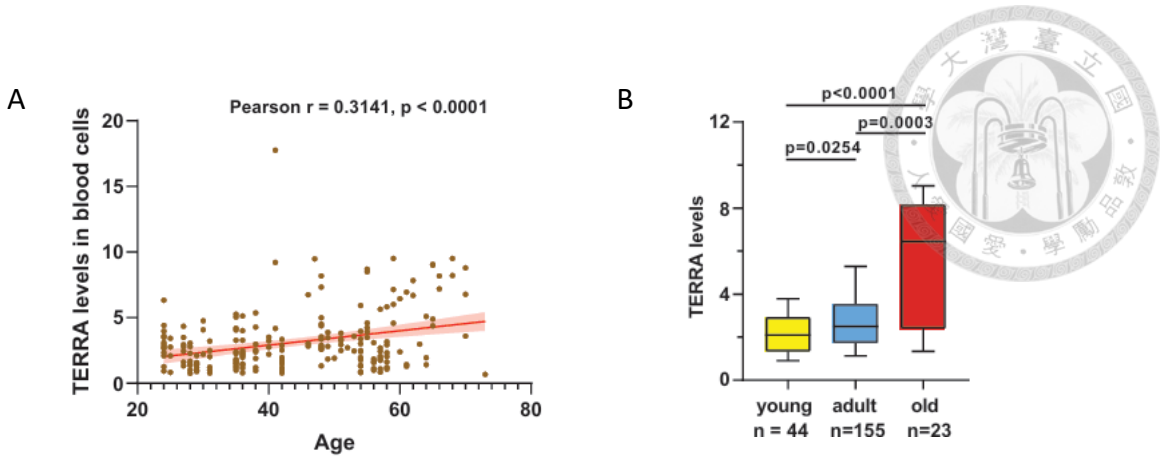
doi:10.6342/NTU202504254



Chapter 2 Figure 7. Polyadenylation profiles of TERRA across chromosome ends

(A) Workflow of the bioinformatics pipeline for TERRA expression quantification. Reads mapped to TERRA transcription regions were counted, and normalization across different tissues was performed using the YARN package. **(B)** Log_2 ratios of poly(A)+ to poly(A)− TERRA reads mapped to individual chromosome ends and interstitial telomeric sequences (ITSs) in U2OS and HeLa cells. **(C)** Scatter plots showing the correlation between U2OS and HeLa cells in poly(A)+/poly(A)− TERRA ratios. This data was published in 2025 in NAR [19]

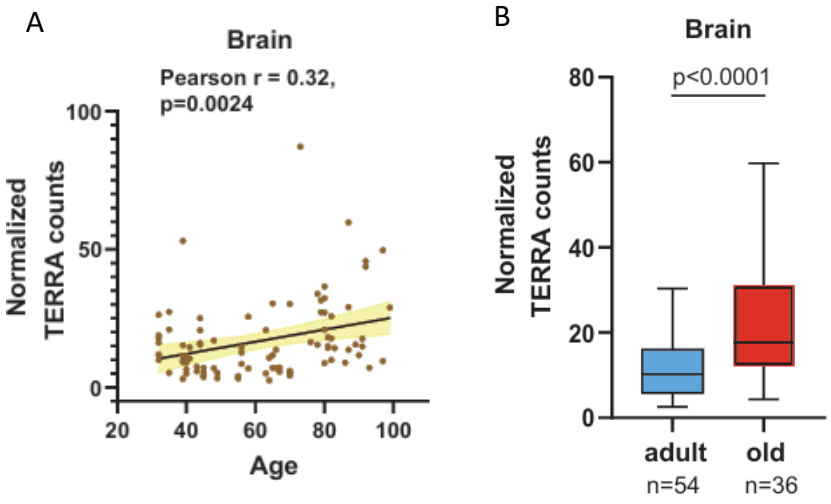
doi:10.6342/NTU202504254



Chapter 2 Figure 8. TERRA expression increases with age in human blood cells

(A) Scatter plot showing the correlation between TERRA levels and age in blood cells; each dot represents total normalized TERRA reads from all chromosome ends per individual. P-values were calculated using Pearson's correlation **(B)** Normalized TERRA counts in blood cells grouped by age: young (<30 years), adult (30–59 years), and old (≥60 years). Bars represent median with interquartile range. P-values were calculated using the Mann–Whitney U test. This data was published in 2025 in NAR [19]

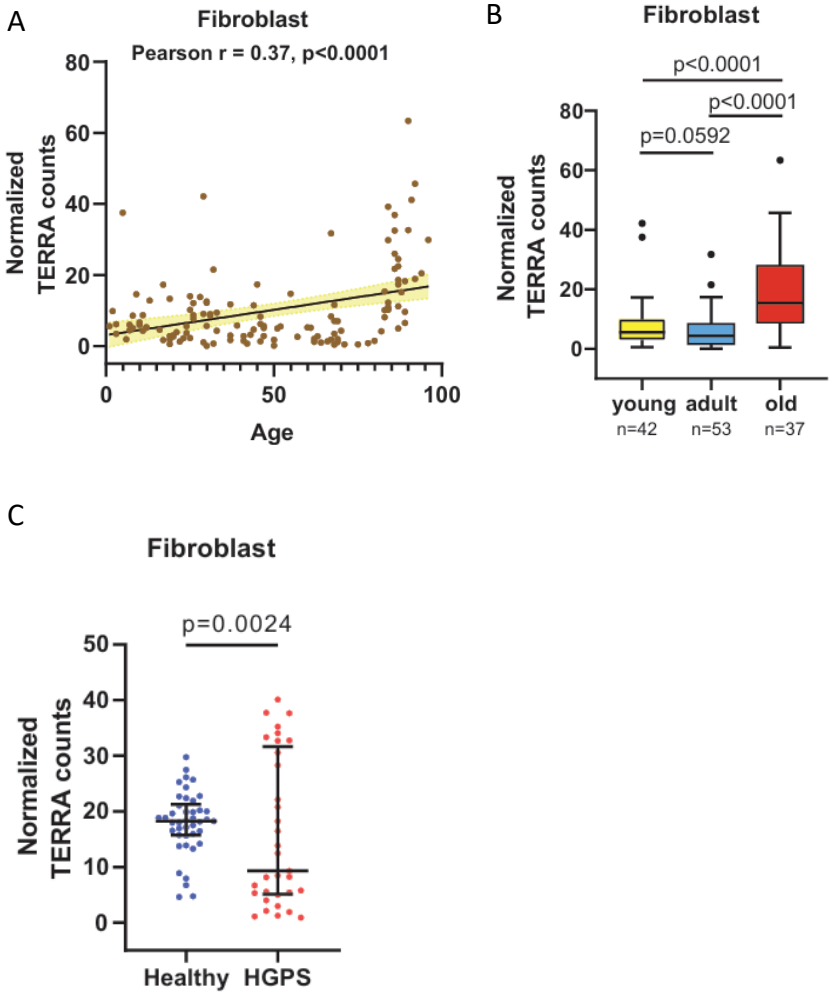
doi:10.6342/NTU202504254



Chapter 2 Figure 9. TERRA expression increases with age in human brain cells

(A) Scatter plot showing the correlation between TERRA levels and age in brain cells. P-values were calculated using Pearson's correlation **(B)** Normalized TERRA counts in brain cells grouped by age. P-values were calculated using the Mann–Whitney U test. This data was published in 2025 in NAR [19]

doi:10.6342/NTU202504254



Chapter 2 Figure 10. TERRA expression increases with age in human fibroblast cells and abnormal TERRA profiles in Hutchinson–Gilford progeria syndrome patients

(A) Scatter plot showing the correlation between TERRA levels and age in fibroblasts cells. P-values were calculated using Pearson's correlation **(B)** Normalized TERRA counts in fibroblasts cells grouped by age. P-values were calculated using the Mann–Whitney U test. **(C)** TERRA expression in fibroblasts from HGPS patients. Each dot represents the TERRA level of an individual. Bars indicate median with interquartile range. P-values were calculated using the Kolmogorov–Smirnov test. This data was published in 2025 in NAR [19]

doi:10.6342/NTU202504254

Primer ID	Forward	Reverse
CHM13-1q	CCGGTTGGTGCCCTGAATAA	GCCTTGGGAGAATCTCGGTG
CHM13-3q	GCGCAGTCATTCTCAACACC	ACAGGTGAACCCCTGGAGAT
CHM13-4q	AAAGCGGGAAACGAAAAGCC	GCTGCATTTGAGTTGCGACA
CHM13-8p	TGCTGCATTTGTGTTCCGAC	GGGAAACCAAAAGCCCCTCT
CHM13-13q	TGCCTTTGGGATAACTCGGG	GGTGCGCAGGATTCAAAGAG
CHM13-15q	AGGGGCTGCATTAAAGGGTC	TAGGACTGCTGCATTTGCGT
CHM13-22q	GGCTGCATGGACGGTGAATA	TTCGTTTCCCGCTTTCCACA

Chapter 2 Table 1. TERRA PCR primers sequence

doi:10.6342/NTU202504254

REFERENCES

1. de Lange T. How telomeres solve the end-protection problem. Nat Rev Genet 2009;10:299–305.
2. Baur JA, Zou Y, Shay JW, Wright WE. Telomere position effect in human cells. Science 2001;292:2075–2077.
3. Azzalin CM, Reichenbach P, Khoriauli L, et al. Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 2007;318:798–801.
4. Rudenko G, van der Ploeg LHT. Transcription of telomere repeats in trypanosomes. Cell 1989;57:351–361.
5. Vrbsky J, Akimcheva S, Watson JM, et al. siRNA-mediated methylation of Arabidopsis telomeres. EMBO J 2010;29:588–599.
6. Azzalin CM, Lingner J. Telomeric repeat-containing RNA: a link between telomeres and genome integrity. Cell Cycle 2008;7:1176–1181.
7. Nergadze SG, Farnung BO, et al. CpG-island subtelomeric promoters drive formation of TERRA transcripts. RNA 2009;15:1807–1815.
8. Feretzaki M, Pospisilova B, Lingner J. CTCF promotes telomere transcription and telomere protection. RNA 2019;25:805–819.
9. Graf M, Bielak-Zmijewska A, et al. Telomere R-loops and telomere fragility in ALT cancer cells. Cell 2017;170:1135–1147.
10. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles. Nat Struct Mol Biol 2015;22:17–25.
11. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres. Nat Cell Biol 2008;10:228–236.
12. Chawla R, Redon S, Raftopoulou C, et al. Telomerase RNA levels are regulated by telomere length in ALT cells. Nat Struct Mol Biol 2011;18:85–90.
13. Chu HP, Cifuentes-Rojas C, Kesner B, et al. TERRA RNA antagonizes ATRX and modulates heterochromatin. Nat Commun 2017;8:15466.
14. Clynes D, Higgs DR, Gibbons RJ. ATRX and TERRA interactions at repetitive chromatin. Nat Commun 2015;6:7623.

doi:10.6342/NTU202504254

15. Savoca V, Patsenker J, et al. RALY binds and stabilizes TERRA lncRNA. Cell Rep 2023;42:112230.
16. Wanat JJ, Lamarche BJ, et al. TERRA overexpression accelerates senescence in telomerase-null yeast. PLoS One 2018;13:e0203779.
17. Aguado J, Sola P, Jimenez-Gomez DE, et al. Telomere dysfunction induces TERRA up-regulation in Hutchinson–Gilford progeria. Nat Commun 2019;10:5281.
18. Tsai RX, Fang KC, Yang PC, Hsieh YH, Chiang IT, Chen Y, et al. TERRA regulates DNA G-quadruplex formation and ATRX recruitment to chromatin. *Nucleic Acids Research.* 2022 Nov 28;50(21):12217-12234.
19. Hsieh YH, Tai CH, Yeh MT, Chen YC, Yang PC, Yen CP, *et al.* Telomeric repeat-containing RNA increases in aged human cells. *Nucleic Acids Research.* 2025 Jul 8;53(13):gkaf597.

Appendix

Newly Published Paper Related to This Thesis: Yu-Hung Hsieh, Chin-Hua Tai, Meng-Ting Yeh, Yu-Chen Chen, Po-Cheng Yang, Chien-Ping Yen, Hong-Jhih Shen, Chan-Hsien Yeh, Hung-Chih Kuo, Der-Sheng Han, Hsueh-Ping Catherine Chu, Telomeric repeat-containing RNA increases in aged human cells, *Nucleic Acids Research*, Volume 53, Issue 13, 22 July 2025, gkaf597

doi:10.6342/NTU202504254

Nucleic Acids Research, 2025, **53**, gkaf597
https://doi.org/10.1093/nar/gkaf597
Genomics

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,*], Hsueh-Ping Catherine Chu[1,*]

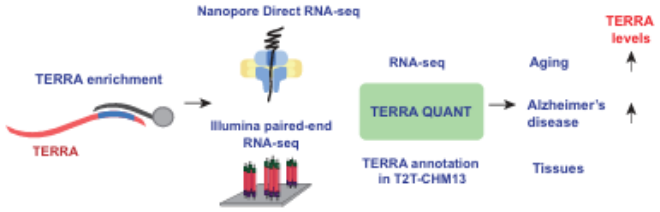
[1]Institute of Molecular and Cellular Biology, National Taiwan University, No. 1 Sec. 4 Roosevelt Road, Taipei 10617, Taiwan
[2]Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei 10051, Taiwan
[3]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei 11529, Taiwan
[4]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Bei-Hu Branch, Taipei 10051, Taiwan
[5]Department of Physical Medicine and Rehabilitation, College of Medicine, National Taiwan University, Taipei 10051, Taiwan

[*]To whom correspondence should be addressed. Email: cchu2017@ntu.edu.tw
Correspondence may also be addressed to Der-Sheng Han. Email: dshan1121@yahoo.com.tw
[†]The first three authors should be regarded as Joint First Authors.

Abstract

Telomeric repeat-containing RNA (TERRA), transcribed from subtelomeric regions toward 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 reverse transcription quantitative polymerase chain reaction. 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.

Graphical abstract



Introduction

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

Received: November 14, 2024. **Revised:** June 1, 2025. **Editorial Decision:** June 3, 2025. **Accepted:** June 12, 2025

doi:10.6342/NTU202504254

TERRA exhibits binding capabilities with telomeric DNA binding proteins [22–24] and non-telomeric DNA binding proteins such as chromatin modifiers [24]. It binds to both telomeric and non-telomeric chromatin, regulating telomere integrity and gene expression *in cis* and *in trans* [23–25]. For instance, TERRA interacts with ATRX, a protein involved in histone H3.3 recruitment and H3K9me3 formation [26], and prevents ATRX from binding to chromatin [24, 27]. TERRA depletion leads to increased ATRX recruitment to repetitive sequences such as rDNA, retrotransposons, subtelomeric sequences, and telomeric repeats, as well as the regions close to transcription start sites [24, 27]. Furthermore, TERRA plays a crucial role in modulating DNA G-quadruplex (G4) structures around the transcription start sites, suggesting a function of TERRA beyond telomere biology [27].

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

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

Materials and methods

Cell culture

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

TERRA capture for Illumina RNA-seq

To enrich TERRA, total RNA was extracted from 1×10^7 cells using the TRIzol™ Reagent (Invitrogen, Cat# 15596018) following the manufacturer's protocol. The total RNA was treated with DNase I (Invitrogen, AM2238) and supplied with ribonucleoside vanadyl complexes (VRC, New England Biolabs, E7760S) at 37°C for 15 min to eliminate DNA (0.4 unit/μl DNase I and 10 mM VRC). The enzyme was inactivated by 5 mM of EDTA. The DNase I-treated RNA sample was then subjected to a second round of purification using TRIzol™ reagent. For one TERRA-capture experiment, a total of 20 μg of the DNase I-treated RNA was used. The RNA sample was dissolved in 90 μl of 6× SSC buffer (0.9 M NaCl, 0.9 M sodium citrate). The RNA sample and 10 μl of 1 μM biotinylated TERRA antisense probe [5′-(CCCATT)$_5$-3′BioTEG] were individually denatured at 70°C for 2 min, and then were mixed together (final: 20 μg RNA, 2× SSC buffer, 100 nM probe) for an additional 8 min at 70°C. The RNA-probe annealing was carried out by gradually decreasing the temperature from 70°C to 44°C, followed by incubation at 44°C for an additional 30 min. To capture TERRA RNA, 100 μl of streptavidin beads (MyOneC1, Invitrogen, Cat# 65002) were prepared and mixed with the sample. The mixture was then incubated on a rotating wheel at 37°C for 15 min. The supernatant was separated from beads by a magnet and carefully removed. Beads were washed four times with 2× SSC buffer containing 0.1% NP-40 (0.3 M NaCl, 0.3 M sodium citrate, 0.1% NP-40) at 37°C, followed by two washes with 1× SSC buffer containing 0.1% NP40 (0.15 M NaCl, 0.15 M sodium citrate, 0.1% NP-40), once at 37°C and once at room temperature. Finally, beads were washed once with 1× SSC buffer without NP40 (0.15 M NaCl, 0.15 M sodium citrate) at 37°C. Washed beads were resuspended in 100 μl of nuclease-free water and incubated at 70°C to elute RNA. The elution step was repeated twice. The eluted RNA sample was concentrated by precipitating in 3× volume of 100% ethanol (300 μl), containing 0.1× volume of 3 M NaOAc (40 μl, Thermo Fisher Scientific, AM9740) and 2 μl of GlycoBlue™ (Invitrogen, AM9515) overnight at −30°C. The RNA was precipitated by centrifugation at 4°C and 21 100 × *g*. The RNA pellet was washed twice with 70% EtOH and eluted in 30 μl of nuclease-free water. RNA concentration was measured using Qubit RNA HS kit (Invitrogen, Q32852). The captured RNA was converted into complementary DNA (cDNA), followed by paired-end Illumina sequencing and RT-qPCR (reverse transcription quantitative polymerase chain reaction).

Illumina RNA-seq library preparation

The cDNA library was constructed using the NEBNext® Ultra II Directional RNA Library Prep Kit (New England Biolabs, E7760S) according to the manufacturer's instruction with some modifications. The first-strand synthesis step was carried out by SuperScript IV (Thermo Fisher Scientific, Cat# 18090010). The TERRA-captured RNA sample (30 μl) was divided into two equal halves for the reverse transcription reaction. Half of the TERRA-captured sample was reverse transcribed using 2 μM random hexamers (Thermo Fisher Scientific, 48190011) and the other half using 1 μM of telom-

doi:10.6342/NTU202504254

eric C-rich primer 5′-(CCCTAA)$_5$-3′. These two groups are referred to as the random and specific groups, respectively, followed by reverse transcription, second-strand synthesis, end repair, adaptor ligation, and size selection. Libraries were amplified using NEBNext Ultra II Q5 Master Mix (New England Biolabs, #M0544L) with optimal cycles determined by qPCR, and size-selected twice using AMPure XP beads. Finally, the DNA was eluted in 15 µl of 0.1× TE buffer and stored at −20°C. The library size distribution was assessed using the Agilent 2100 Bioanalyzer. Each library was quantified using the NEBNext Library Quant Kit (New England Biolabs, #E7630S). The pooled libraries (fragment size 300–500 bp) were subjected to paired-end 150 bp sequencing on Illumina MiSeq or NovaSeq platforms.

Reverse transcription and quantitative PCR

TERRA enrichment was assayed by RT-qPCR using cDNA. The reverse transcription was performed using random hexamers (Thermo Fisher Scientific, #48190011) and SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, Cat# 18090200) (final: 1 µM random hexamer, 400 µM dNTP, 1× SuperScript IV buffer, 8 units/µl SuperScript IV reverse transcriptase, 4 mM DTT, 1.6 units/µl RNaseOUT). The qPCR analysis was conducted using a 1:5 dilution of TERRA-enriched cDNA and a serial dilution of control cDNA, ranging from 1× to 1000× dilution (1% to 0.001% of input), to generate a standard curve. Subtelomeric primer sequences (hg38-2q, CHM13-1q, CHM13-3q, CHM13-8p, and CHM13-15q) are listed in Supplementary Table S6. The qPCR reaction was performed with IQ SYBR Green Supermix (Bio-Rad, Cat# 170-8882) using the CFX Real-Time PCR Detection System. The thermal cycling conditions were as follows: initial heat activation of polymerase at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 15 s, annealing and extension at 58°C for 1 min. Data analysis was carried out using CFX Maestro software v1.1 (Bio-Rad). For Illumina samples, the standard curve was generated from a series of dilutions from the input. TERRA RNA abundance was normalized to GAPDH. TERRA-captured samples were compared to the control sample (no capture) to calculate TERRA enrichment. For Illumina samples, TERRA enrichment ratio was calculated using the ddCt method, where the Ct value of TERRA was normalized to that of GAPDH.

TERRA capture for Nanopore direct RNA-seq

We harvested cells from 200 of 150-mm culture dishes for RNA extraction. The TERRA-capture procedure followed the method described previously with a few modifications. First, the RNA samples were not treated with DNase I. Second, the input RNA amount for each capture was increased to 500 µg. Third, to achieve the required RNA quantity for Nanopore direct RNA-seq, 10 captures were pooled. RNA was concentrated by precipitation with 3× volume of 100% ethanol, 1/10× volume of 3 M NaOAc (Thermo Fisher Scientific, AM9740), and 2 µl of GlycoBlue™ (Invitrogen, AM9515) for 3 days at −20°C. One hundred captures (a total of 50 mg input RNA) were required to acquire a sufficient amount of TERRA-enriched RNA for sequencing. Subsequently, RNA was precipitated by centrifugation at 4°C at 21 100 × *g* for 20 min. The pellet was washed with 70% ethanol and eluted in 20 µl of nuclease-free water. Finally, a total of 3.4 µg RNA was obtained after capture. Next, 1.5 µg of TERRA-enriched RNA was polyadenylated using *E. coli* poly(A) polymerase (New England Biolabs, #M0276S) at a final concentration of 0.25 U/µl in 1× poly(A) polymerase reaction buffer with 1 mM dATP at 37°C for 30 min. To prevent RNA degradation, the reaction included 2 units/µl RNaseOUT (Thermo Fisher Scientific, Cat# 10777019). The polyadenylation reaction was terminated by adding 10 mM EDTA. Polyadenylated RNA was purified using 2× sample volume of AMPure XP RNA beads (Beckman Coulter, A66514), following the manufacturer's protocol. RNA was eluted in 20 µl of nuclease-free water. RNA concentration was measured using Qubit RNA high-sensitivity assay (Thermo Fisher Scientific, Q32852). The fragment size was assessed using Bioanalyzer (Agilent RNA 6000 Pico Kit). TERRA-enriched RNA was subjected to Nanopore direct RNA sequencing (Cat# SQK-RNA002) with an R9.4.1 flow cell (FLO-MIN106).

Analysis of TERRA-capture Illumina RNA-seq

The fastq files were subjected to a quality filter ($Q > 20$), and Illumina adaptor sequences were trimmed using Trimgalore v0.6.7 (trimgalore –illumina –pair Read1.fastq Read2.fastq -o /path/to/output). The reads were aligned to the T2T-CHM13 genome using the STAR aligner v2.7.9a (star –runMode alignReads –genomeDir /path/to/T2T-CHM13 –outSAMtype BAM SortedByCoordinate –outFileNamePrefix output_path/id –readFilesIn trimmed_Read1.fastq trimmed_Read2.fastq). The aligned BAM files were indexed by samtools v1.17 (samtools index -b exp.bam -o exp.bam.bai) and converted to coverage files in bigwig format by deepTools v3.3.1 bamCoverage, with RPKM as the normalization method (bamCoverage -b BAM_files –normalizeUsing RPKM –binSize 30 -of bigwig -o output.bw). Furthermore, a comparison was performed between TERRA-captured groups and no-captured RNA-seq data (control) using $\log_2$ ratio analysis by bigwigCompare (deepTools). The coverage files (bigwig) were visualized on the Integrative Genomics Viewer (IGV) v2.16.0. TERRA transcription regions were manually assigned according to TERRA enrichment, Nanopore reads, and CAGE (cap analysis of gene expression) tags at chromosome ends of the T2T-CHM13 genome. The BED and GTF files of TERRA transcription regions were used for TERRA quantification. To quantify TERRA read counts, the alignments were extracted from the BAM files using samtools v1.17 (samtools view -b -h -M -L TERRA_region.bed –fetch-pairs input.bam -o TERRA_region_alignment.bam), and filtered with MAPQ $\geq$ 30.

Analysis of TERRA Nanopore direct RNA-seq

Fast5 files were processed using Guppy base caller (guppy_GPUv5.0.7) with a default quality filter of $Q > 7$ (guppy -i /path/to/fast5 -c rna_r9.4.1_70bps_hac.cfg -s /path/to/output_fastq). After base calling, reads were aligned to CHM13 reference genome using the minimap2 aligner v2.22 (minimap2 -ax splice -uf -k14 /path/to/CHM13.mmi TERRA_Nanopore.fastq > TERRA_Nanopore_aligned.sam). The alignment was converted to BAM format and sorted using samtools v1.17 (samtools view -b TERRA_Nanopore_aligned.sam -o TERRA_Nanopore_aligned.bam; sam-

doi:10.6342/NTU202504254

tools sort TERRA_Nanopore_aligned.bam -o TERRA_Nanopore_aligned.sorted.bam). A bam index was created by samtools v1.17 (samtools index TERRA_Nanopore_aligned.sorted.bam -o TERRA_Nanopore_aligned.sorted.bam.bai). The aligned reads displayed on the genome browser were manually examined to assist in defining TERRA transcription regions. Reads aligned to TERRA transcription regions were selected using samtools v1.17 with the TERRA transcription region bed file (samtools view -b -h -M -L TERRA_region.bed TERRA_Nanopore_aligned.sorted.bam -o Nanopore_TERRA_regions_aligned.bam), and the counts of each TERRA region were quantified. For the analysis of TERRA from individual chromosome ends (Fig. 1D), the unique reads were filtered based on MAPQ ($\geq$1), FLAG values (FLAG = 0, or 16 for primary alignments, FLAG = 2048, or 2064 for supplementary alignments) provided by minimap2 (Supplementary Table S5.1). The supplementary alignments with the same read IDs were removed (Supplementary Table S5.2).

Analysis of ChIP-seq datasets

The ChIP-seq datasets are listed as the follows: U2OS H3K4me3 (GSE114703) [38], U2OS RNA polymerase II (GSE87324) [39], U2OS CTCF (GSE87831) [40], U2OS MeDIP-seq (GSE81165) [41], U2OS DRIP-seq (GSE115957) [42], and HeLa CAGE-seq (GSE121351). Additionally, the annotated CpG island BED file was downloaded from UCSC (assembly: T2T-CHM13; group: Expression and Regulation; track: CpG Islands; table: hub_3671779_cpgIslandExtUnmasked). Reads with quality scores below Q30 were filtered out, and adaptor sequences were trimmed using TrimGalore v0.6.7 (trimgalore –illumina –pair -q 30 Read1.fastq Read2.fastq -o /path/to/output). The processed reads were aligned to the CHM13 genome using bowtie2 with default settings. The resulting Sequence Alignment Map (SAM) files were converted into Binary Alignment Map (BAM) format, sorted, duplicate reads were removed, and the files were indexed using samtools v1.17 (samtools view -b exp.sam -o exp.bam, samtools sort exp.bam -o exp.sorted.bam, samtools rmdup exp.sorted.bam -o exp.sorted.rmdup.bam, samtools index exp.sorted.rmdup.bam -o exp.sorted.rmdup.bam.bai). ChIP-seq coverage files (bigwig) were generated using deepTools v3.3.1 bamCompare, normalizing using CPM method and subtracting to input signal (bamCompare -b1 ChIP_BAM_file -b2 Input_BAM_file –normalizeUsing CPM –binSize 30 –operation subtract -of bigwig -o output.bw). Meta-analysis surrounding the 61–29-37 nt repeat regions was performed using deepTools v3.3.1 computeMatrix and plotHeatmap (computeMatrix reference-point –referencePoint center -a 500 -b 500 -S ChIP-seq_coverage.bw -R 61_29_37_region.bed -o Matrix.gz, plotHeatmap -m Matrix.gz –colorMap bwr -o output_meta-analysis.pdf). For visualization on CHM13 genome, the bigwig files were examined in IGV.

Analysis of TERRA expression by TERRA-QUANT

RNA-seq datasets were downloaded and pre-processed using the previously described methods (SRAToolkit and TrimGalore commands). The processed reads were aligned to the T2T-CHM13v1.1 genome using STAR aligner with default settings. Annotated TERRA transcription regions were divided into subtelomeric and telomeric repeat portions and incorporated into the CHM13 GTF file. Read counts for both human genes and TERRA were calculated using HTseq-count (htseq -f bam -s no/yes/reverse -t transcript –idattr gene_name -m intersection-nonempty –nonunique all BAM_file CHM13_GTF_with_TERRA_regions > output_count.txt). For quantifying TERRA from individual chromosome ends, reads were extracted with MAPQ $\geq$ 30 for paired-end reads, or with MAPQ = 255 for single-end reads by samtools. Only reads mapped to the subtelomeres within defined TERRA transcription regions were calculated for chromosome-arm-specific TERRA. For total TERRA quantification, reads were aligned by STAR with default settings, and duplicates were removed and filtered with primary alignments (MAPQ $\geq$ 1) by samtools. The reads mapped to TERRA regions and all human genes were then analyzed together and normalized using the YARN package [43] in the R language. The total TERRA levels in human tissues were calculated based on reads mapped to all TERRA transcription regions with proximal TSSs located at chromosome ends (Type I and Type II, but not Type III), including both subtelomeric regions and pure telomeric repeats. "TERRA levels" refer to normalized TERRA read counts, representing the relative expression of TERRA transcripts compared to a control gene set across all samples. Counts for other genes were obtained using HTseq-count, and all counts were merged for downstream normalization. For normalization, gene counts—including TERRA and other genes quantified using HTseq-count—were processed using the YARN R package, which supports multi-sample normalization across diverse tissues and cell types.

NGS sequencing for TERRA RT-qPCR products

RNA was extracted using TRIzol™ Reagent (Invitrogen, Cat# 15596018). Reverse transcription was performed using the SuperScript™ IV system (Thermo Fisher Scientific, Cat# 18090200) with random hexamers (2.5 nM). The RT-qPCR reaction was conducted by using IQ SYBR Green Supermix (Bio-Rad, Cat# 170-8882). The design of the primer sequences was based on the CHM13 genome (Supplementary Table S6). TERRA RT-qPCR products were purified using Gel/PCR Purification Mini Kit (FAVORGEN, Cat# FAGCK 001-1) before DNA library preparation. Eluted DNA was subjected to library construction using NEBNext Ultra II DNA Library Prep Kit (NEB, Cat# E7645S). Paired-end 150 bp reads were obtained from an Illumina NovaSeq system. The sequencing reads of TERRA RT-qPCR products underwent quality filtration (Q > 30) and adapters pruning using TrimGalore (v0.6.3) (https://github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to the CHM13 genome using STAR aligner, filtering with MAPQ $\geq$ 30 by samtools. SAM files were transformed to BAM files, and were deduplicated using samtools (v.1.18). Bigwig files were generated by DeepTools (v.3.3.1) bamCoverage (bamCoverage –normalizeUsing CPM, –centerReads, and –extendReads 150). To assess the proportion of PCR products in the TERRA transcription regions, we calculated the read counts of PCR products on TERRA transcription regions using samtools (v.1.13) view [samtools view –fetch-pairs, -h, -b, -M (–use-index), and

doi:10.6342/NTU202504254

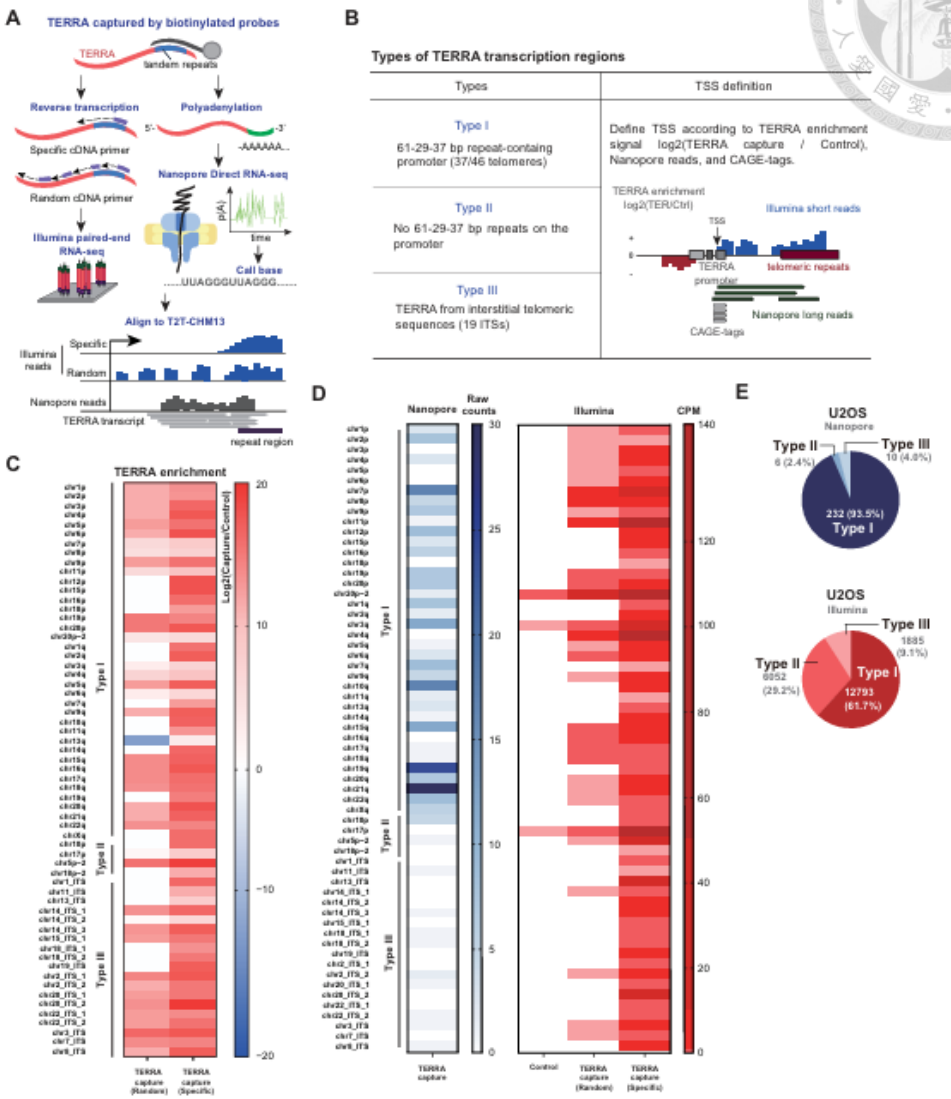


Figure 1. Identification of TERRA transcription regions by short-read and long-read RNA-seq in human cells. (**A**) The flowchart of TERRA-capture RNA-seq. TERRA was captured by using biotinylated antisense oligos. Captured TERRA RNAs were subjected to Illumina or Oxford Nanopore long-read RNA direct sequencing. (**B**) Definition of three types of TERRA transcription regions. TERRA transcription start sites were determined by TERRA enrichment from Illumina short-reads, Nanopore long reads, and CAGE-seq. (**C**) Heatmap represents TERRA enrichment (log_2 values) at different chromosome ends and interstitial telomeric sequences (ITSs) based on Illumina read counts of TERRA transcripts in U2OS cells. cDNA was synthesized using random or telomeric specific primers for library contruction. Capture: TERRA capture RNA-seq. Control: regular RNA-seq with ribosomal RNA depletion. (**D**) Heatmap represents TERRA read counts from different chromosome ends and ITSs in U2OS cells. Reads from Illumina or Nanopore sequencing were mapped to T2T-CHM13. CPM: counts per million mapped reads. (**E**) Pie charts show the numbers of Type I, II, and III TERRA reads from Nanopore or Illumina sequencing in U2OS cells.

doi:10.6342/NTU202504254

-L (–target-file)] with a bed file including TERRA transcription regions.

Poly(A)+ and poly(A)− TERRA capture sequencing and analysis

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

Research participants for collecting blood cells

Blood samples were collected as previously described [44]. The study was approved by the institutional review board (IRB No. 201601091RIND) of the National Taiwan University Hospital. All participants were provided with written informed consent before participation in the study. Trial registration: ClinicalTrials.gov: NCT02779088. Registered 20 May 2016, https://clinicaltrials.gov/ct2/keydates/NCT02779088.

Telomere length measurement in blood cells

Blood samples were collected from the participants as previously described [44]. Briefly, genomic DNA was extracted from frozen human buffy coats using the QIAamp® DNA Mini Kit (QIAGEN, Hilden, Germany, #Cat 51306). Quantitative analysis was performed by quantitative polymerase chain reaction (qPCR). The relative telomere length for each participant was calculated using the telomere-to-single copy gene (T/S) ratio, defined as the number of telomeric repeats (T) divided by a standard reference DNA (S). The *36B4* gene was used as the single-copy DNA. A standard curve was generated from a series of dilutions of genomic DNA from HeLa cells, and was used for calibration for each qPCR plate. Primers for telomeric DNA and *36B4*: *36B4* forward primer sequence: 5′-CAGCAAGTGGGAAGGTGTAATCC-3′ and *36B4* reverse primer sequence: 5′-CCCATTCTATCATCAACGGGTACAA-3′; telomeric DNA forward primer sequence: 5′-GGTTTTTGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT-3′ and telomeric DNA reverse primer sequence: 5′-TCCCGACTATCCCTATCCCTATCCCTATCCCTATCCCTA-3′.

Quantification of TERRA by RT-qPCR in blood samples

RNA was extracted from buffy coat samples using TRIzol (Thermo Fisher Scientific, Cat# 15596018) and purified by acid–phenol:chloroform (Thermo Fisher Scientific, Cat# AM9722). RNA was treated with ezDNase™ Enzyme (Thermo Fisher Scientific) to eliminate contaminating genomic DNA. Reverse transcription was carried out using the SuperScript™ IV system (Thermo Fisher Scientific, Cat# 18090200). Random (2.5 nM) hexamers combined with telomeric repeat primers (0.25 nM) to enrich TERRA cDNA. The protocol for total TERRA qPCR was similar to that for the telomere qPCR using telomeric repeat primers. Beta-2 microglobulin (*B2M*) was used as the reference gene. B2M primers: forward primer sequence, 5′-CTATCCAGCGTACTCCAAAG-3′; reverse primer sequence, 5′-GAAAGACCAGTCCTTGCTGA-3′. TERRA RNA extracted from U2OS cells was used to obtain a standard curve for calibration between samples from different qPCR plates. The relative TERRA expression level was analyzed using the TERRA RNA/B2M RNA ratio. Subtelomeric primers (hg38-2q) for TERRA RT-qPCR in blood cells are listed in Supplementary Table S6. Data points from samples with poor RNA quality—assessed by RT-qPCR analysis of the B2M control were excluded from the analysis.

Analysis of TERRA expression across different tissues

Publicly available RNA-seq datasets were gathered from NCBI GEO datasets [32, 45, 46]. Sample sources are documented in Supplementary Table S7. To compare TERRA levels across different tissues or cell lines, we employed the YARN R package [43]. The whole gene counts derived from all RNA-seq datasets were merged in the same expression table for the following process. The filtration of low-expressed genes was performed based on an expression threshold (CPM < 1). TERRA counts for each sample were determined using TERRA-QUANT as described previously. Normalization of all data was performed using the qsmooth method [47].

Single-cell and single-nucleus RNA-seq analyses

Neuronal single-cell RNA-seq datasets (GSE186698) and PBMC single-cell RNA-seq datasets (GSE157007) were downloaded as previously described (Supplementary Table S8). Alignment, barcode assignment, and unique molecular identifier (UMI) counting were performed using CellRanger v7.2.0 [48] with a default setting (MAPQ = 255). Downstream analysis of CellRanger outputs was performed using Seurat v4.4.0 [49]. The low-quality cells were filtered out if the number of genes detected or mitochondria gene reads were three mean absolute deviations away from the average. Genes expressed in fewer than three cells are ex-

doi:10.6342/NTU202504254

cluded. Two thousand highly variable genes were identified using the variance stabilizing transformation method and were used for principal component analysis. The top 20 principal components were selected for cell clustering. The result was visualized with uniform manifold approximation and projection (UMAP). Data integration was performed by using FindIntegrationAnchors() and IntegrateData() functions according to the union of the top 2000 highly variable genes of each data. Plots of TERRA, ATRX, and MAP2 were generated by scCustomize v2.0.1 (Samuel Marsh, Maëlle Salmon, and Paul Hoffman, 2023, https://doi.org/10.5281/zenodo.10161832). In neuronal differentiation samples, cells positive for CD29, CD44, and CD105 but negative for CD14, CD34, and CD45 were classified as mesenchymal stem cells; cells expressing SOX2 and Nestin were identified as neuroepithelial cells, while those positive for MAP2 were categorized as mature neurons. Cells in PBMC were annotated by the expression of representative cell markers: $CD4^+$ T cells (CD4, IL7R, and CD3G), NK cells (NCAM1 and NKG7), $CD8^+$ T cells (CD8A and CD8B), B cells (CD19 and MS4A1), monocytes (CD14 and CD16), and dendritic cells (IL3RA). Single-nucleus RNA-seq datasets were obtained from the Rush Alzheimer's Disease Center (RADC) Research Resource Sharing Hub at Synapse (https://www.synapse.org/#!Synapse:syn18485175) with DOI: 10.7303/syn18485175. The ROSMAP metadata can be found at https://www.synapse.org/#!Synapse:syn3157322. Details of individual samples, including those with no pathology (13 individuals), AD-early pathology (8 individuals), and AD-late pathology (5 individuals), utilized in the analysis of this study are outlined in Supplementary Table S9. In Alzheimer's disease samples, we assigned cell types by marker genes: excitatory neurons (NRGN), inhibitory neurons (GAD1), astrocytes (AQP4), oligodendrocytes (MBP), microglia (CSF1R and CD14), and oligodendrocyte progenitor cells (VCAN).

Alzheimer's disease patient-derived iPSC differentiation to neurons

To generate cortical neurons, induced pluripotent stem cells (iPSCs) derived from an Alzheimer's disease (AD) patient carrying the P117L mutation in presenilin-1 (*PS1*)—the catalytic subunit of γ-secretase—and mutation-corrected isogenic controls were used. iPSCs were dissociated with dispase and cultured in suspension using human embryonic stem cell (hESC) medium lacking bFGF, but supplemented with 100 nM LDN193189 and 10 μM SB431542 to initiate neural induction. Cells were then transferred to N2 medium and cultured until neural rosette structures formed. These neural rosettes were manually isolated and further cultured in NI medium to promote neural sphere formation. The resulting neural spheres were plated onto Matrigel-coated six-well plates and cultured in ND medium to induce cortical neuron differentiation. The differentiated cortical neurons were subsequently collected for RT-qPCR analysis.

Results

Nanopore direct RNA sequencing for TERRA transcripts

To investigate TERRA transcripts in human cells, we captured RNA molecules containing UUAGGG repeats by using biotinylated antisense oligos in U2OS cells (Fig. 1A). The TERRA-enriched RNA samples were subjected to two different sequencing methods: Illumina paired-end sequencing and Nanopore direct RNA sequencing. For Illumina sequencing, TERRA-enriched RNA was reverse transcribed using either random hexamers (random group) or telomere-specific primers (specific group) for cDNA library construction. For Nanopore direct RNA sequencing, enriched TERRA molecules were polyadenylated by poly(A) polymerase before library preparation (Fig. 1A). Approximately, a few hundred- to over 2000-fold enrichment was observed after TERRA capture (Supplementary Fig. S1A). The reads from both methods were aligned to the T2T-CHM13 human genome reference, which includes nearly complete coverage of the entire genome including subtelomeric and telomeric regions [50].

TERRA transcription regions at chromosome ends

The criteria for identification of TERRA transcription regions included the observations of Nanopore reads, CAGE tags, and TERRA enrichment within 100 kb of chromosome ends (Fig. 1B and Supplementary Fig. S1B). Compared to ribosomal depleted RNA-seq (control), TERRA-capture RNA-seq (enriched by antisense oligos) exhibited an increase in read counts derived from chromosome ends in U2OS cells, shown by TERRA enrichment in $\log_2$ values from Illumina sequencing data (Fig. 1C). In cases where no CAGE-seq tags mapped to the subtelomeric regions, we estimated transcription start sites (TSSs) based on subtelomeric regions with Nanopore long reads and TERRA enrichment. We identified 39 chromosome ends that contain TERRA transcription (Supplementary Tables S1 and S2). Notably, chr5p, chr18p, and chr20p each potentially contain two TERRA transcription start sites (one proximal and one distal to telomere), while other chromosome ends have a single TERRA transcription start site located near the telomeric repeat tracts (Supplementary Fig. S1C). The distance of the identified TERRA TSS from the telomeric repeat tract per chromosome end is shown in Supplementary Fig. S2. These results demonstrate the presence of TERRA transcription across a majority of chromosome termini in U2OS cells (Fig. 1D), aligning with a recent study showing that TERRA is detectable at most chromosome ends in human cell lines [51].

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

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

We classified three distinct types of TERRA transcription regions according to their locations and promoters (Fig. 1B). The majority of telomeres (37 out of 46 chromosome ends) were characterized as Type I TERRA transcription regions by the presence of 61–29-37 bp repeats at the promoter (Supplementary Fig. S4). Only chr10p and chr17p termini were characterized as Type II TERRA transcription region, which lacks 61–29-37 bp repeats (Fig. 1B and

doi:10.6342/NTU202504254

Supplementary Fig. S4). TERRA transcription regions with 61–29-37 bp repeats or CAGE tags are summarized in Supplementary Tables S1 and S2. Not all Type I TERRA promoters encompass all three repetitive motifs. Sixteen chromosome ends are composed of all three repetitive motifs, while 21 chromosome ends lack either one or two repetitive motifs (Supplementary Fig. S5A and Supplementary Table S3). To investigate the impact of the 29 bp repeat element on TERRA expression, we compared TERRA levels at chromosome ends with or without the 29 bp repeats. Our analysis revealed no significant difference in TERRA expression between these regions (Supplementary Fig. S5B), implying the presence of an additional promoter that lacks the 29 bp repeats for TERRA transcription.

TERRA transcription from interstitial telomeric repeats

In addition to TERRA transcription at chromosome ends, TERRA can be transcribed from ITSs, classified as Type III TERRA transcription regions (Fig. 1B–D, Supplementary Fig. S4, and Supplementary Table S4). We quantified Type I, II, and III TERRA transcription and found that most TERRA transcripts in U2OS cells originate from chromosome ends (Type I + II) (Fig. 1E and Supplementary Fig. S3B). Type III accounts for only ∼4% of Nanopore and 9.1% of Illumina TERRA reads in U2OS but is more prominent in HeLa cells, comprising 46% of Nanopore and 17.7% of Illumina reads. In the T2T-CHM13 genome, we identified 79 ITS sites containing telomeric repeats longer than 200 base pairs (Supplementary Fig. S5C). Among these, 19 sites showed TERRA enrichment and Nanopore reads containing long telomeric repeats at the 3′ ends. Notably, Type III TERRA transcription regions exhibit CAGE tags, and the promoters usually comprise the 37 bp repeats without 29 bp repeats (Supplementary Fig. S4). These results indicate that most TERRA transcripts originated from the 37 bp repeat regions, regardless of whether TERRA derived from chromosome ends or ITS. Genome browser views demonstrated some examples of TERRA transcription regions carrying these features (Supplementary Fig. S4).

Interestingly, we observed antisense TERRA transcripts, named ARIA [52] in Nanopore direct RNA-seq data. ARIA reads were mapped to chromosome 2, a region containing an ancestral telomere–telomere fusion site [53]. The 5′ ends of ARIA reads comprise C-rich telomeric repeats near the junction of the telomere–telomere fusion site, following the unique sequences aligned to chromosome 2 (Supplementary Fig. S5D). TERRA transcripts were also observed near the junction but were not directly connected to ARIA transcripts (Supplementary Fig. S5D).

The length of telomeric repeats in TERRA transcripts

We sought to elucidate the length of telomeric repeat tracts within TERRA molecules by analyzing Nanopore sequencing reads that aligned to subtelomeric regions and extended into the telomeric repeats. Among TERRA reads mapping to chromosome ends, the maximum lengths are 2478 bp in U2OS cells and 4639 bp in HeLa cells (Fig. 2A and B). The mean lengths of bulk TERRA reads were 721 bp in U2OS cells and 982 bp in HeLa cells. For the pure telomeric tracts in TERRA reads, the maximum lengths observed were 1080 bp in U2OS cells and 1487 bp in HeLa cells, while the mean lengths were 222 and 284 bp in U2OS and HeLa cells, respectively. Overall, telomeric repeat tracts within TERRA reads ranged from several hundred nucleotides to over 1000 nucleotides. Dot plots of Nanopore TERRA reads per chromosome arm, grouped by types of TERRA transcription regions, revealed considerable variability in length of bulk TERRA reads and telomeric repeats across chromosomes in both U2OS and HeLa cells (Fig. 2C). The maximum bulk TERRA read lengths were longer in Type I compared to Type II and Type III in both cells. Mean telomeric repeat lengths in TERRA reads were slightly shorter in Type III in both cell lines. Given the fragility of RNA molecules, it is possible that some TERRA transcripts were partially degraded during sample preparation, potentially resulting in shorter observed lengths and increased variation in transcript size. Nevertheless, our findings are consistent with previous reports indicating that TERRA transcripts generally exceed 1 kb in length and that the average telomeric repeat length within TERRA is ∼200–300 bp [32, 54].

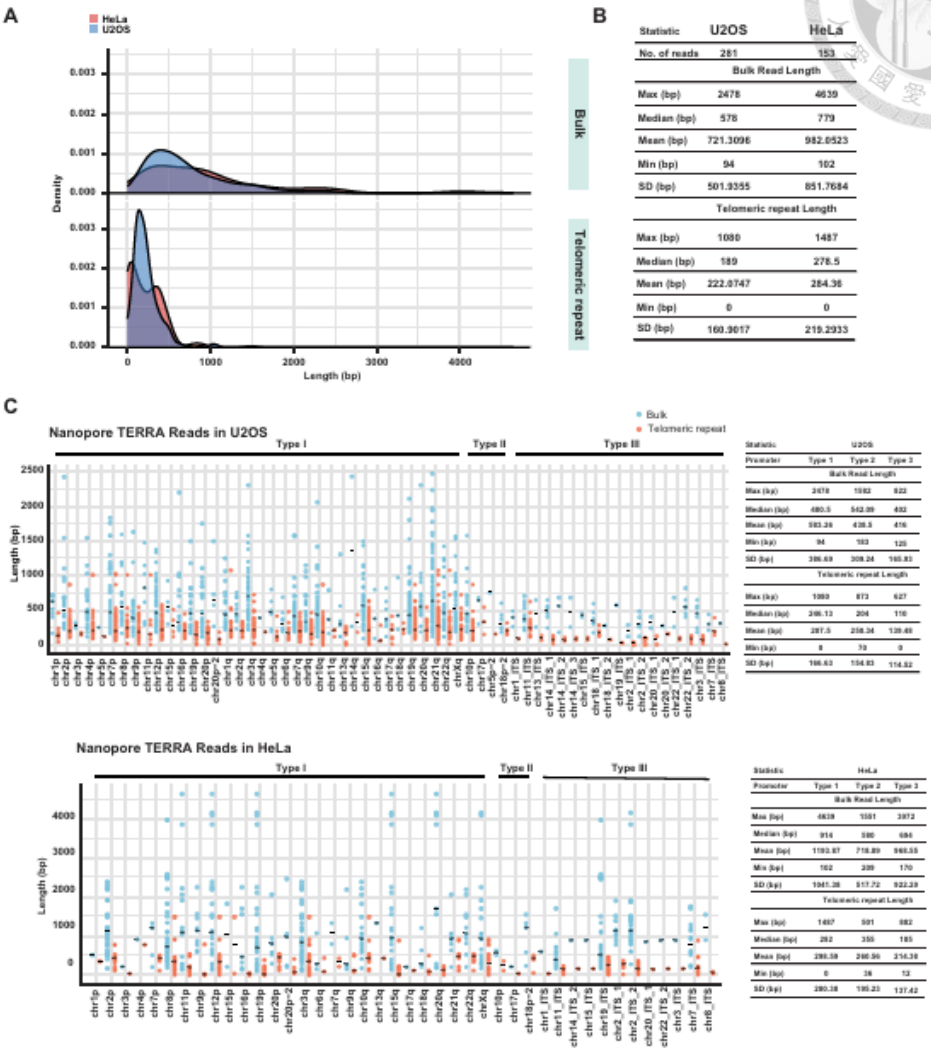
TERRA promoters are associated with H3K4me3, Pol II, CpGs, and R-loops at distinct repeat elements in Type I TERRA transcription regions

Next, we searched for common features among the transcription start sites of TERRA. ChIP-seq data for RNA polymerase II [39], H3K4me3 [38], CTCF [40], and DNA methylation [41] were analyzed using T2T-CHM13 reference genome. Additionally, DRIP-seq data [42] were investigated to examine the profiles of R-loops surrounding the 61–29-37 bp repeats of TERRA promoters located in the subtelomeric regions of the Type I TERRA transcription regions. The meta-analysis revealed epigenetic profiles corresponding to each repeat element (Fig. 3A). The 61 bp repeat element exhibited an enrichment of CTCF, RNA polymerase II, and R loops. The 29 bp repeat element showed an enrichment of DNA methylation compared to other repeats. Remarkably, CAGE tags and H3K4me3 were enriched at the 37 bp repeat regions. The distribution of analyzed epigenetic marks along the TERRA promoter is, illustrated in Fig. 3B, and an example of the genomic view of these marks is shown in Fig. 3C. We conclude that DNA methylation predominantly occurs at 29 bp repeats, while RNA polymerase II and H3K4me3 are enriched at 61 and 37 bp repeat, respectively, in Type I TERRA transcription regions.

Our observations suggest that TERRA are transcribed mainly from the 61–29-37 bp repeat regions. Only chr10p and chr17p do not contain the 61–29-37 repeats adjacent to TERRA transcription regions. It is worth noting that while certain TERRA transcription regions lack 29 bp repeats, they still consist of 61-37 bp repeats. Intriguingly, even in regions without 29 bp repeats, DNA methylation was still detected in the CpG islands between the 61 and 37 bp repeat tracts (Supplementary Fig. S6A).

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

doi:10.6342/NTU202504254



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	1192	822
Median (bp)	680.5	542.09	492
Mean (bp)	983.36	436.5	416
Min (bp)	94	193	125
SD (bp)	386.63	309.24	165.83
Telomeric repeat Length			
Max (bp)	1080	873	627
Median (bp)	286.13	204	110
Mean (bp)	287.5	258.34	139.49
Min (bp)	8	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	684
Mean (bp)	1193.87	718.89	980.55
Min (bp)	102	209	170
SD (bp)	1041.38	317.72	922.39
Telomeric repeat Length			
Max (bp)	1487	501	982
Median (bp)	282	350	185
Mean (bp)	280.58	260.96	214.38
Min (bp)	0	38	13
SD (bp)	280.18	195.23	137.42

Figure 2. TERRA length distribution from Nanopore reads in U2OS and HeLa cells. (**A**) Density plots showing the distribution of TERRA bulk read lengths and telomeric repeat lengths in TERRA reads in U2OS and HeLa cells. Nanopore reads mapped to T2T-CHM13 TERRA transcription regions were calculated. (**B**) Statistic analysis of TERRA bulk read lengths and telomeric repeat lengths in TERRA reads. (**C**) Dot plots show TERRA bulk read lengths and telomeric repeat lengths in TERRA reads at each chromosome end and ITS. Each dot represents each Nanopore read. Solid bar, median.

doi:10.6342/NTU202504254

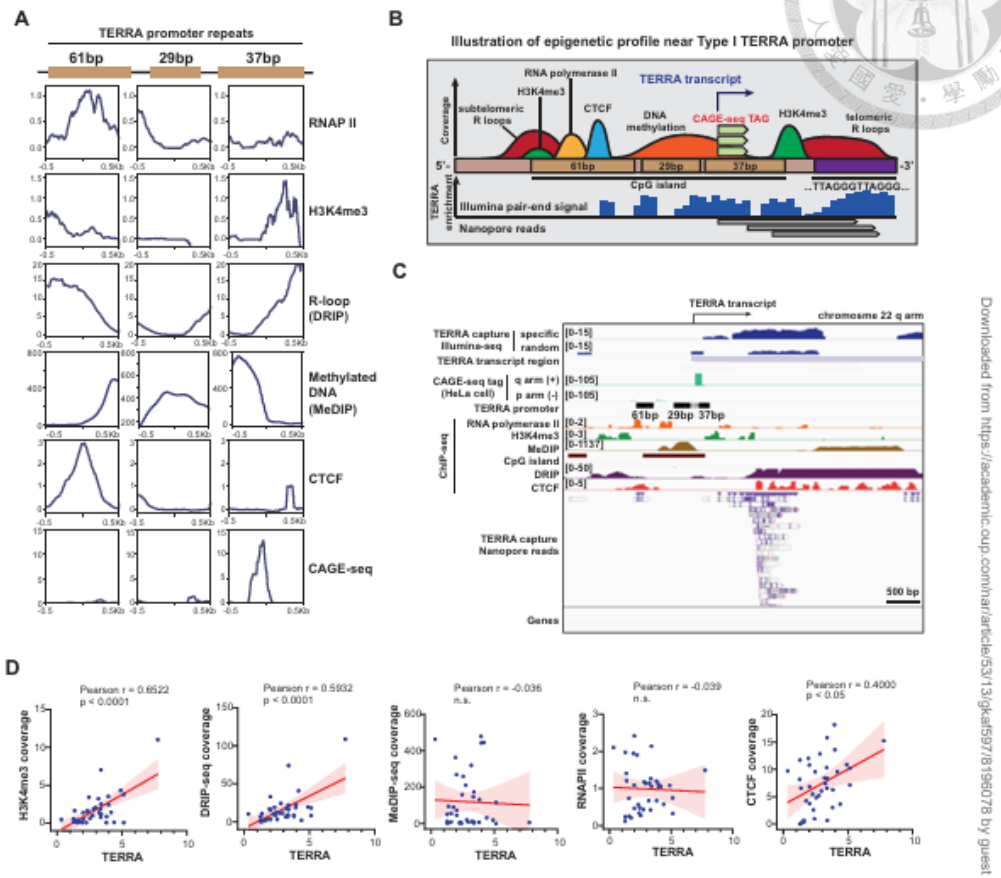


Figure 3. Enrichments of H3K4me3, RNA Pol II, CpGs, DNA methylation, and R-loop at TERRA promoters. (**A**) Meta-analysis showing the distribution of epigenetic marks on 61–29-37 bp repeats located in the subtelomeric regions (Type I promoter). Each plot represents the coverage of indicated marks. (**B**) A schematic model of the epigenetic profiles at Type I TERRA promoter near telomeres. (**C**) Genome browser view showing the coverage of indicated epigenetic marks at TERRA promoter on chr22q arm. (**D**) Scatter plots of TERRA versus H3K4me3, R-loop (DRIP-seq), DNA methylation (MeDIP-seq), CTCF, or RNAPII near TERRA transcription start sites. Each dot indicates TERRA enrichment and epigenetic marks at individual chromosome ends (Type I + Type II TERRA transcription regions) in U2OS cells. TERRA enrichment ($\log_2$ ratio) was calculated by comparing TERRA capture and no capture. *P*-values by Pearson's correlation.

TERRA-QUANT for measuring TERRA expression

To analyze TERRA expression using publicly available short-read RNA-seq datasets, we established a bioinformatics tool, dubbed TERRA-QUANT, for measuring TERRA levels based on the transcription regions defined in this study (Fig. 4A). TERRA-QUANT quantifies reads mapped to TERRA transcription regions, including reads from Types I, II, and III. For total TERRA levels from chromosome ends, only TERRA reads mapped to Type I and II TERRA transcription regions were selected and subsequently underwent YARN normalization. For chromosome-end-specific TERRA analysis, we separated reads from subtelomeric regions and pure telomeric repeat tracts, and only quantified subtelomeric reads and excluded pure telomeric repeat reads.

To test the reliability of this tool, we applied TERRA-QUANT to analyze TERRA expression in HeLa cells using Illumina sequencing data obtained from a previous study [32]. The previous study enriched TERRA RNA by using biotinylated antisense oligos to capture TERRA from nuclear RNA. Notably, TERRA read counts were detected at the majority of chromosome ends in the TERRA capture group, whereas they were absent at multiple chromosome ends in the control group without TERRA capture (Fig. 4B). TRF2 deficiency led to increased TERRA read counts at multiple chromosome ends

doi:10.6342/NTU202504254

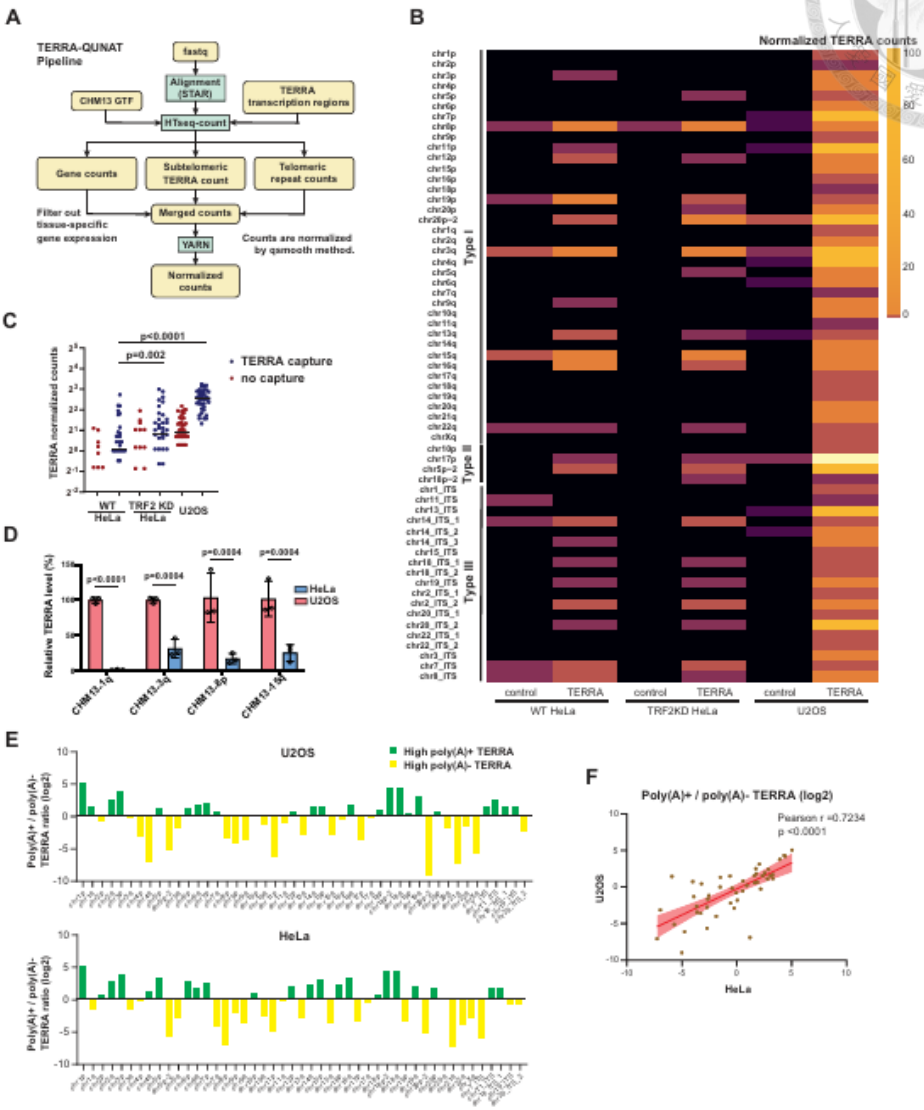


Figure 4. TERRA-QUANT for measuring TERRA expression. (**A**) The workflow of bioinformatics pipeline for quantification of TERRA expression. Reads that mapped to TERRA transcription regions were counted. YARN was used for normalization between different tissues. (**B**) Heatmap showing TERRA expression from individual chromosome ends and ITSs in HeLa and U2OS cells. TERRA capture, using antisense oligos to enrich TERRA. Control, no TERRA capture. (**C**) TERRA normalized counts were analyzed by TERRA-QUANT. Each dot indicates TERRA counts at individual chromosome ends and ITSs. TRF2 deletion (TRF2KD) in HeLa cells increases TERRA expression. *P*-values by Wilcoxon matched-pair signed rank test. Bars, median. (**D**) RT-qPCR to detect TERRA in HeLa and U2OS cells. Subtelomeric primers for TERRA were designed based on the T2T-CHM13 genome sequence. Data from three biological replicates. *P*-values by two-tailed Student's *t*-test. Bars, mean ± SD. (**E**) Log_2 ratios of poly(A)+ to poly(A)− TERRA reads mapped to individual chromosome ends and ITSs in U2OS and HeLa cells. (**F**) Scatter plots showing the correlation between U2OS and HeLa cells in the poly(A)+/poly(A)− TERRA ratios. Each dot represents the ratio (log_2) of poly(A)+/poly(A)− reads at each TERRA transcription region.

doi:10.6342/NTU202504254

compared to wild-type cells (Fig. 4B). Additionally, TERRA read counts from chromosome ends in HeLa cells were significantly lower than those in U2OS cells (Fig. 4C). These results are consistent with earlier studies showing elevated TERRA levels upon TRF2 depletion [4, 32], affirming the robustness of the TERRA-QUANT methodology.

To further validate the annotations for TERRA transcription regions, we designed subtelomeric primer sets specific to CHM13-1q, 3q, 8p, and 15q chromosome ends based on the T2T-CHM13 genome sequence (Supplementary Table S6). We used these subtelomeric primers to detect TERRA following TERRA capture and observed over 1000-fold enrichment (Supplementary Fig. S6B), indicating that the primers efficiently amplified TERRA transcripts. Notably, RT-qPCR analysis using these subtelomeric primers revealed increased TERRA levels from various chromosome ends in U2OS compared to HeLa cells (Fig. 4D). Subsequently, the RT-qPCR products were subjected to Illumina sequencing to validate their sequences. We observed that the reads of RT-qPCR products were predominantly mapped to their target chromosome ends (Supplementary Fig. S6C). These results indicate the specificity of these subtelomeric primers.

Poly(A)+ and non-poly(A) TERRA capture sequencing

Since most publicly available RNA-seq datasets are generated from polyadenylated RNA—and only a portion of TERRA transcripts are polyadenylated—we sought to investigate the genomic origins of TERRA molecules with and without poly(A) tails. To achieve this, we performed TERRA capture followed by an additional poly(A) selection step to separate poly(A)+ and poly(A)− fractions. Both poly(A)+ and poly(A)− TERRA fractions were subjected to Illumina sequencing. We analyzed the ratio of poly(A)+ to poly(A)− TERRA reads in both U2OS and HeLa cells and found that certain chromosome ends, such as chr1p, chr2q, chr3p, and 18q, produced more poly(A)+ TERRA transcripts in both cell lines (Fig. 4E and Supplementary Fig. S6D). In contrast, other chromosome ends, including chr8p, chr11p, chr15q, and chr21q, produced more poly(A)− TERRA transcripts. Notably, we observed a strong positive correlation (Pearson $r = 0.7$, $P < .0001$) between U2OS and HeLa cells in the ratio of poly(A)+ to poly(A)− across different chromosome ends (Fig. 4F), suggesting that polyadenylation patterns in TERRA are largely consistent across cell types.

ALT-positive osteosarcomas show elevated TERRA levels

We gathered rRNA-depleted RNA-seq datasets from a previous study, which included three biological replicates of 13 human osteosarcoma cell lines, with 4 cell lines categorized as ALT negative ($n = 12$) and 9 cell lines classified as ALT positive ($n = 27$) [45]. The RNA-seq reads of these datasets were aligned to the T2T-CHM13 genome and processed using TERRA-QUANT. Total TERRA levels were then determined by cumulating TERRA reads from all chromosome ends. Notably, ALT-positive cells exhibited a higher level of total TERRA, and a lower RNA level of telomerase reverse transcriptase (TERT) (Fig. 5A and B). These results agree with the previous finding showing ALT cells with elevated TERRA expression [2, 54].

TERRA increases along with human aging in blood cells

To compare TERRA expression across different ages, we collected poly(A)-enriched RNA-seq datasets from a previous study [46], which comprised 222 human blood samples for analysis. These blood samples cover a range of ages from 24 to 73 years old and were separated into three groups: young (<30 years old), mid-age adult (30–59 years old), and old groups (≥60 years old). Interestingly, TERRA-QUANT analysis demonstrated that total TERRA levels accumulated from all chromosome ends were positively correlated with age (Fig. 5C). When grouped by different ages, total TERRA levels were significantly increased in old individuals, compared to young and mid-age individuals (Fig. 5D). For chromosome-end-specific analysis, reads mapping to individual chromosome ends were low, preventing statistical analysis (Supplementary Fig. S7A). Notably, TERRA reads were predominately mapped to Type I transcription regions (Supplementary Fig. S7B).

Next, we confirmed the association of TERRA levels with aging by collecting blood samples from individuals of various ages and conducted quantitative PCR to measure telomere length and TERRA levels. The samples were grouped by adult (21–60 years) or old (>60 years) ages. Genomic DNA was extracted from buffy coats, and telomere length (T/S ratio) was analyzed by quantitative PCR. We observed that telomere lengths significantly decreased in elderly people in both males and females (Fig. 5E). Interestingly, total TERRA RNA levels in blood samples determined by RT-qPCR were significantly elevated in old males and females (Fig. 5F), and were negatively correlated with telomere lengths (Fig. 5G). Using subtelomeric primers for RT-qPCR also showed a significant increase in TERRA expression in the old group compared to the adult group (Fig. 5H), confirming an elevation in TERRA levels in aged leukocytes.

TERRA is upregulated in the aged brain

To explore the relationship between TERRA and human aging, we analyzed published poly(A)-enriched RNA-seq datasets obtained from various human tissues including brains, hearts, and ovaries [46] by TERRA-QUANT. The scatter plots displayed a significant positive correlation between total TERRA levels and age in brains, but not in ovaries or hearts (Fig. 6A). Remarkably, total TERRA levels in brain tissues exhibited an upregulation in the old group (≥60 years old), compared to the adult group (30–59 years old) (Fig. 6B). Considering that ATRX is important for neuronal development and functionality [55, 56, 57], we analyzed ATRX expression in human brains. Interestingly, the levels of ATRX, which is a TERRA-interacting protein [24], declined in old individuals (Supplementary Fig. S7C) and displayed an anticorrelation with TERRA expression in brain tissues (Supplementary Fig. S7D).

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

Next, we quantified TERRA expression in fibroblasts obtained from healthy individuals of various ages and patients diagnosed with HGPS, a genetic disorder characterized by premature aging features in childhood. Total TERRA levels, analyzed by TERRA-QUANT using poly(A)-enriched

doi:10.6342/NTU202504254

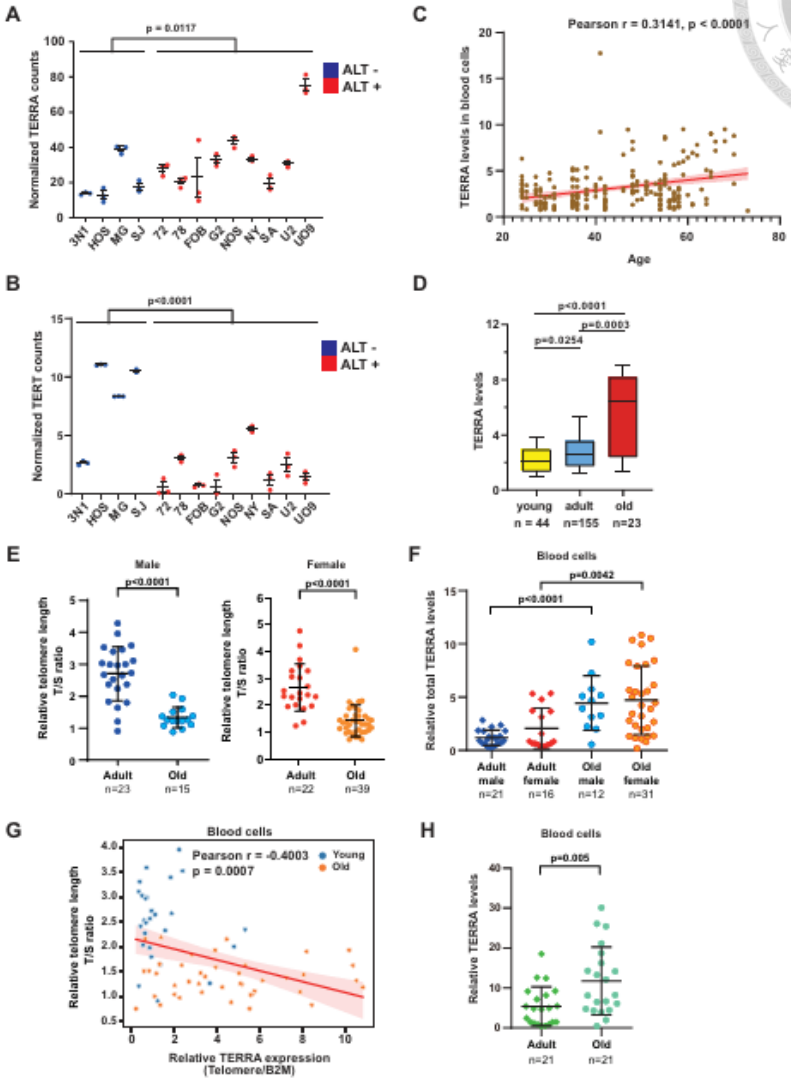


Figure 5. TERRA levels increase with age in blood cells. (**A**) TERRA expression in ALT-positive (+) or negative (−) cell lines was quantified by TERRA-QUANT. Total TERRA counts at all chromosome ends were accumulated. Each dot indicates the normalized TERRA read counts of each RNA-seq dataset. Bars, mean ± SD. *P*-values, by Mann–Whitney *U* test. (**B**) TERT expression of each RNA-seq data from ALT+ or ALT− cells. Bars, mean ± SD. *P* -values, by Mann–Whitney *U* test. (**C**) TERRA expression in human blood cells was analyzed by TERRA-QUANT using RNA-seq datasets. Scatter plots showing the correlation of TERRA versus age in blood. Each dot indicates the total TERRA normalized reads from all chromosome ends of each individual. *P* -values, by Pearson's correlation. (**D**) Normalized TERRA counts in blood cells of different ages: young (<30 years); adult (30–59 years); and old (≥60 years). Bars, median with interquartile. *P*-values, by Mann–Whitney *U* test. (**E**) The T/S ratio represents the relative telomere length (T) to the single-copy gene (S, 36B4 gene). Each dot represents an individual T/S ratio: adult (21–59 years); old (≥60 years). (**F**) RT-qPCR to detect total TERRA levels in blood cells using telomeric repeat primers. Each dot indicates each individual. (**G**) Scatter plot showing negative correlation between TERRA and telomere length in blood cells. *P* -values, by Pearson's correlation. (**H**) RT-qPCR to detect TERRA levels using subtelomeric primers (hg38-2q). Each dot indicates each individual. (E, F, and H) *P* -values, by two-tailed Student's *t*-test. *n* = sample size. Error bar, SD.

doi:10.6342/NTU202504254

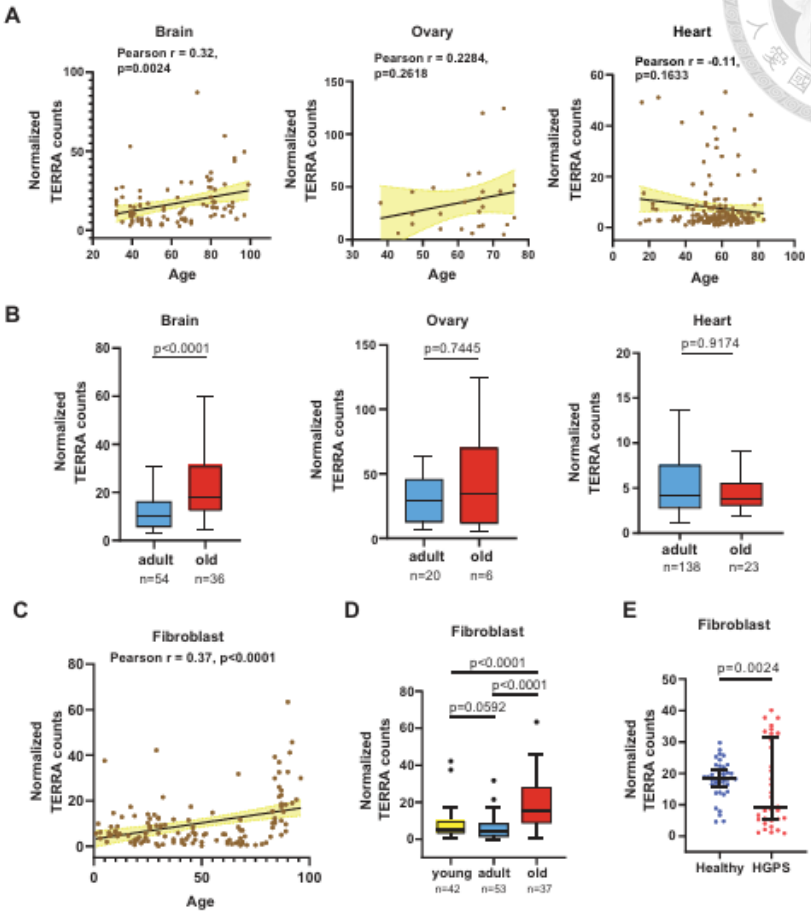


Figure 6. TERRA expression increases with age in brain tissues and fibroblasts. (**A**) Scatter plots showing the correlation of TERRA versus age in brain, ovary, and heart tissues. Each dot indicates the total TERRA level from all chromosome ends in an individual. *P* values, by Pearson's correlation. (**B**) TERRA levels in various tissues with different ages: Adult (21–59 years); and old (≥60 years). Bars, median with interquartile. *P* values, by Mann–Whitney *U* test. (**C**) TERRA levels in human fibroblasts derived from healthy individuals of different ages. *P* values, by Pearson's correlation. (**D**) Boxplots showing the upregulation of TERRA levels in old human fibroblasts: Young (<30 years); adult (30–59 years); and old (≥60 years). Bars, median with interquartile. *P* values, by Mann–Whitney *U* test. (**E**) Fibroblasts derived from HGPS patients display abnormal TERRA levels. Each dot represents the TERRA level of each individual. Bars, median with interquartile. *P* values, by Kolmogorov–Smirnov test.

RNA-seq datasets, displayed a significant upregulation in fibroblasts derived from healthy older individuals (Fig. 6C and D), with an exponential increase observed in those over 75 years old. Chromosome-end-specific analysis showed that TERRA increased at multiple chromosome ends in the old group (Supplementary Fig. S8A). Type I TERRA transcription was dominant in fibroblasts compared to Types II and III (Supplementary Fig. S8B). Considering that HGPS patients were below 30 years old, we compared TERRA levels with those of their age-matched controls. Notably, HGPS patients displayed an aberrant and diverse TERRA expression pattern compared to healthy individuals (Fig. 6E).

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

We assessed TERRA levels across various cell types by analyzing published RNA-seq datasets from human tissues, in-

doi:10.6342/NTU202504254

cluding adipose, blood, bone, brain, heart, lung, ovary, pancreas, and retina (Supplementary Table S7). Among these tissues, TERRA levels were particularly high in blood, bone, and ovary (Fig. 7A). Given the highest level in blood cells, we sought to identify the specific cell types exhibiting high TERRA expression. To achieve this, we analyzed single-cell RNA-seq datasets [58] from peripheral blood mononuclear cells obtained from healthy young individuals. Notably, dendritic cells displayed the highest TERRA level among various immune cells (Fig. 7B).

To investigate the potential role of TERRA in neurons, we examined TERRA levels using single-cell RNA-seq datasets obtained from hESCs undergoing neuronal differentiation [59]. The hESCs were converted to neuromesodermal progenitors (NMPs) with distinct HOX profiles along rostrocaudal axis by modulation of Wnt, FGF, and retinoic acid signaling, and were further differentiated into ventral neuron phenotypes. Six different NMP cultures from hESCs were generated corresponding to HOX patterning periods: 24, 48, 72, 120, 168, and 216 h. The cultures exhibited a progressive expression of caudal HOX paralogs, corresponding to cervical (HOX1-8; observed in 24, 48, and 72 h), thoracic (HOX1-9; observed in 120 h), lumbar (HOX1-11; observed in 168 h), and lumbosacral spinal regions (HOX1-13; observed in 216 h). Cells were clustered based on gene expression into mesenchymal cells, neuroepithelial cells, and neurons. Notably, higher levels of TERRA were observed in neurons compared to neuroepithelial cells and mesenchymal cells (Fig. 7C). Additionally, TERRA levels displayed a concordant expression pattern with ATRX during neuronal differentiation (Fig. 7D–F).

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

Next, we sought to analyze TERRA levels using TERRA-QUANT in neuronal disorder diseases. We analyzed published single-nucleus RNA-seq datasets from the prefrontal cortex of patients with Alzheimer's disease (AD) [60]. Postmortem human brain samples were obtained from participants enrolled in the Religious Order Study and Rush Memory and Aging Project (ROSMAP) [61]. Individuals with AD pathology were categorized into two subgroups according to the previous study [60], reflecting distinct stages of AD pathological progression: (i) early-pathology characterized by amyloid burden, modest neurofibrillary tangles, and cognitive impairment; and (ii) late-pathology marked by higher amyloid levels, elevated neurofibrillary tangles, increased global pathology, and cognitive impairment. Notably, increased TERRA levels were observed in various types of brain cells in the early stage of AD pathology compared to no AD pathology, with more profound TERRA levels in excitatory neurons (Fig. 7G and Supplementary Fig. S9A). To further investigate TERRA in AD neurons, induced pluripotent stem cells (iPSCs) derived from an AD patient, along with mutation-corrected isogenic control cells, were generated. The AD patient carried the P117L mutation in presenilin-1 (PS1mut), the catalytic subunit of γ-secretase. Notably, RT-qPCR demonstrated that AD-PS1mut neuronal cells differentiated from iPSCs exhibited significantly higher TERRA levels compared to control cells (Fig. 7H), further supporting the upregulation of TERRA expression in AD neurons.

Discussion

In this study, we investigated TERRA transcripts in human cells using Nanopore direct RNA-seq and TERRA capture RNA-seq. We identified TERRA transcription regions at 39 chromosome ends and utilized these defined regions to quantify TERRA expression across various tissues. Our bioinformatics pipeline is versatile, accommodating different RNA-seq datasets, including single-cell RNA-seq experiments. Notably, substantial mapped reads per cell were observed in single-cell RNA-seq datasets without YARN normalization (Supplementary Fig. S9A–C), indicating that TERRA can be detected at the single-cell level.

Some discrepancies were observed between our findings and those reported by Rodrigues *et al.* [51] regarding TERRA expression at specific chromosome ends in HeLa and U2OS cells. A comparison of our and their results revealed consistent identification of TERRA expression at 31 chromosome ends, while 3 chromosome ends were consistently found to lack detectable TERRA expression (Supplementary Fig. S9D). Our analysis revealed eight additional chromosome ends expressing TERRA that were not reported by Rodrigues *et al.*, while their study identified four ends absent from our dataset. We attribute these differences primarily to the distinct mapping strategies applied. Rodrigues *et al.* used a limited reference consisting of 25 kb subtelomeric regions, whereas we utilized the entire genome for read alignment. When we applied 25 kb subtelomeric regions as a reference genome for mapping our data, the results aligned closely with their findings. These observations suggest that the choice of mapping strategy impacts the sensitivity and specificity of TERRA detection at individual chromosome ends.

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

Utilizing TERRA-QUANT for quantifying TERRA levels across diverse tissues, we detected TERRA expression in most tissues examined. Notably, high TERRA levels were observed in blood and bone tissues, raising the possibility of a role

doi:10.6342/NTU202504254

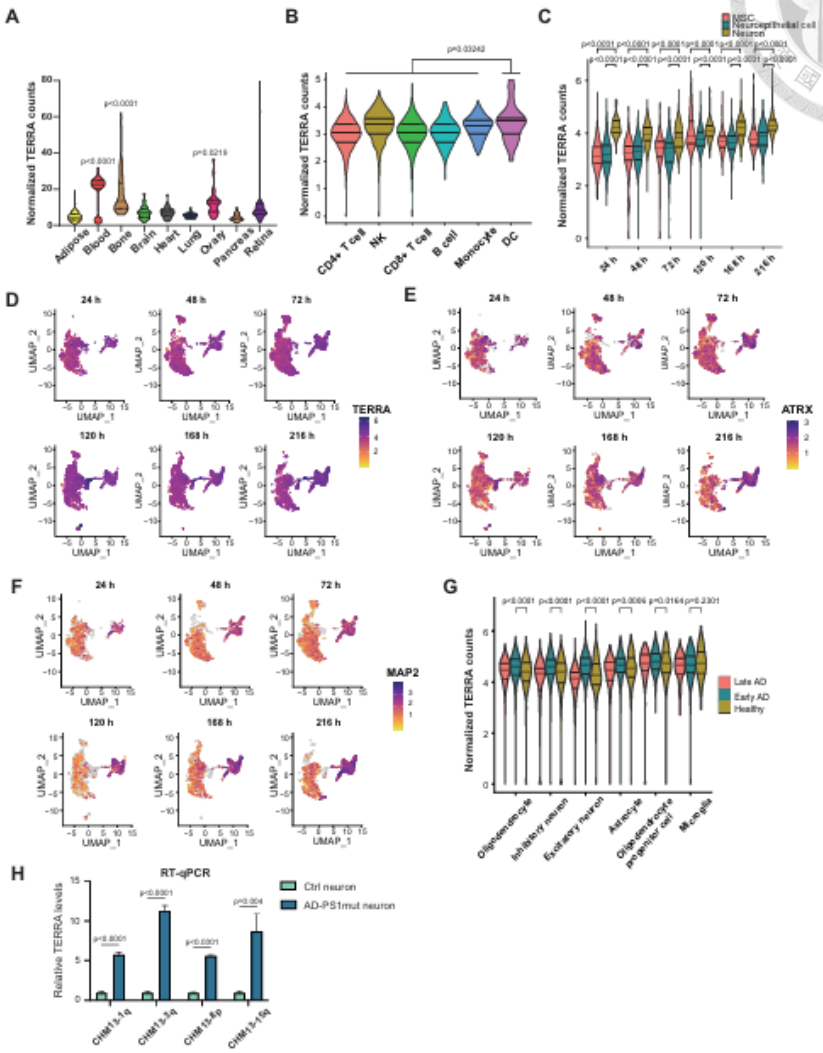


Figure 7. Single-cell analysis of TERRA levels undergoing neuronal differentiation and in Alzheimer's disease. (**A**) Violin plot showing normalized TERRA counts in various human tissues. The solid lines indicate the median and the dashed lines are interquartile range. *P* values, by Mann–Whitney *U* test. (**B**) Single-cell RNA-seq showing normalized TERRA counts in various immune cells from PBMC. NK, nature killer cell; DC, dendritic cells. Solid lines, median. Dashed lines are interquartile range. *P* values, by Mann–Whitney *U* test. (**C**) Single-cell RNA-seq analysis of normalized TERRA counts in hESCs undergoing neuronal differentiation with various HOX patterning periods (24, 72, 48, 120, 168, and 216 h). MSC, mesenchymal stem cells. Solid lines, median. Dashed lines are interquartile range. *P* values, by Mann–Whitney *U* test. (**D**–**F**). UMAP plots of single-cell RNA-seq datasets from hESCs undergoing neuronal differentiation using various HOX patterning periods. Colors according to TERRA, ATRX, or neuron marker MAP2 expression levels. (**G**) TERRA levels in cells isolated from the prefrontal cortex of individuals with Alzheimer's disease (AD). Single-nucleus RNA-seq datasets were obtained from the ROSMAP project, and grouped into healthy (no AD pathology), early-AD pathology, and late-AD pathology. Solid lines, median. Dashed lines are interquartile range. *P* values, by Mann–Whitney *U* test. (**H**) RT-qPCR analysis showing elevated TERRA levels in AD neurons. AD-PS1mut or Ctrl neurons were differentiated from iPSC. AD-PS1mut carrying P117L mutation in *PS1* gene. Ctrl: mutation-corrected isogenic control. *P* values, by two-tailed Student's *t*-test. $n = 3$. TERRA levels normalized to GAPDH.

doi:10.6342/NTU202504254

for TERRA in immune-related processes. The earlier research has shown that TERRA can be secreted via extracellular exosomes to stimulate inflammatory cytokines during telomere dysfunction [62, 63]. TERRA also interacts with ZBP1 to activate immune response under conditions of replicative crisis [64]. However, it is important to note that such ZBP1–TERRA complex-mediated immune activation is unlikely to occur under normal physiological conditions in healthy blood cells. Notably, we observed elevated total TERRA levels in fibroblasts, brain, and blood cells as individuals aged. The telomere shortening may contribute to TERRA upregulation during aging. However, neurons are postmitotic cells that are not expected to undergo telomere shortening. Therefore, the increase of TERRA in the brain is unlikely to result from telomere shortening, but may instead be associated with oxidative stress, neuronal damage, and telomere dysfunction in the aging brain. It has been previously reported that telomeric transcription is induced by DNA damage at telomeres in HGPS fibroblasts [33]. In our analysis, fibroblasts derived from HGPS patients demonstrated abnormal TERRA expression, with some individuals showing high levels while others displayed lower levels. We reason that the diverse TERRA counts observed in HGPS patients could be attributed to variable telomere lengths or defects in telomere integrity. Previous studies have shown that telomere length in HGPS patients is generally shorter, yet exhibits variability [65, 66]. Low TERRA counts in some HGPS patients could result from either low TERRA expression or the loss of subtelomeric and telomeric DNA at chromosome ends.

Examination of single-nucleus RNA-seq datasets and RT-qPCR analysis of TERRA levels unveiled an upregulation of TERRA levels in Alzheimer's disease (AD) neurons. AD is associated with several cellular stress responses—such as oxidative stress and chronic inflammation [67–69]—that are also known to upregulate TERRA expression [37, 70]. Moreover, the accumulation of DNA G4 structures has been reported in AD neurons [71], and TERRA has also been implicated in promoting G4 formation [27]. These results indicate elevated TERRA levels correlate with increased DNA G4 formation and oxidative stress, which are aligned with the observations in AD neurons. Taken together, these connections provide a compelling rationale for further investigation into the role of TERRA in AD pathogenesis.

Upregulation of TERRA and ATRX during neuronal differentiation implies the potential function of TERRA in neuronal cells. Interestingly, the inverse relationship between TERRA and ATRX in the aged brain was observed, suggesting that low levels of ATRX might contribute to the dysfunction of the aged neurons. Previous research indicated that the absence of ATRX leads to cognitive impairment and promotion of cellular senescence [72], further highlighting the significance of exploring the interplay between TERRA, ATRX, and aging-related cellular processes in the brain and neuronal tissues. It is plausible that proper levels of TERRA and ATRX are crucial for maintaining normal gene expression in neuronal cells, and dysregulation of TERRA and ATRX may impact neuron function during aging.

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

Acknowledgements

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

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by National Science and Technology Council (NSTC) grants NSTC 112-2628-B-002-008 (H.-P.C.C.), NSTC 112-2320-B-002-058 (H.-P.C.C.), NSTC 113-2320-B-002-009 (H.-P.C.C.), and NSTC 113-2628-B-002-010-MY3 (H.-P.C.C.); National Taiwan University grants NTU-111L7880 (H.-P.C.C.), NTU-AS-112L104312 (H.-P.C.C), NTU-CDP-112L7721 (H.-P.C.C.), NTU-CDP-113L7705 (H-P.C.C.); and National Health Research Institutes grants NHRI-EX111-11107SI, NHRI-EX112-11107SI (H.-P.C.C.), NHRI-EX113-11107SI (H.-P.C.C.), and NHRI-EX114-11411SI. Funding to pay the Open Access publication charges for this article was provided by National Science and Technology Council.

Data availability

TERRA-capture RNA-seq datasets including Illumina short-read RNA-seq, Nanopore direct RNA-seq, poly(A)+ and poly(A)− TERRA capture RNA-seq, and the sequencing data of TERRA RT-qPCR products are accessible at GEO (GEO accession GSE250303) https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE250303. The published sequencing datasets analyzed in this study are described in Supplementary Tables S7 --9. A copy of the custom code utilized for quantification of TERRA is accessible on GitHub (https://github.com/ls807terra/TERRA_RNA-seq_pipeline) and Zenodo (https://doi.org/10.5281/zenodo.15598541). Publicly available packages were used in this study as indicated in the "Materials and methods" section.

doi:10.6342/NTU202504254

References

1. Azzalin CM, Reichenbach P, Khoriauli L *et al.* Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 2007;**318**:798–801. https://doi.org/10.1126/science.1147182
2. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* 2008;**10**:228–36. https://doi.org/10.1038/ncb1685
3. Feretzaki M, Renck Nunes P, Lingner J. Expression and differential regulation of human TERRA at several chromosome ends. *RNA* 2019;**25**:1470–80. https://doi.org/10.1261/rna.072322.119
4. Nergadze SG, Farnung BO, Wischnewski H *et al.* CpG-island promoters drive transcription of human telomeres. *RNA* 2009;**15**:2186–94. https://doi.org/10.1261/rna.1748309
5. Graf M, Bonetti D, Lockhart A *et al.* Telomere length determines TERRA and R-loop regulation through the cell cycle. *Cell* 2017;**170**:72–85. https://doi.org/10.1016/j.cell.2017.06.006
6. Azzalin CM, Lingner J. Telomeres: the silence is broken. *Cell Cycle* 2008;**7**:1161–5. https://doi.org/10.4161/cc.7.9.5836
7. Savoca V, Rivosecchi J, Gaiatto A *et al.* TERRA stability is regulated by RALY and polyadenylation in a telomere-specific manner. *Cell Rep* 2023;**42**:112406. https://doi.org/10.1016/j.celrep.2023.112406
8. Rossi A, Moro A, Tebaldi T *et al.* Identification and dynamic changes of RNAs isolated from RALY-containing ribonucleoprotein complexes. *Nucleic Acids Res* 2017;**45**:6775–92. https://doi.org/10.1093/nar/gkx235
9. Yehezkel S, Segev Y, Viegas-Pequignot E *et al.* Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum Mol Genet* 2008;**17**:2776–89. https://doi.org/10.1093/hmg/ddn177
10. Sagie S, Toubiana S, Hartono SR *et al.* Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat Commun* 2017;**8**:14015. https://doi.org/10.1038/ncomms14015
11. Deng Z, Campbell AE, Lieberman PM. TERRA, CpG methylation and telomere heterochromatin: lessons from ICF syndrome cells. *Cell Cycle* 2010;**9**:69–74. https://doi.org/10.4161/cc.9.1.10358
12. Deng Z, Wang Z, Stong N. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J* 2012;**31**:4165–78. https://doi.org/10.1038/emboj.2012.266
13. Lovejoy CA, Li W, Reisenweber S. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet* 2012;**8**:e1002772. https://doi.org/10.1371/journal.pgen.1002772
14. Diman A, Boros J, Poulain F *et al.* Nuclear respiratory factor 1 and endurance exercise promote human telomere transcription. *Sci Adv* 2016;**2**:e1600031. https://doi.org/10.1126/sciadv.1600031
15. Tutton S, Azzam GA, Stong N. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J* 2016;**35**:193–207. https://doi.org/10.15252/embj.201490880
16. Koskas S, Decottignies A, Dufour S *et al.* Heat shock factor 1 promotes TERRA transcription and telomere protection upon heat stress. *Nucleic Acids Res* 2017;**45**:6321–33. https://doi.org/10.1093/nar/gkx208
17. Mazzolini R, Gonzalez N, Garcia-Garijo A *et al.* Snail1 transcription factor controls telomere transcription and integrity. *Nucleic Acids Res* 2018;**46**:146–58. https://doi.org/10.1093/nar/gkx958
18. Arora R, Lee Y, Wischnewski H *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun* 2014;**5**:5220. https://doi.org/10.1038/ncomms6220
19. Yu TY, Kao YW, Lin JJ. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc Natl Acad Sci USA* 2014;**111**:3377–82. https://doi.org/10.1073/pnas.1307415111
20. Guh CY, Shen HJ, Chen LW. XPF activates break-induced telomere synthesis. *Nat Commun* 2022;**13**:5781. https://doi.org/10.1038/s41467-022-33428-0
21. Silva B, Arora R, Bione S *et al.* TERRA transcription destabilizes telomere integrity to initiate break-induced replication in human ALT cells. *Nat Commun* 2021;**12**:3760. https://doi.org/10.1038/s41467-021-24097-6
22. Deng Z, Norseen J, Wiedmer A *et al.* TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* 2009;**35**:403–13. https://doi.org/10.1016/j.molcel.2009.06.025
23. Mei Y, Deng Z, Vladimirova O *et al.* TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. *Sci Rep* 2021;**11**:3509.
24. Chu HP, Cifuentes-Rojas C, Kesner B *et al.* TERRA RNA antagonizes ATRX and protects telomeres. *Cell* 2017;**170**:86–101.e16. https://doi.org/10.1016/j.cell.2017.06.017
25. Flynn RL, Centore RC, O'Sullivan RJ *et al.* TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* 2011;**471**:532–6. https://doi.org/10.1038/nature09772
26. Voon HP, Hughes JR, Rode C *et al.* ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. *Cell Rep* 2015;**11**:405–18. https://doi.org/10.1016/j.celrep.2015.03.036
27. Tsai RX, Fang KC, Yang PC *et al.* TERRA regulates DNA G-quadruplex formation and ATRX recruitment to chromatin. *Nucleic Acids Res* 2022;**50**:12217–34. https://doi.org/10.1093/nar/gkac1114
28. Arnoult N, Van Beneden A, Decottignies A. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1alpha. *Nat Struct Mol Biol* 2012;**19**:948–56. https://doi.org/10.1038/nsmb.2364
29. Cusanelli E, Romero CA, Chartrand P. Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. *Mol Cell* 2013;**51**:780–91. https://doi.org/10.1016/j.molcel.2013.08.029
30. Misino S, Busch A, Wagner CB *et al.* TERRA increases at short telomeres in yeast survivors and regulates survivor associated senescence (SAS). *Nucleic Acids Res* 2022;**50**:12829–43. https://doi.org/10.1093/nar/gkac1125
31. Caslini C, Connelly JA, Serna A *et al.* MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. *Mol Cell Biol* 2009;**29**:4519–26. https://doi.org/10.1128/MCB.00195-09
32. Porro A, Feuerhahn S, Delafontaine J *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* 2014;**5**:5379. https://doi.org/10.1038/ncomms6379
33. Aguado J, Sola-Carvajal A, Cancila V. Inhibition of DNA damage response at telomeres improves the detrimental phenotypes of Hutchinson–Gilford progeria syndrome. *Nat Commun* 2019;**10**:4990. https://doi.org/10.1038/s41467-019-13018-3
34. Maestroni L, Reyes C, Vaurs M *et al.* Nuclear envelope attachment of telomeres limits TERRA and telomeric rearrangements in quiescent fission yeast cells. *Nucleic Acids Res* 2020;**48**:3029–41. https://doi.org/10.1093/nar/gkaa043
35. Balk B, Maicher A, Dees M *et al.* Telomeric RNA–DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol* 2013;**20**:1199–205. https://doi.org/10.1038/nsmb.2662
36. Wanat JJ, Logsdon GA, Driskill JH *et al.* TERRA and the histone methyltransferase Dot1 cooperate to regulate senescence in budding yeast. *PLoS One* 2018;**13**:e0195698. https://doi.org/10.1371/journal.pone.0195698
37. Liu B, Maekawa T, Yoshida K *et al.* Telomere shortening by transgenerational transmission of TNF-alpha-induced TERRA via ATF7. *Nucleic Acids Res* 2019;**47**:283–98. https://doi.org/10.1093/nar/gky1149

doi:10.6342/NTU202504254

38. Jang SM, Zhang Y, Utani K. The replication initiation determinant protein (RepID) modulates replication by recruiting CUL4 to chromatin. *Nat Commun* 2018;**9**:2782. https://doi.org/10.1038/s41467-018-05177-6
39. Gorthi A, Romero JC, Loranc E. EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature* 2018;**555**:387–91. https://doi.org/10.1038/nature25748
40. Ibarra A, Benner C, Tyagi S *et al.* Nucleoporin-mediated regulation of cell identity genes. *Genes Dev* 2016;**30**:2253–8. https://doi.org/10.1101/gad.287417.116
41. Wang Y, Khan A, Marks AB. Temporal association of ORCA/LRWD1 to late-firing origins during G1 dictates heterochromatin replication and organization. *Nucleic Acids Res* 2017;**45**:2490–502. https://doi.org/10.1093/nar/gkw1211
42. De Magis A, Manzo SG, Russo M *et al.* DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc Natl Acad Sci USA* 2019;**116**:816–25. https://doi.org/10.1073/pnas.1810409116
43. Paulson JN, Chen CY, Lopes-Ramos CM *et al.* . Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics* 2017; **18**:437. https://doi.org/10.1186/s12859-017-1847-x
44. Chang KV, Chen YC, Wu WT *et al.* Expression of telomeric repeat-containing RNA decreases in sarcopenia and increases after exercise and nutrition intervention. *Nutrients* 2020;**12**:3766. https://doi.org/10.3390/nu12123766
45. Mason-Osann E, Dai A, Floro J *et al.* Identification of a novel gene fusion in ALT positive osteosarcoma. *Oncotarget* 2018;**9**:32868–80. https://doi.org/10.18632/oncotarget.26029
46. Shokhirev MN, Johnson AA. Modeling the human aging transcriptome across tissues, health status, and sex. *Aging Cell* 2021;**20**:e13280. https://doi.org/10.1111/acel.13280
47. Hicks SC, Okrah K, Paulson JN *et al.* Smooth quantile normalization. *Biostatistics* 2018;**19**:185–98. https://doi.org/10.1093/biostatistics/kxx028
48. Zheng GX, Terry JM, Belgrader P. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**:14049. https://doi.org/10.1038/ncomms14049
49. Hao Y, Stuart T, Kowalski MHea. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* 2023;**42**:293–304.
50. Miga KH, Koren S, Rhie A. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 2020;**585**:79–84. https://doi.org/10.1038/s41586-020-2547-7
51. Rodrigues J, Alfieri R, Bione S *et al.* TERRA ONTseq: a long-read-based sequencing pipeline to study the human telomeric transcriptome. *RNA* 2024;**30**:955–66. https://doi.org/10.1261/rna.079906.123
52. Azzalin CM, Lingner J. Telomere functions grounding on TERRA firma. *Trends Cell Biol* 2015;**25**:29–36. https://doi.org/10.1016/j.tcb.2014.08.007
53. IJdo JW, Baldini A, Ward DC *et al.* Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci U S A* 1991;**88**:9051–5. https://doi.org/10.1073/pnas.88.20.9051
54. Rosso I, Jones-Weinert C, Rossiello F. Alternative lengthening of telomeres (ALT) cells viability is dependent on C-rich telomeric RNAs. *Nat Commun* 2023;**14**:7086. https://doi.org/10.1038/s41467-023-42831-0
55. Gibbons RJ, Picketts DJ, Villard L *et al.* Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* 1995;**80**:837–45. https://doi.org/10.1016/0092-8674(95)90287-2
56. Quesnel KM, Martin-Kenny N, Berube NG. A mouse model of ATRX deficiency with cognitive deficits and autistic traits. *J Neurodev Disord* 2023;**15**:39. https://doi.org/10.1186/s11689-023-09508-7
57. Berube NG, Mangelsdorf M, Jagla M *et al.* The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. *J Clin Invest* 2005;**115**:258–67. https://doi.org/10.1172/JCI200522329
58. Luo OJ, Lei W, Zhu G *et al.* . Multidimensional single-cell analysis of human peripheral blood reveals characteristic features of the immune system landscape in aging and frailty. *Nat Aging* 2022; **2**:348–64. https://doi.org/10.1038/s43587-022-00198-9
59. Iyer NR, Shin J, Cuskey S *et al.* Modular derivation of diverse, regionally discrete human posterior CNS neurons enables discovery of transcriptomic patterns. *Sci Adv* 2022;**8**:eabn7430. https://doi.org/10.1126/sciadv.abn7430
60. Mathys H, Davila-Velderrain J, Peng Z. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 2019;**570**:332–7. https://doi.org/10.1038/s41586-019-1195-2
61. Bennett DA, Buchman AS, Boyle PA *et al.* Religious orders study and rush memory and aging project. *J Alzheimers Dis* 2018;**64**:S161–89. https://doi.org/10.3233/JAD-179939
62. Wang Z, Lieberman PM. The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol* 2016;**13**:690–5. https://doi.org/10.1080/15476286.2016.1203503
63. Wang Z, Deng Z, Dahmane N. Telomeric repeat-containing RNA (TERRA) constitutes a nucleoprotein component of extracellular inflammatory exosomes. *Proc Natl Acad Sci USA* 2015;**112**:E6293–300. https://doi.org/10.1073/pnas.1505962112
64. Nassour J, Aguiar LG, Correia A. Telomere-to-mitochondria signalling by ZBP1 mediates replicative crisis. *Nature* 2023;**614**:767–73. https://doi.org/10.1038/s41586-023-05710-8
65. Decker ML, Chavez E, Vulto I *et al.* Telomere length in Hutchinson–Gilford progeria syndrome. *Mech Ageing Dev* 2009;**130**:377–83. https://doi.org/10.1016/j.mad.2009.03.001
66. Li Y, Zhou G, Bruno IG *et al.* Transient introduction of human telomerase mRNA improves hallmarks of progeria cells. *Aging Cell* 2019;**18**:e12979. https://doi.org/10.1111/acel.12979
67. Klein JA, Ackerman SL. Oxidative stress, cell cycle, and neurodegeneration. *J Clin Invest* 2003;**111**:785–93. https://doi.org/10.1172/JCI200318182
68. Ionescu-Tucker A, Cotman CW. Emerging roles of oxidative stress in brain aging and Alzheimer's disease. *Neurobiol Aging* 2021;**107**:86–95. https://doi.org/10.1016/j.neurobiolaging.2021.07.014
69. Akiyama H, Barger S, Barnum S. Inflammation and Alzheimer's disease. *Neurobiol Aging* 2000;**21**:383–421. https://doi.org/10.1016/S0197-4580(00)00124-X
70. Galigniana NM, Charo NL, Uranga R *et al.* Oxidative stress induces transcription of telomeric repeat-containing RNA (TERRA) by engaging PKA signaling and cytoskeleton dynamics. *Biochim Biophys Acta Mol Cell Res* 2020;**1867**:118643. https://doi.org/10.1016/j.bbamcr.2020.118643
71. Hanna R, Flamier A, Barabino A *et al.* G-quadruplexes originating from evolutionary conserved L1 elements interfere with neuronal gene expression in Alzheimer's disease. *Nat Commun* 2021;**12**:1828. https://doi.org/10.1038/s41467-021-22129-9
72. Kovatcheva M, Liao W, Klein ME *et al.* ATRX is a regulator of therapy induced senescence in human cells. *Nat Commun* 2017;**8**:386. https://doi.org/10.1038/s41467-017-00540-5

Received: November 14, 2024. **Revised:** June 1, 2025. **Editorial Decision:** June 3, 2025. **Accepted:** June 12, 2025

doi:10.6342/NTU202504254