Supplemental information for

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

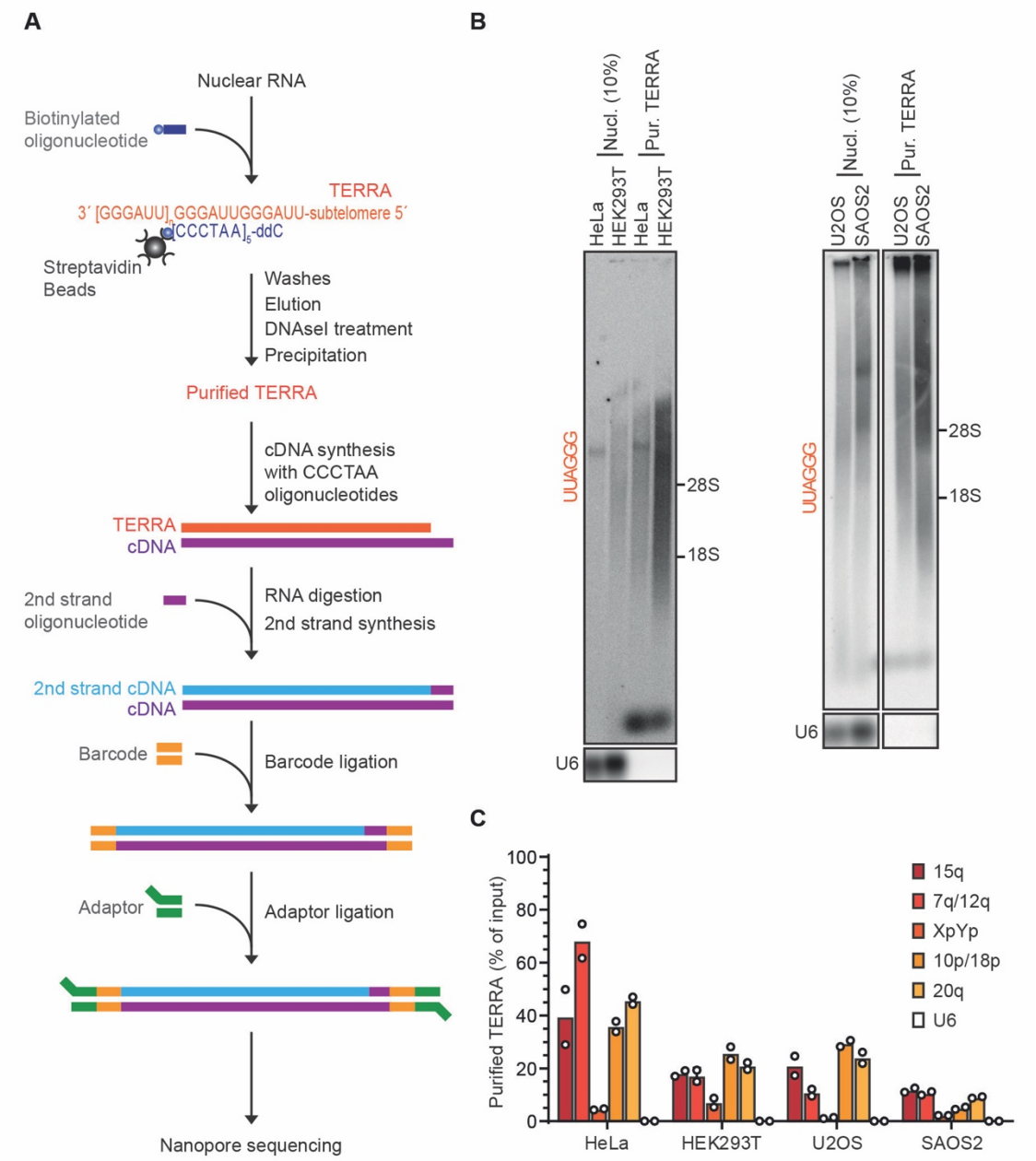
JOANA RODRIGUES,[1] ROBERTA ALFIERI,[2,3] SILVIA BIONE,[2] and CLAUS M. AZZALIN[1]

[1] Instituto de Medicina Molecular João Lobo Antunes (iMM), Faculdade de Medicina da Universidade de Lisboa, Lisbon, 1649-028, Portugal.

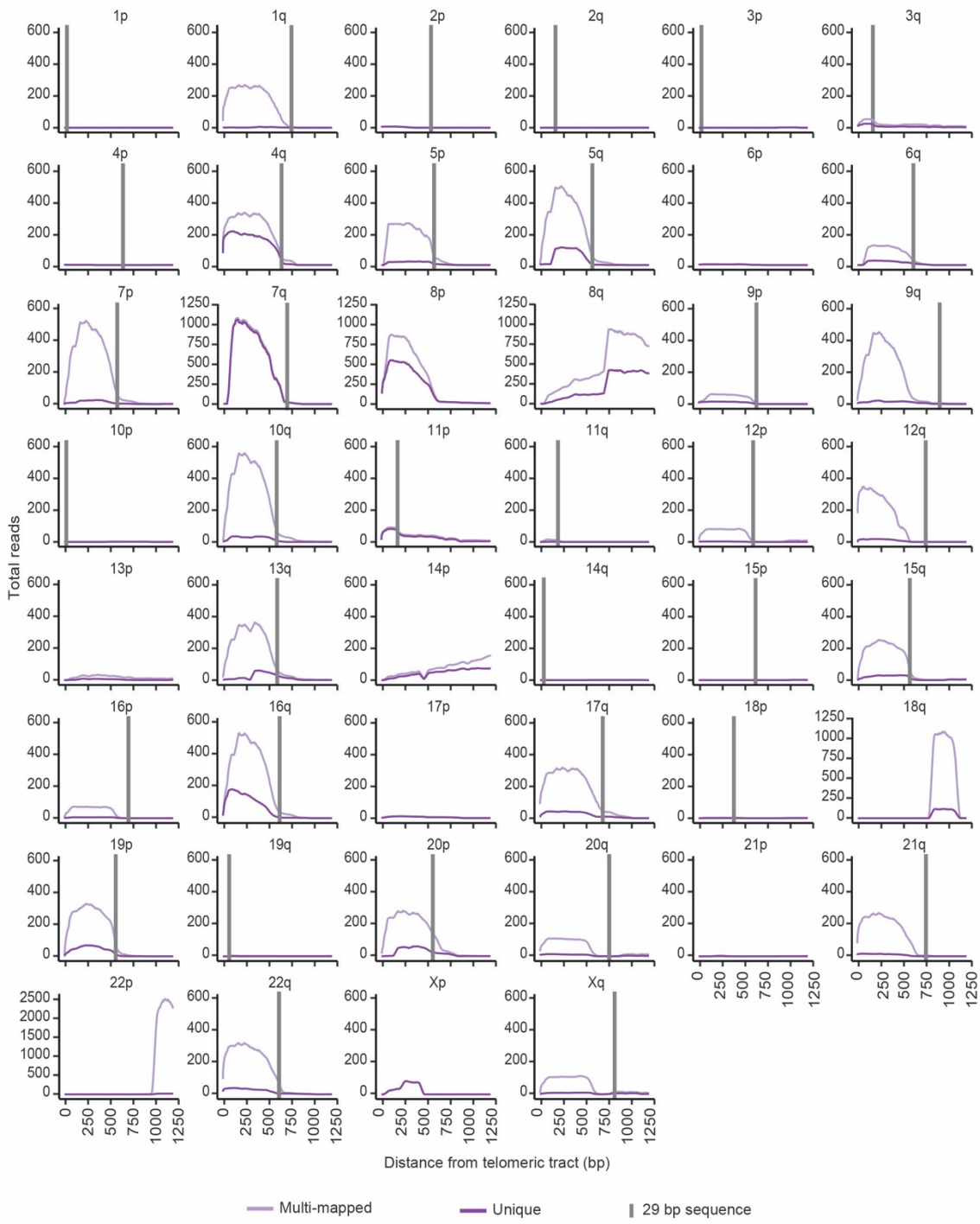
[2] Istituto di Genetica Molecolare Luigi Luca Cavalli-Sforza - Consiglio Nazionale delle Ricerche, Pavia, 27100, Italy

[3] Istituto di Tecnologie Biomediche - Consiglio Nazionale delle Ricerche, Segrate (MI), 20054, Italy

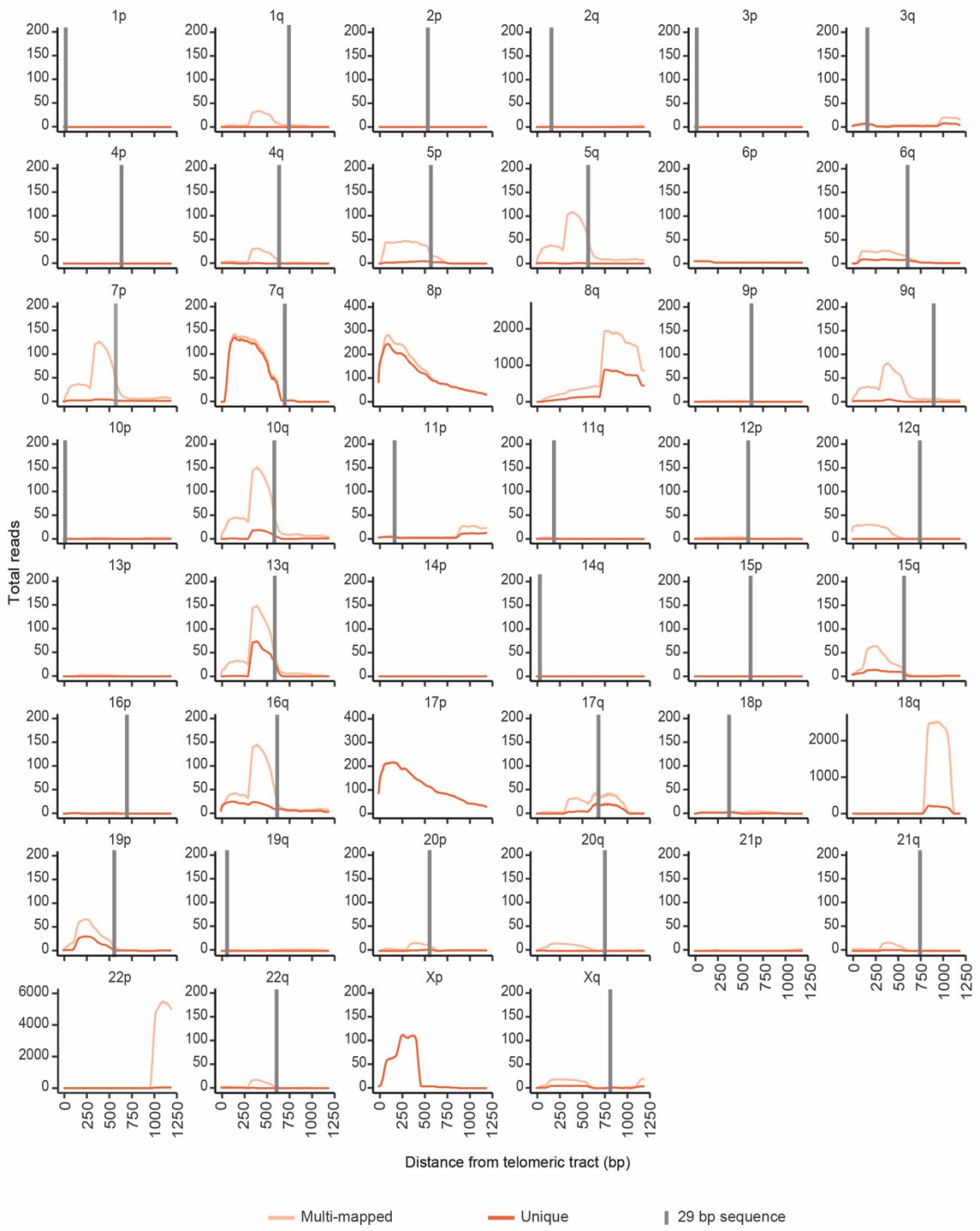
SUPPLEMENTAL FIGURES



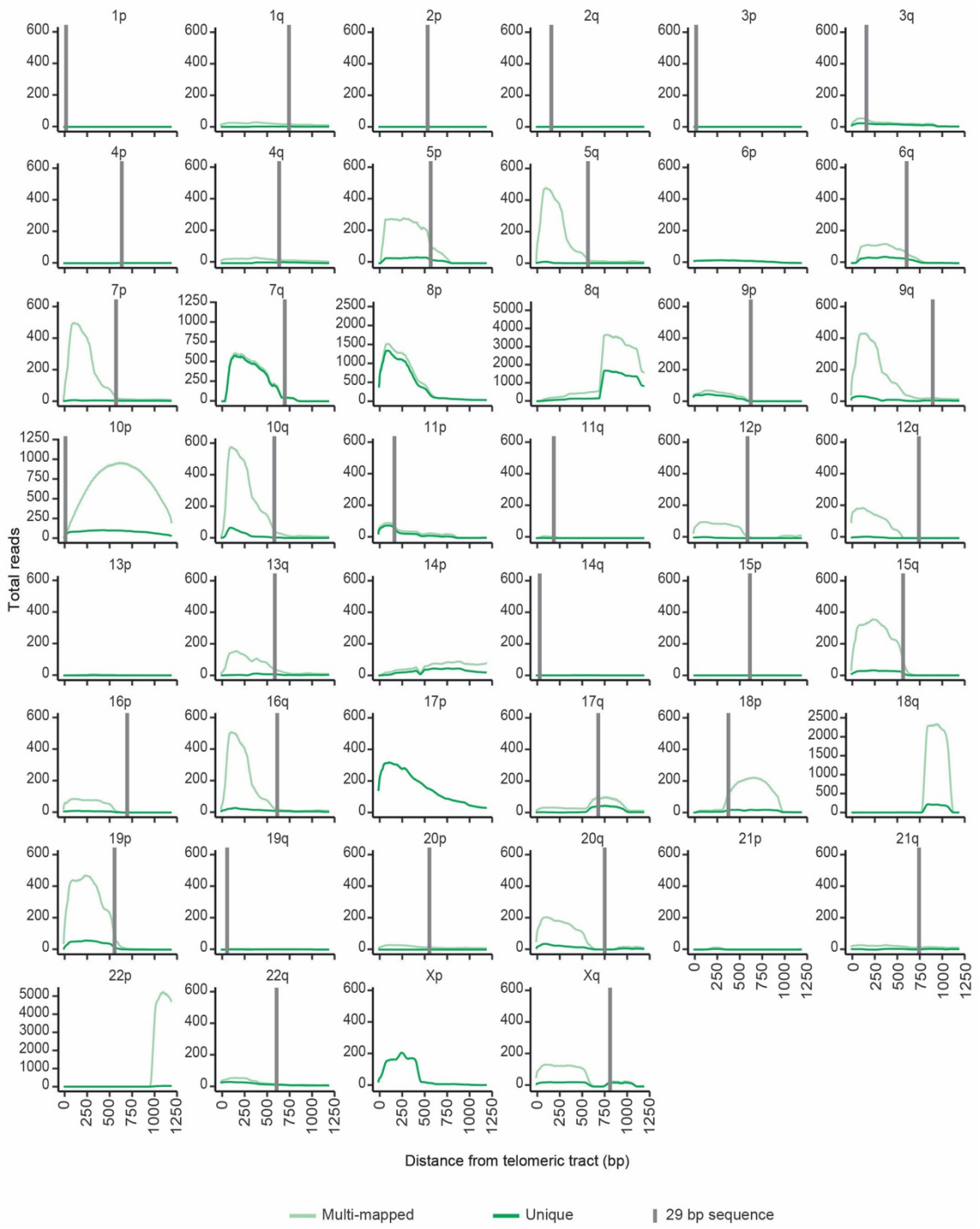
Supplemental Figure S1: TERRA purification and sequencing. (A) Schematic representation of the TERRA purification and ONT sequencing pipeline. (B) TERRA-enriched nuclear RNA from the indicated cell lines was analyzed by northern blotting using a probe complementary to the UUAGGG repeats (total TERRA) or U6. (C) RT-qPCR quantifications of TERRA transcripts or U6 in TERRA-enriched nuclear RNA as in B. Values are expressed as fraction of the input. Bars are the averages from two independent experiments.



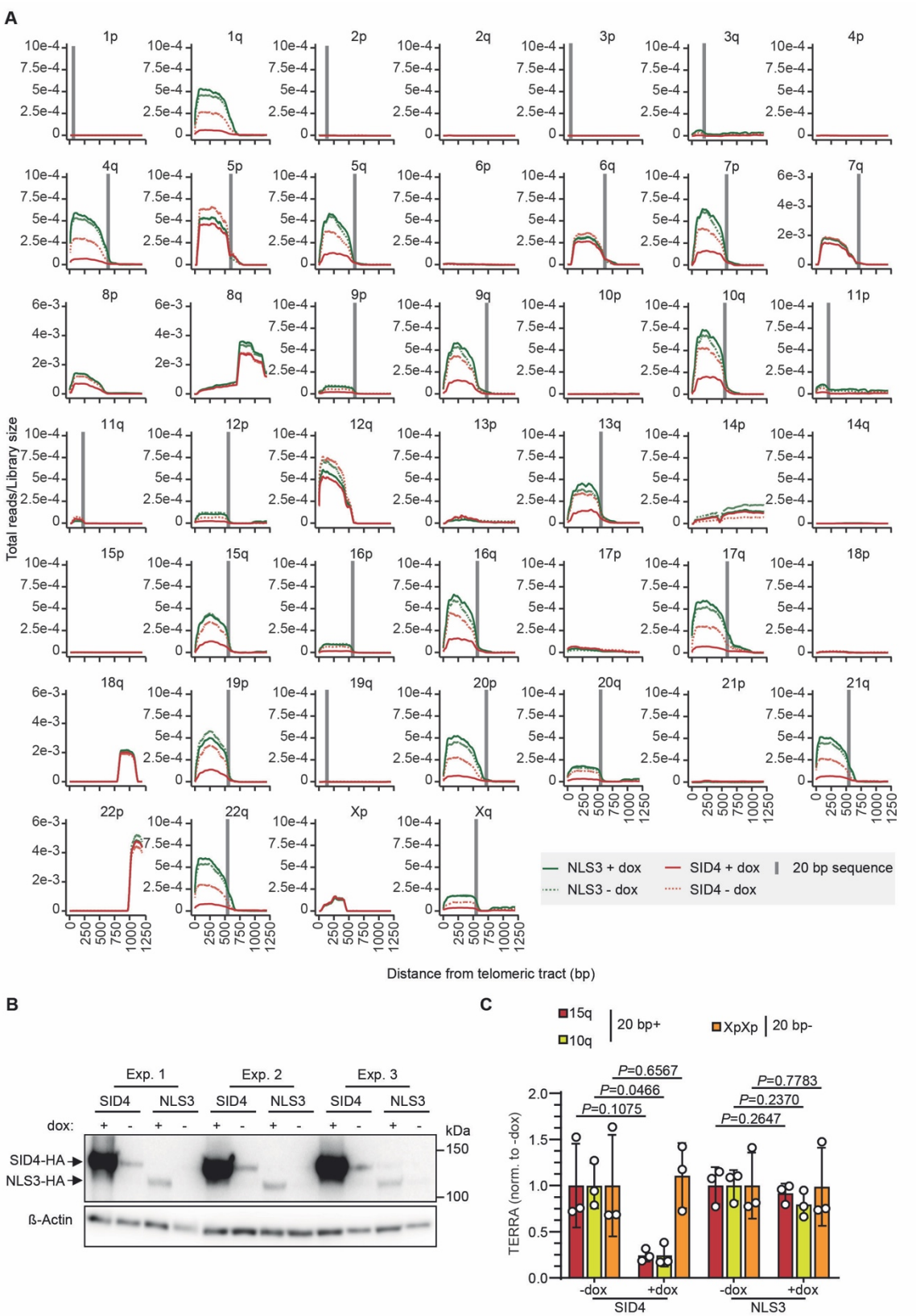
Supplemental Figure S2: TERRA ONTseq mapping for U2OS cells. ONT reads from U2OS cells mapped to the T2T CHM13v2.0/hs1 subtelomeric reference genome. Read coverage results for the most telomere-proximal 1250 bp of subtelomeres are shown. Light and dark purple lines represent multi-mapped and unique reads, respectively. Vertical grey lines indicate the position of the most telomere-proximal 29 bp-like repeat.



Supplemental Figure S3: TERRA ONTseq mapping for HeLa cells. ONT reads from HeLa cells mapped to the T2T CHM13v2.0/hs1 subtelomeric reference genome. Read coverage results for the most telomere-proximal 1250 bp of subtelomeres are shown. Light and dark orange lines represent multi-mapped and unique reads, respectively. Vertical grey lines indicate the position of the most telomere-proximal 29 bp-like repeat.

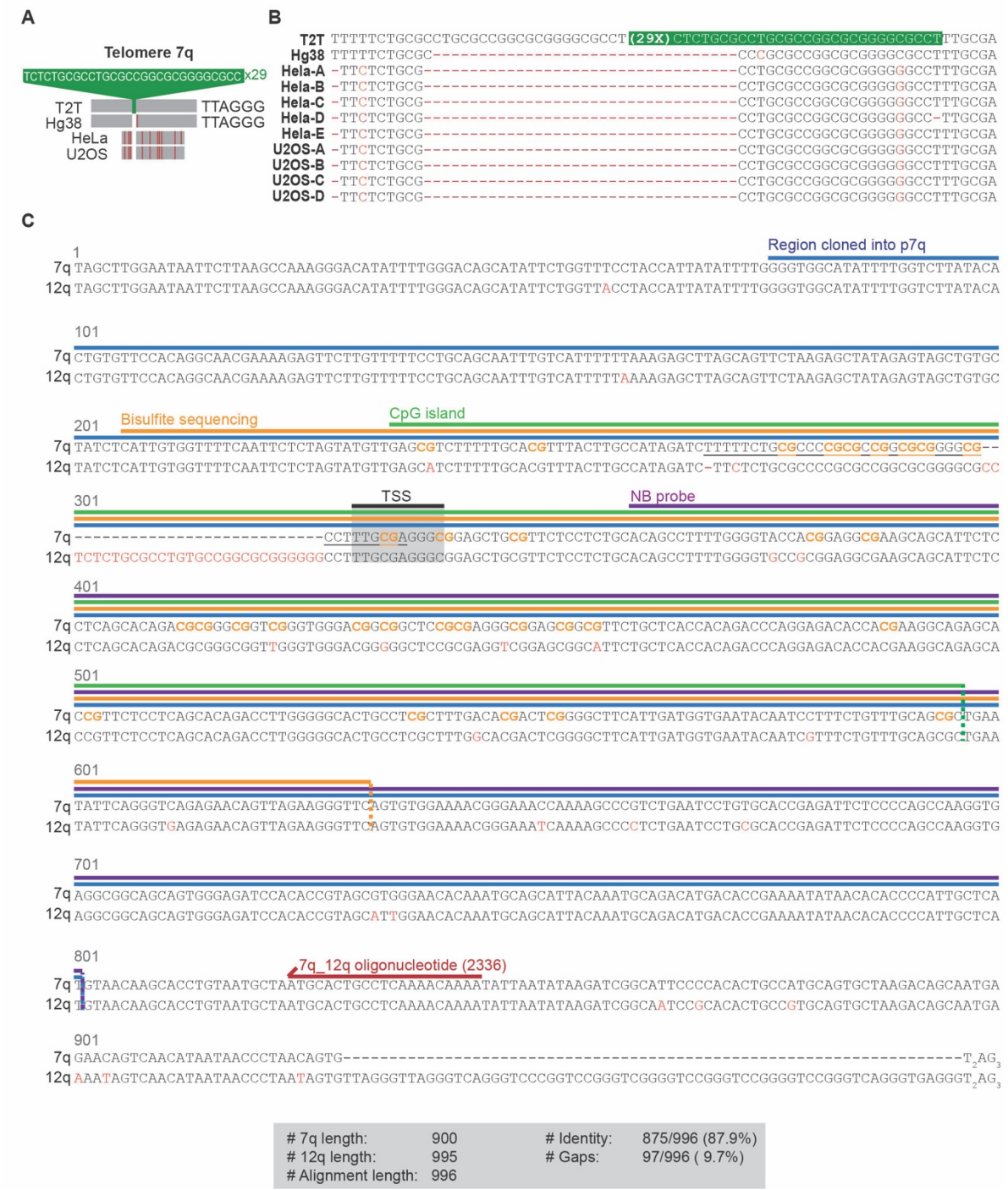


Supplemental Figure S4: HEK293T TERRA ONTseq mapping for HEK293T cells. ONT reads from HEK293T cells mapped to the T2T CHM13v2.0/hs1 subtelomeric reference genome. Read coverage results for the most telomere-proximal 1250 bp of subtelomeres are shown. Light and dark green lines represent multi-mapped and unique reads, respectively. Vertical grey lines indicate the position of the most telomere-proximal 29 bp-like repeat.



Supplemental Figure S5: TERRA ONTseq mapping for T-TALE cell lines. (A) ONT reads from T-TALE cells mapped to the T2T CHM13v2.0/hs1 subtelomeric reference genome. Read coverage results for the most telomere-proximal 1250 bp of subtelomeres are shown. Green lines represent the reads from NLS3 cells treated with dox for 24 h and dotted light green lines from untreated NLS3 cells; red lines represent the reads from SID4 cells treated with dox and dotted light red lines from untreated SID4 cells. Vertical grey lines indicate the position of the most telomere-proximal 20 bp T-TALE target

(100% identity), which is contained within the 29 bp-like repeat TERRA promoters. (B) Western blot analysis of total protein extracts from the same cells as in A using anti-HA antibodies to detect the HA tag fused to the N-terminus of the T-TALEs. Eta (β) Actin serves as loading control. (C) RT-qPCR quantifications of TERRA transcripts from subtelomeres containing (20 bp+; target) or devoid of (20 bp−; non-target) the 20 bp sequence recognized by the T-TALEs in cells as in A. After normalization through U6, values from dox-treated samples (+dox) were expressed as fold change over untreated (-dox). Bars and error bars are averages and SDs from three independent experiments. P values (Student's t test) are indicated.



Supplemental Figure S6: Genomic structure of the 7q subtelomere.

(A) Schematic representation of 7q subtelomere regions from the T2T reference genome (coordinates: chr7:160,563,685-160,564,877), the corresponding region from the GRCh38/hg38 reference genome (coordinates: chr7:159,335,185-159,335,537) and HeLa and U2OS genomic DNA sequenced in this study. The DNA insertion in the T2T is in green and the single mismatches to the T2T sequence are in red. (B) Sequence alignments between a 7q subtelomeric fragment from the T2T reference genome, the same region from the GRCh38/hg38 reference genome and multiple PCR products obtained from HeLa and U2OS genomic DNA. Mismatches to the T2T sequence are in red. The sequence repeated in tandem and present only in the T2T reference sequence is boxed in green. (C) Sequence alignment between a 900 bp long fragment from the subtelomere of chromosome 7q (coordinates chr7:159,334,969-159,335,868 in the GRCh38/hg38 reference genome) and a highly similar sequence from the subtelomere of chromosome 12q (coordinates chr12:133,263,945-133,264,939 in the

GRCh38/hg38 reference genome). Key features are indicated and include the fragment cloned into the p7q reporter plasmid (in blue), the fragment amplified for bisulfite sequencing (orange), the predicted CpG island (green), the TSS identified by TERRA ONTseq (black), the fragment corresponding to the probe used for northern blotting (purple), the 7q_12q oligonucleotide used in RNAseH experiments (red), and the region aligned in B (underlined). The 31 CpG dinucleotides present in the predicted CpG island are in orange.

SUPPLEMENTAL TABLE LEGENDS

Supplemental Table S1. List of primers used in this study.

Supplemental Table S2. Read counts for TERRA ONTseq from U2OS, HeLa, HEK293T, U2OS SID4 (plus and minus dox) and U2OS NLS3 (plus and minus dox) cells.

Supplemental Table S3. 10 bp regions containing putative TERRA transcription start sites derived from TERRA ONTseq.

Supplemental Table S4. List of the different variants of the 29 bp-repeats and comparison of the 29 bp-promoter positions in the GRCh38/hg38 and the T2T CHM13v2.0/hs1 reference genomes.