

1 **Genetic regulation of *TERT* splicing contributes to reduced or elevated cancer risk by**
2 **altering cellular longevity and replicative potential**

3 Oscar Florez-Vargas¹, Michelle Ho¹, Maxwell Hogshead¹, Chia-Han Lee¹, Brenen W
4 Papenberg¹, Kaitlin Forsythe¹, Kristine Jones², Wen Luo², Kedest Teshome², Cornelis
5 Blauwendraat³, Kimberly J Billingsley³, Mikhail Kolmogorov⁴, Melissa Meredith⁵, Benedict
6 Paten⁵, Raj Chari⁶, Chi Zhang², John S. Schneekloth⁷, Mitchell J Machiela⁸, Stephen J Chanock⁹,
7 Shahinaz Gadalla¹⁰, Sharon A Savage¹⁰, Sam M Mbulaiteye¹¹, Ludmila Prokunina-Olsson^{1, #}

8
9 ¹Laboratory of Translational Genomics, DCEG, National Cancer Institute, Rockville, MD, USA,
10 ²Cancer Genomic Research Laboratory, Leidos Biomedical Research, Frederick National
11 Laboratory for Cancer Research, Frederick, MD, USA, ³Center for Alzheimer's and Related
12 Dementias, National Institute of Aging and National Institute of Neurological Disorders and
13 Stroke, Bethesda, MD, USA, ⁴Cancer Data Science Laboratory, CCR, National Cancer Institute,
14 Bethesda, MD, USA, ⁵UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA, ⁶Genome
15 Modification Core, Laboratory Animal Sciences Program, Leidos Biomedical Research,
16 Frederick National Laboratory for Cancer Research, Frederick, MD, USA, ⁷Chemical Biology
17 Laboratory, CCR, National Cancer Institute, Frederick, MD, USA, ⁸Integrative Tumor
18 Epidemiology Branch, DCEG, National Cancer Institute, Rockville, MD, USA, ⁹Laboratory of
19 Genetic Susceptibility, DCEG, National Cancer Institute, Rockville, MD, USA, ¹⁰Clinical
20 Genetics Branch, DCEG, National Cancer Institute, Rockville, MD, USA, ¹¹Infections and
21 Immunoepidemiology Branch, DCEG, National Cancer Institute, Rockville, MD, USA

22
23 # corresponding author

24 Ludmila Prokunina-Olsson, PhD
25 Laboratory of Translational Genomics,
26 Division of Cancer Epidemiology and Genetics,
27 National Cancer Institute,
28 9615 Medical Center Dr,
29 Rockville, MD, 20850, USA
30 prokuninal@mail.nih.gov

31
32
33
34
35
36
37

38 **ABSTRACT**

39 The chromosome 5p15.33 region, which encodes telomerase reverse transcriptase (*TERT*),
40 harbors multiple germline variants identified by genome-wide association studies (GWAS) as
41 risk for some cancers but protective for others. We characterized a variable number tandem
42 repeat within *TERT* intron 6 (VNTR6-1, 38-bp repeat unit) and observed a strong association
43 between VNTR6-1 alleles (Short: 24-27 repeats, Long: 40.5-66.5 repeats) and GWAS signals
44 within *TERT* intron 4. Specifically, VNTR6-1 fully explained the GWAS signals for rs2242652
45 and partially for rs10069690. VNTR6-1, rs10069690 and their haplotypes were associated with
46 multi-cancer risk and age-related telomere shortening. Both variants reduce *TERT* expression
47 through alternative splicing and nonsense-mediated decay: rs10069690-T increases intron 4
48 retention and VNTR6-1-Long expands a polymorphic G quadruplex (G4, 35-113 copies) within
49 intron 6. Treatment with G4-stabilizing ligands decreased the fraction of the functional
50 telomerase-encoding *TERT* full-length isoform, whereas CRISPR/Cas9 deletion of VNTR6-1
51 increased this fraction and apoptosis while reducing cell proliferation. Thus, VNTR6-1 and
52 rs10069690 regulate the expression and splicing of *TERT* transcripts encoding both functional
53 and nonfunctional telomerase. Altered *TERT* isoform ratios might modulate cellular longevity
54 and replicative potential at homeostasis and in response to environmental factors, thus selectively
55 contributing to the reduced or elevated cancer risk conferred by this locus.

56

57 **INTRODUCTION**

58 *TERT* encodes the catalytic subunit of telomerase, a reverse transcriptase that extends telomeric
59 repeats at chromosome ends, ensuring the maintenance of telomere length, genome integrity, and
60 cell proliferation¹. Telomere dysfunction has been implicated in many human diseases². At least
61 ten independent pleiotropic multi-cancer GWAS signals within the ~100 kb genomic region on
62 chromosome 5p15.33 harboring *TERT* and *CLPTMIL* have been associated with cancer risk or
63 protection³⁻⁶. The associated variants might be causal or tag some known or yet unknown
64 functional polymorphisms. Identifying these variants and the mechanisms underlying their
65 associations may improve the understanding of the etiology and biological mechanisms of these
66 cancers, leading to optimized cancer risk prediction, prevention, and treatment.

67 Several variable number tandem repeats (VNTRs) have been reported within the 5p15.33
68 region^{7,8} but only minimally characterized due to their high variability, complexity, and length of
69 genomic fragments extended by repeat copies. Recent advances in long-read genome sequencing
70 and assembly⁹ have closed many genomic gaps and facilitated the discovery and exploration of
71 complex regions, such as VNTRs, for which copy numbers analysis using short-read sequencing
72 or PCR-based methods is challenging. Recent examples¹⁰ have shown that VNTRs might
73 account for or contribute to GWAS signals for cancer and other human traits, expanding the list
74 of previously unknown functional genetic variants to be explored.

75 We hypothesized that VNTRs might be responsible for some of the reported GWAS signals
76 within the *TERT* region. Here, we explored two VNTRs within *TERT* intron 6 in relation to the
77 cancer-related GWAS signals reported in this region. Among those signals, a strong association
78 was observed only between VNTR6-1 and two single nucleotide polymorphisms (SNPs) -
79 rs2242652 and rs10069690 - within *TERT* intron 4. Specifically, VNTR6-1 Long alleles (40.5-
80 66.5 repeats) in contrast to Short alleles (24-27 repeats) were preferentially linked with
81 rs2242652-A and rs10069690-T alleles, both of which are associated with a reduced risk of
82 bladder⁶ and prostate cancer¹¹ but an elevated risk of glioma¹², breast cancer^{13,14} and ovarian
83 cancer¹⁵. We present a comprehensive genetic and functional analysis of VNTR6-1 and GWAS
84 signals within this region. These results provide new insights into the etiology and genetic
85 susceptibility of multiple cancers and telomere biology.

86

87 RESULTS

88 VNTR6-1 is linked with GWAS leads rs2242652 and rs10069690

89 We explored two previously reported but only minimally characterized VNTRs^{7,8} within the
90 *TERT* intron 6 in relation to all cancer-related GWAS signals within the multi-cancer 5p15.33
91 region³⁻⁶. For this, we analyzed 452 phased long-read genome assemblies from 226 controls of
92 diverse ancestries generated by the Human Pangenome Reference Consortium (HPRC)⁹ and the
93 Center for Alzheimer's and Related Dementias (CARD)¹⁶. The strongest associations were
94 detected for VNTR6-1 (38-bp repeat unit, range 24-66.5 copies in the assemblies). Specifically,
95 more VNTR6-1 copies were detected in assemblies with alleles of *TERT* intron 4 SNPs -
96 rs2242652-A ($p=5.93E-19$) and rs10069690-T ($p=5.40E-11$) compared with the alternative
97 alleles at these SNPs (**Figure 1a, b, Figure S1, Table S1**). In contrast, the copies of VNTR6-2
98 (36-bp repeat unit, range 8-155 copies in the assemblies) were only moderately associated with
99 rs2242652-A allele ($p=7.66E-04$) but not with rs10069690 ($p=0.84$, **Figure 1c**), or other GWAS
100 signals (**Table S1**). Thus, we focused on VNTR6-1 as a potential functional proxy for GWAS
101 leads rs2242652 and rs10069690.

102 We performed targeted PacBio sequencing of the VNTR6-1 amplicon (2126-3750 bp) in various
103 samples (**Table S2**). This analysis confirmed concordance in repeat scoring in targeted vs.
104 whole-genome long-read sequencing of 5 HPRC controls with available long-read assemblies⁹,
105 in targeted sequencing of 5 pairs of tumor vs. tumor-adjacent normal bladder tissues, as well as
106 Mendelian segregation in HapMap samples from 30 European (Central European from Utah,
107 CEU) and 30 African (Yoruba, YRI) family trios. Two SNPs, rs56345976 and rs33961405
108 SNPs, were covered by the PacBio amplicon and their genotypes perfectly matched those
109 determined by long-read genome assemblies and our TaqMan genotyping confirming that
110 VNTR6-1 is a stable germline polymorphism that can be confidently genotyped by long-read
111 sequencing (**Table S2**). Despite its reliability, long-read sequencing remains an expensive,
112 laborious, and low-throughput method that requires a significant amount of high-quality DNA.

113 The availability of more convenient approaches to analyze VNTR6-1 would facilitate its testing
114 in association studies.

115 We noted that both in assemblies and in HapMap samples with targeted PacBio sequencing data
116 (**Figure 1, Table S2**), the VNTR6-1 alleles clustered in 2 main groups. In HapMap samples,
117 these groups comprised Short alleles (CEU: 25.8 ± 1.0 copies, 83.3% and YRI: 27.0 ± 2.0 copies,
118 80%) and Long alleles (CEU: 16.7%, 40.5 ± 0 copies, and YRI: 43.75 ± 8.8 copies, 20%). The
119 Long alleles included an uncommon allele detected in YRI only: 66.5 ± 0 copies, 2.5% (**Table**
120 **S2**). Hypothesizing that short-read sequencing might help with VNTR analysis, we examined the
121 genomic depth profiles of aligned short reads generated by whole-genome sequencing in all
122 3,201 individuals of diverse ancestries from the 1000 Genomes Project (1000G). The reference
123 human genome has the Short VNTR6-1 allele (27 copies) and in 1000G samples carrying only
124 the Short alleles (24-27 copies based on HPRC assemblies or targeted PacBio), the genomic
125 profiles were flat. In contrast, noticeable read pileups were observed in 1000G samples with at
126 least one copy of the Long allele (40.5 or 66.5 copies); however, the Long alleles and their
127 heterozygous or homozygous state could not be distinguished based on the genomic profiles
128 (**Figure S2**). We applied machine-learning methods to evaluate short-read profiles and classify
129 all the 1000G samples as carriers of at least one copy of the VNTR6-1-Long allele (Long/any
130 genotype) vs. the Short/Short genotype (**Table S3**). By treating the classification based on
131 genomic profiles as true genotypes, we performed random forest analysis of the 1000G samples
132 using all SNPs within the 400 kb genomic region (GRCh38 chr5:1,100,000-1,500,000) derived
133 from short-read whole-genome sequencing and identified two SNPs, rs56345976 and
134 rs33961405, that most effectively predicted VNTR6-1 groups in all populations despite
135 differences in LD profiles (**Figure S3, Table S4, Table S5**). Although these SNPs were not
136 sufficiently informative on their own, the rs56345976-A/rs33961405-G haplotype separated the
137 carriers with VNTR6-1-Long alleles (40.5 or 66.5 copies) from the carriers of the VNTR6-1-
138 Short/Short genotypes based on three other haplotypes (AA, GG, and GA) (**Figure S4, Table**
139 **S3**).

140 The classification of all 1000G samples into groups with VNTR6-1 Short/Short and Long/any
141 genotypes based on the rs56345976/rs33961405 haplotypes matched the scoring based on
142 genomic profiles, with an area under the curve (AUC) of 0.98 (**Figure S4**). The VNTR6-1
143 scoring based on rs56345976/rs33961405 haplotypes was concordant with the repeat sizes
144 determined by assemblies and targeted PacBio sequencing (**Figure S5, Table S6**). To test
145 whether the VNTR6-1 can be confidently imputed, we created a custom reference panel of the
146 region (400 kb) in all 3,201 samples from the 1000G dataset. VNTR6-1 was incorporated into
147 this panel as a biallelic marker with Short/Long alleles determined based on phased
148 rs56345976/rs33961405 haplotypes (**Table S3**). We randomly split the 1000G dataset into two
149 equal groups and used the first group as a reference for imputing VNTR6-1 in the second group.
150 The imputation showed a 99.3% concordance with the predetermined VNTR6-1 genotypes,
151 further supporting the robustness of the VNTR6-1 allele assignment (Short/Long) based on

152 rs56345976/rs33961405 haplotypes. These results demonstrated that VNTR6-1 could be used as
153 a germline biallelic marker and confidently imputed into datasets with good coverage of the
154 region through whole-genome sequencing or array genotyping and imputation. Because
155 rs56345976/rs33961405 can be genotyped individually by targeted assays (such as by TaqMan
156 assays), the VNTR6-1 genotypes can be inferred in any samples, even without genome-wide
157 sequencing or genotyping data. In a subset of 1000G samples from Europeans (1000G-EUR), the
158 VNTR6-1-Long allele was most strongly linked with rs2242652-A ($r^2=0.62$) and rs10069690-T
159 ($r^2=0.48$, **Table S5, Figure S6**), suggesting that it might contribute to the associations detected
160 for the GWAS signals for rs2242652 and rs10069690.

161 **VNTR6-1 creates an expandable G quadruplex that modulates *TERT* splicing**

162 VNTR6-1 is located ~3.5 kb upstream of *TERT* exon 7 (**Figure 2a**). The simultaneous inclusion
163 or skipping of exons 7 and 8 defines the *TERT* full-length (*TERT-FL*) or *TERT-β* isoform,
164 respectively¹⁷. To assess the functional effect of VNTR6-1 on *TERT* splicing, we deleted the
165 entire VNTR6-1 region (2,241 bp in the reference human genome) by CRISPR/Cas9 editing.
166 Partial deletion of this highly repetitive genomic region was technically impossible. We
167 established three stable isogenic knockout clones (V6.1-KO) in UMUC3, a bladder cancer cell
168 line with high *TERT* expression (DepMap transcripts per million [TPM]=6.78 (**Figure 2a,**
169 **Figure S7a-b**) and two clones in A549, a lung cancer cell line with moderate *TERT* expression
170 (TPM=3.63, **Figure S8a**). Deletion of VNTR6-1 increased the inclusion of exons 7 and 8,
171 shifting the ratio of *TERT-FL* from ~45% to 71% in UMUC3 (**Figure 2b, 2e, Figure S7c-f**) and
172 from 34% to 49% in A549 (**Figure S8b-c**). These results suggest that VNTR6-1 acts as a
173 splicing switch between the *TERT-FL* (expressed at a higher fraction in V6.1-KO cells) and
174 *TERT-β* (expressed at a higher fraction in WT cells).

175 We found no evidence of differential DNA methylation (**Figure S9**) or long-range chromatin
176 interactions (**Figure S10**) involving the VNTR6-1 region. However, we noted a high G content
177 within the 38-bp consensus repeat sequence of VNTR6-1 (5'-
178 GGTGGGGATCTGTGGGATTGGTTTTTCATGTGTGGGGTA-3'). Based on G4Hunter
179 analysis and G4-ChiP-seq, we predicted that VNTR6-1 could adopt a G quadruplex (G4)
180 structure in the *TERT*-sense orientation, creating 35-113 G4s per allele with conserved core G-
181 containing motifs (**Figure 2a, Figure S11a, b**).

182 As a single invariable G4 upstream of VNTR6-1 has been implicated in *TERT-β* splicing¹⁸, we
183 hypothesized that the variation in the number of G4s, created by VNTR6-1-Short vs. Long
184 alleles, could affect splicing and the *TERT-FL*:*TERT-β* isoform ratio. We treated our isogenic
185 UMUC3 and A549 cell lines—WT, representative of the VNTR6-1-Long allele (**Figure 2c, 2f,**
186 **Figure S11c-d** for UMUC3; **Figure S8d, e, h, i** for A549) and V6.1-KO, representative of
187 VNTR6-1-Short allele (**Figure 2d, g, Figure S11e-f** for UMUC3; **Figure S8f, g, j, k** for A549),
188 —with several G4-stabilizing ligands. In cDNA from the treated and untreated cells, we
189 quantified the expression of exons 6-9 (*TERT-β*) and 7-8 (*TERT-FL*) and total *TERT*. Treatment

190 with G4 ligands Pidnarulex (CX5461)¹⁹ or PhenDC3²⁰ decreased the *TERT-FL* fraction while
191 increasing the *TERT-β* fraction in both the WT and V6.1-KO cell lines, likely by stabilizing
192 VNTR6-1-G4s (**Figure 2f, g**). In UMUC3, CX5461 increased the *TERT-β* fraction 4.7-fold in
193 the WT cells compared to 2.9-fold in the V6.1-KO cells (**Figure S11c, S11e**). PhenDC3 also
194 significantly increased total *TERT* expression in WT cells, whereas no changes were observed in
195 V6.1-KO (**Figure S11d, f**). These results suggest that while invariable G4s in intron 6 might
196 influence the baseline level of *TERT* splicing, the G4s formed by VNTR6-1 further modulate
197 these splicing ratios and total *TERT* expression. A novel splicing isoform with exon 8 skipping
198 (*TERT-Δ8*, **Figure S12**) was observed in V6.1-KO and WT cells after ligand treatment.

199 **rs10069690-T and VNTR6-1-Long alleles affect *TERT* expression and splicing**

200 *TERT* expression is generally low in normal human tissues (The Genotype-Tissue Expression
201 (GTEx) Project, median TPM=0.00-2.73) and is not associated with the GWAS leads rs2242652
202 and rs10069690 (**Table S7**). However, *TERT* expression is generally higher in tumors (The
203 Cancer Genome Atlas (TCGA), median TPM=0.02-5.71; **Table S7**) and is associated with these
204 SNPs in some tumor types (kidney chromophobe, KICH and head and neck squamous
205 carcinoma, HNSC; **Table S7**). We detected high *TERT* expression (mean TPM=59.7, **Figure 3a**,
206 **Table S8**) in a set of 78 Burkitt lymphoma (BL) tumors²¹. BL is an aggressive pediatric cancer
207 originating from germinal center B cells, in which high *TERT* expression is necessary for the
208 longevity of memory B cells²². Two hotspot somatic mutations in the *TERT* promoter, C228T (-
209 124 bp) and C250T (-146 bp), upregulate *TERT* expression in many tumors^{23,24}, but these
210 mutations are absent in non-Hodgkin lymphomas, including BL²⁵ and our set of BL tumors²⁵.
211 The combination of high *TERT* expression in the absence of upregulating promoter mutations in
212 BL tumors provides an opportunity to explore the regulation of *TERT* expression by germline
213 variants.

214 In BL tumors (**Table S8**), *TERT* expression decreased with the rs10069690-T allele ($\beta=-13.95$,
215 $p=0.035$; **Figure 3b**) but not with the rs2242652-A allele ($\beta=2.53$, $p=0.83$; **Figure 3c**), with a
216 suggestive trend for decreased *TERT* expression associated with the VNTR6-1-Long allele ($\beta=-$
217 16.97 , $p=0.053$; **Figure 3d**). These variants are in high LD in 1000G-EUR but in low LD in
218 1000G-AFR and our set of BL tumors (88% from African patients, **Figure S13**), suggesting
219 independent effects of rs10069690 and VNTR6-1 on *TERT* expression. Based on the LD profiles
220 and association with *TERT* expression in BL tumors, we functionally prioritized rs10069690 and
221 VNTR6-1 for further analyses.

222 The functional role of the rs10069690-T allele has previously been attributed to the creation of
223 an alternative splicing site in the *TERT* intron 4, resulting in the coproduction of telomerase-
224 functional *TERT-FL* and a truncated telomerase-nonfunctional INS1b isoform²⁶. However, due
225 to the low *TERT* expression in most human tissues, this relationship has not been explored in
226 relation to genetic variants in the 5p15.33 region²⁶. In BL tumors, 26.2% of all RNA-seq reads
227 between exons 4 and 5 were retained within intron 4, in contrast with neighboring introns 3 and 5

228 (with 10.5% and 8.1% of the retained reads, respectively) (**Figure S14a**). The rate of *TERT*
229 intron 4 retention was more statistically significantly associated with rs10069690 ($p=5.36E-09$,
230 **Figure S14b**) than with rs2242652 ($p=5.0E-03$, **Figure S14c**). We analyzed four splicing events
231 between exons 4 and 5, one with canonical intron 4 splicing and three with intron 4 retention
232 (isoforms *INSI*^{17,26}, *INSIb*²⁶ and with unspliced intron 4, **Figure S15a-d**). The fraction of
233 canonical intron 4 splicing decreased (68.3%, 63.8% and 57.3% of the reads in the rs10069690-
234 CC, CT and TT genotype groups, respectively; $p=1.65E-05$; **Figure S15e**). These results were
235 consistent with previous observations on the association of the rs10069690-T allele with *INSIb*-
236 type splicing²⁶; in BL tumors, the fraction of *INSIb* splicing increased (0% to 3.8% and 7.0% of
237 all reads between exons 4 and 5 in the rs10069690-CC, CT and TT genotype groups,
238 respectively; $p=3.07E-09$; **Figure S15e**). The fraction of *INSI* was decreased, and that of
239 unspliced intron 4 (excluding reads for *INSI* and *INSIb* isoforms) was increased with the
240 rs10069690-T allele (**Figure S15h**). These results suggest that *INSI*- and *INSIb*-type splicing is
241 minor and could be secondary to intron retention, which increases with the rs10069690-T allele.

242 Several other common *TERT* isoforms have been reported¹⁷ (**Figure S16**). The *TERT- α* isoform
243 involves in-frame 36-bp skipping within exon 6 ($\Delta 6_{(1-36)}$), causing partial loss of the reverse
244 transcriptase domain¹⁷. As discussed above, *TERT- β* ($\Delta 7-8$)¹⁷ results from the simultaneous
245 skipping of exons 7 and 8 (182 bp), terminating the frameshifted protein in exon 10.
246 Additionally, *TERT- $\alpha\beta$* results from concurrent $\Delta 6_{(1-36)}$ and $\Delta 7-8$ splicing events. The expression
247 of these *TERT* isoforms was not significantly associated with rs10069690, rs2242652, or
248 VNTR6-1 in BL tumors (**Table S8**). Transcripts truncated by premature termination codons
249 (**Figure S16**), including *INS* (truncated within exon 5), *INSIb* (intron 4), and *TERT- β* or *TERT- $\alpha\beta$*
250 (*exon 10*), are likely to be eliminated by nonsense-mediated decay (NMD), reducing total
251 *TERT* expression. Escaping NMD would result in alternative TERT proteins without telomerase
252 activity but still binding the telomerase RNA component (hTR), thus producing dominant-
253 negative competitors of the telomerase-functional TERT-FL.

254 Due to premature termination codons (in intron 4 for *INSIb* and in exon 10 for *TERT- β*), both
255 rs10069690 and VNTR6-1 increase the fraction of NMD-targeted transcripts encoding
256 telomerase-nonfunctional proteins, decreasing total *TERT* expression and the fraction of the
257 telomerase-encoding *TERT-FL* isoform. To assess the combined effects of these variants, we
258 analyzed *TERT* expression based on the VNTR6-1/rs10069690 haplotypes (**Figure 3e**).
259 Compared to the Short-C haplotype, the *TERT* expression was decreased by the Short-T ($\beta=-$
260 12.2, $p=0.10$) and Long-C ($\beta=-15.92$, $p=0.36$) haplotypes, with a greater decrease occurring
261 when both the VNTR6-1-Long and rs10069690-T alleles were included in the same haplotype
262 (Long-T, $\beta=-24.18$, $p=0.027$, **Figure 3e**, **Table S8**). Thus, the decrease in total *TERT* expression
263 and increase in alternative isoforms through different splicing events are independently
264 contributed by both alleles, VNTR6-1-Long and rs10069690-T, with a stronger effect when these
265 events are combined.

266 **VNTR6-1 contributes to proliferative and anti-apoptotic responses to external stimuli**

267 Differences in the activities and ratios of TERT-FL to TERT- β isoforms could contribute to
268 variability in intracellular dynamics and alter cell phenotypes. We compared cell proliferation
269 rates of WT and V6.1-KO UMUC3 cell lines by two methods: a real-time cellular impedance
270 system and a flow cytometry assay based on the dilution of an intracellular dye. To mimic some
271 potentially relevant environmental exposures, cell lines were cultured in media supplemented
272 with either full fetal bovine serum (full serum) or charcoal-stripped serum (CS serum, depleted
273 of hormones and growth factors), (**Figure 4a-c**). Under both serum conditions, the V6.1-KO
274 cells grew slower than WT cells. While both cell lines grew faster in full serum than in CS
275 serum, the increase in proliferation stimulated by full serum was significantly lower in V6.1-KO
276 than that of WT cells (**Figure 4a, Table S9**). Proliferation rate differences between WT and
277 V6.1-KO cells were also apparent after a short period (24 hours) of serum starvation prior to
278 seeding (**Figure S17c-d**), further suggesting that VNTR6-1 functions are sensitive to
279 environmental signals in culture media. Continuous culturing resulted in an increased *TERT- β*
280 fraction (**Figure S17e-h**), potentially reducing proliferation due to limited nutrients as the cells
281 reached confluency. Total *TERT* expression was not affected by the increase in cell density in
282 WT cells but decreased in V6.1-KO cells (**Figure S17f, h**). This finding suggested that VNTR6-
283 1 might modulate cellular proliferation in response to environmental signals, including hormones
284 and growth factors, and cell density.

285 Additionally, flow-based apoptosis analyses demonstrated a significant increase in the
286 percentage of apoptotic cells in V6.1-KOs compared to WT cells. This was evident when cells
287 were in full serum and stimulated to grow or upon cisplatin treatment, which induces cell death
288 (**Figure 4d, 4e**). RNA-seq analysis also demonstrated enrichment in both apoptosis and
289 proliferation pathways in V6.1-KO compared to WT cells (**Figure 4f, Table S10, S11**).

290 TERT- β is a dominant-negative competitor of TERT-FL for telomerase function²⁷, but the
291 contributions of these isoforms to cell proliferation are less clear. We monitored cell proliferation
292 after transient overexpression of the *TERT-FL* and *TERT- β* isoforms in 5637 cells, a bladder
293 cancer cell line with low *TERT* expression (DepMap TPM=1.23). Overexpression of both
294 isoforms similarly increased cell proliferation compared to that of the GFP control (**Figure S18,**
295 **Table S12**). In the co-transfection experiments, cell proliferation increased the most at a 50:50%
296 *TERT-FL:TERT- β* ratio, followed by at a 20:80% ratio (**Figure S18, Table S12**). The 50:50%
297 ratio appears to be optimal, potentially because it offers a balance between promoting
298 proliferation and reducing cell death.

299 Structured illumination microscopy fluorescence imaging of A549 cells (a cell line with a large
300 cytoplasm, facilitating visualization) co-transfected with both TERT isoforms revealed stronger
301 mitochondrial co-localization for TERT- β than for TERT-FL (**Figure 5a, Figure S19**). These
302 results further suggest the role of TERT- β in mitochondrial-localized processes, one of which
303 may be protection from apoptosis. The effects on proliferation and apoptosis achieved by
304 VNTR6-1 knock-out (**Figure 5b**) suggest that the ratio of TERT-FL to TERT- β may affect both

305 cellular longevity (by protecting cells from apoptosis) and replicative potential (by altering
306 proliferation in response to environmental conditions).

307 **VNTR6-1, rs10069690 or their haplotypes account for pleiotropic cancer GWAS** 308 **associations**

309 Because we established that VNTR6-1 is linked with GWAS leads rs2242652-A ($r^2=0.62$) and
310 rs10069690-T ($r^2=0.48$) in the 1000G-EUR populations, we next sought to compare the
311 associations of these markers with cancer risk. Having validated the rs56345976/rs33961405
312 haplotypes as a confident predictor of VNTR6-1 Short vs Long alleles (**Table S3, S4, Figure S3,**
313 **S4**), we used these haplotypes to infer VNTR6-1 alleles in various sets. Specifically, we inferred
314 VNTR6-1 and the composite marker VNTR6-1/rs10069690 because it captured functional
315 effects from both variants. Using these markers, we performed association analyses in
316 individuals of European ancestry from the Prostate, Lung, Colorectal and Ovarian (PLCO)
317 cohort of cancer-free controls ($n=73,085$) and 29,623 patients with 16 cancer types²⁸. The PLCO
318 association results for the VNTR6-1-Long allele and VNTR6-1/rs10069690 were comparable to
319 those for the rs10069690-T and rs2242652-A alleles; these alleles were associated with a
320 reduced risk of bladder and prostate cancer but an elevated risk of breast, endometrial, ovarian,
321 and thyroid cancer and glioma (**Figure 6a, Table S13**). Only for prostate cancer the association
322 was stronger for rs10069690 than for VNTR6-1 and rs22942652, perhaps suggesting a more
323 important role of decreased *TERT* expression due to intron 4 retention and *INS1b*-splicing in the
324 molecular mechanism of this cancer. Compared to the reference Short-C haplotype, the strongest
325 positive or negative cancer-specific associations were for the Long-T haplotype (**Figure S20**).
326 The SNPs capturing VNTR6-1 status (rs56345976 and rs33961405) were moderately associated
327 with some cancer types (**Table S13**).

328 **Associations of *TERT* isoforms and genetic variants with telomerase-related metrics**

329 *TERT-β*, which encodes a telomerase-nonfunctional protein, is the major *TERT* isoform in both
330 normal and tumor tissues (**Figure S20**). In 30 normal GTEx tissue types, the *TERT-FL* and
331 *TERT-β* isoforms represented on average 17.7% and 58.5%, respectively, of the total *TERT*
332 expression, while they represented 38.4% and 41.0%, respectively, in 33 tumor types in TCGA
333 (**Table S14**). To further explore the functional differences between *TERT-FL* and *TERT-β*, we
334 assessed four telomerase-related metrics: EXpression based Telomerase ENzymatic activity
335 Detection (EXTEND)²⁹, stemness (mRNAsi)³⁰, the telomerase signature score³¹ and telomere
336 length in primary tumors³¹ (**Figure S21, Table S15**). In GTEx, significant correlations with the
337 EXTEND signature (positive for *TERT-FL* and negative for *TERT-β*) were observed in four
338 tissues (blood, colon, esophagus and testis). Similarly, in TCGA, most tumors with significant
339 correlations across all metrics showed positive values for *TERT-FL* and negative values for
340 *TERT-β*. In TCGA, of the four metrics, telomere length in tumors showed the weakest
341 correlations with *TERT* isoform expression (**Figure S21**), potentially due to somatic events,
342 including *TERT*-upregulating promoter mutations^{23,24}.

343 We next tested whether VNTR6.1, rs10069690 or their haplotypes are associated with relative
344 leukocyte telomere length (rLTL). We inferred VNTR6-1 and VNTR6-1/rs10069690 as
345 described above in cancer-free individuals of European ancestry (n=339,103) from the UK
346 Biobank (UKB)³²; the Short-C haplotype was associated with shorter rLTLs ($\beta=-0.049$,
347 $p=8.75E-78$, **Figure 6b, Figure S20b, Table S16**). Significant associations were also observed
348 with several known markers^{33,34} within *TERT* intron 2, including rs7705526 ($\beta=-0.079$, $p=1.02E-$
349 219 ; **Table S16**); adjustment for rs7705526 eliminated the association of the rLTL with VNTR6-
350 1/rs10069690 ($p=0.64$). Notably, regression slopes for these markers differed by genotypes
351 (**Figure 6b**). Interaction analysis in 5-year interval groups revealed a significantly slower
352 decrease in the rLTL in younger individuals and a faster decrease in older individuals without the
353 Short-C haplotype (a greater fraction of telomerase-nonfunctional TERT) than in those with this
354 haplotype ($p_{\text{int}}=1.39E-02$, **Table S16**). This effect remained unchanged after adjustment for
355 rs7705526 ($p_{\text{int}}=1.38E-02$, **Table S16**), which had its own significant interaction ($p_{\text{int}}=3.21E-03$,
356 **Table S16**). The rLTL association pattern was consistent in a smaller set of healthy individuals
357 with lymphocyte telomere length measured by flow FISH,³⁵ but interaction analysis was limited
358 by sample size and age range (**Table S17**).

359 **The VNTR6-1-Short and rs10069690-C alleles are human-specific variants**

360 The VNTR6-1 genomic region is absent in non-primate species (**Figure S22**). In primates, the
361 VNTR6-1 consensus repeat sequences in chimpanzee and bonobo are nearly identical to those in
362 humans and more divergent in orangutan, with all these primates carrying the Long-T haplotype
363 (**Figure S22**). In the genomes of archaic humans (Neandertal and Denisova), only Long-T
364 haplotypes were observed (**Figure S23**). Thus, the VNTR6-1-Short and rs10069690-C alleles, as
365 well as the Short-C haplotype that increase the fraction of the telomerase-functional TERT-FL
366 isoform, are human-specific and major or common in all modern human populations (**Table**
367 **S18**). In cancer-free controls of European ancestry, the Short-C haplotype frequencies were
368 comparable (71.36-72.07%) across 40- to 80-year-old age groups in the UKB and PLCO but
369 decreased to 67% in individuals aged 98-108 years (**Figure S24**). Decreased frequencies of both
370 the rs10069690-C and the VNTR6-1-Short alleles contributed to this difference between
371 centenarians and 40- to 80-year-olds.

372

373 **DISCUSSION**

374 Cancer risk is influenced by complex interactions between genetic and environmental factors and
375 further depends on the replicative potential of stem cells³⁶⁻³⁸. We showed that reduced or
376 elevated cancer risk associated with a multi-cancer GWAS locus at chr5p15.33 marked by the
377 SNPs rs2242652 and rs10069690 is related to the genetic regulation of *TERT* splicing by
378 rs10069690 and VNTR6-1, a 38-bp intronic tandem repeat. In all populations, VNTR6-1 status
379 can be confidently inferred as a biallelic marker with a Short allele (24-27 copies) vs. a Long
380 allele (40.5-66.5 copies) based on haplotypes of two common SNPs, rs56345976 and

381 rs33961405. We inferred VNTR6-1 (Short vs Long alleles) alleles and explored their
382 distributions in controls from 1000G populations and in cancer patients and controls of European
383 ancestry.

384 In Europeans, the GWAS associations for rs2242652-A allele were fully explained by the linked
385 VNTR6-1-Long allele ($r^2=0.62$), but the associations for rs10069690-T ($r^2=0.48$) were partially
386 explained, suggesting that rs10069690 may have an individual functional effect. While we did
387 not detect any functional properties for rs2242652, we demonstrated that both the VNTR6-1-
388 Long and rs10069690-T alleles independently and their combined haplotype (i.e., the Long-T
389 haplotype) alter the *TERT* isoform ratios by reducing telomere-functional TERT-FL and
390 increasing alternative telomerase-nonfunctional INS1b and TERT- β isoforms. We propose that
391 genetically regulated levels of expression and the ratios of these isoforms affect cellular
392 longevity by protecting cells from apoptosis and altering their replicative potential at
393 homeostasis or in response to environmental factors (**Figure 7**).

394 The alleles/haplotypes reducing the TERT-FL and increasing the TERT- β isoform were
395 associated with an elevated risk of cancers originating from tissues (e.g., brain, thyroid, ovary)
396 with no/very low replicative potential at homeostasis and no regenerative replication, i.e. cell
397 growth to repair the tissue damage. Notably, the *TERT- β* isoform accounts for ~80% of the total
398 *TERT* expression in these normal tissues (**Table S14**). The anti-apoptotic effect of the TERT- β
399 isoform may extend the lifespan of these cells, allowing for the accumulation of somatic
400 mutations over time, increasing cancer risk.

401 However, the same alleles/haplotypes were associated with a reduced risk of cancers originating
402 from tissues with low homeostatic proliferation but potentially high regenerative proliferation to
403 repair tissue damage caused by environmental exposures and stressors, including pathogens,
404 hormones and reactive metabolites (e.g., bladder and prostate cancer). In these cases, the anti-
405 apoptotic effect of the TERT- β isoform may reduce the extent of regenerative proliferation, thus
406 limiting mutagenesis caused by replication errors.

407 The reduced fraction of the telomerase-encoding TERT-FL isoform leads to a decreased
408 availability of the TERT-TERC complex for telomere maintenance, making it less likely for
409 mutated cells to achieve immortalization and expansion. Tumorigenesis in tissues with low
410 replicative potential requires driver mutations, such as *TERT*-upregulating promoter mutations³⁹,
411 which can be acquired through replicative or environmental mutagenesis. Rare cells with high
412 TERT expression (TERT-high cells) can serve as stem cells to support tissue regeneration⁴⁰ and
413 initiate tumorigenesis after acquiring driver mutations.

414 We did not detect GWAS associations for the same alleles/haplotypes for cancers originating
415 from tissues with high homeostatic proliferation (e.g., the gastrointestinal tract). High
416 proliferation rates in stem cells of these tissues, combined with cell death induced by critical
417 telomere shortening in differentiated cells, prevent cells from reaching a malignant state and thus
418 act as a tumor-suppressive mechanism⁴¹. For cancer types with no or marginal associations for

419 these alleles/haplotypes, TERT-related mechanisms might be more heterogeneous and dependent
420 on cell specificity, tumor subtype, and timing, as well as the type and intensity of environmental
421 exposure.

422 While the telomerase activity of TERT is essential for cellular homeostasis, non-canonical roles
423 have also been recognized, including roles in the DNA damage response and radiosensitivity⁴⁴.
424 TERT protects mitochondrial function and reduces the production of reactive oxygen species
425 (ROS)⁴². Telomere shortening is accelerated under oxidative stress conditions⁴², potentially
426 through increased damage to telomere DNA or/and the relocation of TERT from the nucleus to
427 mitochondria⁴³. Our data and previous observations²⁷ of the predominant localization of
428 telomerase-nonfunctional TERT- β to mitochondria support the role of this isoform in cancer risk
429 or protection through telomerase-independent regulation of apoptosis.

430 Telomere length has been extensively studied in relation to cancer and non-cancer
431 conditions^{45,46}. Mendelian randomization analysis revealed an association between genetically
432 predicted longer telomeres and the risk of 8 of 22 cancer types tested, especially for rare cancers
433 and cancers of tissues with low replicative potential⁴⁷. Our analysis in the UKB showed a strong
434 association for the VNTR6-1/rs10069690 haplotypes with rLTL but weaker than for the other
435 *TERT* rLTL markers (rs7705526, rs2736100, and rs2853677) used for predicting telomere
436 length^{33,34}. We noted a greater degree of telomere shortening in older than in younger individuals
437 without the Short-C haplotype. This might be due to a greater proportion of circulating
438 lymphocytes originating from stem cells and their progenitors that have experienced more cell
439 divisions due to increased cellular longevity provided by the anti-apoptotic TERT- β isoform.

440 The alleles associated with an increased ratio of telomerase-encoding *TERT-FL* isoform,
441 VNTR6-1-Short, rs10069690-C, and their Short-C haplotype are human-specific variants with a
442 high frequency in 40- to 80-year-old European individuals but a lower frequency in centenarians.
443 The emergence and retention of these alleles might be consistent with the disposable soma theory
444 of ageing, which postulates that evolution favors factors supporting reproductive fitness and
445 growth at the expense of longevity, which requires substantial maintenance to repair the somatic
446 damage that accumulates with increasing age⁴⁸. Female fertility strongly depends on ovarian
447 telomerase⁴⁹, and telomere shortening is considered an evolutionary cost of reproductive trade-
448 offs⁵⁰. The evolutionary selection of genetic variants that increase the ratio of the telomerase-
449 encoding *TERT-FL* isoform might provide this reproductive fitness benefit while decreasing
450 longevity later in life, perhaps due to elevated cancer risk.

451 Some limitations of our study include the lack of longitudinal rLTL data and not considering
452 other possible tissue-specific effects of *TERT* splicing regulation by VNTR6-1 and rs10069690.
453 Further studies are warranted to explore our findings in the context of other cancer GWAS
454 signals within the 5p15.33 region^{3,4}. In conclusion, we demonstrate that the multi-cancer GWAS
455 locus at 5p15.33 marked by rs10069690 and rs2242652 can be genetically and functionally
456 accounted for by a combination of the SNP rs10069690 within intron 4 and a VNTR within

457 intron 6 (VNTR6-1) of *TERT*. These variants independently regulate *TERT* splicing and
458 expression, with a stronger combined effect. The *TERT* isoform ratios affect cellular longevity
459 and replicative potential, which can be further modulated by environmental factors, thus
460 contributing to reduced or elevated cancer risk.

461

462 **Acknowledgments**

463 This work was supported by the Intramural Research Programs of the Division of Cancer
464 Epidemiology and Genetics (DCEG) and the Center for Cancer Research (CCR), National
465 Cancer Institute, and the Center for Alzheimer's and Related Dementias (CARD) within the
466 Intramural Research Program of the National Institute on Aging and the National Institute of
467 Neurological Disorders and Stroke (1ZIAAG000538). BLGSP was funded in part by the
468 Foundation for Burkitt Lymphoma Research (<http://www.foundationforburkittlymphoma.org>)
469 and with U.S. Federal funds from the National Cancer Institute, National Institutes of Health,
470 under Contract No. HHSN261200800001E and Contracts No. HHSN261201100063C and
471 No. HHSN261201100007I (DCEG). The presented results are in part based upon data generated
472 by the TCGA Research Network. The work was conducted using the UK Biobank resource
473 (application 92005). The UK Biobank was established by the Wellcome Trust, the Medical
474 Research Council, the United Kingdom Department of Health, and the Scottish Government. The
475 UK Biobank has also received funding from the Welsh Assembly Government, the British Heart
476 Foundation, and Diabetes UK. The CIBMTR is supported primarily by the Public Health Service
477 U24CA076518 from the NCI, the National Heart, Lung and Blood Institute (NHLBI), and the
478 National Institute of Allergy and Infectious Diseases (NIAID); 75R60222C00011 from the
479 Health Resources and Services Administration (HRSA); and N00014-23-1-2057 and N00014-
480 24-1-2057 from the Office of Naval Research. The Cancer Genomics Research (CGR)
481 Laboratory and Genome Modification Core are funded with Federal funds from the National
482 Cancer Institute under Contract No. 75N910D00024. BP and MM acknowledge the support of
483 the Chan Zuckerberg Initiative and the National Institutes of Health grants U24HG011853 and
484 OT2OD033761 to BP. Max Hogshead was supported by the NCI Intramural Continuing
485 Umbrella for Research Experiences (iCURE) program. We thank Drs. Helen Piontkivska, and
486 the members of the Laboratory of Translational Genomics for comments and discussions. We
487 thank Dr. Tatiana Karpova, Optical Microscopy Core (NCI/CCR/LRBGE), for help with super-
488 resolution imaging. The opinions expressed by the authors are their own and should not be
489 interpreted as representing the official viewpoint of the U.S. Department of Health and Human
490 Services, the National Institutes of Health or the National Cancer Institute.

491 **Author contributions**

492 O. F-V and LP-O conceived the study; O. F-V, C-H. L and CZ performed the data analysis; MH,
493 MH, BWP and KF performed the experiments; CB, KJB, MK, MM and BP generated the long-
494 read genome assemblies; KF, MH, KJ, WL and KT performed the targeted PacBio sequencing;

495 RC, JS, MJM, SJC, SG, SAS and SMM provided reagents, data, samples and interpretations of
496 the results; O. F-V and LP-O wrote the manuscript with the input of all the authors; and LP-O
497 supervised the project.

498 **Competing interests**

499 None declared.

500 **Data availability**

501 All sequencing data generated in this study (PacBio targeted sequencing, PacBio-WGS, HiChIP,
502 RNA-seq) are deposited in the Sequence Read Archive (SRA) under accession numbers
503 (PRJNA1134701 and PRJNA1134698 available at publication). The publicly available datasets
504 used in the study are listed in **Table S19**. The derived data for the public datasets (1000G) are
505 provided in the supplementary tables.

506 **Code availability**

507 The pipeline and script used for the analysis of genome assemblies are available at GitHub
508 (<https://github.com/oflorez/HumanGenomeAssemblies>).

509 Supplementary Information is available for this paper.

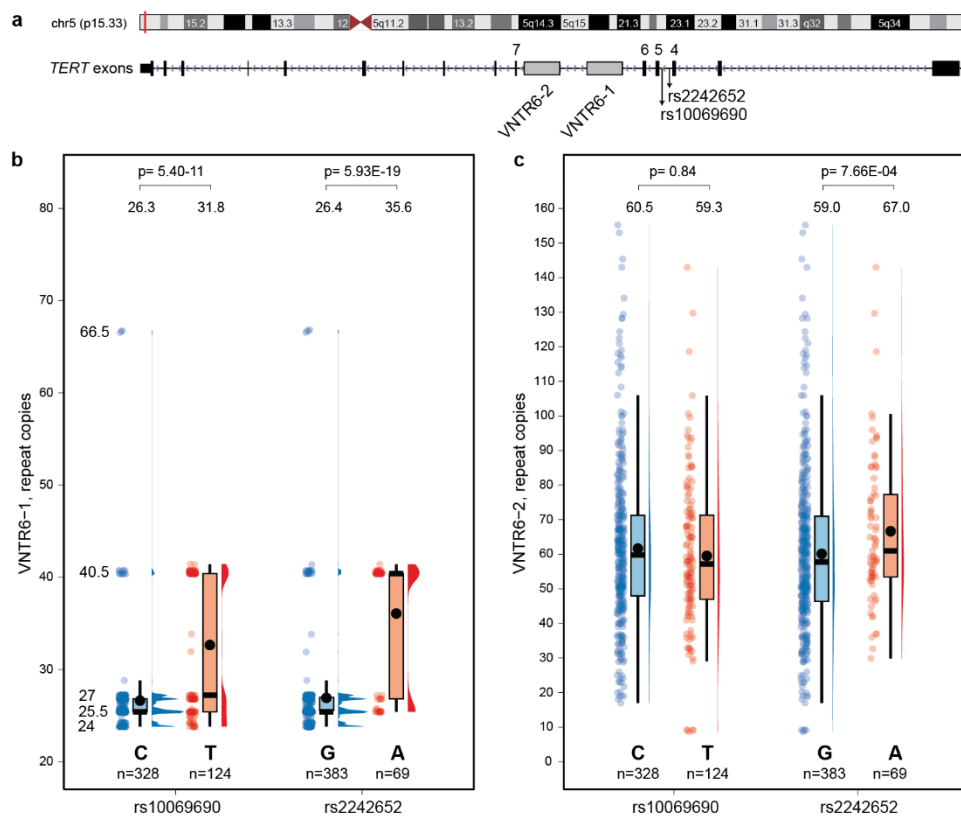
510 Correspondence and requests for materials should be addressed to LP-O.

511

512 **REFERENCES**

- 513 1. Roake, C.M. & Artandi, S.E. Regulation of human telomerase in homeostasis and disease. *Nat*
514 *Rev Mol Cell Biol* **21**, 384-397 (2020).
- 515 2. Rossiello, F., Jurk, D., Passos, J.F. & d'Adda di Fagagna, F. Telomere dysfunction in ageing and
516 age-related diseases. *Nat Cell Biol* **24**, 135-147 (2022).
- 517 3. Rafnar, T. *et al.* Sequence variants at the TERT-CLPTM1L locus associate with many cancer types.
518 *Nat Genet* **41**, 221-7 (2009).
- 519 4. Wang, Z. *et al.* Imputation and subset-based association analysis across different cancer types
520 identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33.
521 *Hum Mol Genet* **23**, 6616-33 (2014).
- 522 5. Chen, H. *et al.* Large-scale cross-cancer fine-mapping of the 5p15.33 region reveals multiple
523 independent signals. *HGG Adv* **2**, 100041 (2021).
- 524 6. Koutros, S. *et al.* Genome-wide Association Study of Bladder Cancer Reveals New Biological and
525 Translational Insights. *Eur Urol* **84**, 127-137 (2023).
- 526 7. Leem, S.H. *et al.* The human telomerase gene: complete genomic sequence and analysis of
527 tandem repeat polymorphisms in intronic regions. *Oncogene* **21**, 769-77 (2002).
- 528 8. Szutorisz, H. *et al.* Rearrangements of minisatellites in the human telomerase reverse
529 transcriptase gene are not correlated with its expression in colon carcinomas. *Oncogene* **20**,
530 2600-5 (2001).
- 531 9. Liao, W.W. *et al.* A draft human pangenome reference. *Nature* **617**, 312-324 (2023).

- 532 10. Lee, O.W. *et al.* Targeted long-read sequencing of the Ewing sarcoma 6p25.1 susceptibility locus
533 identifies germline-somatic interactions with EWSR1-FLI1 binding. *Am J Hum Genet* **110**, 427-
534 441 (2023).
- 535 11. Schumacher, F.R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate
536 cancer susceptibility loci. *Nat Genet* **50**, 928-936 (2018).
- 537 12. Melin, B.S. *et al.* Genome-wide association study of glioma subtypes identifies specific
538 differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet* **49**,
539 789-794 (2017).
- 540 13. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**,
541 92-94 (2017).
- 542 14. Milne, R.L. *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative
543 breast cancer. *Nat Genet* **49**, 1767-1778 (2017).
- 544 15. Phelan, C.M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial
545 ovarian cancer. *Nat Genet* **49**, 680-691 (2017).
- 546 42. Ahmed, S. *et al.* Telomerase does not counteract telomere shortening but protects
547 mitochondrial function under oxidative stress. *J Cell Sci* **121**, 1046-53 (2008).
- 548 43. Haendeler, J., Hoffmann, J., Brandes, R.P., Zeiher, A.M. & Dimmeler, S. Hydrogen peroxide
549 triggers nuclear export of telomerase reverse transcriptase via Src kinase family-dependent
550 phosphorylation of tyrosine 707. *Mol Cell Biol* **23**, 4598-610 (2003).
- 551
- 552



553

554 **Figure 1. Analysis of VNTR6-1 and VNTR6-2 within *TERT* intron 6 in relation to GWAS**
 555 **leads rs10069690 and rs2242652.**

556 **a**, The chr 5p15.33 genomic region with GWAS leads rs10069690 and rs2242652 within *TERT*
 557 intron 4 and VNTRs within intron 6. **b**, Distribution of repeat copies of VNTR6-1 (38-bp repeat
 558 unit) and **c**, VNTR6-2 (36-bp repeat unit) in 452 phased long-read genomic assemblies from 226
 559 controls of diverse ancestries. The dots represent repeat copies for each chromosome assembly;
 560 box plots show the overall and interquartile range, median (horizontal black line), and mean
 561 (black dot, with values above corresponding plots). Half-violin plots show the density
 562 distribution of the data. Five VNTR6-1 alleles—24, 25.5, 27, and 40.5 repeat copies—were
 563 observed above the 5% frequency threshold and accounted for 90.04% of all alleles in the set;
 564 VNTR6-2 alleles were scattered between 8 and 155 repeat copies, all under the 5% threshold. P
 565 values were calculated for unpaired two-sample Wilcoxon–Mann–Whitney tests comparing the
 566 number of repeat copies between the genotype groups.

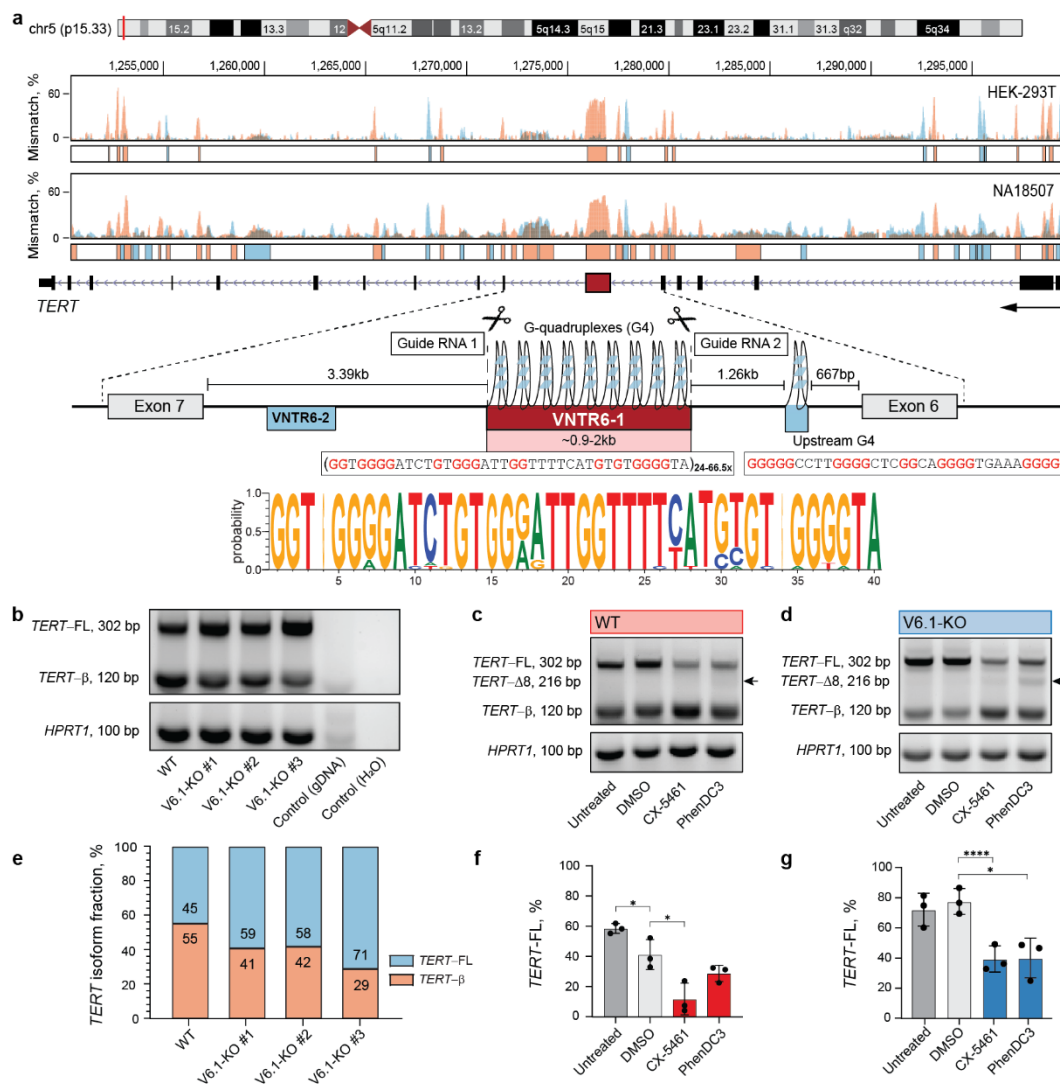
567

568

569

570

571



572

573 **Figure 2. VNTR6-1 affects the splicing ratios of the *TERT-FL* and *TERT-β* isoforms.**

574 **a**, G4-ChIP results within the *TERT* region in the HEK-293T (VNTR6-1: Long/Long) and
575 NA18507 (YRI, 1000G, VNTR6-1: Short/Short) cell lines display mismatches (%) during DNA
576 synthesis, reflecting polymerase stalling after G stabilization in both the plus (blue) and minus
577 (orange, direction of *TERT* transcription) genome strands. The genomic region of *TERT* intron 6
578 shows VNTR6-1 (24-66.5 copies of the 38-bp repeat unit), VNTR6-2, G4s in the minus strand
579 (polymorphic G4 within VNTR6-1 and constitutive upstream G4), and CRISPR/Cas9 guide
580 RNAs for excising VNTR6-1. The sequence logo shows the consensus of the 38-bp VNTR6-1
581 repeat unit in UMUC3 cells based on PacBio long-read WGS. **b**, Agarose gels of RT-PCR
582 products amplified from cDNA of corresponding samples; gDNA—genomic DNA was used as a
583 negative control; *HPRT1* was used as a normalization control. **e**, Densitometry results of the PCR
584 amplicons in plot **b**. The differences in the *TERT* isoform ratios are further explored in **Figure**
585 **S11**. Experiments in UMUC3 cells comparing *TERT* splicing and isoform-specific expression
586 after 72 hrs of treatment with G4 stabilizing ligands, normalized to *HPRT1* as an endogenous
587 control in the WT (**c, f**) and V6.1-KO (**d, g**) cell lines. **c, d**, A representative agarose gel of

588 SYBR-Green RT-qPCR products detecting several isoforms with primers located in exons 6 and
589 9. The extra PCR band, marked by an arrow in panels **c** and **d**, is further explored in **Figure S12**.
590 **f, g**, Densitometry analysis of the corresponding agarose gels evaluating the *TERT-FL* (%)
591 relative to the total PCR products. All analyses are based on three experiments, with one
592 representative gel shown. Comparisons were made against the vehicle control (DMSO).
593 Statistical significance is indicated as follows: ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$,
594 Student's T-test.

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

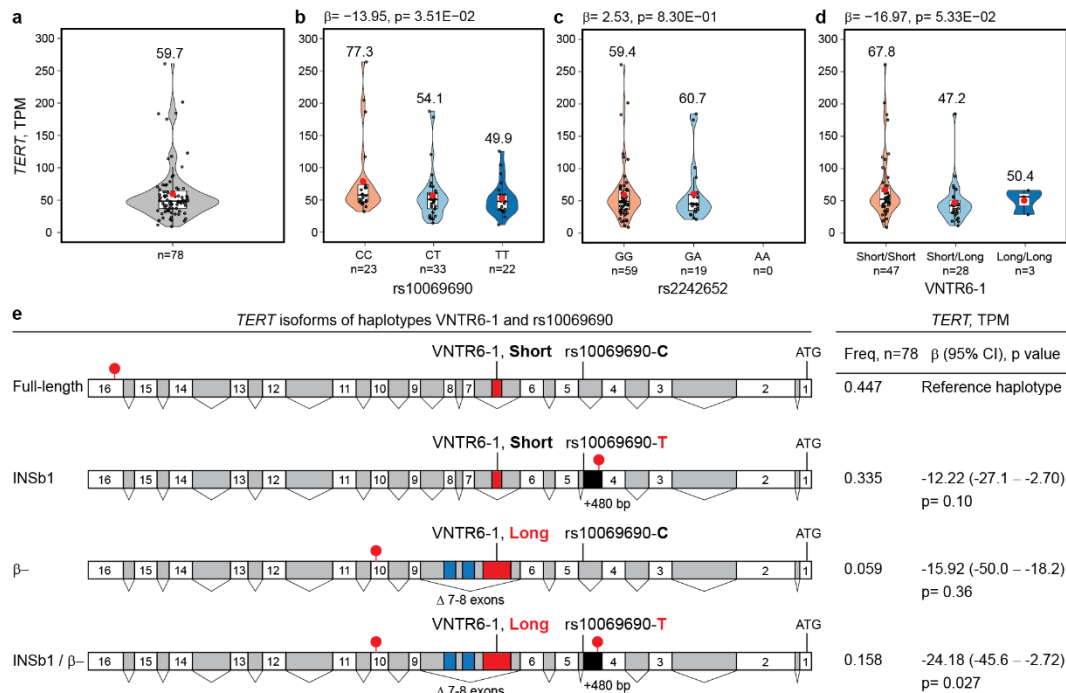
613

614

615

616

617



618

619 **Figure 3. Analysis of *TERT* expression in 78 Burkitt lymphoma (BL) tumors.**

620 Total *TERT* expression analyzed as transcripts per million (TPMs) **a**, overall and in relation to
 621 the **b**, rs10069690, **c**, rs2242652 and **d**, VNTR6-1 genotypes. Group means are shown as red dots
 622 with values above corresponding violin plots. **e**, association of the VNTR6-1 and rs10069690
 623 haplotypes with total *TERT* expression. The reference Short-C haplotype corresponds to the
 624 telomerase-encoding *TERT-FL* isoform, while the *INS1* and *TERT-β* isoforms encode truncated
 625 proteins without telomerase activity. Effect alleles in haplotypes are marked in red; white boxes
 626 – exons; gray boxes – introns; black boxes – intron 4 retention; blue boxes – alternative exons 7
 627 and 8; and red lollipops – stop codons. The direction of the *TERT* exons is from right to left,
 628 corresponding to the minus strand, as presented in the UCSC browser. “ATG” indicates
 629 translation start codons. P values and β-values are for linear regression models adjusted for sex
 630 and age.

631

632

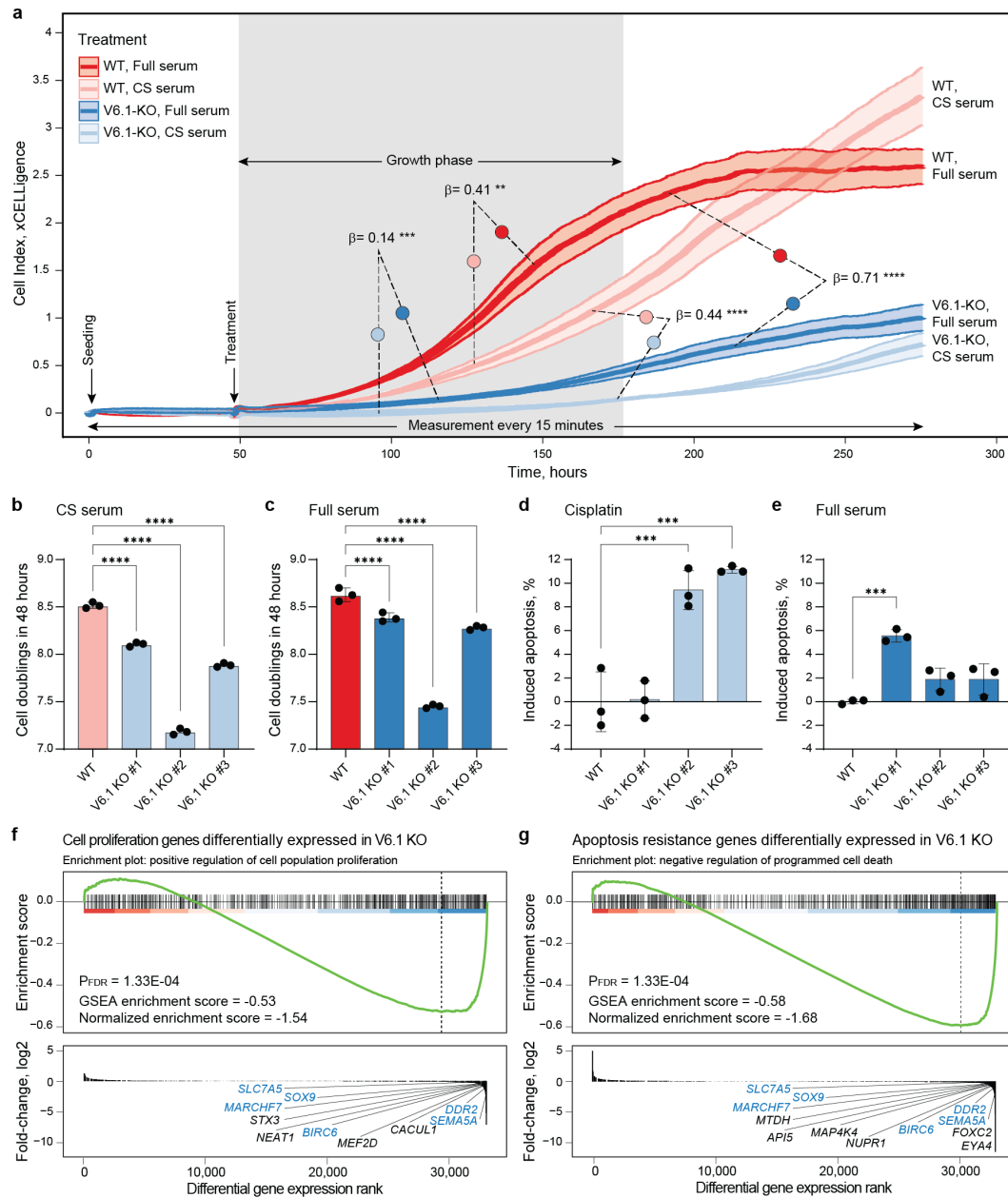
633

634

635

636

637



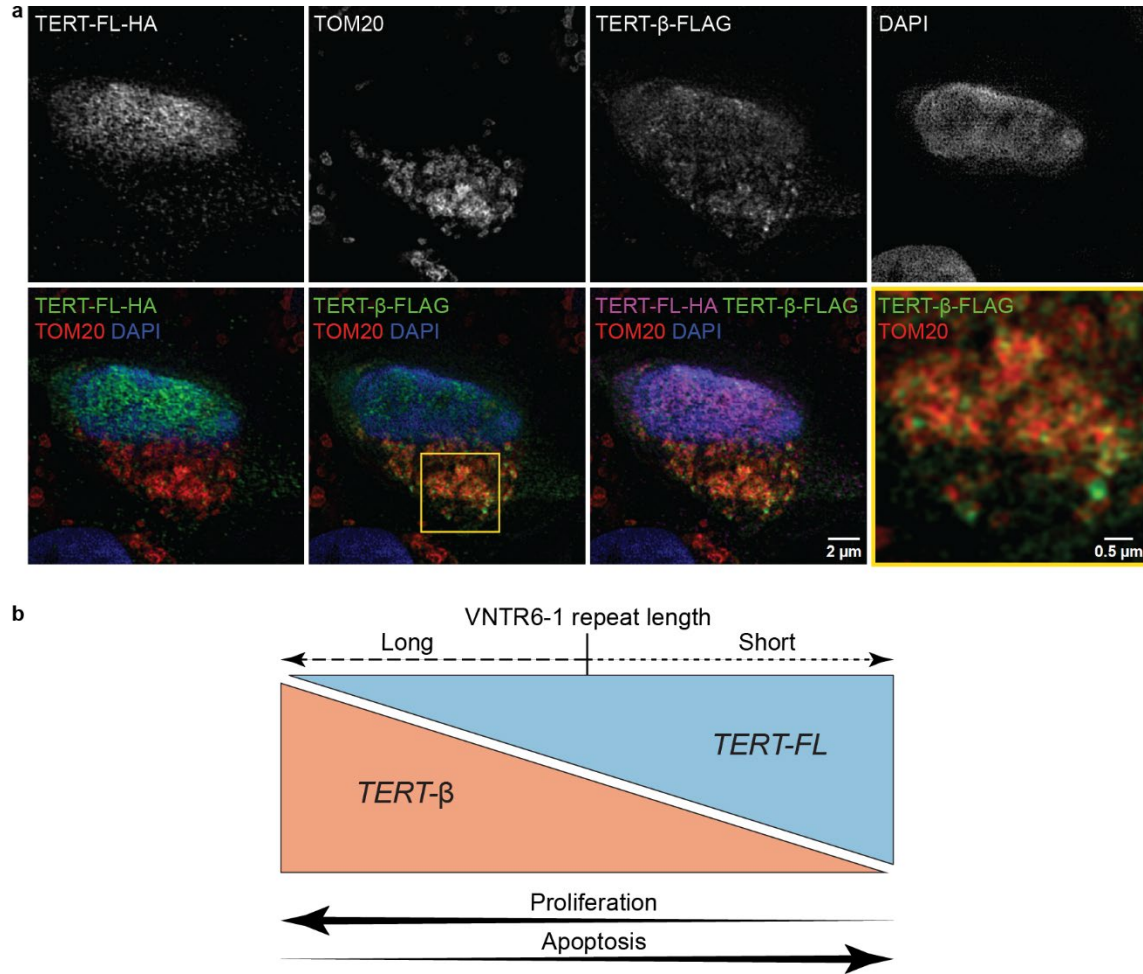
638

639 Figure 4. VNTR6-1 affects the proliferation and apoptosis of UMUC3 cells

640 **a**, Real-time monitoring of cell population growth dynamics (cell index) for 283 hrs in UMUC3
 641 WT and V6.1-KO cells cultured in media supplemented with full serum or charcoal-stripped
 642 (CS) serum revealed significantly greater proliferation rates in WT cells than in V6.1-KO cells
 643 under both culture conditions. Statistical significance and β -values for differences in the cell
 644 index during the visually determined growth phase (gray highlighting between 50 and 183 hours)
 645 were calculated using linear mixed-effects models based on six replicates, ** $p < 0.01$, *** $p <$
 646 0.001 , **** $p < 0.0001$. **b**, Quantification of cell doubling events in CFSE-stained cells cultured
 647 with CS serum or **c**, full serum medium for four days in three replicates (**** $p < 0.0001$,
 648 Student's *t* test). **d**, Cells were treated with 10 μ M cisplatin or **e**, full serum medium for 48 hrs,

649 followed by Annexin V-FITC staining to determine the percentage of apoptotic cells in three
650 replicates (**p < 0.001, Student's t test). Differential expression of genes involved in pathways
651 related to the downregulation of **f**, cell proliferation (positive regulation of cell population
652 proliferation pathway, GO:0008284, **Table S11**) and **g**, apoptosis resistance (negative regulation
653 of programmed cell death pathway, GO:0043069, **Table S11**) according to RNA-seq analysis of
654 V6.1-KO UMUC3 cells compared to WT UMUC3 cells. Genes highlighted in blue are common
655 to both pathways. The data shown in all panels except **f** and **g** represent one of three independent
656 experiments.

657



658

659 **Figure 5. The functional differences between the TERT-FL and TERT-β isoforms**

660 **a**, Structured illumination microscopy images of A549 cells co-transfected with TERT-FL and
 661 TERT-β expression constructs at a 50:50% ratio. For individual channels, staining is shown as
 662 black/white images for better contrast. On tri-color merged panels, green – FLAG (TERT-β) or
 663 HA (TERT-FL), blue – DAPI (nuclei). On the quad-color merged panel, purple - HA (TERT-
 664 FL), green - FLAG (TERT-β), red - TOM20 (mitochondria), blue - DAPI (nuclei). The yellow
 665 inset in the TERT-β-FLAG panel is shown at a higher magnification to demonstrate
 666 colocalization with mitochondria (yellow staining). **b**, The overview of the VNTR6-1, *TERT*-
 667 FL:*TERT*-β ratio, proliferation, and apoptosis.

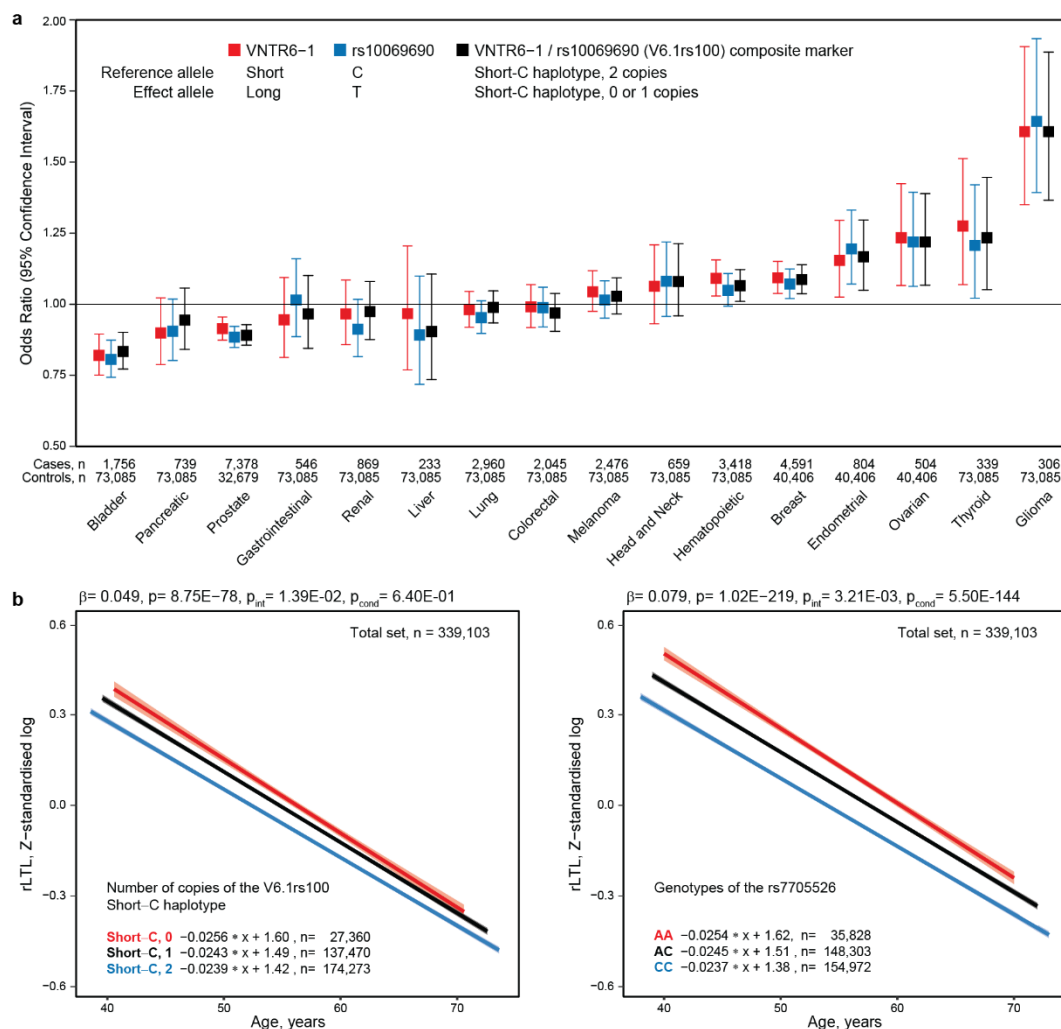
668

669

670

671

672

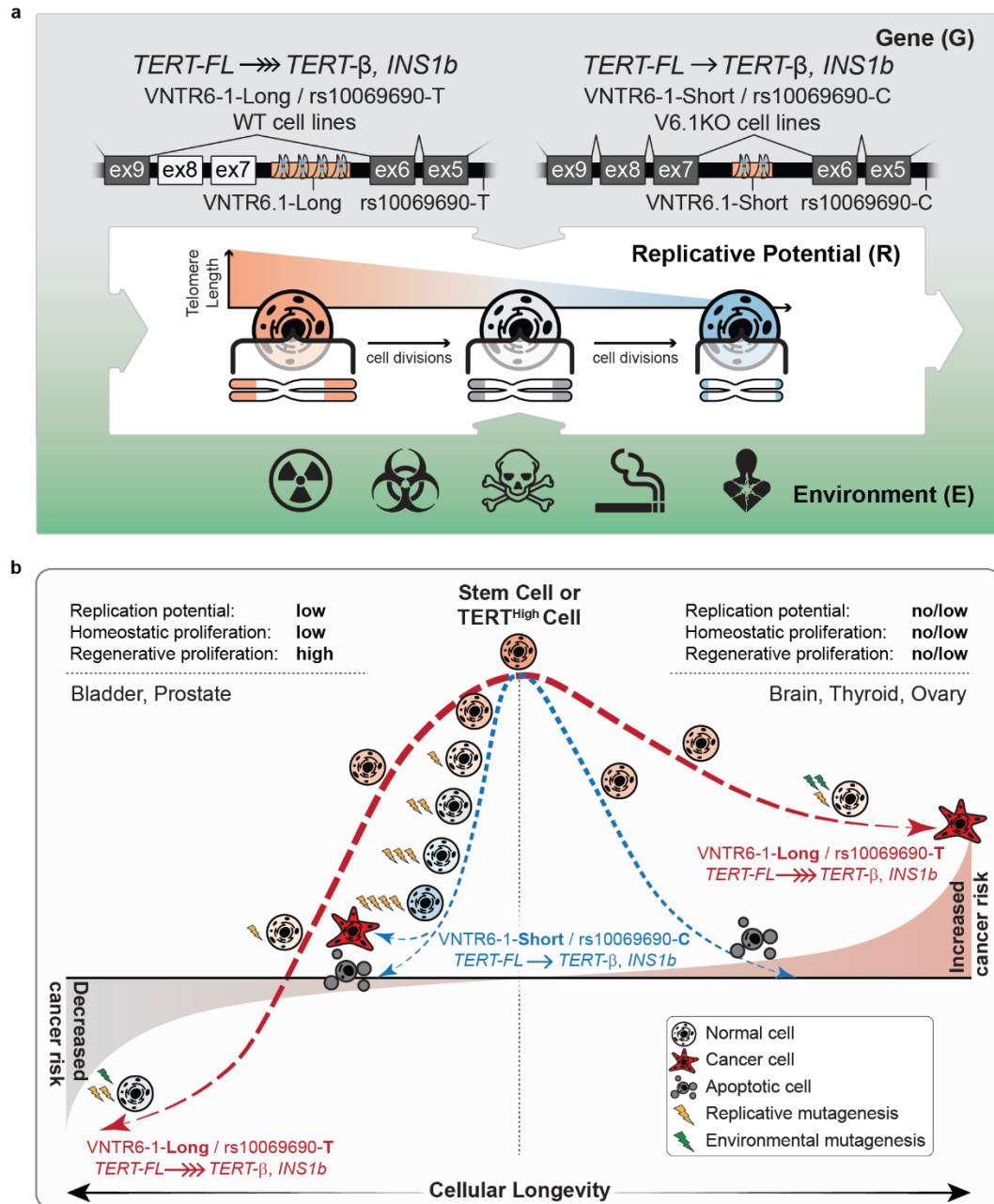


673

674 **Figure 6. Association analyses for cancer risk in PLCO and relative leukocyte telomere**
 675 **length (rLTL) in UKB cancer-free individuals.**

676 **a**, Evaluation of cancer risk associated with the VNTR6-1-Long and rs10069690-T alleles and
 677 the composite marker (VNTR6-1/rs10069690) in the PLCO dataset (n=102,708). Odds ratios
 678 (ORs) with 95% confidence intervals (CIs) were calculated for comparisons between patients
 679 with the indicated cancers and a common group of cancer-free controls using logistic regression
 680 analysis with an additive genetic model adjusted for sex and age. **b**, Evaluation of the
 681 relationships between rLTL and VNTR6-1/rs10069690 and rs7705526 in UKB cancer-free
 682 individuals (n=339,103, LD, $r^2=0.33$ between VNTR6-1/rs10069690 and rs7705526). P values
 683 and β -values were derived from linear regression models adjusted for sex, age, and smoking
 684 status. P_{int} represents the interaction between genotypes and 5-year age groups; P_{cond} represents
 685 the mutual adjustment for rs7705526 or VNTR6-1/rs10069690. The graphs display regression
 686 lines with 95% confidence intervals and regression equations. The analysis revealed a decrease
 687 in the rLTLs with more copies of the Short-C haplotype. The results of sex-specific analyses of
 688 VNTR6-1/rs10069690 are presented in **Figure S20b**.

689



690

691 **Figure 7. The interaction model of factors affecting cancer risk**

692 **a**, *TERT* genetic variants VNTR6-1 and rs10069690 and environmental factors define the
 693 relative ratios of the isoforms encoding telomerase-functional *TERT-FL* and telomerase-
 694 nonfunctional *TERT-β* and *INS1b* isoforms. These isoforms affect cell proliferation, apoptosis
 695 and telomere length, thus modulating cellular longevity and replicative potential, including
 696 homeostatic proliferation, which maintains tissue self-renewal, and regenerative proliferation,
 697 which responds to environmental factors and tissue damage. **b**, Cancer risk as a product of G x E
 698 x R interactions. The VNTR6.1-Long and rs10069690-T alleles, or their haplotype (Long-T), are
 699 associated with reduced cancer risk in tissues with low homeostatic but high regenerative
 700 potential (e.g., bladder). The anti-apoptotic effect of the *TERT-β* isoform reduces the need for

701 regenerative proliferation, thus decreasing the risk of acquiring mutations due to replicative
702 mutagenesis. In tissues with no/low homeostatic and regenerative proliferation (e.g., brain,
703 thyroid, ovary), the same alleles and Long-T haplotype are associated with elevated cancer risk.
704 The anti-apoptotic effect of *TERT-β* extends cellular longevity, allowing the accumulation of
705 more mutations due to environmental mutagenesis, such as through exposure to reactive oxygen
706 species (ROS), cellular metabolites, etc.

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731 ONLINE METHODS

732 Human samples used for targeted PacBio-seq and TaqMan genotyping of select SNPs:

733 DNA samples for HapMap I (CEU panel for CEPH Utah residents with ancestry from Northern
734 and Western Europe, n=90), HapMap III (YRI panel for Yoruba in Ibadan, Nigeria, n=90), select
735 samples from the Human Pangenome Reference Consortium (HPRC, n=10), and the European
736 panel of the Georgia Centenarian Collection (n=100) were purchased from the Coriell Institute
737 for Medical Research; deidentified tissue samples for bladder tumors and matching adjacent
738 normal samples (n=5 pairs) were purchased from Asterand Bioscience and used for DNA
739 extraction and genotyping. Flow FISH telomere length samples (n=77) were obtained from
740 donors of hematopoietic cell transplants from the Center for International Blood and Marrow
741 Transplant Research (CIBMTR; <https://cibmtr.org>) biorepository. Telomere length was measured
742 for total lymphocytes and lymphocyte subsets using the flow FISH assay in a previous study¹.
743 For the current analysis, the samples were selected to represent a wide range of telomeres (4.5 to
744 11.2 kb) and telomere length was analyzed in relation to *TERT* genetic variants using linear
745 regression models adjusted for age and sex.

746 **Cell lines:** The urinary bladder cell lines UMUC3 (CRL-1749), 5637 (HTB-9), HT1376 (CRL-
747 1472), RT4 (HTB-2), T24 (HTB-4), and SCaBER (HTB-3), as well as the Burkitt lymphoma cell
748 line Raji (CCL-86) and the lung cancer cell line A549 (CCL-185), were purchased from ATCC
749 (Manassas) and maintained in the recommended media supplemented with 10% FBS (unless
750 specified otherwise) and 1% antibiotics. All cell lines were regularly tested for Mycoplasma
751 contamination using the MycoAlert Mycoplasma Detection Kit (Lonza) and authenticated by
752 Identifiler (Thermo Fisher). For experiments with cells grown in media containing charcoal-
753 stripped (CS) serum, 3-4 days prior to the experiments, the media was changed to phenol red-
754 free EMEM supplemented with 10% CS FBS, 1% GlutaMAX, and 1% antibiotics.

755 **Analyses of BL tumors:** RNA-seq and DNA-WGS data (Illumina) for Burkitt lymphoma (BL)
756 tumors were obtained from the National Cancer Institute (NCI) Cancer Genome Characterization
757 Initiative (CGCI): Burkitt Lymphoma Genome Sequencing Project (BLGSP)^{2,3}, dbGaP
758 phs000527.v6.p2. The datasets were accessed through the National Cancer Institute Genomic
759 Data Commons (GDC, <https://gdc.cancer.gov/>). RNA-seq BAM files were analyzed using read
760 counts based on the R package FeatureCounts (v2.0.6). Splicing events between *TERT* exons 4
761 and 5 were annotated based on a custom GTF annotation file to perform read summarization at
762 the feature level, generating a raw count matrix. The total number of reads was determined by
763 counting the reads mapped to the splicing junction between exons 4 and 5 and those that
764 extended into intron 4 by at least 20 bp. Read counts were calculated for the splicing events *INSI*
765 (a 38-bp extension of exon 4 into intron 4), *INSIb* (a 480-bp extension of exon 4 into intron 4)
766 and unspliced intron 4 (total reads between exons 4 and 5 minus reads for *INSI* and *INSIb*) as
767 fractions of the read counts for these events within total read counts. BAM files were also used to
768 estimate the overall expression of *TERT* isoforms $-\alpha$, $-\beta$, and $-\alpha-\beta$, which were indexed in a
769 GTF file from ENSEMBL and analyzed using MISO (v0.5.4) with default parameters.

770 Transcripts per million (TPM) values for bulk *TERT* RNA-seq data were downloaded from the
771 GDC data portal. eQTL analyses were performed under additive genetic models using the ‘lm’
772 function in R (v4.3.0), with adjustments for sex and age. *TERT* intron retention was analyzed
773 with IRFinder (v2.0.1) with default settings using the GRCh38 reference genome FASTA file
774 and transcriptome GTF file for annotation.

775 **Analysis of long-read sequences:** VNTR6-1 and VNTR6-2 within *TERT* intron 6 were explored
776 based on long-read sequencing data. Phased genome assemblies for 47 individuals (94
777 chromosomes) were downloaded in FASTA format from the Human Pangenome Reference
778 Consortium (HPRC)⁴. Additionally, we used 358 long-read sequencing (R9, Oxford Nanopore)
779 DNA assemblies generated by the Center for Alzheimer's and Related Dementias (CARD) of the
780 National Institute on Aging. Input DNA was extracted from the brain tissue of 179
781 neurologically normal individuals of European ancestry ([dbGaP_phs001300.v4.p1](#)) and phased
782 assemblies were generated using the Napu pipeline⁵.

783 From the assemblies, the genomic sequences were extracted in FASTA format using Cutadapt
784 (v4.0) based on two sets of nested sequences flanking the region of interest, ~9 kb, GRCh38,
785 chr5:1,271,950-1,281,050. The extracted sequences were aligned to the GRCh38 reference
786 genome using minimap2 (v2.26) and combined in one BAM file, with each individual
787 represented by two sequences, one for each chromosome. In this BAM file, SNPs were scored
788 using SAMtools with mpileup flag (v1.17), and VNTRs were scored using Straglr (v1.4) with
789 default settings. The pipeline is available at
790 <https://github.com/oflorez/HumanGenomeAssemblies>.

791 **Targeted PacBio-seq:** PCR amplicons for targeted PacBio sequencing of VNTR6-1 were
792 generated using the LA Taq Hot-Start DNA Polymerase Kit (Takara) and M13-tagged primers
793 VNTR6-1-M13F) and VNTR6-1-M13R (**Table S20**). In the reference human genome, these
794 primers capture a genomic fragment of 2,241 bp. The optimized 20 µl reactions included 4%
795 DMSO, 0.3 µl of LA Taq DNA Polymerase, 2.5 µl of 10x LA Taq PCR Buffer, 4 µl of 2.5 mM
796 dNTPs, 0.5 µl of each 10 µM primer, and 25 ng of genomic DNA. The PCR conditions included
797 denaturation for 1 min at 94°C, 36 cycles of denaturation for 10 s at 98°C and combined
798 annealing/extension for 3.5 min at 68°C, followed by a final extension for 10 min at 72°C.

799 The controls included 1000G DNA samples purchased from the Coriell Institute for Medical
800 Research and selected to represent various repeat lengths determined based on HPRC assemblies
801 (HG00741, HG01358, HG01891, HG02080, HG02622, HG02717, HG02723, HG03453,
802 HG03492, and NA18906) (**Table S2**). For technical validation of the first and second rounds of
803 PCR, all products were quantified with the Quant-iT PicoGreen dsDNA Assay (Invitrogen), and
804 5% of the products were analyzed with the TapeStation D5000 Kit (Agilent). The second round
805 of PCR was performed with the LA Taq Hot-Start DNA Polymerase Kit and the SMRTbell
806 Barcoded Adapter Complete Prep Kit (PacBio), and the M13 tags incorporated by the first PCR
807 were used to attach unique barcodes to each sample with primers M13F and M13R, where “N”

808 represents the unique barcode (**Table S20**). The 25 μ l PCRs included 4% DMSO, 0.4 μ l of LA
809 Taq DNA Polymerase, 2.5 μ l of 10x LA Taq PCR Buffer II, 4 μ l of 2.5 mM dNTPs, 1.0 μ l of
810 each 3 μ M barcoded M13 primer, and 25 ng of product from the first PCR. The PCR conditions
811 included denaturation for 1 min at 94°C, 10 cycles of denaturation for 10 s at 98°C and
812 combined annealing/extension for 3.5 min at 68°C, followed by a final extension for 10 min at
813 72°C. The final amplicons from three 96-well PCR plates (288 samples) were pooled, processed
814 with the Sequel II binding kit 3.1 (PacBio), and sequenced on one SMRT Cell on the Sequel II
815 System (PacBio).

816 **PacBio amplicon analysis:** The high-fidelity (HiFi) reads were assembled by circular consensus
817 sequencing (CCS) within SMRT Link (PacBio), demultiplexed with Lima, and aligned to the
818 reference genome GRCh38 with minimap2. The VNTR6-1 amplicons had an average read
819 coverage of ~10,000 reads per sample. The resulting BAM files were scored for rs56345976 and
820 rs33961405 using SAMtools with mpileup flag (v1.17) and for VNTR6-1 using Straglr (v1.4).
821 The analysis was restricted by reads fully covering the amplicon (GRCh38, chr5:1275500-
822 1277500), excluding outputs from partial reads using SAMtools with the ampliconclip flag
823 (v1.17). Phased haplotypes of rs56345976 and rs33961405 were constructed based on PacBio
824 reads.

825 **DNA genotyping:** TaqMan genotyping assays for *TERT* SNPs rs56345976 (C__88595060_10),
826 rs33961405 (C__34209972_10), rs10069690 (C__30322061_10), rs2242652
827 (C__16174622_20), rs7705526 (C__189441058_10), rs2736100 (C__1844009_10) and
828 rs2853677 (C__1844008_10) were purchased from Thermo Fisher. The samples were genotyped
829 in 384-well plates on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems) using
830 2x TaqMan Genotyping Master Mix (Thermo Fisher) in 5- μ L reactions with 4 ng of genomic
831 DNA per reaction.

832 **Analyses in the 1000 Genomes Project:** High-coverage (30x) short-read whole-genome
833 sequencing (WGS) data in CRAM format and phased genetic variants for 3,201 individuals from
834 the 1000G populations⁶ were downloaded from <https://www.internationalgenome.org/data-portal/data-collection/30x-grch38>
835 for the 400 kb genomic region (GRCh38 chr5:1,100,000-
836 1,500,000). The depth of coverage of aligned short-sequencing reads within the 2,290 bp
837 genomic region corresponding to VNTR6-1 (GRCh38 chr5:1,275,210-1,277,500) was analyzed
838 by calculating median coverage within consecutive 50-base windows using Mosdepth (v0.2.5).
839 All the samples were classified into VNTR6-1-Short/Short genotypes (24-27 copies) and
840 Long/any genotypes (with one or two Long alleles of 40.5 or 66.5 copies) by applying a machine
841 learning strategy with the tidymodels framework and based on the R package ‘glmnet’ (v4.1-7)⁷.
842 First, a total of 605 samples (18.89%) were randomly selected from the set, representing all
843 1000G super-populations, and visually examined and assigned to the Short or Long groups based
844 on the coverage profiles in IGV. Then, the dataset was split into training (60%) and testing
845 (40%) sets. Fivefold cross-validation was used during the training process to develop and
846 evaluate the prediction model. The model demonstrated stable performance in accurately

847 classifying VNTR6-1 into the Short and Long categories, with 96.8% specificity, 92.8%
848 sensitivity, an F score of 0.95, and an area under the ROC curve (AUC) of 0.98 (**Figure S4**).

849 To identify variants predictive of VNTR6-1-Short/Long status, all 12,338 biallelic SNPs from
850 the 1000G phased genetic variant data across the 400 kb genomic region (GRCh38
851 chr5:1,100,000-1,500,000) were extracted and filtered for MAF > 5%, resulting in 1,473 SNPs
852 for analysis. Based on Chi-square tests, 594 of these SNPs were significantly associated with the
853 VNTR6-1 Short and Long categories ($p < 0.05$). A random forest model was then applied using
854 the R package ‘randomForest’ (v4.7-1.1) to identify the predictive value of the significant SNPs
855 for VNTR6-1 categories, selecting the top 10% based on mean decrease in Gini scores. A total of
856 60 SNPs were identified as highly informative, with rs56345976 and rs33961405 showing the
857 highest combined predictive probabilities for VNTR6-1 classification.

858 To map the haplotypes of rs56345976 and rs33961405 with the profile of coverage distribution
859 across the genomic region GRCh38 chr5:1,275,210-1,277,500, we applied unsupervised
860 hierarchical clustering using the core ‘hclust’ function in R (v4.3.0) with the Euclidean distance
861 metric and complete linkage method. The rs56345976-A/rs33961405-G haplotypes captured the
862 VNTR6-1 Long allele (Cohen's Kappa coefficient of 0.78 and agreement of 0.90), while all the
863 remaining haplotypes captured the VNTR6-1 Short allele (**Figure S4**). Phased data from our
864 long-read sequencing, including assemblies and targeted PacBio sequencing, was used to
865 confirm the co-segregation of rs56345976 and rs33961405 with VNTR6-1 (**Table S2**).

866 We created a custom 1000G reference panel that included all markers within the 400 kb genomic
867 region (GRCh38 chr5:1,100,000-1,500,000). In this region, VNTR6-1 was used as a biallelic
868 marker, with Short and Long alleles determined by the rs56345976/rs33961405 haplotypes at
869 position chr5:1,275,400 (**Table S3**). To evaluate the scoring performance, the 1000G dataset
870 ($n=3,201$) was randomly partitioned into two groups, which served as a reference panel
871 ($n=1,601$) and a test panel ($n=1,600$) to perform phasing with SHAPEIT4 (v4.2.0) and
872 imputation with IMPUTE2 (v2.3.2) with default settings. VNTR6-1 was confidently scored in all
873 test panel samples (imputation quality score⁸, IQS = 0.98), with an overall concordance of 99.3%
874 compared to the predetermined genotypes across the entire dataset. Population-specific
875 concordance rates for VNTR6-1 imputation were as follows: EUR 99.7% ($n=321$), AMR 99.6%
876 ($n=243$), AFR 99.1% ($n=456$), SAS 98.99% ($n=299$), and EAS 98.32% ($n=281$).

877 **Analyses in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial:**
878 PLCO⁹ is a large population-based cohort that includes 155,000 participants enrolled between
879 November 1993 and July 2001. The individual-level data, including genotyped and imputed
880 variants and phenotype data, were provided by PLCO upon approved application. The dataset of
881 individuals of European ancestry comprised 102,708 individuals, including 73,085 common
882 cancer-free controls and 29,623 patients with 16 cancer types. All the variants within the 400 kb
883 region (GRCh38 chr5:1,100,000-1,500,000) were phased using SHAPEIT4 (v4.2.0) and then
884 VNTR6-1 genotypes (Short or Long) were assigned based on phased rs56345976/rs33961405

885 haplotypes. Logistic regression analyses were conducted with the logit link function for binary
886 outcomes using the ‘glm’ function in R (v4.3.0), adjusting for sex and age.

887 **Analyses in the UK BioBank:** Associations between genetic markers and relative telomere
888 length (rTL) in peripheral blood lymphocytes were assessed in the UK Biobank (UKB)
889 (<https://www.ukbiobank.ac.uk/>), a population-based prospective study in the United Kingdom¹⁰.
890 The analysis included 351,634 cancer-free participants of European ancestry with both SNP data
891 and rTL measurements. VNTR6-1 was scored as described above for PLCO. We used linear
892 regression models to assess the association between the technically adjusted rLTLs (log_e and Z-
893 transformed)¹¹ and the genetic markers. This analysis was performed using the ‘lm’ function in R
894 (v4.3.0) and adjusting for sex, age, and smoking status. A conditional linear model was tested by
895 independently adding SNPs (rs2736100, rs2853677, and rs7705526) that are strongly associated
896 with telomere length in multiple populations. To account for trend differences in rLTLs across
897 all ages, the conditional linear model included an interaction term between the genetic markers
898 and 5-year age groups, that were used to avoid age-heaping bias while maintaining a sufficient
899 sample size for each age class.

900 **Analyses in The Cancer Genome Atlas (TCGA):** Blood-derived germline data for 9,610
901 TCGA participants across 33 cancer types were accessed through the National Cancer Institute
902 Genomic Data Commons (GDC, <https://gdc.cancer.gov/>). Controlled access genotype calls
903 generated from Affymetrix SNP6.0 array intensities using BIRDSUITE¹² were retrieved from the
904 genomic region GRCh37, chr5:335,889-2,321,650. In this region, in addition to the 5,453
905 initially genotyped variants, we imputed approximately 57,000 additional variants with
906 imputation quality scores exceeding 0.8 using the TOPMed Imputation Server, which includes
907 data from more than 97,000 participants¹³. The imputation quality scores across cancer types
908 were as follows: mean (min–max) $r^2=0.83$ (0.78-0.89) for rs56345976, $r^2=0.85$ (0.75-0.92) for
909 rs33961405, $r^2=0.85$ (0.76-0.94) for rs10069690, and $r^2=0.84$ (0.74-0.92) for rs2242652. Direct
910 genotyping from germline WGS files for 387 BLCA downloaded from GDC showed high
911 concordance rates between imputed and WGS-genotyped markers: 89.90% for rs56345976,
912 86.79% for rs33961405, 91.17% for rs10069690 and 92.75% for rs2242652.

913 Transcripts per million (TPM) for bulk *TERT* RNA-seq data were downloaded from the GDC
914 within the Pan-Cancer Atlas publications¹⁴. The TPMs for the *TERT*-β and *TERT*-FL transcripts
915 were downloaded from the UCSC Xena platform (<https://xenabrowser.net/datapages/>) within the
916 UCSC toil RNA-seq Recompute Compendium, cohort TCGA Pan-Cancer (PANCAN). We used
917 pre-computed telomerase-related metrics, including the expression-based telomerase enzymatic
918 activity detection (EXTEND) scores based on a 13-gene signature¹⁵, stemness indices calculated
919 via a predictive model using one-class logistic regression on mRNA expression¹⁶, a telomerase
920 signature score estimated from a 43-gene panel, and the telomere length scores calculated using
921 TelSeq based on WGS¹⁷.

922 eQTL analysis was conducted using TPMs for bulk RNA-seq *TERT* expression data and genetic
923 markers (additive genetic model) using the ‘lm’ function in R (v4.3.0), with adjustments for sex
924 and age. Spearman rank correlations between *TERT* expression (*TERT*-β and *TERT*-FL) and
925 telomerase-associated metrics for each cancer type were determined using the ‘rcorr’ function of
926 the Hmisc package in R (v4.3.0).

927 **Analyses in the Genotype-Tissue Expression (GTEx) project:** TPMs for the *TERT*-β and
928 *TERT*-FL transcripts were downloaded from the GTEx Portal
929 (<https://gtexportal.org/home/downloads/>) within the bulk tissue expression database, GTEx
930 Analysis V8 RNA-seq. Pre-computed EXTEND scores based on a 13-gene signature were
931 obtained from the Supplementary Information of the corresponding publication²⁹. Spearman rank
932 correlations between *TERT* expression (*TERT*-β and *TERT*-FL) and EXTEND scores for each
933 tissue type were determined using the ‘rcorr’ function of the Hmisc package in R (v4.3.0). The
934 eQTLs for rs10069690, rs2242652 and *TERT* expression were assessed through the GTEx portal.

935 **CFSE proliferation assay:** For each condition, cells (9.6×10^5) were stained with a 5 μM
936 solution of carboxyfluorescein succinimidyl ester (CFSE) dye (CellTrace CFSE Cell
937 Proliferation Kit, Thermo Fisher) for 15 min at 37°C. Culture media containing 10% CS FBS
938 was added to an equal volume of staining solution to quench excess dye. CFSE-stained cells
939 (1.2×10^5) were seeded into each well of a 6-well plate in CS serum medium and incubated at
940 37°C and 5% CO₂. The remaining CFSE-stained cells were analyzed on an AttuneNxT (Thermo
941 Fisher) flow cytometer to determine the day 0 (maximal) CFSE intensity. Seeded cells were
942 grown for 48 h in CS serum medium to allow all cell lines to reach a sufficient level of
943 attachment for a medium change and then switched to either full serum or CS serum medium.
944 Forty-eight hours after the media were changed, the cells were harvested with 0.05% trypsin-
945 EDTA and analyzed by flow cytometry to determine the final CFSE intensities. The data were
946 analyzed using FlowJo v10. The CFSE mean fluorescence intensity (MFI) was determined by
947 taking the geometric mean of fluorescence (collected on the BL1 channel, 530/30 nm) after
948 gating live single cells.

949 **CRISPR/Cas9 genome editing:** CRISPR/Cas9 guide RNAs flanking the VNTR6-1 region
950 (1,267 bp in the reference genome) were designed using sgRNA Scorer 2.0¹⁸. Annealed
951 oligonucleotides corresponding to two guide RNAs (**Table S20**) were cloned using Golden Gate
952 Assembly cloning into PDG458 (ref¹⁹, Addgene plasmid #100900;
953 <http://n2t.net/addgene:100900>; RRID:Addgene 100900, a gift from Paul Thomas). The cells
954 (1×10^6) were transiently transfected with CRISPR/Cas9-expressing plasmids using the Amaxa
955 4D nucleofection system (Lonza), a 100 μl SF cell line kit, and the CM-130 program (A549
956 profile settings were used for all cell lines). GFP-positive cells were enriched by FACS 48 hours
957 post-transfection using an SH800 sorter (Sony). The enriched population was further single-cell
958 sorted in 96-well plates to isolate pure knockout populations. Genomic DNA from the expanded
959 clones was screened by PCR with primers VNTR6-1F and VNTR6-1R (**Table S20**). These
960 primers generate a 2,241-bp PCR product (based on the reference genome sequence) and a 974-

961 bp PCR product after knockout. Three independent knockout clones (V6.1-KOs) were selected
962 for functional analyses. Clones that were exposed to CRISPR reagents but did not result in
963 knockout were compared with parental controls (WT, no CRISPR treatment) by RNA-seq
964 analysis. CRISPR treatment had negligible effects on gene expression, and statistical analysis of
965 RNA-seq data was performed comparing V6.1-KOs to WT.

966 **Cloning:** The pCMV-TERT-FL-HA expression construct was generated with high-fidelity Q5
967 polymerase (NEB), starting from the plasmid for *TERT*-FL (GenScript OHu25394), using a
968 forward primer with an AgeI recognition site and a reverse primer with HA-tag and BsrGI
969 recognition sites (**Table S20**). PCR fragments were isolated by electrophoresis and a gel
970 extraction kit (Qiagen) and cloned into an mEGFP-N1 expression vector (Addgene #54767)
971 using AgeI and BsrGI restriction enzymes (NEB), replacing mEGFP. The pCMV-TERT- β -
972 3xFLAG expression construct was generated using two separate Q5 PCRs from the same *TERT*-
973 FL plasmid. The first PCR utilized the same AgeI forward primer and a reverse primer with a
974 native BamHI recognition site within *hTERT* exon 9. The second PCR utilized the BamHI site in
975 its forward primer and a reverse primer with 3xFLAG-tag and a BsrGI recognition site (**Table**
976 **S20**). These two PCR fragments were isolated by electrophoresis and a gel extraction kit
977 (Qiagen), cloned into pCR4 Blunt-TOPO (Invitrogen), and subcloned into the mEGFP-N1
978 expression vector using AgeI+BamHI and BamHI+BsrGI, replacing mEGFP.

979 **RNA extraction:** Cell lysates were harvested from culture plates using 350 μ l of RLT lysis
980 buffer/well and stored at -80°C before extraction. RNA was extracted with the Qiagen RNeasy
981 Mini RNA kit using QIAcube with standard on-column DNase treatment (Qiagen). RNA
982 concentrations were quantified with a Qubit RNA High Sensitivity Kit (Invitrogen).

983 **cDNA synthesis:** 7.5 μ g of RNA from each sample was used in 20 μ l reactions with the iScript
984 Advanced cDNA Synthesis Kit (Bio-Rad). The cDNA was concentrated overnight by ethanol
985 precipitation and resuspended in 37.5 μ l of water, resulting in an RNA input concentration of
986 200 ng/ μ l.

987 **Expression assays:** Expression of the *TERT*- β and *TERT*-FL transcripts was quantified with two
988 custom TaqMan gene expression assays (Thermo Fisher, **Table S20**) designed to target specific
989 exons and splice junctions. Reactions were multiplexed to include both targets and a custom
990 human *HPRT1* endogenous control (NED/MGB probe, primer limited, Assay ID:
991 Hs99999909_m1, Thermo Fisher). TaqMan reactions were run in technical quadruplicate in 384-
992 well plates on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). Each 6 μ l
993 reaction included 2 μ l of cDNA diluted to 100 ng/ μ l from a 200 ng/ μ l RNA input. All assays
994 (individually and in multiplexed reactions) were validated using the *TERT*-FL-HA and *TERT*- β -
995 3xFLAG plasmids in a 5x 10-fold dilution series (from 100 pM to 10 fM). All assays had
996 experimentally determined PCR efficiencies of 72-100%. The identities of the PCR products
997 were confirmed by cloning into TOPO-pCR4 vector (Invitrogen) and Sanger sequenced using
998 M13_TOPO primers (**Table S20**).

999 SYBR Green RT-qPCR assays were performed with iTaq Universal SYBR Green Supermix
1000 (Bio-Rad). The samples were run in 5 µl reactions with 2 µl of cDNA diluted to 50 ng/µl from
1001 the RNA input in 12 technical replicates on a QuantStudio 7 Flex Real-Time PCR System. The
1002 primers (10 mM, Thermo Fisher) used were identical to those used in the TaqMan assays.
1003 *HPRT1* controls (**Table S20**) were run in parallel reactions. For visualization, technical replicates
1004 of selected RT-qPCR products were pooled and resolved on 2% agarose gels, along with a low-
1005 molecular-weight DNA ladder (NEB). Gel images were captured on a Bio-Rad ChemiDoc
1006 Imaging System and analyzed using Image Lab Software v6.1.0 (Bio-Rad). The ratios of *TERT*
1007 isoforms were calculated based on gel densitometry of the PCR products (120 bp and 302 bp).

1008 Total *TERT* expression was measured in 5 µL reactions using TaqMan assays (FAM, exons 3-4)
1009 with *TERT*-Hs00972650_m1 multiplexed with the endogenous control *HPRT1* (VIC, primer-
1010 limited, Assay ID: Hs99999909_m1) and TaqMan Gene Expression Buffer (all from Thermo
1011 Fisher).

1012 **RNA-seq:** RNA quality (all RINs>9.0) was verified using the Bioanalyzer (Agilent) and an RNA
1013 6000 Nano Kit (Agilent). For each sample, 200 ng of total RNA was used to prepare an adapter-
1014 ligated library with the KAPA RNA HyperPrep kit with RiboErase (HMR) (KAPA Biosystems)
1015 using xGen Dual Index UMI Adapters (IDT). The multiplexed libraries with 250-350 bp inserts
1016 were sequenced on a NovaSeq 6000 (Illumina) to generate 279 to 418 million paired-end 150 bp
1017 reads per sample. Quality assessment of RNA-seq data was conducted using MultiQC (v1.16)²⁰.
1018 Quantification of transcript abundance was performed using Salmon (v0.14.1) in count mode
1019 with `—validateMappings` flag and expressed as transcripts per million (TPM). The raw RNA-
1020 sequencing reads were aligned with STAR²¹ based on the reference genome GRCh38 and
1021 GENCODE annotation (v36). Differential expression analysis was conducted with DESeq2
1022 (v1.40.2) based on the estimated counts obtained from Salmon quantification, controlling for the
1023 false discovery rate (FDR). Gene-level transcript abundances were estimated with
1024 ‘lengthScaledTPM’ in the R package ‘tximport’ (v1.28.0). Gene Ontology (GO) analysis and
1025 gene set enrichment analysis (GSEA) on differentially expressed genes was conducted with
1026 clusterProfiler (v4.8.3).

1027 **G4 Hunter prediction analysis:** Analysis was performed with G4Hunter
1028 (<https://bioinformatics.ibp.cz>)²². PacBio-generated DNA sequences for UMUC3 (24 repeat
1029 copies per each allele) and HG03516 (27 and 66.5 repeat copies per allele) were used as inputs
1030 flanked by 120 bp on each side of the V6.1 region.

1031 **G4-seq analysis:** For the lymphoblastoid cell line NA18057 (VNTR6-1-Short/Short genotype,
1032 24 and 27 repeat copies), ChIP-seq data for G quadruplexes (G4) detected in forward and reverse
1033 orientations were downloaded from BED files from the GEO dataset GSE63874 (ref²³, files
1034 GSE63874_Na_K_minus_hits_intersect.bed.gz and
1035 GSE63874_Na_K_plus_hits_intersect.bed.gz). These files were merged into a single BED file
1036 and converted to the UCSC BED format. The G4 mismatch quantification bedGraph files

1037 GSE63874_Na_K_12_minus.bedGraph.gz and GSE63874_Na_K_12_plus.bedGraph.gz were
1038 downloaded and converted into bigwig format using the bedGraphToBigWig tool.

1039 Similarly, for the 293T normal embryonic kidney cell line (V6.1-Long/Long genotype), the G4-
1040 seq data were downloaded from GSE110582 (ref²⁴, files GSM3003539_Homo_all_w15_th-
1041 1_minus.hits.max.K.w50.25.bed.gz and GSM3003539_Homo_all_w15_th-
1042 1_plus.hits.max.K.w50.25.bed.gz) and processed as above. The G4 mismatch quantification
1043 values were downloaded from GSM3003539_Homo_all_w15_th-1_minus.K.bedGraph.gz and
1044 GSM3003539_Homo_all_w15_th-1_plus.K.bedGraph.gz. The G4-seq tracks for NA18057 and
1045 293T cells were visualized through the UCSC Genome Browser (GRCh37).

1046 **Evaluation of G4 ligands:** Five G4 stabilizing ligands were tested for their ability to stabilize
1047 TERT G4. Ligands: PhenDC3, TMPyP4, BRACO-19 and Pyridostatin were provided by Dr.
1048 John Schneekloth. Pidnarulex (CX5461) was selected from the literature²⁵ and obtained from
1049 Selleck Chem. For optimization, UMUC3 cells (4×10^5) were seeded into each well of a 6-well
1050 plate. After adhering for 24 hours, the cells were treated for 24, 48, or 72 hours with ligands at
1051 0.1 μ M, 0.3 μ M, 1 μ M, 3 μ M, 10 μ M, or 30 μ M dissolved in DMSO, with DMSO vehicle alone
1052 and untreated control samples included on each plate. In the 72-hour group, the media was
1053 replaced at 48 hours, and the cells were harvested at 72 hours. The viability of the treated cells
1054 was evaluated by cell counting with a Lionheart FX automated microscope (Agilent) every 24
1055 hours. Pidnarulex (CX5461) and PhenDC3 at 3 μ M for 72 hours were identified as the most
1056 effective treatments for modulating *TERT* exon 7-8 skipping and were used in subsequent
1057 experiments. UMUC3 WT and V6.1 KO cells were treated in technical replicates in three
1058 independent experiments.

1059 **Western blot:** BCA-normalized protein samples and 10 μ L of SeeBlue Plus2 ladder were loaded
1060 and run on gels using 1X Bolt running buffer at 165 V for 1 hour and transferred to nitrocellulose
1061 membranes using an iBlot2 dry transfer instrument (Invitrogen). The membranes were blocked
1062 with 5% milk in 1X TBST for 1 hour at room temperature. The membranes were incubated
1063 overnight at 4°C with primary antibodies in 2.5% milk in 1X TBST (anti-GFP: Invitrogen A-
1064 11122; anti-HA: Novus NB600-362; anti-FLAG: Sigma M2; anti-GAPDH: Abcam ab9485).
1065 After three 5-min washes with 1X TBST, the membranes were incubated at room temperature
1066 for 1 hour with secondary antibodies (anti-rabbit: Cell Signaling 7074; anti-mouse: Cell
1067 Signaling 7076; anti-goat: Santa Cruz sc-2304) and imaged using Pico and Femto ECL reagents
1068 (Thermo).

1069 **Structured illumination microscopy fluorescence imaging:** A549 cells were chosen for
1070 imaging of mitochondria because this highly transfectable cell line has a larger cytoplasmic area
1071 than UMUC3, allowing better visualization. The cells were seeded in a 12-well plate at 1.25×10^5
1072 cells/ml and co-transfected with pCMV-TERT-FL-HA or pCMV-TERT- β -3xFLAG expression
1073 constructs at a 50:50% isoform ratios. Transfections were performed using Lipofectamine 3000
1074 for 4 hrs. Transfected cells were washed with DPBS, dissociated using Accutase (StemPro), and

1075 counted. The cells were then diluted and seeded onto CultureWell Chambered Coverglass
1076 (Intigrogen). After 48 hours, the coverslips were fixed with 4% formaldehyde in PBS for 10 min,
1077 permeabilized with 0.03% Triton-X 100 for 10 min, and blocked with blocking buffer (5% BSA
1078 + 0.01% Triton-X 100 in PBS) for 30 min. The coverslips were incubated at 4°C overnight with
1079 the following primary antibodies: anti-FLAG (Sigma M2, mouse, 1:400 dilution), anti-HA
1080 (Novus NB600-362, goat, 1:400 dilution), and anti-TOM20 (Proteintech 11802-1-AP, rabbit,
1081 1:1000 dilution) diluted in blocking buffer, followed by incubation at room temperature for 30
1082 min with the following secondary antibodies: anti-mouse-AlexaFluor488 (Thermo Fisher
1083 A21202, 1:500 dilution), anti-goat-AlexaFluor647 (Thermo Fisher A32849, 1:500 dilution), and
1084 anti-rabbit-AlexaFluor555 (Thermo Fisher A31572, 1:500 dilution) diluted in blocking buffer.
1085 Three washes were performed with PBS between all staining steps; after the final wash, the cells
1086 were counterstained with 3 µg/ml DAPI. The coverslips were then mounted onto glass slides
1087 with ProLong Gold Antifade Mountant (Invitrogen) and sealed with clear nail polish. Super-
1088 resolution structured illumination microscopy fluorescence images were obtained using ZEN
1089 Black software on an ELYRA PS.1 Super Resolution (SR) microscope (Carl Zeiss, Inc.) with a
1090 Plan-Achromat 63X/1.4 NA oil objective and a Pco.edge sCMOS camera, 405 nm/488 nm/561
1091 nm/633 nm laser illumination and standard excitation and emission filter sets. Raw data were
1092 acquired by projecting grids onto the sample generated from the interference from a phase
1093 grating with 23 µm, 28 µm, and 34 µm spacings for 405 nm, 488 nm and 561 nm excitation,
1094 respectively (3 grid rotations and 5 grid shifts for a total of 15 images per super-resolved z-plane
1095 per color). The raw images were processed with ZEN black software. For publication, images
1096 were scaled to 8-bit RGB identically with a linear LUT and exported in TIFF format using
1097 ImageJ. Figures were made from the TIFF images in Adobe Illustrator without any change in
1098 resolution, except for the inset zoomed images.

1099 **Apoptosis assay:** Cells (1.2×10^5) were seeded in each well of 6-well plates (Corning), and the
1100 media was changed 48 h later to complete serum, CS serum medium or medium supplemented
1101 with 10 µM cisplatin. Cells were harvested with 0.05% trypsin-EDTA 48 h after media change,
1102 pelleted at 500 ×g for 5 min, and washed with 1 mL of PBS. The cells were stained with an
1103 Annexin V-FITC conjugate (Thermo Fisher) and propidium iodide (Thermo Fisher) in Annexin
1104 V staining buffer (Thermo Fisher) according to Rieger et al²⁶. FITC (ex.488 nm/em.517 nm) and
1105 PI (ex.488 nm/em. 617 nm) fluorescence were analyzed by flow cytometry on an Attune NxT
1106 with a CytKick Autosampler (Thermo Fisher). Unstained cells, Annexin V-FITC-stained cells,
1107 and PI-stained UMUC3 cells were used as compensation controls. Apoptosis was determined by
1108 the percentage of FITC-positive cells.

1109 **Cell density analysis:** UMUC3 and three V6.1-KO clones were seeded in 6-well plates (Falcon)
1110 at 4×10^4 cells/well in EMEM. After adherence for 24 hours, the cells were grown for 72, 96, 120,
1111 or 144 hours, with one time point per plate. The cells were washed with 1 mL of PBS and
1112 harvested on plates with RLT lysis buffer (Qiagen). RNA was extracted using the QiaCube

1113 RNeasy Mini protocol (Qiagen), followed by cDNA preparation, *TERT* qPCR, and gel
1114 densitometry as described above.

1115 **xCELLigence Real-Time Cell Analysis (RTCA):** For RTCA, 5637 cells were seeded in a 12-
1116 well plate at 1.25×10^5 cells/ml and transfected with either GFP, pCMV-TERT-FL-HA, or
1117 pCMV-TERT- β -3xFLAG expression constructs either as single transfection or co-transfection at
1118 different ratios of isoforms (80:20%, 50:50% and 20:80%). Transfections were performed using
1119 Lipofectamine 3000 for 4 hrs. Transfected cells were washed with DPBS, dissociated using
1120 Accutase (StemPro), and counted. The cells were then diluted and seeded into an xCELLigence
1121 E-Plate 16 microplate (Agilent) at 1.0×10^3 cells/well and placed on an xCELLigence RTCA DP
1122 system (Agilent). The data were collected every 15 minutes in RTCA software for 288 hours and
1123 then exported for analysis.

1124 In a separate experiment, 1.0×10^3 UMUC3-WT or UMUC3-V6.1-KO cells grown in CS serum
1125 medium were seeded into each well of an E-Plate 16 (Agilent). Cell label-free impedance in the
1126 E-Plate (correlated with cell proliferation) was measured every 15 minutes for 283 hours using
1127 an xCELLigence RTCA DP system. Two days after seeding, the medium was changed to either
1128 full serum medium or CS serum medium (control).

1129 Linear mixed models were applied to the impedance data obtained from the xCELLigence
1130 system, where the treatment type was considered a fixed effect term and the technical replicate
1131 was considered a random effect term. Maximum likelihood estimation procedures were
1132 employed to conduct joint effects likelihood-ratio tests, while restricted maximum likelihood
1133 estimation was utilized for more precise estimation of effect sizes as beta coefficients using the
1134 linear mixed-effects function in the R package ‘nlme’ (v3.1–162).

1135 **HiChIP analysis:** The H3K27Ac HiChIP libraries for the bladder cancer cell lines T24 and RT4
1136 were generated using the Arima-HiChIP protocol (Arima Genomics, A101020). Briefly, 1×10^6
1137 cells/replicate were collected for chromatin cross-linking followed by digestion with a restriction
1138 enzyme cocktail, biotin labeling, and ligation. The samples were then purified, fragmented, and
1139 enriched. Pulldown was performed using an antibody against H3K27ac (Cell Signaling, #8173).
1140 The Arima-HiChIP libraries that passed QC were sequenced using an Illumina NovaSeq 6000 to
1141 generate raw FASTQ files for each sample. The paired-end reads were aligned to the GRCh37
1142 genome using the HiC-Pro pipeline (v3.1.0, <https://github.com/nservant/HiC-Pro>). The
1143 confirmed interaction reads were used as input for significant loop calling via the FitHiChIP tool
1144 (v.11.0, <https://github.com/ay-lab/FitHiChIP>) with default settings. The HiChIP loop and ATAC
1145 peak calling files for the GM12878 and normal bladder samples were downloaded from the Gene
1146 Expression Omnibus (GSE188401). The interactions were visualized through the UCSC genome
1147 browser.

1148 **PacBio DNA methylation analysis:** Freshly collected genomic DNA (5 μ g) from the HT1376,
1149 RT4, T24, SCAber, UMUC3, and Raji cell lines was sheared using Covaris g-tubes at 4800
1150 rpm, followed by size selection using PippinHT. Three SMRT flow cells were run for each

1151 sample library on the PacBio Sequel II platform. The sequence reads were transformed into
1152 FASTQ and aligned to the GRCh38 reference genome using the default settings of the SMRT-
1153 Link workflow. 5mC DNA methylation analysis was a part of the SMRT-Link pipeline, and the
1154 corresponding information specifying the positions and probabilities of 5mC methylation at CpG
1155 sites was integrated into the output file.

1156 **Oxford Nanopore cDNA-seq:** cDNA libraries were generated using the PCR cDNA
1157 Sequencing Kit SQK-DCS109 (Oxford Nanopore Technologies), starting with 100 ng of poly-A
1158 RNA. Libraries were loaded onto R9.4.1 PromethION flow cells mounted on a P2 Solo and run
1159 for 96 hours. Basecalling was performed using MinKNOW software with the high-accuracy
1160 model on a GridION sequencer (Oxford Nanopore Technologies). Reads were aligned to
1161 GRCh38 via Minimap2 (v2.26) and SAMtools (v1.5). UMUC3 yielded 25,827,200 reads, with
1162 46 reads aligning to *TERT*, whereas UMUC3 V6.1 KO yielded 18,709,848 reads, with 62 reads
1163 aligning to *TERT*.

1164 **Analysis of sequence conservation in non-human species:** Haplotype-resolved Telomere-to-
1165 Telomere (T2T) assemblies of primates were downloaded from GenomeArk
1166 (<https://www.genomeark.org/>). The FASTA sequences were aligned to the human GRCh38
1167 reference genome using Minimap2 (v2.26) with the '-ax asm10' flag and converted to a BAM
1168 file using SAMtools (v1.5). The *TERT* V6.1 repeat units were analyzed with Tandem Repeat
1169 Finder (<https://tandem.bu.edu/trf/trf.html>). The BAM files of Neandertal (n=3) and Denisova
1170 (n=1) individuals were downloaded from the Max Planck Institute for Evolutionary
1171 Anthropology resource (http://cdna.eva.mpg.de/neandertal/Vindija/bam/Pruefer_et_al_2017/ and
1172 <http://cdna.eva.mpg.de/neandertal/Chagyrskaya/>) and visualized using IGV.

1173 **Statistical analysis:** Unless indicated, analyses were performed with R Studio (v4.3.0),
1174 GraphPad Prism (v.8) and FlowJo (v9); p values are for unpaired two-sided Student's t tests or
1175 linear regression adjusted for the indicated covariates.

1176 This work utilized the computational resources of the NIH HPC Biowulf cluster
1177 (<http://hpc.nih.gov>).

1178

1179 **METHODS-ONLY REFERENCES**

- 1180 1. Gadalla, S.M. *et al.* Donor telomere length and causes of death after unrelated hematopoietic
1181 cell transplantation in patients with marrow failure. *Blood* **131**, 2393-2398 (2018).
- 1182 2. Thomas, N. *et al.* Genetic subgroups inform on pathobiology in adult and pediatric Burkitt
1183 lymphoma. *Blood* **141**, 904-916 (2023).
- 1184 3. Grande, B.M. *et al.* Genome-wide discovery of somatic coding and noncoding mutations in
1185 pediatric endemic and sporadic Burkitt lymphoma. *Blood* **133**, 1313-1324 (2019).
- 1186 4. Liao, W.W. *et al.* A draft human pangenome reference. *Nature* **617**, 312-324 (2023).

- 1187 5. Kolmogorov, M. *et al.* Scalable Nanopore sequencing of human genomes provides a
1188 comprehensive view of haplotype-resolved variation and methylation. *Nat Methods* **20**, 1483-
1189 1492 (2023).
- 1190 6. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000
1191 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440 e19 (2022).
- 1192 7. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via
1193 Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
- 1194 8. Lin, P. *et al.* A new statistic to evaluate imputation reliability. *PLoS One* **5**, e9697 (2010).
- 1195 9. Hasson, M.A. *et al.* Design and evolution of the data management systems in the Prostate, Lung,
1196 Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* **21**, 329S-348S (2000).
- 1197 10. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
1198 **562**, 203-209 (2018).
- 1199 11. Codd, V. *et al.* Measurement and initial characterization of leukocyte telomere length in 474,074
1200 participants in UK Biobank. *Nat Aging* **2**, 170-179 (2022).
- 1201 12. Korn, J.M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy
1202 number polymorphisms and rare CNVs. *Nat Genet* **40**, 1253-60 (2008).
- 1203 13. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
1204 *Nature* **590**, 290-299 (2021).
- 1205 14. Gao, G.F. *et al.* Before and After: Comparison of Legacy and Harmonized TCGA Genomic Data
1206 Commons' Data. *Cell Syst* **9**, 24-34 e10 (2019).
- 1207 15. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an association
1208 with cancer stemness and proliferation. *Nat Commun* **12**, 139 (2021).
- 1209 16. Malta, T.M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic
1210 Dedifferentiation. *Cell* **173**, 338-354 e15 (2018).
- 1211 17. Barthel, F.P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer
1212 types. *Nat Genet* **49**, 349-357 (2017).
- 1213 18. Chari, R., Yeo, N.C., Chavez, A. & Church, G.M. sgRNA Scorer 2.0: A Species-Independent Model
1214 To Predict CRISPR/Cas9 Activity. *ACS Synth Biol* **6**, 902-904 (2017).
- 1215 19. Adikusuma, F., Pfitzner, C. & Thomas, P.Q. Versatile single-step-assembly CRISPR/Cas9 vectors
1216 for dual gRNA expression. *PLoS One* **12**, e0187236 (2017).
- 1217 20. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for
1218 multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-8 (2016).
- 1219 21. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
- 1220 22. Brazda, V. *et al.* G4Hunter web application: a web server for G-quadruplex prediction.
1221 *Bioinformatics* **35**, 3493-3495 (2019).
- 1222 23. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P. & Balasubramanian, S.
1223 High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat*
1224 *Biotechnol* **33**, 877-81 (2015).
- 1225 24. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species.
1226 *Nucleic Acids Res* **47**, 3862-3874 (2019).
- 1227 25. Li, G. *et al.* Alternative splicing of human telomerase reverse transcriptase in gliomas and its
1228 modulation mediated by CX-5461. *J Exp Clin Cancer Res* **37**, 78 (2018).
- 1229 26. Rieger, A.M., Nelson, K.L., Konowalchuk, J.D. & Barreda, D.R. Modified annexin V/propidium
1230 iodide apoptosis assay for accurate assessment of cell death. *J Vis Exp* (2011).