# Human telomere length is chromosome specific and conserved across individuals

Kayarash Karimian[1,2], Aljona Groot[3], Vienna Huso[1,2], Ramin Kahidi[4], Kar-Tong Tan[5,6,7], Samantha Sholes[1,2,8], Rebecca Keener[9], John F. McDyer[10], Jonathan K. Alder[10], Heng Li [11,12], Andreas Rechtsteiner[3], Carol W. Greider[1,3,*]

[1]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

[2]Biochemistry, Cellular and Molecular Biology Graduate Program, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

[3]Department of Molecular Cell and Developmental Biology, University of California, Santa Cruz,

[4] University of Calgary, Calgary, AB, Canada

[5] Harvard Medical School, Department of Genetics, Boston, MA

[6] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA

[7] Broad Institute, Cancer Program, Cambridge, MA,

[8] Present address Merck & Co., 770 Sumneytown Pike, West Point, PA 19486

[9] Department of Biomedical Engineering, Johns Hopkins University,

[10] Pulmonary, Allergy, Critical Care, and Sleep Medicine Division, Department of Medicine, University of Pittsburgh

[11] Dana-Farber Cancer Institute, Department of Data Sciences, Boston, MA,

[12] Harvard Medical School, Department of Biomedical Informatics, Boston, MA

*Corresponding Author

Carol W Greider, PhD
Department of Molecular, Cell & Developmental Biology
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
E-mail: cgreider@ucsc.edu

Short telomeres cause age-related disease and long telomeres predispose to cancer; however, the mechanisms regulating telomere length are unclear. Current methods for telomere length measurement are not direct, precise, or widely accessible. Here we describe a direct nanopore telomere profiling assay using an Oxford Nanopore Technology (ONT) MinION that is easy to implement, precise, and cost effective with broad applications in research, and the clinic. Telomere length measurement is currently used in clinical settings to make highly consequential treatment decisions. Using patient samples our method returned similar results to the clinical, FlowFISH assay. Telomere profiling enables mapping of the telomere to specific chromosomes, and we identified both chromosome-specific and haplotype-specific telomere length distribution with a remarkable 6kb difference between some telomere lengths. Further, we found that specific chromosome ends were consistently shorter or longer than the average length across 150 individuals. The presence of conserved chromosome end-specific telomere lengths suggests there are new paradigms in telomere biology that are yet to be explored. Understanding these mechanisms will allow deeper insights into telomere biology that can lead to new approaches to disease.

## Introduction

Human health is profoundly affected by telomere length, yet the detailed mechanism of length regulation is poorly understood. Telomere length is maintained as an equilibrium distribution with constant shortening at each round of DNA replication, which is counterbalanced by *de novo* addition of new telomere repeats by the enzyme telomerase [1]. Failure to maintain the length distribution leads to inherited Short

60 Telomere Syndromes and age-related degenerative disease such pulmonary fibrosis,
61 immunodeficiency, and bone marrow failure [2]. Conversely long telomeres predispose
62 people to cancer [3], and a cluster of mutations that increases telomerase activity is one
63 of the most common mutational signatures in cancer [4,5].
64
65 Precisely how telomerase action maintains a length equilibrium is of great interest. The
66 prevailing 'protein counting' model for length maintenance [6,7] proposes that proteins that
67 bind telomeric TTAGGG repeats negatively regulate telomere elongation in *cis.* This is
68 supported by evidence that telomerase stochastically elongates short telomeres more
69 frequently than long telomeres [8]. Together these studies propose that the length
70 equilibrium is maintained by telomerase lengthening short telomeres to precisely
71 counterbalance shortening of all telomeres. An implication of this model is that all
72 telomeres will be regulated around a similar mean length distribution.
73
74 Methods for measuring telomere length have had significant influence on length
75 regulation models. Southern blotting was first established to measure the length of all
76 telomeres in a large population of cells simultaneously [9] Szostak, 1982 #170}.
77 Telomere Southerns probed with the telomere repeat sequence reveal a distribution of
78 lengths of every chromosome end in every cell in the population. The distribution of
79 lengths is difficult to quantitate, and absolute lengths differ significantly between labs [10].
80 The protein counting model was based on data from Southern blots which explains the
81 focus on the regulation of the distribution of lengths across all telomeres [11-13]. The
82 adoption of FlowFISH as the clinical standard for diagnosis of telomere diseases [14-16]

83 may have reinforced the thinking in the field that telomeres are maintained around a
84 global length distribution. FlowFISH measures the average of all the TTAGGG repeats
85 in patients' lymphocytes and is normalized to the median length of all telomeres on a
86 Southern blot [17]. The fact that this method is robust and accurately identifies telomere
87 mediated disease may imply to some that the global telomere length average is the
88 biologically relevant measurement, when in fact it might not be.
89
90 qFISH allows measurement of individual telomeres by fluorescence by *in situ*
91 hybridization on each telomere in a metaphase spread [18]. qFISH experiments have
92 suggested that telomeres on all chromosome arms are not globally regulated around a
93 common length distribution [19-21], however this data in not yet reconciled with the
94 prevailing protein counting model of length regulation. This may be because
95 sophisticated microscopy and image analysis is required for qFISH and so this method
96 is not readily accessible. Here we describe nanopore telomere profiling, which
97 measures the length of each individual telomere in the cell at nucleotide resolution. With
98 this new method we establish that indeed individual telomeres on specific chromosome
99 ends are maintained around their own unique length distributions, and the means of
100 these distributions can differ by more than 6kb. Nanopore telomere profiling using the
101 ONT MinION is easy to implement, precise, cost effective, and will be of broad use in
102 research, clinical, and biotechnology settings. Telomere profiling represents a paradigm
103 shift in telomere analysis and will enable exploration of entirely new areas of telomere
104 biology.
105

## Results

To determine whether human telomeres are maintained around a common length distribution across all chromosomes or if specific chromosome ends maintain their own unique length distributions, we developed a method to enrich telomere so that we could sequence them by Oxford Nanopore Technology (ONT) long read sequencing. We ligated the telomeric ends with a biotinylated oligonucleotide (TeloTag) that contains a multiplexing barcode and restriction enzyme sites. Following ligation, we pulled down the tagged telomeres with streptavidin beads and released them by restriction enzyme digestion prior to sequencing (Fig. 1a). To assess the enrichment efficiency, we prepared libraries from both telomere enriched and non-enriched samples and sequenced them on an ONT MinION. Enrichment was efficient, recovered ~17% of total telomere input (Extended Data Fig. 1. b, c), and resulted in a ~3400-fold increase in sequenced telomeres (Extended Data Fig.1d and methods). We routinely multiplexed samples and generated ~50,000 telomere reads per flow cell with an average fragment length (subtelomere + telomere) of ~20kb. The cost per sample when multiplexing was approximately $75 (see methods).

## Bioinformatic analysis of telomere length

We developed a bioinformatic pipeline to determine both “bulk length” (all telomeres) as well as chromosome end-specific telomere length. We used ONT Guppy base calling and filtered for reads containing telomere repeats (Methods). We initially determined telomere length using a method we used in yeast [22] that has been previously used in human cells [23-25]. Reads are first mapped to a reference genome with a defined

129 telomere/subtelomere boundary and telomere length is defined as base pairs from the
130 boundary to the end of the read (Methods). However, we found heterogeneity in
131 subtelomere sequences among individuals, where sometimes the subtelomere was
132 slightly longer or shorter than in the reference genome which caused overcalling or
133 under calling of telomere lengths (Extended Data Fig. 2). To overcome this, we
134 developed TeloNP, an algorithm to define the subtelomere boundary and measure
135 telomere length directly from the nanopore sequencing reads while taking systematic
136 nanopore basecalling errors [26] into account. The subtelomere boundary was set at the
137 start of a sustained discontinuity in the telomere repeat pattern when scanning from the
138 end of the read towards the subtelomere (Fig. 1b and Extended Data Fig. 3) (see
139 methods). Telomere length was defined as the base pairs from the TeloTag to the
140 subtelomere boundary on the reads.
141
142 To examine whether telomere length determined by TeloNP after Guppy base calling
143 accurately represents the true length of the telomere repeats, we examined the
144 electrical current signals from the flow cells. We developed TeloPeakCounter to count
145 the repeated current peaks from the (TTAGGG)n in the nanopore Fast5 signal data to
146 test base calling by Guppy (Extended Data Fig. 4a, b) (see methods). Length
147 determined by TeloNP after Guppy base calling was in good agreement with
148 TeloPeakCounter. We note the new ONT Dorado base caller overestimated telomere
149 length compared with both TeloPeakCounter and Guppy (Extended Data Fig. 4 c, d).
150 We therefore adopted Guppy base calling for our analyses.
151

152 **Nanopore telomere profiling accurately and reproducibly reports telomere length**

153 We performed telomere profiling by sequencing DNA from PBMCs (see methods) of
154 people ranging from 0 years to 90 years of age (Fig. 1d) and found agreement with
155 telomere lengths on a Southern blot (Fig. 1c). To test reproducibility, we measured
156 telomere length of DNA from one individual on the same flow cell (intra-assay) (Fig. 1e
157 and Extended Data Fig. 1e) (correlation of variation, CV 1.3%) or on different flow cells
158 (inter-assay) Fig. 1f and Extended Data Fig. 1f (CV 2.4%). This low variability compares
159 well with FlowFish, the that has an inter-assay CV of 2.2%, and considerably
160 outperformed the frequently used qPCR assay that has an inter-assay CV of 25.0% [15].
161 In addition, we tested inter-lab variability by measuring telomere length of seven
162 samples where the same DNA was enriched and sequenced by two different people in
163 two different labs (Johns Hopkins and UCSC) and found highly reproducible results
164 (mean difference of 104.7 bp with SEM of +/- 34 bp) (Fig. 1g). To determine whether
165 any fragment length bias of nanopore sequencing could skew telomere length
166 determination, we compared restriction enzyme cutting with a combination of *BamHI*
167 and *EcoRI* which generates fragments ~9 kb, or with *AsiSI* and *PvuI,* which generate
168 fragments ~25 kb. We found similar telomere lengths (Extended Data Fig. 5a, b)
169 indicating fragment length in these size ranges did not have detectable bias on telomere
170 length determination.

171

172 **Telomere profiling determines telomere shortening with age at nucleotide**
173 **resolution**

174 Telomere length is known to shorten with age [15,27-30], however previous methods could
175 not measure telomere length at nucleotide resolution. To test the dynamic range, we
176 first applied nanopore telomere profiling to DNA samples of 11 individuals from 0-84
177 years of age and ranked them based on descending order for telomere length. We then
178 did a Southern blot on the same DNA. Telomere profiling predicted the rank order of
179 telomere lengths and captured the wide dynamic range of the Southern blot (Fig. 2a and
180 b). Southern blotting does not measure the shortest telomeres because telomere
181 repeats are required for probe hybridization on a Southern. We plotted the 1st, 10th and
182 50th percentile of telomere length as determined by telomere profiling and observed a
183 decrease of the 50th and 10th percentile as the mean length shortened. However, the
184 1st percentile telomere length did not decrease suggesting there is a threshold length in
185 PBMC's of ~1,100 bp below which telomeres cannot be maintained (Fig 2c).
186
187 To directly compare nanopore telomere profiling to FlowFISH, we conducted nanopore
188 telomere profiling on whole blood and PBMCs (see methods) of 132 healthy individuals
189 ranging from 0 to 90 years of age (Fig. 2d). Using the mean telomere length for each
190 individual, we defined 90$^{th}$, 50$^{th}$ and 10$^{th}$ intervals of telomere length at each age using
191 statistical methods used for FlowFISH [15]. While the shapes of the curves are very
192 similar between telomere profiling and FlowFISH, the absolute lengths of the telomeres
193 are longer for FlowFISH. Nanopore telomere profiling of cord blood showed a telomere
194 length distribution with a mean of 7,986 +/- 245 bp (Fig. 2d) across 18 samples. This is
195 shorter than the average for FlowFISH and with less variance (Fig. 2e) [15]. Average cord
196 blood telomere length estimates measured by FlowFISH vary from ~18kb [31] to ~9kb [32]

197 to ~11 kb [15]. FlowFISH fluorescence signal is normalized to Southern blots, which
198 includes some subtelomeric sequences and this may account for the longer telomere
199 lengths of FlowFISH. Furthermore, Southern blot estimated telomere lengths are known
200 to vary between laboratories [10]. In contrast, nanopore telomere profiling offers a precise
201 readout in base pairs that can be directly compared between laboratories.
202
203 To compare our method directly to FlowFISH, we sequenced 5µg of archived DNA from
204 blinded samples of IPF patients previously diagnosed with Short Telomere Syndromes.
205 Telomere profiling showed that bulk telomere length in most IPF samples were similar
206 to the FlowFISH measurement (Fig. 2f). FlowFISH uses flow cytometry and can
207 distinguish telomere lengths in specific cell types from whole blood samples [33,34] and
208 some samples have discordant lymphocyte and granulocyte telomere lengths [15,35,36],
209 Nanopore telomere profiling will report the average length from all cell types in the
210 samples. Thus, while nanopore telomere profiling likely can be used for diagnosing
211 short telomere patients in the future, additional development such as isolation of specific
212 cell types may help to capture heterogeneity of clinical samples.
213
214 **Human telomeres have chromosome end-specific length and haplotype-specific**
215 **length differences**
216 To determine whether humans have chromosome end-specific telomere length, we first
217 examined telomeres from the diploid HG002 cell line for which a high-quality reference
218 genome is available [37]. Human subtelomeres contain many blocks of homology shared
219 between different telomeres (paralogy blocks) [23,38]. Simulation of long read data from

220 CHM13 references genome showed that *minimap2* [39,40] can assign simulated reads to
221 the correct telomere with high accuracy using 10kb of subtelomere sequence [26]. We
222 isolated DNA from HG002 cell line, sequenced the telomeres and mapped reads with
223 an average total length of 16.4 kb (4.6 kb telomere repeats and 11.8 kb sub-telomeric
224 sequence, on average) to the HG002 reference genome using *minimap2* using a
225 customized filtering pipeline (methods). Seventy-seven chromosome ends passed our
226 quality filters, and we found 66 ends had significant differences in length distribution
227 from the grand mean (Fig. 3. a, b). In addition to chromosome end-specific lengths, we
228 also found that some telomeres showed significant differences between the maternal
229 and paternal haplotypes. In some cases, remarkably, there was more than 6kb
230 difference in mean length, for example for chromosome 1p Maternal (1pM) and 1p
231 Paternal (1pP). Thus, like in yeast [22], humans have chromosome end-specific telomere
232 length distributions.
233
234 **Chromosome specific telomere lengths are conserved across individuals.**
235 To determine whether chromosome end-specific differences were conserved across a
236 broad population, we used *minimap2* to map telomere reads from 150
237 individuals to the subtelomere sequences from the recently released pangenome
238 containing 47 high quality T2T assemblies representing 94 haploid genomes [41] and
239 filtered for reads with >2kb of alignment (see methods).
240 We removed the acrocentric and X Y chromosome ends because the
241 high rate of meiotic recombination between these ends across a population would not
242 allow them to map uniquely [42].

243

244 *Minimap2* map quality score (mapq) reports quality of alignment when mapping reads to
245 a single reference genome but is not optimized for mapping to the multiple genomes
246 present in the pangenome as most reads have multiple near identical alignments and
247 thus get low mapq. We compared the pangenome alignment of reads from our 300
248 haploid genomes to the alignment in three high quality haploid reference T2T genomes
249 CHM13, HG002 maternal and HG002 paternal to establish if reads reproducibly
250 mapped to the same subtelomere in each reference genome. The reads were mapped
251 to the respective haploid genome using mapq of 60 (Fig. 4 a, b, c). Pangenome
252 chromosomes are shown by the columns in the matrix heatmaps and T2T haploid
253 genome chromosomes are shown by the rows. The color in each square indicates what
254 fraction of reads mapping to chromosome in the respective haploid T2T genome. The
255 diagonal indicates the fraction of reads mapping to the same chromosome end in the
256 pangenome and the respective haploid genomes. Of all reads, 87% mapped to the
257 same chromosome end in the pangenome and CHM13, 90% in the pangenome and
258 HG002 maternal and 88% in the pangenome and HG002 paternal. We also quantified
259 the percent of reads that mapped to the same chromosome end in the pangenome and
260 all three haploid reference genomes (Fig. 4d and Extended Data Fig. 5 a, b, c). For 33
261 of the 39 chromosomes ends, 100-60% of the reads mapped to the same chromosome
262 end in the pangenome and all three haploid genomes. Six chromosome ends had
263 between 10-20% of reads map to the same chromosome end in the pangenome and all
264 three haploid genomes. When we added back the acrocentric chromosomes, we found
265 0 reads amongst this group mapped back to the same chromosome end in all three

266 references (Extended Data Fig 5d), as expected for reads known to map to different
267 chromosome ends across a population. Together this data suggests that the reads we
268 found mapping to a certain pangenome chromosome map with high confidence.
269
270 To compare the telomere length of each chromosome end across the aging population,
271 we established the relative mean telomere length. For each individual, their grand
272 mean telomere length was subtracted from the their chromosome specific telomere
273 length. The resulting mean centered chromosome telomere lengths were plotted with
274 zero indicating no difference from the individual's grand mean for that chromosome (Fig
275 4e). We ranked the chromosome ends by their relative telomere lengths and found that
276 17p, 20q and 12p tended to be the shortest telomeres in the population while 4q, 12q
277 and 3p tended to be the longest (Fig. 4e). Thus, while haplotype specific differences in
278 telomere length are seen in a single individual (Fig. 3a), across a population, on
279 average, certain chromosome ends are more likely to be shorter while others are more
280 likely to be longer than the grand mean. Remarkably, previous work using qFISH to
281 measure telomere length on metaphase spreads in 10 individuals also found 17p, 20q
282 and 12p among the top 4 shortest and 4q, 12q and 3p among the top 8 longest ends [20]
283 strengthening the conclusion that some chromosome ends are reproducibly shorter or
284 longer than the grand mean.
285
286 To determine whether chromosome end-specific telomere lengths are present at birth,
287 we mapped the reads from cord blood to the pangenome and calculated the relative
288 mean telomere lengths as described above (Fig. 4f). While we had fewer cord blood

289 samples, and therefore fewer chromosome ends met our quality filters, we found again
290 that 17p, 20q and 12p were shorter while 4q, 12q and 3p were longer than the grand
291 mean. This supports previous work [43] that suggested that telomere length at birth is
292 maintained with age.
293
294
295 **Discussion**
296 A fundamental understanding of the mechanisms that regulate telomere length is
297 essential to develop future disease treatments. When the telomere length distribution
298 shifts to shorter lengths, some telomeres become critically short, initiating senescence,
299 [44-47] and can cause age-related degenerative disease in humans [16]. Inherited mutations
300 that shift to a longer equilibrium predispose people to cancer [3,48] and the most frequent
301 somatic mutations in cancer increase telomerase levels and lengthen telomeres [4,5].
302 Nanopore telomere profiling will enable the dissection of how individual telomere
303 lengths on specific chromosomes are maintained and may play a role triggering
304 senescence, and ultimately in disease.
305
306 **Chromosome end-specific telomere length equilibria imply new regulatory**
307 **mechanisms**
308 The predominant protein counting model for telomere length maintenance proposes that
309 telomere proteins that bind TTAGGG repeats repress the elongation of a given telomere
310 in *cis* [6,49] and longer telomeres have more repression, allowing shorter telomeres to be
311 preferentially elongated [8]. This model represents a robust way to maintain a length
312 equilibrium [50]. However, since all telomeres have the same TTAGGG repeats, the

313 model predicts that all telomeres would be regulated around a shared equilibrium
314 length. The demonstration of end-specific lengths indicates that other, yet unknown
315 factors, can play a key role modifying the set point for each unique telomere length
316 distribution.
317
318 In yeast that lack telomerase, all chromosome end-specific length distributions
319 shortened at similar rates [22], suggesting telomere elongation, not shortening, is the
320 major influence on chromosome specific length. Telomere elongation is the sum of the
321 frequency of elongation of any given end (telomerase recruitment) and number of
322 repeats added per elongation event (telomerase processivity). When the sum of these
323 events, on average, equals the rates of telomere shortening, the equilibrium point is set.
324 However, given end-specific length distributions, it is clear that this simple view does not
325 represent the full complexity of the system. There must be factors at specific
326 chromosome ends that regulate telomerase recruitment, processivity, or both, to
327 establish end specific lengths. In addition, stochastic shortening such as telomere rapid
328 deletion [51] or replication fork collapse [52,53] may play yet unknown roles in establishing
329 telomere length equilibrium.
330
331 **Mechanisms that may influence end specific telomere length**
332 Subtelomeric sequences are obvious candidates to regulate end-specific telomere
333 lengths. In yeast, subtelomere DNA binding proteins can affect telomere length [54],
334 although the mechanism is not yet understood. The subtelomeric TAR1 element [55]
335 present in paralogy block 23 [25,38] was proposed to regulate telomere length, possibly

336 through binding CTCF and regulating expression of the lncRNA, TERRA [56-58]. Previous
337 studies suggested that the absence of TAR1 may correlate with shorter telomeres [25].
338 However, we did not find a direct relationship of the shortest telomere with those ends
339 described by Dubocanin *et. al.* that lack TAR1 (8q,13p,14p, 17p, 21p, 22p Xp) in our
340 data set. Future comprehensive analysis of the subtelomere sequences adjacent to long
341 and short telomeres will lead to new testable models for establishment of telomere
342 length equilibria.
343
344 Epigenetic modifications of DNA or histones may influence telomere length [59]. Human
345 and mouse subtelomeric regions are known to be methylated at CpG sites [60] and
346 experiments in mice suggest that loss of DNA methyltransferases results in shorter
347 telomeres [61]. Sequences in the subtelomere could recruit chromatin modifying enzymes
348 that might influence length regulation. Subtelomere sequences may also influence other
349 mechanisms that have been proposed to regulate telomere length including replication
350 timing and tethering to the nuclear periphery [62]. The availability of nanopore telomere
351 profiling will allow exploration of the role of these factors in establishing telomere length
352 equilibria.
353
354 **Chromosome end-specific length differences are present at birth and maintained**
355 **as telomeres shorten with age**
356 Telomere length is inherited from parent to child. Evidence of this comes from the
357 genetic anticipation in patients with Short Telomere Syndromes; short telomeres are
358 passed down to each generation, and the severity of disease increases across

359 generations [63]. Similarly, in mice heterozygous for telomerase deletion, short telomeres
360 are progressively passed down across 6 generations causing progressive severity of
361 disease [64]. Twin studies have also documented the inheritance of telomere length in
362 humans [65].
363
364 Analysis of chromosome end-specific telomere lengths across 150 individuals showed
365 specific telomeres tend to be the longest or shortest, supporting a previous study using
366 qFISH on 10 individuals that identified a similar set of chromosomes as the longest and
367 shortest [20]. Cord blood also showed that 17p, 20q and 12p were among the shortest
368 and 4q, 12q and 3p were among the longest ends suggesting that telomere length
369 differences present at birth are maintained over decades. This establishment of
370 chromosome-end specific telomere length equilibria at birth [43] and maintenance of the
371 equilibria after birth leaves little room for proposed effects of life history, psychological,
372 or environmental exposures [66] on telomere length. The similarity of our data with
373 Martens *et al.* qFISH analysis is remarkable, and our method will enable future studies
374 to explore the biological significance of this finding. We did not prospectively choose our
375 samples to be representative of the diversity of the human population, but rather to span
376 a wide age range. However, future studies could be powered to examine whether
377 certain chromosome ends are consistently the shortest or longest more broadly in a
378 diverse human population.
379
380 **Implications for human disease**

381 Being able to accurately measure chromosome-end specific telomere length has
382 important implications for human disease. Nanopore telomere profiling determines
383 nucleotide resolution of the length distribution and can distinguish the length of specific
384 chromosome ends unlike Southern blots, qPCR, or FlowFISH assay. Telomere profiling
385 employs the accessible MinION instrument that can be used in-house in any research
386 or clinical lab, with very low start-up costs, allowing for equitable access to telomere
387 length determination methods. This method provides the opportunity to prospectively
388 develop clinical standards analogous to those for FlowFISH and may allow clinical
389 length measurements in samples other than blood. In addition, having a highly
390 reproducible assay that can be easily automated will enable experimental approaches
391 to define new regulators of telomere length. The role telomere elongation in the
392 immortalization of cancer cells has been known since 1990 [28,67,68]. Having a precise tool
393 that can be automated, will allow new approaches that may exploit telomere length
394 modulation in cancer treatment. Finally, the identification of conserved chromosome
395 end-specific telomere lengths implies that new, undiscovered biological mechanisms
396 influence telomere length. Nanopore telomere profiling will empower the field as a whole
397 to dissect these mechanisms, leading to new discoveries in telomere biology.
398

399 **Methods**

400 **Cells and Cell Lines**

401 PMBCs were purchased from Stem Cell Technologies, ZenBio Inc, and Precision for Medicine.
402 Samples were chosen from the repository based on age to span the distribution from 0 to 95.
403 Cord blood was purchased from Stem Cell Technologies to represent the youngest individuals
404 in the population. Blood samples used to initially calibrate the assay were de-identified excess

405 samples from Johns Hopkins Hospital, certified as exempt by the John Hopkins IRB. HG002
406 cells were cultured in RPMI 1640 media (Gibco, Cat.11875093) supplemented with 2g/L
407 glucose 2mM L-glutamine (Glutamax, Gibco, Cat.35050061) 15% fetal bovine serum (Gibco,
408 Cat.26140079) and 1% penicillin-streptomycin (Gibco, Cat.15140122). PBMCs were counted
409 using the Luna II hemocytometer (VitaScientific, Cat.LGBD10029).

410 **Telomere Southern blot analysis**

411 Genomic DNA was isolated using the Promega Wizard gDNA kit (Cat.A1120, Promega) and
412 quantified by QuBit 3.0 (Thermo Fisher) using the DNA kit (Q32853; Thermo Fisher).
413 Approximately 1 µg of genomic DNA was restricted with *Hinf*I (NEB, Cat.R0155M) and *Rsa*I
414 (NEB, Cat.R0167L,) and resolved by 0.8% Tris-acetate-EDTA (TAE) agarose gel
415 electrophoresis (Invitrogen, Cat.EA0375BOX). 10 ng of a 1kB Plus DNA ladder (NEB,
416 Cat.N3200) was included as a size reference. Following denaturation (0.5 M NaOH, 1.5M NaCl)
417 and neutralization (1.5 M NaCl, 0.5 M Tris-HCL, pH 7.4) the DNA was transferred in 10x SSC
418 (3M NaCl, 0.35 M NaCitrate) to a Nylon membrane (GE Healthcare, Cat. RPN303B) by vacuum
419 blotting (Boekel Scientific). The membrane was UV crosslinked (Stratagene), prehybridized in
420 Church buffer (0.5M Na2HP04, pH7.2, 7% SDS, 1mM EDTA, 1% BSA), and hybridized
421 overnight at 65°C using a radiolabeled telomere fragment and ladder, as previously described
422 (Morrish and Greider 2009). The membrane was washed twice with a high salt buffer (2x SSC,
423 0.1% SDS) and twice with a low salt buffer (0.5X SSC, 0.1% SDS) at 65°C, exposed to a
424 Storage Phosphor Screen (GE Healthcare), and scanned on a Storm 825 imager (GE
425 Healthcare). The images were copied from ImageQuant TL (GE Life Sciences) to Adobe
426 PhotoShop CS6, signal was adjusted across the image using the curves filter, and the image
427 was saved as a .tif file.

428 **Preparation of HMW DNA**

429 A modified DNA extraction protocol was used to produce high molecular weight DNA based on
430 the Lucigen/EpiCentre's MasterPureTM Complete DNA and RNA Purification Kit A (Biosearch
431 Technologies, Cat MC85200). For HG002 cell line, fresh or frozen cell pellets were osmotically
432 lysed in presence of 150mL of Nuclei Prep Buffer (NEB, Cat.T3052) supplemented with 5.5 mL
433 of Rnase A (NEB, Cat. T3018L) and 5.5 mL of RNase If (NEB, Cat. M0243L) per million cells for
434 15 seconds and mixed by flicking. For PBMC or fresh blood samples, an optional PBS wash
435 followed by Red Blood Cell lysis step was included (10 mins at RT) prior to hypotonic lysis with
436 Nuclei Prep Buffer (NEB, Cat.T3052) and Rnase digestion. Nuclei from 1 million cells were
437 lysed with 300 mL of lysis buffer supplemented with 20 mL of Proteinase K (20mg/mL)
438 (ThermoFisher, Cat. 25530049). Lysates were incubated at 50 degrees C for a minimum of 24
439 hours overnight with periodic vortexing at low speeds (minimum speed to achieve swirling of the
440 solution). 150 mL of MPC Protein Precipitation Reagent solution from the Lucigen/EpiCentre's
441 MasterPureTM Complete DNA and RNA Purification Kit A was added to precipitate proteins
442 followed by centrifugation at 2000 x g for 30 mins. DNA was precipitated by adding 500 mL of
443 cold isopropanol (100%) (Supply Store, Cat.100209) and pelleted by centrifugation (2000 x g for
444 20 mins). DNA pellets were washed 3X with 70% ethanol and hydrated in pre-warmed (37°C)
445 Elution buffer (Qiagen, 10 mM Tris-Cl, pH 8.5. Cat. 19086) and incubated on HulaMixer™
446 Sample Mixer (Thermo Fisher Scientific, Cat. 15920D) at 37°C incubator overnight at 1rpm end
447 over end mixing.

448 **Annealing of TeloTags for duplex barcode assembly:**

449 TeloTags were prepared in 100µlL reactions with 5mM of each of the 6 permutations of
450 telomere splint Extended Data Fig.1A) and 30 mM of biotinylated adapter in HiFi Taq DNA
451 Ligase Reaction Buffer (NEB, Cat. M0647S). Annealing was done by heating to 99 degrees and
452 slowly decreasing the temperature 1°C /min in a Veriti™ 96-Well Thermal Cycler (Applied
453 Biosystems, Cat. 4375786). After annealing, reactions were diluted 1:100 in 1x Taq buffer and

454 kept at 4°C. The sequences of the TeloTag and splint adapter is listed in Extended Data Table
455 2)
456
457 **Telomere Tagging**
458 High molecular weight genomic DNA (gDNA) was quantified using the Qubit dsDNA BR assay
459 kit (Thermo Fisher Scientific, Cat.Q32850). A total of 40 µg of gDNA was incubated with 3µl of
460 ClaI (NEB, Cat. R0197S) or AsiSI (NEB, Cat.R0630L) or PmeI (NEB, Cat.R0560L), or BamHI
461 (NEB, Cat.R0136M) for 2 hours at 37°C, with gentle flicking every 20 mins. Subsequently, the
462 enzyme was heat-inactivated at 65°C for 20 mins. Ligations were carried out using 4 ug of DNA
463 per 50 µl reaction.
464 Tagging reactions were done in 50 µL volume for each reaction in a MicroAmp™ TriFlex Well
465 PCR Reaction Plate (Applied Biosystems, Cat. A32811), with 4µl/reaction of 0.3µM duplex
466 TeloTag adapter, 5µl/reaction of 10X HiFi Taq DNA Ligase Reaction Buffer (NEB, Cat.
467 M0647S), and 1µl/reaction of HiFi Taq DNA Ligase (NEB, Cat. M0647S). The TeloTagging
468 reactions were incubated for 5 mins at 65°C in a Veriti™ 96-Well Thermal Cycler (Applied
469 Biosystems™, Cat.4375786). Ligations were done through 15 cycles of denaturing at 65°C for 1
470 min, followed by annealing and ligating at 45°C for 3 mins with a 15% ramp down of rate
471 between steps.
472
473 **Telomere Enrichment and Nanopore Sequencing**
474 For chromosome-specific telomere length measurements, we typically used 30-40 µg of DNA
475 per sample. A standard 3 mL tube of blood or 30 million PBMC produced ~200 ug of DNA. For
476 bulk telomere length measurements, as little as 5-10 µg of starting gDNA was employed. All
477 pipetting was performed using wide bore pipette tips to minimize DNA shearing, except for
478 addition of SPRI beads where accurate volume ratios are extremely important for successful
479 cleanups. All the Telomere Tagging reactions were pooled in DNA LoBind (Eppendorf,

480 Cat.0030108523) tube. Cleanup and removal of excess TeloTag adapters was done using SPRI
481 beads (Beckman Coulter, Cat. B23318) a ratio SPRI beads to DNA of 45 µL:100 µL was used.
482 The samples were incubated with SPRI beads rotating end over end on a Hula mixer for 20
483 mins at 10rpm. SPRI beads were then separated using a DynaMag™-2 Magnet (Thermo Fisher
484 Scientific, Cat. 12321D) and washed while on the magnet twice with freshly made 85% ethanol.
485 DNA was eluted using heated (65 °C) 1X rCutsmart Buffer. The volume of elution volume was
486 calculated to achieve 150 ng/ml final concentration based on input DNA amount. The eluting
487 SPRI beads were incubated for 20 mins at 65°C with gently flicking every 5 mins. SPRI beads
488 were removed using a DynaMag™-2 Magnet.
489 The gDNA recovery was quantified using the Qubit dsDNA BR assay kit. Tagged gDNA was
490 enriched using Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific, Cat. 65001).
491 The beads were allowed to room temperature while being resuspended on a HulaMixer™
492 Sample Mixer (Thermo Fisher Scientific) at 3 rpm for 1h. A ratio streptavidin to DNA of 1
493 µg:250ng was used. The beads were washed once in Binding Buffer from Dynabeads™
494 kilobaseBINDER™ Kit and resuspended in equal volume binding buffer as eluted DNA volume.
495 The beads were then added to the gDNA sample and incubated at room temp at 1 rpm on a
496 HulaMixer™ Sample Mixer for 20 mins. Reactions can be scaled up or down as needed, though
497 the maximum volume of beads + gDNA + binding buffer should not exceed 1.4 ml for a single
498 1.5 ml Protein LoBind tube (Eppendorf, Cat. 30108442). Multiple tubes can be used and pooled
499 at the restriction enzyme digest step. After binding to streptavidin, the beads were washed
500 using the following sequence to remove background genomic DNA: 2x kilobaseBinder wash
501 buffer, 2x Elution buffer (Qiagen, 10 mM Tris-Cl, pH 8.5), 1x rCutsmart Buffer.
502 To release telomeres, the streptavidin bead-telomere complex was resuspended in 72 µl of 1X
503 rCutsmart, 3 µl of PvuI (NEB, Cat. R3150S) PacI (NEB, Cat. R0547L) or EcoRI (NEB,
504 Cat.R0101M), and incubated at 37°C for 30 min, with periodic gentle flicking. The sample was
505 then heated at 65°C for 20 mins to release any bound telomeres. If multiple tubes were used,

506 sequential rounds of digestion can be used by adding restriction enzyme to the eluted telomere
507 solution from the first step and incubating with streptavidin-telomere beads in the second tube.
508 Recovered tagged gDNA was quantified using the Qubit dsDNA HS assay kit. The expected
509 recovery was approximately 0.1-0.01% of the starting gDNA sample.
510 Enriched telomeres were carried forward into the standard Nanopore library prep protocol from
511 ONT. All reactions were prepared using Ligation Sequencing Kit V10 (SQK-LSK114) kits and
512 sequenced on R9.4.1 (Oxford Nanopore Technologies, FLO-MIN106D) flow cells. Libraries
513 were eluted in 40 ml of elution buffer (Qiagen, 10 mM Tris-Cl, pH 8.5) with optional 15 mins of
514 incubation at 37°C to recover long molecules. Each library was split into 3 reactions. Each
515 reaction was sequenced on a flow cell for ~18 hours before flow cell flushing/washing using flow
516 cell wash kit (Oxford Nanopore Technologies EXP-WSH004) and loading of the remaining
517 fraction. Reads were collected using MinKNOW software (5.7.5) without live basecalling.
518
519 **Determination of telomere/subtelomere boundary position and telomere length**

520 To determine the length of the telomere repeats, we tested two methods. One method is based
521 on determining the junction of the telomeres and subtelomeres in the respective reference
522 genome (CHM13, HG002 maternal and HG002 paternal) and the second method determines
523 the subtelomere to telomere junction in every read. For method 1, to determine the junction in a
524 reference genome, we developed a Python algorithm named TeloBP (Telomere Boundary
525 Point). TeloBP employs a rolling window approach, scanning from the telomere into the
526 chromosome, identifying the telomere-subtelomere junction by detecting a discontinuity in a
527 user defined telomeric pattern. The algorithm's default telomeric pattern is a sequence where at
528 least 50% of nucleotides are "GGG". As the window moves along at six nucleotide intervals, it
529 scores telomere similarity in 100-nucleotide segments. Variants of the telomere repeats known
530 to be in the subtelomere do not significantly change sequence content. The junction is defined

531 when the sequence content no longer matches a telomere like sequence content. This is
532 calculated by averaging the similarity of a sequence with a 500 bp window, marking the start of
533 a 50% deviation, then scanning until the increase in discontinuity plateaus, marking the
534 subtelomere boundary. After reads are mapped to the reference genome, for each read the
535 telomere length is determined as the number of base pairs from subtelomere junction in the
536 reference to the TeloTag. This method incorporated many variant telomere repeats into the
537 telomere that are not incorporated by identifying the boundary as 4X TTAGGG (Extended Data
538 Fig 3).

539 In the second method we determined the subtelomeres/telomeres boundary in each read. We
540 developed a version of TeloBP that considers common errors in the nanopore Guppy base
541 calling. These patterns are set by default based on findings in [26],
542 "[^GGG]GGG|[^AAA]AAA|TTAGG." for G strand and "CTTCTT|CCTGG|CCC…" C strands. But
543 the patterns can be user defined as new base callers are developed. We named this algorithm
544 TeloNP (Telomere NanoPore). Both TeloBP and TeloNP are available on Github
545 (https://github.com/GreiderLab).

546 **Custom genome for mapping telomeres**

547 For mapping reads to the T2T genomes CHM13 and HG002 we generated custom reference
548 genomes. We first extracted the terminal 500kb of chromosome end for each genome, then
549 removed the telomere repeats (as determined by TeloBP) from the reference genome to allow
550 for maximized weighing of subtelomere information for read mapping.

551

552 **Bioinformatic filtering of telomere reads**

553 Reads were first filtered for any of the following telomere patterns ["TTAGGGTTAGGG",
554 "TTAAAATTAAAATTAAAA", "CCCTCCGATA", "TGGCCTGGCCTGGCC"] based on findings in
555 previous literature [26]. To identify reads with a TeloTag at the end and to demultiplex samples we

556 performed a pairwise alignment of the 24bp barcodes with the terminal 300bp of each read
557 using the pairwise Alignment function in the Bio Strings package of Bioconductor
558 (doi:10.18129/B9.bioc.Biostrings , R package version 2.68.1,
559 https://bioconductor.org/packages/Biostrings). The alignment score cutoff was set so the false
560 discovery rate for our nanopore reads was < 1% based on random 24bp barcode sequences
561 and unused ONT barcode sequences. We used Minimap2 with the -x map-ont option to map
562 our reads to the custom genomes HG002 and CHM13 [40]. We only considered primary
563 alignments that started within 1 kb of the subtelomere boundary.
564
565 **Peak calling to measure telomere length with** TeloPeakCounter
566 To examine whether the Guppy and Dorado base caller accuracy call the telomere length in
567 correctly, we developed an algorithm, TeloPeakCounter, to count repeated peaks, or waves, in
568 the electrical signal data measured by the nanopore device. These distinct repeated waves
569 found in the telomere region of reads correspond to the TTAGGG telomere repeat sequences.
570 TeloPeakCounter analyzes and counts these distinctive, periodic wave patterns in the electric
571 signal data, and enables a direct measurement of telomere length. Assuming each wave
572 represents a 6-nucleotide telomere repeat, we can compute estimated telomere lengths for a
573 read. The code for TeloNP and TeloPeakCounter is available at GitHub
574 (https://github.com/GreiderLab)
575
576 **Mapping HG002 subtelomere to maternal and paternal alleles**
577 For the diploid HG002 genome some maternal and paternal subtelomere sequences are very
578 similar and correct assignment of reads becomes difficult. We developed a two-step mapping
579 procedure for mapping HG002 reads. In the first step, reads are mapped to the HG002 diploid
580 genome. Mapq mapping confidence scores are set low for these mappings, as the mapper can
581 have difficulty deciding between very similar maternal and paternal subtelomere sequences. We

582 applied a relatively low mapq filter cutoff of 10 to the diploid mapping. In a second step we
583 mapped the reads also to the maternal and paternal haploid genomes separately. Mapq scores
584 generally increase for the alignments to the haploid genomes. To identify high confidence
585 alignments, we applied a mapq cutoff of 30 to the haploid genome alignments. A read needed to
586 map to the same chr end in both mappings and pass the two mapq cutoffs to be considered
587 correctly assigned. There were different numbers of reads for specific chromosome ends due to
588 the restriction enzyme sometimes cutting very near a telomere. To minimize this, we used two
589 different restriction enzymes for both the initial cutting and for the release. This allowed mapping
590 of more reads for some chromosome ends.
591
592 **Pangenome based mapping for chromosome assignment of telomeres from diverse**
593 **individuals**
594 We mapped reads to the pangenome to efficiently capture telomere length across the diverse
595 population. A references file of 500kb of subtelomere sequences was assembled from each of
596 the genomes [41] in the pangenome. We mapped our reads from 150 individuals to this
597 reference. We filtered for reads that had a minimum of 2kb alignment to the pangenome
598 reference. To compare telomere length across individuals we removed acrocentric
599 chromosomes (13p, 14p, 15p, 21p, 22p) and X and Y subtelomeres (XpYp, and Xq Yq) which
600 recombine in the population.
601
602 **Statistical Analysis**
603 To determine whether HG002 chromosome specific telomere lengths were significantly different
604 from the individual's grand mean telomere length, we used Analysis of the Mean (ANOM).
605 Statistics were calculated using the R package rstatix (v0.7.0) and ANOM (v0.2)
606 (https://cran.rproject.org/web/packages/rstatix/index.html).
607

## Data Availability

Data generated during the study will be made available in public sequence repository Sequence Read Archive (SRA) https://www.ncbi.nlm.nih.gov/sra and is available upon request.

## Code availability

The python code for TeloBP, TeloNP and TeloPeakCounter is available at https://github.com/GreiderLab


## Acknowledgements

We thank Dr. Rachel Green for provision of laboratory space at Johns Hopkins as well as discussions. Drs. Brendan Cormack and Christine Gao and for reviewing the manuscript. Carla Connelly, Julie Brunelle, and Margaret Strong provided experimental and logistical assistance. We thank Dr. Mary Armanios and her lab for help in the early stages of assay development and Dr. Ludmilla Danilova for statistical analysis. This work was supported by NIH grant R35CA209974 to CWG and Johns Hopkins Bloomberg Distinguished Professorship to CWG.


## Author contributions

These authors contributed to the following aspects of this work. Conceptualization CWG, KK, SS; Data curation, AG, KK, RK, RWK, AR; Formal analysis, KK, RK, AR; Funding acquisition CWG, HL; Investigation AG, VH, KK, RWK, Methodology,CWG, VH, KK, RK, RWK, AR, SS; Project administration CWG; Resources JA, CWG, JFM, KK; Software KK, RK, HL, AR ,K-TT; Supervision CWG, HL; Validation CWG, AG, KK, RK, AR; Visualization JA, KK, RK, AR; Writing original draft CWG, KK Writing – review & editing JA, CWG, AG, VH, KK, AR, SS

## Competing interest declaration

CWG and KK are inventors of US Patent PCT/US2023/073375 titled "methods for telomere length measurement".

641 **References**


642 1 Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal
643 transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405-413 (1985).

644 2 Armanios, M. Syndromes of telomere shortening. *Annu Rev Genomics Hum Genet* **10**,
645 45-61 (2009).

646 3 Armanios, M. The Role of Telomeres in Human Disease. *Annu Rev Genomics Hum*
647 *Genet* **23**, 363-381 (2022).

648 4 Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science*
649 **339**, 959-961 (2013).

650 5 Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma.
651 *Science* **339**, 957-959 (2013).

652 6 Marcand, S., Gilson, E. & Shore, D. A protein-counting mechanism for telomere length
653 regulation in yeast. *Science* **275**, 986-990 (1997).

654 7 Smogorzewska, A. *et al.* Control of human telomere length by TRF1 and TRF2. *Mol Cell*
655 *Biol* **20**, 1659-1668 (2000).

656 8 Teixeira, M. T., Arneric, M., Sperisen, P. & Lingner, J. Telomere length homeostasis is
657 achieved via a switch between telomerase- extendible and -nonextendible states. *Cell*
658 **117**, 323-335 (2004).

659 9 Blackburn, E. H. & Chiou, S. S. Non-nucleosomal packaging of a tandemly repeated
660 DNA sequence at termini of extrachromosomal DNA coding for rRNA in Tetrahymena.
661 *Proc Natl Acad Sci U S A* **78**, 2263-2267 (1981).

662 10 Aviv, A. *et al.* Impartial comparative analysis of measurement of leukocyte telomere
663 length/DNA content by Southern blots and qPCR. *Nucleic acids research* **39**, e134
664 (2011).

665 11 Marcand, S., Wotton, D., Gilson, E. & Shore, D. Rap1p and telomere length regulation in
666 yeast. *Ciba Found Symp* **211**, 76-93; discussion 93-103 (1997).

667 12 Wotton, D. & Shore, D. A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to
668 regulate telomere length in Saccharomyces cerevisiae. *Genes Dev* **11**, 748-760 (1997).

669 13 Levy, D. L. & Blackburn, E. H. Counting of Rif1p and Rif2p on Saccharomyces
670 cerevisiae telomeres regulates telomere length. *Molecular and cellular biology* **24**,
671 10857-10867 (2004).

672 14 Aubert, G., Hills, M. & Lansdorp, P. M. Telomere length measurement-caveats and a
673 critical assessment of the available technologies and tools. *Mutation research* **730**, 59-
674 67 (2012).

675 15 Alder, J. K. *et al.* Diagnostic utility of telomere length testing in a hospital-based setting.
676 *Proc Natl Acad Sci U S A* **115**, E2358-E2365 (2018).

677 16 Armanios, M. Telomeres and age-related disease: how telomere biology informs clinical
678 paradigms. *J Clin Invest* **123**, 996-1002 (2013).

679 17 Rufer, N., Dragowska, W., Thornbury, G., Roosnek, E. & Lansdorp, P. M. Telomere
680 length dynamics in human lymphocyte subpopulations measured by flow cytometry [see
681 comments]. *Nat Biotechnol* **16**, 743-747 (1998).

682 18 Lansdorp, P. M. *et al.* Heterogeneity in telomere length of human chromosomes. *Hum*
683 *Mol Genet* **5**, 685-691 (1996).

684 19 Zijlmans, J. M. *et al.* Telomeres in the mouse have large inter-chromosomal variations in
685 the number of T2AG3 repeats. *Proc Natl Acad Sci U S A* **94**, 7423-7428 (1997).

686 20 Martens, U. M. *et al.* Short telomeres on human chromosome 17p. *Nat Genet* **18**, 76-80
687 (1998).

688 21 Londono-Vallejo, J. A., DerSarkissian, H., Cazes, L. & Thomas, G. Differences in
689 telomere length between homologous chromosomes in humans. *Nucleic Acids Res* **29**,
690 3164-3171 (2001).

691 22 Sholes, S. L. *et al.* Chromosome-specific telomere lengths and the minimal functional
692 telomere revealed by nanopore sequencing. *Genome Res* **32**, 616-628 (2022).

693 23 Grigorev, K. *et al.* Haplotype diversity and sequence heterogeneity of human telomeres.
694 *Genome Res* (2021).

695 24 Tham, C. Y. *et al.* High-throughput telomere length measurement at nucleotide
696 resolution using the PacBio high fidelity sequencing platform. *Nat Commun* **14**, 281
697 (2023).

698 25 Dubocanin, D. & Stergachis, A. B. Single-molecule architecture and heterogeneity of
699 human telomeric DNA and chromatin. *BioRxiv* (2022).

700 26 Tan, K. T., Slevin, M. K., Meyerson, M. & Li, H. Identifying and correcting repeat-calling
701 errors in nanopore sequencing of telomeres. *Genome Biol* **23**, 180 (2022).

702 27 Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of
703 human fibroblasts. *Nature* **345**, 458-460 (1990).

704 28 Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing.
705 *Nature* **346**, 866-868 (1990).

706 29 Vaziri, H. *et al.* Loss of telomeric DNA during aging of normal and trisomy 21 human
707 lymphocytes. *Am. J. Hum. Genet* **52**, 661-667 (1993).

30 Aubert, G., Baerlocher, G. M., Vulto, I., Poon, S. S. & Lansdorp, P. M. Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. *PLoS Genet* **8**, e1002696 (2012).

31 Rufer, N. *et al.* Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *J Exp Med* **190**, 157-167 (1999).

32 Martens, U. M. *et al.* Telomere maintenance in human B lymphocytes. *Br J Haematol* **119**, 810-818 (2002).

33 Van Ziffle, J. A., Baerlocher, G. M. & Lansdorp, P. M. Telomere length in subpopulations of human hematopoietic cells. *Stem Cells* **21**, 654-660 (2003).

34 Baerlocher, G. M. & Lansdorp, P. M. Telomere length measurements in leukocyte subsets by automated multicolor flow-FISH. *Cytometry A* **55**, 1-6 (2003).

35 Alder, J. K. & Armanios, M. Telomere-mediated lung disease. *Physiol Rev* **102**, 1703-1720 (2022).

36 Lakota, K. *et al.* Short lymphocyte, but not granulocyte, telomere length in a subset of patients with systemic sclerosis. *Ann Rheum Dis* **78**, 1142-1144 (2019).

37 Garg, S. *et al.* Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol* **39**, 309-312 (2021).

38 Stong, N. *et al.* Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res* **24**, 1039-1050 (2014).

39 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

40 Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572-4574 (2021).

41 Liao, W. W. *et al.* A draft human pangenome reference. *Nature* **617**, 312-324 (2023).

42 Antonarakis, S. E. Short arms of human acrocentric chromosomes and the completion of the human genome sequence. *Genome Res* **32**, 599-607 (2022).

43 Graakjaer, J. *et al.* The relative lengths of individual telomeres are defined in the zygote and strictly maintained during life. *Aging Cell* **3**, 97-102 (2004).

44 Hemann, M. T., Strong, M. A., Hao, L.-Y. & Greider, C. W. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. *Cell* **107**, 67-77 (2001).

741 45 d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated
742 senescence. *Nature* **426**, 194-198 (2003).

743 46 Enomoto, S., Glowczewski, L. & Berman, J. MEC3, MEC1, and DDC2 are essential
744 components of a telomere checkpoint pathway required for cell cycle arrest during
745 senescence in Saccharomyces cerevisiae. *Mol Biol Cell* **13**, 2626-2638 (2002).

746 47 IJpma, A. & Greider, C. W. Short telomeres induce a DNA damage response in
747 Saccharomyces cerevisiae. *Mol Biol Cell* **14**, 987-1001 (2003).

748 48 McNally, E. J., Luncsford, P. J. & Armanios, M. Long telomeres and cancer risk: the
749 price of cellular immortality. *J Clin Invest* **130**, 3474-3481 (2019).

750 49 Smogorzewska, A. & de Lange, T. Regulation of telomerase by telomeric proteins. *Annu*
751 *Rev Biochem* **73**, 177-208 (2004).

752 50 Shore, D. & Bianchi, A. Telomere length regulation: coupling DNA end processing to
753 feedback regulation of telomerase. *Embo J* **28**, 2309-2322 (2009).

754 51 Lustig, A. J. Clues to catastrophic telomere loss in mammals from yeast telomere rapid
755 deletion. *Nat Rev Genet* **4**, 916-923 (2003).

756 52 Zimmermann, M., Kibe, T., Kabir, S. & de Lange, T. TRF1 negotiates TTAGGG repeat-
757 associated replication problems by recruiting the BLM helicase and the TPP1/POT1
758 repressor of ATR signaling. *Genes Dev* **28**, 2477-2491 (2014).

759 53 Paschini, M. *et al.* Spontaneous replication fork collapse regulates telomere length
760 homeostasis in wild type cells. *bioRxiv* (2020).

761 54 Arneric, M. & Lingner, J. Tel1 kinase and subtelomere-bound Tbf1 mediate preferential
762 elongation of short telomeres by telomerase in yeast. *EMBO reports* **8**, 1080-1085
763 (2007).

764 55 Brown, W. R. A. *et al.* Structure and polymorphism of human telomere-associated DNA.
765 *Cell* **63**, 119-132 (1990).

766 56 Nergadze, S. G. *et al.* CpG-island promoters drive transcription of human telomeres.
767 *RNA* **15**, 2186-2194 (2009).

768 57 Deng, Z. *et al.* A role for CTCF and cohesin in subtelomere chromatin organization,
769 TERRA transcription, and telomere end protection. *Embo J* **31**, 4165-4178 (2012).

770 58 Feretzaki, M., Renck Nunes, P. & Lingner, J. Expression and differential regulation of
771 human TERRA at several chromosome ends. *RNA* **25**, 1470-1480 (2019).

772 59 Achrem, M., Szucko, I. & Kalinka, A. The epigenetic regulation of centromeres and
773 telomeres in plants and animals. *Comp Cytogenet* **14**, 265-311 (2020).

774 60 Toubiana, S. & Selig, S. Human subtelomeric DNA methylation: regulation and roles in
775 telomere function. *Curr Opin Genet Dev* **60**, 9-16 (2020).

776 61 Gonzalo, S. *et al.* DNA methyltransferases control telomere length and telomere
777 recombination in mammalian cells. *Nat Cell Biol* **8**, 416-424 (2006).

778 62 Arnoult, N. *et al.* Replication timing of human telomeres is chromosome arm-specific,
779 influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genet*
780 **6**, e1000920 (2010).

781 63 Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat Rev Genet* **13**, 693-704
782 (2012).

783 64 Hao, L. Y. *et al.* Short telomeres, even in the presence of telomerase, limit tissue
784 renewal capacity. *Cell* **123**, 1121-1131 (2005).

785 65 Slagboom, P. E., Droog, S. & Boomsma, D. I. Genetic determination of telomere size in
786 humans: a twin study of three age groups. *Am J Hum Genet* **55**, 876-882 (1994).

787 66 Conklin, Q. A., Crosswell, A. D., Saron, C. D. & Epel, E. S. Meditation, stress processes,
788 and telomere biology. *Curr Opin Psychol* **28**, 92-101 (2019).

789 67 Greider, C. W. Telomeres, telomerase and senescence. *Bioessays* **12**, 363-369 (1990).

790 68 de Lange, T. *et al.* Structure and variability of human chromosome ends. *Mol. Cell. Biol.*
791 **10**, 518-527 (1990).

792
793

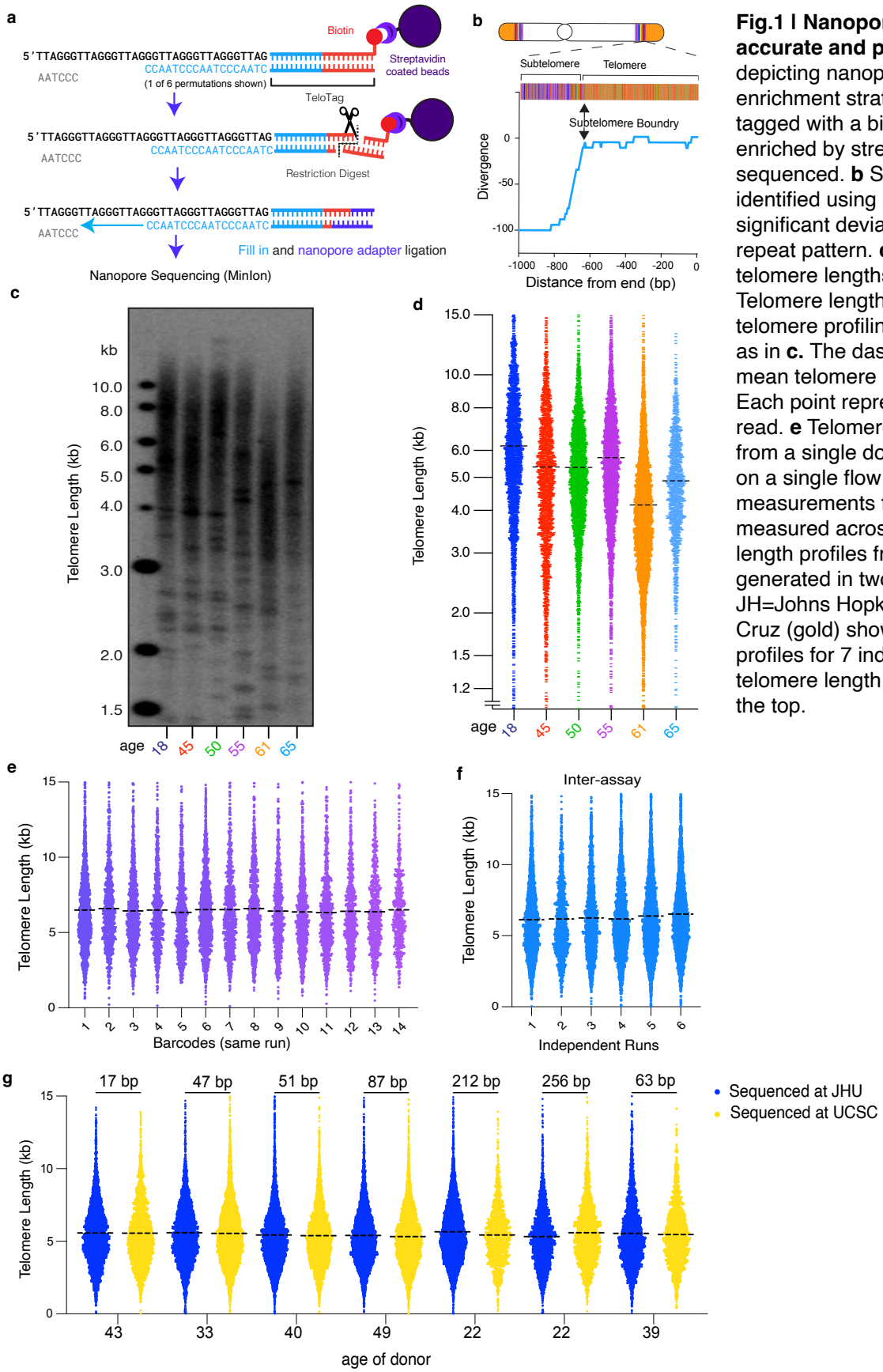



**Fig.1 | Nanopore telomere profiling is accurate and precise. a** Schematic depicting nanopore telomere profiling enrichment strategy. Telomeres are tagged with a biotin adapter (TeloTag), enriched by streptavidin pull down, and sequenced. **b** Subtelomere boundary is identified using an algorithm that detects significant deviation from the telomere repeat pattern. **c.** Southern blot of telomere lengths from 6 individuals. **d**. Telomere length measured by Nanopore telomere profiling for the same individuals as in **c.** The dashed line represents the mean telomere length of each distribution. Each point represents a single telomere read. **e** Telomere length measurements from a single donor measured 14 times on a single flow cell. **f** Telomere length measurements from a single donor measured across 6 flow cells **g.** Telomere length profiles from the same samples generated in two different laboratories JH=Johns Hopkins (blue), SC=UC Santa Cruz (gold) showing reproducible length profiles for 7 individuals. The difference in telomere length in base pairs is shown at the top.

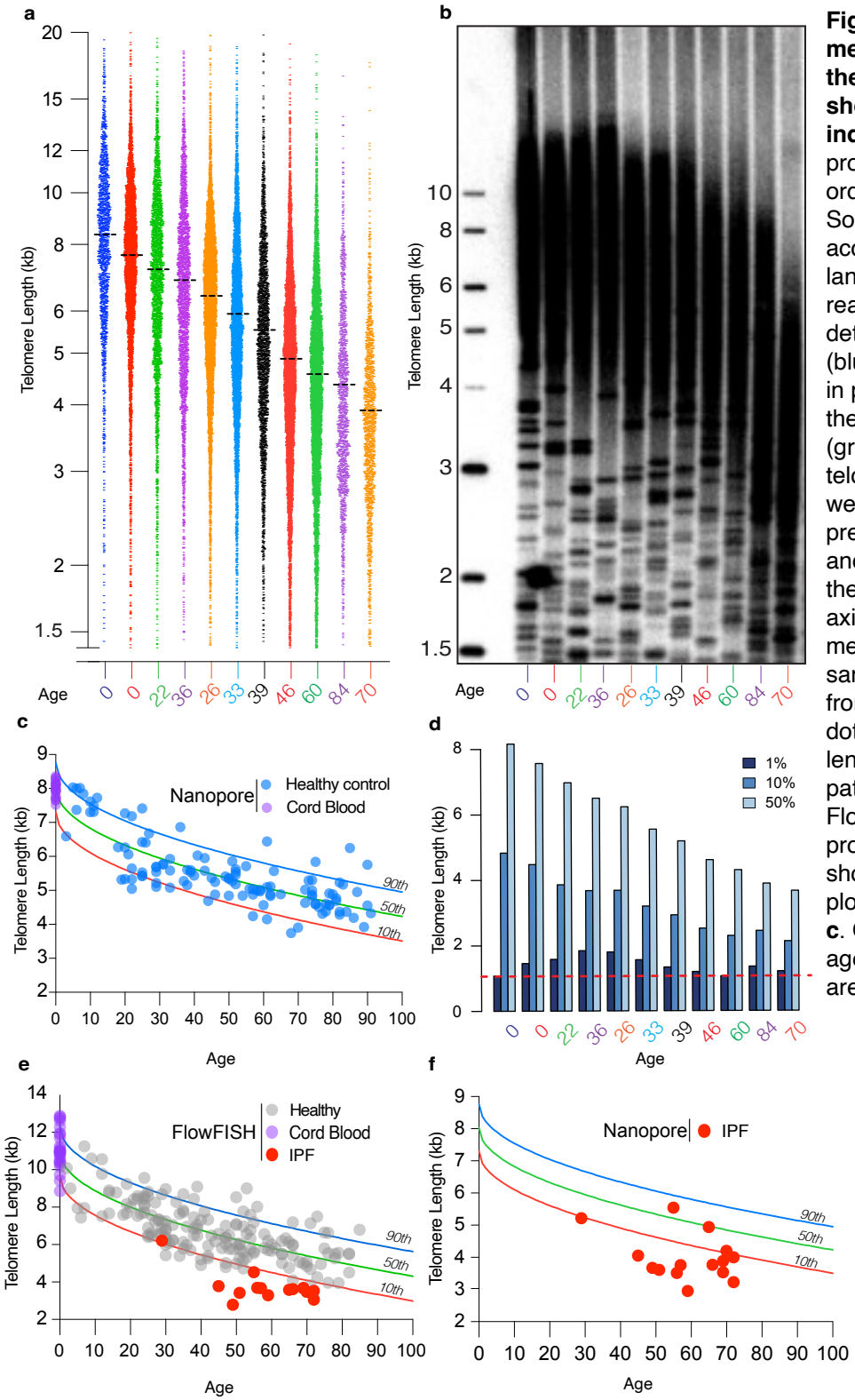



**Fig. 2 | Nanopore telomere profiling measures telomere length dynamics in the aging population and separates short telomere patients from healthy individuals. a.** Nanopore telomere length profiles of 11 samples ranked in decreasing order. Age of individual noted at bottom. **b.** Southern blot of 11 samples loaded according to the order calculated in **a**, each lane is a different person. Each dot is a read. **c.** The mean telomere length was determined for 132 individuals aged 0 to 90 (blue dots). Cord blood lengths are shown in purple. The population distribution shows the 90th percentile (blue), the median (green line) and 10th percentiles (red) for telomere length in this population. Intervals were derived using parameters established previously for FlowFISH [15]. **d**. The 50th 10th and 1st percentile of telomere lengths for the data in **a**. are plotted with age on the X axis as in **b**. The dotted line represents the mean length of the 1st percentile for all samples **e.** Lymphocyte telomere length from FlowFISH data from Alder et al. (gray dots) and cord blood (purple dots). The lengths for 15 short telomere syndrome patients with IPF (red) were determined by FlowFISH. **f**. Nanopore telomere length profiles from the same 15 patients with short telomere syndrome shown in **e.** plotted against population distribution from **c**. One point represents two individuals aged 50 have nearly identical length and are indistinguish-able in the figure.

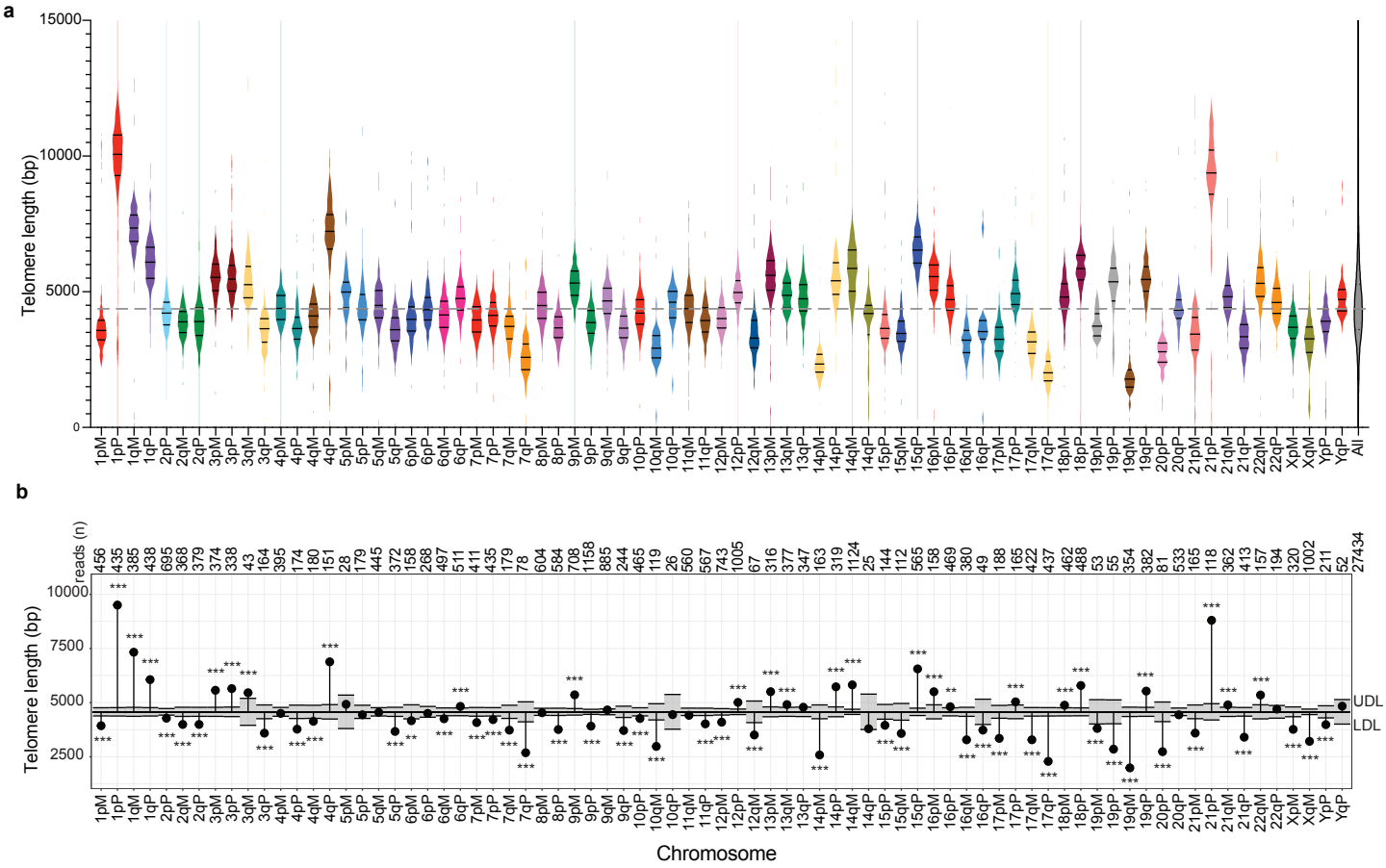



**Fig. 3 l Chromosome end-specific telomere lengths a.** Violin plots of the distribution of telomere lengths for 77 telomeres from HG002 that mapped with confidence and passed our filters (see methods). Each end is labeled with the chromosome number and p for the short and q for long arms. The haplotypes for each chromosome end are labeled Maternal (M) and Paternal (P). The mean, 90th and 10th percentile for each distribution are shown with short horizontal black lines in each violin plot. The grand mean of all telomeres is at the far right (all). The dashed line represents the grand mean of all telomeres **b**. Analysis of the means (ANOM) multiple contrast test of each telomere length distribution against the grand mean of all telomere lengths for data in **a.** The number of reads for each chromosome end is shown at the top. P-values were adjusted for multiple hypotheses testing using the Bonferroni method. Chromosome ends with length profiles reaching outside of the shaded gray region between the upper decision limit (UDL) and lower decision limit (LDL) are considered significantly different from the grand mean. (*) $p \leq 0.05$, (**), $p \leq 0.01$. (***), $p \leq 0.001$; nonsignificant differences have no stars.

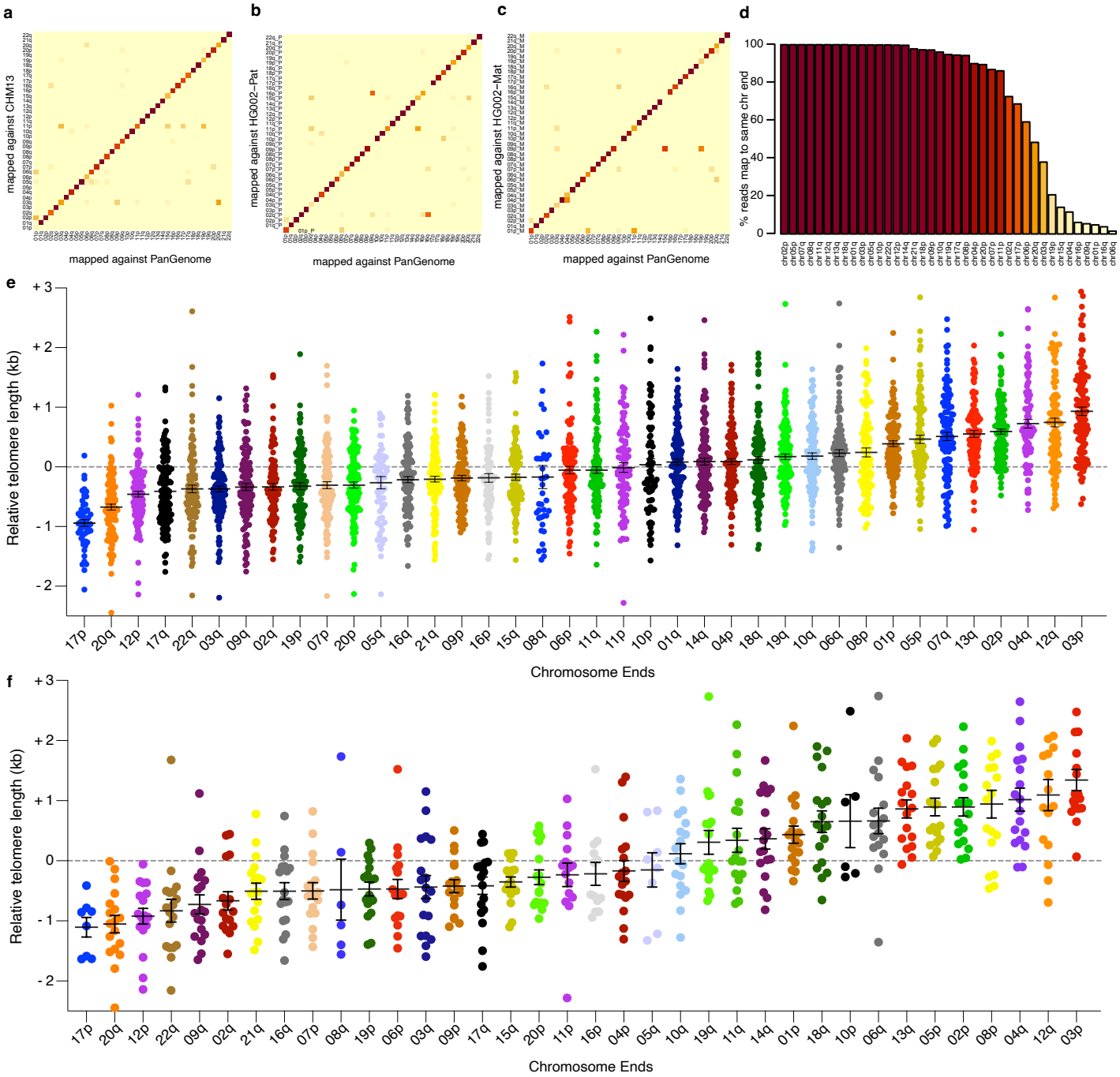
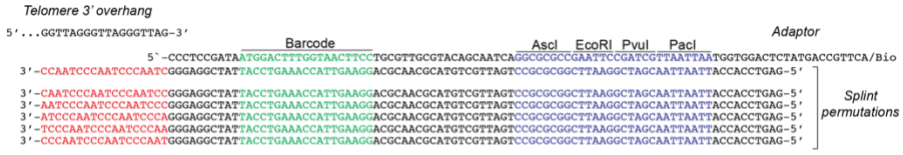
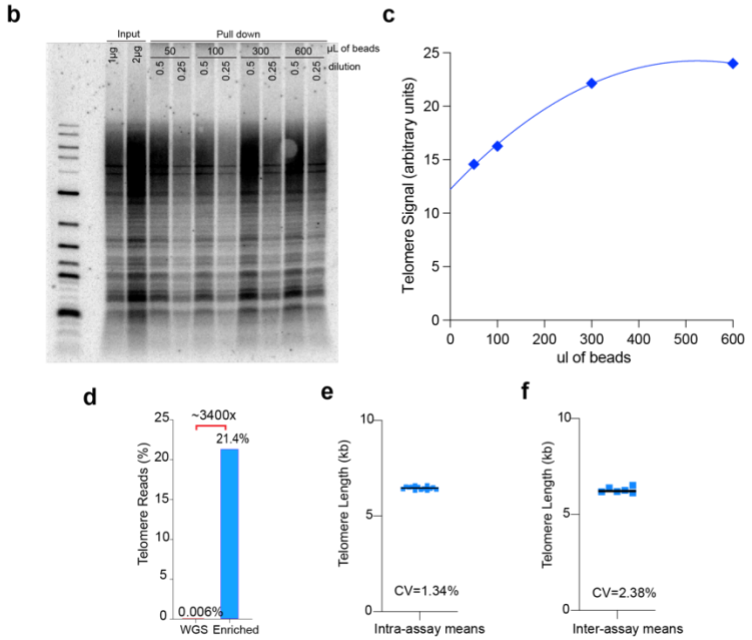



**Fig. 4 | Chromosome-specific telomere lengths are conserved and established at birth.** We used the pangenome reference to assign chromosome end status for (~720,000) telomere reads obtained from 150 individuals. Telomere reads were mapped to the pangenome with a requirement of 2kb of alignment. **a.** Matrix heat-map showing pairwise comparison of chromosome assignments of telomere reads when mapped to the pangenome or to the CHM13 reference genome. Heatmap shows what fraction of reads mapping to a given chromosome end in the pangenome (column) map to each chromosome end in CHM13 (rows). Light yellow indicates 0% and dark red indicates 100% of reads mapping to the respective CHM13 chr. end. **b**. As in **a**. but reads mapped to the HG002 paternal reference **c.** As in **a.** but reads mapped to the HG002 maternal reference genome. **d.** Bar graph showing the fraction of reads that mapped for each Chr end in the pangenome to the same Chr end in all three haploid genomes (CHM13, HG002 maternal and HG002 paternal). Colors are the same as in the heatmaps **a**. **b**. and **c**. **e**. For any given individual, we calculated Relative telomere length. For each chromosome end in each individual, we calculated the mean telomere length and subtracted it from the individual's grand mean telomere length. Zero indicates no difference between the chromosome end's mean telomere length and the individual's grand mean telomere length. Bars in the violin plots represent the mean length of all of the specific chromosome end in the population and the whiskers represent the standard error of the mean. **f**. Same as in **e**. but for cord blood samples only**.**

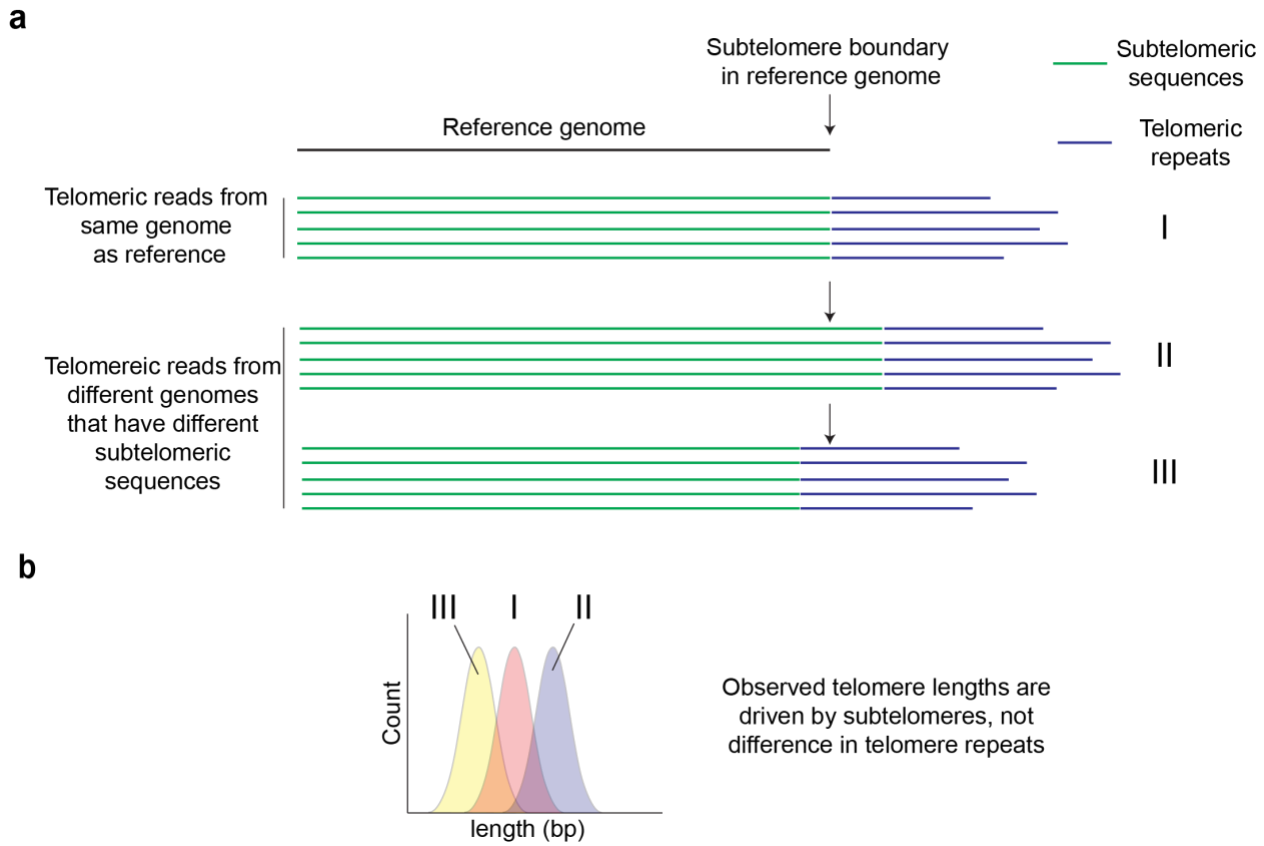
## Extended Data Figures

**a**

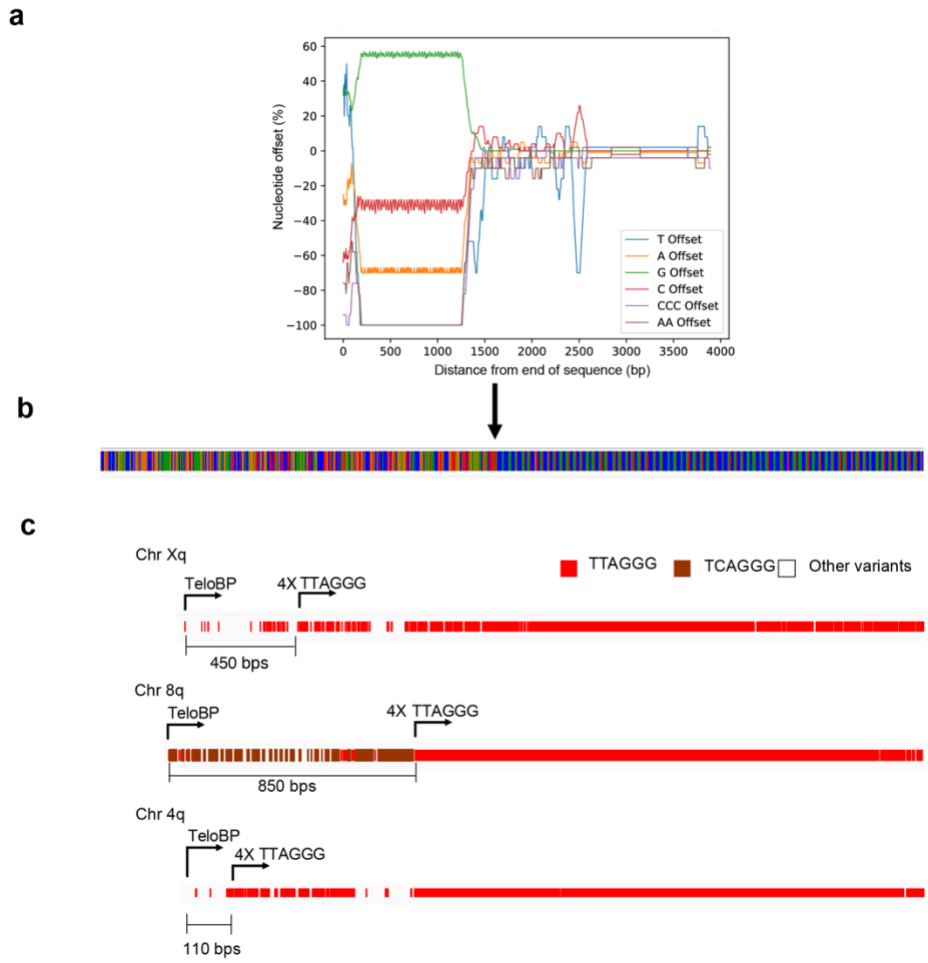


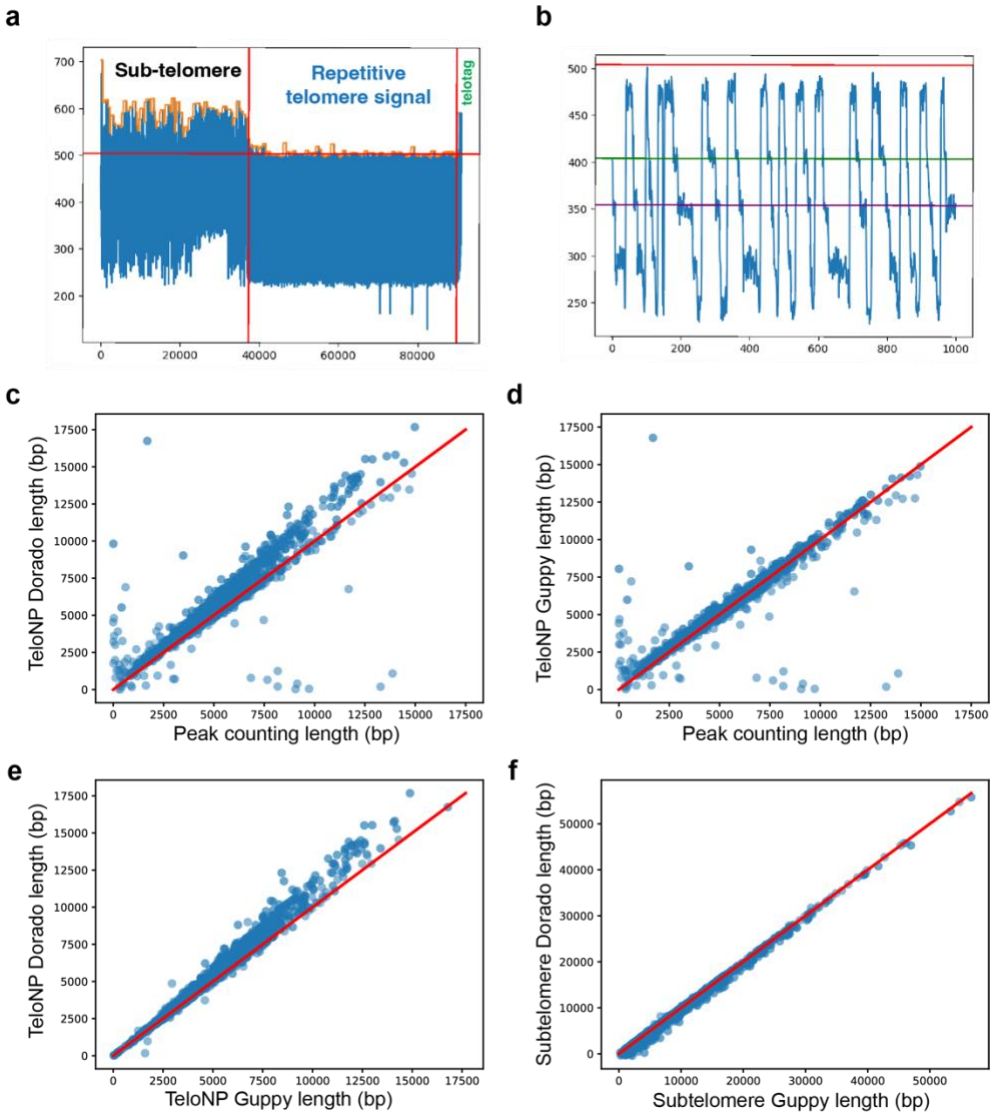

**Extended Data Fig. 1. Quantitation of enrichment and assay reproducibility. a.** Sequence of one representative TeloTag adapter. The barcoded adaptor (top strand) is annealed to a mixture of splints that have all 6 permutations of the CCCTAA sequence to improve chances of in-frame annealing to the telomere 3' overhang. **b**. Southern blot of telomeres recovered after biotin pull down using different volumes of streptavidin bead enrichment. **c.** Quantification of the efficiency of enrichment using increasing ratio of streptavidin beads to DNA. **d.** Enrichment of telomeric reads using biotin pull down relative to WGS. **e.** Intra-assay coefficient of variation (CV) of one single with different barcodes measured multiple times on the same flow cell. **f**. Inter-assay coefficient of variation (CV) of one sample measured multiple times across different flow cells. Mean telomere length of a single sample measured on multiple different runs.



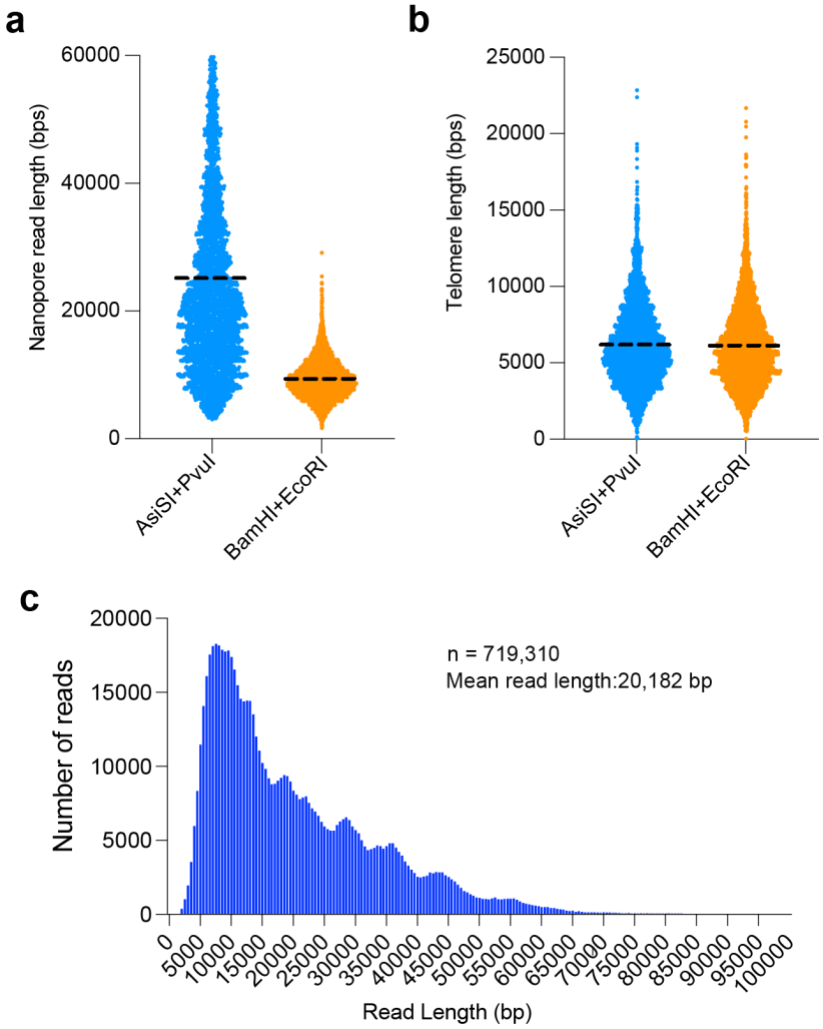

**Extended Data Fig. 2: Heterogeneity in human subtelomere sequence means the telomere subtelomere boundary point can differ in sequence reads from diverse genomes and the reference genome a**. Telomere reads from the DNA identical to the reference genome will align at the boundary point in the reference. However, for some individuals a telomeric read will map well but there is extra sequence past the reference boundary point. For others there may be less subtelomere sequence on the read b. When telomere length is determined by mapping to the reference sequence boundary point, this can lead to incorrectly longer (II) or incorrectly shorter (III) telomere length distributions.



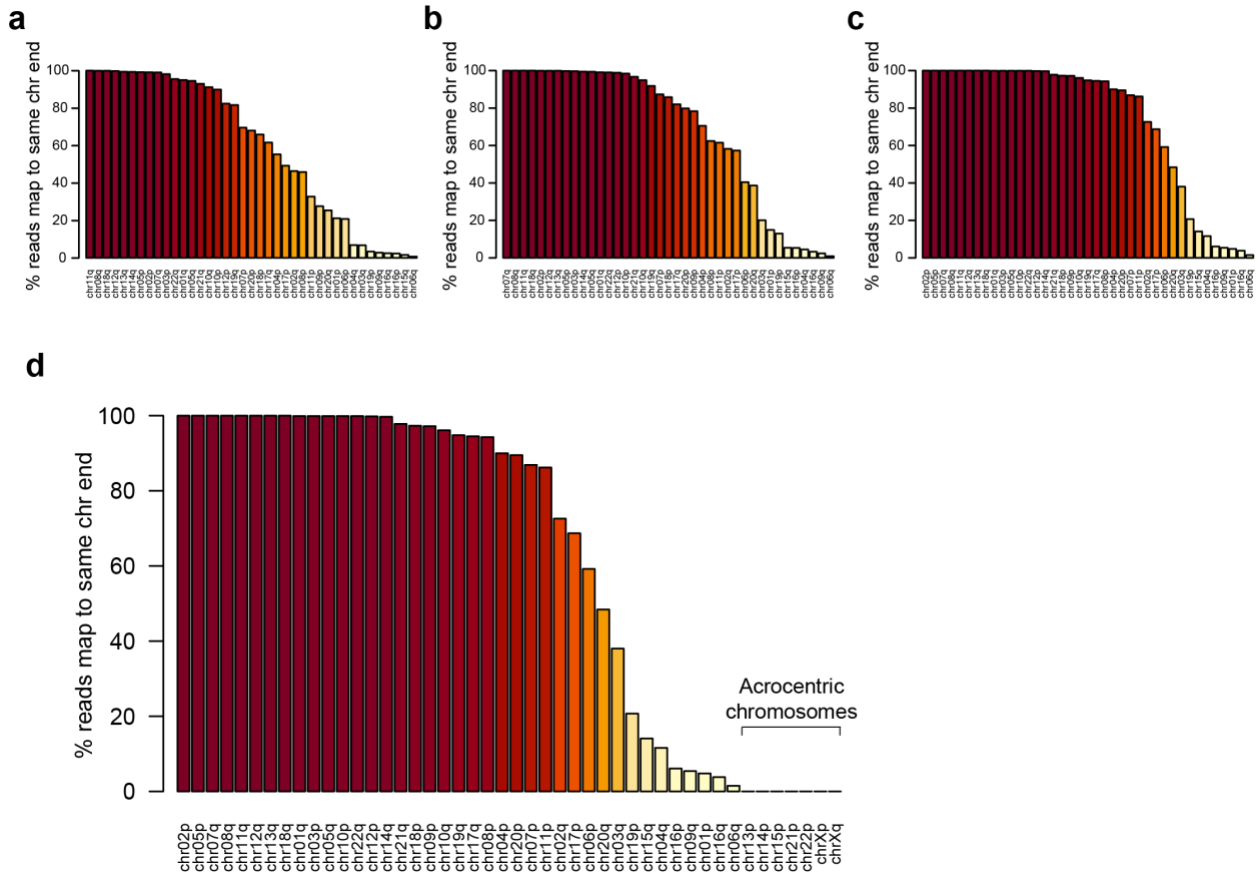

**Extended Data Fig. 3: Establishing telomere boundary points with TeloBP algorithm a**. Representation of the nucleotide offsets for several different parameters as a rolling window scanning from telomere end on right (see methods). **b.** IGV view of the telomere sequence and where the boundary is called **c.** Example of where TeloBP incorporates variant repeats into the telomere, compared to method setting a boundary of 4 consecutive repeats of TTAGGG.




**Extended Data Fig 4: Analysis of telomere length by TeloPeakCounter**. **a**. Representation of the subtelomere and telomere sequence electrical signal **b**. High resolution image of peaks in the telomere repeats in electrical signal. **Comparison of Guppy versus Dorado base caller**. Each blue dot represents an individual telomere read. 2435 read were examined from one data set (F63) from Fig 2. **c**. Comparison of telomere length determined by the peak counting vs Dorado base calling. **d**. Comparison of telomere length determined by the peak counting vs Guppy base calling. **e**. Comparison of Guppy telomere length by TeloNP vs Dorado. **f.** Comparison of subtelomere length with Guppy vs Dorado.




**Extended Data Fig 5**. **Length of fragments does not affect telomere length determination.** **a.** The length of the fragments when genomic DNA is cut with AsiSI and PvuI is shown in blue. The length of the fragments when cut with BamHI and EcoRI is shown in orange. **b.** The telomere length of fragments cut with AsiSI and PvuI is in blue and BamHI and EcoRI is shown in orange. **c.** The distribution of fragment lengths for all reads: the Y axis is the number of reads and the X axis is the length in base pairs.




**Extended Data Fig. 6. Concordance of reads mapped to the pangenome with mapping to CHM13 and HG002 Mat and HG002 Pat.** We used different mapq scores to quantitate the fraction of reads that mapped to the same chromosome ends as the pangenome and the three referenced genomes **a.** Mapq score of 1 **b.** Mapq score of 30 **c.** Mapq score of 60. **d**. In previous analysis the acrocentric were omitted. Here they were included 13p,14p,15p, 21p and 22p and show less than 0 reads mapped to the same chromosome ends for these acrocentric.

**Extended Data Table 1. Samples included in comparison of FlowFISH and Telomere Profiling**

| *Sample* | *age* | *FlowFISH lymphocyte length (bps)* | *FlowFISH granulocyte length (bps)* | *Telomere profiling length (bps)* |
|---|---|---|---|---|
| *1* | 29 | 6210 | 5480 | 5045 |
| *2* | 45 | 3790 | 3860 | 4054 |
| *3* | 49 | 2790 | 3640 | 3678 |
| *4* | 51 | 3420 | 4020 | 3609 |
| *5* | 55 | 4520 | 4860 | 5545 |
| *6* | 56 | 3690 | 3990 | 3519 |
| *7* | 57 | 3670 | 2830 | 3759 |
| *8* | 59 | 3300 | 3600 | 2954 |
| *9* | 65 | 3580 | 4250 | 4945 |
| *10* | 66 | 3630 | 4430 | 3774 |
| *11* | 69 | 3670 | 3990 | 3541 |
| *12* | 69 | 3670 | 4300 | 3883 |
| *13* | 70 | 3480 | 4500 | 4203 |
| *14* | 72 | 3060 | 3090 | 3236 |
| *15* | 72 | 3530 | 4000 | 4003 |