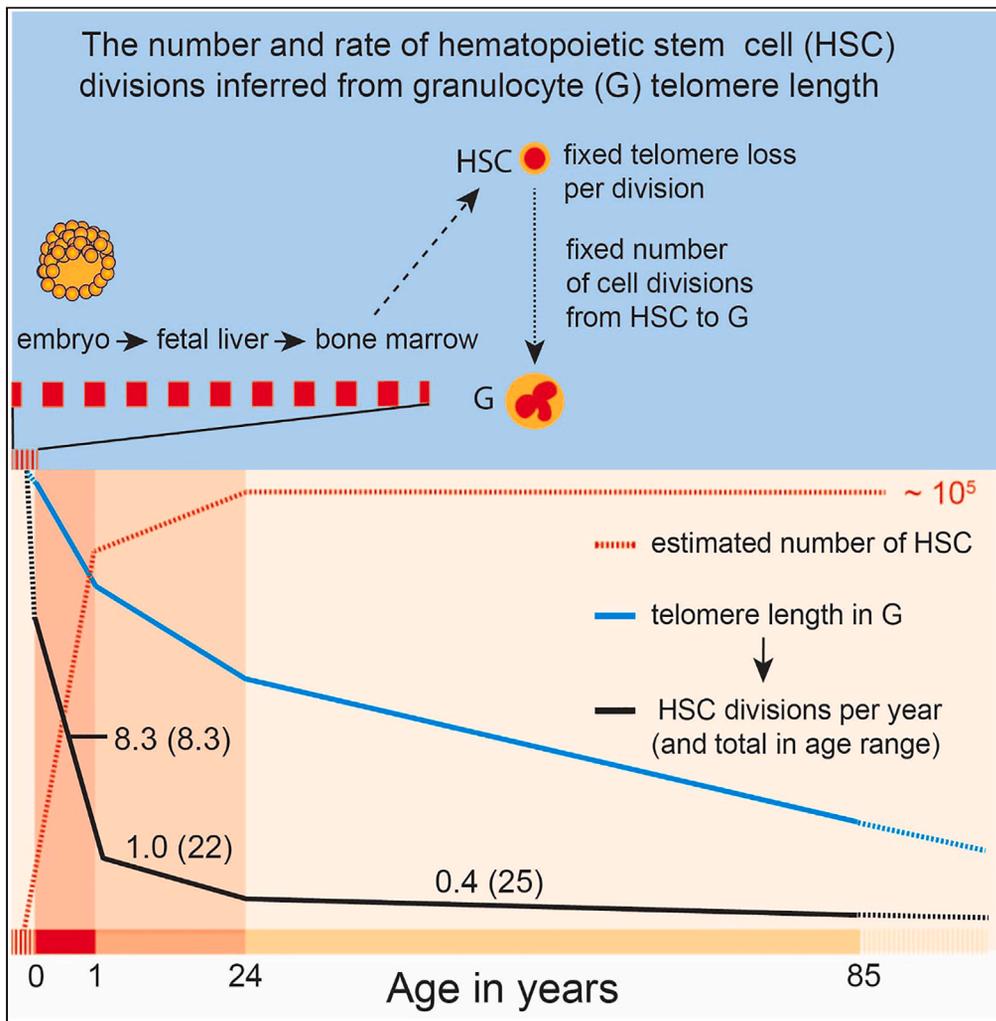


Article

# Predicting the number of lifetime divisions for hematopoietic stem cells from telomere length measurements



Cole Boyle, Peter M. Lansdorp, Leah Edelstein-Keshet

cab@student.ubc.ca

Highlights

Hematopoietic stem cells (HSCs) go through 3 distinct division rate phases

HSCs divide on average 56 times over a normal human lifespan

Roughly half of an HSC's divisions occur in the first 24 years of life

Adult stem cells divide less than once every two years

Boyle et al., iScience 26, 107053  
July 21, 2023 © 2023 The Authors.  
<https://doi.org/10.1016/j.isci.2023.107053>



## Article

## Predicting the number of lifetime divisions for hematopoietic stem cells from telomere length measurements

Cole Boyle,<sup>1,4,\*</sup> Peter M. Lansdorp,<sup>2,3</sup> and Leah Edelstein-Keshet<sup>1</sup>

## SUMMARY

**How many times does a typical hematopoietic stem cell (HSC) divide to maintain a daily production of over  $10^{11}$  blood cells over a human lifetime? It has been predicted that relatively few, slowly dividing HSCs occupy the top of the hematopoietic hierarchy. However, tracking HSCs directly is extremely challenging due to their rarity. Here, we utilize previously published data documenting the loss of telomeric DNA repeats in granulocytes, to draw inferences about HSC division rates, the timing of major changes in those rates, as well as lifetime division totals. Our method uses segmented regression to identify the best candidate representations of the telomere length data. Our method predicts that, on average, an HSC divides 56 times over an 85-year lifespan (with lower and upper bounds of 36 and 120, respectively), with half of these divisions during the first 24 years of life.**

## INTRODUCTION

The number of times stem cells in self-renewing tissues divide over a lifetime is not known. Estimates of stem cell numbers and cell division rates require accurate identification of stem cells, a major challenge in all studies of stem cell biology.<sup>1</sup> The problem of stem cell identification is exemplified in the hematopoietic system, the most well-studied self-renewing tissue in both humans and mice.

In the early part of the 20th century, hematopoietic stem cells (HSCs) in human bone marrow were identified in smears and tissue sections using histological dyes in combination with wishful thinking. The observation that lethally irradiated animals could be rescued by transplantation of bone marrow enabled a new, transplantation-based approach to identify HSCs in the 1950's. This approach was pioneered by Till and McCullough, who showed that single murine bone marrow cells upon transplantation into irradiated recipients could give rise to visible colonies in the spleen.<sup>2</sup> Such colonies contained a variety of cell types as well as cells that could again give rise to colonies upon secondary transplantation.<sup>3</sup> These observations founded the general concept that stem cells are endowed with variable "self-renewal" properties. Transplantation assays in combination with increasingly sophisticated labeling and clonal marking strategies have been at the heart of HSC research ever since.<sup>4,5</sup>

New insight into the clonal composition of the hematopoietic system has come from DNA sequencing studies.<sup>6–8</sup> In this approach, genome-wide somatic mutations in nucleated blood cells are used to reconstruct phylogenetic trees and estimate the number of HSCs. In human adults, this number was estimated to be between 20,000 and 200,000 HSCs.<sup>6</sup>

The low frequency of HSCs contrasts sharply with the estimated daily marrow output of over  $10^{11}$  nucleated<sup>9</sup> and over  $10^{11}$  red blood cells,<sup>10</sup> suggesting a very deep hierarchy of increasingly more differentiated progenitor cells. The number of layers in this hierarchy and the frequency of cell divisions in HSCs at the top remain to be determined, a major experimental challenge in view of the low frequency and lack of unique markers of HSCs. In principle, acquired mutations in HSCs that provide a proliferative advantage can disrupt polyclonal hematopoiesis. To prevent rapid clonal dominance of such clones, cell divisions at the level of HSCs are probably limited as was proposed for stem cells in the skin<sup>11</sup> and bone marrow.<sup>12</sup> Whether all stem cells, including those in the digestive system and testis, are organized in an economical hierarchy is currently not clear.<sup>13,14</sup>

<sup>1</sup>Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2 Canada

<sup>2</sup>Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC V5Z 1L3, Canada

<sup>3</sup>Department of Medical Genetics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>4</sup>Lead contact

\*Correspondence: cab@student.ubc.ca

<https://doi.org/10.1016/j.isci.2023.107053>



Telomeres are noncoding repetitive DNA sequences at the ends of eukaryotic chromosomes (reviewed by Lansdorp<sup>12</sup> and references therein). In mammals, telomeres consist of a variable number of TTAGGG repeats and associated proteins. Telomeric DNA repeats are lost with each cell division as DNA polymerase is unable to completely replicate the lagging strand of DNA during replication (for a review see 12 and 15). Without compensation for the inevitable loss of telomere repeats, erosion limits the number of times somatic cells can divide.<sup>15</sup> This was clearly demonstrated *in vitro* for human fibroblasts<sup>16</sup> and T lymphocytes<sup>17</sup> where it was shown that overexpression of telomerase prevents replicative senescence and thereby extends the replicative lifespan. Similar data for HSCs are not available, but telomerase activity in HSCs also appears to be unable to prevent the loss of telomere repeats with each cell division.<sup>12</sup>

Through successive cell divisions during hematopoietic differentiation, telomere length in HSCs and their nucleated blood cell progeny are linked. This link can be used to describe HSC turnover in terms of the attrition of telomeres in nucleated blood cells, such as granulocytes, that are separated from HSC by a relatively constant number of cell divisions.<sup>12</sup> Previous attempts to describe this link have used assumptions about the telomere loss per cell division in HSCs<sup>18–20</sup> and/or mathematical modeling of the underlying HSC population dynamics, about which little is known.<sup>21,22</sup> Generally, these models require strong assumptions about the asymmetric and symmetric divisions of HSCs. In our approach, by contrast, we recover the number of HSC divisions directly from telomere shortening data in granulocytes, with minimal assumptions about the details of how HSCs divide. We revisit published telomere length data with a systematic method of estimating the HSC division rates, the number of distinct phases and timing of changes in those rates, and, ultimately, the mean total number of HSC divisions over a human lifetime. In carrying out our procedure, we assumed 30 to 100 base pairs (bp) of telomeric DNA are lost in HSCs with each cell division, as originally proposed by Vaziri et al.<sup>23</sup> We took a loss of 65 bp per division as a likely possibility for HSCs since a similar loss was measured in T cells that could be clonally propagated.<sup>17</sup>

## RESULTS

### Telomere length data

Telomere shortening at individual chromosomes is affected by many factors, including oxidative damage to telomeric DNA, the levels and efficiency of telomerase, and unknown stochastic interactions between a myriad of known and unknown factors. To mitigate such complexity at the level of individual chromosomes, we consider only the average length of telomere repeats in a cell. As eukaryotic human cells have a total of 92 telomeres that occupy the ends of 46 chromosomes, this average provides a reasonable estimate of the average length of telomere repeats at any given chromosome end in a cell.

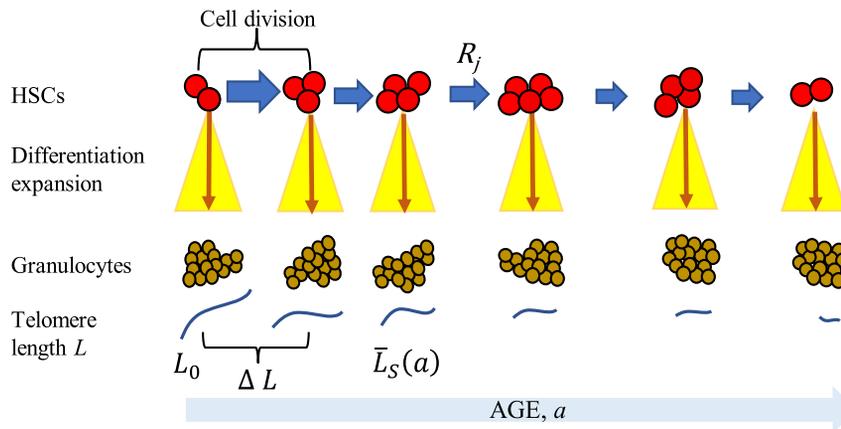
Ideally, when discussing telomere attrition in a nucleated blood cell population of a person, we need longitudinal measurements of telomere length over time. However, measurements tracking single persons over a lifetime are not feasible or available, so we must interpret existing cross-sectional data as though they reflect “typical” longitudinal data. This practice is reasonable since measurements are reproducible<sup>24</sup> and sufficient data are available to mitigate the telomere length diversity at any given age.

Measurements of the telomere length in the leukocyte populations of 835 healthy individuals were reported by Aubert et al.<sup>25</sup> using fluorescence *in situ* hybridization and flow cytometry (Flow-FISH). By including reference control samples in every measurement and avoiding DNA amplification, the Flow-FISH technique is accurate and e.g. suitable for the diagnosis of patients with telomere biology disorders.<sup>24</sup> Due to the presence of outliers in telomere fluorescence histograms of thousands of measured cells, the sample median for measurements of the average telomere length in several different types of leukocytes was used as a proxy for the average telomere length in that population of cells.

### Mathematical models and setup

Consider HSCs in the bone marrow of an individual at age  $a$ . We define  $N(a)$ ,  $N_i(a)$  as the total number of HSCs and the number of HSCs that have gone through  $i$  cell divisions, respectively.  $N_i$  represents the size of the  $i$ ’th HSC division class. Let  $\bar{D}(a)$  be the number of divisions of a “typical” HSC in a person of age  $a$ . Our goal is to estimate this average.

We make several assumptions to connect leukocyte telomere length data to the average number of HSC divisions at a given age.



**Figure 1. Schematic diagram of our method**

We use data for telomere length in human granulocytes over age (bottom, terminally differentiated cells, brown) to infer the rate of division  $R_j$  and total number of divisions in hematopoietic stem cells (HSCs, top, red). It is assumed that the differentiation and expansion of the HSC progeny (yellow funnels) are the same at every age so that telomere loss  $\Delta L$  observed in granulocytes reflects HSC cell division. We find that the HSC cell division rate (blue arrows) is roughly constant over each of two or three phases (e.g., infancy, youth, and adult for a 3-phase model).

1. Telomerase expression in HSCs is suppressed at birth (so telomere length reflects the number of cell divisions). The precise timing of this suppression is unknown.<sup>26</sup>
2. HSCs start out with approximately the same average telomere length,  $L_0$  (once telomerase is suppressed).
3. Each HSC cell division results in roughly the same telomeric DNA loss,  $\Delta L$ . (We take  $30 \leq \Delta L \leq 100$  bp<sup>23</sup>). In contrast to Rodriguez-Brenes and Peskin<sup>27</sup> we do not consider telomerase-mediated lengthening of telomeres.
4. At any given time, the number of divisions ( $d$ ) between an HSC and a leukocyte progeny is fixed for a given leukocyte type.
5. Fluctuations in  $d$  are transient (e.g., associated with infections, disease, or injury) and short relative to a normal human lifespan.

The above assumptions imply that cell division classes are equivalent to telomere length classes. Note that we make no further assumptions about specific types or rates of cell division. Rather, we use available telomere length data to infer rates of HSC division. See Figure 1 for a schematic diagram of our method.

Based on the assumptions, the telomere length at age  $a$  is simply the initial length ( $L_0$ ) discounted by the amount lost per division ( $\Delta L$ ) multiplied by the number of divisions  $i$ . Hence the mean telomere length  $\bar{L}_S(a)$  at age  $a$  is

$$\bar{L}_S(a) = \frac{1}{N(a)} \sum_i (L_0 - i \Delta L) N_i(a).$$

(where we have simply averaged the telomere length associated with each HSC division class  $i$ , over the entire HSC population at age  $a$ .) This equation is essentially the same as Equation S5 in Werner et al.<sup>21</sup> but in contrast with their approach (and those of Shepherd et al.<sup>19</sup> and Edelstein-Keshet et al.<sup>22</sup>); we do not attempt to compute  $\bar{L}_S(a)$  from a hypothetical model that tracks the distribution of  $N_i(a)$  over time. Rather, we expand the sum, obtaining

$$\bar{L}_S(a) = L_0 - \frac{\Delta L}{N(a)} \sum_i i N_i(a) = L_0 - \Delta L \bar{D}(a), \text{ where } \bar{D}(a) = \frac{1}{N(a)} \sum_i i N_i(a).$$

We use this last version to relate the mean number of HSC divisions,  $\bar{D}(a)$ , directly to the mean telomere length  $\bar{L}_S(a)$  from the data.

### Granulocyte telomere length data mirrors HSC division schedule

We specifically choose to use telomere length data from granulocytes, for which assumptions 4 and 5 can be justified, allowing us to link the population average telomere length of these differentiated blood cells to that of their HSC ancestors. Granulocytes do not undergo cell divisions and hence satisfy assumption 4. This contrasts with T cells, which undergo division while performing their function (introducing variation in  $d$  between distinct T cells). As for assumption 5, temporary changes in  $d$  will occur during infections or injury when more granulocytes are needed, but  $d$  will eventually return to a baseline during normal homeostasis. This allows us to assume that  $d$  is constant when averaging over the entire lifespan. Under these assumptions, it follows that the average telomere length in the granulocyte population ( $\bar{L}_G$ ) differs by some constant,  $c$ , from that of HSCs,

$$\bar{L}_G(a) = \bar{L}_S(a) - c = L_{G,0} - \Delta L \bar{D}(a),$$

where  $L_{G,0}$  is the initial average length of telomeres in granulocytes. Therefore, with our setup, the rate of telomere attrition in granulocytes directly corresponds to the division rate of HSCs. Note that, while Aubert et al.<sup>25</sup> also statistically fit granulocyte telomere loss data, they did not infer an implied HSC division rate or total number of divisions.

### Telomere data suggest phases with distinct HSC division rates through life

The decline in leukocyte telomere length is very profound in the first year of life both in baboons<sup>28</sup> and humans,<sup>25</sup> in comparison with more modest and nearly constant gradual loss in adulthood.<sup>21,25,28,29</sup> Assuming telomere attrition reflects stem cell divisions, as we have argued above, this implies that HSC division rates change from birth to later ages.

The data appear to consist of distinct “phases”, each lasting years, over which the shortening rate is roughly constant. Restated, the data appear to be describable by a set of lines with distinct slopes, connected to one another at certain ages (technically speaking, a “piecewise-linear” function). Indeed, Aubert et al.<sup>25</sup> used a piecewise-linear fit with breakpoints at years 1 and 18, for their data; i.e., they assumed a 3-phase model. Here, we generalize that idea to consider  $n$  – phase models of the average telomere length in granulocytes, where, in each phase, the mean division rate of an HSC is constant.

Therefore, we assume that the average telomere length,  $\bar{L}_G(a)$ , is of the form

$$\bar{L}_G(a) = \begin{cases} L_{G,0} - s_1 \cdot a & (0 \leq a \leq a_1) \\ L_1 - s_2 \cdot (a - a_1) & (a_1 < a \leq a_2) \\ \vdots & \vdots \\ L_{n-1} - s_n \cdot (a - a_{n-1}) & (a_{n-1} < a) \end{cases}$$

where  $a_i$ 's are the ages at which a phase transition occurs,  $L_i$ 's are the average telomere length at these transition points, and  $s_i$ 's are the slopes of the lines, representing average telomere length loss rate along each of the phases. Recall that  $s_j = \Delta L \cdot R_j$  where  $R_i = \frac{d}{da} \bar{D}(a)$  is the division rate of HSCs during the  $i$ <sup>th</sup> phase. As we do not expect telomere lengths to jump at the transition points, we also impose the continuity condition that the lines meet at the transition points,  $L_i = L_{i-1} - s_j \cdot (a_i - a_{i-1})$ . Once the number of phases,  $n$ , is given, we can fit the  $n$ -phase model to estimate the division rate of HSCs during each phase and determine the phase-transition points.

This method of modeling telomere loss captures the significant age intervals over which HSCs have noticeably higher division rates. If models with more phases fit the data better, then we would conclude that HSCs undergo a greater number of changes in their division rate over a lifespan.

This piecewise-linear model has sharp transitions between phases. We also considered polynomials and shifted logarithmic fits to the data. We found that, while not significantly improving the predictive quality over the piecewise-linear fits, those alternate models result in very similar estimates for lifetime HSC divisions (see the [STAR Methods](#) for details).

### Model estimation and selection

For a given  $n$ -phase model we use the standard R library Segmented<sup>30,31</sup> to determine the model's optimal parameters. In the case of our model, which is continuous at the transition points, Segmented has been shown to perform better than other regression model estimators.<sup>32</sup>

**Table 1. Results from fitting 1-, 2-, 3-, and 4-phase models to the Aubert et al. data**

	Estimated value (SE)			
<b>1-phase model</b>				
Intercept (kbp)	9.60 (0.07)			
Slope (kbp/yr)	−0.0357 (0.0013)			
R <sup>2</sup>	0.4785			
AIC	2489.98			
BIC	2504.06			
<b>2-phase model</b>				
Breakpoint	20.9 (3.2)			
Intercept (kbp)	10.12 (0.11)			
Slopes (kbp/yr)	−0.081 (0.010)	−0.0265 (0.0021)		
R <sup>2</sup>	0.5055			
AIC	2451.03			
BIC	2474.51			
<b>3-phase model</b>				
Breakpoints	0.95 (2.3)	24.2 (5.2)		
Intercept (kbp)	10.34 (0.15)			
Slope (kbp/yr)	−0.6 (1.2)	−0.063 (0.012)	−0.0265 (0.0023)	
R <sup>2</sup>	0.5084			
AIC	2450.26			
BIC	2483.13			
<b>4-phase model</b>				
Breakpoints	0.6 (1.9)	25 (12)	28 (20)	
Intercept (kbp)	10.33 (0.15)			
Slope (kbp/yr)	−0.8 (2.2)	−0.067 (0.011)	0.02 (0.64)	−0.0272 (0.0024)
R <sup>2</sup>	0.5083			
AIC	2454.45			
BIC	2496.71			

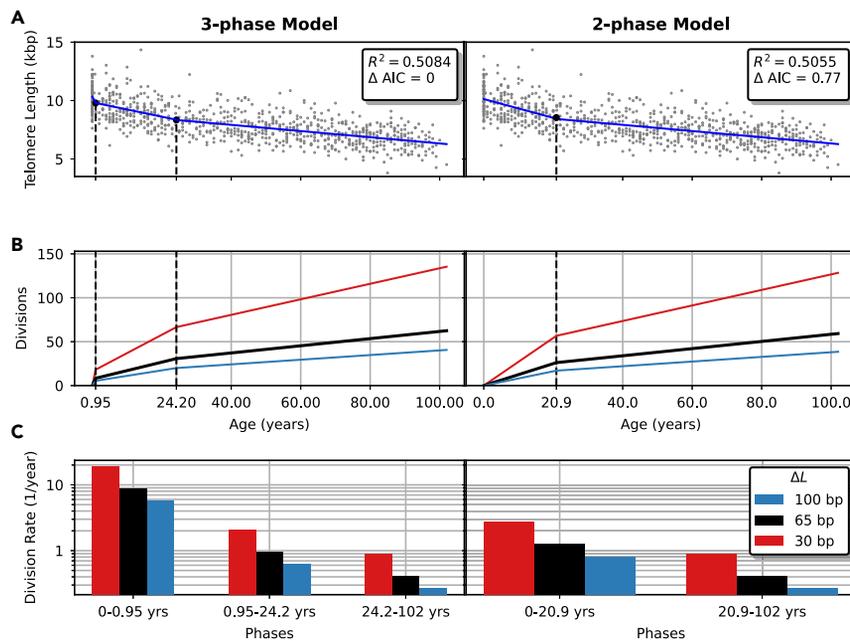
Fittings were done using the R library Segmented. Estimated values are reported with their standard error (SE). The slope values for each model correspond to the rate of telomere attrition during each phase in kbp/year.

Once the best estimates for the  $n$ -phase models are computed, we employ the Akaike information criterion (AIC) as a means of rejecting the sub-optimal models. The AIC is a standard tool that appraises whether increasing the number of parameters (in our case, phases) is justified by the data without the risk of overfitting. AIC rewards goodness of fit and penalizes the number of parameters to be fit.<sup>33</sup> The model with the lowest AIC is “optimal”, while, as a rule of thumb, those within 2 AIC points of that optimal model share a similar level of support.<sup>34</sup>

### Best-fit parsimonious model predicts 2 or 3 phases

We considered models with  $n = 1, 2, 3,$  and  $4$  phases and found that the 3-phase model had the lowest AIC, followed closely (within 1 point) by the 2-phase model. As the AIC score for the 1-phase and 4-phase models were considerably larger (by 44 and 4 points, respectively) we can safely reject those models. While the AIC (and previous work by Aubert et al.<sup>25</sup>) suggests that the 3-phase model is optimal, it turned out that the Bayesian information criterion (BIC), another model selection score that penalizes the number of parameters more heavily, favors the 2-phase model. Results of these criteria are summarized in Table 1.

As the AIC and BIC numerical criteria are less than decisive for the given data, we avoid claiming optimality of one or the other model. Rather, we calculate and compare results for both 2- and 3-phase



**Figure 2. Best-fit parsimonious model predicts 2 or 3 phases**

(A) Optimal 3-phase and 2-phase models of average telomere length in granulocytes, fit to telomere measurement data from Aubert et al.<sup>25</sup>  $\Delta AIC$  denotes the difference in AIC score with respect to the minimal AIC. The 2-phase model still has strong support in comparison.

(B) Predictions for the number of divisions that the average HSC has accrued by a given age. Colors in B and C correspond to the distinct possible choices for the amount of telomeric DNA lost in a division,  $\Delta L = 30, 65, 100$  bp, shown in panel C.

(C) Estimated division rate for an average HSC during each phase. Note that the y axis is in logarithmic scale.

models, showing that the total lifetime divisions of a typical HSC are comparable for both, even though the division schedules are distinct. Results are provided in Figure 2 and Table 2. The estimated division rates are recovered from the fitted slopes on each phase by dividing our assumed telomere loss per division ( $\Delta L$ ). Then, multiplying these predicted rates by the length of the phase gives an estimate for the number of divisions that the average HSC has accrued over the duration of the given phase.

### Significance for total lifetime HSC divisions

According to the 3-phase model (given the timing of its fitted breakpoints), a relatively high rate of HSC divisions is predicted during the first year of life, tapering at adolescence, and then decreasing further to a stable division rate for adults over 24 years old. This prediction appears to be consistent with Aubert et al.<sup>25</sup> and with recent evidence that the HSC pool expands very rapidly to some stable level in the early years of life.<sup>6,7</sup> The much higher division rate seen in the first year of life may reflect increased self-renewal divisions that grow the HSC pool to its long-term carrying capacity. The continued growth of the HSC pool during adolescence may be due to increased demand for blood cells as the individual grows. The 2-phase model misses this early trend and predicts only the decelerating rate of division after about 21 years of life.

### Confirming conclusions with distinct datasets

We asked whether our conclusions so far are highly dependent on the specific dataset we employed. To address this question, we used the same method on two independent datasets. One for granulocyte telomere length values reported for the LifeLines Deep cohort of normal individuals studied in the Netherlands.<sup>35</sup> A second dataset, by Alder et al.,<sup>24</sup> is from healthy individuals in the Baltimore area. Both granulocyte telomere length datasets were gathered with Flow-FISH, the same technique used by Aubert et al. The data from Alder et al. consist of measurements in 140 individuals from birth to 82 years old, whereas the LifeLines data are for 1,072 individuals from 18 to 81 years old.

**Table 2. Estimated division rates and total divisions from 2- and 3-phase models**

Age Range	Estimated Divisions/Year (Range)	Number of Divisions (Range)
<b>3-phase model</b>		
0 to 0.95	8.78 (5.71–19.03)	8.34 (5.42–18.1)
0.95 to 24.2	0.96 (0.63–2.09)	22.4 (14.5–48.5)
24.2 to 85	0.41 (0.27–0.88)	24.8 (16.1–53.8)
0 to 85		55.5 (36.0–120)
<b>2-phase model</b>		
0 to 20.9	1.25 (0.813–2.71)	26.1 (17.0–56.6)
20.9 to 85	0.41 (0.27–0.88)	26.2 (17.0–56.7)
0 to 85		52.3 (34.0–113)

Estimates for HSC division rates and accumulation of divisions over each phase for the average HSC in the 3- and 2-phase models, respectively. The estimates assume a loss of 65 bp of telomeric DNA per cell division. The lower (upper) bound on each interval is obtained from assuming 100 (respectively 30) bp of telomeric DNA is lost per HSC division.

Results for both datasets are shown in [Figure 3](#). For the LifeLines data, we find, that a 1-phase model fits best, scoring 3.5 and 6.8 points below the 2- and 3-phase models' AICs, respectively.

The range of adulthood HSC division rates obtained from the 1-phase model is consistent with our previous analysis, predicting 25 (with a range of 16–55) divisions in the age range 18–81. Recall that our previous 3-phase model predicted 25 (with a range of 16–54) divisions in the age range of 24–85.

With the Alder et al. data, we find that a 2-phase model with a phase-transition point at 26 years is optimal, scoring 12.1, 3.3, and 3.7 points below the 1-, 3-, and 4-phase models' AIC scores, respectively. The Alder et al. dataset consists of significantly fewer adolescent measurements than the Aubert et al. dataset. This is likely the cause of the slight discrepancy in the predicted adulthood phase-transition point. The 2-phase model predicts 50 lifetime divisions (with a range of 32–107) for the average HSC, consistent with our 2-phase model for the Aubert et al. data.

### Related approaches—A comparison

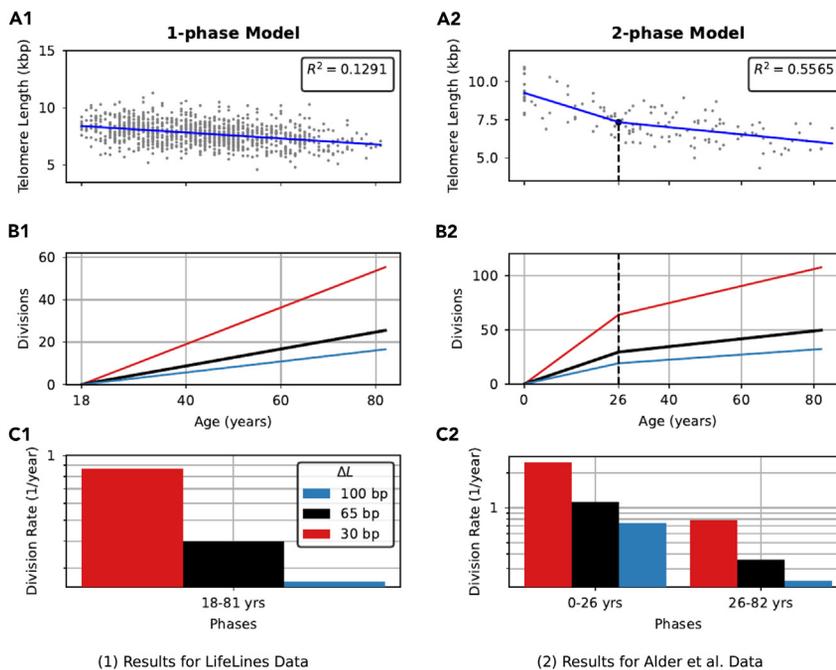
Werner et al.<sup>21</sup> proposed a mathematical model of HSC divisions. In contrast to our “top-down” approach (from data to inference about the underlying HSC division schedule), that model made specific assumptions about the progression of HSCs through division classes. For example, an assumption was made that the HSC rate of division is inversely proportional to the total HSC pool size. Predictions of that model, including analytic approximations, were then appraised against telomere length measurement data, including the dataset we have used.

We sought to compare the results of Werner et al. with our own. With parameters reported by Werner et al. (assuming 30 to 100 bp telomere loss per division), we used both full simulations of their HSC turnover dynamics model as well as the approximate telomere length formula they reported (a shifted logarithm) to obtain *de novo* estimates for HSC divisions (see the [STAR Methods](#) for details).

We report the upper bound of the model predictions by setting the maximum possible number of HSC divisions to be so large that relatively few HSCs reach this state given the Werner parameters, simplifying the model interpretation. The resulting progression for the average number of divisions in the stem cell pool is graphed in [Figure 4](#). Even in the absence of an upper bound on HSC divisions, we find that Werner et al. model predicts significantly fewer lifetime divisions than our method. In our hands, fitting the predictions of Werner et al. to the telomere data of Aubert et al. results in parameters and a final estimate with even fewer total lifetime HSC divisions than our methods predict.

## DISCUSSION

In this study, we used several granulocyte telomere length datasets for healthy humans from infancy to old age to infer the turnover of stem cells in the hematopoietic system over a lifetime. Our data-driven



**Figure 3. Confirming conclusions with distinct datasets**

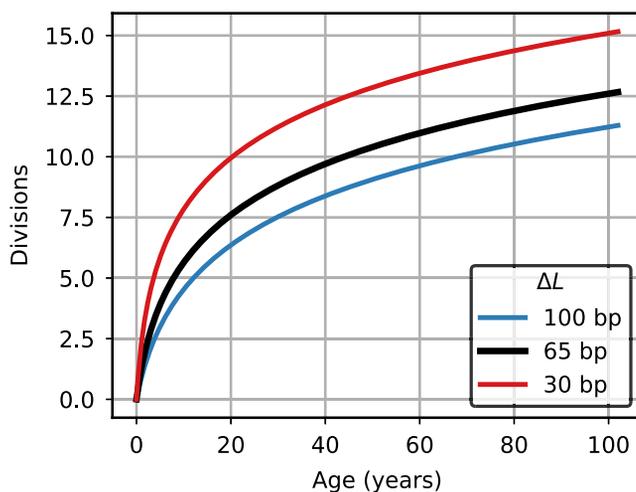
As in Figure 2 but for telomere length datasets provided by Andreu-Sánchez et al.<sup>35</sup> (1) and Alder et al.<sup>24</sup> (2). The LifeLines dataset in (1) consists of 1,072 individuals aged 18 to 81. The 1-phase model best fits these data. The adulthood HSC division rates are consistent with those obtained from the Aubert et al. data (see Table 2). (2) A two-phase model fits the Alder et al. data best and predicts 50 lifetime HSC divisions (with a range of 32–107), consistent with the 2-phase model for the Aubert et al. data.

approach provides estimates for the rate of HSC divisions at various phases of life, as well as the number and timing of those phases. The AIC allowed us to reject some models in favor of those most parsimonious models that best fit the data.

Our results suggest between 30 and 120 HSC divisions over a lifetime, with roughly half occurring by an individual's early 20s. In adults (20+ years-old) we find that HSCs divide once every 1 to 4 years. The same estimate holds using any of the three independent telomere length datasets that were made available to us (see Figure 3). We found AIC and BIC support for the 3- and 2-phase models over and above the 1- and 4-phase models. The 3-phase model predicts the most rapid turnover in infancy, which has some independent biological support.<sup>6,7</sup> We conjecture that such an accelerated turnover period in the first year of life could reflect an increased HSC symmetric division rate that serves to grow the stem cell pool toward some lifetime carrying capacity of around 50,000 to 200,000 cells.<sup>6,7</sup> However, the evidence for 3 rather than 2 phases is not decisive. The adolescent HSC division rate is reduced by over half of its value during the first year of life but is more than double that of adults over the age of 20. This observation may reflect increased HSC turnover at young ages due to increased demand for blood cells in these stages of growth.

Our predicted total number of HSC divisions over a lifetime is surprisingly low given previous estimates of a division every 2 weeks to 2 months, a frequency that would surpass 500 total cell divisions over a lifetime.<sup>1</sup> Based in part on such estimates, it typically has been assumed that HSCs must express telomerase to avoid replicative senescence or apoptosis (the "Hayflick limit") imposed by progressive loss of telomere repeats with each cell division.<sup>12</sup>

Our models are based on several assumptions. The first is that the number of cell divisions between an HSC and a granulocyte is more or less constant over a lifetime (see also Supplementary Note 9 by Abascal et al.<sup>8</sup>). This notion seems difficult to reconcile with the effect of growth factors such as granulocyte colony stimulating factor (G-CSF) that are known to stimulate transient granulocyte production. We propose that



**Figure 4. Related approaches—a comparison**

Progression of the average number of divisions in the HSC pool over 102 years predicted by the Werner et al. model. Colors correspond to various assumptions for the telomeric DNA loss per division,  $\Delta L = 30, 65, 100$ . In each case, we used the fitted parameters reported by Werner et al. Since we assumed that a significant number of HSCs do not reach a maximal division number, these results are an upper bound on what the model predicts for lifetime divisions.

such transient stimulation of granulocyte production by cytokines induced by infection, injection, or injury primarily triggers extra cell divisions in committed progenitor cells, rather than significantly impacting HSC turnover. Such transient responses likely contribute to noise in the data. Possible cytokine effect on the “mobilization” or anatomical localization of HSCs is also assumed to have a limited effect on the turnover of HSCs. In contrast, loss of HSCs, e.g., by irradiation, toxins, or telomerase deficiency, would increase HSC turnover, potentially contributing to eventual bone marrow failure.<sup>12</sup> Such increases in HSC turnover are reflected in the telomere length in nucleated blood cells.<sup>36</sup>

The second assumption in our model is that telomere repeats are lost in HSCs at a more or less constant rate with each cell division. For this purpose, we took the original assumption of 30–100 bp lost with each HSC division.<sup>23</sup> Such estimates were based on the measured loss of telomere repeats per population doubling in cultured fibroblasts and T cells. Serially cloned T cells were found to lose 60–70 bp of telomeric DNA per division, in line with these previous estimates.<sup>17</sup> In general, the role of telomerase in HSCs has been puzzling.<sup>12</sup> Whereas telomerase is limited by expression levels of the telomerase RNA template gene in early embryos,<sup>26,37</sup> the expression of the telomerase reverse transcriptase gene hTERT limits telomerase expression in most somatic cells.<sup>38</sup> The suppression of telomerase varies between somatic cells and is not observed in somatic cells from short-lived animals such as laboratory mice.<sup>15</sup> Where and when telomerase becomes unable to maintain telomeres in human HSCs is currently not known. In our model, we assume that telomerase activity in HSCs cannot prevent overall telomere erosion with each cell division but nevertheless plays a crucial role in the repair of sporadic damage to telomeric DNA as was recently proposed.<sup>12</sup> Here, we assumed that telomerase levels in HSCs of newborns and the very old are suppressed to the same degree. This proposition requires experimental validation. Given that telomerase is upregulated during DNA replication, both the low frequency and low turnover of HSCs are major challenges for the experimental verification of this hypothesis.

#### Limitations of the study

The results produced by our method are limited by the assumptions we have discussed above regarding the structure of the hematopoietic hierarchy and the amount of telomeric DNA lost with each stem cell division. Further advances in experimental techniques for studying HSCs more directly are needed to validate these assumptions and test the accuracy of our methods and results. Moreover, we have been limited to cross-sectional data of telomere length with age. Measuring changes in telomere length in individuals over significant periods of time is needed to better understand differences in HSC turnover over early life as well as between individuals.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
  - Lead contact
  - Materials availability
  - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
- [METHOD DETAILS](#)
  - Alternatives to phase-based piecewise-linear model
  - Comparison with the werner et al. model
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.107053>.

## ACKNOWLEDGMENTS

C.B. has been supported as an undergraduate summer research associate (USRA) by the Natural Sciences and Engineering Research Council of Canada (NSERC) Canada. L.E.K. is supported by an NSERC Discovery grant. Work in the laboratory of P.M.L. is funded by a Program Project Grant (#1074) from the Terry Fox Research Institute, a Project Grant (#PJT-159787) from the Canadian Institutes of Health Research, and a grant (#40044) from the Canadian Foundation for Innovation and the Government of British Columbia. Telomere length data for the LifeLines Deep cohort were generously provided by the authors of Andreu-Sánchez et al.<sup>35</sup> We thank the anonymous reviewers for suggestions that improved the paper.

## AUTHOR CONTRIBUTIONS

Research was designed by P.M.L., L.E.K., and C.B.; the research was carried out by C.B.; the paper was written by C.B., L.E.K., and P.M.L.; the paper was edited by C.B., L.E.K., and P.M.L.; research was funded by L.E.K. and P.M.L.

## DECLARATION OF INTERESTS

P.M.L. is a founding shareholder of Repeat Diagnostics Inc., a company specializing in clinical telomere length measurements since 2006.

Received: September 21, 2022

Revised: May 9, 2023

Accepted: June 1, 2023

Published: June 8, 2023

## REFERENCES

1. Tomasetti, C., and Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347, 78–81. <https://doi.org/10.1126/science.1260825>.
2. Till, J.E., and McCulloch, E.A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* 14, 213–222. <https://doi.org/10.2307/3570892>.
3. Siminovitch, L., McCulloch, E.A., and Till, J.E. (1963). The distribution of colony-forming cells among spleen colonies. *J. Cell. Comp. Physiol.* 62, 327–336.
4. Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553, 418–426. <https://doi.org/10.1038/nature25022>.
5. Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. *Cell Stem Cell* 10, 120–136. <https://doi.org/10.1016/j.stem.2012.01.006>.
6. Mitchell, E., Spencer Chapman, M., Williams, N., Dawson, K.J., Mende, N., Calderbank, E.F., Jung, H., Mitchell, T., Coorens, T.H.H., Spencer, D.H., et al. (2022). Clonal dynamics of haematopoiesis across the human lifespan. *Nature* 606, 343–350. <https://doi.org/10.1038/s41586-022-04786-y>.
7. Lee-Six, H., Øbro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R.J., Huntly, B.J.P., Martincorena, I., Anderson, E., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature* 561, 473–478. <https://doi.org/10.1038/s41586-018-0497-0>.
8. Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., Russell, A.J.C., Alcantara, R.E., Baez-Ortega, A., Wang, Y., et al. (2021). Somatic mutation landscapes at single-molecule resolution. *Nature* 593, 405–410. <https://doi.org/10.1038/s41586-021-03477-4>.
9. Summers, C., Rankin, S.M., Condliffe, A.M., Singh, N., Peters, A.M., and Chilvers, E.R. (2010). Neutrophil kinetics in health and

- disease. *Trends Immunol.* 31, 318–324. <https://doi.org/10.1016/j.it.2010.05.006>.
10. Higgins, J.M. (2015). Red blood cell population dynamics. *Clin. Lab. Med.* 35, 43–57. <https://doi.org/10.1016/j.cl.2014.10.002>.
  11. Derényi, I., and Szöllösi, G.J. (2017). Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations. *Nat. Commun.* 8, 1–8. <https://doi.org/10.1038/ncomms14545>.
  12. Lansdorp, P.M. (2022). Telomeres, aging, and cancer: the big picture. *Blood. The Journal of the American Society of Hematology* 139, 813–821. <https://doi.org/10.1182/blood.2021014299>.
  13. Shivdasani, R.A., Clevers, H., and de Sauvage, F.J. (2021). Tissue regeneration: reserve or reverse? *Science* 371, 784–786. <https://doi.org/10.1126/science.abb6848>.
  14. Scally, A. (2016). Mutation rates and the evolution of germline structure. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150137. <https://doi.org/10.1098/rstb.2015.0137>.
  15. Lansdorp, P. (2022). Telomere length regulation. *Front. Oncol.* 12, 943622. <https://doi.org/10.3389/fonc.2022.943622>.
  16. Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *science* 279, 349–352. <https://doi.org/10.1126/science.279.5349.349>.
  17. Rufer, N., Migliaccio, M., Antonchuk, J., Humphries, R.K., Roosnek, E., and Lansdorp, P.M. (2001). Transfer of the human telomerase reverse transcriptase (tert) gene into t lymphocytes results in extension of replicative potential. *Blood* 98, 597–603. <https://doi.org/10.1182/blood.V98.3.597>.
  18. Brümmendorf, T.H., Mak, J., Sabo, K.M., Baerlocher, G.M., Dietz, K., Abkowitz, J.L., and Lansdorp, P.M. (2002). Longitudinal studies of telomere length in feline blood cells: implications for hematopoietic stem cell turnover in vivo. *Exp. Hematol.* 30, 1147–1152. [https://doi.org/10.1016/s0301-472x\(02\)00888-3](https://doi.org/10.1016/s0301-472x(02)00888-3).
  19. Shepherd, B.E., Gutter, P., Lansdorp, P.M., and Abkowitz, J.L. (2004). Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths. *Exp. Hematol.* 32, 1040–1050. <https://doi.org/10.1016/j.exphem.2004.07.023>.
  20. Shepherd, B.E., Kiem, H.P., Lansdorp, P.M., Dunbar, C.E., Aubert, G., LaRochelle, A., Seggewiss, R., Gutter, P., and Abkowitz, J.L. (2007). Hematopoietic stem-cell behavior in nonhuman primates. *Blood. The Journal of the American Society of Hematology* 110, 1806–1813. <https://doi.org/10.1182/blood-2007-02-075382>.
  21. Werner, B., Beier, F., Hummel, S., Balabanov, S., Lassay, L., Orlikovsky, T., Dingli, D., Brümmendorf, T.H., and Traulsen, A. (2015). Reconstructing the in vivo dynamics of hematopoietic stem cells from telomere length distributions. *Elife* 4, e08687. <https://doi.org/10.7554/eLife.08687>.
  22. Edelstein-Keshet, L., Israel, A., and Lansdorp, P. (2001). Modelling perspectives on aging: can mathematics help us stay young? *J. Theor. Biol.* 213, 509–525. <https://doi.org/10.1006/jtbi.2001.2429>.
  23. Vaziri, H., Dragowska, W., Allsopp, R.C., Thomas, T.E., Harley, C.B., and Lansdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric dna with age. *Proc. Natl. Acad. Sci. USA* 91, 9857–9860. <https://doi.org/10.1073/pnas.91.21.9857>.
  24. Alder, J.K., Hanumanthu, V.S., Strong, M.A., DeZern, A.E., Stanley, S.E., Takemoto, C.M., Danilova, L., Applegate, C.D., Bolton, S.G., Mohr, D.W., et al. (2018). Diagnostic utility of telomere length testing in a hospital-based setting. *Proc. Natl. Acad. Sci. USA* 115, E2358–E2365. <https://doi.org/10.1073/pnas.1720427115>.
  25. Aubert, G., Baerlocher, G.M., Vulto, I., Poon, S.S., and Lansdorp, P.M. (2012). Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. *PLoS Genet.* 8, e1002696. <https://doi.org/10.1371/journal.pgen.1002696>.
  26. Lansdorp, P.M. (2022). Sex differences in telomere length, lifespan, and embryonic dyskerin levels. *Aging Cell* 21, e13614. <https://doi.org/10.1111/accel.13614>.
  27. Rodriguez-Brenes, I.A., and Peskin, C.S. (2010). Quantitative theory of telomere length regulation and cellular senescence. *Proc. Natl. Acad. Sci. USA* 107, 5387–5392. <https://doi.org/10.1073/pnas.0914502107>.
  28. Baerlocher, G.M., Rice, K., Vulto, I., and Lansdorp, P.M. (2007). Longitudinal data on telomere length in leukocytes from newborn baboons support a marked drop in stem cell turnover around 1 year of age. *Aging Cell* 6, 121–123. <https://doi.org/10.1111/j.1474-9726.2006.00254.x>.
  29. Rufer, N., Brümmendorf, T.H., Kolvraa, S., Bischoff, C., Christensen, K., Wadsworth, L., Schulzer, M., and Lansdorp, P.M. (1999). 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. <https://doi.org/10.1084/jem.190.2.157>.
  30. Muggeo, V.M.R. (2003). Estimating regression models with unknown break-points. *Stat. Med.* 22, 3055–3071. <https://doi.org/10.1002/sim.1545>.
  31. Muggeo, V.M. (2008). Segmented: an r package to fit regression models with broken-line relationships. *R. News* 8, 20–25.
  32. Chen, C.W.S., Chan, J.S.K., Gerlach, R., and Hsieh, W.Y.L. (2011). A comparison of estimators for regression models with change points. *Stat. Comput.* 21, 395–414. <https://doi.org/10.1007/s11222-010-9177-0>.
  33. Cavanaugh, J.E., and Neath, A.A. (2019). The akaike information criterion: background, derivation, properties, application, interpretation, and refinements. *WIREs Comp. Stat.* 11, e1460. <https://doi.org/10.1002/wics.1460>.
  34. Burnham, K.P., and Anderson, D.R. (2004). Multimodel inference: understanding aic and bic in model selection. *Socio. Methods Res.* 33, 261–304. <https://doi.org/10.1177/0049124104268644>.
  35. Andreu-Sánchez, S., Aubert, G., Ripoll-Cladellas, A., Henkelman, S., Zhenakova, D.V., Sinha, T., Kurilshikov, A., Cenit, M.C., Jan Bonder, M., Franke, L., et al. (2022). Genetic, parental and lifestyle factors influence telomere length. *Communications Biology* 5, 1–14. <https://doi.org/10.1038/s42003-022-03521-7>.
  36. Alter, B.P., Rosenberg, P.S., Giri, N., Baerlocher, G.M., Lansdorp, P.M., and Savage, S.A. (2012). Telomere length is associated with disease severity and declines with age in dyskeratosis congenita. *Haematologica* 97, 353–359. <https://doi.org/10.3324/haematol.2011.055269>.
  37. Chiba, K., Johnson, J.Z., Vogan, J.M., Wagner, T., Boyle, J.M., and Hockemeyer, D. (2015). Cancer-associated tert promoter mutations abrogate telomerase silencing. *Elife* 4, e07918. <https://doi.org/10.7554/eLife.07918>.
  38. Shay, J.W., and Wright, W.E. (2019). Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* 20, 299–309. <https://doi.org/10.1038/s41576-019-0099-1>.
  39. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
  40. Baerlocher, G.M., Vulto, I., De Jong, G., and Lansdorp, P.M. (2006). Flow cytometry and fish to measure the average length of telomeres (flow fish). *Nat. Protoc.* 1, 2365–2376. <https://doi.org/10.1038/nprot.2006.263>.
  41. R Core Team (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). <https://www.R-project.org/>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Software and algorithms</i>		
R version 4.2.2	R Foundation for Statistical Computing	<a href="https://r-project.org">https://r-project.org</a>
Segmented version 1.6–2	Muggeo, <sup>30</sup> Muggeo <sup>31</sup>	<a href="https://CRAN.R-project.org/package=segmented">https://CRAN.R-project.org/package=segmented</a>
Python version 3.10.8	Python Software Foundation	<a href="https://www.python.org">https://www.python.org</a>
SciPy version 1.8.1	Virtanen et al. <sup>39</sup>	<a href="https://scipy.org">https://scipy.org</a>
Code used in this paper	This paper	<a href="https://doi.org/10.5281/zenodo.7939383">https://doi.org/10.5281/zenodo.7939383</a>
<i>Other</i>		
Telomere length data	Aubert et al. <sup>25</sup>	N/A
Telomere length data	Alder et al. <sup>24</sup>	N/A
Telomere length data	Andreu-Sánchez et al. <sup>35</sup>	N/A

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Cole Boyle ([cab@student.ubc.ca](mailto:cab@student.ubc.ca)).

#### Materials availability

This study did not generate new unique reagents or cell lines.

#### Data and code availability

- The telomere length datasets produced by Aubert et al.<sup>25</sup> and Alder et al.<sup>24</sup> are publicly available in each paper's respective supplementary information. The dataset of Andreu-Sánchez et al.<sup>35</sup> was kindly provided to us by the authors.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

No new human nor animal studies and no new experiments were used in this research. All data was obtained from previously published sources. A brief summary of that published data is given below.

Telomere length data in Aubert et al.<sup>25</sup> were obtained using nucleated blood cells analyzed using the flow-FISH technique.<sup>40</sup> With this technique, the average length of telomere repeats in cells is measured using fluorescent *in situ* hybridization (FISH) with peptide nucleic acid (PNA) probes specific for telomere repeats in combination with fluorescence measurements by flow cytometry. Accurate measurements of telomere length are obtained by including internal control cells with long telomeres (bovine thymocytes) in every sample and by correlating the results of fluorescence measurements in selected cell populations with results obtained by Southern analysis of control cells.<sup>40</sup>

In the study by Vaziri et al.<sup>23</sup> CD34<sup>+</sup> cells from fetal liver, cord blood and adult bone marrow were cultured for several weeks. Upon culture, CD34 cells were sorted to reinitiate cultures. The telomere length in cultured cells was measured using Southern blot analysis. The telomere length was longest in fetal liver cells, intermediate in cord blood cells and shortest in adult bone marrow cells. Strikingly, with each passage, a noticeable decline in telomere length was observed in all three cultures.

Finally, the data for granulocyte telomere length from Alder et al.<sup>24</sup> consists of measurements in 140 individuals from birth to 82 years old, whereas the Lifelines data is for 1072 individuals from 18 to 81 years old.

## METHOD DETAILS

### Alternatives to phase-based piecewise-linear model

Instead of piecewise linear fits to the average telomere length in granulocytes,  $\bar{L}_G$ , we also considered polynomial and logarithmic fits to the Aubert et al. dataset. These alternate fits predict similar lifetime HSC divisions and have AIC similar to our 3-phase piecewise linear model. (The latter had an absolute AIC of 2450.26 and predicted 56 (with a range of 36–120) lifetime divisions for the average HSC.) The predicted number of HSC divisions for both models are shown in Figure S1. Fittings were done using R. Fitted parameters and AIC for each model are shown in Table S1.

Fitting  $\bar{L}_G$  as a  $n^{\text{th}}$  degree polynomial leads to  $n+1$  parameters, the coefficients  $a_n, \dots, a_0$  in

$$\bar{L}_G(t) = L_0 - \Delta L \bar{D}(t) = a_n t^n + \dots + a_1 t + a_0 \quad (\text{Equation 1})$$

with  $a_0 = L_0$ . We obtain these coefficients by a least-squares fit of the polynomial to the data. Then, rearranging the above relationship, the average number of HSC divisions by time  $t$  is simply:

$$\bar{D}(t) = - \frac{a_n t^n + \dots + a_1 t}{\Delta L}. \quad (\text{Equation 2})$$

We report the predicted number of divisions for the third-degree polynomial fit (optimal in terms of AIC) in Figure S1 and the parameters  $a_0, a_1$ , and  $a_3$  in Table S1.

We can also model  $\bar{L}_G$  as a logarithmic function of time,

$$\bar{L}_G(t) = L_0 - \Delta L \bar{D}(t) = a_0 + a_1 \log(a_2 t + 1). \quad (\text{Equation 3})$$

A nonlinear least squares fit results in values for  $a_0 = L_0, a_1$ , and  $a_2$ . The predicted average number of divisions is then given by

$$\bar{D}(t) = - \frac{a_1}{\Delta L} \log(a_2 t + 1). \quad (\text{Equation 4})$$

To obtain convergence during the fitting process, we start with  $a_0$  in the interval  $[5, 15]$ ,  $a_1$  in  $[-10, 0]$ , and  $a_2$  in  $[0, 1]$ .

### Comparison with the werner et al. model

Werner et al.<sup>21</sup> provided an ordinary differential equation (ODE) model of HSC asymmetric and symmetric divisions with six parameters: the average telomere length at birth,  $c$ , the telomeric length lost per HSC division,  $\Delta c$ , the probability  $p$  of a symmetric HSC division (versus asymmetric,  $1 - p$ ), the HSC proliferation rate  $r$ , the size of the HSC pool at birth,  $N_0$ , and the maximum number of times an HSC can divide  $n$ . (Here we have used  $n$  in place of Werner's notation,  $c$ ). The authors assumed that the rate of HSC divisions is inversely proportional to the size of the stem cell pool. The model they proposed is then given by the ODEs:

$$\frac{d}{dt} N_i = \begin{cases} -\frac{r N_i}{N_0 + r p t} & (i = 0) \\ \frac{r}{N_0 + r p t} (-N_i + (1+p)N_{i-1}) & (0 < i < n) \\ \frac{r N_{i-1}}{N_0 + r p t} & (i = n) \end{cases} \quad (\text{Equation 5})$$

where  $N_i(t)$  is the number of HSCs that have undergone  $i$  divisions by time  $t$ . Werner et al. used (5) to approximate the expected telomere length  $\bar{L}(t)$  at a time  $t$  after birth, stating that:

$$\bar{L}(t) \approx c - \Delta c \frac{1+p}{p} \log\left(p \frac{r}{N_0} t + 1\right). \quad (\text{Equation 6})$$

The above approximation holds for  $n$  large enough that few cells divide  $n$  times. It appears that Werner et al. then used a binomial approximation  $\left(p \frac{r}{N_0} t + 1\right)^{\Delta c} \approx p \frac{r \Delta c}{N_0} t + 1$  to further simplify their formula to the form

$$\bar{L}(t) \approx c - \frac{1+p}{p} \log\left(p \frac{r\Delta c}{N_0} t + 1\right). \quad (\text{Equation 7})$$

They then fit (7) to their own granulocyte telomere length dataset (the Werner et al. original dataset was no longer available and could not be included in our study) and reported the following parameters

$$c = 10.2 \text{ kbp}, p = 0.44, \frac{r\Delta c}{N_0} = 68 \text{ bp/year}. \quad (\text{Equation 8})$$

We found that for these parameters (and  $\Delta c \approx 30 - 100$  bp), the approximation of Equation 6 by Equation 7 fails - deviates significantly from the actual trend, even on a short timescale (see Figure S2). For this reason, we avoided this route.

In order to obtain the parameter values needed for simulating the dynamics of the Werner model (5), we divide  $\frac{r\Delta c}{N_0}$  by several values of  $\Delta c$  in the range 30–100 bp. Each choice of  $\Delta c$  results in a unique simulation. Then we obtain a value for  $r$  by taking  $N_0 = 100$ , without loss of generality. Changing this value merely results in a proportional change in the value of  $r$ . Finally, we implement the model taking  $n = 35$  since at this value we get an upper bound for predicted lifetime divisions (at this value no cells reach the  $n^{\text{th}}$  division class so taking  $n$  larger does not change the resulting division estimates). Reducing  $n$  significantly results in limiting the possible number of divisions that an HSC can go through, hence reducing the number of divisions that the average HSC accrues over a lifetime.

We solve the system of differential equations in (5) with the Python library Scipy's<sup>39</sup> solve\_ivp function over a time interval of 0–102 years. From a simulation, we interpolate the numerical solution for the  $N_i$  at 300 evenly spaced time points (sufficient to obtain accuracy) in the range 0–102. We then obtain the average number of divisions in the stem cell pool at one of these time points,  $t$  by dividing the sum  $\sum_{i=0}^n iN_i(t)$  by the total number,  $N(t)$  of stem cells ( $N(t) = N_0(t) + \dots + N_n$ ).

The average number of divisions is then plotted versus  $t$ , to obtain Figure 4 in the main text.

We can also obtain the same results from Equation 6. Dividing the value of  $\frac{r\Delta c}{N_0}$  reported by Werner et al., by a range of  $\Delta c$  values (30 – 100 bp), results in a range of values for  $\frac{r}{N_0}$ . Plugging these values and the Werner et al. reported value of  $p$  into Equation 6 we get:

$$\bar{L}(t) = c - \Delta c \bar{D}(t) = c - \Delta c \frac{1+p}{p} \log\left(p \frac{r}{N_0} t + 1\right) \Rightarrow \bar{D}(t) = \frac{1+p}{p} \log\left(p \frac{r}{N_0} t + 1\right)$$

Plotting this relationship for  $\bar{D}(t)$ , for each value of  $\Delta c$  gives the same results as in our simulations.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was done with R (4.2.2).<sup>41</sup> Purely linear models were fit using the lm function in R. Continuous piecewise linear models with 1 or more transition points were fit using the segmented function from the R package Segmented (1.6–2).<sup>30,31</sup> For each dataset, we fit models with a differing number of transition points and compute their AIC and BIC using the AIC and BIC functions in R. We reject models with the largest AIC, and use the general rule of thumb that models that differ by less than 2 AIC points are equally likely.<sup>34</sup>

For the Aubert et al.<sup>25</sup> data, segmented does not return a result for a 5-phase model. However, as the AIC continues to increase for 6 and 7-phase models, we are confident that the 2 or 3-phase models are near optimality.