SI Appendix

Here we derive the results for the models discussed in the main text. Let the random variable $N(t)$ represent the number of transcripts of p16 present at time t. The average rates at which transcripts are produced and degraded are $u_0 + u\, N(t)$ and $w_0 + w\, N(t)$, respectively, while the average rate at which an individual is removed from the population is $d_0 + d\, N(t)$. We assume that the initial transcript level $N(0)=n_0$ is also a random number with a mean μ_0 and variance σ_0^2. Finally, let $p_n(t)$ = Prob[$N(t) = n$]. That is, $p_n(t)$ is the probability that at time t an individual has n transcripts.

MODEL I

A simple model that includes only forward, $u_k = uk + u_0$, and backward, $w_k = wk + w_0$, transition rates can produce expression levels that initial grow exponentially, but eventually saturate. To capture this behavior, u_k has to be larger than w_k for small k and smaller than w_k for large k. These conditions are satisfied if $u_0 > w_0$ and $u < w$. The mean steady-state value $n_s = \dfrac{u_0 - w_0}{w - u}$ corresponds to the state in which the rates balance $u_{n_s} = w_{n_s}$.

The equations for the probabilities $p_k(t)$ are:

$$\begin{aligned} &\dot{p}_k(t) = -[(u_0 + w_0) + k(u + w)]p_k(t) + (uk + u_0 - u)p_{k-1}(t) + (wk + w_0 + w)p_{k+1}(t), \\ &p_k(0) = \delta_{kn_0} \end{aligned} \tag{1}$$

We make use of the following notation: $q(t) \equiv \overline{N}(t) = \sum_{k=-\infty}^{\infty} kp_k$, $r(t) = \sum_{k=-\infty}^{\infty} k^2 p_k$, and $s(t) = \sum_{k=-\infty}^{\infty} k^3 p_k$. Because there are no absorbing states in this model, the probabilities $p_k(t)$ sum to one $\sum_{k=-\infty}^{\infty} p_k = 1$. From which it follows: $\sum_{k=-\infty}^{\infty} kp_{k-1} = q + 1$, $\sum_{k=-\infty}^{\infty} kp_{k+1} = q - 1$, $\sum_{k=-\infty}^{\infty} k^2 p_{k-1} = r + 2q + 1$, $\sum_{k=-\infty}^{\infty} k^2 p_{k+1} = r - 2q + 1$, $\sum_{k=-\infty}^{\infty} k^3 p_{k-1} = s + 3r + 3q + 1$, and $\sum_{k=-\infty}^{\infty} k^3 p_{k+1} = s - 3r + 3q - 1$. Multiplying Eq. (1) by k and k^2, summing over all k, and using the above expressions we find that:

$$\begin{aligned} &\dot{q} = q(u - w) + (u_0 - w_0), \\ &\dot{r} = 2r(u - w) + 2q(u_0 - w_0) + q(u + w) + (u_0 + w_0), \end{aligned} \tag{2}$$

with the initial conditions $q(0) = n_0$ and $r(0) = n_0^2$. Solving Eq. (2) we obtain:

$$\begin{aligned} &q(t) = n_0 e^{(u-w)t} + \frac{u_0 - w_0}{u - w}\left[e^{(u-w)t} - 1\right], \\ &r(t) = n_0^2 e^{2(u-w)t} + n_0 \frac{u + w + 2u_0 - 2w_0}{u - w}\left[e^{2(u-w)t} - e^{(u-w)t}\right] \\ &+ \frac{(u + w + 2u_0 - 2w_0)(u_0 - w_0)}{2(u - w)^2}\left[e^{(u-w)t} - 1\right]^2 + \frac{u_0 + w_0}{2(u - w)}\left[e^{2(u-w)t} - 1\right] \end{aligned} \tag{3}$$

Averaging the above expressions over all possible n_0, produces the following expressions for mean and variance of the expression level:

$$\left\langle \bar{N}(t)\right\rangle = \left\langle q(t)\right\rangle = \mu_0 e^{-(w-u)\,t} + \frac{u_0 - w_0}{w-u}\left[1 - e^{-(w-u)t}\right] \tag{4}$$

$$\begin{aligned}\sigma^2(t) = \left\langle r(t)\right\rangle - \left\langle q(t)\right\rangle^2 &= \sigma_0^2 e^{-2(w-u)t} + \mu_0 \frac{u+w}{w-u} e^{-(w-u)t}\left[1 - e^{-(w-u)t}\right] \\ &+ \frac{\left[1 - e^{-(w-u)t}\right]}{(w-u)^2}\left[(wu_0 - uw_0) - (uu_0 - ww_0)e^{-(w-u)t}\right]\end{aligned} \tag{5}$$

As expected, for large *t*, the mean approaches $\left\langle \bar{N}(t)\right\rangle \approx n_s$, while for small *t* we find that $\left\langle \bar{N}(t)\right\rangle \approx \mu_0 + t(w-u)(n_s - \mu_0)$.

Limiting case for *t* >> 1/(*w* - *u*)

In the limit of large *t*, we find $\left\langle \bar{N}(t)\right\rangle \approx \frac{u_0 - w_0}{w-u}$ and $\sigma^2(t) \approx \frac{wu_0 - uw_0}{(w-u)^2}$, in which case the coefficient of variation (standard deviation/mean) is given by $c_v \approx \frac{\sqrt{wu_0 - uw_0}}{|u_0 - w_0|}$. It is natural to consider two cases:

l) Suppose $w >> u$, then if $n_s >> 1$, as measure for p16 levels, we have $u_0 >> w_0$. In this case, for large *t* we find that $\left\langle \bar{N}(t)\right\rangle \approx \frac{u_0}{w}$ and $c_v \approx \sqrt{\frac{w}{u_0}} = \frac{1}{\sqrt{\left\langle \bar{N}\right\rangle}} << 1$. Therefore, in this limit the model is unable to account for the measured variability in p16 expression levels.

2) In the opposite limit when $w \sim u$, we find that the coefficient of variation is not necessarily small $c_v \approx \sqrt{\frac{w}{u_0 - w_0}}$, because $(u_0 - w_0)$ can be chosen to be of the order of *w*. For example, if we choose $(u_0 - w_0) = w = \frac{n_s u}{n_s - 1}$, then at large *t* we get $\left\langle \bar{N}(t)\right\rangle \approx n_s$ and $c_v \approx 1$. However, in this limit, on the time scales that correspond to the p16 measurements (<80 years) the dependence is approximately linear.

MODEL II

In this section we derive the results for the model in which saturation of p16 levels occur due to the removal of individuals from the sample population at an average rate of $d_0(t) + d\,N(t)$. This model also produces expression levels that initially grow exponentially then eventually saturate. In this case $p_n(t)$ satisfies the following equation:

$$\frac{dp_n(t)}{dt} = -\left[nu + nd + u_0 + d_0(t)\right] p_n(t) + \left[u(n-1) + u_0\right] p_{n-1}(t) \qquad (6)$$

Initially, we assume that at $t = 0$, the transcript level is given by a fixed value n_0. That is, $p_{n_0}(0) = 1$. Below we relax this assumption and allow the initial transcript level to be a random variable. Because in this model we do not consider events that decrease the expression level, $p_n(t) \equiv 0$ for $n < n_0$. We simplify Eq. (6) using the following transformation

$$p_n(t) = f_n(t) \exp\left[-\left(n_0 u + n_0 d + u_0\right)t - \int_0^t d_0(t')\,dt'\right] \qquad (7)$$

so that $f_n(t) = 0$ for $n < n_0$, $f_{n_0}(t) = 1$ and for $n > n_0$ we have

$$\dot{f}_n(t) = -(n-n_0)(u+d)\,f_n(t) + \left[u(n-n_0) + u(n_0-1) + u_0\right] f_{n-1} \qquad (8)$$

We look for solutions of Eq. (8) of the form

$$f_{n_0+s} = A_s h^s(t) \text{ for } s = 0, 1, 2, \ldots \qquad (9)$$

From the initial condition, it follows that $A_0 = 1$. Inserting Eq. (9) into Eq. (8), we find that for s > 0

$$sA_s\left[\dot{h}(t) + (u+d)h(t)\right] = A_{s-1}\left[us + u(n_0-1) + u_0\right] \qquad (10)$$

Therefore,

$$A_s = \frac{us + u(n_0-1) + u_0}{sH} A_{s-1} \qquad (11)$$

and

$$\begin{aligned} &\dot{h}(t) + (u+d)h(t) = H, \\ &h(0) = 0, \end{aligned} \qquad (12)$$

where H is an arbitrary non-zero constant. Solving Eq. (11) recursively, we find

$$A_s = \frac{u^s}{s!H^s}\prod_{m=0}^{s-1}\left[m+\left(n_0+\frac{u_0}{u}\right)\right], \tag{13}$$

while the solution of Eq. (12) is

$$h(t) = \frac{H}{u+d}\left[1-e^{-(u+d)t}\right]. \tag{14}$$

Finally, we obtain

$$f_{n+s}(t) = \frac{1}{s!}\left[\frac{uh(t)}{H}\right]^s \prod_{m=0}^{s-1}\left[m+\left(n_0+\frac{u_0}{u}\right)\right] = \frac{\left[g(t)\right]^s}{s!}\prod_{m=0}^{s-1}(m+a), \tag{15}$$

where $g(t) \equiv \frac{u}{u+d}\left[1-e^{-(u+d)t}\right]$ and $a \equiv n_0 + \frac{u_0}{u}$.
We first compute the mean expression level $\overline{N}(t)$ for fixed n_0:

$$\overline{N}(t) = \frac{\sum_{k=0}^{\infty} k p_k(t)}{\sum_{k=0}^{\infty} p_k(t)} = \frac{\sum_{s=0}^{\infty}(n_0+s)\, f_{n+s}(t)}{\sum_{s=0}^{\infty} f_{n+s}(t)} = n_0 + \frac{ag(t)}{1-g(t)} = \frac{1}{1-g(t)}\left[n_0 + g(t)\frac{u_0}{u}\right] \tag{16}$$

Similarly,

$$\begin{aligned}\overline{N^2}(t) &= \frac{\sum_{k=0}^{\infty} k^2 p_k(t)}{\sum_{k=0}^{\infty} p_k(t)} = \frac{\sum_{s=0}^{\infty}(n_0+s)^2\, f_{n_0+s}(t)}{\sum_{s=0}^{\infty} f_{n_0+s}(t)} = n_0^2 + 2n_0\frac{ag}{1-g} + \frac{ag(1+ag)}{(1-g)^2} \\ &= \frac{1}{(1-g)^2}\left[n_0^2 + n_0 g\left(1+2\frac{u_0}{u}\right) + g\frac{u_0}{u}\left(1+g\frac{u_0}{u}\right)\right]\end{aligned}. \tag{17}$$

Given n_0 has a mean value of μ_0 and variance of σ_0^2, averaging with respect to n_0 produces the following expressions for the mean, variance and coefficient of variation of *N(t)*:

$$\left\langle \overline{N}(t)\right\rangle = \frac{1}{1-g(t)}\left[\mu_0 + g(t)\frac{u_0}{u}\right], \tag{18}$$

$$\sigma^2(t) = \left\langle \overline{N^2}(t)\right\rangle - \left\langle \overline{N}(t)\right\rangle^2 = \frac{1}{\left[1-g(t)\right]^2}\left[\sigma_0^2 + g(t)\left(\mu_0 + \frac{u_0}{u}\right)\right], \tag{19}$$

$$c_v \equiv \frac{\sigma}{\left\langle \overline{n}(t)\right\rangle} = \frac{\sqrt{\sigma_0^2 + g(t)\left(\mu_0 + \frac{u_0}{u}\right)}}{\mu_0 + g(t)\frac{u_0}{u}}. \tag{20}$$

Results for the limit *t* >> 1/(*u* + *d*)

For older people $g(t) \approx \frac{u}{u+d}$, so that $\langle \bar{N} \rangle \approx \frac{u+d}{d}\left[\mu_0 + \frac{u_0}{u+d}\right]$ and $c_v \approx \frac{\sqrt{\sigma_0^2 + \frac{u}{u+d}\left(\mu_0 + \frac{u_0}{u}\right)}}{\mu_0 + \frac{u_0}{u+d}}$

1) For the case of a constant growth rate $u_k = u_0$ (i.e. $u = 0$), in the long time limit, we have $\langle \bar{N} \rangle \approx \mu_0 + \frac{u_0}{d}$ and $c_v \approx \frac{\sqrt{\sigma_0^2 + u_0/d}}{\mu_0 + u_0/d}$. p16 expression levels grow significantly with time, (i.e. $\{\mu_0, \sigma_0\} << u_0/d$). This produces the following scaling for older people: $\langle \bar{N} \rangle \approx \frac{u_0}{d}$ and $c_v \approx \frac{1}{\sqrt{\langle \bar{N} \rangle}} << 1$. Thus, this limit cannot produce the large variability seen at later times in the measure p16 levels.

2) In contrast when $u >> d, u_0$, we have $\langle \bar{N} \rangle \approx \mu_0 \frac{u}{d}$ and $c_v \approx \frac{\sqrt{\mu_0 + \sigma_0^2}}{\mu_0} \sim 1$, which is consistent with the variability observed in the measurements of p16 expression. This is the limit for which the best fits to the data have been obtained.

GENERALIZATIONS OF MODEL I

Previously we obtained exact formulas for the first two moments of the down-regulation model with production and degradation rates, u_k and w_k, that depend linearly on the p16 level. This analysis showed that such a model fails to capture both the mean and variability of the experimental data. To prove that this failure is not a result of the linear forms of u_k and w_k, we define a more general class of models and show that within this class, the larger the coefficient of variation (std/mean) the longer it takes to reach saturation. This property makes this class of models inadequate to explain the data.

Steady-state probability distribution and coefficient of variation

Let us assume that u_k and w_k are two arbitrary functions of k, such that $u_k > w_k$ for $k < n_s$, $u_k < w_k$ for $k > n_s$, and $u_{n_s} = w_{n_s}$. This definition ensures that with time the system saturates near the value n_s. Some examples of the functional forms of this class of models are shown in Fig. S2.

First, we find the exact expression for the steady-state distribution:

$P_k = \lim_{t \to \infty} p_k(t)$, where $P_k = 0$ for $k < 0$.

From the Master equation

$$\dot{p}(t) = -(u_k + w_k)p_k(t) + u_{k-1}p_{k-1}(t) + w_{k+1}p_{k+1}(t) \tag{21}$$

we have

$$0 = -(u_k + w_k)P_k + u_{k-1}P_{k-1} + w_{k+1}P_{k+1}.$$

Solving this equation recursively, we find

$$P_1 = \frac{u_0 + w_0}{w_1}P_0 = \left(\frac{u_0 + w_0}{w_0}\right)\frac{w_0 P_0}{w_1},$$

$$P_2 = \frac{u_1 + w_1}{w_2}P_1 - \frac{u_0}{w_2}P_0 = \left(\frac{w_0 w_1 + w_0 u_1 + u_0 u_1}{w_0 w_1}\right)\frac{w_0 P_0}{w_2},$$

$$P_3 = \frac{u_2 + w_2}{w_3}P_2 - \frac{u_1}{w_3}P_1 = \left(\frac{w_0 w_1 w_2 + w_0 w_1 u_2 + w_0 u_1 u_2 + u_0 u_1 u_2}{w_0 w_1 w_2}\right)\frac{w_0 P_0}{w_3}$$

$$\vdots$$

Therefore, for $k = 1, 2, \cdots$, we can write

$$\boxed{P_k = w_0 P_0 \frac{1 + \sum_{m=0}^{k-1}\prod_{i=0}^{m} w_i/u_i}{w_k \prod_{i=0}^{k-1} w_i/u_i}}. \tag{22}$$

Using the normalization condition: $\sum_{k=-\infty}^{\infty} P_k = 1$, we obtain

$$P_0 = \left(1 + w_0 \sum_{k=1}^{\infty} \frac{1 + \sum_{m=0}^{k-1}\Gamma_m}{w_k \Gamma_{k-1}}\right)^{-1}$$

Formally, Eq. (22) can be used to compute values for any steady-state moment $\langle k^m \rangle$, from which the standard deviation is computed as

$$\sigma_s = \sqrt{\sum\left(k^2 P_k\right) - \left(\sum k P_k\right)^2}, \tag{23}$$

It is useful to derive an approximate but simple and explicit expression for the standard deviation. To achieve this goal, we expand u_k and w_k in the vicinity of n_s as

$$u_k = u_{n_s} + (k - n_s)u' + O\left[(k - n_s)^2\right], \text{ where } u' = \left.\frac{du_k}{dk}\right|_{k=n_s} \tag{24}$$

$$w_k = w_{n_s} + (k - n_s)w' + O\left[(k-n_s)^2\right], \text{ where } w' = \left.\frac{dw_k}{dk}\right|_{k=n_s} \text{ and } u_{n_s} = w_{n_s}. \quad (25)$$

In the limit $\sigma_s << n_s$ (i.e., the limit of small fluctuations), we can use the linear approximations given by Eqs. (24) and (25) and Eq. (5) to find that

$$\sigma_s^2 = \frac{w u_0 - u\, w_0}{(w-u)^2} = \frac{w'(u_{n_s} - n_s u') - u'(w_{n_s} - n_s w')}{(w'-u')^2} = \frac{w_{n_s}}{w'-u'}.$$

Therefore, steady state coefficient of variation can be approximated as

$$\boxed{c_v^s = \frac{1}{n_s}\sqrt{\frac{w_{n_s}}{w'-u'}}} \quad (26)$$

This expression is exact when u_k and w_k are linear functions of k, and remains accurate as long as u' and w' are not too close to each other (see examples below).

Time to reach steady state (saturation)

We need to quantify the time it takes to approach steady state. Using Eq.(21) we find that for arbitrary u_k and w_k the time-dependent mean $q(t)$ satisfies the equation

$$\begin{aligned}\dot{q}(t) &= \sum_{k=-\infty}^{\infty} k\, \dot{p}_k(t) = -\sum_{k=-\infty}^{\infty} k\,(u_k + w_k)p_k(t) + \sum_{k=-\infty}^{\infty} k\, u_{k-1} p_{k-1}(t) + \sum_{k=-\infty}^{\infty} k\; w_{k+1} p_{k+1}(t) \\ &= -\sum_{k=-\infty}^{\infty} k\,(u_k + w_k)p_k(t) + \sum_{k=-\infty}^{\infty}(k+1)u_k\, p_k(t) + \sum_{k=-\infty}^{\infty}(k-1)\, w_k\, p_k(t) \\ &= \sum_{k=-\infty}^{\infty} u_k\, p_k(t) - \sum_{k=-\infty}^{\infty} w_k\, p_k(t)\end{aligned} \quad (27)$$

For cases in which the probabilities $p_k(t)$ are sharply peaked around the mean value $q(t)$, Eq.(27) can be approximated as

$$\frac{dq}{dt} = \sum_{k=-\infty}^{\infty} u_k\, p_k(t) - \sum_{k=-\infty}^{\infty} w_k\, p_k(t) \approx \left.(u_k - w_k)\right|_{k=q(t)}$$

Therefore, we estimate the mean time to reach a certain mean value $\langle N \rangle$ as

$$\boxed{t(\langle N \rangle) = \int_0^{\langle N \rangle} \frac{dq}{u_q - w_q}}. \quad (28)$$

As before, this estimation is exact if u_k and w_k are linear functions of k and is valid as long as the coefficient of variation is relatively small.

Examples:

The following examples demonstrate quantitatively that to obtain a wide distribution ($c_v^s \sim 1$) at steady-state, u_k and w_k must be close to each other for a range of *k*'s, but in this case because of the slow drift $(u_k - w_k)$, the time to reach saturation is large. Therefore, this class of models fails to satisfy simultaneously the following two properties of the data: $c_v^s \sim 1$ and $t(\langle N\rangle \rightarrow n_s) < 80$ years old.

For each model, we determine parameters that can be tuned to achieve large values of the coefficient of variation and vary them to explore the model's behavior. In particular, we plot three representative values of the following quantities:

A. The forward drift, $u_k - w_k$.
B. The time to reach a mean value $\langle N\rangle$ as a function of $\langle N\rangle / n_s$ according to Eq. (28).
C. The steady-state coefficient of variation, $c_v^s = \sigma_s / n_s$, as a function of an appropriate parameter, both according to the exact formulas (Eqs. 22 and 23) and the approximate expression given by Eq. (26).
D. The steady-state distributions.

Because the expression level of p16 must not be negative, for all models we assume that $w_0 = 0$. We also make sure that u_k and w_k are written in a form consistent with the condition $u_{n_s} = w_{n_s}$.

Model-I:

$w_k = wk$, so that $w' = w$,
$u_k = (w-u)n_s + uk$, so that $u' = u$.

$$c_v^s = \frac{1}{n_s}\sqrt{\frac{wn_s}{w-u}} = \frac{1}{\sqrt{n_s}}\sqrt{\frac{1}{1-u/w}} \quad . \tag{29}$$

In this case, the steady-state coefficient of variation is a function of the ratio, u/w. Therefore, The steady-state distribution is broad, when this ratio is close to unity.

Model-Ia:

$w_k = \dfrac{wk^h}{n_s^h + k^h}$, so that $w' = \dfrac{wh}{4n_s}$,

$u_k = u_0 + uk = u_0 + \dfrac{k}{n_s}\left(\dfrac{w}{2} - u_0\right)$, so that $u' = \dfrac{1}{n_s}\left(\dfrac{w}{2} - u_0\right)$.

$$c_v^s = \frac{1}{n_s}\sqrt{\frac{\frac{w}{2}}{\frac{wh}{4n_s} - \frac{1}{n_s}\left(\frac{w}{2} - u_0\right)}} = \frac{1}{\sqrt{n_s}}\sqrt{\frac{2}{h - 2 + 4u_0/w}}$$

In this case, the steady-state coefficient of variation depends on both the Hill coefficient, h, and the ratio, u_0/w. The steady-state distribution is wide, when the Hill coefficient approaches the value, $2 - 4u_0/w$. For a fixed h, c_v^s is maximal when $u_0 = 0$.

Model-Ib:

$w_k = \dfrac{wk^h}{n_s^h + k^h}$, so that $w' = \dfrac{wh}{4n_s}$,

$u_k = w\left[1 - \exp\left(-\dfrac{k\ln 2}{n_s}\right)\right]$, so that $u' = \dfrac{w\ln 2}{2n_s}$.

$$c_v^s = \frac{1}{n_s}\sqrt{\frac{\frac{w}{2}}{\frac{wh}{4n_s} - \frac{w\ln 2}{2n_s}}} = \frac{1}{\sqrt{n_s}}\sqrt{\frac{2}{h - 2\ln 2}}$$

Here the steady-state coefficient of variation only depends on the Hill coefficient, h, so that the steady-state distribution becomes wider, as the Hill coefficient approaches $2\ln 2$.

Model-Ic:

$w_k = \dfrac{wk^{h_1}}{n_s^{h_1} + k^{h_1}}$, so that $w' = \dfrac{wh_1}{4n_s}$,

$u_k = \dfrac{w-u}{2} + \dfrac{u\,k^{h_2}}{n_s^{h_2} + k^{h_2}}$, so that $u' = \dfrac{u\,h_2}{4n_s}$.

$$c_v^s = \frac{1}{n_s}\sqrt{\frac{\frac{w}{2}}{\frac{wh_1}{4n_s} - \frac{u\,h_2}{4n_s}}} = \frac{1}{\sqrt{n_s}}\sqrt{\frac{2}{h_1 - h_2\,u/w}}$$

In this case, the steady-state coefficient of variation depends on the ratio, u/w, and the difference between the Hill coefficients, h_1 and h_2. The steady-state distribution is wide, when h_1/h_2 is close to u/w. For fixed Hill functions, say $h_1 = h_2$, c_v^s grows as u/w approaches to the unity.

Model-Id:

$w_k = wk = \dfrac{uk}{2n_s}$, so that $w' = \dfrac{u}{2n_s}$,

$u_k = \dfrac{u n_s^h}{n_s^h + k^h}$, so that $u' = -\dfrac{uh}{4n_s}$.

$$c_v^s = \frac{1}{n_s}\sqrt{\frac{\frac{u}{2}}{\frac{u}{2n_s} + \frac{uh}{4n_s}}} = \frac{1}{\sqrt{n_s}}\sqrt{\frac{2}{h+2}}$$

For this type of model $w' > 0$ while $u' < 0$, and hence, c_v^s is bounded from above. Here, for the Hill functions with $h > 1$, $c_v^s < \sqrt{2/3n_s} \approx 0.067$. Therefore, this model Id, as well as its variations Ie and If, are even less plausible than the ones considered previously.

SKEWNESS OF THE DISTRIBUTION

In this section we explore the skewness of the p16 distributions as a function of age for both the experimental data and Model II. The skewness is defined as

$$\text{skewness} = \frac{\left\langle \left(N - \langle N \rangle\right)^3 \right\rangle}{\left(\langle N^2 \rangle - \langle N \rangle^2\right)^{3/2}}.$$

Unfortunately, the number of data points is too small to draw any conclusions based on a standard binning of the data. For bins of length 4 years we have on average roughly 10 data points per bin. Although 10 points is a reasonable number to estimate the mean and variance of the data, it is not sufficient to accurately estimate the overall distribution. In Fig. S8 we show 2D histograms (age × p16) of the data for both linear and $\log_2$ scales with bins of size 4×20 and 4×0.5, respectively. Binned in this way the data have skewness values that range from –1 to +2.5 on the linear scale and from –1 to +1 on the log scale.

To better estimate the distribution and its skewness we used a moving bin method analogous to the procedure used to estimate the mean values. In particular, we count data points falling within a box of large size, 20×100, and move the box in small steps, 4 and 20, along the 'age' and 'p16' axes. As a result the number of data points collected for each increment significantly increased and the 2D distribution appears much smoother.

Figs. S9 and S10 that show that:

A. On a linear scale the data have high positive skewness varying between 1.0 and 2.5 with the distribution maximum less than the statistical mean.
B. On a $\log_2$ scale the data are more symmetric with the skewness ranging from 0.0 to 0.7.

Fig. S11 shows contour plots of the 2D distribution for Model II. To qualitatively compare with the distribution estimated from the data, we also plot the distributions at fixed ages of 20, 40 and 60 years.

Because the distribution for the log of a random variable is not equal to the log of the distribution, we used the model distribution to generate a large number ($\sim 10^7$) of 'theoretical' data points, and then plotted the distributions of log of these numbers. The skewness of the theoretical distribution at various time points for both linear and $\log_2$ scales are shown in Fig. S12.

Taking into account the limited amount of data for p16 levels, the predicted distributions from Model II are in good agreement with the data.