Aging clocks, entropy, and the limits of age-reversal

Andrei E. Tarkhov,[1,2,3] Kirill A. Denisov,[1,4] and Peter O. Fedichev[1,3,*]

[1] *Gero PTE, 60 Paya Lebar Road # 05-40B, Paya Lebar Square, 409051, Singapore*
[2] *Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, Moscow, Russia, 121205*
[3] *Moscow Institute of Physics and Technology, 141700, Institutskii per. 9, Dolgoprudny, Moscow Region, Russia*
[4] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 1-73 Leninskie Gory, Moscow, Russia, 119991*

We analyze aging signatures in the DNA-methylation and Electronic Medical Records from the UK Biobank datasets and observe that aging is driven by a large number of individually rare and independent transitions between metastable states in a vast configuration space. The compound effect of the configuration changes can be captured by a single stochastic variable, the thermodynamic biological age (tBA), tracking the entropy produced and hence the information lost in the aging process. We show that tBA increases with age, causes the linear aging drift of physiological state variables, reduces resilience, and drives the exponential acceleration of the risks of chronic diseases and death. The entropic character of the aging drift sets limits on possibilities of age-reversal. However, the universal features of configuration transitions suggest practical ways of controlling the rate of aging and thus promising the strongest possible life-extension effects.

I. INTRODUCTION

Aging is a complex process manifesting itself across different organismal levels (see hallmarks of aging [1]) and leading to the exponential acceleration of incidence of chronic diseases [2] and mortality [3]. It is both practically and intellectually appealing to reduce the effects of the multitude of phenotypic changes to a few, or even better a single actionable indicator, most commonly referred to as "biological age". Biological age (BA) models can be trained either to predict chronological age or mortality risks of an individual from different sources of biomedical data, ranging from DNA methylation [4–11] to physical activity records from wearable devices [12, 13]. The excessive BA (or biological age acceleration, BAA) is associated with all-cause mortality, the prevalence, future incidence, and severity of chronic [14–16] and transient diseases, including COVID-19 [13, 17, 18]. Hence, BA predictors have been increasingly gaining traction in clinical trials [19–21].

The relation between the variation of BA and aging is not entirely clear. For example, epigenetic aging drift and hence the methylation age increase may happen without an appreciable increase in all-cause mortality in negligible senescent species [22, 23]. Moreover, even in the most healthy individuals, BA levels can transiently change in response to stress factors and lifestyle choices such as smoking [16, 24], The recovery rate of the BA fluctuations progressively decreases as a function of age [16]. The number of individuals exhibiting slow recovery increases exponentially and doubles approximately every 8 years, which is close to the mortality doubling time for humans [13]. Further applications of BA models in aging research and medicine would require a better understanding of the dynamics and causal relation between underlying biological and physiological variations of the organism state captured by various BA indicators, on one hand, and mortality, prevalence, and severity of diseases and effects of medical interventions, on the other.

To address these fundamental questions, we reviewed the common features of aging signatures in biomedical data. We performed the principal component analysis (PCA) in a large cross-sectional white-blood-cells DNA methylation dataset [25] and in the longitudinal Electronic Medical Records (EMRs) from UK Biobank [26]. In both cases, we observed a large number of rare transitions between the respective states representing the methylation of individual sites or the incidence of specific diseases in the course of the aging process. At the same time, most of the variance in the data could be explained by the stochastic evolution of a single factor linearly increasing with age and demonstrating the strongest correlation to the Horvath's methylation age and the number of chronic diseases in the DNA-methylation and EMR datasets, respectively.

To explain the dynamic properties of the aging signatures, we put forward a semi-quantitative model of aging in a complex regulatory network. In the spirit of Hayflick's proposal [27], we assumed that living organisms are the collections of a vast number of interacting physiological units, each set to some metastable state at the end of development. Aging then results from the stochastic relaxation of an organism state towards equilibrium via a sequence of activation transitions between the metastable microscopic states in a very high-dimensional configuration space. If the number of activations into irregular configurations is sufficiently large, then the total number of configuration transitions is a stochastic variable with linearly increasing mean, variance, and possessing most expected properties of the biological age. Such a BA measure, henceforth referred to

* Correspondence and requests for materials should be addressed to P.O.F. email peter.fedichev@gero.ai

as the thermodynamic BA or tBA, is closely related to the dominant principal component (PC) score and the configuration entropy attained in the course of the aging process. The fluctuations of the organism state in humans are dynamically stable and quickly adjust to the slowly changing biological context defined by the increasing BA. Eventually, at the end of life, the dynamic stability of the organism state fluctuations is lost, which leads to the development of most dangerous diseases and death. We demonstrate, that most of the age-related physiological changes in humans are entropic and can indeed be captured by the dynamics of a properly defined BA, also driving the mortality rate acceleration. If the hypothesis is correct, then the a large part of age-related changes is thermodynamically irreversible. We predict, however, that most configuration transitions responsible for the increase in tBA share universal features, which could be exploited for future medical interventions controlling the rate of aging and thus providing non-incremental extension of human healthspan.

II. RESULTS

A. Aging and the principal component analysis of biomedical data

We start by analyzing age-dependent features in a white-blood-cells DNA-methylation dataset, GSE87571 [25]. The reported methylation level $\bar{\sigma}_i$ of a specific CpG site i in a given sample is the average of the binary "methylation on/off" variable σ_i over a vast number of cells. Hence $\bar{\sigma}_i$ is the probability of observing a particular CpG site methylated. Most of the significantly correlated with age DNA-methylation sites were initially "polarized" (or methylated, $\bar{\sigma}_i > 0.9$, 26% of all sites in the age bin $20-25$) or de-polarized ($\bar{\sigma}_i < 0.1$, 28% of the sites).

The number of CpG sites in each sample (approximately 450k) is very large compared to the number of patients (samples) in the dataset (less than 800) that leads to the so-called "curse of dimensionality" [28]. Hence, to improve the stability and focus the principal component analysis (PCA) on aging, we reduced the number of features by selecting the sites significantly correlated with age (after Bonferroni correction for multiple comparisons at the level of $p = 0.005/\text{total number of methylation sites in the signal}$). According to this definition, approximately 100k (almost 25% of all reported) sites were significantly correlated with age.

Due to the binary character of the methylation signal and following the analogy from statistical physics (see Section II B), we turned to the analysis of the convenient proxy variables, the "regulatory fields" h_i, controlling the probability of observing the average methylation level of a site i according to the Boltzmann distribution: $h_i \sim \ln(\bar{\sigma}_i/(1-\bar{\sigma}_i))$.

The PCA of the regulatory fields reveals a few principal components (PCs) associated with age. The dominant, the first PC (DNAm-PC1), evolves approximately lin-

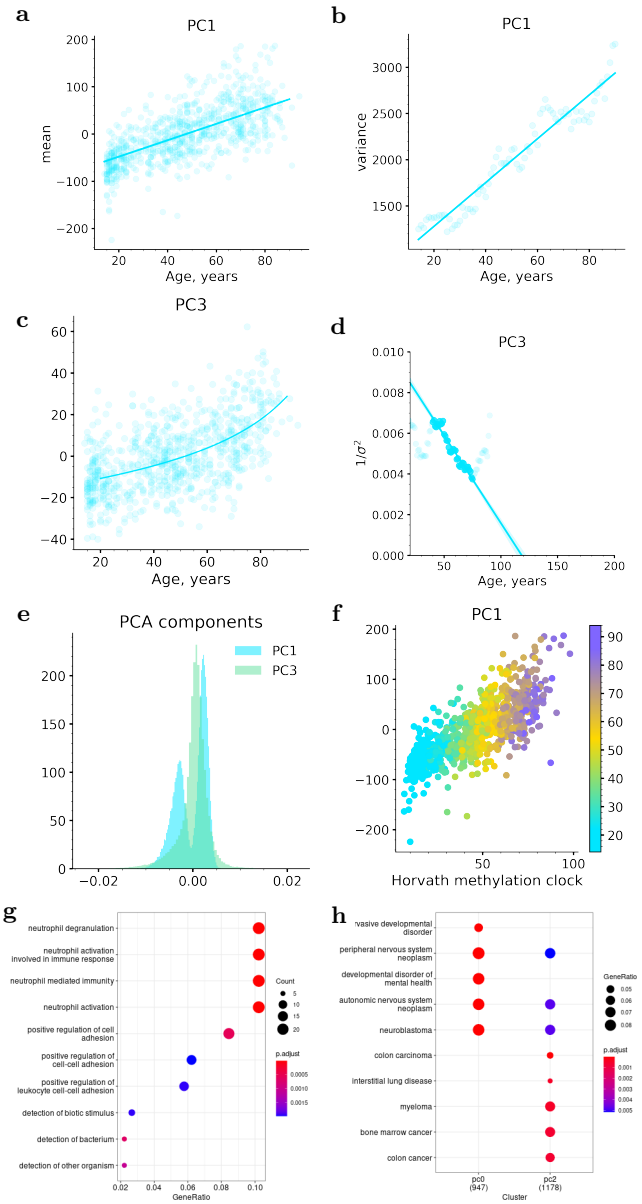

FIG. 1. PCA of age-dependent methylation profiles from white-blood-cells samples from GSE87571. (a) DNAm-PC1 and (b) the variance of DNAm-PC1 increase, on average, linearly as the functions of age. (c) DNAm-PC3 increases faster than linearly with age; (d) the inverse variance of DNAm-PC3, the extrapolation for the range of 40 to 75 years vanishes at 120 y.o. (e) The distribution of PC vector components. (f) The DNAm-PC1 scores correlate with Horvath's methylation clock. The color-bar represents the chronological age of the patients. (g) Gene set enrichment analysis of CpG sites best associated with DNAm-PC3 suggests involvement in regulating genes associated with the innate immune response. (h) Methylation profiles driven by DNAm-PC1 and DNAm-PC3 can be associated with the clusters of developmental and mental diseases, and internal organs' diseases.

early as the function of age of the patients (Fig. 1a; Pearson's $r = 0.68$, $p = 3 \cdot 10^{-98}$). The dynamics of DNAm-PC1 are most probably stochastic since the variance of DNAm-PC1 also increases linearly with age (Fig. 1b).

Aside from DNAm-PC1, the best correlation with the chronological age was produced by the third PC (DNAm-PC3). The corresponding Pearson's correlation coefficient was $r = 0.56$ ($p = 3 \cdot 10^{-62}$). DNAm-PC3 in subsequent age-adjusted bins increased faster than at a linear pace as the function of age (Fig. 1c). The variance of DNAm-PC3 also increased, also faster than linearly so that the inverse variance of signal decreased approximately linearly in the patients older than 40 years old (Fig. 1d). By extrapolation, the inverse variance of DNAm-PC3 would approach zero (and hence the variance would diverge) at some age within the range of 120–150 years.

The loading vectors corresponding to DNAm-PC1 and -PC3 describe two distinct methylation profile changes with age. The distribution function (the histogram) of the PC1 loading vector components is non-Gaussian and bi-modal. Hence the dominant aging signature in DNA methylation data consists of the two large groups of methylation sites (Fig. 1e) changing their methylation status ("polarization") with age in the opposite directions. The first PC score is then proportional to the total number of the polarization transitions.

On the contrary, the distribution of the loading vector components from DNAm-PC3 has a single peak and clear leading contributions from non-Gaussian tails (see Fig. 1e). Gene set analysis of methylation regions best associated with the PC3 variation reveals pathways involved in innate immunity and cancer (Figs. 1g and h).

The dominant PC scores computed from the GSE87571 samples demonstrated the best correlation with Horvarth's methylation clock from [4] (Fig. 1f). The corresponding Pearson's correlation coefficients were $r = 0.75$ ($p = 2 \cdot 10^{-131}$) and $r = 0.52$ ($p = 10^{-52}$) for DNAm-PC1 and DNAm-PC3, respectively (see also Figs. S.1 and S.2 for the summary of the correlations of the PC scores to the chronological age and to Horvath's methylation age, respectively).

To confirm the stochastic character of the dominant aging signature in humans, one would need to analyze a large longitudinal dataset. We did not have access to a high-quality set of longitudinal DNA-methylation measurements. Instead, we turned to an extensive Electronic Medical Records (EMRs) collection from UK Biobank. Irrespective of the age at the first assessment, the EMRs provided information on the prevalence of chronic diseases from birth untill the end of the follow-up (slightly more than ten years, on average). We represented each of UK Biobank patients by a vector of the respective binary variables indicating the presence or absence of a disease (see Section VI C).

Most of UK Biobank's subjects are healthy early in life. Hence, the states representing the presence of the diseases are initially polarized ($\sigma_i = 0$). The majority

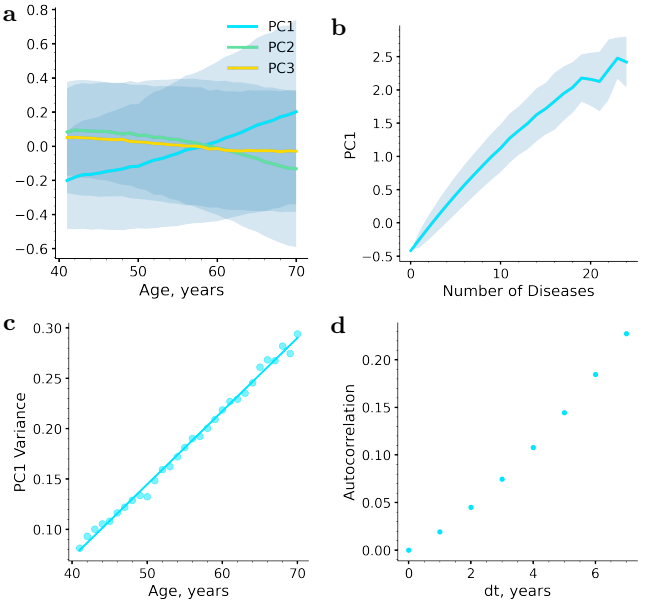

FIG. 2. PCA of disease-state vectors from the Electronic Medical Records (EMR) from UK Biobank. (a) Age dependence of the first three PC scores. (b) PC1 correlates with the total number of chronic diseases. (c) The variance of PC1 increases linearly with age (d) The autocorrelation function of PC1 increases linearly as a function of the time lag, $dt = 1$ year.

of the states stay polarized for life, with only a small fraction of patients exhibiting the depolarization transitions leading to the incidence of specific diseases. Indeed, chronic diseases are individually rare: the most prevalent diseases are metabolic disorders (15%), joint disorders (14%), and arthrosis (12%).

PCA of the binary-valued vectors representing the health state according to EMRs of UKB subjects at the time of the first assessment look pretty similar to the PCA results from the white-blood-cells DNA methylation study above. This time, we observed only two PCs significantly associated with age (see the blue and the green lines and the respective ranges marking one standard deviations in Fig. 2a).

The dominant aging signature, the first PC in the UKB EMR data (EMR-PC1), evolves approximately linearly as a function of age and is linearly associated with the total number of the diagnosed diseases (Fig. 2b). Hence, in line with the results of our DNA-methylation analysis above, the first PC correlates with the total number of depolarization transitions (this time being equal to the disease burden at the time of the measurement).

As expected, the variance of EMR-PC1 increased linearly with age (Fig. 2c), which is a hallmark of a stochastic process. This time, however, due to the longitudinal nature of the dataset of EMRs, we could make a stronger statement by computing the autocorrelation function of EMR-PC1. We observed that the autocorrelator increases linearly as the function of the time lag

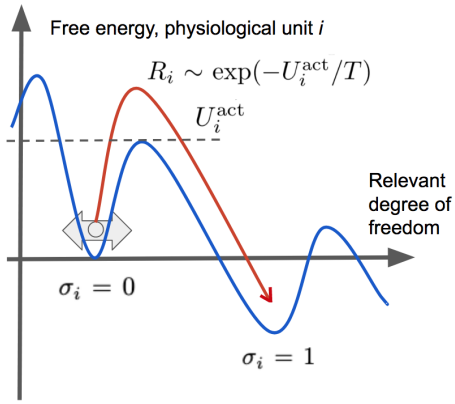

FIG. 3. Schematic representation of relaxation dynamics of a physiological unit (PU) i. The subsequent metastable states are labeled by the "polarization" $\sigma_i = 0$ and $\sigma_i = 1$. The initial, "polarized" state is protected by the activation barrier characterized by the activation energy $U_i^{\rm act}$. The configuration transition rates R_i are presumed small and depend exponentially on the effective temperature T.

between the observations, which is typical of a solution of a diffusion equation (Fig. 2d, see the discussion in the section below).

B. Aging in a complex regulatory network

To explain the key features of aging signatures, let us think of an organism as a collection of interacting physiological units (PU). Each of the units can be observed in multiple states of varying physiological capacity. We will focus on the two most important configurations at any given moment (Fig. 3). These may be the most prevalent state in a population at a young age, on the one hand, and a state corresponding to the closest adjacent potential well in the free energy landscape shaped by regulatory interactions, on the other hand.

So far, we have analyzed the transitions among the methylation and among the disease states. However, there must be other examples of configuration transitions, and hence, for the sake of generality, we do not specify the nature of the PUs at this time. We encode the pair of the most relevant states of a specific PU i by a binary variable taking values of $\sigma_i = 0$ and $\sigma_i = 1$, respectively. It is important to understand, that the total number of PUs comprising the organism is practically infinite and hence, most of the PUs can not be in principle observed directly in any given experiment.

We propose to model the effect of the interactions among PUs with the help of the auxiliary variables – the effective "regulatory fields" h_i evolving over time according to

$$\dot{h}_i = \sum_j k_{ij}\sigma_j + \sum_{jk} g_{ijk}\sigma_j\sigma_k + J_i^{(0)} + f_i, \qquad (1)$$

where k_{ij} and g_{ikj} describe the first linear and first order non-linear interaction between the individual units. The force terms $J_i^{(0)}$ and f_i represent the effects of constant (such as smoking or diets) and stochastic (social status, deleteriousness of the environment [29]) factors, respectively. For simplicity, we assume that the noise factors have zero mean and are not correlated over time (see Section VII). The states of individual PUs i can be observed depending on the regulatory field h_i according to the Boltzmann distribution: $\bar{\sigma}_i = (1 + \exp(-h_i/T))^{-1}$, where T is the effective temperature. We note that the human organism is a non-equilibrium system, hence, the effective temperature in this presentation shall not be confused the body or environmental temperature. Instead, T depends on the statistical properties of the noise term f_i in Eq. 1 (Methods VII), and measures the level of deleteriousness of the environment [29].

Most PUs are initially (i.e., at some age t_0 roughly corresponding to the end of the development) polarized. However, the polarized states are not necessarily ground states and, hence, are metastable. The data suggests, that in most cases, the configuration transitions are rare (fewer than one transition between the states occurs over the lifetime $\bar{t}$ of the organism). This means, that transition rates R_i are small ($R_i\bar{t} \ll 1$). Therefore, we expect that the activation energy $U_i^{\rm act}$ characterizing the transition greatly exceeds the effective temperature so that the depolarization rates, $R_i \sim \exp(-U_i^{\rm act}/T)$, are exponentially small (Fig. 3).

Stochastic forces cause "jumps" towards the equilibrium, the probability of the depolarization of every PU $\Delta\sigma_i = \bar{\sigma}_i(t) - \bar{\sigma}_i(t_0)$ changes at a small, but, most crucially, age-independent rate R_i, $\langle\Delta\sigma_i(t)\rangle = \epsilon_i R_i(t-t_0) \ll 1$, where $\epsilon_i = \pm 1$ is the direction of the depolarization transition (fully defined by the initial state $\bar{\sigma}_i(t_0)$). Accordingly, the average depolarization in samples of approximately the same age t is

$$\langle\sigma_i\rangle = \epsilon_i\langle R_i(t-t_0)\rangle \approx \epsilon_i Z/N, \qquad (2)$$

where $Z \approx NR(t-t_0)$ is the total number of configuration transitions, R is the characteristic (average) depolarization rate, and N is the total number of PUs. For simplicity, henceforth we count the age from the end of the development and hence drop t_0.

The PCA of a dataset modelled by Eq. 2, would produce the first PC, which is directly proportional to the total number of the configuration (depolarization) transitions Z (which is itself proportional to the age, t). This is exactly what happens in the PCA of both the DNA methylation (Fig. 1a) and EMR data (Fig. 2a) above. The distribution of the first PC vector components would be bi-modal according to the corresponding values of depolarization transition directions ϵ_i (Fig. 1e).

If the total number of the configuration transitions Z has a chance to grow large, Z becomes a random quantity, obeying a diffusion equation (with a drift). This means that the variance of Z (and hence the first PC in the data) in age-adjusted bins should increase linearly with

age. This was indeed observed in the DNA methylation (Fig. 1b) and EMR (Fig. 2c) data, respectively.

More evidence in favor of stochastic character of Z (and hence of the first PC scores in the data) could be produced by the investigation of the autocorrelation function $C(\tau) = \langle (Z(t+\tau) - Z(t))^2 \rangle \sim \tau$ (here $\langle ... \rangle$ stands for the averaging first along the individual trajectory and then over all the patients). The autocorrelation function of the leading PC in the EMR data increased linearly as a function of the time lag in the range between 2 and 10 years (Fig. 2d). The diffusion coefficients estimates from the variance and autocorrelation increase turned out to be close: 0.012 and 0.009 per year, respectively, thus confirming the association of the leading PC score with the increasing number of configuration transitions Z.

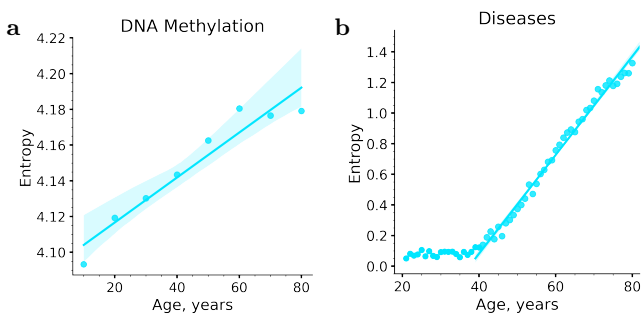

FIG. 4. (a) Configuration entropy computed from the distributions of the DNA methylation and (b) EMRs features in subsequent age-matched cohorts increases linearly as a function of age after 40 years.

To highlight the stochastic character of the aging drift in the model, let us note the relation between the number of depolarization transitions Z and the configuration entropy in the aging organism. Assuming that we start from highly polarized states, $\bar{\sigma}_i \approx 1 - R_i t$, we find that

$$S = \sum_i S_i = N \langle R_i t \log(1/(R_i t)) \rangle \sim Z$$

up to a proportionality coefficient. As expected, the configuration entropy increases along with the number of the depolarization transitions understood as DNAm changes or incidence of chronic diseases or age (Figs. 4a and b, respectively).

The probability of depolarization for each PU is small, but the total number of PUs available for the configuration transitions is very large. Hence, if enough time is available in the course of life of an organism, the combined effect of the configuration transitions on the whole organism, and each of the PUs does not need to be small. Most importantly, the shift of the regulatory fields is a result of a large number of contributions from each of the already depolarized units (see Eq. 1). In this case, the central limit theorem [30] ensures that the net effect is proportional to the total number of the configuration transitions Z. Hence, the effect of aging on each of the units can be modeled by a "mean-field" proportional to Z.

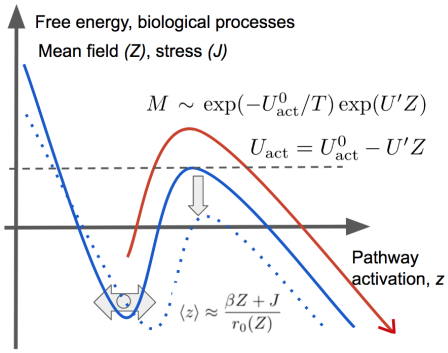

FIG. 5. Aging drift (Z) and stress (J) modify the shape of the free energy surface controlling the pathway (the solid and the dashed blue lines). This leads to pathway activations $\langle z \rangle$ and alteration of the activation energy of the barrier separating the dynamically stable (healthy, left) and unstable (disease, right) states. The mean-field effect of the aging drift $Z \sim t$ drives the exponential acceleration of the incidence of diseases with age by lowering the activation threshold.

The mean-field produces the most significant effects in physiological compartments (modules or pathways) – large clusters of correlated PUs. Mathematically, this can be seen by expanding the solutions of Eq. 1 in the vicinity of a metastable state and representing the solution as $\delta h_i \approx \sum_A z_A b_i^A$, where z_A are the pathway activations, and b_i^A are the right eigenvectors of the interaction matrix corresponding to the smallest eigenvalues r_A (see Section VII). The components of the vector b_i^A characterize the participation of the PU i in the pathway A.

The quantities r_A have the meaning of the recovery rates and characterize the resilience, which is the time required for the organism state to recover from a perturbation affecting the relevant cluster of PUs [16]. The mean pathway activations averaged over time scales greatly exceeding the equilibration times, $\sim 1/r_A$, follow the effects of aging (Z) and the stress (J):

$$\langle z_A \rangle \approx \frac{\beta_A Z + J_A}{r_A(Z)}, \quad \sigma_z^2 \sim \frac{B}{r_A(Z)}. \tag{3}$$

Both quantities are inversely proportional to the age-adjusted recovery rate, $r_A(Z) = r_A - r'_A Z$, where J_A, β_A, and r'_A are pathway-specific parameters depending on the regulatory interactions and the nonlinearity in Eq. 1.

Eq. 3 explain why small fluctuations of the organism state are dominated by the dynamics of clusters of PUs characterized by the lowest recovery rates. The cluster participation vectors b_i^A and the pathway activation variables z_A should approximately coincide with the leading PC loading vectors and scores, respectively.

The collective variables z_A experience stochastic fluctuations around the mean value of $\langle z_A \rangle$ in the potential well that is shaped by the regulatory interactions (Fig. 5). Aging in the form of the increasing number of depolariza-

tion transitions Z induces progressive shifts of the pathway activation variables precisely in the same way as any constant stress J. More subtly, aging affects the recovery rate (see the denominator in Eq. 3), and hence the pathway activations may depend on age in a non-linear fashion.

If the recovery rate is sufficiently small, the denominator characterizing the cluster of features with large fluctuations (and hence corresponding to one of the leading PC in the data) in Eq. 3 may vanish: $r(Z) = r_0 - r'Z = r_0(1 - t/t_{\max})$ at some point late in life at the age $t_{\max}$. This is the critical point corresponding to the loss of resilience (inability of the system to retain its homeostasis equilibrium, incompatible with survival [16]). In cross-sectional data, we may expect both the mean pathway activations and their variance in subsequent age-matched bins to diverge at the same age.

Our analysis of the DNA methylation dataset produced one of the leading PCs, DNAm-PC3, best associated with the innate immunity function and increasing at a faster than linear pace as a function of the chronological age. The fit of the DNAm-PC3 scores to the singular solution for the average z_A from Eq. 3 gives $t_{\max} \approx 130$ years (see the solid line in Fig. 1c and Section VI A for the details of the calculations). The extrapolation suggests that the inverse variance of DNAm-PC3 hits zero at about the same age (see the solid line in Fig. 1d). The results of the two extrapolations are comfortably close. Hence, our calculations support the existence of a critical point at some age in the range of $100 - 150$ years from birth.

III. DISCUSSION

We put forward a semi-quantitative model of aging in a complex regulatory network and applied it to the analysis of aging signatures in humans in a cross-sectional white-blood-cells DNA methylation dataset [25] and the extensive collection of longitudinal Electronic Medical Records (EMRs) from UK Biobank [26]. We demonstrated that the dynamics of physiological indices and hence the organism state is driven by massive and thermodynamically irreversible configuration changes accompanied by entropy increase.

First, we observed that rates characterizing the transitions among microscopic states different by the methylation status or the incidence of specific diseases are small. In most cases, fewer than a single transition occurs on average throughout the lifetime. However, even though the rates of the configuration changes may be low, the total number of the configuration transitions between the states is vast: no less than 25% methylation sites exhibited age-related dynamics. Hence, the overall number of concurrently occurring transitions is large, and the massive uncorrelated changes together dominate the dynamics of the physiological state.

We observed that the dominant aging signature, which is the first PC score explaining most variance in the data, increased on average linearly with age in the PCA of the DNA-methylation and EMR data. In both cases, the first PC score was proportional to the total number of the configuration transitions (the number of DNAm level changes or the total number of the chronic diseases). Simultaneously, the variance of the dominant PC score grew linearly with age in both datasets, as is expected for a stochastic quantity, a product of a large number of independent relaxation transitions.

Thus, the total number of the observable configuration transitions Z, on average, increases linearly with age and explains most of the variance of the physiological variables. Accordingly, we propose using Z as the quantitative measure of the compound effect of the slow configuration changes in the aging organism – the thermodynamic biological age (tBA $\sim Z$). The dominant aging signature in the data (the first PC score) is then an estimate of tBA from the specific data. Most comfortably, DNAm-PC1 and hence tBA exhibited the strongest correlation to Horvath's methylation age.

Configuration transitions provide the thermodynamic arrow of time in an aging organism. We demonstrated, that tBA is directly related to the configuration entropy produced (or the information regarding the healthy state lost) in the course of aging. The association can be understood since the biological age is a single number capturing the result of individual transitions characterized by different transition rates depending on the activation energy of individual physiological units and the effective temperature T. The particular depolarization patterns (configuration) changes may differ in various cells or subjects. In contrast, the total number of transitions should be close and hence characterize the overall state of the organism in relation to aging.

Due to the data availability, we could only analyze aging signatures in DNA-methylation states and chronic diseases. However, there must be innumerable examples of physiological units experiencing configuration changes over time. We may think about (but not limit ourselves to) conformational or chemical modifications of macromolecules, including DNA damage, etc. Since all the configuration changes happen simultaneously and increment tBA, we must not consider any single kind of them as causing the aging drift.

The configuration transitions change the organism's state and affect all other biological processes. Since the number of configuration changes is large, the details of the individual transitions are not important. The compound effect of the aging drift manifests itself as a "mean-field" causing the shift of physiological indices that is proportional to the number of the configuration transitions to date (and hence to tBA itself).

The mean-field produces the strongest effects on large clusters of physiological units characterized by long recovery times and hence exhibiting strong fluctuations (hence dominating the leading PCs other than PC1 $\sim$ tBA in biological signals). We show that the effect of aging drift on such modules or pathways is indistinguishable

from effects of stress (such as smoking or diet). Since the BA increases linearly with age, we expect all pathways to "follow" the aging process by increasing (or decreasing) activations approximately linearly with age.

The total number of transitions is large, and hence tBA increases linearly with age to a very high degree of accuracy (according to the central limit theorem [30]). That may be the mathematical reason why it is almost always possible to build a very accurate predictor of chronological age from different sources of biomedical data (see examples [4, 5, 7]).

Evidence suggests that the mean-field increases enough to let the non-linearity of regulatory interactions produce significant deviations of the mean pathway activations from a simple linear dependence as the functions of age (see DNAm-PC3 and EMR-PC2 in the PCA of DNA methylation and EMR data, respectively).

Nonlinear regulatory interactions may also alter the recovery times. If the recovery rate is particularly small, this may lead to divergence of the organism state fluctuations at some advanced age corresponding to the critical point, where the recovery rate vanishes. This happens with the cluster of DNA methylation features associated with the DNAm-PC3. By extrapolation, we observed both the mean and the variance of DNAm-PC3 diverging at the age close to $t_{\max} \approx 130$.

Gene set enrichment analysis (GSEA) of genes regulated by CpG sites involved in DNAm-PC3 reveals the association with innate immunity. Recently, we demonstrated that linear log-mortality predictors built from complete blood counts (CBC) and physical activity [16] also exhibited diverging fluctuations and vanishing recovery rate at about the same limiting age $t_{\max} \approx 130$. We, therefore, infer that the white-blood-cells DNA methylation, blood composition and even physical activity all substantially depend on a common factor related to innate immunity and all-cause mortality.

In reality, the disintegration of the organism state happens well before reaching the criticality at the limiting age $t_{\max}$. Stress factors and the mean-field Z do not merely shift the mean pathway activation levels. Both factors may also decrease the activation energy separating the organism state from the disintegration and death (Fig. 5). In the linear regime, the activation energy linearly depends on the mean-field, $U_{\text{act}} = U^0_{\text{act}} - U'Z$, and hence the probability of the barrier crossing per unit time, the mortality in the model, is $M \sim \exp(-U^0_{\text{act}}/T)\exp(U'Z/T)$. Thus we explain how the aging drift in the form of the massive configuration transitions registered by the tBA can drive the exponential acceleration of all-cause mortality with age: $M \sim \exp(\Gamma t)$. The mortality doubling rate Γ in the model depends on the details of the regulatory interactions (through U'), the rate of the aging drift (through Z) and on the effective temperature T.

The prediction of mortality (or the remaining lifespan) in humans hence requires an estimate for tBA $\sim Z$ and for a few most crucial pathway activations (also, on average, proportional to Z). Hence, no single biological age measure fully describes longevity in humans. We expect that the biological age models trained to predict the chronological age from the data should yield better estimates of tBA. On the other hand, the models trained to predict the remaining lifespan (such as PhenoAge [8], GrimAge [15], DOSI [16], etc.) should return a combination of the pathway activations associated with prevalence of diseases and accelerated mortality [31] and hence better suited for the detection of reversible effects of diseases, lifestyles and medical interventions.

PCA of human data is peculiar since it produced more than a single age-dependent feature. This is not the case in simpler animals such as worms [32], flies [33] or mice [34], where aging could be explained by a simple dynamic instability leading to the exponential disintegration of the organism state [33, 35]. We expect that the entropic contribution to aging has no time to develop in such cases. The biological age is then a dynamic factor, and effects of aging may be reversible [34].

This work, along with the direct dynamics stability analysis of the organism state fluctuations in longitudinal biomedical data [13, 16] shows that humans (and probably other long-lived mammals, such as naked mole-rats) evolved so that the fully grown subjects are metastable until very late in life. The loss of stability is the result of a loss of resilience due to a combined effect of many configuration transitions.

We show that aging in humans has a very significant entropic component. If the proposition is accurate, we must expect that although the hallmarks of aging (features or activations of specific pathways leading to mortality and morbidity acceleration [1]) can in principle be reverted, the expected effects on lifespan should be limited. Any attempts to reduce the dominant aging signature, tBA, would, however, run against the tendency of complex systems to increase their entropy. Any working strategy would require the availability and timely application of an immense number of precise interventions. This is, to say the least, technologically challenging. Accordingly, we must think that aging in humans can be reversed only partially.

Significant rejuvenation may thus remain a remote perspective. Our model, however, suggests that there must be a practical way to intercept aging, that is to reduce the rate of aging dramatically. The rates controlling configuration transitions between any two states depend exponentially on the effective temperature. Hence, even minor alterations of the parameter may cause a dramatic drop in the rate of aging. In condensed matter physics, this situation is known as glass transition, where the viscosity and relaxation times may grow by ten to fifteen orders of magnitude in a relatively small temperature range [36–43]. We note, of course, that living organisms are non-equilibrium open systems, and hence the effective temperature must not coincide with the body or environment temperatures. Rather, the effective temperature is a measure of deleteriousness of the environment [29].

We speculate that the evolution of long-lived mammals may have provided an example of tuning the effective temperature. Naked mole rats are known for their exceptional stress resistance, DNA repair efficacy [44–47], and translational fidelity [48, 49]. Both factors should reduce the noise in the regulatory circuits and lower the effective temperature of the system. Most recent studies indicate that naked mole-rats breeders age slower than their non-breeding peers, at least according to the DNA-methylation clock [22].

Social status and mental health also impact the aging rate measured by the DNA methylation and other clocks in humans [50, 51], possibly via neuroendocrine system. Socioeconomic status somewhat counter-intuitively significantly increases the mortality doubling rate and simultaneously reduces age-independent mortality in such a way that the mortality in the top and the lowest income group converge at an age close to our $t_{\max}$ estimates [52]. Such behavior of mortality is consistent with reducing the effective temperature in the top-earning cohorts in our model.

Future studies should help establish the best ways to "cool down" the organism state and reduce the rate of aging in humans. The simple linear PCA exemplified here may only help gain a qualitative understanding of underlying processes. We expect that increasing availability of high-quality longitudinal biomedical data will lead to a better understanding of the most critical factors behind the kinetics of aging and diseases, including those controlling the entropy production in the course of the aging process. This should lead to the discovery of actionable targets influencing the rate of the aging drift, help slow down aging and thus produce a dramatic extension of the human healthspan.

IV. ACKNOWLEDGEMENTS

We would like to thank Anastasia Velikanova and Dmitry Kriukov from Skolkovo Institute of Science and Technology, K. Avchaciov and T. Pyrkov from Gero for insightful discussions and help with the data preparation; Maxim Kholin and Alexey Kadet from Gero for stimulating discussions and comments on the manuscript. The work was funded by Gero LLC (Singapore)

V. COMPETING INTERESTS

P.O.F. is a shareholder of Gero PTE. A.E.T., K.A.D., and P.O.F. were employed by Gero PTE during the work on the manuscript. The study was funded by Gero PTE.

VI. MATERIALS AND METHODS

A. PCA of the DNA methylation data

We took the white-blood-cells methylation data from GSE87571 dataset [25]. It contains 729 samples (more than 440k features each) collected from patients of both genders (341 males and 388 females) covering the age range between 14 and 94 years of age.

To focus the analysis on aging, we filtered out the patients younger than 20 y.o. (620 samples remaining). We filtered out the CpG sites according to Pearson's correlation between the DNA methylation levels and the chronological age at the level of $p < 0.005/N$ (where N is the total number of the reported features), thus obtaining 96536 sites. We performed and reported the results of the PCA on the resulting data.

We computed Horvarth's methylation age as described in [4]. A few CpG sites (cg17099569, cg00431549, cg11025793, and cg14409958) were not present in the data, and hence we had to exclude them from the calculation.

DNAm-PC3 increased with age at a faster than linear pace. We collected all the pairs of the DNAm-PC3 scores and the chronological age for every patient n in the dataset and used the available age-range to produce a fit of the data to average from Eq. (3):

$$\mathrm{DNAm-PC3}^n \approx \frac{a}{t_{\max} - t^n} + b \cdot t^n + c$$

with the uniform Gaussian error and $t_{\max}$, a, b and c being the parameters of the fit. The calculation returned $t_{\max} = 129.9$ years. We also performed the linear fit of the inverse variance of DNAm-PC3 and obtained 90% CI $[114.5, 122.2]$ for $t_{\max}$.

B. Gene set enrichment analysis

We collected the CpG sites best associated with DNAm-PC1 and DNAm-PC3 according to the values of the respective vector components. We retrieved the gene IDs from Illumina's 450k methylation arrays documentation. Finally, we performed Gene Ontology (GO) and disease ontology (DO) enrichment with the help of the R "clusterProfiler4.0" package [53].

C. Pre-processing of EMRs from UK Biobank

To avoid using the disease labels corresponding to the transient diseases, we selected 111 chronic diseases diagnoses using Chronic Condition Indicators for ICD-10 [54]. Overall, in the EMR dataset are 389494 patients, of mostly Caucasian origin (366715 or 94%), of both sexes (179032 males and 210462 females) in the age range $38 - 74$).

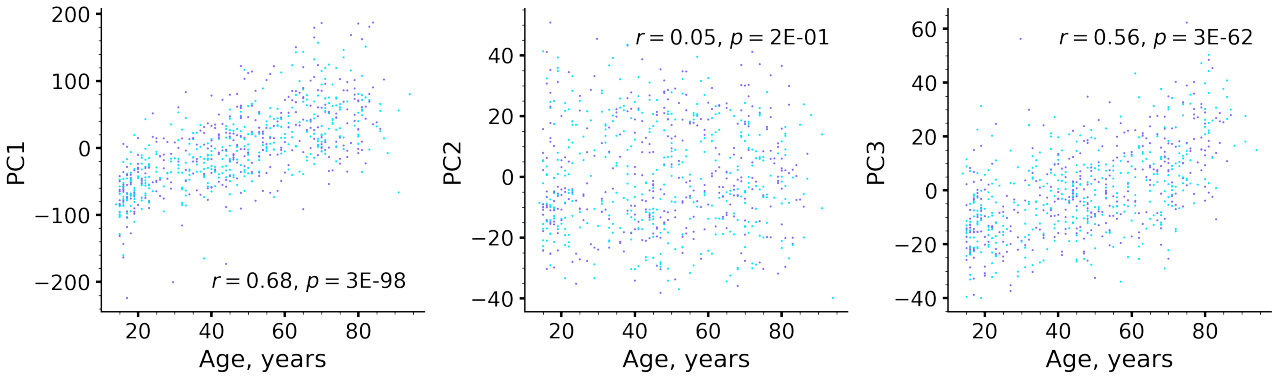

FIG. S.1. Association of the PC scores in regulatory fields with chronological age.

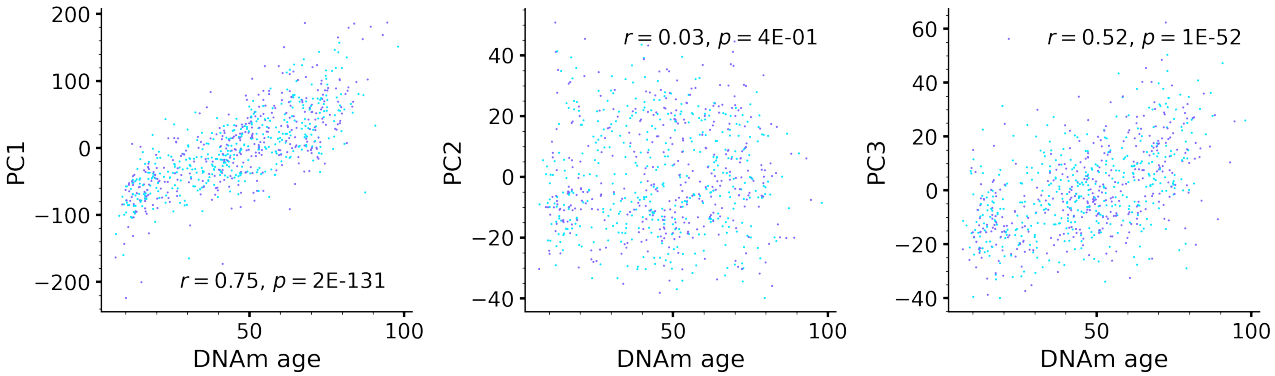

FIG. S.2. Association of the PC scores in regulatory fields with Horvath's methylation clock (DNAm age).

D. Entropy/entropy production rate determination

For the practical calculation of entropy, we used a Python library *scipy.stats.entropy* [55], which was applied to the individual distributions of methylation levels and to the distributions of EMR vectors averaged over the population in age-binned cohorts.

VII. SUPPLEMENTARY: THEORY

We start from Eqs. 1 and observe, that the regulatory fields change over time in response to the deterministic (the direct linear and the higher-order non-linear interactions between the units) and stochastic forces f_i. We naturally assume that the stochastic force terms are not correlated over long time intervals: $\langle f(t)f(t')\rangle = B\delta(t-t')$ with B is the power of the stochastic noise, $\langle ...\rangle$ stands for the averaging along the individual trajectory and over all specimen, and $\delta(t)$ is the Dirac delta-function.

In spite of apparent simplicity, the Eqs. 1 are non-linear and may have highly non-trivial solutions leading to applications in condensed matter physics [56] and neurophysiology [57]. For our discussion, it is important that the stochastic noise drives the system towards equilibrium at an effective temperature controlled by the power of the noise $T \sim B$.

The data suggests that there is a large "bulk" of units characterized by excessive lifetimes. Mechanistically this may be explained by operating within a vicinity of a metastable state with a very high activation energy $U_{\rm act}$ relative to the effective temperature, $U_{\rm act} \ll T$ (Fig. 3).

We will assume that the effects of aging are small on the scale of U and hence the depolarization rates $R_i \sim U_{\rm act}/T$ are not only very small, but also do not considerably depend on age. Accordingly, the depolarization is on average a linear function of age and the total number of configuration transitions Z: $\langle \Delta\sigma_i\rangle = R_i t \propto Z$ and $\langle \Delta\sigma_i^2\rangle \propto R_i t \propto Z$.

Let us think that the aging drift in the form of simultaneously occurring configuration transitions progresses slowly compared to fast functional responses in the organism. We linearize the equations for the regulatory

fields next to the youthful state $\bar{h}_i$:

$$\dot{\delta h}_i = \sum_j K_{ij}\delta h_j + \\ + \sum_j K_{ij}\langle\Delta\sigma_j\rangle + \sum_j g_{ijk}\langle\Delta\sigma_j^2\rangle + \\ + \sum_{j,k} g_{ijk}\frac{d\sigma_j}{dh_j}\delta h_j\langle\Delta\sigma_k\rangle + J_i^{(0)} + f_i, \tag{S.1}$$

where $\delta h_i = h_i - \bar{h}_i$ and $\Delta\sigma_j$ variables describe the deviations of the fields and depolarization of the units, whereas the averages $\langle ... \rangle$ involve the averaging over the "bulk" uncorrelated states only.

The solutions of the linearized Eq. S.1 can be best understood with the help of a linear decomposition: $\delta h_i \approx \sum_A z_A b_i^A$, where z_A are the pathway activations, and b_i^A are the right eigenvectors of the interaction matrix corresponding to the smallest eigenvalues r_A (the matrix K is non-symmetric and hence its complete eigensystem must include the left, $a^A K = -r_A a^A$, and the right, $K b^A = -r_A b^A$, eigenvectors). The components of the vector b_i^A characterize the participation of the PU i in the pathway A.

Substituting the solution into the equation and multiplying both sides by the corresponding left eigenvector, we find, that

$$\dot{z}_A = -(r_A - r'_A Z)z + \beta_A Z + J_A + f_A, \tag{S.2}$$

where $J_A = a^A J$, $f_A = a^A f$. The effect of aging comes through the mean field on the pathway activation $\beta_A Z = a^A K\langle\Delta\sigma\rangle + \mathcal{O}(g)$ and the non-linear correction to the eigenvalue $r_A \approx r_A - r'_A Z$.

It is important to understand, that all the relevant vectors and constants can not be derived and could only be measured experimentally. The large number of configuration transitions ensures by virtue of the central limit theorem that the effect of the mean field is exactly linear in Z.

Qualitatively, the net effect of the rare transitions and the associated mean field Z together produce a persistent pathway activation, on average, slowly increasing with age. This is often referred to as an enslavement principle: stochastic depolarization transitions produce a slowly evolving mean field Z that disturbs pathways characterized by fast relaxation times having thus enough time to adjust to the current value of Z.

REFERENCES

1. López-Otın, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153,** 1194–1217 (2013).
2. Zenin, A. *et al.* Identification of 12 genetic loci associated with human healthspan. *Communications biology* **2,** 1–11 (2019).
3. Gompertz, B. XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical transactions of the Royal Society of London,* 513–583 (1825).
4. Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14,** 1–20 (2013).
5. Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49,** 359–367 (2013).
6. Levine, M. E. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **68,** 667–674 (2013).
7. Petkovich, D. A. *et al.* Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell metabolism* **25,** 954–960 (2017).
8. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10,** 573 (2018).
9. Levine, M. E. *Assessment of epigenetic clocks as biomarkers of aging in basic and population research* 2020.
10. Levine, M. *et al.* A rat epigenetic clock recapitulates phenotypic aging and co-localizes with heterochromatin. *Elife* **9,** e59201 (2020).
11. Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *Nature Aging* **1,** 1189–1201 (2021).
12. Pyrkov, T. V. *et al.* Quantitative characterization of biological age and frailty based on locomotor activity records. *Aging (Albany NY)* **10,** 2973 (2018).
13. Pyrkov, T. V., Sokolov, I. S. & Fedichev, P. O. Deep longitudinal phenotyping of wearable sensor data reveals independent markers of longevity, stress, and resilience. *Aging (Albany NY)* **13,** 7900 (2021).
14. Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8,** 1844 (2016).
15. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)* **11,** 303 (2019).
16. Pyrkov, T. V. *et al.* Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts human lifespan limit. *Nature communications* **12,** 1–10 (2021).
17. Santesmasses, D. *et al.* COVID-19 is an emergent disease of aging. *Aging Cell* **19,** e13230 (2020).
18. Kuo, C.-L. *et al.* COVID-19 severity is predicted by earlier evidence of accelerated aging. *Medrxiv* (2020).
19. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging cell* **18,** e13028 (2019).
20. *Rejuvant™ Safety and Biomarker Study* `https://ClinicalTrials.gov/show/NCT04821401`.
21. Nielsen, J., Bakula, D. & Scheibye-Knudsen, M. Clinical Trials Targeting Aging. *Front. Aging* **3,** 820215 (2022).

22. Horvath, S. *et al.* DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. *Nature Aging,* 1–14 (2021).

23. Kerepesi, C. *et al.* Epigenetic aging of the demographically non-aging naked mole-rat. *Nature Communications* **13,** 355 (2022).

24. Bergstedt, J. *et al.* Factors Driving DNA Methylation Variation in Human Blood (2021).

25. Johansson, Å., Enroth, S. & Gyllensten, U. Continuous aging of the human DNA methylome throughout the human lifespan. *PloS one* **8,** e67378 (2013).

26. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018).

27. Hayflick, L. *Entropy explains aging, genetic determinism explains longevity, and undefined terminology explains misunderstanding both* 2007.

28. Teschendorff, A. E. Avoiding common pitfalls in machine learning omic data science. *Nature materials* **18,** 422–427 (2019).

29. Strehler, B. L. & Mildvan, A. S. General Theory of Mortality and Aging: A stochastic model relates observations on aging, physiologic decline, mortality, and radiation. *Science* **132,** 14–21 (1960).

30. Fischer, H. *A history of the central limit theorem: From classical to modern probability theory* (Springer Science & Business Media, 2010).

31. Porter, H. L. *et al.* Many chronological aging clocks can be found throughout the epigenome: Implications for quantifying biological aging. *Aging Cell* **20,** e13492 (2021).

32. Tarkhov, A. E. *et al.* A universal transcriptomic signature of age reveals the temporal scaling of Caenorhabditis elegans aging trajectories. *Scientific reports* **9,** 1–18 (2019).

33. Kogan, V., Molodtsov, I., Menshikov, L. I., Reis, R. J. S. & Fedichev, P. Stability analysis of a model gene network links aging, stress resistance and negligible senescence. *Scientific reports* **5,** 1–12 (2015).

34. Avchaciov, K. *et al.* Identification of a blood test-based biomarker of aging through deep learning of aging trajectories in large phenotypic datasets of mice. *bioRxiv* (2020).

35. Podolskiy, D. *et al.* Critical dynamics of gene networks is a mechanism behind ageing and Gompertz law. *arXiv preprint arXiv:1502.04307* (2015).

36. Kob, W. & Barrat, J.-L. Aging effects in a Lennard-Jones glass. *Physical review letters* **78,** 4581 (1997).

37. Yue, Y., Jensen, S. & deC. Christiansen, J. Physical aging in a hyperquenched glass. *Applied Physics Letters* **81,** 2983–2985 (2002).

38. Lubchenko, V. & Wolynes, P. G. Theory of aging in structural glasses. *The Journal of chemical physics* **121,** 2852–2865 (2004).

39. Sellitto, M. & Kurchan, J. Shear-thickening and entropy-driven reentrance. *Physical review letters* **95,** 236001 (2005).

40. Lunkenheimer, P., Wehn, R., Schneider, U. & Loidl, A. Glassy aging dynamics. *Physical review letters* **95,** 055702 (2005).

41. Rittigstein, P. & Torkelson, J. M. Polymer–nanoparticle interfacial interactions in polymer nanocomposites: confinement effects on glass transition temperature and suppression of physical aging. *Journal of Polymer Science Part B: Polymer Physics* **44,** 2935–2943 (2006).

42. Greinert, N., Wood, T. & Bartlett, P. Measurement of effective temperatures in an aging colloidal glass. *Physical review letters* **97,** 265702 (2006).

43. Jabraoui, H., Ouaskit, S., Richard, J. & Garden, J.-L. Determination of the entropy production during glass transition: Theory and experiment. *Journal of Non-Crystalline Solids* **533,** 119907 (2020).

44. Delaney, M., Nagy, L., Kinsel, M. & Treuting, P. Spontaneous Histologic Lesions of the Adult Naked Mole Rat (Heterocephalus glaber) A Retrospective Survey of Lesions in a Zoo Population. *Veterinary pathology* **50,** 607–621 (2013).

45. Tian, X. *et al.* High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499,** 346–349 (2013).

46. Grimes, K. M., Reddy, A. K., Lindsey, M. L. & Buffenstein, R. And the beat goes on: maintained cardiovascular function during aging in the longest-lived rodent, the naked mole-rat. *American Journal of Physiology-Heart and Circulatory Physiology* **307,** H284–H291 (2014).

47. Evdokimov, A. *et al.* Naked mole rat cells display more efficient excision repair than mouse cells. *Aging (Albany NY)* **10,** 1454 (2018).

48. Buffenstein, R. *et al.* The naked truth: a comprehensive clarification and classification of current 'myths' in naked mole-rat biology. *Biological Reviews* **97,** 115–140 (2022).

49. Azpurua, J. *et al.* Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. *Proceedings of the National Academy of Sciences* **110,** 17350–17355 (2013).

50. Protsenko, E. *et al.* "GrimAge," an epigenetic predictor of mortality, is accelerated in major depressive disorder. *Translational psychiatry* **11,** 1–9 (2021).

51. Ng, T. K. S. *et al.* Association between housing type and accelerated biological aging in different sexes: moderating effects of health behaviors. *Aging (Albany NY)* **13,** 20029 (2021).

52. Chetty, R. *et al.* The association between income and life expectancy in the United States, 2001-2014. *Jama* **315,** 1750–1766 (2016).

53. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2,** 100141 (2021).

54. *Chronic Condition Indicators for ICD-10-CM* `https://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp`.

55. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

56. Amit, D. J., Gutfreund, H. & Sompolinsky, H. Spin-glass models of neural networks. *Physical Review A* **32,** 1007 (1985).

57. La Camera, G. The mean field approach for populations of spiking neurons. *arXiv preprint arXiv:2109.01279* (2021).