Senescence of multicellular individuals: imbalance of epigenetic and non-epigenetic information in histone modifications

Felipe A. Veloso [iD] *

Facultad de Ciencias, Universidad Mayor, Santiago, Chile.
(Dated: April 27, 2018)

Cellular aging has been progressively elucidated by science. However, aging at the multicellular-individual level is still poorly understood. A recent theory of individuated multicellularity describes the emergence of crucial information content for cell differentiation. This information is mostly conveyed in the non-epigenetic constraints on histone modifications near transcription start sites. According to this theory, the non-epigenetic content emerges at the expense of the information capacity for epigenetic content. However, it is unclear whether this "reassignment" of capacity continues after adulthood. To answer this question, I analyzed publicly available high-throughput data of histone H3 modifications and mRNA abundance in human primary cells. The results show that the "reassignment" continues after adulthood in humans. Based on this evidence, I present a falsifiable theory describing how continued "reassignment" of information capacity creates a growing epigenetic/non-epigenetic information imbalance. According to my theoretical account, this imbalance is the fundamental reason why individuated multicellular organisms senesce.

Keywords: aging; ageing; cancer; constraints; naked mole-rat; bristlecone pine; *Turritopsis*; gene regulation; epigenetics; teleodynamics

Our intellectual endeavors have entertained the prospect of unlimited lifespan for centuries [1], and the scientific endeavor has been no exception [2]. In the 1950s, the immortality of cultured somatic cells was indeed a widely-held belief [3]. That changed only when Hayflick & Moorhead showed that cultured human somatic cells do stop dividing and become less viable once their divisions reach a certain number [4], a phenomenon known today as the Hayflick limit [3]. This loss of replicative capacity and, in general, the process of aging at the cellular level, have been found to correlate with telomere length [3,5,6]. Yet, the number of times human cells can divide in culture exceeds the number of times cells divide throughout our lifespan; there is no significant correlation between human cell replicative capacity and cell donor age [7]. That is, we—and individuated multicellular organisms in general—age before most of our cells do [8,9]. The outstanding question is why.

Theoretical descriptions of senescence or aging at the multicellular-individual level have been classified into two categories: programmed senescence and senescence caused by damage/error [10]. Recently it has been argued, however, that senescence is not programmed nor is it ultimately a consequence of damage or error in the organism's structure/dynamics [11]. Instead, it may be a byproduct of maintenance and/or developmental dynamics [11,12], themselves underpinned in part by intracellular signaling pathways such as the cell-cycle-related PI3K/AKT/mTOR pathway [11]. These pathways have been shown to modulate aging at the cellular level in species such as the yeast *Saccharomyces cerevisiae* [13].

The analogous notion of aging at the multicellular-individual level as a byproduct of certain functional signaling pathways [11] is, in principle, supported by the fact that the deficiency of mTOR kinase—a key component of the PI3K/AKT/mTOR pathway—can double the lifespan of the roundworm *Caenorhabditis elegans* [14]. However, the fundamental dynamics that make individuated multicellular organisms senescent after adulthood remain unclear and largely lack falsifiable scientific theories. Falsifiability—the possibility of establishing a hypothesis or theory as false by observation and experiment [15]—allows the objective rejection of existing scientific theories, fosters the development of new ones, and constitutes the most widely accepted demarcation between science and non-science [16].

Using publicly available high-throughput data of histone H3 modifications and mRNA abundance in human primary cells to look for proof of concept, the issue of senescence can also be approached from the angle of theoretical biology. Thus, I conducted a statistical data analysis in this study, which revealed that proof of concept exists for the human species. These findings provide empirical grounds for my theoretical work, suggesting that senescence is a byproduct of functional developmental dynamics as first described by a recently proposed theory of individuated multicellularity [17]. Specifically, I show that the byproduct is a post-ontogenetic, growing imbalance between two different information contents conveyed respectively in two different types of constraints on histone post-translational modifications near transcription start sites (TSSs). Constraints are here understood as the local and level-of-scale specific thermodynamic boundary conditions required for energy to be released as work as described by Atkins [18]. The concept of constraint is crucial because, according to the theory of individuated multicellularity, a higher-order constraint (i.e., a constraint on constraints) on changes in histone modifications harnesses critical work that

———

* Correspondence: veloso.felipe.a@gmail.com

2

regulates transcriptional changes for cell differentiation at the multicellular-individual level.

Under the theory of individuated multicellularity, the intrinsic higher-order constraint is the simplest multicellular individual in fundamental terms. In addition, the dynamics of the lower-order constraints must be explicitly unrelated to each other (i.e., statistically independent) in order to elicit the emergence of the intrinsic higher-order constraint. Along with the emergence of this intrinsic higher-order constraint, the theory of individuated multicellularity describes the emergence of critical information content, named in the theory hologenic content, which is about the multicellular individual as a whole in terms of developmental self-regulation. Thus, for the sake of brevity, I here refer to the theory of individuated multicellularity as the hologenic theory.

The constraints on the combinatorial patterns of histone modifications are generally known as histone crosstalk [19,20]. Histone modifications are also known to be relevant for epigenetic changes [21], which are defined as changes in gene expression that cannot be explained by (i.e., that are explicitly unrelated to) changes in the DNA sequence [22]. This relevance is underpinned by the capacity of histone modifications to convey information content, which has allowed the prediction of mRNA levels from histone modification profiles near TSSs with high accuracy [23].

Based on these considerations and the properties of the nonnegative measure of multivariate statistical association known as total correlation [24] or multiinformation [25] (symbolized by C and typically measured in bits), the overall observable histone crosstalk can be decomposed. That is, histone crosstalk, if measured as a total correlation C, is the sum of two explicitly unrelated C components: one epigenetic (i.e., explicitly related to changes in gene expression) and the other non-epigenetic (i.e., explicitly unrelated to changes in gene expression). This sum can be expressed as follows:

$$\underbrace{C(X_1,\ldots,X_n)}_{\substack{\text{Overall histone crosstalk} \\ \text{(total correlation} \\ \text{of } X_1,\ldots,X_n)}} = \underbrace{C_Y(X_1,\ldots,X_n,Y)}_{\substack{\text{Epigenetic histone crosstalk} \\ \text{(total correlation of } X_1,\ldots,X_n \\ \text{that is explicitly related to } Y), \\ \text{depends on DNA-nucleosome} \\ \text{interactions (Ref. 17), and conveys} \\ \text{epigenetic information content}}} + \underbrace{C(X_1,\ldots,X_n|Y),}_{\substack{\text{Non-epigenetic histone crosstalk} \\ \text{(total correlation of } X_1,\ldots,X_n \\ \text{that is explicitly unrelated to } Y), \\ \text{depends on protein/RNA-nucleosome} \\ \text{interactions (Ref. 17), and conveys} \\ \text{hologenic information content}}} \tag{1}$$

where $X_1,\ldots,X_n$ are random variables representing n histone modification levels in specific genomic positions with respect to the TSS and Y is a random variable representing either gene expression level, transcription rate, or mRNA abundance level associated with the TSS. These levels are equivalent for the decomposition because of the strong correlation that exists between them ([26] and references therein).

The hologenic theory describes how the epigenetic component of histone crosstalk (represented by $C_Y(X_1,\ldots,X_n,Y)$ in the sum decomposition of Eq. 1) conveys information about each cell's transcriptional profile. This component is, in information content terms, the dominating component for any eukaryotic colonial species (such as the alga *Volvox carteri* [27]) and, importantly, also for undifferentiated stem cells.

The second, non-epigenetic component of histone crosstalk (represented by $C(X_1,\ldots,X_n|Y)$ in Eq. 1) is known to grow in magnitude during development until the organism's mature form is reached [17]. This component is described by the hologenic theory as conveying information about the multicellular individual as a whole—starting from the moment said individual emerges as an intrinsic higher-order constraint on the early embryo's proliferating cells.

Importantly, the overall observable histone crosstalk magnitude (represented by $C(X_1,\ldots,X_n)$ in Eq. 1) is not infinite. In other words, the overall histone crosstalk has a finite information capacity, which can be measured in bits. Moreover, the sum decomposition in Eq. 1 implies that the growth in magnitude (bits) of the hologenic (i.e., non-epigenetic) component must be accompanied by a decrease in magnitude of the epigenetic component. That is, the capacity (in bits) for hologenic information content in histone crosstalk is bound to grow at the expense of the capacity for epigenetic information content.

The hologenic theory also maintains that a necessary condition for the evolution of individuated multicellular lineages was the appearance of a class of molecules synthesized by the cells—called Nanney's extracellular propagators (symbolized by $F_N^{\rightarrow}$) in the theory [17]. These $F_N^{\rightarrow}$ molecules are predicted to be, in a given tissue and time period, (i) secretable into the extracellular space, (ii) once secreted, capable of eliciting a significant incremental change (via signal transduction) in the magnitude of the non-epigenetic histone crosstalk (i.e., the $C(X_1,\ldots,X_n|Y)$ summand in Eq. 1) within other cells' nuclei, and (iii) affected in their extracellular diffusion dynamics by the geometrical complexity of the extracellular space (i.e., constraints on diffusion at the multicellular-individual level, which cannot be reduced to constraints at the cellular level). Also under the hologenic theory, for the multicellular individual to develop and survive, both hologenic (developmental self-regulation of

3

the multicellular individual overall) and epigenetic (each cell's transcriptional profile) contents must coexist.

One final but important consideration regarding histone crosstalk is that it is the result of constraints which, as mentioned previously, are level-of-scale specific. To exemplify this specificity, consider the example of an internal combustion engine: a single molecule in a cylinder wall does not embody a constraint on the expansion of the igniting gas, yet the cylinder-piston ensemble does. For this reason, histone crosstalk constraints were expected to have relevance for senescence but only at a specific level of scale. The specific level of scale in histone crosstalk that is relevant for human senescence has not been studied in detail before.

To investigate from a theoretical angle if the "reassignment" of information capacity for epigenetic and non-epigenetic (i.e., hologenic) content stops when development reaches the multicellular individual's mature form or instead continues without interruption, one also needs to investigate the "reassignment" (if any) in cancer cells. One of the corollaries of the hologenic theory is a significant loss of hologenic content in cancer cells, because they are no longer constrained by the multicellular individual that normal (i.e., non-cancerous) cells serve and are constrained by. Thus, I developed a falsifiable theory of senescence based on the post-ontogenetic continuation of this "reassignment" process in human histone crosstalk as proof of concept.

To test this theory, I formalized the proof of concept into the following two hypotheses: (i) within genomic regions adjacent to TSSs in primary normal cells, the log-ratio between the non-epigenetic and epigenetic histone H3 crosstalk magnitudes is significantly and positively correlated with cell donor age (over a range of 0-90 years old) and (ii) no such statistically significant correlation exists for primary cancer cells (see Fig. 1).
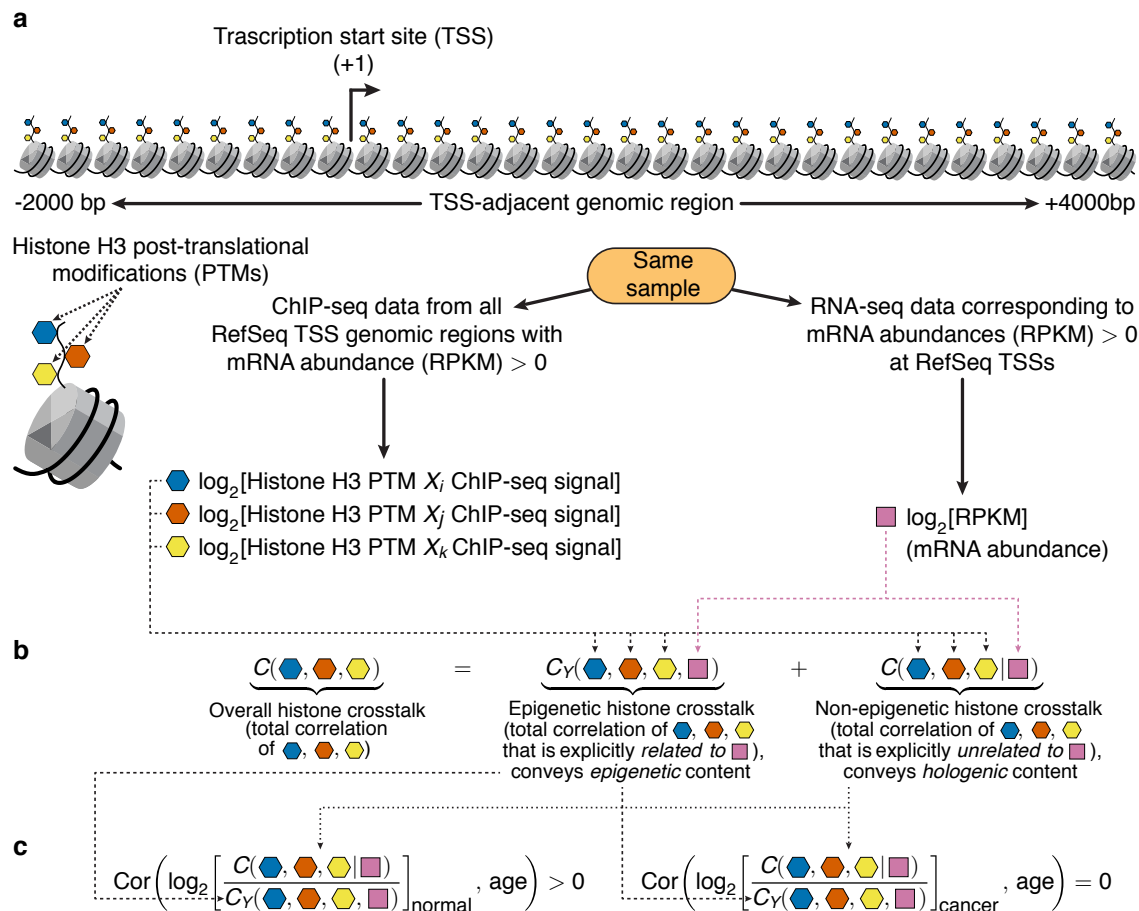
Fig. 1. **Schematic for proof-of-concept hypotheses and computational analysis for testing.** Publicly available ChIP-seq (chromatin immunoprecipitation followed by high-throughput DNA sequencing) and RNA-seq (transcriptome high-throughput sequencing) data for human primary cell samples allowed the computation, for each TSS, of position-specific histone H3 modification levels (at every 200bp) and its associated mRNA abundance level **(a)**. After log-transforming these levels and taking into account all TSSs, the TSS-adjacent histone H3 crosstalk (triad-wise crosstalk depicted here) was represented as a total correlation [24] or information capacity in bits, which in turn was decomposed as the sum of two measurable and explicitly unrelated components: one epigenetic (explicitly related to transcriptional changes) and the other non-epigenetic (explicitly unrelated to said changes) **(b)**. Taking into account all samples, the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude was hypothesized to be positively correlated with cell donor age in normal cells **(c, left)** and also to be uncorrelated with cell donor age in cancer cells **(c, right)**. The subsequent rejection of the statistical null hypothesis in **(c, left)** and the failure to reject the statistical null hypothesis in **(c, right)** provided proof of concept for the theory of senescence proposed in this paper.

4

RESULTS

To test the two proof-of-concept hypotheses, I used publicly available ChIP-seq (chromatin immunoprecipitation followed by high-throughput DNA sequencing) and RNA-seq (transcriptome high-throughput sequencing) data for human primary cell samples, which were obtained from different individuals with ages ranging from 0 to +90 years old. I computed log-transformed ChIP-seq signal magnitudes for each primary cell sample from ChIP-seq data of different position-specific (in bp relative to the TSS) histone H3 modifications. Similarly, I log-transformed mRNA abundance values from RNA-seq assay data associated with each ChIP-seq assay for each primary cell sample.

Using these tandem ChIP-seq and RNA-seq data, I quantified the non-epigenetic and epigenetic histone H3 crosstalk magnitudes (Eq. 1) for triads of variables $\{X_i, X_j, X_k\}$. These variables represented position-specific histone H3 modification levels, i.e., $C(X_i, X_j, X_k|Y)$ and $C_Y(X_i, X_j, X_k, Y)$ for the non-epigenetic and epigenetic histone crosstalk components, respectively, where Y represents mRNA abundance. Triads (as opposed to pairs or tetrads) were first analyzed because a triad constitutes the number of variables (i.e., position-specific histone modification levels) found to possess both significant predictive power and predictive synergy to resolve the statistical uncertainty about the mRNA abundance level associated with a given TSS (see details in Methods).

The log-ratio (base 2) between the non-epigenetic and epigenetic histone H3 crosstalk magnitudes was thus computed as the dimensionless quantity

$$\log_2\left[\frac{C(X_i, X_j, X_k|Y)}{C_Y(X_i, X_j, X_k, Y)}\right]. \tag{2}$$

Importantly, total correlation C captures all possible associations in the set of variables $\{X_i, X_j, X_k\}$ that may exist starting from the pairwise level.

The log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude is positively correlated with cell donor age in normal cells

ChIP-seq data for five histone H3 modifications were used in all analyses: H3K4me1 (histone H3 lysine 4 monomethylation), H3K9me3 (histone H3 lysine 9 trimethylation), H3K27ac (histone H3 lysine 27 acetylation), H3K27me3 (histone H3 lysine 27 trimethylation), and H3K36me3 (histone H3 lysine 36 trimethylation). The ChIP-seq signals for these modifications were computed for 30 200bp-long genomic bins across a 6,000bp-long TSS-adjacent region (see Fig. 1). Thus, a total of 150 variables X_i representing
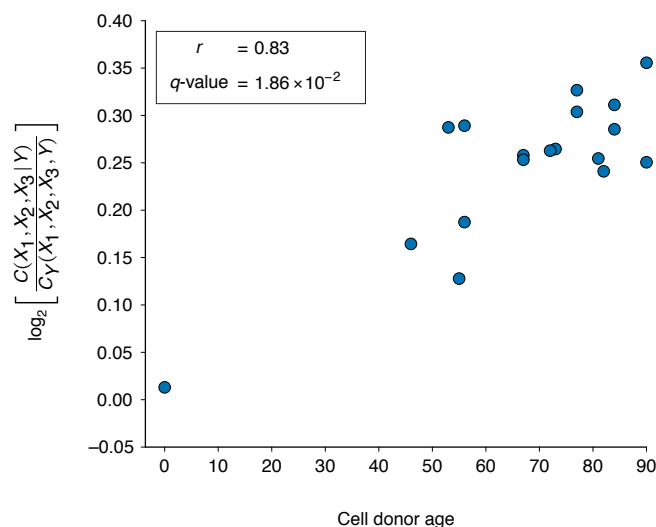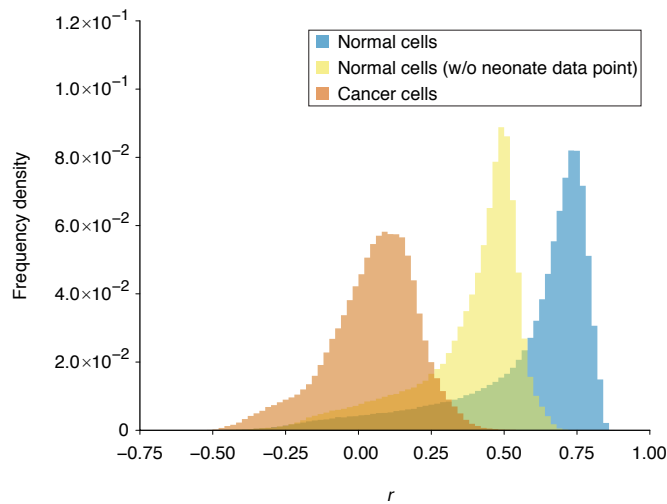

Fig. 2. **Positive correlation between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude and cell donor age for one triad of position-specific histone H3 modification levels in normal cells.** In this triad X_1 represents H3K27ac (at –1000bp), X_2 represents H3K36me3 (at +1000bp), and X_3 represents H3K4me1 (at +3200bp), which together comprise all TSSs. Regardless of the specific triad, Y always represents the mRNA abundance profile comprising all TSSs. Each data point in the figure corresponds to a primary cell sample.

position-specific histone H3 modification levels—each variable with signals for 18,220 RefSeq TSSs—were used when analyzing each cell sample. A total of 18 normal cell samples and 17 cancer cell samples was included in the analysis. The Pearson correlation coefficient r between the log-ratio and the cell donor age was obtained for each of the $\binom{150}{3} = 551,300$ possible $\{X_i, X_j, X_k\}$ triads. The 551,300 p-values (one-sided Student's t-test) associated to these r values were then corrected for multiple testing (Benjamini-Yekutieli correction, see Methods), obtaining q-values.

To determine whether the hypothesized positive correlation between the non-epigenetic/epigenetic histone H3 crosstalk log-ratio and cell donor age exists, and also to illustrate the concept of positive correlation in normal cells vs. no correlation cancer cells, I obtained all possible 551,300 correlation values for triads (one-sided Student's t-test). To exemplify, the results for the triad {H3K27ac (at –1000bp), H3K36me3 (at +1000bp), H3K4me1 (at +3200bp)} are shown here, where the correlation was positive ($r=0.83$) and highly significant ($q=1.86\times10^{-2}$), as seen in Fig. 2, indicating that the hypothesized correlation holds for this triad.

Altogether, the 551,300 correlation values had a mean value $\bar{r}=0.58$, a median value $\tilde{r}=0.67$, and a standard deviation value $\sigma_r=0.24$ (see statistical distribution of r in Fig. 3). From these correlation values, only 24,185 (i.e., ~4%) were nonpositive and none of them was statistically significant (i.e., where $r\leq0$, $q>0.05$). In contrast, it was found that for 315,378 triads

Fig. 3. Statistical distribution of the correlation coefficient (r) between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude (triad-wise) and cell donor age. Histograms represent r values for all 551,300 possible triads in normal cells (blue), normal cells excluding the neonate data point (yellow), and cancer cells (orange).
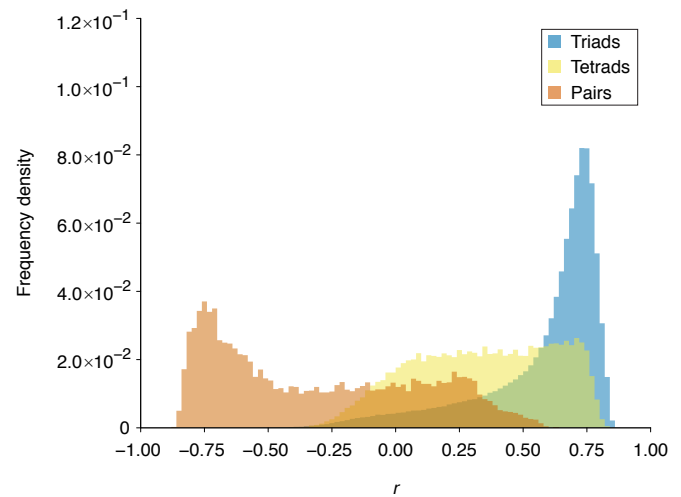
Fig. 4. Statistical distribution of the correlation coefficient (r) between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude and cell donor age for triads, tetrads, and pairs of position-specific histone H3 modification levels. Histograms represent r values for all 551,300 possible triads (blue), 50,000 random tetrads (yellow), and all 11,175 possible pairs (orange).

(i.e., ~57%) the correlation values were positive and statistically significant (i.e., $r>0$ and $q\leq0.05$).

Importantly, I also found that the hypothesized positive correlation between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk and cell donor age verified for triads of position-specific histone H3 modifications in normal cells loses its strength for tetrads ($\bar{r}=0.34$; $\tilde{r}=0.35$; $\sigma_r=0.27$). It is also no longer greater than zero for pairs ($\bar{r}=-0.30$; $\tilde{r}=-0.38$; $\sigma_r=0.40$) (Fig. 4). These results for tetrads and pairs indicate that the predicted positive correlation only holds for triads (and it was predicted in the second proof-of-principle hypothesis not to hold in cancer cells). Such specificity was expected because if senescence can be explained in terms of an imbalance of information-conveying constraints that are level-of-scale specific like other thermodynamic constraints, the imbalance itself also must be level-of-scale specific.

The log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude does not correlate with cell donor age in cancer cells

When I analyzed the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude and cell donor age for cancer cells using the same exemplary triad {H3K27ac (at −1000bp), H3K36me3 (at +1000bp), H3K4me1 (at +3200bp)}, I found that no significant correlation exists between those two variables ($r=-0.2$; $q=1$; see Fig. 5), as hypothesized.

For the 551,300 correlation values corresponding to all triads of position-specific histone H3 modifications in cancer cells, the mean and median were close to zero ($\bar{r}=0.05$; $\tilde{r}=0.07$), and the standard deviation was
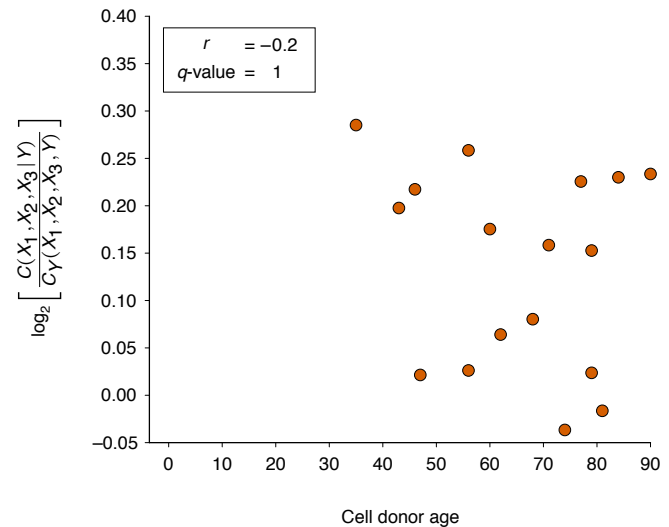
Fig. 5. No significant correlation between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude and cell donor age for one triad of position-specific histone H3 modification levels in cancer cells. This is the same triad of Fig. 2, i.e., X_1 representing H3K27ac (at −1000bp), X_2 representing H3K36me3 (at +1000bp), and X_3 representing H3K4me1 (at +3200bp).

$\sigma_r=0.15$ (see statistical distribution of r in Fig. 3). All associated p-values (two-sided Student's t-test) were corrected and the resulting q-values were all equal to 1 and hence non-significant. Similar results—i.e., all q-values equal to 1—were obtained for all 11,175 pairs of position-specific histone H3 modification levels ($\bar{r}=-0.04$; $\tilde{r}=-0.05$; $\sigma_r=0.23$) and for all 50,000 random tetrads ($\bar{r}=0.08$; $\tilde{r}=0.11$; $\sigma_r=0.15$). These results suggest that, as predicted, no significant correlation exists between the log-ratio of non-epigenetic to epigenetic histone H3 crosstalk magnitude and cell donor age in cancer cells.

6

I also evaluated whether the stark difference of the correlation values between normal—i.e., r markedly positive—and cancer cells—i.e., r close to zero—was only attributable to the data point (for normal cell samples) that corresponds to a neonate, with coordinates $(0, 0.01)$ in Fig. 2. In other words, whether the neonate data point was simply a statistical outlier that created an otherwise nonexistent difference between normal and cancer cells in the analysis.

For this purpose, I recomputed all 551,300 correlation values corresponding to normal cells, excluding the neonate data point. The mean, median, and standard deviation values obtained were $\bar{r}=0.38$, $\tilde{r}=0.44$, and $\sigma_r=0.19$, respectively (see distribution of r in comparison with that for cancer cells in Fig. 3). This difference between r values for normal cells (neonate data point excluded) and cancer cells was further tested and shown to be highly significant (Mann-Whitney U test: $U=2.7\times10^{11}$, $p<2.2\times10^{-16}$). These findings suggest that the neonate data point is not a statistical outlier among normal cell samples let alone explains the difference between normal and cancer cells in terms of the correlation values obtained.

The total information capacity of triad-wise histone H3 crosstalk does not correlate with cell donor age

Finally, I assessed whether the total information capacity (represented by $C(X_1, \ldots, X_n)$ in Eq. 1 and measured in bits) of overall histone H3 crosstalk (triad-wise) is significantly correlated with age, in particular, whether it is positively correlated. This potential correlation is important, because if total information capacity increases with cell donor age, an age-correlated decrease of the proportion available for epigenetic content would not be necessarily a problem. That is, a proportionally smaller and smaller information capacity for epigenetic content within histone crosstalk would not generate an information content imbalance—hypothesized in the Introduction—as long as a growing total capacity provides enough room for epigenetic content in absolute terms.

To test this possibility, the correlation value r between cell donor age and total information capacity (in bits) of TSS-adjacent histone H3 crosstalk, computed as

$$C_Y(X_i, X_j, X_k, Y) + C(X_i, X_j, X_k|Y), \qquad (3)$$

was obtained for all 551,300 triads of position-specific histone H3 modifications for normal cells.

The analysis revealed that the correlation coefficients r have mean, median, and standard deviation values $\bar{r}=0.21$, $\tilde{r}=0.21$, and $\sigma_r=0.24$, respectively, and that all associated q-values were equal to 1 and thus non-significant. For cancer cells, all correlation values were also non-significant ($q=1$). Their mean, median, and standard deviation values were $\bar{r}=0.01$, $\tilde{r}=0.04$, and $\sigma_r=0.18$, respectively. These results suggest that senescence would indeed be an information capacity "reassignment" problem—creating in turn an information-content imbalance, as hypothesized—rather than a "total capacity contraction" problem.

Taken together, the statistical strength of all the results obtained—notwithstanding the heterogeneous origin of the primary cell samples analyzed given the different tissues from different individuals—provides proof of concept and underpins a strong falsifiable prediction for a theory of senescence presented in the Discussion.

DISCUSSION

The successful testing of the two proof-of-concept hypotheses in the present work provides empirical grounds for the following falsifiable theory of senescence as a byproduct of developmental dynamics: Given that the "reassignment" process for information capacity in histone crosstalk—i.e., a progressive gain of capacity for hologenic information content at the expense of that for epigenetic content—continues without interruption throughout the multicellular individual's lifespan, a growing and ultimately lethal information content imbalance is created in the cells' nuclei. Importantly, this "reassignment" process is underpinned by constraints on the extracellular diffusion of $F_{\vec{N}}$ molecules, and the constraints are embodied only at the multicellular-individual level. That is, in histone crosstalk there is a time-correlated loss of capacity for epigenetic information (i.e., less and less epigenetic constraints on histone crosstalk), which causes a global and progressive impairment of biological functions at the multicellular-individual level, eventually causing the death of the individual.

The nature of the epigenetic constraints on histone crosstalk strongly implicates this time-correlated loss of capacity for epigenetic information content (and concurrent gain of that for hologenic content) as the fundamental cause of senescence. Epigenetic constraints are explicitly related to transcriptional/gene expression changes and represented by the $C_Y(X_1, \ldots, X_n, Y)$ summand in Eq. 1. Because they depend on the interactions between the histone-modified nucleosomes and the DNA wrapped around them—allowing or preventing transcription—the epigenetic information content they embody allows precise mRNA (and, ultimately, gene expression) levels from histone modification patterns.

This age-correlated hologenic/epigenetic information imbalance in histone crosstalk can also be understood in terms of an imbalance between the accuracy and precision of transcription in the cells with respect to the needs of the multicellular individual. That is, more accuracy (i.e., closeness of the mean mRNA level to the mean level functional for the multicellular individual) is reached with age at the expense of precision (i.e., closeness of the resulting mRNA levels from the same pattern

of histone modifications). This trade-off is unavoidable because (i) the relative growth of $C(X_1, \ldots, X_n | Y)$ implies an increasing constraint on (i.e., regulation of) histone modification patterns with respect to the multicellular individual [17], thus making transcription more accurate and (ii) the concurrent relative decrease of $C_Y(X_1, \ldots, X_n, Y)$ means histone modification patterns become worse and worse predictors of mRNA levels, in turn making transcription less and less precise to the point of dysfunctionality with respect to the multicellular individual (see schematic in Fig. 6a).

Thus, we can characterize senescence under this theory as a global transcriptional over-regulation with respect to the multicellular individual's needs—as opposed to cancer, where the dysfunctional effect is typically characterized in terms of the dysregulation of transcription and gene expression [28,29].

The following general prediction applies to the falsifiability of the theory of senescence: Within genomic regions adjacent to TSSs in primary normal cells from any given tissue in any individuated multicellular species, a significant positive correlation will be observed between the log-ratio of non-epigenetic to epigenetic histone crosstalk magnitude and the age of the individual from whom the cells were obtained. The specific level of crosstalk—i.e., number of position-specific histone modifications involved—at which this correlation exists may vary among species. It is predicted to be the level that possesses both significant predictive power and predictive synergy (see Methods) on mRNA levels. Moreover, since hologenic information content is described as emerging locally and independently in each developmental process [17], the statistical strength of the predicted positive correlation will be further increased—and underpinned by a monotonically increasing function—if all primary cell samples are obtained from the same tissue of the same individual throughout its lifespan.

The notable exceptions to be made for the prediction above are a few species able to undergo reverse developmental processes from adult to juvenile stages. One such species is the jellyfish *Turritopsis nutricula* [30], which is predicted to display an analogous negative correlation in the processes, i.e., "reassignment" in reverse. Another exception for the prediction are species displaying extremely slow or potentially negligible senescence processes [31]. Examples of these are the bristlecone pine *Pinus longaeva* [32], the freshwater polyp *Hydra vulgaris* [33], and the naked mole-rat *Heterocephalus glaber* [34], which, after adulthood, are predicted to display a significant but very weak positive correlation (in cases where senescence is extremely slow), or an hologenic/epigenetic log-ratio invariant with age (i.e., no correlation in cases where senescence is truly negligible; Fig. 6b).

Senescence is widely regarded as an evolutionary consequence of the relaxation of selection on traits that maintain/repair the multicellular individual's functions in later life, because later life would have been rarely realized in the wild with the hazards it imposes [35]. However, under the falsifiable theory presented in this paper, this consensus is fundamentally incorrect. Indeed, senescence at the multicellular-individual level is, I suggest, not the result of relaxed selection but instead an intrinsic developmental byproduct that would have been already observable theoretically in the emergence of the very first individuated multicellular organisms as described by the hologenic theory [17]. In other words, had the first individuated multicellular organisms been free from any extrinsic hazard in the wild, they would have begun to senesce significantly after reaching a mature form in their development, as opposed to displaying extremely slow or negligible senescence as can be inferred from the relaxed-selection hypothesis.

If correct, the evolutionary account of senescence suggested here underscores the need for modern evolutionary theory to incorporate the effects of the few yet crucial events where unprecedented forms of biological individuality have emerged throughout the history of life on Earth. One of these events—as discussed here—is the emergence of the individuated multicellular organism [17] with senescence as its developmental byproduct, and its influence on the population renewal process. Other emergence events where new forms of individuality can arise with significant evolutionary consequences include the origin of life—explicitly excluded by Darwin from the scope of his original theory [36]—with its unprecedented self-regulating and self-reproducing dynamics that first enabled natural selection [44], and the emergence of the mind [37], which—through synthetic biology—could at some point elicit the appearance of new species in the evolutionary process without any involvement of natural selection. These latter two events, and potentially others, remain to be fully elucidated along with their evolutionary consequences.

Any theory of senescence is bound to address the question of whether aging at the multicellular-individual level can be dynamically stopped. The answer suggested here is that achieving a dynamical arrest of senescence is not a fundamental impossibility but it may well be a technical impossibility because of a therapeutic safety issue. From a fundamental point of view, methods could be developed to, for example, artificially increase the dynamical range of nucleosome-DNA interactions (thus increasing the capacity for epigenetic information content in histone crosstalk at the expense of that for hologenic content).

Yet, the hologenic theory also predicts that a significant loss of hologenic content is a necessary condition for the onset of cancer. If this is correct, a potentially unsurmountable safety problem arises: also under hologenic theory, the *in vivo* balance between hologenic and epigenetic information content is predictably "fine-tuned" as it is individual-specific, cell-type-specific, and also confined to small functional ranges. Thus, there could be an inherent high risk of greatly increasing cancer incidence with the slightest

8
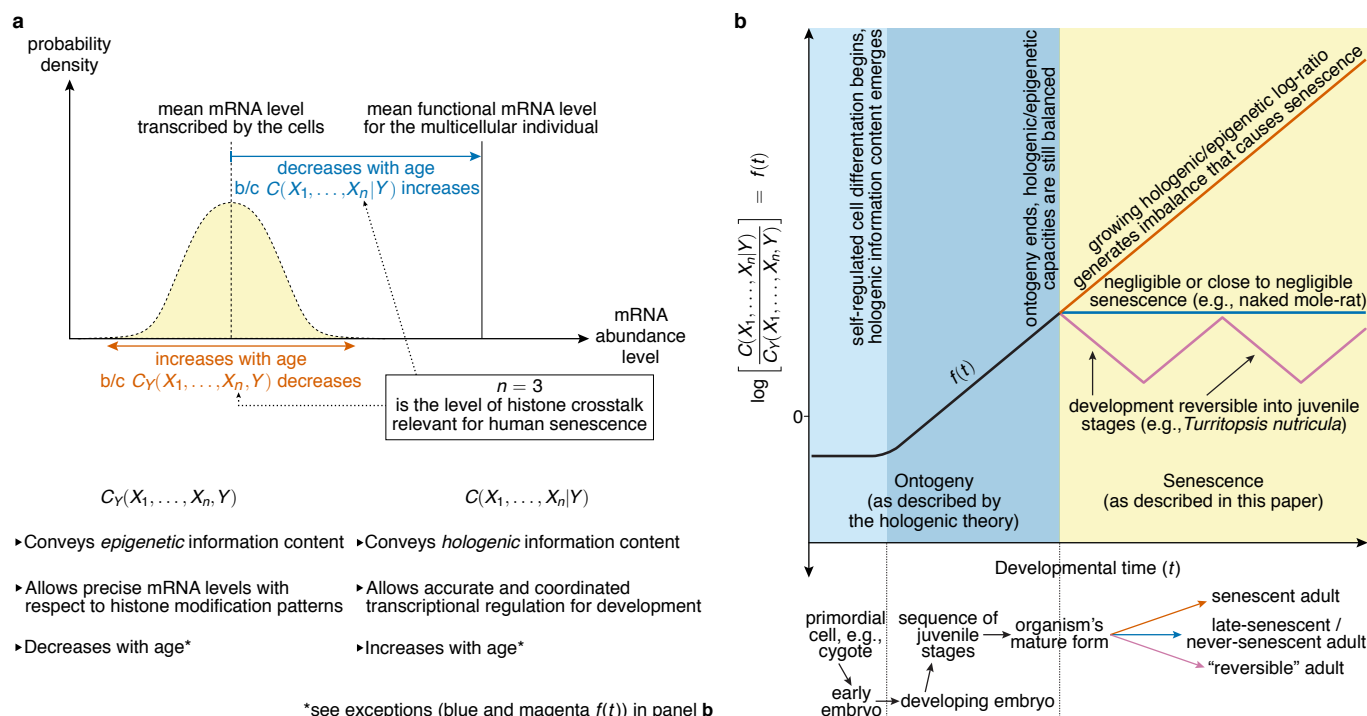
Fig. 6. Schematic of the imbalance between information capacity for hologenic content and epigenetic content in TSS-adjacent histone modification crosstalk as the cause of senescence. **(a)** The age-correlated increase in $C(X_1, \ldots, X_n | Y)$ and concurrent decrease in $C_Y(X_1, \ldots, X_n, Y)$ over-regulates the transcriptional response with respect to the multicellular individual's needs, i.e., the response becomes more and more accurate (blue) but less and less precise (orange) to the point of dysfunctionality. This unavoidable trade-off is explained by histone modification patterns becoming more constrained by regulation at the multicellular-individual level while at the same time becoming worse predictors of mRNA levels. The specific critical level of histone modification crosstalk (i.e., the value of n in $\{X_1, \ldots, X_n\}$) at which this phenomenon occurs is $n=3$ (i.e., triads of position-specific histone modifications) in humans, but may vary for other species. **(b)** The log-ratio of non-epigenetic to epigenetic histone crosstalk magnitude increases during development as the embryo grows (black $f(t)$ in darker blue area). After the organism reaches its mature form (yellow area), the log-ratio continues to increase (orange $f(t)$)—with a few notable exceptions (blue and magenta $f(t)$). This continuous increase in turn creates an increasing dysfunctional imbalance of information contents that translates into senescence and, eventually, into death.

547 extrinsic attempt to correct for the hologenic/epigenetic
548 content imbalance. This problem resides in that hologenic
549 constraints, whose growth in magnitude has senescence
550 as a byproduct, are the very constraints preventing an
551 otherwise likely onset of cancer [17].

552   Based on a mathematical model of intercellular
553 competition, Nelson and Masel have argued that stopping
554 senescence, even if possible, will always elicit the onset of
555 cancer and that senescence is ultimately inevitable [38].
556 Nevertheless, the existence of individuated multicellular
557 species such as *Turritopsis nutricula* demonstrates that
558 development can be reversed at least into juvenile
559 developmental stages [30] and that of the naked
560 mole-rat suggests that senescence is reversible in some
561 cases and negligible or close to negligible in others,
562 however exceptional.

563   The delicate balance between hologenic and epigenetic
564 information described here may shed light on the
565 well-known positive correlation between cancer incidence
566 and age [39]: if the senescent multicellular individual
567 attempts to correct its growing hologenic/epigenetic
568 content imbalance too strongly, it may elicit the onset
569 of cancer. Thus, age-related cancer would be the result
570 of a strong enough "pushback" from the multicellular

571 individual against its own senescence. Although the
572 specific dynamics that would underpin the "pushback"
573 are beyond the scope of this paper, this hypothesis is
574 indeed falsifiable by means of the following secondary
575 prediction: the observed log-ratio of non-epigenetic to
576 epigenetic histone crosstalk magnitude in the normal
577 (i.e., non-cancerous) cells closest to an age-related stage I
578 malignant tumor will be significantly lower than said
579 log-ratio observed in the other (i.e., tumor-nonadjacent)
580 normal cells of the same tissue. (Note: The falsification of
581 this secondary prediction does not imply the falsification
582 of the theory as a whole.)

583   In turn, the "pushback"-against-senescence hypothesis
584 for age-related cancer has, if correct, an implication we
585 should not overlook. Namely, stopping senescence and
586 eliminating the incidence of age-related cancer should be
587 one and the same technical challenge. In this respect,
588 it is worth noting that in the naked mole-rat both
589 senescence [34] and cancer incidence [40,41] have been
590 described as negligible or close to negligible.

591   Rozhok and DeGregori have recently highlighted
592 the explanatory limitations [42] of the Armitage-Doll
593 multistage model of carcinogenesis, which regards the
594 accumulation of genetic mutations as the cause of

age-related cancer [43]. They further argued that age-related cancer should rather be understood as a function of senescence-related processes [42]. However, their description of age-related cancer is based on Darwinian processes and thus differ from the account suggested here, which can be understood within the concept of teleodynamics [37,44]—a framework of biological individuality based on the emergence of intrinsic higher-order constraints, such as that described in the hologenic theory [17].

Apart from the proof of concept presented here, if the main prediction of this paper resists falsification attempts consistently, further research will be needed to elucidate the specific molecular dynamics embodying hologenic and epigenetic constraints within histone crosstalk completely. Such insights will be necessary to decide whether the hologenic/epigenetic information content imbalance can be corrected without compromising the multicellular individual's health or survival.

METHODS

Data collection

The genomic coordinates and associated transcript lengths of all annotated RefSeq mRNA TSSs for the hg19 (*Homo sapiens*) assembly were downloaded from the UCSC (University of California, Santa Cruz) database [45]. All ChIP-seq and RNA-seq data downloaded, processed, and analyzed in this work were generated by the Canadian Epigenetics, Epigenomics, Environment and Health Research Consortium (CEEHRC) initiative funded by the Canadian Institutes of Health Research (CIHR), Genome BC, and Genome Quebec. CEEHRC protocols and standards can be found at http://www.epigenomes.ca/protocols-and-standards, and specific details on ChIP-seq antibody validation can be found on this link. Further information about the CEEHRC and the participating investigators and institutions can be found at http://www.cihr-irsc.gc.ca/e/43734.html. For a full list of source data files with their respective URLs for downloading, see Supplementary Information.

Cell sample data sets in the CEEHRC database were selected based on the following criteria: (i) only data sets with associated age were included and (ii) among these data sets, the group (for both normal and cancer cells) that maximized the number of specific histone H3 modifications present in all data sets was chosen.

ChIP-seq datafile processing

The original ChIP-seq binary datafile format was bigWig. For mapping its ChIP-seq signal into the hg19 assembly, each datafile was processed with standard bioinformatics tools [46–48] in the following pipeline:

```
bigWigToWig → wig2bed --zero-indexed →
sort -k1,1 -k2,2n → bedtools map -o median
 -null 0 -a hg19_all_tss.bed/hg19_all_tss_control.bed
```

to generate an associated BED (Browser Extensible Data) file. (Note: The hg19_all_tss.bed file is a 200bp-per-bin BED reference file with no score values to perform the final ChIP-seq histone modification data mapping onto the 6,000bp-long TSS-adjacent genomic regions. The hg19_all_control.bed file is an analogous BED reference file for mapping the ChIP-seq input data onto 200-bp, 1-kbp, 5-kbp, and 10-kbp genomic windows, see ChIP-seq read profiles and normalization.)

ChIP-seq read profiles and normalization

To quantify and represent ChIP-seq read signal profiles for the histone H3 modifications, data were processed with the same method used in the EFilter multivariate algorithm [23] to predict mRNA levels with high accuracy ($R \sim 0.9$). Steps in this method comprise (i) dividing the genomic region from 2 kbp upstream to 4 kbp downstream of each TSS into 30 200-bp-long bins, in each of which ChIP-seq reads were later counted; (ii) dividing the read count signal for each bin by its corresponding control (ChIP-seq input) read density to minimize artifactual peaks; (iii) estimating the control read density within a 1-kbp window centered on each bin, if the 1-kbp window contained at least 20 reads; otherwise, a 5-kbp window, or else a 10-kbp window was used if the control reads were less than 20. When the 10-kbp length was insufficient, a pseudo-count value of 20 reads per 10 kbp was set as the control read density. This implies that the denominator (i.e., control read density) is at least 0.4 reads per bin.

RNA-seq datafile processing

For each strand in the DNA, original datafiles contained mRNA abundances in RPKM (reads per kilobase of transcript per million mapped reads) in bigWig format. These datafiles were thus processed analogously to the ChIP-seq datafiles, i.e., using the pipeline

```
bigWigToWig → wig2bed --zero-indexed →
sort -k1,1 -k2,2n → bedtools map -o median
 -null 0 -a refseq_pos.bed/refseq_neg.bed
```

to obtain associated BED files. (Note: The refseq_pos.bed and refseq_neg.bed files are BED reference files for each strand, with no score values, to perform the final RPKM calculation for each RefSeq mRNA in the hg19 assembly.)

When two or more mRNAs shared the same TSS (i.e., transcription start site with same genomic position and strand) the mean of the respective RPKM values was computed and associated with the corresponding TSS.

ChIP-seq/RNA-seq signal data tables

Using the RPKM values processed in this work, a subset TSS_{def} of all RefSeq mRNA TSSs displaying measured abundance (i.e., RPKM > 0) in all normal and cancer samples was determined. The number of TSSs in this subset TSS_{def} was 18,220, indicating that ∼70% of the 26,048 RefSeq mRNA TSSs annotated in the hg19 assembly had an associated mRNA abundance greater than zero in all (i.e., both normal and cancer) samples. The obtained TSS_{def} subset thus provided the data analysis with a common basis for all samples that comprises most protein-coding genes annotated in the human genome.

For each sample data entry, 30 genomic bins were defined and denoted by the distance (bp) between their 5′-end and their respective TSS_{def} genomic coordinate: "-2000", "-1800", "-1600", "-1400", "-1200", "-1000", "-800", "-600", "-400", "-200", "0" (TSS_{def} or '+1'), "200", "400", "600", "800", "1000", "1200", "1400", "1600", "1800", "2000", "2200", "2400", "2600", "2800", "3000", "3200", "3400", "3600", and "3800". Then, for each sample data entry, the ChIP-seq read signal was computed for all bins and for all histone modifications (30 bins×5 modifications=150 signal values) in all TSS_{def} genomic regions. Data input tables—comprising histone H3 modifications H3K4me1, H3K9me3, H3K27ac, H3K27me3, and H3K36me3—were thus generated for each sample entry as exemplified next:

$$
\#TSS_{def} \left\{ \begin{array}{cccccccc}
\text{H3K4me1\_-2000} & \ldots & \text{H3K36me3\_-2000} & \cdots & \text{H3K4me1\_3800} & \ldots & \text{H3K36me3\_3800} & \text{RPKM} \\
4.68 & \ldots & 7.94 & \cdots & 1.32 & \ldots & 12.15 & 35.63 \\
\vdots & & \vdots & & \vdots & & \vdots & \vdots \\
2.13 & \ldots & 4.97 & \cdots & 6.33 & \ldots & 3.06 & 17.44
\end{array} \right.
$$

The tables were then written to tab-delimited datafiles, which were subsequently classified into two groups: normal and cancer cells (see Table 1).

group	datafiles
normal (n = 18)	CEMT0032.nm.dat CEMT0033.nm.dat CEMT0034.nm.dat CEMT0040.nm.dat CEMT0042.nm.dat CEMT0044.nm.dat CEMT0050.nm.dat CEMT0051.nm.dat CEMT0052.nm.dat CEMT0053.nm.dat CEMT0054.nm.dat CEMT0055.nm.dat CEMT0056.nm.dat CEMT0057.nm.dat CEMT0058.nm.dat CEMT0059.nm.dat CEMT0060.nm.dat CEMT0061.nm.dat
cancer (n = 17)	CEMT0004.nm.dat CEMT0005.nm.dat CEMT0006.nm.dat CEMT0019.nm.dat CEMT0021.nm.dat CEMT0025.nm.dat CEMT0026.nm.dat CEMT0027.nm.dat CEMT0028.nm.dat CEMT0029.nm.dat CEMT0030.nm.dat CEMT0047.nm.dat CEMT0063.nm.dat CEMT0064.nm.dat CEMT0065.nm.dat CEMT0066.nm.dat CEMT0067.nm.dat

Table 1. Datafiles generated in this work containing normalized ChIP-seq signal values and RPKM values. The 'nm' suffix in the filename refers to the 'NM' RefSeq label for messenger RNAs, as opposed to non-coding RNAs.

Shannon measures of statistical uncertainty and statistical association

Shannon measures of statistical uncertainty and statistical association were used in this work in order to quantify histone H3 crosstalk at TSSs and its relationship with mRNA levels.

Statistical uncertainty

C.E. Shannon's seminal work, among other things, introduced the notion of—and a measure for—the uncertainty about discrete random variables [49]. For a discrete random variable X with probability mass function $P(X)$ its uncertainty (also known as Shannon entropy) is defined as

$$
H(X) := -\sum_{x \in X} P(x) \log_b[P(x)], \tag{4}
$$

where $P(x)$ is the probability of $X=x$ and b is the logarithm base. When $b=2$ (the base used in this work), the unit for this measure is the bit. $H(X)$ can also be interpreted as the amount of information necessary to resolve the uncertainty about the outcome of X. Shannon uncertainty was the measure used to estimate the uncertainty about the mRNA abundance level to be resolved in normal cells.

$H(X)$ is typically called marginal uncertainty because it involves only one random variable. In a multivariate scenario, the measure $H(X_1, \ldots, X_n)$ is called the joint uncertainty of the set of discrete random variables $\{X_1, \ldots, X_n\}$, and it is analogously defined as

$$
H(X_1, \ldots, X_n) := -\sum_{x \in X_1} \cdots \sum_{x \in X_n} P(x_1, \ldots, x_n) \log_b[P(x_1, \ldots, x_n)]. \tag{5}
$$

Another measure important to this work is the conditional uncertainty about a discrete random variable Y, with probability mass function $P(Y)$, given that the value of another discrete random variable X is known. This conditional uncertainty $H(Y|X)$ can be expressed as

$$
H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log_b \left[\frac{P(x)}{P(x, y)} \right], \tag{6}
$$

where $P(x, y)$ is the joint probability of $X=x$ and $Y=y$. Importantly, any measure of Shannon uncertainty (or any other derived Shannon measure) that is conditional on a random variable X can also be understood as said measure being explicitly unrelated to, or statistically independent from, the variable X.

Statistical association

A classic Shannon measure of statistical association of any two discrete random variables X and Y is that of mutual information I, defined as

$$I(X;Y) := -\sum_{x \in X}\sum_{y \in Y} P(x,y)\log_b\left[\frac{P(x,y)}{P(x)P(y)}\right] \quad (7)$$

$$= H(X) + H(Y) - H(X,Y) \quad (8)$$

$$= H(Y) - H(Y|X). \quad (9)$$

Note that if and only if X and Y are statistically independent then $I(X;Y)=0$, $H(X,Y)=H(X)+H(Y)$, and $H(Y|X)=H(Y)$. To analyze the magnitude of histone H3 crosstalk at TSSs, the two best known multivariate generalizations of mutual information were used in this work. The first is interaction information [50] or co-information [51], also symbolized by I, which is defined analogously to Eq. 8 for a set V of n discrete random variables as

$$I(V) := \sum_{U \subseteq V} (-1)^{|U|+1} H(U), \quad (10)$$

where $|U|$ is the cardinality (in this case, the number of random variables) of the subset U. In the case of interaction information I, Shannon uncertainty H is thus summed over all subsets of V (the uncertainty of the empty subset is $H(\varnothing) = 0$). Importantly, the interaction information of the random variables $\{X_1,\ldots,X_n\}$ can be decomposed with respect to another random variable Y as follows:

$$I(X_1;\ldots;X_n) = I(X_1;\ldots;X_n;Y) + I(X_1;\ldots;X_n|Y). \quad (11)$$

Interaction information $I(X_1;\ldots;X_n)$ captures the statistical association of all variables $\{X_1,\ldots,X_n\}$ taken at once, i.e., excluding all lower-order associations, and it can also take negative values in some cases. Interaction information was used in this work as a means to compute total correlation values.

To specifically quantify the magnitude of histone H3 crosstalk, the second multivariate generalization of mutual information used in this work was total correlation [24] (symbolized by C) or multiinformation [25], which is defined as

$$C(X_1,\ldots,X_n) := \left[\sum_{i=1}^n H(X_i)\right] - H(X_1,\ldots,X_n), \quad (12)$$

i.e., as the sum of the marginal uncertainties of the random variables $\{X_1,\ldots,X_n\}$ minus their joint uncertainty. Importantly, and unlike interaction information I, total correlation C captures all possible statistical associations including lower-order associations or, equivalently, all possible associations between any two or more random variables in the set $\{X_1,\ldots,X_n\}$. This is because the definition of interaction information I in Eq. 10 allows

total correlation C to be rewritten as a sum of quantities I for all possible combinations of variables in $\{X_1,\ldots,X_n\}$:

$$C(X_1,\ldots,X_n) = \sum_{i,j} I(X_i;X_j) + \sum_{i,j,k} I(X_i;X_j;X_k) + \ldots + I(X_1;\ldots;X_n). \quad (13)$$

This expression for total correlation C as a sum of interaction information quantities I along with the sum decomposition of I in Eq. 11 allows C to be decomposed also as a sum:

$$C(X_1,\ldots,X_n) = C_Y(X_1,\ldots,X_n,Y) + C(X_1,\ldots,X_n|Y), \quad (14)$$

where $C_Y(X_1,\ldots,X_n,Y)$ is the sum (analogous to that of Eq. 13) of all interaction information quantities I but now including the random variable Y in each combination of variables in $\{X_1,\ldots,X_n\}$, i.e.,

$$C_Y(X_1,\ldots,X_n,Y) = \sum_{i,j} I(X_i;X_j;Y) + \sum_{i,j,k} I(X_i;X_j;X_k;Y) + \ldots + I(X_1;\ldots;X_n;Y), \quad (15)$$

and where $C(X_1,\ldots,X_n|Y)$ is the sum of all conditional interaction information quantities I given Y for each combination of variables in $\{X_1,\ldots,X_n\}$, i.e.,

$$C(X_1,\ldots,X_n|Y) = \sum_{i,j} I(X_i;X_j|Y) + \sum_{i,j,k} I(X_i;X_j;X_k|Y) + \ldots + I(X_1;\ldots;X_n|Y). \quad (16)$$

For this work's purposes, total correlation C was chosen as the measure of statistical association to assess TSS-adjacent histone crosstalk because (i) C is non-negative and thus easier to interpret conceptually, (ii) C is equal to zero if and only if all random variables it comprises are statistically independent, (iii) C captures all possible associations up to a given number of variables (in this work, position-specific histone modification levels) and, (iv) C can be decomposed, as shown in Eq. 14, as a sum of two C quantities: one explicitly related to a certain variable Y and the other explicitly unrelated to Y. Property (iv) was useful to decompose the overall histone crosstalk as a sum of an epigenetic and other non-epigenetic component (see Introduction).

An additional Shannon measure of statistical association was used to assess the predictive power of TSS-adjacent histone modification levels on mRNA abundance levels (such power has already been used to predict mRNA levels with high accuracy [23]). The uncertainty coefficient U [52] is defined as

$$U(Y|X_1,\ldots,X_n) := \frac{H(Y) - H(Y|X_1,\ldots,X_n)}{H(Y)}, \quad (17)$$

i.e., $U(Y|X_1,\ldots,X_n)$ is the relative decrease in uncertainty about Y when $\{X_1,\ldots,X_n\}$ are known—or, equivalently, the fraction of bits in Y that can be predicted by $\{X_1,\ldots,X_n\}$—and it can take values from 0 to 1. $U(Y|X_1,\ldots,X_n)=0$ implies the set $\{X_1,\ldots,X_n\}$ has no predictive power on Y, whereas $U(Y|X_1,\ldots,X_n)=1$ implies $\{X_1,\ldots,X_n\}$ can predict Y completely.

12

Levels of possible statistical associations when assessing histone crosstalk magnitudes

An important aspect of quantifying the epigenetic and non-epigenetic histone crosstalk components is the specific range of possible statistical associations. In other words, the choice of the number n of TSS-adjacent, position-specific histone H3 modification levels when computing $C_Y(X_1, \ldots, X_n, Y)$ and $C(X_1, \ldots, X_n | Y)$. To this end, the minimal n able to predict mRNA levels significantly and non-redundantly—which corresponds to the level of histone crosstalk able to convey a non-neglectable amount of epigenetic information content—was first determined. This value is straightforward to assess using the uncertainty coefficient $U(Y | X_1, \ldots, X_n)$, where Y represents mRNA levels.

In effect, $U(Y | X_i, X_j)$ (i.e., where $n=2$) quantifies the predictive power of pairs of position-specific histone modification levels, $U(Y | X_i, X_j, X_k)$ quantifies the predictive power of triads, etc. $U(Y | X_1, \ldots, X_n)$ values were thus computed for singletons, pairs, triads, and tetrads. Singeltons were calculated for descriptive purposes only, because histone crosstalk is not measurable for them. On average, a triad (i.e., when $n=3$) of position-specific histone H3 modification levels was found to have (i) significant predictive power on mRNA levels ($U(Y | X_i, X_j, X_k)=0.63$) and, importantly, (ii) at least 2.3 times more predictive power than all possible singletons (3) and pairs (3) that exist within a triad taken together, i.e.,

$$\frac{U(Y | X_i, X_j, X_k)}{\sum_i U(Y | X_i) + \sum_{i,j} U(Y | X_i, X_j)} \geq 2.3, \qquad (18)$$

a phenomenon known as synergy of a set of predictor variables [53] (see Table 2).

level	set	$U(Y|V)$ [mean]	κ	$U(Y|V)$ measured in
singleton	$V=\{X_i\}$	0.02	154	all 150 singletons
pair	$V=\{X_i, X_j\}$	0.07	441	all 11,175 pairs
triad	$V=\{X_i, X_j, X_k\}$	0.63	325	all 551,300 triads
tetrad	$V=\{X_i, X_j, X_k, X_l\}$	0.95	379	50,000 random tetrads

Table 2. **Predictive power (quantified as $U(Y|V) \in [0, 1]$) of different sets $V=\{X_1, \ldots, X_n\}$ of TSS-adjacent, position-specific histone H3 modification levels on mRNA levels (represented by Y) in normal cells.** For each instance of the set V, $U(Y|V)$ was averaged over the 18 normal cell samples analyzed. The distribution of $U(Y|V)$ in the interval $[0, 1]$ for each family V of sets (singletons, pairs, triads, and tetrads) was parameterized and described here in terms of the mean $\bar{x}$ and the concentration parameter κ, which are derived from the respective beta distribution Beta(α, β), with $\kappa=\hat{\alpha}+\hat{\beta}$ [54]. Point estimators $\hat{\alpha}$ and $\hat{\beta}$ were computed using the method of moments, i.e., $\hat{\alpha} = \bar{x}^2 \left(\frac{1-\bar{x}}{\sigma^2} - \frac{1}{\bar{x}} \right)$ and $\hat{\beta} = \hat{\alpha} \left(\frac{1}{\bar{x}} - 1 \right)$, where σ^2 is the variance.

Pairs (i.e., when $n=2$) were also found to possess predictive synergy, but this synergy is smaller than that found for triads $\left(1.75 \leq \frac{U(Y | X_i, X_j)}{\sum_i U(Y | X_i)} < 2.3 \right)$. The average predictive power of pairs on mRNA levels is also substantially lower ($U(Y | X_i, X_j)=0.07$). On the other hand, tetrads (i.e., when $n=4$) were found to have high predictive power ($U(Y | X_i, X_j, X_k, X_l)=0.95$) but they possess no synergy whatsoever and display instead what is called redundancy [53]. Based on previous work [23], high predictive power on mRNA levels and yet no synergy are thus expected to happen with a large enough n. From all possible singletons (4), pairs (6), and triads (4) that exist within a tetrad, the explanatory power on mRNA levels of a non-redundant set of only one triad and five pairs already exceeds the explanatory power of the tetrad.

(Note: In previous work it has been argued that RPKM may not always be a suitable unit of mRNA abundance when studying differential gene expression. Specifically, it was shown that, if transcript size distribution varies significantly among the samples, RPKM might introduce significant biases [55]. To overcome this problem, an alternative abundance unit TPM (transcripts per million)—which is an invertible linear transformation of the RPKM value for each sample—was introduced [55]. Nonetheless, this issue was not a problem for the present work because Shannon measures are invariant under any invertible transformation of the discrete random variables.)

Theoretical methods

The elaboration of the main falsifiable prediction took into account two observations for human primary cells in this work. Namely, (i) the uniqueness of triads of position-specific histone modification levels in terms of significant predictive power and predictive synergy and (ii) the *post hoc* result that triads constitute precisely the level n at which the predicted correlation between the non-epigenetic/epigenetic histone H3 crosstalk log-ratio and cell donor age actually exists. In this way, the main prediction was formulated with explicit dependence on the level of scale: "For any given tissue in any individuated multicellular species a positive correlation between the non-epigenetic/epigenetic histone H3 crosstalk log-ratio and cell donor age will be observed at the level n of histone crosstalk that possesses both significant predictive power and predictive synergy on mRNA levels."

Statistical tests

The statistical significance of each Pearson correlation coefficient r obtained was assessed using the statistic t defined as

$$t := r \sqrt{\frac{n-2}{1-r^2}}, \qquad (19)$$

which is known to follow a Student's t-distribution with $n-2$ degrees of freedom, and where n is the number of data pairs [56]. For the hypothesized positive correlation between the non-epigenetic/epigenetic

histone H3 crosstalk log-ratio and age, the statistical null hypothesis was tested against the alternative hypothesis that the correlation is greater than zero (i.e., one-sided Student's t-test). For the hypothesized non-significant correlation between the overall histone H3 crosstalk magnitude and age, the statistical null hypothesis was tested against the alternative hypothesis that the correlation is greater or less than zero (i.e., two-sided Student's t-test).

On the other hand, the distribution of correlation coefficients (r) is known to be non-Gaussian [57], which can be easily appreciated in Fig. **3**. For this reason, the statistical comparison of r for normal cells (neonate data point excluded) and cancer cells was performed using the non-parametric Mann-Whitney U test [58].

Correction for multiple testing

The analysis of histone crosstalk involved 5 histone H3 modifications $\times$ 30 genomic bins = 150 TSS-adjacent, position-specific histone H3 modification levels. Thus, assessing the statistical significance of the correlation values involved a large number of tests of the null hypothesis (for triads, tetrads, and pairs) under general dependence. This dependence derives from the fact that different histone modification levels are known to be highly correlated (this is the phenomenon of histone crosstalk itself).

The resampling-based procedure by Benjamini and Yekutieli [59] provides control of the false discovery rate (FDR) [60] under general dependence conditions. This was the method thus used in this work in order to correct for multiple testing.

Code availability

Standard bioinformatics tools [46–48] and the Perl language were used to process the ChIP-seq and RNA-seq source data and to generate the *.nm.dat files displayed in Table **1**. The R software [61] and its infotheo package [62] were used for the computation of Shannon measures of statistical uncertainty and statistical association from the *.nm.dat files. Marginal and joint Shannon uncertainties and all the other derived Shannon measures were computed using maximum likelihood (ML) estimation [63] and bias-corrected with the Miller-Madow method [64]. All the R code and the *.nm.dat files necessary for a full reproduction of the results are available as Supplementary Information.

Sample metadata

sample ID	disease status	cell donor age	sex	cell type/tissue
CEMT0032	normal	0	N/A	hematopoietic (cord)
CEMT0033	normal	82	F	colon
CEMT0034	normal	73	M	colon
CEMT0040	normal	67	F	thyroid
CEMT0042	normal	46	F	thyroid
CEMT0044	normal	55	M	thyroid
CEMT0050	normal	53	M	colon
CEMT0051	normal	67	M	colon
CEMT0052	normal	72	M	colon
CEMT0053	normal	81	F	colon
CEMT0054	normal	90+	F	colon
CEMT0055	normal	90+	F	colon
CEMT0056	normal	84	F	colon
CEMT0057	normal	84	F	colon
CEMT0058	normal	56	F	colon
CEMT0059	normal	56	F	colon
CEMT0060	normal	77	M	colon
CEMT0061	normal	77	M	colon
CEMT0004	cancer	74	M	peripheral blood
CEMT0005	cancer	68	F	peripheral blood
CEMT0006	cancer	60	M	peripheral blood
CEMT0019	cancer	46	F	brain
CEMT0021	cancer	35	M	brain
CEMT0025	cancer	62	F	peripheral blood
CEMT0026	cancer	79	M	peripheral blood
CEMT0027	cancer	71	M	peripheral blood
CEMT0028	cancer	79	F	peripheral blood
CEMT0029	cancer	47	M	colon
CEMT0030	cancer	56	F	peripheral blood
CEMT0047	cancer	43	M	brain
CEMT0063	cancer	81	F	colon
CEMT0064	cancer	90+	F	colon
CEMT0065	cancer	84	F	colon
CEMT0066	cancer	56	F	colon
CEMT0067	cancer	77	M	colon

Table 3. **Metadata for each primary cell sample analyzed.** Note: Age entries originally tabulated as 90+ were entered as 90 into the computational analysis. Metadata source: CEEHRC.

14

REFERENCES

[1] S. Katz, Imagining the life-span: from premodern miracles to postmodern fantasies, in: Images of Aging: Cultural Representations of Later Life, Routledge, 1995, pp. 59–74.

[2] J. Vijg, J. Campisi, Puzzles, promises and a cure for ageing, Nature 454 (2008) 1065–1071. DOI:10.1038/nature07216.

[3] J. W. Shay, W. E. Wright, Hayflick, his limit, and cellular ageing, Nat. Rev. Mol. Cell Biol. 1 (2000) 72–76. DOI:10.1038/35036093.

[4] L. Hayflick, P. S. Moorhead, The serial cultivation of human diploid cell strains, Exp. Cell Res. 25 (3) (1961) 585–621. DOI:10.1016/0014-4827(61)90192-6.

[5] C. B. Harley, A. B. Futcher, C. W. Greider, Telomeres shorten during ageing of human fibroblasts, Nature 345 (1990) 458–460. DOI:10.1038/345458a0.

[6] R. C. Allsopp, H. Vaziri, C. Patterson, S. Goldstein, E. V. Younglai, A. B. Futcher, C. W. Greider, C. B. Harley, Telomere length predicts replicative capacity of human fibroblasts, Proc. Natl. Acad. Sci. U.S.A. 89 (21) (1992) 10114–10118. DOI:10.1073/pnas.89.21.10114.

[7] V. J. Cristofalo, R. G. Allen, R. J. Pignolo, B. G. Martin, J. C. Beck, Relationship between donor age and the replicative lifespan of human cells in culture: a reevaluation, Proc. Natl. Acad. Sci. U.S.A. 95 (18) (1998) 10614–10619. DOI:10.1073/pnas.95.18.10614.

[8] A. Lorenzini, M. Tresini, S. N. Austad, V. J. Cristofalo, Cellular replicative capacity correlates primarily with species body mass not longevity, Mech. Ageing Dev. 126 (10) (2005) 1130–1133. DOI:10.1016/j.mad.2005.05.004.

[9] C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging, Cell 153 (6) (2013) 1194–1217. DOI:10.1016/j.cell.2013.05.039.

[10] K. Jin, Modern biological theories of aging, Aging Dis. 1 (2) (2010) 72–74. Available from: http://www.aginganddisease.org/EN/Y2010/V1/I2/72.

[11] M. V. Blagosklonny, Aging is not programmed: genetic pseudo-program is a shadow of developmental growth, Cell Cycle 12 (24) (2013) 3736–3742. DOI:10.4161/cc.27188.

[12] S. Horvath, K. Raj, DNA methylation-based biomarkers and the epigenetic clock theory of ageing, Nat. Rev. Genet. (2018) DOI:10.1038/s41576-018-0004-3.

[13] R. W. Powers 3rd, M. Kaeberlein, S. D. Caldwell, B. K. Kennedy, S. Fields, Extension of chronological life span in yeast by decreased TOR pathway signaling, Genes Dev. 20 (2) (2006) 174–184. DOI:10.1101/gad.1381406.

[14] T. Vellai, K. Takacs-Vellai, Y. Zhang, A. L. Kovacs, L. Orosz, F. Müller, Genetics: Influence of TOR kinase on lifespan in *C. elegans*, Nature 426 (2003) 620. DOI:10.1038/426620a.

[15] K. Popper, The Logic of Scientific Discovery, Routledge, 2005.

[16] G. Ellis, J. Silk, Scientific method: defend the integrity of physics, Nat. News 516 (7531) (2014) 321–323. Available from: http://www.nature.com/news/scientific-method-defend-the-integrity-of-physics-1.16535.

[17] F. A. Veloso, On the developmental self-regulatory dynamics and evolution of individuated multicellular organisms, J. Theor. Biol. 417 (2017) 84–99. DOI:10.1016/j.jtbi.2016.12.025.

[18] P. W. Atkins, The Second Law, Scientific American Library, 1984.

[19] J.-S. Lee, E. Smith, A. Shilatifard, The language of histone crosstalk, Cell 142 (5) (2010) 682–685. DOI:10.1016/j.cell.2010.08.011.

[20] Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, W. Peng, M. Q. Zhang, K. Zhao, Combinatorial patterns of histone acetylations and methylations in the human genome, Nat. Genet. 40 (7) (2008) 897–903. DOI:10.1038/ng.154.

[21] A. Lennartsson, K. Ekwall, Histone modification patterns and epigenetic codes, Biochim. Biophys. Acta - Gen. Subj. 1790 (9) (2009) 863–868. DOI:10.1016/j.bbagen.2008.12.006.

[22] V. E. A. Russo, R. A. Martienssen, A. D. Riggs, Epigenetic Mechanisms of Gene Regulation, Cold Spring Harbor Laboratory Press, 1996.

[23] V. Kumar, M. Muratani, N. A. Rayan, P. Kraus, T. Lufkin, H. H. Ng, S. Prabhakar, Uniform, optimal signal processing of mapped deep-sequencing data, Nat. Biotechnol. 31 (7) (2013) 615–622. DOI:10.1038/nbt.2596.

[24] S. Watanabe, Information theoretical analysis of multivariate correlation, IBM J. Res. Dev. 4 (1) (1960) 66–82. DOI:10.1147/rd.41.0066.

[25] M. Studený, J. Vejnarova, The multiinformation function as a tool for measuring stochastic dependence, in: Learning in Graphical Models, 1998, pp. 261–297. DOI:10.1007/978-94-011-5014-9_10.

[26] G. Csárdi, A. Franks, D. S. Choi, E. M. Airoldi, D. A. Drummond, Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast, PLoS Genet. 11 (5) (2015) e1005206. DOI:10.1371/journal.pgen.1005206.

[27] S. E. Prochnik, J. Umen, A. M. Nedelcu, A. Hallmann, S. M. Miller, I. Nishii, P. Ferris, A. Kuo, T. Mitros, L. K. Fritz-Laylin, U. Hellsten, J. Chapman, O. Simakov, S. A. Rensing, A. Terry, J. Pangilinan, V. Kapitonov, J. Jurka, A. Salamov, H. Shapiro, J. Schmutz, J. Grimwood, E. Lindquist, S. Lucas, I. V. Grigoriev, R. Schmitt, D. Kirk, D. S. Rokhsar, Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*, Science 329 (5988) (2010) 223–226. DOI:10.1126/science.1188800.

[28] K. Malik, K. W. Brown, Epigenetic gene deregulation in cancer, Br. J. Cancer 83 (12) (2000) 1583–1588. DOI:10.1054/bjoc.2000.1549.

[29] T. J. Gonda, R. G. Ramsay, Directly targeting transcriptional dysregulation in cancer, Nat. Rev. Cancer 15 (11) (2015) 686–694. DOI:10.1038/nrc4018.

[30] S. Piraino, F. Boero, B. Aeschbach, V. Schmid, Reversing the life cycle: medusae transforming into polyps and cell transdifferentiation in *Turritopsis nutricula* (Cnidaria, Hydrozoa), Biol. Bull. 190 (3) (1996) 302–312. DOI:10.2307/1543022.

[31] C. E. Finch, Update on slow aging and negligible senescence–a mini-review, Gerontology 55 (3) (2009) 307–313. DOI:10.1159/000215589.

[32] R. M. Lanner, K. F. Connor, Does bristlecone pine senesce?, Exp. Gerontol. 36 (4-6) (2001) 675–685. DOI:10.1016/S0531-5565(00)00234-5.

[33] R. Schaible, A. Scheuerlein, M. J. Dańko, J. Gampe, D. E. Martínez, J. W. Vaupel, Constant mortality and fertility over age in *Hydra*, Proc. Natl. Acad. Sci. U.S.A. 112 (51) (2015) 15701–15706. DOI:10.1073/pnas.1521002112.

[34] J. G. Ruby, M. Smith, R. Buffenstein, Naked mole-rat mortality rates defy gompertzian laws by not increasing with age, Elife 7 (2018) e31157. DOI:10.7554/eLife.31157.

[35] D. Gems, L. Partridge, Genetics of longevity in model organisms: debates and paradigm shifts, Annu. Rev. Physiol. 75 (1) (2013) 621–644. DOI:10.1146/annurev-physiol-030212-183712.

[36] C. Darwin, On the Origin of Species, 1859, Routledge, 2004.

[37] T. W. Deacon, Incomplete Nature: How Mind Emerged from Matter, WW Norton & Company, 2011.

[38] P. Nelson, J. Masel, Intercellular competition and the inevitability of multicellular aging, Proc. Natl. Acad. Sci. U.S.A. 114 (49) (2017) 12982–12987. DOI:10.1073/pnas.1618854114.

[39] F. Kamangar, G. M. Dores, W. F. Anderson, Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world, J. Clin. Oncol. 24 (14) (2006) 2137–2150. DOI:10.1200/JCO.2005.05.2308.

[40] R. Buffenstein, Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species, J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 178 (4) (2008) 439–445. DOI:10.1007/s00360-007-0237-5.

[41] M. A. Delaney, L. Nagy, M. J. Kinsel, P. M. Treuting, Spontaneous histologic lesions of the adult naked mole rat (*Heterocephalus glaber*): a retrospective survey of lesions in a zoo population, Vet. Pathol. 50 (4) (2013) 607–621. DOI:10.1177/0300985812471543.

[42] A. I. Rozhok, J. DeGregori, The evolution of lifespan and age-dependent cancer risk, Trends in Cancer 2 (10) (2016) 552–560. DOI:10.1016/j.trecan.2016.09.004.

[43] P. Armitage, R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis, Br. J. Cancer 8 (1) (1954) 1–12. DOI:10.1038/bjc.1954.1.

[44] T. W. Deacon, A. Srivastava, J. A. Bacigalupi, The transition from constraint to regulation at the origin of life, Front. Biosci. 19 (2014) 945–957. DOI:10.2741/4259.

[45] D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, W. J. Kent, The UCSC Table Browser data retrieval tool, Nucleic Acids Res. 32 (Database issue) (2004) D493–D496. DOI:10.1093/nar/gkh103.

[46] W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: enabling browsing of large distributed datasets, Bioinformatics 26 (17) (2010) 2204–2207. DOI:10.1093/bioinformatics/btq351.

[47] S. Neph, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J. A. Stamatoyannopoulos, BEDOPS: high-performance genomic feature operations, Bioinformatics 28 (14) (2012) 1919–1920. DOI:10.1093/bioinformatics/bts277.

[48] A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (2010) 841–842. DOI:10.1093/bioinformatics/btq033.

[49] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423. DOI:10.1002/j.1538-7305.1948.tb01338.x.

[50] W. McGill, Multivariate information transmission, Trans. IRE Prof. Gr. Inf. Theory 4 (4) (1954) 93–111. DOI:10.1109/tit.1954.1057469.

[51] A. J. Bell, The co-information lattice, in: Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA, Vol. 2003, 2003.

[52] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, Cambridge Univ. Press, 1992.

[53] D. Chicharro, S. Panzeri, Synergy and redundancy in dual decompositions of mutual information gain and information loss, Entropy 19 (71) (2017) 1–29. DOI:10.3390/e19020071.

[54] J. Kruschke, Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Academic Press, 2014.

[55] G. P. Wagner, K. Kin, V. J. Lynch, Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples, Theory Biosci. 131 (4) (2012) 281–285. DOI:10.1007/s12064-012-0162-3.

[56] R. E. Walpole, Essentials of Probability & Statistics for Engineers & Scientists, Pearson, 2013.

[57] F. Kenney, E. S. Keeping, Mathematics of Statistics: Part Two, D. Van Nostrand Company, Inc., 1951.

[58] H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1) (1947) 50–60. DOI:10.1214/aoms/1177730491.

[59] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. (2001) 1165–1188. Available from: http://www.jstor.org/stable/2674075.

[60] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B 50 (1) (1995) 289–300. Available from: http://www.jstor.org/stable/2346101.

[61] R Core Team, R: A Language and Environment for Statistical Computing (2017). Available from: https://www.r-project.org/.

[62] P. E. Meyer, infotheo: Information-Theoretic Measures (2014). Available from: https://cran.r-project.org/package=infotheo.

[63] L. Paninski, Estimation of entropy and mutual information, Neural Comput. 15 (6) (2003) 1191–1253. DOI:10.1162/089976603321780272.

[64] G. A. Miller, Note on the bias of information estimates, in: Information Theory in Psychology: Problems and Methods II-B, The Free Press of Glencoe, 1955, pp. 95–100.