



Review

Biohorology and biomarkers of aging: Current state-of-the-art, challenges and opportunities

Fedor Galkin^{a,b}, Polina Mamoshina^c, Alex Aliper^a, João Pedro de Magalhães^{b,e},
Vadim N. Gladyshev^d, Alex Zhavoronkov^{a,e,f,*}

^a InSilico Medicine, Science Park, Hong Kong

^b Integrative Genomics of Ageing Group, Institute of Ageing and Chronic Disease, University of Liverpool, UK

^c Deep Longevity, Science Park, Hong Kong

^d Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

^e Biogerontology Research Foundation, London, UK

^f Buck Institute for Research on Aging, Novato, USA

ARTICLE INFO

Keywords:

Aging
Biogerontology
Aging clock
Deep learning
Neural network

ABSTRACT

The aging process results in multiple traceable footprints, which can be quantified and used to estimate an organism's age. Examples of such aging biomarkers include epigenetic changes, telomere attrition, and alterations in gene expression and metabolite concentrations. More than a dozen aging clocks use molecular features to predict an organism's age, each of them utilizing different data types and training procedures. Here, we offer a detailed comparison of existing mouse and human aging clocks, discuss their technological limitations and the underlying machine learning algorithms. We also discuss promising future directions of research in biohorology — the science of measuring the passage of time in living systems. Overall, we expect deep learning, deep neural networks and generative approaches to be the next power tools in this timely and actively developing field.

1. Introduction

The XXth century has seen the rise of multiple aging theories. Today, we have the means to inspect and manipulate biological systems with precision unavailable to our predecessors, yet the mystery of aging remains unsolved. Neither of the existing theories has been decisively proven and elaborated to the point of clear understanding of what the ultimate cause of aging is. Putting together a conclusive theory of aging has been difficult due to the inability to properly quantify and define aging. Consequently, the efficacy of various geroprotective interventions remains subject to controversy. Without general agreement as to what constitutes aging and biological age (BA), and how to measure their progression, conclusions on the benefits of particular therapies are likely to be biased.

Meanwhile, the very existence of a reliable way to measure BA remains under question. August Weismann proposed in 1881 the existence of an evolved mechanism of aging, which is selected for at the group level and facilitates resource redistribution within species from the elderly to the young. He also suggested that such a mechanism of “programmed death” is probably realized via an intrinsic, species-specific limit of somatic cell division. Indeed, the limit originally

conceptualized by Weismann can be related to the Hayflick limit. While it would be infinitely convenient for aging to be akin to developmental processes, such as puberty, and possess specific checkpoints and well-established regulatory pathways, subsequent studies proved Weismann wrong (Gavrilov and Gavrilova, 2002). Firstly, programmed aging contains an implicit contradiction with observations, since it requires group selection for elderly elimination to be stronger than individual selection for increased lifespan. Secondly, in order for the mechanism to come into place, natural populations should contain a significant fraction of old individuals, which is not observed either (Williams, 1957). Finally, if aging is a program, it may be disrupted by gene mutations, but this has not been observed in any individuals and in any species (Gladyshev, 2016). All this does not exclude a possibility of programmed aging in some species/contexts, but contradicts the universal nature of programmed aging.

One alternative to the Weismann's idea of programmed death would be aging as an evolutionary neutral or antagonistically pleiotropic trait that still presents a single point of attack to biogerontologists. This hypothesis is also referred to as “shortsighted watchmaker”. Processes that are required for an organism to reach maturity persist until they become a liability due to weaker selection in late life. Eventually these

* Corresponding author at: Buck Institute for Research on Aging, Novato, USA.

E-mail address: alex@insilicomedicine.com (A. Zhavoronkov).

<https://doi.org/10.1016/j.arr.2020.101050>

Received 14 September 2019; Received in revised form 6 February 2020; Accepted 22 March 2020

Available online 06 April 2020

1568-1637/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

processes result in nothing but harm we interpret as aging symptoms (De Magalhães, 2012). This hypothesis can be illustrated by the “Golden Antelope” fairy tale, in which a greedy raja captures a magical antelope that strikes golden coins with each step. At first the raja is glad to see his prisoner galloping to enrich him more and more, right until he starts getting buried under the weight of the coins. But alas, the antelope cannot be stopped (Abramov and Atamanov, 1954). Various mechanistic theories of aging proposed many different antelopes, among them: ROS, telomere attrition and inflammaging. These theories can serve a surprisingly fruitful basis for the development of anti-aging interventions. For example, ROS theory of aging has yielded mitochondria targeted SkQ antioxidants that have been proven to protect organisms in a variety of age-related scenarios (Skulachev et al., 2009). However, none of the reductionist theories of aging has managed to provide a cogent explanation on how a single factor gives rise to all the distinctive aging phenotypes.

This presents the research community with a troubling alternative that aging has no distinct genetic signature and is in essence a multitude of simultaneous damage accumulation processes. If that is true, BA as a concept is unlikely to be a property of objective reality but should be treated as an artificial construct. Such a construct would heavily rely on socioeconomic and cultural differences among people as well as on the scientific consensus on what detrimental processes should be considered aging. Without such a consensus, gerontology may be continuously torn apart, as scientists claiming to study the very same “aging” will in fact be elaborating on its fundamentally different aspects — a situation known as “semantic barrier”.

If there is indeed no singular process behind all the manifestations of aging, measuring BA is infinitely harder than in the case of single-source aging. It would require either (a) finding a common denominator for the majority of aging-related processes or (b) arbitrarily weighing all such processes according to their perceived importance to form the final “age score”. The problem of multiple aging processes is further confounded by individual variability. There are indications that people age according to different trends that can be grouped into several “ageotypes” defined by the hierarchical clustering of their biomarkers. Moreover, the fluidity of individual ageotypes can cause unstable performance of an aging score within any singular individual (Ahadi et al., 2020).

Nonetheless, a reliable and universally agreed upon way to quantify BA is a necessity for modern biogerontology. Most importantly, it would enable flexible experimental designs and in many cases would remove the need to follow up human subjects for decades to evaluate the benefits of a geroprotective intervention. Moreover, it could be used as a criterion to test the relevance of specific diseases, pathways or processes in the context of aging research. Hopefully, accurate BA measurement could bring around new hypotheses on the nature of aging and be the first step towards a paradigm shift in biogerontology.

Today there is no shortage of either biomarkers – ranging from DNA methylation (DNAm) to neuroimaging – or methods to translate various interpretations of BA — the so-called aging clocks (Gialluisi et al., 2019; Xia et al., 2017). This review is focused only on the clocks constructed using molecular or cellular-level data. First, we provide information on what aspects of aging they cover, their technological challenges, user availability and a short historical reference as well as compare their performance. Next, we discuss more technical aspects of biohorology – the science of measuring BA – and provide a short overview of state-of-the-art deep learning techniques that could be used to increase our understanding of aging processes and to design geroprotective interventions.

2. Biomarkers of aging and existing aging clocks

Multiple features and traits within an organism, on all levels of biological organisation, undergo transformation during aging and can be called biomarkers of aging. In this review, we will cover the subset of

biomarkers, which satisfy the following criterion: to qualify for an aging clock, a biomarker has to be a reflection of a ubiquitous aging related process, present both in humans and model organisms as well as in their multiple tissues (Butler et al., 2004; Johnson, 2006; Jylhävä et al., 2017). This criterion rules out, for instance, all visual biomarkers of aging, which nonetheless can achieve formidable precision as demonstrated by PhotoAgeClock (Bobrov et al., 2018), whose mean absolute error (MAE) of 2.3 years remains unparalleled by any of the clocks described here in more detail. More generally, this requirement implies that universal aging clocks should be based on cellular or subcellular level biomarkers since aging processes that take place on higher orders of organization may be incomparable between tissues or species. For example, left ventricular filling rate depends on age and could be used to estimate the likelihood of a fatal stroke (Strait and Lakatta, 2012), yet this organ-level feature cannot be used to estimate the BA of other organs or a whole organism (yet may be useful when placed in the context of other biomarkers). Moreover, organ- or tissue-specific aging biomarkers that require invasive procedures (e.g. biopsy for histological analysis) impose multiple limitations on experimental settings and applications.

While complex organisms accumulate innumerable footprints of aging, very few of them have the potential to become a biomarker that could be used to measure the passage of time across different tissues and species. In this review, we will focus on such aging biomarkers: telomere length, genomic instability, epigenetic marks, biochemical compounds and gene expression levels.

2.1. Defining the target

Defining aging clock as a method to predict an individual's age is unlikely to cause much discord in the scientific community. Meanwhile, the definition of age per se and what is it that should be predicted may lead to a heated argument.

The two concepts: chronological and biological ages (CA and BA) — intertwine and are sometimes used interchangeably in the literature, implicitly and explicitly. The definition of CA is trivial: the amount of time passed since one's birth (or inception in case of gestational age). Meanwhile BA is a fluid, borderline placeholder concept used to refer to the time-dependent component of an organism's overall health condition and is frequently juxtaposed with CA. One might quite reasonably doubt the necessity of such a concept that creates semantic barriers.

The long unsettled question of “When aging starts?” is of great utility for demonstrating the confusion surrounding BA. This question has many different answers, and each of them implies a different definition of what aging and BA are. Some prefer to use the BA defined by mortality risk and claim that it is most practical to consider aging onset to coincide with the Gompertz (age-mortality) curve minimum, which approximately corresponds to the start of reproductive maturity period (Dolejs and Marešová, 2017). Others point out that this definition of aging and biological age as a substitute to mortality produces a circular argument: aging starts at the minimum of the mortality curve, and the onset of aging then causes mortality to increase. Some scientists believe that mortality minimum and the age of reproductive capacity onset coinciding implies the existence of a deeper evolutionary connection between the two. But their fluctuations across centuries and nations show that this is unlikely the case (Milne, 2006). The supporters of “aging as damage accumulation” concept suggest that BA starts earlier than that, as even embryos are not immune to oxidation, somatic mutations, telomere attrition and other forms of damage (Milne, 2006). Indeed, if aging is defined by age-related deleterious changes, such as molecular damage, it begins very early in life. A more physiological view on aging based on the intensity of bodily functions puts BA starting point at 25–35 years, when many age-related conditions first manifest: decrease in reproductive function, sarcopenia, onset of cardio-vascular diseases, reduced neurogenesis, etc (Forbes and Reina, 1970; Kempermann, 2015; Nelson et al., 2013; Strait and Lakatta,

2012).

All these points of view imply different definitions of BA and its ticking rate at different life periods, yet BA is often referred to as a universally agreed upon metrics of aging. It is highly important to not fall for this, and always pay attention to how exactly the authors derive BA. In the field of biohorology BA has to be quantifiable, which adds a new layer of mathematical ambiguity. Most aging clocks base their BA definitions either on CA or mortality risk, with variations discussed in the sections below. Mortality risk in its turn is derived from demographic tables and can be assumed to be a function of CA in most animals, including human. Thus, aging clocks are ultimately treating CA as a substitute BA with the caveat that deviations from the actual CA signify better or worse physical fitness when compared to age matched controls.

Such a design has several flaws. First of all, it ties the produced age predictions to the historical and geographical context it was created in (Levine and Crimmins, 2018). Some clocks use extra metavariables during training and prediction to cover for the differences in the aging patterns among different nations, races and countries (Mamoshina et al., 2018a).

Another concern is the clock's biological relevance. A clock that predicts CA reliably will fail to grasp the health status variance, especially in the elderly, when the division between "normal agers", "underagers" and "overagers" becomes apparent. Some aging clocks acknowledge this drawback and incorporate multiple measures of physical fitness apart from CA to derive BA, e.g. PhenoAge discussed in Section 2.1 (Levine et al., 2018). In other cases, aging clock developers determine their biological relevance post-factum. The most widely used DNA methylation clocks (Section 2.4) owe their popularity to producing BA estimates that are highly correlated with physical and cognitive fitness in the elderly, as well as an extensive list of age-related diseases, cancers, overall frailty and biomarkers of aging (Degerman et al., 2017; Marioni et al., 2015).

These examples prove that CA-based BA definitions are capable of grasping the whole scale of interpersonal BA variance. One should bear in mind that this circumstance does not certify this approach's unconditional efficiency. Measuring different aspects of aging notwithstanding, the clocks still inherit all their training set's biases, requiring careful sample selection. They rely greatly on the definitions, such as "healthy" and "disease", as well as on data annotation. In many publicly available data sets originating from case-control studies, it is quite common for the entries to contain no information on comorbidities. As a result, control samples in, for example, a study on Alzheimer's disease may in fact be a wild mix of diabetics, cancer survivors, subjects with cardio-vascular conditions and other considerably unhealthy individuals. One may assume that the relative number of subjects with serious diseases is typically low and their effect on an aging clock performance is minimal. But then it is important to make sure that all CA groups in the training set are evenly covered both in terms of sample and study numbers, to avoid importing a specific study's bias in control selection criteria.

2.2. Chronological age as an (in)dependent variable

Prior to examining specific implementations, we should address a frequently overlooked aspect of biohorology. While it is most reasonable to treat CA as a target variable, which should never leak into the predictor variables, this is not always the case. Most modern aging clocks indeed treat CA merely as a selection criterion, but there is also a class of clocks where it is treated as an aging marker.

To put it in mathematical terms, the former approach is equivalent to solving $BA = W \cdot x + \varepsilon$ for W (a vector of unknown coefficients), where x is a matrix of dependent features (e.g. clinical blood parameter levels) and ε is noise. Assuming that CA of healthy individuals is close to their BA and therefore represents a sort of a norm, the BA prediction problem can be solved by minimizing the difference between estimated

BA and CA. Meanwhile, the alternative approach is described by the following equation: $BA = W \cdot (x + CA) + \varepsilon$. In this case, CA cannot be used as a benchmark for a solution's quality to avoid producing a degenerate model $BA = CA$. Thus, BA should be defined in a different fashion. For example, in (Borkan and Norris, 1980) BA is defined in a way that makes it a metrics of similarity to a person's CA peers.

The first instance of age being treated strictly as a dependent variable is probably (Nakamura et al., 1988). This study criticizes the previously used methods and provides sound arguments against them. It points out that they lead to a situation, when "a perfect equation, which predicts the dependent variable correctly, always predicts the identical CA". In other words, it carries the risk of producing mathematically and biologically trivial models.

The authors then follow to demonstrate their alternative. Having information on 30 physiological parameters (such as total cholesterol, pulse and blood hemoglobin) for 462 people, they used PCA and correlation analysis to select a set of 11 parameters most characteristic of aging. In the end, they present a regression model with a standard error of 7.49 years. The model was then used to show that diabetic and hypertensive people age faster than healthy population.

Today most aging clocks are Nakamura-like in that they avoid using age to predict age, although some of them are significantly more complex. However, there are exceptions.

One such exception is PhenoAge (Levine et al., 2018). It belongs to a family of methylation (DNAm) clocks, described in more detail below (Section 2.4). PhenoAge replaces CA benchmark with "phenotypical age", which is derived from a 10-dimensional distribution for mortality risk in the training set. The ten dimensions include nine blood markers and CA itself. Although strongly predictive of mortality risk, smoking status and BMI, when it comes to age estimation PhenoAge represents one of the caveats mentioned by Nakamura, namely: "the distortion of ... BA at the regression edges", – it indeed overestimates the age of younger adults (Liu et al., 2018).

Another clock, DNAm GrimAge, uses a two-stage training procedure to estimate time to death. Firstly, DNAm biomarkers were used to estimate concentrations of 88 plasma proteins and smoking pack-years (Lu et al., 2019). The resulting DNAm-based surrogates for 12 of them, as well as pack-years, significantly correlated with actual concentrations (Pearson $R > 0.35$). Secondly, the surrogates, gender and CA are all used to regress time to death — the "observed GrimAge". As in PhenoAge, the target variable is defined using CA, as well as plasma protein levels, gender and reported pack-years. Such a complicated procedure was shown to offer improvements, compared to the single stage method, when DNAm markers are directly used to estimate time to death.

Strictly speaking, neither GrimAge nor PhenoAge are aging clocks: their definitions of age acceleration are not intended to approximate CA. Since the score they predict is derived from the concept of mortality and disease risks, "death and disease timer" or "health timer" would be more fitting names, and this is whether their utility lies. As their mortality risk definition has CA embedded in it, it makes them ideologically similar to the models debated by Nakamura. Such an approach used for purposes not intended by their design may lead to target leaking and mathematical degeneracy. Including CA among independent features limits certain applications, e.g. for age identification purposes in forensics. Rejuvenation focused settings may pose yet another hypothetical limitation for models that use CA both as a predictor and as a target. Provided that a rejuvenative intervention's efficiency is CA-invariant, such a model would assign greater BA to older people who underwent therapy, compared to the younger people who were subjected to it as well. However, it is still unclear how the other kind of aging clocks acts in such circumstances and what should be the benchmark to compare the two strategies applied to rejuvenated organisms. Both strategies are viable, but one should always keep the future applications and the associated pitfalls in mind while choosing one.

2.3. Telomere length

Telomeres are DNA regions located at the ends of linear chromosomes. Due to the specifics of eukaryotic DNA replication mechanism, telomeres shorten with each replication event but are indispensable to this process, which effectively makes telomere length a factor limiting the maximum number of cell divisions (Hayflick limit) and the regenerative potential of an organism.

Telomerase is an enzyme that elongates telomeres and its expression is essential for stem cell proliferation and differentiation (Feron et al., 2009; Hiyama and Hiyama, 2007). While some animals (e.g., pika, mouse and to a greater extent lobster and rainbow trout) exhibit telomerase activity in somatic cells throughout their lives, human telomerase is generally inactive in somatic cells (Gorbunova and Seluanov, 2009; Klapper et al., 1998b, 1998a). Telomerase suppression is believed to have evolved as a means to reduce the risk of malignant transformation by paying the cost of replicative senescence (Gorbunova and Seluanov, 2009). In humans replicative senescence is ubiquitous and leads to senescent cell accumulation, which ultimately is reflected in multiple aging phenotypes (Faragher and Kipling, 1999). Telomere involvement in this process makes it a promising biomarker in the aging clock context.

Telomere attrition relevance as an aging biomarker is further justified by numerous studies displaying correlation between longer telomeres and increased lifespan both in model organisms and in humans (Bekaert et al., 2020; Mitteldorf, 2013). Decreased telomere length in humans has been associated with such age-associated conditions as Alzheimer's disease (Cai et al., 2013; Liu et al., 2016), female reproductive aging (Kalmbach et al., 2013), blood composition shifting (Lin et al., 2015) and the risk of cardiovascular disease (Haycock et al., 2014). Senior individuals (60–97 years) with shorter telomeres detected in their blood samples have significantly higher all-cause mortality rate¹ (Cawthon et al., 2003). But in (Martin-Ruiz et al., 2005) no such correlation within an elderly cohort (85–101 years) passed statistical tests. Assessing associations between telomere length and age-related conditions also failed in most cases. Differences between sample sizes (143 vs 598 people, respectively), telomere measuring methods (qPCR and TRF — terminal restriction fragment, respectively) or age distribution could explain the discord between these two studies. This small example illustrates, how studies focused on telomere length predictive value regarding BA quite frequently show inconsistent results.

Multiple researchers have estimated the speed of telomere attrition to be 40–50 bp/year in blood cells. However, both the initial telomere length and the identified attrition rate vary greatly between different data sets (Sanders and Newman, 2013). Moreover, DNA lesions caused by oxidative stress are repaired less efficiently in telomeric regions, which causes frailty and subsequent telomere shortening (Coluzzi et al., 2014; Reichert and Stier, 2017; von Zglinicki, 2000). While replication frequency in hematopoietic stem cells (whose progeny in blood is most commonly used to measure telomere length) remains constant at approximately 0.6 times per year after puberty (Sidorov et al., 2009), oxidative stress levels may fluctuate due to habitat, life style, inflammatory diseases – factors that do not necessarily represent replicative clock ticking.

So far, telomere aging clock remains a hypothetical concept. Oxidative and inflammatory noise masking replicative attrition signal in blood samples is a major obstacle to creating an accurate age predictor, yet it is not the only one (Aviv, 2008). Existing telomere length measurement methods may also obscure useful information by introducing errors that are hard to account for. Two methods are most

frequently used for measuring telomere length: TRF and qPCR. The latter requires much less DNA and is more suitable for high throughput settings, but it introduces errors that can completely overwhelm the actual changes in telomere length (Sanders and Newman, 2013). Choosing qPCR also narrows experimental design space, as it allows only relative measurement. Considering these limitations, other approaches have been proposed to estimate telomere length (STELA, TESLA). These methods overcome limitations of qPCR and TRF at the cost of being labor intensive and low throughput. These and other telomere length measuring techniques are described in much more detail in the following review (Lai et al., 2018).

There are more factors that contribute to the problem of accurate telomere length analysis in blood. Lymphocyte subpopulations have been shown to have telomeres of different length. For example, naïve T-cells are reported to have telomeres 1.4kbp longer than memory T-cells (Weng et al., 1995). Meanwhile, lymphocyte abundance profile is not stable throughout a person's life, i.e. it changes in response to age and disease. This blood composition shift has to be accounted for in studies designed to display age as a function of telomere length (Wang and Navin, 2015). When mixed populations of blood cells are tested for telomere attrition, it remains unclear whether the discovered changes are due to telomere attrition *per se* or are due to differences in naïve T-cell content in samples (Lin et al., 2015). Besides, tracking blood cell telomere length in individuals shows that it does not decrease at a constant rate and can in fact go up. Telomere length typically fluctuates within $\pm 2 - 4\%$ per month. This led scientists to hypothesize that telomere attrition is an oscillatory process (Svenson et al., 2011). Its dynamic nature cannot be ignored and indicates that its interpretation as a biomarker of aging requires longitudinal design. Also, recently telomerase has been verified to be active in heart and endothelium (Haendeler et al., 2004; Richardson et al., 2012). Interestingly, its activity in these tissues, as well as in leukocytes, can be elevated by endurance training (Werner et al., 2009). These findings further challenge the prior hypothesis of steady rate telomere shortening.

The studies on the involvement of telomeres in cell senescence indicate that this process may not always be driven by telomere attrition. Telomere damage, however, clearly plays a role in it. DNA damage is repaired at slower rate in telomeres than other regions, which may cause the cell-cycle arrest. The situation is further confounded by the technical limitations of telomere-senescence connection research. The use of telomerase-deficient mice and non-exclusive senescence markers might obscure our understanding of the cell senescence phenomenon (de Magalhães and Passos, 2018).

All these issues preclude the creation of a robust telomere-based aging clock. The development of methods to measure telomere length is most essential for this cause. Additionally, telomere attrition should be separated from other factors affecting telomere length by either processing specific cell subpopulations or explicitly subtracting their effects. This, however, obviates the reductionist ideal of a telomere “candle clock”, as it turns an alluring one-dimensional biomarker into a convoluted mixture of multiple biomarkers. Perhaps telomere length could be another variable in a fully-functional multi-dimensional aging clock, but its independent performance for such purposes is highly doubtful.

2.4. Epigenetic marks

Genetic information in animals is expressed according to the instructions contained both within (promoters, enhancers, etc) and outside its DNA sequence. Such instructions are realized in the repertoire of covalent DNA modifications, proteins organizing DNA into a 3D structure (chromatin) and their respective modification. Among these, DNA methylation is the most studied feature in the context of aging biomarkers. DNA methylation most commonly takes place in CpG sites, which despite being a simple dinucleotide motif is much rarer in mammals than expected (1% observed frequency against 4.4 %

¹ Mortality rate ratio between senior people from the lower and the upper half of telomere length distribution reaches 8.5 for infectious diseases and 3.2 for heart conditions, while all-cause mortality ratio is 1.7.

expected in humans) (Babenko et al., 2017; Han et al., 2008). While in invertebrate model organisms (*Drosophila melanogaster* and *Caenorhabditis elegans*) CpG distribution is uniform across the genome, mammalian CpG sites are located mostly in CG-island (CGI) regions, where their frequency reaches 18 %. Interestingly, CGIs often coincide with gene promoter regions where their methylation status affects transcription levels (Babenko et al., 2017; Deaton and Bird, 2011).

Despite DNA methylation (DNAm) being an extensively studied epigenetic mark, its great significance in the aging process long remained unrecognized (Gibbs, 2014). In a 2011 study conducted on saliva samples from 34 twin pairs 88 CpG sites were reported to be significantly associated with age. Methylation status at 2 of these sites was used to build a regression model that predicted CA of a donor with just 5.2 root mean square error (RMSE) (Bocklandt et al., 2011). Building on this work he had taken part in, Steve Horvath published a seminal 2013 article, where he presented 353 CpG sites whose methylation status allowed to predict CA with 2.9 years median error across multiple tissues (minimum being <1 year in peripheral blood mononuclear cells and maximum — 12 years in dermal fibroblasts). These results were derived from analyzing 7844 samples and tested in various settings including chimpanzee tissues, cancer tissues, embryonic stem cells and cell-cultures. Interestingly, the DNAm clock does not reflect replicative or cellular senescence (which is the case for telomere length) as it provides accurate predictions for both post-mitotic and immortalized tissues (Horvath, 2013).

Another independent DNAm clock was introduced in 2012 by Hannum et al. (Hannum et al., 2013). While this clock was developed on fewer samples (656) and contained fewer (71) CpG sites it achieved RMSE of 3.9 years. Despite the model being trained on blood samples, it can be extrapolated to other tissues with linear offset adjustments to predict age with 5.71 years RMSE.

The creation of DNAm clocks has greatly benefited biogerontology. Since 2013, they have been extensively tested and used to show connections between organismal aging and dozens of diseases, human phenotypes and lifestyle choices including obesity, gender, insulin, blood glucose levels and most importantly — all-cause mortality. High biological relevance has proven DNAm based models to go beyond mere CA estimation and now their predictions are frequently used as BA measures. A detailed comparison and discussion on DNAm clock technology is available elsewhere (Horvath and Raj, 2018).

Apart from DNAm clocks' applications *in vivo*, they also have become extremely useful for *in vitro* experimental designs, where the problem of assessing the age of a cell culture could not be solved previously. For instance, DNAm clocks have proven to be indispensable in induced pluripotent stem cell (iPSC) experiments, where they were used to display rejuvenation phenomenon upon cell dedifferentiation (Frobel et al., 2014).

Most recent development in DNAm-based tools of age prediction include a variety of mouse aging clocks (Meer et al., 2018; Petkovich et al., 2017; Stubbs et al., 2017; Wang et al., 2017). One of them utilizes 435 CpG sites and shows 53 days MAE when validated in multiple tissues (Meer et al., 2018). This error is in line with the performance of human DNAm clocks and amounts to 4–5 % of mouse average lifespan.

Their established and potential applications aside, DNAm aging clocks raise a number of fundamental questions regarding what aspects of aging they reflect and why certain epigenetic marks correlate with CA so significantly. The comparison of existing DNAm aging clocks, surprisingly, shows little overlap between the identified marker CpG sites. Among 1143 CpG sites used in 4 different mouse aging clocks none are shared by all four. Besides, among 605 genes containing the clocks' sites only 12 appear in the intersection (Meer et al., 2018). This issue was directly addressed in (Thompson et al., 2018), where a variety of training procedures was tested to conclude that “the construction of epigenetic clocks is highly degenerate”, i.e. there are multiple equally good solutions for the same set of features.

Human DNAm clocks are not free of this issue either, as only 6 CpG

sites are shared between Hannum and Horvath clocks. Moreover, testing these two clocks in independent data sets may result in significantly different error distribution (Armstrong et al., 2017; Zhang et al., 2017). Previously mentioned PhenoAge, which utilizes 513 CpG sites, shares only 41 with Horvath's clock and only 6 – with Hannum's. In spite of this heterogeneity, all three clocks showed predictive association with all-cause mortality in a Cox regression analysis with PhenoAge slightly outperforming other clocks (Levine et al., 2018). The factors contributing to DNAm clock degeneracy may involve batch effect, training bias, training tissue selection, and intrinsic inter-site correlation. These do not detract from the applicability of DNAm clocks, but pose a number of interesting questions, such as: Is there one most succinct and descriptive set of CpGs? Are the sites in different clocks regulated by the same epigenetic mechanisms? Is there a practical limit to the number of DNAm clocks?

Yet another blood DNAm aging clock was reported in 2014 by Weidner et al. (2014). The clock reached MAE of 3.34 years when utilizing 102 age related CpG sites. However, using only 3 most descriptive sites reduced the accuracy to only 4.5 years MAE. Among these three only one was present among the 353 CpG sites of Horvath's clock and none among the 71 CpG sites of Hannum's clock (Fig. 1). The clock was verified in an iPSC setting, tested for invariance against blood composition and showed increased aging rate in donors with dyskeratosis and aplastic anemia as well as in people with high alcohol consumption. This clock was later reworked into a 99-CpG clock to contain only the sites present in both Illumina 450 K and 27 K platforms (Lin et al., 2016).

A similar feat in feature reduction was achieved for murine DNAm clocks, when in 2018 a model based on 3 CpG sites allowed to estimate mouse age with 35 days of MAE (Han et al., 2018). The fact that only 3 CpG sites could be used to produce a model of comparable accuracy to those that require much more information is utmost astonishing.

Small overlap between features of different DNAm clocks with comparable accuracy requires careful examination. One of the initial hypotheses behind DNAm clock remarkable performance posited that age related changes in DNAm reflect weakening control over spontaneous modification. Methylation at specific sites has been shown to drift towards moderate levels with age: highly methylated regions steadily grow less methylated, while undermethylated ones, on the other hand, steadily gain methylation marks (Hannum et al., 2013; Petkovich et al., 2017). In this case, DNAm clocks indicate the triumph

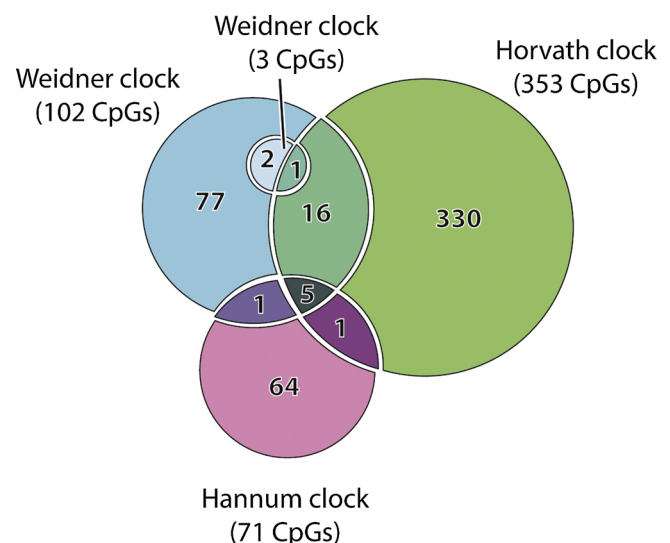


Fig. 1. Venn diagram of the CpG sites used in Hannum, Horvath and two Weidner clocks (all have <5 years of error, see Table 1). The intersection between their features is rather thin and indicates that DNAm clocks can achieve similar accuracy while relying on different sets of methylation sites.

of entropy in complex live systems and there could be innumerable solutions to quantify it using one of the countless subsets of methylation sites. In this case, all such clocks can be used interchangeably since they are all based on this particular and universal aspect of aging.

However, an alternative hypothesis would be that methylation at different CpG sites corresponds to fundamentally different aging processes. If an organism's DNAm profile is not directly linked to the thermodynamic root of aging but instead is a downstream product of competing processes, the applicability of DNAm aging clock methodology is at risk. In this case different aging clocks may not be equally good for different experiment settings and may propagate semantic barriers instead of removing them.

Experiments conducted in mice show that while genetic, pharmacological and dietary interventions with proven effect on life expectancy change the methylation state of the age-associated CpG sites, they do so in different ways. For instance, long-lived Ames dwarf mice contain three times fewer aging associated CpG sites than wild type mice. Meanwhile, caloric restriction is more efficient in preventing methylation loss at hypomethylated sites and methylation gain at hypermethylated sites than rapamycin (Cole et al., 2017). These findings imply that DNAm profiles do not simply gravitate towards the average with age and that there is no single pathway through which all aging processes are imbued into an organism's epigenetic landscape. In addition, it was pointed out that the location of predictive CpG sites near genes regulating development and proliferation is consistent with program-like behavior (Petkovich et al., 2017).

Theoretical implications of aging clocks are indeed very valuable, and their discussion dominates the field of biohorology. In the meantime, the technical approaches used to obtain and process DNAm data receive relatively little attention. The choice of the platform for DNAm screening is critical for a clock's applicability. While CpG microarrays are highly reproducible, bisulfite sequencing approaches suffer from coverage issues and testing clocks using it (e.g. all the mentioned mouse clocks) on independent data sets may be impossible due to missing values for important CpGs (Wagner, 2017).

Machine learning algorithms used in biohorology also require careful examination. All the DNAm clocks described in this review so far are based on linear regression with coefficient regularization (Table 1). Regularization techniques are used to tackle multicollinearity and feature selection problems which may cause overfitting. Lasso regularization does that by imposing L_1 penalty on coefficients, which can result in assigning zero weights to certain features. Ridge regularization, on the other hand, makes use of L_2 penalty and is more likely to reduce model complexity by shrinking the weights instead of dropping them (Fig. 2). Elastic net regularization combines both L_1 and L_2 penalties in a user-defined manner, which is often presented as "taking the best of two worlds". Caveats regarding the application of these methods for training a DNAm clock remained unexplored for a long time. In a 2018 study clocks built with Lasso, Ridge and Elastic Net were compared in the same data set. The training set contained information on methylation status of 193,651 CpG sites in multi-tissue samples of 1189 mice. Lasso regularization showed the poorest performance of three. Meanwhile, more accurate Elastic Net based clocks turned out to be inferior to Ridge clocks, since they were less sensitive to the effects of certain anti-aging interventions (e.g. failing to detect slower aging in dwarf long living mice) (Thompson et al., 2018).

While DNAm is the most well-studied epigenetic mark in the context of biohorology, it is not the only one. Protein structures encapsulating DNA and regulating its accessibility (chromatin and histones) have also been shown to change with age. Moreover, DNAm machinery and histone modifications are interlinked and change throughout aging concordantly. For example, DNA methyltransferases are attracted by the H3K36me mark. With aging it is less tightly regulated, and thus, more sporadic DNAm occurs, which ultimately translates to epigenetic clock ticking (Martin-Herranz et al., 2019).

The idea of integrating DNAm with other chromatin features into a

single aging clock has high potential, but is hard to implement. Apart from histone marks, chromatin properties that are age-dependent include: senescence-associated heterochromatin foci (SAHF) (Aird and Zhang, 2013), irregularly spaced nucleosomes (Ishimi et al., 1987), reduced histone biosynthesis and turnover (O'Sullivan et al., 2010), changes in nucleosome occupancy at specific sites (Bochkis et al., 2014) and others (Feser and Tyler, 2011). Many such age-related features are extremely labor intensive to detect and require chromatin immunoprecipitation (ChIP), antibody staining, microscopy and sometimes sequencing on top of that. Experimental pipelines used to study chromatin may have numerous noise-introducing stages (e.g. the antibody wild card) and may only allow for relative measurement.

Hopefully, the advent of new methodologies in chromatin research will increase measurement reproducibility and provide biogerontologists access to aging related information contained within chromatin structure. Such high-throughput methodologies as mass cytometry have already been successfully used to describe histone modifications differing between the young and the elderly (Cheung et al., 2018).

Currently, multiple compounds affecting DNAm and chromatin machinery are approved or are seeking approval for malignant pathologies (Eckschlager et al., 2017; Johnson et al., 2012). The existing clinical data on these drugs makes them top candidates for being repurposed for geroprotective therapies, which further increase the importance of epigenetic research in the biogerontological context.

2.5. Transcriptomics

Numerous studies have mapped various aging phenotypes to changes in transcriptome and more than 1000 of transcripts have been shown to be differentially abundant in people of different age groups (Harris et al., 2017; Peters et al., 2015). Some studies even identify dozens of genes whose age-related expression patterns are preserved across humans and rodents (de Magalhães et al., 2009). Cell senescence also has a distinct transcriptomic signature involving dozens of genes (Casella et al., 2019). Other studies draw bridges between senescence and aging by discovering the common signatures between the two (Chatsirisupachai et al., 2019). Compared to epigenetic profiles, age-associated transcriptome changes present a much more solid foothold for geroprotective drug target discovery. Transcriptome analysis allows to highlight specific up- or down-regulated pathways and proteins associated with aging, while immense body of literature and experimental data makes it practical to draw and test hypotheses regarding various interventions *in silico*. State-of-the-art deep learning algorithms can be employed for the purpose of finding disease targets. Neural-network derived latent representations of expression data, protein interactions and gene annotations can be aggregated in a modular fashion to assess the likelihood of a gene being involved in the pathogenesis of a certain disease (Fabris et al., 2019). While such tools allow for analyses of previously unseen scale and accuracy, they deal with statistically prepared data, while transcriptomic aging clocks have the power to inspect the age ticking rate on individual level and thus complement the described approach within the paradigm of personalized medicine.

In 2015, the first transcriptomic aging clock was published that was trained on blood RNA profiles from 8847 people. In this study, 1497 genes were reported to significantly change with age, yet the predictor used expression values obtained with exon microarrays from 11,908 genes. The clock achieved 7.8 years MAE and its biological relevance was further proved by displaying that increased predicted age is associated with higher blood pressure, blood glucose and cholesterol levels. The clock was then released as an online tool: Transcriptomic Age Prediction (TRAP) (Peters et al., 2015).

In 2018, a transcriptome aging clock was published that was trained on skin fibroblast profiles from 133 people (1–94 years old) and exhibited MAE of 7.7 years. Multiple machine learning techniques are compared in the study including elastic net, random forest, kNN, Gaussian naïve Bayes and linear discriminant analysis (LDA) classifier

Table 1
Summary of studies on biological age assessment.

Reference	Species	Regression model	Accuracy reported	N of Input Feat.	Sample types	Number of samples (overall)	Age range	Data Profiling Technology
(Bocklandt et al., 2011)	Human	Lasso	MAE of 5.2y	88	Saliva	128	18–70y	HumanMethylation27
(Horvath, 2013)		Elastic net	MedAE of 3.6y	353	Multi-tissue	7844 (+ 5826 cancer)	0–100y	HumanMethylation27, HumanMethylation450
(Hannum et al., 2013)		Elastic net, filtering features based on bootstrap frequency. Plus gender, age-adjusted BMI, diabetes status, ethnicity, and batch included in the model but not regularized.	RMSE of 3.9y	71	Blood, other tissues with linear offset adjustment	656	19–101y	HumanMethylation450
(Weidner et al., 2014)		Correlation selection + Linear regression	3.34y 4.5y	102 3	Blood	575	0–78y	HumanMethylation27
(Lin et al., 2016)			MedAE of 3.45y	99		2765	0–103y	HumanMethylation27, HumanMethylation450
(Peters et al., 2015)		Ridge	MAE of 7.8y	11,908	Blood	8847	18–90y	Exon microarrays
(Putin et al., 2016)		DNN	MAE of 5.55y	42	Blood	62,419	0–95y	Clinical blood tests
(Fleischer et al., 2018)		LDA ensemble	MAE of 7.7y, MedAE of 4y	4852	Skin	133	1–94y	RNAseq
(Levine et al., 2018)		Elastic net (CA used as a feature)	NA	513	Blood	9620	22–93y	HumanMethylation27, HumanMethylation450, Illumina EPIC array, Blood tests
(Mamoshina et al., 2018b)		DNN	MAE of 6.2y; mAUC of 0.80	7682	Muscle	1109	19–89y	Exon microarray, RNAseq
(Mamoshina et al., 2018a)		DNN	5.9y	18–21	Blood	198,130	20–90y	Clinical blood tests
(Mamoshina et al., 2019)		DNN	MAE of 5.7y – 5.9y	14–23	Blood	149,000	20–90y	Clinical blood tests
(Lu et al., 2019)	Mouse	Elastic net (CA used as a feature)	NA	1115	Blood	10,091	46–98y	HumanMethylation450, Illumina EPIC array
(Petkovich et al., 2017)		Elastic net	R ² = 0.90	90	Blood	141	3–35mo	RRBS
(Stubbs et al., 2017)		Elastic net	MedAE of 3.33w	329	Multi-tissue	318	0.2–9.5mo	RRBS
(Wang et al., 2017)		Elastic net	MAE of 4.2mo	148	Liver	102	0.2–26mo	RRBS
(Meer et al., 2018)		Elastic net	MAE of 53–64d	435	Multi-tissue	416	0.2–35mo	RRBS
(Han et al., 2018)		Lasso	MAE of 4.86–7.1w	3	Blood	97	3–109mo	RRBS
(Thompson et al., 2018)		Ridge	MedAE of 0.3mo	193,641	Multi-tissue	1189	2–21mo	RRBS
		Elastic net	MedAE of 1.2mo	463				
		Ridge	MedAE of 3.5mo	952				
		Elastic net	MedAE of 2.5mo	247				

Note: CA for chronological age; LDA for linear discriminant analysis; MAE for mean absolute error; MedAE for median absolute error; RMSE for root mean squared error; mAUC for multiclass area under the receiver operating characteristic curve for the accuracy; y for years; mo for month; RNAseq for RNA sequencing or whole transcriptome sequencing using next generation sequencing; RRBS for reduced representation bisulfite sequencing.

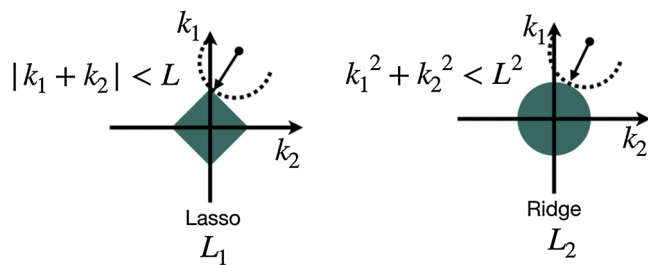


Fig. 2. Lasso (L_1) and Ridge (L_2) regularization in a two-dimensional space. A constraint is posed on two predetermined model coefficients (k_1 , k_2): their values should be within the green area defined by the regularization penalty function. The final value of the coefficients is determined by finding an optimal position on the plane, where the combined regularization and prediction error (uniform here, dashed line) penalties are minimal.

ensembles. Among these neither one reached the less than a decade accuracy of the LDA ensemble. Each model in the ensemble predicted whether a sample belonged to a 20 year long shifting age bin and the final prediction was defined as the year that belongs to the most chosen bins. Authors of this study suggest that their LDA approach could produce predictions with less than 5 or even 3 years of error with only 32 more samples in a training set, judging from the LDA learning curve extrapolation (Fleischer et al., 2018).

Another transcriptome aging clock published in 2018 showed that such a steep increase in accuracy is unlikely. In this study, 545 gene expression profiles (19–89 years old) from muscle tissue were used to construct a variety of age predictors with the best performance belonging to a Deep Neural Network (DNN) with a MAE of 6.24 years (Mamoshina et al., 2018b). Transcriptome age predictors have not yet reached the <5 year accuracy of their DNAm counterparts, but with the high-throughput RNAseq technology in place this feat is most likely a matter of time. Among the three presented transcriptomic clocks, the clock based on DNN is the most accurate, which may indicate that transcriptomic age prediction requires more complex machine learning techniques than those commonly used in DNAm clocks.

2.6. Biochemical compounds

All previously mentioned biomarkers were based on biological macromolecules, be it DNA, RNA or proteins. But all these are not separate entities, they are involved in various feedback loops that incorporate innumerable small molecules both as upstream signals and as downstream responses. Many of these small molecules have long been used as indicators of aging-associated diseases and have well established, reproducible and robust laboratory screening pipelines, which make them an attractive basis for aging clock development.

Despite the affordable price of clinical biochemistry screening and vast amounts of corresponding data, their potential as aging biomarker was not assessed for a long time. The first aging clock based on blood biochemistry was introduced in 2016 (Putin et al., 2017). The clock was trained and validated on a massive data set containing 62,419 people. Although the clock used just 42 features (such as calcium, cholesterol, glucose, urea and others, plus sex) it displayed remarkably small MAE of 5.55 years. The method that was used in this study is quite different from any other discussed above: the predictor consists of an ensemble of 21 DNNs whose predictions are then stacked by an elastic net model to produce the final age prediction (Fig. 3). The clock is also available as a free online service called Aging.AI. Its elemental models can also be used to produce age predictions independently with the best performing DNN having 6.07 years MAE.

In 2018, an updated version of Aging.AI was released (Mamoshina et al., 2018a). The article accompanying this release featured extensive cross-population analysis of Aging.AI performance. The aggregated data set used for the new Aging.AI training and validation consisted of

almost 200,000 blood profiles from Canada, Eastern Europe and South Korea. The DNN trained over all three populations displayed MAE of 5.94 years, which is slightly higher than 5.55 years of the previous version. However, the new predictor is much less complex: it uses half the features (sex, geographical location plus 19 blood markers) and is a single DNN instead of a DNN ensemble. Interestingly, feature importance analysis included in the study displays that sex and population are among the most important variables and excluding them from the model increases the error to 6.23 years. Nonetheless, Aging.AI was shown to be useful for assessing mortality risks in an independent National Health and Nutrition Examination Survey (NHANES) data set: hazard ratios were consistently bigger in people with higher than actual predictions, and similarly, smaller in people with lower than actual predictions. Moreover, similar DNN-driven approach was later shown to detect higher aging rates in smokers by the same group of authors (Mamoshina et al., 2019).

High accessibility, standardized protocols and affordable price make clinical biochemistry tests a promising source of aging biomarkers that could be used independently or to complement other methods of age prediction.

3. Adding biological interpretation

Following decades of failed attempts to develop biomarkers of aging, the aging clock boom started in 2013 and has been growing ever since. On the one hand, this reflects ongoing progress in the field. On the other, this signals an unsatisfied need in the community. Horvath's DNAm clock is still regarded as a golden standard and a go-to solution in biogerontology, despite it being a first-generation model. Arguably, the whole biohorology sector is still experiencing its first generation. While all aging clocks mentioned in Table 1 show significant predictive power, there is little concordance between their predictions. An elaborate comparison of seven measures of BA (including three DNAm clocks and telomere length) shows that they are only loosely correlated. For example, Pearson coefficients for the DNAm predictions lie within 0.3–0.5 range. All the measures also display different effects on healthspan-related characteristics such as grip strength, cognitive function and facial aging (Belsky et al., 2017).

Such discrepancies show that any clock based on a single biomarker of aging will likely miss some of its aspects. It may be more practical to combine biomarkers of different nature into one model. But increasing the procedural complexity of a predictor inevitably raises the cost of measurement. From this point of view, clinical blood tests offer the most cost-effective means to increase another model's performance. Besides, recent feats in age prediction show that such tests can be used as standalone solutions (Mamoshina et al., 2018a). PhenoAge creators have chosen an original approach to combine different biomarkers and succeeded. By training a DNAm clock to predict a substitute age metrics derived from blood parameters they have built a model that in some cases outperforms all other DNAm clocks trained to predict CA (Levine et al., 2018). But this approach does not always work as intended. Despite perceived facial age being clearly associated with all-cause mortality, training a DNAm clock to predict this measure of BA has failed (Marioni et al., 2018). Simply incorporating more biomarkers into one model does not necessarily guarantee its increased performance. Neither does it guarantee that the clock will reflect more aspects of aging.

Assessing the biological relevance of a clock and that of the corresponding aging biomarkers requires extensive laboratory research, and feature importance analysis is essential to establish a starting point. In case of regression, model coefficients provide insight into which biomarkers have a greater effect on the prediction. But when it comes to comparing how each feature contributes to the predictor accuracy in more complex models, such a straightforward approach is likely to be unavailable. This issue gets even more pressing when there is a need to compare how fundamentally different models treat the same variables.

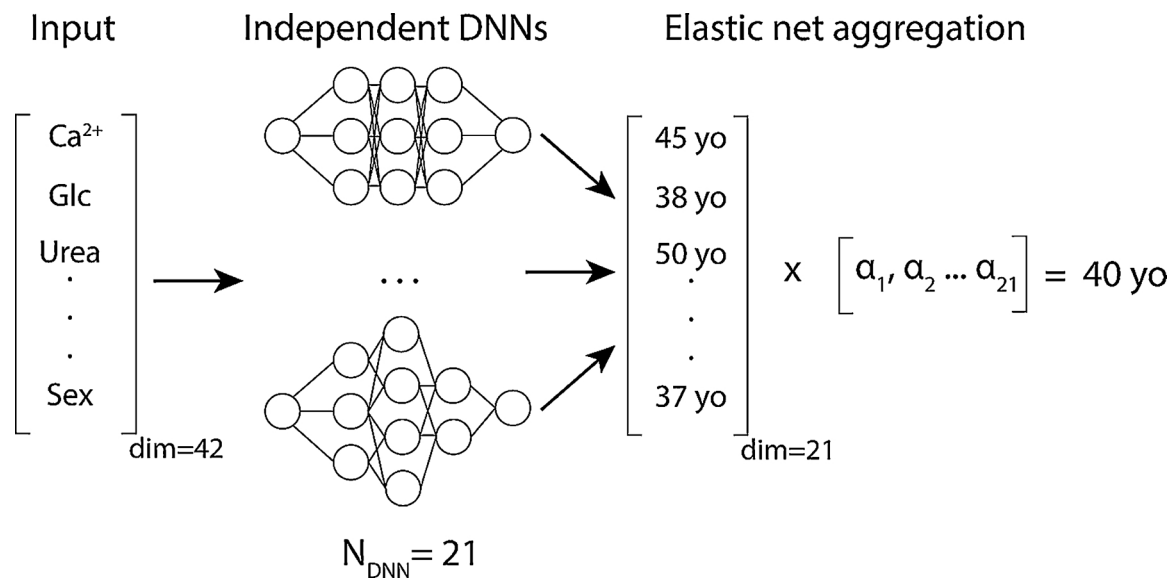


Fig. 3. Tandem DNN and elastic net approach used in Aging. AI blood chemistry aging clock. Predictions from 21 independent DNNs are aggregated by an elastic net regressor to produce the prediction better than any separate DNN can provide.

For example, such a need may rise while compiling a random forest and a DNN into an ensemble, or while investigating why a particular class of predictors outperforms any other one.

In such scenarios permutation feature importance score (PFI) might help provide the required answers. PFI is based on the idea that a model's accuracy metrics changes more significantly while permuting its important features rather than the unimportant ones. The score itself is the average change in accuracy over multiple iterations and its P-value (Altmann et al., 2010). While PFI is a useful tool, it should be used carefully when features are not independent, as in this case random permutations may produce unrealistic mock vectors. For example, some PFI implementations when applied to blood tests data can produce mock vectors with low blood glucose and high glycated hemoglobin, while these two parameters are actually positively correlated (Makris and Spanou, 2011). As a result, feature importance analysis can be uninformative or misleading. To avoid this pitfall, prior correlation analysis and feature selection should be performed as well as a careful examination of the assumptions taken in any particular PFI implementation. It also makes sense to explore other model-agnostic methods of feature importance analysis such as partial dependence plots, local interpretable explanations, accumulated local effects and SHAP values (Apley, 2016; Lundberg and Lee, 2017; Ribeiro et al., 2016; Zhao and Hastie, 2017). The latter two can be especially useful, as apart from being resistant to data collinearity they also show how exactly a model responds to changes in feature values.

For DNN-centric approaches deep feature selection (DFS) is the preferred method of importance analysis (Li et al., 2016). DFS models contain a linear one-to-one layer receiving the input, and the weights assigned to it upon training can be interpreted as importance scores to select an optimal subset of features. While DFS scores cannot be calculated for shallow models, they still can be compared to model-agnostic scores by shifting from absolute score values to their ranks. Furthermore, such rank scores can be aggregated by a voting system (e.g. Borda count) to provide a consensus list of most important features. Interestingly, when applied to a variety of transcriptomic aging clocks, Borda count produces a consensus importance list that is the closest to DFS importance ranks (Mamoshina et al., 2018b).

Proper feature importance analysis should always accompany an aging clock. It can lead to the detection of data contamination, uncovering algorithm-specific biases, reduction of model complexity and/or increase in its overall accuracy. Moreover, it provides useful insights into the nature of aging and can be used to produce novel hypotheses.

4. Generative biology in biogerontology

Currently the field is dominated by regularized regression models introduced in the first DNAm clocks. While this method has been reported to produce the most precise predictions, it was argued that it contains an implicit and possibly overlooked tradeoff between precision and the ability to behave correctly in uncommon settings (Thompson et al., 2018). Another shortcoming of regression models rises from their sensitivity to missing values, which impedes their application in cross-platform studies. For example, a model built on Illumina HumanMethylation27 BeadChip array may contain features absent in the Illumina HumanMethylation450 BeadChip array and thus be incompatible with it. Considering the fast pace of platform development and adoption, clocks built with data from legacy platforms may become obsolete in the near future.

Deep learning techniques provide an attractive alternative to regularized regression. They have been used to reduce the dominance of DNAm in the field with models capable of accurately estimating age from transcriptomic and blood test data (Mamoshina et al., 2018b, 2018a; Putin et al., 2017). Their highly customizable structure makes DNNs more resilient to missing values: by introducing dropout layers the model can be trained on vectors with randomly omitted values, which results in reduced reliance on any particular variable (Fig. 4).

One might argue that elastic net supremacy is not a choice, but a necessity dictated by DNAm data specifics, such as an immense number of dimensions. Training a network that uses all the features identified by modern DNAm screening platforms (e.g. Illumina HumanMethylationEPIC platform profiles >850'000 sites) would require datasets of non-existent magnitude. In such cases it makes sense to preselect a subset of features based on their genomic location, target association or via iteratively resampling the training set. An alternative solution to feature elimination would be feature clustering, for example according to correlation with a target variable or *a priori* biological information. The payoff of choosing a more complex pipeline to pre-process the samples and train a deep network, however, is that the resulting model can serve multiple purposes apart from predicting biological age.

One certain task DNNs could help with is domain adaptation — the process of teaching a model to work with data from a different (but related) distribution than the data it was originally trained on. For example, transferring the data obtained from a state-of-the-art sequencing platform into the domain of a legacy platform in order to

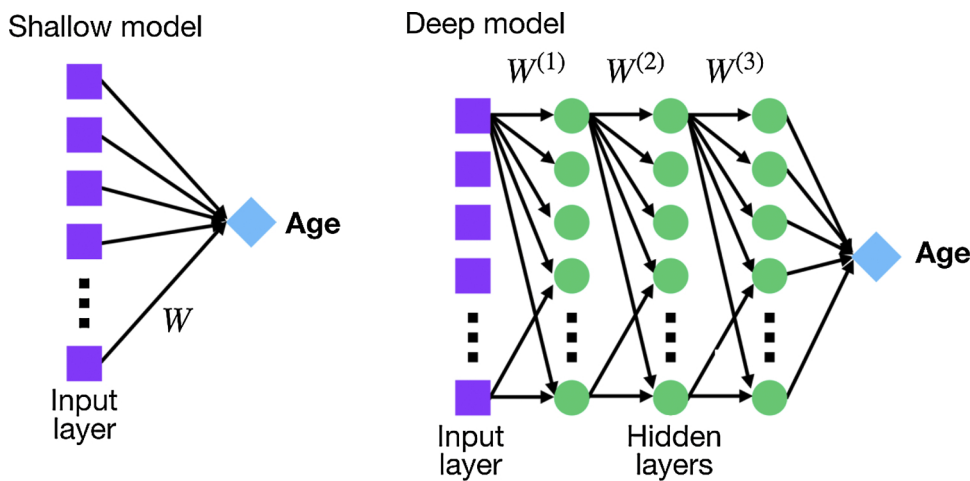


Fig. 4. Structural difference between shallow and deep models. Rigid shallow models are not fit to approximate complex, non-linear functions and are vulnerable to missing values. Meanwhile deep models theoretically can approximate any smooth function, which comes in handy in a variety of classification and regression problems (Ohn and Kim, 2019). However, deep models usually require bigger samples and more resources to train.

make the most recent data compatible with an older, yet outstanding age predictor. Domain adaptation algorithms are being actively studied and used for the purposes of visual learning (Ganin et al., 2015; Hoffman et al., 2017; Laradji and Babanezhad, 2018; Russo et al., 2017; Tzeng et al., 2017). All of them are based on DNNs and many utilize a specific subclass of DNN architectures – generative adversarial networks (GANs). Shortly, this architecture consists of two independent neural networks: discriminator and generator. The discriminator is taught to recognize a target domain, while the generator processes source-domain entries to resemble ones from the target domain. Their emulation continues until discriminator fails to tell real and artificial target entries apart. GANs have proven to be an extremely versatile approach and are used in various remarkable AI feats, such as text to image synthesis (Reed et al., 2016).

Although image processing is the field pushing innovation in domain adaptation methods, there are already some noteworthy examples of applying this technology to biological data. In 2015 DNNs were shown to outperform shallow methods in the task of trans-species domain adaptation. The specific task was to predict human protein phosphorylation states in response to a stimulus, using data obtained for rats. Based on the AUROC metrics, DNN (0.936) turned out to be superior to both elastic net (0.709) and support vector machine (0.724) solutions. The authors also provide rationale behind DNN's outstanding performance, and point out that shallow classification algorithms are created to discern the case-separating objects, while accurate domain adaptation requires the model to grasp the mechanisms producing said objects. Or as the authors put it, domain adaptation should focus on the “common encoding mechanism” instead of the “distinct signaling molecules ... employed by different species” (Chen et al., 2015). Another (probably more relevant in the context of this review) example of DNN-based domain adaptation would be using DNNs in single cell transcriptomics. scRNA data is prone to missing values, intra-cell type variability and on top of that — batch effects. Treating different batches as domains and applying deep learning to them produced a method that can assign cells from different experiments to cell types, with the only assumption that any two batches share at least one cell type (Wang et al., 2019). Similar methods could be adjusted to other data types, which would let biogerontologists overcome the problem of insufficient data needed for training elaborate models such as DNNs. However, many other measures could be taken to make future experiments compatible with each other and reduce the need for complex domain-adaptation algorithms. Such measures include: standardizing protocols and introducing benchmark or calibration samples, extensively and uniformly documenting metavariables in depositories and quantifying the biases introduced by sequencing platforms. All these options require high levels of organization and cooperation within the community, which makes them hard to implement, albeit much more technically

simple than employing DNNs.

GANs in biogerontology can be used not only to reformat data and impute missing values but also to expand training sets, predict future changes in a patient's biomarker profile and construct geroprotective interventions. DNN-based methods can be used to seek mimetics of known anti-aging drugs by screening aging biomarker response to various compounds. Such an approach has already been used to identify natural alternatives to metformin and rapamycin (Aliper et al., 2017). GANs have the potential to take anti-aging drug design one step further: from *in silico* screening libraries for promising compounds to generating libraries of yet undiscovered and possibly beneficial ones. This GANs' application has received a lot of development in recent years. Such *in silico* drug production lines contain a generator trained on a subset of structures that are used to treat a disease and a discriminator that rates the generated molecules (Kadurin et al., 2017; Putin et al., 2018). One specific instance of this technology was used to produce a DDR1-kinase inhibitor with cell culture verified activity in less than a month, demonstrating the impact DNN can have both in academia and industry (Zavoronkov et al., 2019).

5. Conclusion

Biological horology is a rapidly growing field of research. The multitude of aging clocks created over the last decade indicates the great need to measure BA in humans and model organisms. Among other biomarkers of aging DNAm has received the most attention and has been used to develop the most accurate aging clocks (MAE < 4 years) based on molecular profiles. Although there are numerous DNAm clock solutions using almost non-overlapping sets of features, specific aging scenarios exhibit distinct methylome fingerprints. This circumstance needs deeper investigation to determine what unites all the DNAm clocks and understand their limits of application.

Other kinds of clocks based on gene expression and biochemistry profiles offer a new point of view on aging, but are yet to break the MAE < 5 years milestone. These biomarkers are also responsive to interventions (Fig. 5). In the following years, proteomic, genomic or other measures of aging may emerge, but they will require significant advances in the corresponding quantification methods to allow high-throughput solutions.

But further progress in biological horology does not rely solely on the development of novel experimental practices. The creation of new analytical algorithms is just as important, if not even more so. The advent of deep learning techniques in the field ushers a paradigm shift in a field previously dominated by shallow models such as EN. DNNs and GANs specifically have the potential to greatly expand the applications of aging clocks and make them the heart of *in silico* geroprotective research pipelines. Gene expression age trajectories are

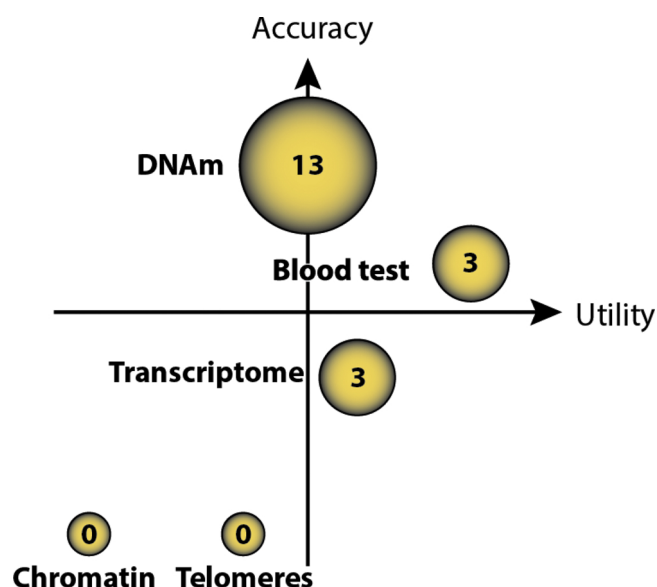


Fig. 5. Biomarkers mentioned in this article placed on an intuitive plane of Accuracy vs Utility. Bubble size depends on the number of clocks based on a corresponding aging biomarker. Currently, DNAm is the most accurate and the most frequently used biomarker in biohorology. However, it is harder to apply a DNAm clock compared to clocks based on clinical blood tests. Moreover, DNAm marks often take a long time to emerge in response to aging interventions. Such biomarkers as chromatin structure and telomeres, while intriguing, are too labor intensive and error-prone to be practical.

especially valuable in that respect as they allow for a direct pharmacological target identification.

Declaration of Competing Interest

FG, PM, AA, AZ work for Insilico Medicine, a for-profit biotechnology company developing the end-to-end target identification and drug discovery pipeline for a broad spectrum of age-related diseases. The company may have commercial interests in this publication. Products of Insilico Medicine include “Young.AI” system mentioned in this article. PM works for Deep Longevity, a for-profit longevity company. JPM is an advisor for Centaura, Longevity Vision Fund and is the founder of Magellan Science Ltd, a company providing consulting services in longevity science.

References

- Abramov, N., Atamanov, L., 1954. *Zolotaya Antilopa (The Golden Antelope)*. Soyuzmultfilm, USSR.
- Ahadi, S., Zhou, W., Schüssler-Fiorenza Rose, S.M., Sailani, M.R., Contrepolis, K., Avina, M., Ashland, M., Brunet, A., Snyder, M., 2020. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* <https://doi.org/10.1038/s41591-019-0719-5>.
- Aird, K.M., Zhang, R., 2013. Detection of senescence-associated heterochromatin foci (SAHF). *Methods Mol. Biol.* 965, 185–196. https://doi.org/10.1007/978-1-62703-239-1_12.
- Aliper, A., Jellen, L., Cortese, F., Artemov, A., Karpinsky-Semper, D., Moskalev, A., Swick, A.G., Zhavoronkov, A., 2017. Towards natural mimetics of metformin and rapamycin. *Aging (Albany NY)* 9, 2245–2268. <https://doi.org/10.18632/aging.101319>.
- Altmann, A., Tološi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26, 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>.
- Apley, D.W., 2016. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models.
- Armstrong, N.J., Mather, K.A., Thalamuthu, A., Wright, M.J., Trollor, J.N., Ames, D., Brodaty, H., Schofield, P.R., Sachdev, P.S., Kwok, J.B., 2017. Aging, exceptional longevity and comparisons of the Hannu and Horvath epigenetic clocks. *Epigenomics* 9, 689–700. <https://doi.org/10.2217/epi-2016-0179>.
- Aviv, A., 2008. The epidemiology of human telomeres: faults and promises. *J. Gerontol. A Biol. Sci. Med. Sci.* 63, 979–983.
- Babenko, V.N., Chadaeva, I.V., Orlov, Y.L., 2017. Genomic landscape of CpG rich

- elements in human. *BMC Evol. Biol.* 17, 19. <https://doi.org/10.1186/s12862-016-0864-0>.
- Bekaert, S., De Meyer, T., Van Oostveldt, P., 2020. Telomere attrition as ageing biomarker. *n.d.*. *Anticancer Res.* 25, 3011–3021.
- Belsky, D.W., Moffitt, T.E., Cohen, A.A., Corcoran, D.L., Levine, M.E., Prinz, J.A., Schaefer, J., Sugden, K., Williams, B., Poulton, R., Caspi, A., 2017. Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: do they measure the same thing? *Am. J. Epidemiol.* 187, 1220–1230. <https://doi.org/10.1093/aje/kwx346>.
- Bobrov, E., Georgievskaya, A., Kiselev, K., Sevastopolsky, A., Zhavoronkov, A., Gurov, S., Rudakov, K., del Pilar Bonilla Tobar, M., Jaspers, S., Clemann, S., 2018. PhotoAgeClock: deep learning algorithms for development of non-invasive visual biomarkers of aging. *Aging (Albany NY)* 10, 3249–3259. <https://doi.org/10.18632/aging.101629>.
- Bochkis, I.M., Przybylski, D., Chen, J., Regev, A., 2014. Changes in nucleosome occupancy associated with metabolic alterations in aged mammalian liver. *Cell Rep.* 9, 996–1006. <https://doi.org/10.1016/j.celrep.2014.09.048>.
- Bocklandt, S., Lin, W., Sehl, M.E., Sánchez, F.J., Sinsheimer, J.S., Horvath, S., Vilain, E., 2011. Epigenetic predictor of age. *PLoS One* 6, e14821. <https://doi.org/10.1371/journal.pone.0014821>.
- Borkan, G.A., Norris, A.H., 1980. Assessment of biological age using a profile of physical parameters. *J. Gerontol.* 35, 177–184.
- Butler, R.N., Sprott, R., Warner, H., Bland, J., Feuers, R., Forster, M., Fillit, H., Harman, S.M., Hewitt, M., Hyman, M., Johnson, K., Kligman, E., McClearn, G., Nelson, J., Richardson, A., Sonntag, W., Weindruch, R., Wolf, N., 2004. Biomarkers of aging: from primitive organisms to humans. *J. Gerontol. A Biol. Sci. Med. Sci.* 59, B560–7.
- Cai, Z., Yan, L.-J., Ratka, A., 2013. Telomere shortening and Alzheimer's disease. *Neuro Mol. Med.* 15, 25–48. <https://doi.org/10.1007/s12017-012-8207-9>.
- Casella, G., Munk, R., Kim, K.M., Piao, Y., De, S., Abdelmohsen, K., Gorospe, M., 2019. Transcriptome signature of cellular senescence. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz555>.
- Cawthon, R.M., Smith, K.R., O'Brien, E., Sivatchenko, A., Kerber, R.A., 2003. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* 361, 393–395. [https://doi.org/10.1016/S0140-6736\(03\)12384-7](https://doi.org/10.1016/S0140-6736(03)12384-7).
- Chatsirisupachai, K., Palmer, D., Ferreira, S., de Magalhães, J.P., 2019. A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell.* <https://doi.org/10.1111/acel.13041>.
- Chen, L., Cai, C., Chen, V., Lu, X., 2015. Trans-species learning of cellular signaling systems with bimodal deep belief networks. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btv315>.
- Cheung, P., Vallania, F., Warsinske, H.C., Donato, M., Schaffert, S., Chang, S.E., Dvorak, M., Dekker, C.L., Davis, M.M., Utz, P.J., Khatri, P., Kuo, A.J., 2018. Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. *Cell* 173, 1385–1397. <https://doi.org/10.1016/j.cell.2018.03.079>. e14.
- Cole, J.J., Robertson, N.A., Rather, M.I., Thomson, J.P., McBryan, T., Sproul, D., Wang, T., Brock, C., Clark, W., Ideker, T., Meehan, R.R., Miller, R.A., Brown-Borg, H.M., Adams, P.D., 2017. Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol.* 18, 58. <https://doi.org/10.1186/s13059-017-1185-3>.
- Coluzzi, E., Colamartino, M., Cozzi, R., Leone, S., Meneghini, C., O'Callaghan, N., Sgura, A., 2014. Oxidative stress induces persistent telomeric DNA damage responsible for nuclear morphology change in mammalian cells. *PLoS One* 9, e110963. <https://doi.org/10.1371/journal.pone.0110963>.
- de Magalhães, J.P., Passos, J.F., 2018. Stress, cell senescence and organismal ageing. *Mech. Ageing Dev.* <https://doi.org/10.1016/j.mad.2017.07.001>.
- de Magalhães, J.P., Curado, J., Church, G.M., 2009. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btp073>.
- De Magalhães, J.P., 2012. Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? *FASEB J.* <https://doi.org/10.1096/fj.12-210872>.
- Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022. <https://doi.org/10.1101/gad.2037511>.
- Degerman, S., Josefsson, M., Nordin Adolffson, A., Wennstedt, S., Landfors, M., Haider, Z., Pudas, S., Hultdin, M., Nyberg, L., Adolffson, R., 2017. Maintained memory in aging is associated with young epigenetic age. *Neurobiol. Aging.* <https://doi.org/10.1016/j.neurobiolaging.2017.02.009>.
- Dolejs, J., Marešová, P., 2017. Onset of mortality increase with age and age trajectories of mortality from all diseases in the four nordic countries. *Clin. Interv. Aging.* <https://doi.org/10.2147/CIA.S119327>.
- Eckslager, T., Plach, J., Stiborova, M., Hrabeta, J., 2017. Histone deacetylase inhibitors as anticancer drugs. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18071414>.
- Fabris, F., Palmer, D., Salama, K.M., de Magalhães, J.P., Freitas, A.A., 2019. Using deep learning to associate human genes with age-related diseases. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btz887>.
- Faragher, R.G.A., Kipling, D., 1999. How might replicative senescence contribute to human ageing? *BioEssays* 20, 985–991. [https://doi.org/10.1002/\(SICI\)1521-1878\(199812\)20:12<985::AID-BIES4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1521-1878(199812)20:12<985::AID-BIES4>3.0.CO;2-A).
- Ferron, S.R., Marques-Torres, M.A., Mira, H., Flores, I., Taylor, K., Blasco, M.A., Farinas, L., 2009. Telomere shortening in neural stem cells disrupts neuronal differentiation and neurogenesis. *J. Neurosci.* 29, 14394–14407. <https://doi.org/10.1523/JNEUROSCI.3836-09.2009>.
- Feser, J., Tyler, J., 2011. Chromatin structure as a mediator of aging. *FEBS Lett.* 585, 2041–2048. <https://doi.org/10.1016/j.febslet.2010.11.016>.
- Fleischer, J.G., Schulte, R., Tsai, H.H., Tyagi, S., Ibarra, A., Shokhirev, M.N., Huang, L., Hetzer, M.W., Navlakha, S., 2018. Predicting age from the transcriptome of human

- dermal fibroblasts. *Genome Biol.* 19, 221. <https://doi.org/10.1186/s13059-018-1599-6>.
- Forbes, G.B., Reina, J.C., 1970. Adult lean body mass declines with age: some longitudinal observations. *Metabolism*. [https://doi.org/10.1016/0026-0495\(70\)90062-4](https://doi.org/10.1016/0026-0495(70)90062-4).
- Frobel, J., Hemeda, H., Lenz, M., Abagnale, G., Jousen, S., Denecke, B., Sarić, T., Zenke, M., Wagner, W., 2014. Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells. *Stem Cell Reports* 3, 414–422. <https://doi.org/10.1016/j.stemcr.2014.07.003>.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Laroche, H., Laviolette, F., Marchand, M., Lempitsky, V., 2015. Domain-Adversarial Training of Neural Networks.
- Gavrilov, L.A., Gavrilova, N.S., 2002. Evolutionary theories of aging and longevity. *Sci. World J.* 2, 339–356. <https://doi.org/10.1100/tsw.2002.96>.
- Gialluisi, A., Di Castelnuovo, A., Donati, M.B., de Gaetano, G., Iacoviello, L., Moli-sani Study Investigators, the M.S., 2019. Machine learning approaches for the estimation of biological aging: the road ahead for population studies. *Front. Med.* 6, 146. <https://doi.org/10.3389/fmed.2019.00146>.
- Gibbs, W.W., 2014. Biomarkers and ageing: the clock-watcher. *Nature* 508, 168–170. <https://doi.org/10.1038/508168a>.
- Gladyshev, V.N., 2016. Aging: progressive decline in fitness due to the rising deleteriousness adjusted by genetic, environmental, and stochastic processes. *Aging Cell* 15, 594–602. <https://doi.org/10.1111/ace.12480>.
- Gorbulonova, V., Seluanov, A., 2009. Coevolution of telomerase activity and body mass in mammals: from mice to beavers. *Mech. Ageing Dev.* 130, 3–9. <https://doi.org/10.1016/j.mad.2008.02.008>.
- Haendeler, J., Hoffmann, J., Diehl, J.F., Vasa, M., Spyridopoulos, I., Zeiher, A.M., Dimmeler, S., 2004. Antioxidants inhibit nuclear export of telomerase reverse transcriptase and delay replicative senescence of endothelial cells. *Circ. Res.* <https://doi.org/10.1161/01.RES.0000121104.05977.F3>.
- Han, L., Su, B., Li, W.-H., Zhao, Z., 2008. CpG island density and its correlations with genomic features in mammalian genomes. *Genome Biol.* 9, R79. <https://doi.org/10.1186/gb-2008-9-5-r79>.
- Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., Izeta, A., de Haan, G., Geiger, H., Wagner, W., 2018. Epigenetic age-predictor for mice based on three CpG sites. *Elife* 7. <https://doi.org/10.7554/eLife.37462>.
- Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., Zhang, K., 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* 49. <https://doi.org/10.1016/j.molcel.2012.10.016>.
- Harris, S.E., Riggio, V., Evenden, L., Gilchrist, T., McCafferty, S., Murphy, L., Wröbel, N., Taylor, A.M., Corley, J., Pattie, A., Cox, S.R., Martin-Ruiz, C., Prendergast, J., Starr, J.M., Marioni, R.E., Deary, I.J., 2017. Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian Birth Cohort 1936. *Aging (Albany NY)* 9, 2489–2503. <https://doi.org/10.18632/aging.101333>.
- Haycock, P.C., Heydon, E.E., Kaptoge, S., Butterworth, A.S., Thompson, A., Willeit, P., 2014. Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* 349. <https://doi.org/10.1136/bmj.g4227>. g4227–g4227.
- Hiyama, E., Hiyama, K., 2007. Telomere and telomerase in stem cells. *Br. J. Cancer* 96, 1020–1024. <https://doi.org/10.1038/sj.bjc.6603671>.
- Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A.A., Darrell, T., 2017. CyCADA: Cycle-Consistent Adversarial Domain Adaptation.
- Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome Biol.* 14, R115. <https://doi.org/10.1186/gb-2013-14-10-r115>.
- Horvath, S., Raj, K., 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19, 371–384. <https://doi.org/10.1038/s41576-018-0004-3>.
- Ishimi, Y., Kojima, M., Takeuchi, F., Miyamoto, T., Yamada, M., Hanaoka, F., 1987. Changes in chromatin structure during aging of human skin fibroblasts. *Exp. Cell Res.* 169, 458–467.
- Johnson, T.E., 2006. Recent results: biomarkers of aging. *Exp. Gerontol.* 41, 1243–1246. <https://doi.org/10.1016/j.exger.2006.09.006>.
- Johnson, A.A., Akman, K., Calimport, S.G., Wuttke, D., Stolz, A., De Magalhães, J.P., 2012. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res.* <https://doi.org/10.1089/rej.2012.1324>.
- Jylhä, J., Pedersen, N.L., Hägg, S., 2017. Biological age predictors. *EBioMedicine* 21, 29–36. <https://doi.org/10.1016/j.ebiom.2017.03.046>.
- Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., Zhavoronkov, A., 2017. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol. Pharm.* 14, 3098–3104. <https://doi.org/10.1021/acs.molpharmaceut.7b00346>.
- Kalmbach, K.H., Fontes Antunes, D.M., Draxler, R.C., Knier, T.W., Seth-Smith, M.L., Wang, F., Liu, L., Keefe, D.L., 2013. Telomeres and human reproduction. *Fertil. Steril.* 99, 23–29. <https://doi.org/10.1016/j.fertnstert.2012.11.039>.
- Kempermann, G., 2015. Activity dependency and aging in the regulation of adult neurogenesis. *Cold Spring Harb. Perspect. Biol.* <https://doi.org/10.1101/cshperspect.a018929>.
- Klapper, W., Heidorn, K., Kühne, K., Parwaresch, R., Krupp, Guido, 1998a. Telomerase activity in 'immortal' fish. *FEBS Lett.* 434, 409–412. [https://doi.org/10.1016/S0014-5793\(98\)01020-5](https://doi.org/10.1016/S0014-5793(98)01020-5).
- Klapper, W., Kühne, K., Singh, K.K., Heidorn, K., Parwaresch, R., Krupp, G., 1998b. Longevity of lobsters is linked to ubiquitous telomerase expression. *FEBS Lett.* 439, 143–146. [https://doi.org/10.1016/S0014-5793\(98\)01357-X](https://doi.org/10.1016/S0014-5793(98)01357-X).
- Lai, T.-P., Wright, W.E., Shay, J.W., 2018. Comparison of telomere length measurement methods. *Philos. Trans. R. Soc. B Biol. Sci.* 373, 20160451. <https://doi.org/10.1098/rstb.2016.0451>.
- Laradi, I., Babanezhad, R., 2018. M-ADDA: Unsupervised Domain Adaptation with Deep Metric Learning.
- Levine, M.E., Crimmins, E.M., 2018. Is 60 the new 50? Examining changes in biological age over the past two decades. *Demography*. <https://doi.org/10.1007/s13524-017-0644-5>.
- Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., Whitsel, E.A., Wilson, J.G., Reiner, A.P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., Horvath, S., 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10, 573–591. <https://doi.org/10.18632/aging.101414>.
- Li, Y., Chen, C.-Y., Wasserman, W.W., 2016. Deep feature selection: theory and application to identify enhancers and promoters. *J. Comput. Biol.* 23, 322–336. <https://doi.org/10.1089/cmb.2015.0189>.
- Lin, Y., Damjanovic, A., Metter, E.J., Nguyen, H., Truong, T., Najarro, K., Morris, C., Longo, D.L., Zhan, M., Ferrucci, L., Hodes, R.J., Weng, N., 2015. Age-associated telomere attrition of lymphocytes in vivo is co-ordinated with changes in telomerase activity, composition of lymphocyte subsets and health conditions. *Clin. Sci. (Lond.)* 128, 367–377. <https://doi.org/10.1042/CS20140481>.
- Lin, Q., Weidner, C.I., Costa, I.G., Marioni, R.E., Ferreira, M.R.P., Deary, I.J., Wagner, W., 2016. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)* 8, 394–401. <https://doi.org/10.18632/aging.100908>.
- Liu, M., Huo, Y.R., Wang, Junwei, Wang, C., Liu, Shuling, Liu, Shuai, Wang, Jinhuan, Ji, Y., 2016. Telomere shortening in Alzheimer's disease patients. *Ann. Clin. Lab. Sci.* 46, 260–265.
- Liu, Z., Kuo, P.-L., Horvath, S., Crimmins, E., Ferrucci, L., Levine, M., 2018. A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: a cohort study. *PLoS Med.* 15, e1002718. <https://doi.org/10.1371/journal.pmed.1002718>.
- Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D., Whitsel, E.A., Assimes, T.L., Ferrucci, L., Horvath, S., 2019. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)* 11, 303–327. <https://doi.org/10.18632/aging.101684>.
- Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.
- Makris, K., Spanou, L., 2011. Is there a relationship between mean blood glucose and glycated hemoglobin? *J. Diabetes Sci. Technol.* 5, 1572–1583. <https://doi.org/10.1177/193229681100500634>.
- Mamoshina, P., Kochetov, K., Putin, E., Cortese, F., Aliper, A., Lee, W.-S., Ahn, S.-M., Uhn, L., Skjold, N., Kovalchuk, O., Scheibye-Knudsen, M., Zhavoronkov, A., 2018a. Population specific biomarkers of human aging: a big data study using South Korean, Canadian, and Eastern European patient populations. *J. Gerontol. Ser. A* 73, 1482–1490. <https://doi.org/10.1093/gerona/gly005>.
- Mamoshina, P., Volosnikova, M., Ozerov, I.V., Putin, E., Skibina, E., Cortese, F., Zhavoronkov, A., 2018b. Machine learning on human diverse transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Front. Genet.* 9. <https://doi.org/10.3389/fgene.2018.00242>.
- Mamoshina, P., Kochetov, K., Cortese, F., Kovalchuk, A., Aliper, A., Putin, E., Scheibye-Knudsen, M., Cantor, C.R., Skjold, N.M., Kovalchuk, O., Zhavoronkov, A., 2019. Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Sci. Rep.* 9, 142. <https://doi.org/10.1038/s41598-018-35704-w>.
- Marioni, R.E., Shah, S., McRae, A.F., Ritchie, S.J., Muniz-Terrera, G., Harris, S.E., Gibson, J., Redmond, P., Cox, S.R., Pattie, A., Corley, J., Taylor, A., Murphy, L., Starr, J.M., Horvath, S., Visscher, P.M., Wray, N.R., Deary, I.J., 2015. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int. J. Epidemiol.* <https://doi.org/10.1093/ije/dyu277>.
- Marioni, R.E., Belsky, D.W., Deary, I.J., Wagner, W., 2018. Association of facial ageing with DNA methylation and epigenetic age predictions. *Clin. Epigenetics* 10, 140. <https://doi.org/10.1186/s13148-018-0572-2>.
- Martin-Herranz, D.E., Aref-Eshghi, E., Bonder, M.J., Stubbs, T.M., Choufani, S., Weksberg, R., Stegle, O., Sadikovic, B., Reik, W., Thornton, J.M., 2019. Screening for genes that accelerate the epigenetic aging clock in humans reveals a role for the H3K36 methyltransferase NSD1. *Genome Biol.* 20, 146. <https://doi.org/10.1186/s13059-019-1753-9>.
- Martin-Ruiz, C.M., Gussekloo, J., van Heemst, D., von Zglinicki, T., Westendorp, R.G.J., 2005. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. *Aging Cell*. <https://doi.org/10.1111/j.1474-9726.2005.00171.x>.
- Meer, M.V., Podolskiy, D.I., Tyshkovskiy, A., Gladyshev, V.N., 2018. A whole lifespan mouse multi-tissue DNA methylation clock. *Elife* 7. <https://doi.org/10.7554/eLife.40675>.
- Milne, E.M.G., 2006. When does human ageing begin? *Mech. Ageing Dev.* <https://doi.org/10.1016/j.mad.2005.11.003>.
- Mitteldorf, J.J., 2013. Telomere biology: cancer firewall or aging clock? *Biochem. J.* 78, 1054–1060. <https://doi.org/10.1134/S0006297913090125>.
- Nakamura, E., Miyao, K., Ozeki, T., 1988. Assessment of biological age by principal component analysis. *Mech. Ageing Dev.* 46, 1–18. [https://doi.org/10.1016/0047-6374\(88\)90109-1](https://doi.org/10.1016/0047-6374(88)90109-1).
- Nelson, S.M., Telfer, E.E., Anderson, R.A., 2013. The ageing ovary and uterus: new biological insights. *Hum. Reprod. Update*. <https://doi.org/10.1093/humupd/dms043>.
- O'Sullivan, R.J., Kubicek, S., Schreiber, S.L., Karlseder, J., 2010. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* 17, 1218–1225. <https://doi.org/10.1038/nsmb.1897>.
- Ohn, I., Kim, Y., 2019. Smooth Function Approximation by Deep Neural Networks with General Activation Functions. <https://doi.org/10.3390/e21070627>.
- Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zernakova, A., Schramm, K., Wilson, Y.A., Kobes, S.,

- Tukiaainen, T., Ramos, Y.F., Göring, H.H.H., Fornage, M., Liu, Y., Gharib, S.A., Stranger, B.E., De Jager, P.L., Aviv, A., Levy, D., Murabito, J.M., Munson, P.J., Huan, T., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., van Rooij, J., Stolk, L., Broer, L., Verbiest, M.M.P.J., Jhamai, M., Arp, P., Metspalu, A., Tserel, L., Milani, L., Samani, N.J., Peterson, P., Kasela, S., Codd, V., Peters, A., Ward-Caviness, C.K., Herder, C., Waldenberger, M., Roden, M., Singmann, P., Zeilinger, S., Illig, T., Homuth, G., Grabe, H.-J., Völzke, H., Steil, L., Kocher, T., Murray, A., Melzer, D., Yaghootkar, H., Bandinelli, S., Moses, E.K., Kent, J.W., Curran, J.E., Johnson, M.P., Williams-Blangero, S., Westra, H.-J., McRae, A.F., Smith, J.A., Kardia, S.L.R., Hovatta, I., Perola, M., Ripatti, S., Salomaa, V., Henders, A.K., Martin, N.G., Smith, A.K., Mehta, D., Binder, E.B., Nylocks, K.M., Kennedy, E.M., Klengel, T., Ding, J., Suchy-Dicey, A.M., Enquobahrie, D.A., Brody, J., Rotter, J.I., Chen, Y.-D.I., Houwing-Duistermaat, J., Kloppenburg, M., Slagboom, P.E., Helmer, Q., den Hollander, W., Bean, S., Raj, T., Bakhshi, N., Wang, Q.P., Oyston, L.J., Psaty, B.M., Tracy, R.P., Montgomery, G.W., Turner, S.T., Blangero, J., Meulenbelt, I., Ressler, K.J., Yang, J., Franke, L., Kettunen, J., Visscher, P.M., Neely, G.G., Korstanje, R., Hanson, R.L., Prokisch, H., Ferrucci, L., Esko, T., Teumer, A., van Meurs, J.B.J., Johnson, A.D., Johnson, A.D., 2015. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* 6, 8570. <https://doi.org/10.1038/ncomms9570>.
- Petkovich, D.A., Podolskiy, D.I., Lobanov, A.V., Lee, S.-G., Miller, R.A., Gladyshev, V.N., 2017. Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell Metab.* 25 <https://doi.org/10.1016/j.cmet.2017.03.016>. 954–960.e6.
- Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., Ostrovskiy, A., Cantor, C., Vijg, J., Zhavoronkov, A., 2016. Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY)* 8, 1021–1033. <https://doi.org/10.18632/aging.100968>.
- Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., Ostrovskiy, A., Cantor, C., Vijg, J., Zhavoronkov, A., 2017. Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY)* 8, 1021–1033. <https://doi.org/10.18632/aging.100968>.
- Putin, E., Asadulaev, A., Ivanenkov, Y., Aladinskiy, V., Sanchez-Lengeling, B., Aspuru-Guzik, A., Zhavoronkov, A., 2018. Reinforced adversarial neural computer for *de novo* molecular design. *J. Chem. Inf. Model.* 58, 1194–1204. <https://doi.org/10.1021/acs.jcim.7b00690>.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H., 2016. Generative Adversarial Text to Image Synthesis.
- Reichert, S., Stier, A., 2017. Does oxidative stress shorten telomeres in vivo? A review. *Biol. Lett.* 13. <https://doi.org/10.1098/rsbl.2017.0463>.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why Should I Trust You? Explaining the Predictions of Any Classifier.
- Richardson, G.D., Breault, D., Horrocks, G., Cormack, S., Hole, N., Owens, W.A., 2012. Telomerase expression in the mammalian heart. *FASEB J.* <https://doi.org/10.1096/fj.12-208843>.
- Russo, P., Carlucci, F.M., Tommasi, T., Caputo, B., 2017. From Source to Target and Back: Symmetric Bi-Directional Adaptive GAN.
- Sanders, J.L., Newman, A.B., 2013. Telomere length in epidemiology: a biomarker of aging, age-related disease, both, or neither? *Epidemiol. Rev.* 35, 112–131. <https://doi.org/10.1093/epirev/mxs008>.
- Sidorov, I., Kimura, M., Yashin, A., Aviv, A., 2009. Leukocyte telomere dynamics and human hematopoietic stem cell kinetics during somatic growth. *Exp. Hematol.* 37, 514–524. <https://doi.org/10.1016/j.exphem.2008.11.009>.
- Skulachev, V.P., Anisimov, V.N., Antonenko, Y.N., Bakeeva, L.E., Chernyak, B.V., Erichev, V.P., Filenko, O.F., Kalinina, N.I., Kapelko, V.I., Kolosova, N.G., Kopnin, B.P., Korshunova, G.A., Lichinitser, M.R., Obukhova, L.A., Pasyukova, E.G., Pisarenko, O.I., Roginsky, V.A., Ruuge, E.K., Senin, I.I., Severina, I.I., Skulachev, M.V., Spivak, I.M., Tashlitsky, V.N., Tkachuk, V.A., Vyssokikh, M.Y., Yaguzhinsky, L.S., Zorov, D.B., 2009. An attempt to prevent senescence: a mitochondrial approach. *Biochim. Biophys. Acta - Bioenergy.* <https://doi.org/10.1016/j.bbabbio.2008.12.008>.
- Strait, J.B., Lakatta, E.G., 2012. Aging-associated cardiovascular changes and their relationship to heart failure. *Heart Fail. Clin.* 8, 143–164. <https://doi.org/10.1016/j.hfc.2011.08.011>.
- Stubbs, T.M., Bonder, M.J., Stark, A.-K., Krueger, F., von Meyenn, F., Stegle, O., Reik, W., Reik, W., 2017. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* 18, 68. <https://doi.org/10.1186/s13059-017-1203-5>.
- Svenson, U., Nordfjäll, K., Baird, D., Roger, L., Osterman, P., Hellenius, M.-L., Roos, G., 2011. Blood cell telomere length is a dynamic feature. *PLoS One* 6, e21485. <https://doi.org/10.1371/journal.pone.0021485>.
- Thompson, M.J., Chwiałkowska, K., Rubbi, L., Lusi, A.J., Davis, R.C., Srivastava, A., Korstanje, R., Churchill, G.A., Horvath, S., Pellegrini, M., 2018. A multi-tissue full lifespan epigenetic clock for mice. *Aging (Albany NY)* 10, 2832–2854. <https://doi.org/10.18632/aging.101590>.
- Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial Discriminative Domain Adaptation.
- von Zglinicki, T., 2000. Role of oxidative stress in telomere length regulation and replicative senescence. *Ann. N. Y. Acad. Sci.* 908, 99–110.
- Wagner, W., 2017. Epigenetic aging clocks in mice and men. *Genome Biol.* 18, 107. <https://doi.org/10.1186/s13059-017-1245-8>.
- Wang, Y., Navin, N.E., 2015. Advances and applications of single-cell sequencing technologies. *Mol. Cell* 58, 598–609. <https://doi.org/10.1016/j.molcel.2015.05.005>.
- Wang, T., Tsui, B., Kreisberg, J.F., Robertson, N.A., Gross, A.M., Yu, M.K., Carter, H., Brown-Borg, H.M., Adams, P.D., Ideker, T., 2017. Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol.* 18, 57. <https://doi.org/10.1186/s13059-017-1186-2>.
- Wang, T., Johnson, T.S., Shao, W., Lu, Z., Helm, B.R., Zhang, J., Huang, K., 2019. BERMUDA: a novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes. *Genome Biol.* <https://doi.org/10.1186/s13059-019-1764-6>.
- Weidner, C., Lin, Q., Koch, C., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D., Jöckel, K.-H., Erbel, R., Mühleisen, T., Zenke, M., Brümendorf, T., Wagner, W., 2014. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* 15, R24. <https://doi.org/10.1186/gb-2014-15-2-r24>.
- Weng, N.P., Levine, B.L., June, C.H., Hodes, R.J., 1995. Human naive and memory T lymphocytes differ in telomeric length and replicative potential. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11091–11094.
- Werner, C., Fürster, T., Widmann, T., Pösch, J., Roggia, C., Hanhoun, M., Scharhag, J., Büchner, N., Meyer, T., Kindermann, W., Haendeler, J., Böhm, M., Laufs, U., 2009. Physical exercise prevents cellular senescence in circulating leukocytes and in the vessel wall. *Circulation.* <https://doi.org/10.1161/CIRCULATIONAHA.109.861005>.
- Williams, G.C., 1957. Pleiotropy, natural selection, and the evolution of senescence. *Evolution (NY)* 11, 398. <https://doi.org/10.2307/2406060>.
- Xia, X., Chen, W., McDermott, J., Han, J.-D.J., 2017. Molecular and phenotypic biomarkers of aging. *F1000Research* 6, 860. <https://doi.org/10.12688/f1000research.10692.1>.
- Zhang, Y., Hapala, J., Brenner, H., Wagner, W., 2017. Individual CpG sites that are associated with age and life expectancy become hypomethylated upon aging. *Clin. Epigenetics* 9, 9. <https://doi.org/10.1186/s13148-017-0315-9>.
- Zhao, Q., Hastie, T.J., 2017. Causal Interpretations of Black-Box Models.
- Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., Volkov, Y., Zhulov, A., Shayakhmetov, R.R., Zhebrak, A., Minaeva, L.I., Zagrebnyy, B.A., Lee, L.H., Soll, R., Madge, D., Xing, L., Guo, T., Aspuru-Guzik, A., 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-019-0224-x>.

Glossary

- BA: Biological age
 CA: Chronological age
 DFS: Deep feature selection
 DNAm: DNA methylation
 DNN: Deep neural network
 EN: Elastic net
 GAN: Generative adversarial network
 LDA: Linear discriminant analysis
 MAE: Mean absolute error
 medAE: median absolute error
 PFI: Permutation feature importance
 RRBS: Reduced representation bisulfite sequencing
 qPCR: quantitative PCR