Report: Predictive accuracy of epigenetic clocks in determining the epigenetic age of older RPE tissues.

Kendrix Kek 16 November 2020

Overview: This report aims to evaluate if the epigenetic clocks can characterise the biological age of eye tissues, specifically the retinal pigment epithelium (RPE). Using the data by Porter et al. (2019) (ArrayExpress submission number: E-MTAB-7183), I will present my main findings and the limitations that I have encountered in trying to reach the aim.

Methods: I put the samples through our stringent pre-processing pipeline, which included extensive technical and biological sample quality checks such as detection p-value, beadcount, M+U median intensities check, technical control metrics, contamination check, sample relations check, sample identity check, outlier detection and sex check. All samples passed quality control so no sample was removed. I performed within-array noob normalisation to complement the BMIQ normalisation in Horvath's online DNAm Age Calculator. After that, I uploaded the normalised data to Horvath's online clock (https://dnamage.genetics.ucla.edu/) and explored the output of multiple different epigenetic clocks (Horvath 2013, Hannum et al. 2013, Zhang et al. 2019, Horvath et al. 2018, Shireby et al. 2020, McEwen et al. 2019, Weidner et al. 2014) in the output file received.

Main Findings:

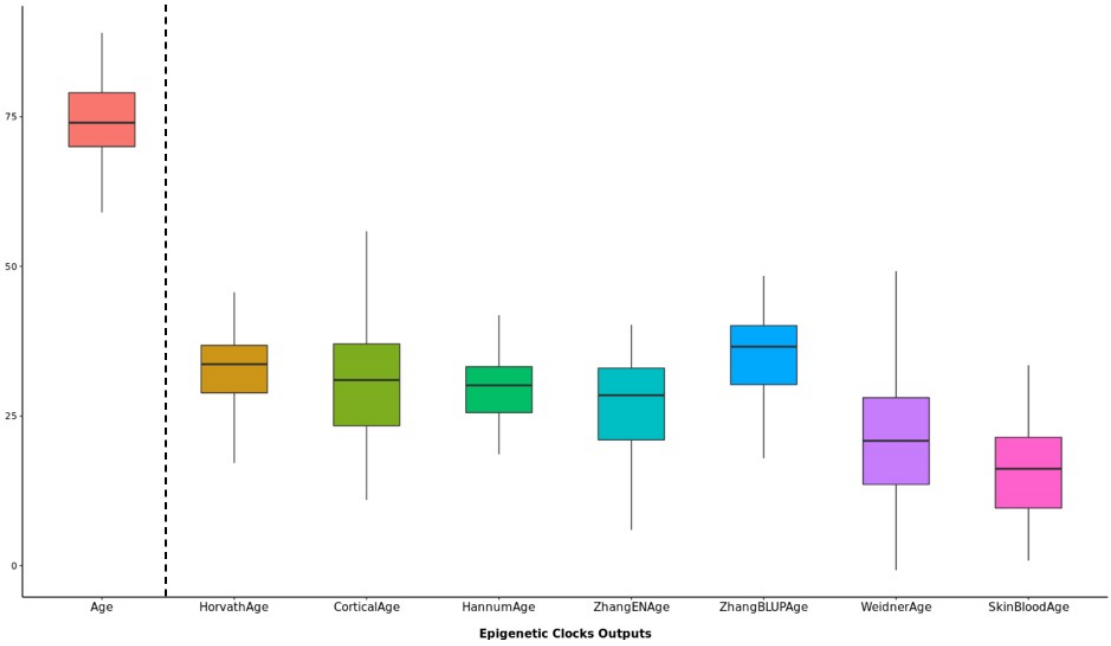

Fig 1. Systematic underestimation of predicted epigenetic age as compared to chronological age when Porter et al.'s (2019) RPE data are applied to various epigenetic clocks.

Observation: The epigenetic clock outputs show a systematic underestimation of the predicted ages as compared to the actual ages of the samples, where the predicted ages of the samples do not go beyond 50 years old. In fact, an evaluation of the RMSE of these predicted ages and the actual ages of the samples shows about 47 years of underprediction across all the clocks used.

Discussion: The existing epigenetic clocks are not trained on any RPE tissue, so the predictive accuracies of the clocks are low with respect to the RPE tissue tested, resulting in the systematic underestimation observed. However, the non-representation of the RPE tissue in the training sets of these clocks is not the sole cause of the underestimation observed here as there has been a consistent pattern of underestimation in literature for other tissues with a similar age range. For example, El Khoury et al. observe a similar systematic underestimation in four brain tissues from elderly individuals using Horvath's and Hannum's clocks, citing (1) the change in

methylation profiles towards saturation (0% or 100% methylation) later in life, (2) the epigenetic maintenance causing developmental changes in cell type as a response to the environment or simply drift, or decay, (3) the inherent properties of the brain tissues with higher levels of hydroxymethylation that are offsetting the predictive accuracy, and (4) the lack of older age groups in testing these clocks (El Khoury et al. 2019). More relevantly, Hewitt et al. observe a similar underestimation (**Fig 2.**) in a small set of ocular tissues with a similar age range (Hewitt et al. 2017). Interestingly, in another study examining the relationship between relative telomere length and epigenetic age, the group sees an overestimation of epigenetic age in younger samples (less than 60 years old) and underestimation in older samples (more than 80 years old) with a predictor that is based on their own cohort that contains a higher representation of samples between 60 to 80 years of chronological age (Banszerus et al. 2019). This observation seems to support the idea that the predictive accuracy of the clocks is highly dependent on the age range that they are trained on, as supported by Zhang et al.'s evaluation of their clock using a similar age range in their training and testing sets (Zhang et al. 2019).

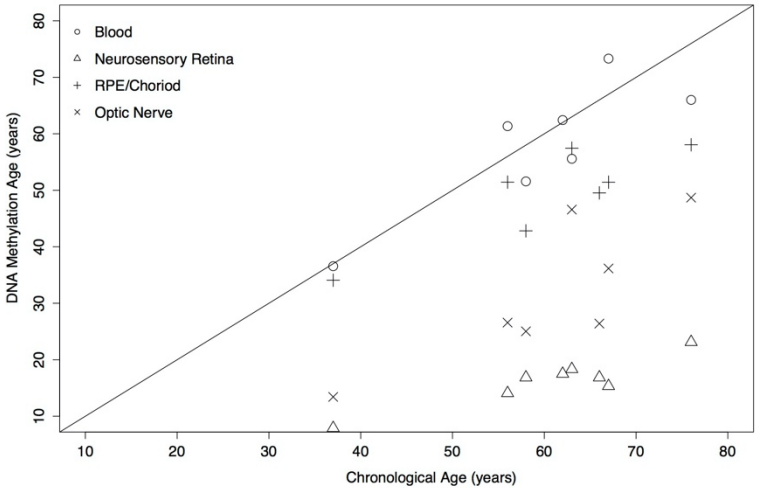

Fig 2. Underestimated epigenetic age of ocular tissues matched with blood of 8 individuals. Adapted from "Differences in epigenetic age of ocular tissue and the implications for eye disease", by Hewitt (2017).

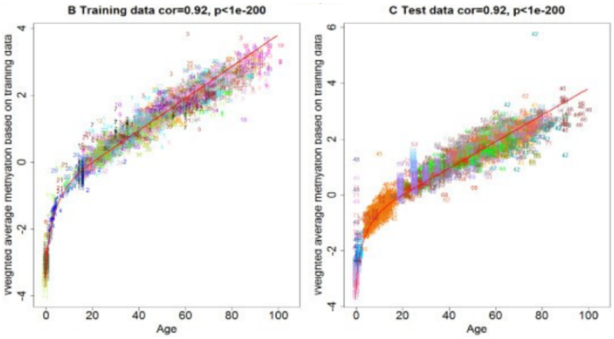

Fig 3. Methylation levels of 353 clock CpGs versus chronological age. Adapted from "DNA methylation age of human tissues and cell types", by Horvath (2013).

Combined, these studies raise an interesting question: Why do we consistently observe an underestimation in the epigenetic age of elderly individuals? The main feature of the epigenetic clocks lies in their ability to leverage our constant biological ticking rate to infer a linear relationship between the predicted epigenetic age and chronological age (**Fig. 3**). In his paper, Horvath shows that this linear relationship is not always prevalent throughout life, showing a "logarithmic dependence" until adulthood and gradual slowing towards linearity (Horvath 2013). It stands to reason that there is a possibility that this "logarithmic dependence" may exist later in life too.

Snir et al. propose that this logarithmic pattern is universally present across lifespan using their unbiased approach to model the trends in epigenetic aging of two different tissue types (blood and brain tissues) (Snir et al. 2019). Using their epigenetic universal pacemaker model that was inspired by evolutionary models of mutational rate change over time, they support Horvath's observation of non-linearity from late fetal stages to adolescence and linearity from early adulthood to old age (Snir et al. 2019). However, taking both tissue types into account across the entire lifespan – from before birth to old age, they observe a universal logarithmic relationship between epigenetic age and chronological age (Snir et al. 2019), with their approximation showing more rapid slowing later in life than Horvath's datasets (**Fig. 4**). This later in life pattern is substantiated by Marioni et

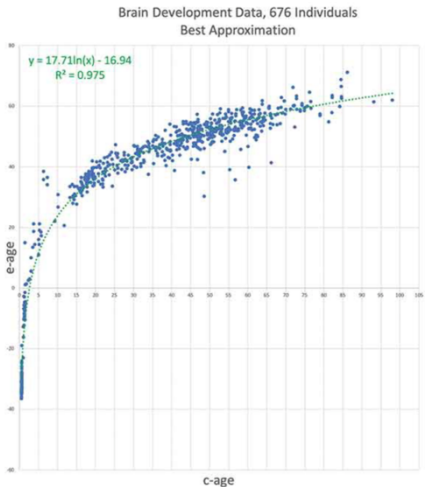

Fig 4. Epigenetic age versus chronological age of brain tissue. Adapted from "Human epigenetic aging is logarithmic with time across the entire lifespan", by Snir et al. (2019).

al.'s meta-analysis of epigenetic aging in longitudinal datasets that show a declining trend in the oldest populations with respect to their Δ_{age} trajectory plots (Marioni et al. 2019). They suggest that this effect may be driven by survival bias, which is substantiated by Nelson et al.'s analysis that shows first-generation epigenetic clocks do not account for mortality selection as they are often developed using cross-sectional data, and the resulting bias causes the machine learning algorithm to select markers that only correlative with aging at a younger age range, thus leaving out individuals with accelerated aging rates who experience higher mortality burden (Marioni et al. 2019, Levine 2020, Nelson et al. 2020).

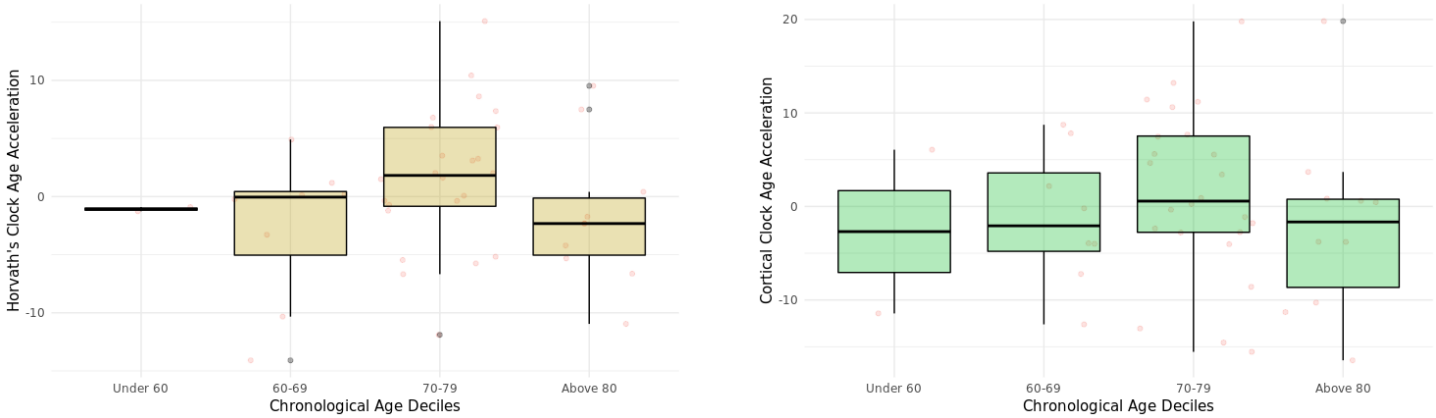

Fig 5. Age acceleration of Porter et al.'s (2019) RPE data using Horvath's (left) and Cortical Clock (right) for each age decile. For both clocks, the age accelerations seem to trend downwards after 80 years old, hinting at a possible non-linearity later in life.

Epigenetic Clock	DNAmAge CpGs	R	*p*-value	RMSE (years)
Horvath (Horvath 2013)	353	0.48	0.00096	41.21
Cortical (Shireby *et al.* 2020)	347	0.43	0.0034	43.93
Skin&Blood (Horvath *et al.* 2018)	391	0.41	0.0057	58.16
Zhang (BLUP) (Zhang *et al.* 2019)	514	0.39	0.0081	39.46
Hannum (Hannum *et al.* 2013)	71	0.36	0.015	45.15
Zhang (ElasticNet) (Zhang *et al.* 2019)	514	0.34	0.025	48.21
Weidner (Weidner *et al.* 2014)	3	0.30	0.046	53.45

Observation: Despite the low predictive accuracy, the predicted epigenetic ages are still significantly correlated with chronological ages across all epigenetic clocks. Horvath's predicted age shows the highest correlation with chronological age, which is unsurprising as Horvath's clock was built with the intention that it can be applied to any type of tissue (even though RPE tissues are not a part of its training set). As with any predictor, the composition of the training data used to build the clock influences the generality of the model, so the wide-ranging sample types, high sample size and large number of age-associated probes that are characteristic of Horvath's clock allow the clock to increase its predictive accuracy – even extending to out-of-sample prediction for tissues that are not part of its training set (Zhang et al. 2019, Shireby et al. 2020). However, there seems to be a limitation to the extent of the generalisation in tissues that are not a part of its training set. Intriguingly, the second most correlated predicted ages come from the tissue-specific Cortical clock. The commonality between these two clocks seems to be the fact that they are both trained on brain tissues (partly with Horvath's clock, and completely with the Cortical clock).

Discussion: The observation that the top two highest correlation coefficients between predicted epigenetic and chronological ages come from clocks that are trained on brain tissues seems to imply an interesting property of the clocks, which is the clocks seem to be able to extend their predictive performances to tissues that share common underlying tissue-specific markers – specifically, brain and RPE tissues. The observation also suggests that the predictive performance of the clocks is more of a result of the sum-of-their-parts rather than the individual age-associated CpG sites, which is supported by Zhang's discovery of redundant sites and the lack of common tissue-agnostic sites in these clocks. To look further into these ideas, I will explore them conceptually

by examining the properties of the RPE tissue in relation to brain tissue and the properties of the clocks in terms of their age-associated CpGs separately:

Properties of the tissue: I propose that the two significant correlations observed using the clocks (that are incidentally trained on brain tissues) are because the RPE tissues in my dataset share some tissue-specific markers with the brain tissues. These markers are then captured in the age-associated probes of these clocks, which allow for the significant predictive accuracy observed in the RPE tissues. Looking at the developmental pathway of the RPE tissue, I discover that it is developed in the same germ line as the other tissues that are used to train Horvath's clock such as neuronal tissues, as shown in **Fig. 6**. The authors suggest that the RPE layer was derived from an ancestral ectodermal cell containing light-sensitive pigments, which becomes the anterior neuroectoderm that eventually derives into the optic neuroepithelium and then the RPE (Martinez-Morales et al. 2004, Fuhrmann et al. 2013). Similarly, this neuroectoderm has also been described as the progenitor of neural tissues such as neurons and glia (Hartenstein & Stollewerk 2015). In fact, the neurotrophic factor, PEDF that is involved in regulating neurogenesis and stemness in the central nervous system (CNS) was first discovered in RPE tissues, and is responsible for the progenitor-like characteristics of RPE (Engelhardt et al. 2005, Brook et al. 2020). While this may not be the tissue-specific marker that allows for the highly correlative prediction observed, it puts a link between the development of neuronal and RPE tissues that potentially results in the retainment of some shared developmental markers in both tissues (Wang et al. 2010). The idea of having some shared developmental markers is further substantiated when a gene-directed reprogramming study of the RPE of chick embryos results in the production of neuron-like cells and neuron-like clusters (Wang et al. 2010).

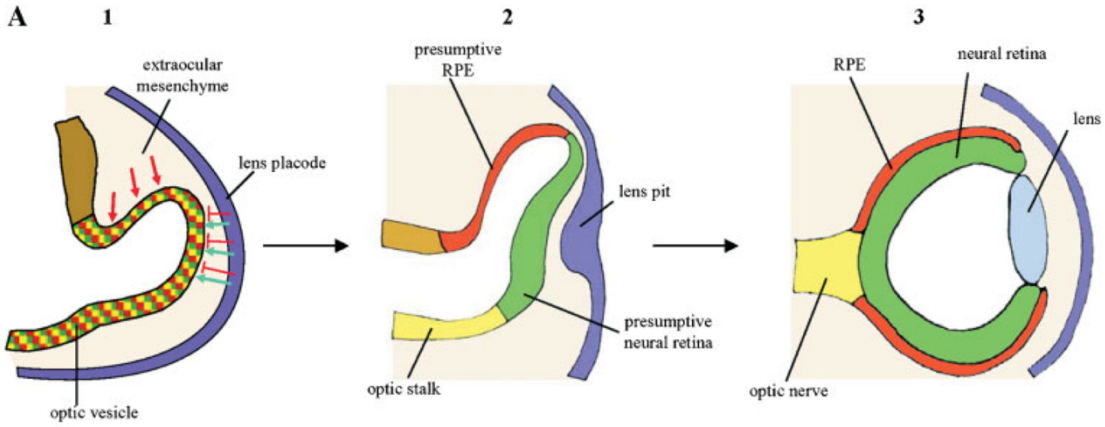

Fig 6. The development from an optic vesicle to the highly differentiated RPE layer. Adapted from "Eye development: a view from the retinal pigment epithelium", by Martinez-Morales et al. (2004).

When Horvath et al. conduct a study to test multiple tissues from the same individuals with his clock, they find that the cerebellum has a lower epigenetic age in 3 independent datasets and 6 individual centenarians (Horvath et al. 2015, Ashapkin et al. 2019). They cite (with caution) the overexpression of RNA helicase genes that may be playing a role in slowing down the epigenetic age of the cerebellum (Horvath et al. 2015). However, the exact mechanism of this action is currently not known as they do not find any prior literature linking the role of RNA helicase to tissue aging (Horvath et al. 2015). Given this shared lineage with RPE tissues, this study may not have uncovered the exact mechanism that causes the slowing in epigenetic age, but it reaffirms my finding of a similar slowing down after 80 years old as seen in **Fig. 5** and **Fig. 7**.

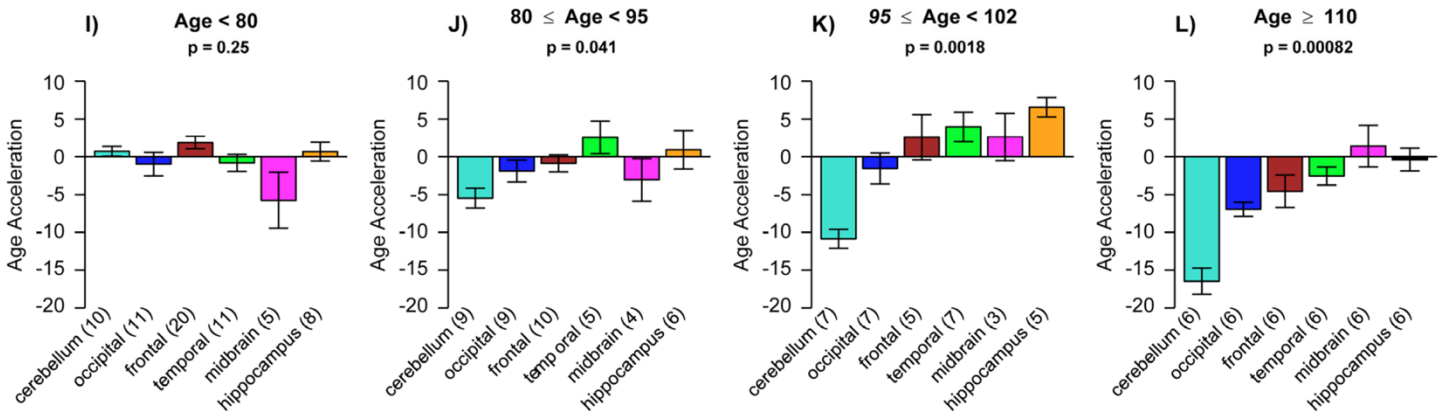

Fig 7. Gradual decrease in age acceleration observed in the cerebellum after 80 years old. Adapted from "The cerebellum ages slowly according to the epigenetic clock", by Horvath et al. (2015).

Properties of the clock: The Horvath's and Cortical clock consist of 353 and 347 age-associated CpG sites respectively, with 5 overlapping sites between them. From past literature, these 5 sites are not known to be associated with aging in brain tissues. More interestingly, Horvath's clock is not known to contain the common tissue-agnostic age-associated sites such as ELOVL2[1] and FHL2, which limit its predictive accuracy in tissues that it is not trained on (Bacalini et al. 2017, Slieker et al. 2018). The lack of these sites suggests that the ability of Horvath's clock to significantly predict the epigenetic age of RPE tissues is mainly due to the cumulative effect of a large number of age-associated sites that are "passengers" rather than key "drivers" of DNA methylation changes that could lead to the expression of aging-related traits – which seems to support Horvath et al.'s hypothesis that DNAm age may be a measure of the Epigenomic Maintenance System (EMS) (Horvath et al. 2015, Tajuddin et al. 2019). Perhaps the strongest evidence comes from Zhang et al.'s removal of some probes that they identify as redundant for age prediction in Horvath's clock and still showing better overall prediction accuracy (Zhang et al. 2019).

Limitations and future directions: Given the small sample size of only 44 samples in this dataset, it is important that the findings here are interpreted with caution. Nevertheless, given the drastic underestimation with significant correlation observed, it hints at a biological underpinning that needs to be explored to explain (1) why certain tissues show slow aging rate as compared to others, and (2) why is there a consistent observation of highly reduced age acceleration in older individuals? This conceptual exploration points to the need for an RPE-specific epigenetic clock to see if the first question is merely an artifact of non-representation of the tissue in the training sets of these currently available clocks. However, to build such a clock, one must address the challenge of a lack of availability of datasets in public repositories such as GEO and ArrayExpress as there are currently only 3 datasets[2] (2 already pre-processed and 1 with idats) available. As I have also discovered that this current dataset is riddled with batch effects, the results from the epigenetic clocks should be interpreted with caution as they may be confounded by batch effects to yield imperfect prediction performance (Zhang et al. 2019). Ultimately, this study may be used to serve as a backbone to explore the limitations of these epigenetic clocks, which may in turn result in a clock that is more applicable towards a wider age range while taking into account the logarithmic dependence in that age range. The success of this potential clock will help to determine the age acceleration of many age-related diseases in the future.

[1] although this argument is disputed by Zhu et al. (2018) who suggest that the clock makes up for the lack of ELOVL2 with almost 80% of the 353 CpG sites being defined as age-associated in more than 3 tissue types and are considered to be tissue-agnostic according to them.
[2] excluding datasets with cell lines.

References:

Ashapkin et al. (2019) Epigenetic Clock: Just a Convenient Marker or an Active Driver of Aging? In: Reviews on Biomarker Studies in Aging and Anti-Aging Research. *Advances in Experimental Medicine and Biology*. Vol. 1178. Springer, Cham.

Bacalini et al. (2017) Systemic Age-Associated DNA Hypermethylation of *ELOVL2* Gene: In Vivo and In Vitro Evidences of a Cell Replication Process. *J Gerontol A Biol Sci Med Sci*. 72(8): 1015 – 1023.

Banszerus et al. (2019) Exploring the Relationship of Relative Telomere Length and the Epigenetic Clock in LipidCardio Cohort. *Int J Mol Sci* 20(12): 3032.

Brook et al. (2020) Pigment epithelium-derived factor regulation of neuronal and stem cell fate. *Exp Cell Res*. 389(2): 111891.

El Khoury et al. (2019) Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biol*. 20(283).

Engelhardt et al. (2005) Adult retinal pigment epithelium cells express neural progenitor properties and the neuronal precursor protein doublecortin. *Brain Res*. 1040(1–2): 98 – 111.

Fuhrmann et al. (2013) Retinal pigment epithelium development, plasticity, and tissue homeostasis. *Exp Eye Res*. 123: 141 – 150.

Hannum et al. (2013) Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell*. 49(2): 359 – 367.

Hartenstein & Stollewerk (2015) The Evolution of Early Neurogenesis. *Dev Cell*. 32(4): 390 – 407.

Hewitt et al. (2017) Differences in epigenetic age of ocular tissue and the implication for eye disease. *ARVO Annual Meeting Abstract*. 58(8).

Horvath (2013) DNA methylation age of human tissues and cell types. *Genome Biol*. 14(3156).

Horvath et al. (2015) The cerebellum ages slowly according to the epigenetic clock. *Aging*. 7(5): 294 – 305.

Horvath et al. (2018) Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. *Aging*. 10(7): 1758 – 1775.

Levine (2020) Assessment of Epigenetic Clocks as Biomarkers of Aging in Basic and Population Research. *J Gerontol A Biol Sci Med Sci*. 75(3): 463 – 465.

Marioni et al. (2019) Tracking the Epigenetic Clock Across the Human Life Course: A Meta-analysis of Longitudinal Cohort Data. *J Gerontol*. 74(1): 57 – 61.

Martinez-Morales et al. (2004) Eye development: a view from the retinal pigment epithelium. *BioEssays*. 26: 766 – 777.

McEwen et al. (2019) The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *PNAS*. 117(38): 23329 – 23335.

Nelson et al. (2020) Biomarkers for Aging Identified in Cross-sectional Studies Tend to be Non-causative. *J Gerontol A Biol Sci Med Sci: Series A*. 75(3): 466 – 472.

Porter et al. (2019) Whole-genome methylation profiling of the retinal pigment epithelium of individuals with age-related macular degeneration reveals differential methylation of the *SKI, GTF2H4,* and *TNXB* genes. *Clin Epigenet*. 11(6).

Shireby et al. (2020) Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain*. awaa334.

Slieker et al. (2018) Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenetics & Chromatin*. 11(25).

Snir et al. (2019) Human epigenetic ageing is logarithmic with time across the entire lifespan. *Epigenetics*. 14(9): 912 – 926.

Tajuddin et al. (2019) Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. *Clin Epigenet*. 11(119).

Wang et al. (2010) Generating retinal neurons by reprogramming retinal pigment epithelial cells. *Expert Opinion on Biol Therapy*. 10(8): 1227 – 1239.

Weidner et al. (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*. 15(R24).

Zhang et al. (2019) Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med*. 11(54).

Zhu et al. (2018) Cell and tissue type independent age-associated DNA methylation changes are not rare but common. *Aging*. 10(11): 3541 – 3557.