Deconvolution of the epigenetic age discloses distinct inter-personal variability in epigenetic aging patterns

Tamar Shahal[1,2,#], Elad Segev[3#], Thomas Konstantinovsky[3,#], Yonit Marcus[1,6], Gabi Shefer[1], Metsada Pasmanik-Chor[4] , Assaf Buch[1], Yuval Ebenstein[2], Paul Zimmet[5], Naftali Stern[1,6]

1. The Sagol Center for Epigenetics of Aging and Metabolism, Institute of Endocrinology, Metabolism and Hypertension, Tel Aviv-Sourasky Medical Center; Sackler Faculty of Medicine, Tel Aviv University, Israel.
2. Department of chemistry, Tel Aviv University, Israel.
3. Department of Applied Mathematics, Holon Institute of Technology, Israel.
4. Bioinformatics Unit, The George S. Wise Faculty of Life Science, Tel Aviv University, Israel.
5. Department of Diabetes, Monas University School of Medicine, Melbourne, Australia
6. The Sackler Faculty of Medicine, Tel-Aviv University, Israel.

#contributed equally

Short title: Epigenetic age: a clue to personalized medicine

Address for correspondence and reprints:

Naftali Stern, MD

The Sagol Center for Epigenetics of Metabolism and Aging

Institute of Endocrinology, Metabolism and Hypertension

Tel Aviv Sourasky Medical Center

Sackler Faculty of Medicine, Tel Aviv University, Israel

Tel: 972-3-6973732 ; Fax: 972-3-6973035

e-mail: naftalis@tlvmc.gov.il

Abstract

Epigenetic age not only correlates with chronological age but predicts morbidity and mortality. We assumed that deconvolution of epigenetic age to its individual components could shed light on the diversity of epigenetic, and by inference, biological aging. Using the Horvath original epigenetic clock, we identified several CpG sites linked to distinct genes that quantitatively explain much of the interpersonal variability in epigenetic aging, with secretagogin and malin showing the most dominant effects. The analysis shows that the same epigenetic age for any given chronological age can be accounted for by variable contributions of identifiable CpG sites; that old epigenetic relative to chronological age is mostly explained by the same CpG sites, mapped to genes showing the highest interindividual variability differences in healthy subjects but not in subjects with type 2 diabetes. This paves the way to form personalized aging cards indicating the sources of accelerated/decelerated epigenetic aging in each examinee, *en route* to targeting specific sites as indicators, and perhaps treatment targets of personal undesirable age drifting.

Introduction

In the past decade, the concept of epigenetic age has attracted growing interest and the number of publications on epigenetic clocks has risen exponentially [1], mostly since it appears to reflect at least some aspects of the biological age. The epigenetic age can be now extrapolated using one of several independently generated epigenetic clocks, each mathematically constructed from time/age related changes in DNA methylation levels at specific multiple CpG sites that collectively, with proper weighting, are highly correlated with chronological age [2–6]. Discrepancies between chronological age and the calculated epigenetic are presumed to represent a measure of biological aging, such that epigenetic age acceleration /deceleration signifies accelerated or a

2

relatively diminished rate of biological aging, respectively. Hence, epigenetic clocks can be compared to individuals' chronological ages to assess inter-individual and/or inter-tissue variability in the rate of aging [1]. How useful and informative this approach could be is exemplified by reports that epigenetic age is a predictor of time of death, mortality rate [5–8] and susceptibility to diseases such as lung cancer [9], breast cancer [10] and cardiovascular events [11].

Time/age related methylation appears to be a rather extensive process as is readily demonstrated by the fact that there are several different epigenetic clocks, each calculated based on several tens or hundreds different methylation sites which mostly do not overlap[12]. Since these clocks vary with respect to their linkage to health outcomes, it is possible that each detects different processes which distinctly contribute to some facets of the biological age.

It is presently unknown whether the changes in methylation profiles which link epigenetic age acceleration to mortality and morbidity are simply aging markers or, perhaps, active players in the aging process. The implication of the latter is that reversal of epigenetic age could comprise a therapeutic target or at least, a measure of therapeutic success achieved by various pharmaceutical means[13] or perhaps lifestyle modification. For example, in murine studies, a reversal of the epigenetic age was achieved by reduced caloric intake [14]. Assuming that the methylation level of the CpG sites, comprising the epigenetic clock, affect specific aging routs through modulation of gene expression, interpersonal differences in the methylation degree of such sites could offer clues to not only differential aging rates, but to variability in aging mechanisms in human subjects.

In the present study we focused on the possibility that the epigenetic age might be individually determined by inter-person differences in the methylation levels of such sites. For example, if the epigenetic age is kept fixed at Z years, which reflects the sum of three CpG sites, A, B and C, how variable is the specific contribution of each of them

to the epigenetic age among different subjects? If A adds most years in one subject, but very little in another, might this reflect important differences in their aging driving mechanisms? What, if any, are the key epigenetic differences between "epigenetically young" and "epigenetically old" subjects of the same chronological age? This study, focuses on the interpersonal variability of the components (methylation levels at specific sites) comprising the epigenetic age, as a potential tool in predicting individual's physiological malfunction, towards the development of personalized medicine. To address this issue we analyzed the epigenetic age of publicly available methylation data of 1441 healthy individuals, of the ages of 40-80 years, retrieved from Illumina methylation arrays. Because of the shorter lifespan of subjects with diabetes mellitus we have also analyzed 135 samples from subjects with type 2 diabetes (DM2). The epigenetic age was calculated by Horvath's clock, which is based on coefficients, calculated by a regression model, relating the methylation status of 353 CpG sites (β- values) to chronological age.

Results

The epigenetic age distribution.

In the aim of assessing interpersonal variability in the aging mechanisms, we first calculated the epigenetic age of the 1441 samples. Figure 1 depicts the overall relation between epigenetic age and chronological age in the entire analyzed data set. Individuals with an epigenetic age of Avg ± 1SD, (between red and orange lines), comprise 71.5% of the data whereas 27% of the data is from people with an epigenetic age between Avg ± 1SD and Avg ± 2.5 SD (between the orange and the purple lines). The 1.5% of the data residing beyond the purple lines were ignored to avoid large effects of potentially uncertain results ("outliers"). The near linear increase of the epigenetic age with chronological age, demonstrates that Horvath's clock is suitable

4

as age predictor for this data set, and that the epigenetic age has a high variability between individuals with the same chronological age, thus implying that individuals may age differently.
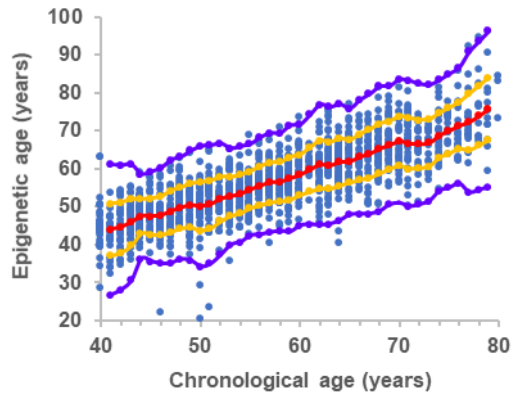

Figure 1: The overall relation between epigenetic age and chronological age. Each blue point represents a single healthy individual. Red dots are the average value of the epigenetic age at each chronological age connected by a regression curve (red line). All dots between the orange and the red line represent individuals with an epigenetic age between Avg and Avg ± 1SD. Dots residing between the orange and the purple lines represent individuals with an epigenetic age between Avg ± 1SD and Avg ± 2.5 SD.

CpG sites with the highest inter-personal variability: In order to find out the possible cause for interpersonal variations in the epigenetic aging, we have searched for CpG sites that were the most variable among subjects, in terms of years added/subtracted by that site to/from the total epigenetic age. Nine CpG sites that were found to be the most variable between individuals in 8 different age groups were selected as described in materials and methods. These sites and the genes within/ close to which they are located are listed in table 1, ranked from the one showing the largest inter-personal variation (overall SD) to the smallest. The two most dominant variable sites are related to secretagogin (SCGN) and malin [NHL Repeat Containing E3 Ubiquitin Protein Ligase 1 (NHLRC1)]. Eight probes out of the 9 most variable probes selected, were found to be independent of the population size (≥80% confidence). Specifically, secretagogin and malin were selected in 100% and 98% of the statistical simulations, respectively, indicating independence of the population size (table 3S in

supplementary file 1). Since there was no difference in the most variable epigenetic age components (CpG sites) between men and women in all chronological ages (supplementary file 4) we have treated the data with no gender distinction.

Table 1: CpG sites with the highest inter-personal methylation variability in healthy population

Related gene symbol	Related Gene definition/ product	Illumina's CpG ID	Contribution to epigenetic age (years)*	SD of the age contribution (years)
SCGN	Secretagogin	cg06493994	7.1	2.1
NHLRC1	Malin	cg22736354	9.3	1.8
MIR7-3HG	MIR7-3 Host Gene	cg02479575	1.8	1.0
FZD9	Frizzled 9	cg20692569	6.8	0.9
SCAP	SREBP cleavage-activating protein	cg26614073	-4.7	0.7
REEP1	Receptor expression enhancing protein 1	cg01968178	2.5	0.7
CSNK1D	Casein kinase 1 delta	cg19761273	-4.3	0.7
FXN	Frataxin	cg07158339	-4.6	0.7
NDUFS5	NADH dehydrogenase Fe-S protein 5	cg07388493	-4.8	0.7

*No sign for the contribution, in years, for each probe means a positive age contribution; a minus sign indicates a negative age contribution
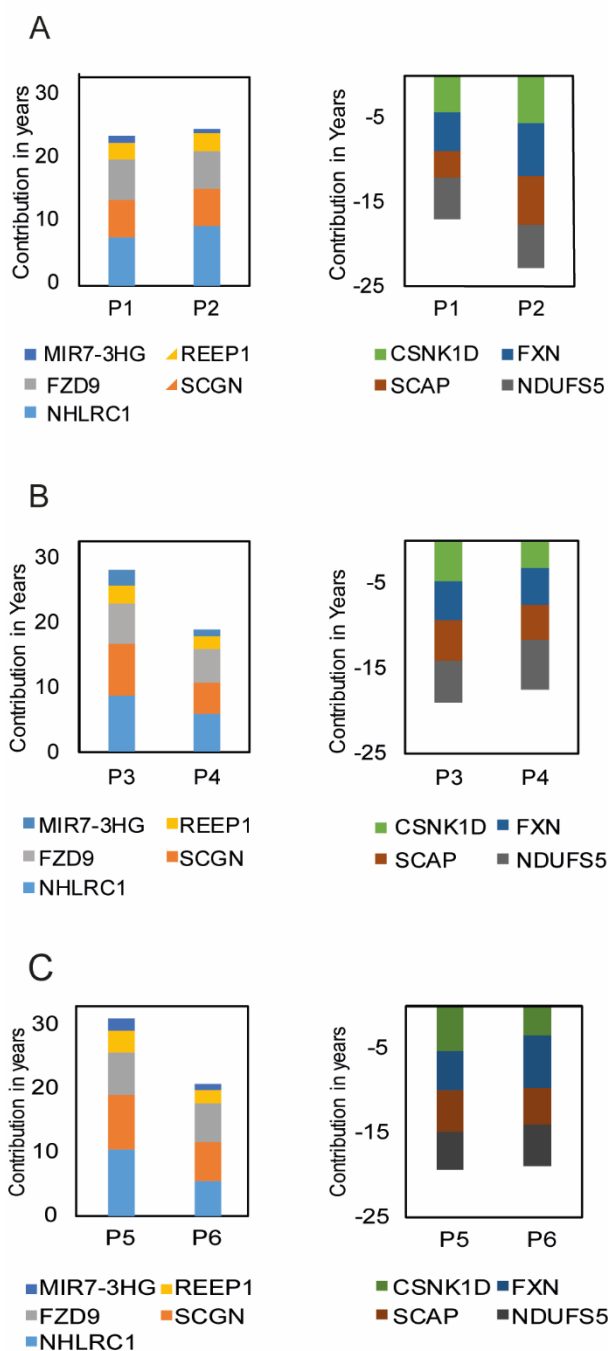
Notably, genes whose CpG sites showed the largest variation (SD) also contributed a sizable positive or negative age years to the calculated epigenetic age. However, inter-subject variability in methylation size effect was not simply a reflection of the magnitude of the contribution to age (in years). For example, CpG site on type 1 hair keratin, protein phosphatase 1 regulatory (inhibitor) subunit 14/7 and testis expressed sequence 286 genes, all of which had an effect exceeding +/-6.3 years, did not add considerably, on the average, to the overall difference (variability) in the makeup of the epigenetic age.

We have examined the correlation between the methylation levels of CpG sites with the highest inter-personal variability (from Horvath's clock) and other CpG sites residing on the same gene. The level of the correlation was found to be related to the proximity of the CpG sites to one another and their location on the gene or its regulatory elements (figure 1S in supplementary file 1, and supplementary file 5). This indicates that the methylation level of the site included in the calculation of the epigenetic age provides good representation of the methylation status of neighboring sites and is therefore likely to be related to gene expression, if it resides on the promoter or another regulatory region.

Large interpersonal variations in the epigenetic age **composition** in subjects with identical epigenetic age

Quantitative variability in the aging vectors was not only found in individuals with identical chronological age and different epigenetic age but also in subjects with identical chronologic and epigenetic age. Figure 2 illustrates the heterogeneity of the contribution to the epigenetic age (in years) of the nine CpG sites that tend to vary the most among subjects. Figure 2A, shows two men, both at the chronological age group of 40-41, with a similar epigenetic age of 40-41. Despite these similarities, malin (NHLRC1) contributes more than 9.5 years to the epigenetic age in man P2 but only 7.5 years in man P1, a difference which is offset by larger contributions of at least two CpG sites- linked to age lowering, which are mapped on SCAP and FXN genes. Next, figure 2B shows two men at a chronological age of 40-41 years. Larger "aging" contribution of all major positive contributors, generated a larger cumulative aging effect in subject P3 compared to subject P4, amounting to a pro-aging effect of 9.5 years. Since both men have an epigenetic age of 45-46 years, this is offset, to some extent, by methylation state of CpG sites that lower the epigenetic age, particularly, in this case, sites mapped on the genes: CSNK1D FXN and SCAP. In the final example (figure 2C), two men of the same chronological age group of 40-41 have a markedly

7

accelerated epigenetic aging of 50-51 years. However, age acceleration is driven by higher aging effect of CpG sites mapped to NHLRC1, SCGN and REEP1 in man P5, with a higher age-reducing effect of CSNK1D and SCAP, which only slightly make up for the stronger pro-aging CpG sites in this person. Presumably, the cumulative effect of other negative age contributors (that are less variable) are responsible for equalizing the epigenetic age in this later pair. This data implies that age acceleration at the same chronological age to the same higher or to the same equal epigenetic age can be reached by a highly variable methylation profile of the genes whose variation in general is the largest. This could signify that different mechanisms underlie the aging processes which people undergo.

Figure 2: Interpersonal variability in the epigenetic age composition of individuals with the same chronological age of 40-41 years old and the same epigenetic age of (A) 40-41 years, (B) 45-46 years and (C) 50-51 years. Left and right panels are for CpG sites on genes which add or subtract from the epigenetic age, respectively. The genes presented according to the color

codes at the bottom of each graph, are related to the CpG sites which belong to the nine sites with the highest variability

Key CpG sites as age accelerators or decelerators

Figure 1 shows that almost a third (27%) of the individuals in our data set, spanning all chronological ages, reside between Avg ± 1SD and Avg ± 2.5SD (dots between orange and purple lines). To detect which CpG sites (and their associated genes) were the major contributors to age acceleration we applied a "greedy algorithm" to the group of "epigenetic old" individuals (individuals whose epigenetic age resides between 1SD to 2.5SD, above the average epigenetic age line) and found that the site responsible for the largest "unfavorable aging" effect in 29% of the "epigenetically old" individuals, is mapped to the secretagogin gene (figure 3A). The remaining 70% of the "epigenetically old" individuals were then tested for the second largest ager, found to be malin. Once malin is consecutively normalized into the average zone, the epigenetic age of 12% of "epigenetically old" individuals is shifted to the average zone. This is followed by frataxin, responsible for 7%, and so on. Since frataxin is a negative age contributor, its effect is depicted by smaller age lowering vector. The entire group of genes related to the CpG sites, responsible for the accelerated aging of 80% of the healthy population, is presented in figure 3A. The CpG sites associated with secretagogin, malin MIR-7 and SCAP were found to be age accelerating components, independent of the population size (≥95% confidence, table 4S in supplementary file 1).

The same key players, with some change in the magnitude and order of their effect and some new effectors', stood out when we searched for the genes associated with the CpG sites responsible for lowering the epigenetic age (age decelerators) from the Avg - 1SD age zone to an epigenetic age of less than -1SD down to – 2.5 SD (Fig. 3B). Secretagogin and malin were the largest contributors to age deceleration, accounting for 25% and 14% of the "epigenetically young" individuals, respectively.

The cumulative percentage of individuals moving from the epigenetic old/young group to the Avg ± 1SD (black line in figure 3A and B) shows that up to 80% of the epigenetically old/young individuals had a single key prominent gene, responsible for accelerated or decelerated aging.
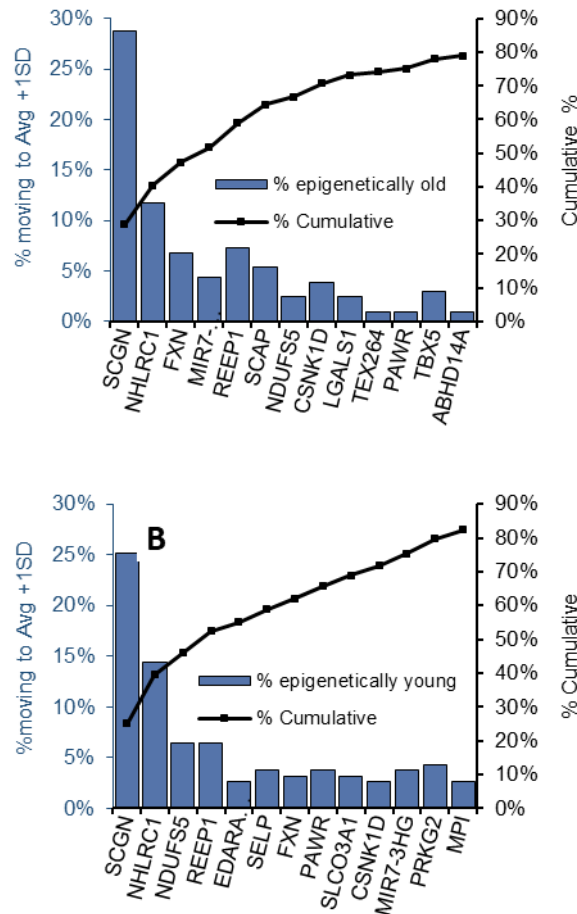

Figure 3: key genes associated with CpG sites, considered as age accelerators/ decelerators. (A) Blue bars present the percentage of individuals from the "epigenetically old" group which moved to the Avg + 1 SD group after setting consecutive CpG sites (presented with the name of their related gene) to their mean epigenetic age contribution, starting from the CpG site which moved the highest number of individuals to the lowest. (B) Blue bars are the percentage of individuals from the "epigenetically young" group who moved to the Avg - 1 SD after setting consecutive CpG sites to their mean epigenetic age contribution, starting from the site which moved the highest number of individuals to the lowest. The black line is the accumulative percentage of individuals moving from the "epigenetically old (A) /young (B)" group to the average group.

Collectively, then, many of the CpG sites or their associated genes that are responsible for interpersonal variation in the makeup of the epigenetic age, are also major players that act as "age accelerators" or "decelerators", depending on their methylation status.

Personalized epigenetic aging card

Finally, a putative "personal epigenetic aging status card" can be produced for each individual tested by the Horvath epigenetic clock. Such card, as shown in figure 4 for 7 individuals, all at the same chronological age of 40-41, grades each subject for the accelerating (orange to red cells, figure 4) or decelerating effect (green cells, figure 4), in years of each of the major CpG sites with the largest SDs. The grades are the relative deviation of the age contribution of each CpG site from its average contribution to the epigenetic age. If the difference between the calculated epigenetic age of a certain individual and the average epigenetic age cannot be significantly explained by these sites, the clock can be further interrogated to reveal other sites, with less common age effect, which might explain deviations from the average epigenetic age. This process may eventually evolve as an individualized panel of aging effects, much like a routine biochemistry panel as presently assessed at the clinician's office to detect indicators of disease, by their actual deviation from the normal range.

EA (years)	Person	NHLRC1	SCGN	FZD9	REEP1	MIR7-3HG	CSNK1D	FXN	SCAP	NDUFS5
41.3	P1	-0.4	-0.2	0.0	0.1	-0.2	0.3	0.4	1.7	0.4
40.6	P2	1.5	-0.4	-0.2	0.2	-0.6	-1.0	-1.2	-0.9	0.2
46.5	P3	0.8	1.5	0.2	0.4	1.0	-0.2	0.4	0.1	0.1
45.4	P4	-2.0	-1.7	-0.9	-0.6	-0.3	1.3	0.7	0.6	-0.7
45.3	P6	1.0	1.7	-0.2	-0.4	0.9	-0.8	0.2	0.3	0.6
50.8	P5	2.4	1.9	0.5	1.0	0.5	-0.8	0.4	-0.2	0.7
50.6	P7	2.0	0.7	0.7	0.0	0.3	-0.8	0.7	-0.1	0.4

Figure 4: Personalized epigenetic gene card: colored cells are the deviation of the epigenetic age contribution (in years) of the CpG sites or their associated genes from their average epigenetic age contribution, for seven individuals at chronological age of 40-41 years (sample number is # GSM). The average epigenetic age for 40-41 years old men, is 45.7 years. We show two samples ~5 years below epigenetic average age (light gray), three samples at average epigenetic age (darker gray) and two samples ~5 years above epigenetic average age (dark gray). Cells with orange to red colors are for genes associated with CpG sites with age contribution above average. Cells in

11

green, or light green associate with CpG sites on genes with age contribution below average.

Aging variability and the CpGs responsible for accelerated aging in a cohort of type 2 diabetes mellitus (DM2)

In order to find the CpG sites responsible for accelerated aging in DM2, we first calculated the epigenetic age of 135 subjects at the ages of 45-79 years, in parallel to the age span of the healthy subjects. Using Horvath's epigenetic clock, there was no significant difference between the average epigenetic age of the diabetes cohort and the healthy cohort (fig. 2S in supplementary file 1). In addition, seven out of the nine most variable CpG sites are shared by the healthy and diabetic samples, with some difference in the order of the size of the standard deviation (in years). Secretagogin, malin and a new player, beta-1,3-Galactosyltransferase 6 (B3GALT6) were the leading genes related to the most variable CpG sites (table 2S in supplementary file 1). Six probes, out of the 9 most variable probes selected, were found to be independent of the population size (≥78% confidence). Specifically, secretagogin and malin were re-selected in 99% of statistical simulations, indicating independence of the population size (table 3S in supplementary file 1). These results emphasize how robust the selection of the CpG sites is for secretagogin and malin as the most interpersonally variable sites in all age groups.

We analyzed the group of "epigenetic old", diabetic subjects (the 21 individuals whose epigenetic age resides between 1SD to 2.5SD, above the average epigenetic age line) by the greedy algorithm method, in order to find the CpG sites accounting for the accelerated ageing of these specific individuals (figure 2S in supplementary file 1). In this analysis, a different group of CpG sites is responsible for accelerating aging in the diabetic cohort compared to the healthy population (Figure 5): 70% of the subjects could have "acquired" an average +1SD epigenetic age once the pro-aging effects of the CpG sites linked to casein kinase 1- delta 1 isoform (CSNK1D), NADH

dehydrogenase (ubiquinone) Fe-S protein 5 (NDUFS5), small nuclear ribonucleotin polypeptide B (SNRPB2) and structure specific recognition protein 1 (SSRP1), are removed. Of these, only CSNK1D and NDUFS5 were present among the CpG sites that contributed to the shifting of age to above Average+1SD in the healthy cohort, but at a different order of magnitude. The independence of these results of the population size was found to be at a confidence of ≤ 25%. This can be explained by the small number of "epigenetically old" individuals in the diabetic cohort. Although these results may not, therefore, reflect the pattern of the diabetic population at large, they do point out a different age acceleration pattern for this specific diabetic cohort relative to the healthy cohort.
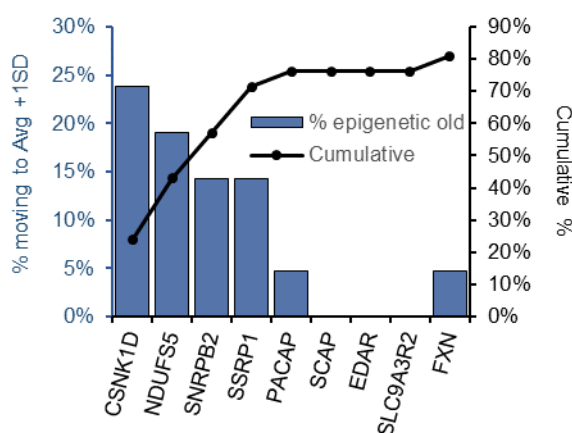

Figure 5: key genes associated with CpG sites, considered as age accelerators in the diabetic cohort. Blue bars present the percentage of individuals from the "epigenetically old" group which moved to the Avg + 1 SD group after setting consecutive CpG sites (presented with the name of their related gene) to their mean epigenetic age contribution, starting from the CpG site which moved the highest number of individuals to the lowest. The black line is the accumulative percentage of individuals moving from the "epigenetically old group to the average group.

Discussion

Following the pioneering reports by Horvath [2] and Hannum [3], it is now widely recognized that epigenetic age and chronological age correlate well with each other in many populations, regardless of the tissue studied [2]. In several publications a potential

13

functional role has been ascribed to the CpG sites that participate in the algorithms developed to derive the epigenetic age [4]. For example, accelerated epigenetic age relative to chronological age is reportedly linked to preferential activation of pro-inflammatory and interferon pathways, along with reduced stimulation of transcriptional/translational machinery, blunted DNA damage response, and weakened mitochondrial signatures [4,15]. This supports the notion that the epigenetic age does not simply mirror randomly the passage of time but reflects specific anti-homeostatic effects that may lead to or indicate specific unfavorable conditions which facilitate disease and affect life span.

The possibility that specific epigenetic aging drivers can be targeted to achieve personalized epigenetic/biological age deceleration, can become a testable approach. In the present study we examined which of the CpG sites included in the original Horvath DNAm algorithm are the major contributors to inter-personal differences in the epigenetic age. We found that the CpG sites related to malin and secretagogin have relatively high contribution to the epigenetic age and the most variable methylation status in between individuals both in the healthy and the diabetes cohorts. Additionally, in the healthy cohort, differences in the methylation status of secretagogin and malin, contributed more than any other methylation loci, to the difference between epigenetically old and epigenetically young subjects, across the entire age span screened by us (40-80yrs) (figure 3). This was not the case, however, in samples from diabetic subjects, in whom neither secretagogin's nor malin's-related methylation loci accounted for the higher epigenetic age in subjects who were epigenetically old. This may indicate that different biological mechanisms are involved in the accelerated aging of "epigenetically old" in the diabetes subjects analyzed, relative to "epigenetically old" healthy subjects (figure 5).

Notably, both the secretagogin and malin-related CpG sites are among the 5 shared by three known DNA-mAge epigenetic clocks: Hannum DNAm Age score (based on

71 methylation sites)[3], Horvath DNAm Age measure (353 sites) [2] and the DNAm PhenoAge score (513 sites) [6] and were also suggested to be the most dominant key age predictor sites [16]. Might these genes, then, be mechanistically involved in aging?

The malin gene encodes a RING type E3-ubiquitin ligase which forms a functional complex with laforin, a glucan phosphatase [17]. Mutations in either malin or laforin in humans lead to the development of Lafora progressive myoclonus epilepsy, a rare fatal neurodegenerative disease with early manifestations in the early childhood. Brain damage is incurred due to deposition underbranched and hyperphosphorylated insoluble glycogen in the brain and peripheral tissues [18–20]. It is notable that glucan deposits have been described in the setting of aging animals and humans [21–23], unrelated to LaFora disease, which raises the possibility of lesser malin activity with age. Indeed, malin appears to participate in a delicate homeostatic network linking neuronal glycogen synthesis and energetic utilization, interacting with autophagy, mitochondrial function and response to thermal stress, which could collectively affect lifespan [17,23–26]. The possibility that malin expression, which is critical for inhibition of polyglucan deposits in neurons, plays a role in healthful longevity in humans is intriguing and requires targeted research. In animal studies malin deficiency can lead to impaired autophagy and accumulation of dysfunctional mitochondria, which eventually promote neurodegeneration, immune disorders, cancer, and accelerated aging [27].

Secretagogin is an intracellular calcium sensor and facilitator of insulin secretion by pancreatic islet beta cells [28]. Recently it was shown that secretagogin play a critical role in the second phase of glucose-stimulated insulin secretion [29], protect against insulin aggregation and enhance peripheral response to insulin [30]. Concordant with this broad role in carbohydrate handling, secretagogin knockout leads to hyperglycemia [31]. Secretagogin is also expressed in neuroendocrine cells where it likely regulates exocytosis and hormone release [32,33]. Concordantly, it is also involved in danger

15

avoidance behavior through the control of post synaptic cell-surface availability of NMDA receptors in the central amygdala [34]. We are not aware, however, of published reports examining the relation between induced changes in secretagogin expression and lifespan or longevity.

Of major interest in the Horvath algorithm are CpG sites with a negative contribution to the epigenetic age, such as frataxin. Frataxin is a nuclear-encoded mitochondrial protein which is part of the Fe-S-cluster-containing proteins acting as an iron chaperone, thereby allowing normal function of the mitochondrial respiratory chain [35]. In our analysis frataxin shows both high interpersonal variability and also partly explains some (~8%) of the calculated age difference between epigenetically old and average subjects (Figure 3A). The fact that higher methylation of frataxin can extend life, as indirectly suggested by its epigenetic age lowering effect is somewhat counterintuitive: defects in the expression of this mitochondrial protein cause the neurodegenerative syndrome of Friedreich's ataxia [36,37], which is also accompanied by cardiomyopathy, diabetes mellitus and reduced life expectancy [38]. However, inactivation of many mitochondrial genes in the nematode Caenorhabditis elegans by RNAi was actually shown to extend lifespan [39]. Ventura et al reported that suppression of the frataxin homolog gene (frh-1) prolonged lifespan in the nematode, along with an altered phenotype of smaller size, diminished fertility and variant responses to oxidative stress. Thus, whereas sizable inactivation of frataxin causes a disabling disease, a more moderate frataxin suppression, such as achieved by RNAi could lead to higher lifespan as seen in C. elegans [40]. There is evidence that frataxin silencing induces mitochondrial autophagy as an evolutionarily conserved response to the ensuing iron starvation [35]. In a broader sense, lesser frataxin availability might comprise a surmountable challenge which elicits mitophagy that eventually preconditions the cell's capacity to sustain future stress, thereby increasing the likelihood of extended lifespan.

16

Overall, our analysis reveals sizable interpersonal differences in the contribution to age of methylation sites of several genes. It is also possible that there is a shift in the epigenetic age vectors in diabetes mellitus patients, that are not necessarily detected by the computation of the mean epigenetic age *per se*. In the healthy cohort, genes such as, but not limited to, secretagogin, malin and frataxin stand out in terms of either the size of their effect on interpersonal differences in the composition of the epigenetic age as well as their influence on the likelihood for an individual to acquire enhanced or delayed epigenetic aging. This analysis also unravels that even healthy subjects with average epigenetic aging could show accelerated aging with respect to some genes. In the same venue, epigenetically healthy older subjects are also heterogeneous and could be pushed to unfavorable epigenetic drifting by different aging vectors. The high epigenetic age of the "epigenetically old" diabetic individuals was driven by a repertoire of genes that are mostly different from the cluster driving accelerated aging in the healthy cohort, potentially implying a different aging mechanism in such conditions. Not necessarily representative of aging in DM2 at large, the findings add further strength to the notion of heterogeneity of epigenetic aging. This paves the way for future attempts to personalize the perception of epigenetic aging by deconvolution, addressing aging not as a general process in search of reversal, but as a collection of individual effects requiring personalized attention.

Materials and methods

Methylation Data

The β values which reflects the methylation status of each CpG site were retrieved from the Gene Expression Omnibus (GEO) Datasets repository. Following filtration for whole blood and healthy subjects at the relevant age range we obtained 2298 samples, from 23 different data sets. All filtered samples were normalized using an R-code

17

provided by Horvath et al.[2]. Samples that failed Horvath's normalization process [2], were removed, leaving a total of 1,441 samples for interrogation, (867 females and 574 males). We also analyzed a distinct cohort of 135 whole blood samples from subjects of the same age range (55 females and 80 males), diagnosed with diabetes, a disease possibly linked to accelerated biological aging [41–43]. Detailed description on the datasets and the processes involved in the selection of the samples we have analyzed can be found in table 1S, detailed materials and methods section, in supplementary file 1.

Epigenetic vs Chronological age:

For each sample we extracted the β values of Horvath's 353 CpGs clock sites and converted them to age contribution based on their coefficients, calculated as explained by Horvath et al. [2]. The epigenetic age of each of the 1,441 individuals is the sum of the contribution plus a constant (representing the intercept of the linear correlation), in years, of all 353 CpGs (supplementary file 3). In order to smooth the average (Avg) and standard deviation (SD), we used a running average and a running standard deviation of the epigenetic age with a window size of 3 (details in materials and methods section of supplementary file 1).

CpG sites with the highest inter-personal variability

We divided the 1441 healthy samples to 8 chronological data sets by age groups spanning from the age of 40 to the age of 80 years regardless of the gender (40-45 (189); 46-50 (215); 51-55 (217); 56-60 (223); 61-65 (220); 66-70 (177); 71-75 (120); 76-80 (80), years (number of subjects)). The 135 diabetic subjects were divided to 7 chronological age groups spanning from the age 45 to the age of 80 regardless of gender (45-49 (6); 50-54 (10); 55-59 (10); 60-64 (6); 65-69 (28); 70-74 (34); and 75-80 (50), years (number of subjects)). At each chronological data set, we recorded the 20 CpG sites from Horvath's clock, with the highest inter-individual variability (standard

18

deviation, SD, in years). We then identified 9 out of these 20 CpG sites with the largest inter-individual variability, which were also consistently present in each of the 8 chronological age data sets of the healthy cohort and 6 of the 7 chronological age data sets of the diabetes cohort. A statistical simulation, examining the relation between the identified CpG sites and the size of the cohort was manifested as explained in table 3S in supplementary file 1.

CpG sites as age accelerators or decelerators

Key epigenetic age accelerators or decelerators were found by looking for the probes with the highest cumulative contribution to the epigenetic age. The data set was divided to: 1) the "epigenetically average" group including all samples with epigenetically age of the running average ± 1 SD (the population in between the two orange lines), 2) the "epigenetically old/young" group, with an epigenetic age between 1 SD and 2.5 SD above or below the average (the population in between the upper/lower orange and purple lines, respectively), 3) the outliers, which have an epigenetic age with more than 2.5 SD from the average (the population above the upper or below the lower purple lines).

For the "epigenetically old" and the "epigenetically young" population, a greedy algorithm was applied. The algorithm calculates, for each probe, the number of individuals that moved from the "epigenetically old" or the "epigenetically young" to the "epigenetically average" group, as a result of setting a particular probe to its mean epigenetic age contribution value (in years). In each iteration, the algorithm selects the probe which moves the largest number of samples into the "average group": In the first iteration, the CpG site selected is the one which moves the highest number of subjects into the average zone, (by setting it to its average value). In the second iteration, the probe selected is the one which moves the most individuals in the residual "epigenetic older/ younger" zone to the average epigenetic age group and so on. The bars in the

graphs shown in figure 3 and 5 presents the percentages of individuals from the entire 1441 population, passing from the "epigenetically old/ young" to the average group when all **consecutive** CpG sites are set to their mean epigenetic age contribution (supplementary file 1)

Personalized epigenetic aging card

A personal epigenetic card is presented for 7 healthy individuals, with chronological age of 40-41 years, as the deviation (in years) from the mean epigenetic age contribution of each of the 9 chosen probes. The mean epigenetic age contribution of each probe is the average addition/subtraction of each probe, to/ from the average epigenetic age at 40-41 years.

Supplementary material:

Supplementary file 1: Detailed materials and method, tables 1S-4S and figures 1S-2S

Supplementary file 2: an R code for the conversion of idat values β values

Supplementary file 3: The age contribution, in years, for each of the 353 CpG sites comprising Horvath's clock, for each of the 1441 individuals

Supplementary file 4: The SD and the rank by SD of the most variable CpG sites from Horvath's clock in three different populations: Men, Woman and gender mixed population

Supplementary file 5: The correlation between the methylation of the CpG site included in Horvath clock and other CpG sites related to the same gene, as a function of their distance, for the nine selected most variable genes found in all age groups.

Author contribution statement

Acknowledgment

References

1.   Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* vol. 19 (2018).

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

3.   Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, (2013).

4.   Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany. NY).* **11**, (2019).

5.   McCrory, C. *et al.* GrimAge Outperforms Other Epigenetic Clocks in the Prediction of Age-Related Clinical Phenotypes and All-Cause Mortality. *Journals Gerontol. Ser. A* (2020) doi:10.1093/gerona/glaa286.

6.   Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany. NY).* **10**, (2018).

7.   Chen, B. H. *et al.* DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging (Albany. NY).* **8**, (2016).

8.   Zhang, Y. *et al.* DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat. Commun.* **8**, (2017).

9.   Levine, M. E. *et al.* DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging (Albany. NY).* **7**, (2015).

10. Ambatipudi, S. *et al.* DNA methylome analysis identifies accelerated epigenetic ageing associated with postmenopausal breast cancer susceptibility. *Eur. J. Cancer* **75**, (2017).

11. Roetker, N. S., Pankow, J. S., Bressler, J., Morrison, A. C. & Boerwinkle, E. Prospective Study of Epigenetic Age Acceleration and Incidence of Cardiovascular Disease Outcomes in the ARIC Study (Atherosclerosis Risk in Communities). *Circ. Genomic Precis. Med.* **11**, (2018).

12. Liu, Z. *et al.* Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell* **19**, (2020).

13. Fahy, G. M. *et al.* Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, (2019).

14. Hahn, O. *et al.* Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol.* **18**, (2017).

15. Irvin, M. R. *et al.* Metabolic and inflammatory biomarkers are associated with epigenetic aging acceleration estimates in the GOLDN study. doi:10.1186/s13148-018-0481-4.

16. Vidaki, A. *et al.* DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Sci. Int. Genet.* **28**,

(2017).

17.   Sanchez-Martin, P., Lahuerta, M., Viana, R., Knecht, E. & Sanz, P. Regulation of the autophagic PI3KC3 complex by laforin/malin E3-ubiquitin ligase, two proteins involved in Lafora disease. *Biochim. Biophys. Acta - Mol. Cell Res.* **1867**, (2020).

18.   Tagliabracci, V. S. *et al.* Abnormal metabolism of glycogen phosphate as a cause for Lafora disease. *J. Biol. Chem.* **283**, 33816–33825 (2008).

19.   Nitschke, F. *et al.* Abnormal glycogen chain length pattern, not hyperphosphorylation, is critical in Lafora disease. *EMBO Mol. Med.* **9**, 906–917 (2017).

20.   Roach, P. J. Glycogen phosphorylation and Lafora disease. *Mol. Aspects Med.* **46**, 78–84 (2015).

21.   Cavanagh, J. B. Corpora-amylacea and the family of polyglucosan diseases. *Brain Research Reviews* vol. 29 (1999).

22.   Cisse, S., Perry, G., Lacoste-Royal, G., Cabana, T. & Gauvreau, D. Immunochemical identification of ubiquitin and heat-shock proteins in corpora amylacea from normal aged and Alzheimer's disease brains. *Acta Neuropathol.* **85**, (1993).

23.   Sinadinos, C. *et al.* Neuronal glycogen synthesis contributes to physiological aging. *Aging Cell* **13**, (2014).

24.   Sun, R. C. *et al.* Nuclear Glycogenolysis Modulates Histone Acetylation in Human Non-Small Cell Lung Cancers. *Cell Metab.* **30**, (2019).

25.   Bilanges, B., Posor, Y. & Vanhaesebroeck, B. PI3K isoforms in cell signalling and vesicle trafficking. *Nature Reviews Molecular Cell Biology* vol. 20 (2019).

26. Sengupta, S., Badhwar, I., Upadhyay, M., Singh, S. & Ganesh, S. Malin and laforin are essential components of a protein complex that protects cells from thermal stress. *J. Cell Sci.* **124**, (2011).

27. Levine, B. & Kroemer, G. Biological Functions of Autophagy Genes: A Disease Perspective. *Cell* vol. 176 (2019).

28. Wagner, L. *et al.* Cloning and expression of secretagogin, a novel neuroendocrine- and pancreatic islet of Langerhans-specific Ca2+-binding protein. *J. Biol. Chem.* **275**, (2000).

29. Yang, S. Y. *et al.* Secretagogin affects insulin secretion in pancreatic β-cells by regulating actin dynamics and focal adhesion. *Biochem. J.* **473**, (2016).

30. Sharma, A. K. *et al.* Secretagogin Regulates Insulin Signaling by Direct Insulin Binding. *iScience* **21**, (2019).

31. Malenczyk, K. *et al.* A TRPV 1-to-secretagogin regulatory axis controls pancreatic β-cell survival by modulating protein turnover . *EMBO J.* **36**, (2017).

32. Dong, Y. *et al.* Secretagogin, a marker for neuroendocrine cells, is more sensitive and specific in large cell neuroendocrine carcinoma compared with the markers CD56, CgA, Syn and Napsin A. *Oncol. Lett.* **19**, (2020).

33. Romanov, R. A. *et al.* A secretagogin locus of the mammalian hypothalamus controls stress hormone release. *EMBO J.* **34**, (2015).

34. Hevesi, Z. *et al.* Secretagogin marks amygdaloid PKCδ interneurons and modulates NMDA receptor availability. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

35. Schiavi, A. *et al.* Iron-Starvation-Induced Mitophagy Mediates Lifespan Extension upon Mitochondrial Stress in C. elegans. *Curr. Biol.* **25**, (2015).

36.   Lynch, D. R., Farmer, J. M., Balcer, L. J. & Wilson, R. B. Friedreich ataxia:
       Effects of genetic understanding on clinical evaluation and therapy. *Archives
       of Neurology* vol. 59 743–747 (2002).

37.   Campuzano, V. *et al.* Friedreich's Ataxia: Autosomal Recessive Disase
       Caused by an Intronic GAA Triplet Repeat Expansion. *Science (80-.).* **271**,
       1423–1427 (1996).

38.   M.H., P., S., B., W., N., C., M. & P., G. Clinical features of Friedreich's ataxia:
       Classical and atypical phenotypes. *J. Neurochem.* **126**, (2013).

39.   SS, L. *et al.* A systematic RNAi screen identifies a critical role for mitochondria
       in C. elegans longevity. *Nat. Genet.* **33**, (2003).

40.   Ventura, N. *et al.* Reduced expression of frataxin extends the lifespan of
       Caenorhabditis elegans. *Aging Cell* **4**, (2005).

41.   Palmer, A. K. *et al.* Cellular senescence in type 2 diabetes: A therapeutic
       opportunity. *Diabetes* **64**, 2289–2298 (2015).

42.   Burton, D. G. A. & Faragher, R. G. A. Obesity and type-2 diabetes as inducers
       of premature cellular senescence and ageing. *Biogerontology* **19**, (2018).

43.   Soriano-Tárraga, C. *et al.* Epigenome-wide association study identifies TXNIP
       gene associated with type 2 diabetes mellitus and sustained hyperglycemia.
       Hum. Mol. Genet. **25**, 609–619 (2016).