Epigenetic profile of Japanese supercentenarians: a cross-sectional study

Shohei Komaki, Masatoshi Nagata, Eri Arai, Ryo Otomo, Kanako Ono, Yukiko Abe, Hideki Ohmomo, So Umekage, Natsuko O Shinozaki, Tsuyoshi Hachiya, Yoichi Sutoh, Yayoi Otsuka-Yamasaki, Yasumichi Arai, Nobuyoshi Hirose, Akio Yoneyama, Hideyuki Okano, Makoto Sasaki, Yae Kanai, Atsushi Shimizu**

Summary

Background Centenarians and supercentenarians with exceptional longevity are excellent models for research towards improvements of healthy life expectancy. Extensive research regarding the maintenance and reduction of epigenetic age has provided insights into increasing healthy longevity. To this end, we explored the epigenetic signatures reflecting hallmarks of exceptional healthy longevity, including avoidance of age-related diseases and cognitive functional decline.

Methods In this cross-sectional study, we enrolled Japanese non-centenarians (eligible participants aged 20–80 years) from the Tohoku Medical Megabank Community-Based Cohort Study and centenarians and supercentenarians (aged 101–115 years) from the Tokyo Centenarian Study and the Japanese Semi-supercentenarian Study. We assessed participants' whole-blood DNA methylation profiles and then developed sex-specific and non-specific first-generation epigenetic clocks by elastic net regression, calculated individuals' epigenetic ages, and assessed their age acceleration. We also screened for age-related CpG sites in non-centenarians by epigenome-wide linear regression analyses and ANOVA. We subsequently investigated which CpG sites in centenarians and supercentenarians had DNA methylation patterns following the age-related findings obtained from non-centenarians and which did not. We further characterised CpG sites with hypermethylation or hypomethylation in the centenarians and supercentenarians using enrichment and protein–protein interaction network analyses.

Findings We enrolled 421 non-centenarians (231 [55%] women and 190 [45%] men; age range 20–78 years), recruited between May 20, 2013, and March 31, 2016, and 94 centenarians and supercentenarians (66 women [70%] and 28 [30%] men; age range 101–115 years), recruited between Jan 20, 2001, and April 17, 2018. Non-sex-specific epigenetic clock showed the highest accuracy (r=0·96) based on which centenarians and supercentenarians had negative epigenetic age acceleration. Epigenome-wide association analyses further showed that centenarians and supercentenarians had younger-than-expected epigenetic states (DNA methylation profiles similar to those of non-centenarians) for 557 CpG sites enriched in cancer-related and neuropsychiatric-related genes, whereas these individuals had advanced (or older) epigenetic states for 163 CpG sites represented by genes related to TGF-β signalling, which is involved in anti-inflammatory responses and known to contribute to healthy ageing.

Interpretation These results indicate that exceptionally healthy longevity depends not only on maintaining young epigenetic states but also on advanced states of specific epigenetic regions.

Funding The Japan Agency for Medical Research and Development, KDDI Research, and Keio University.

Copyright © 2023 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Lancet Healthy Longev 2023; 4: e83–90

See **Comment** page e57

For the Japanese translation of the Summary see **Online** for appendix 1

*Joint senior authors

Division of Biomedical Information Analysis (S Komaki PhD, R Otomo PhD, K Ono MSc, H Ohmomo PhD, S Umekage PhD, N O Shinozaki MSc, T Hachiya PhD, Y Sutoh PhD, Y Otsuka-Yamasaki PhD, Prof A Shimizu PhD), **Iwate Tohoku Medical Megabank Organization** (Prof M Sasaki MD), **Institute for Biomedical Sciences** (S Komaki, H Ohmomo, Prof M Sasaki, Prof A Shimizu), **Iwate Medical University, Iwate, Japan; Healthcare Medical Group, KDDI Research, Saitama, Japan** (M Nagata MSc, A Yoneyama PhD); **Department of Pathology** (E Arai MD, Prof Y Kanai MD), **Center for Supercentenarian Medical Research** (Y Abe BA, Prof Y Arai MD, N Hirose MD), **Department of Physiology** (Prof H Okano MD), **Keio University School of Medicine, Tokyo, Japan**

Correspondence to: Prof Atsushi Shimizu, Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Yahaba, 028-3694, Japan
ashimizu@iwate-med.ac.jp

Introduction

As human longevity increases, limiting illness—both in its duration and number of individuals affected—becomes ever more important.[1] Centenarians (individuals aged 100 years or older) and supercentenarians (individuals aged 110 years or older) are excellent models for the study of healthy longevity, given that they typically do not have age-related diseases and maintain physical and cognitive function for longer than individuals who do not reach such ages.[2–5]

Analysis of the status of DNA methylation, a major epigenetic modification mediating the interplay between genetic and environmental cues, has proven promising in elucidating the mechanisms underlying exceptional healthy longevity. The epigenetic clock, based on DNA methylation, is widely accepted and used to estimate chronological age or identify biological age.[6] Maintaining a young epigenetic age or reducing epigenetic age (through rejuvenation) is gaining attention because it is correlated with biological senescence in health and disease.[7] In Italian semi-supercentenarians (aged 105–109 years), epigenetic age was found to be lower than chronological age,[8] suggesting an epigenetic basis for their exceptional healthy longevity.

Although centenarians and supercentenarians have remarkably young epigenetic ages, which epigenetic

Research in context

Evidence before this study
We searched PubMed using the term "epigenetic age" for studies published from inception to March 22, 2022, with no language restrictions. We found 427 articles, including 40 review articles, most including keywords such as rejuvenation, reversal, and acceleration, indicating an increasing interest in the maintenance of young epigenetic age. 20 reports analysed the epigenetic profiles of centenarians, for whom young epigenetic age has been shown. However, which specific epigenetic regions are maintained young or not in centenarians is not yet known.

Added value of this study
The epigenetic age of centenarians and supercentenarians is known to be young, and we have further explored the genomic regions related to specific functions representing exceptional healthy longevity: genes involved in cancer and cognitive function showed a young methylation state, whereas immune-related genes, such as those involved in TGF-β signalling (linked to healthy ageing), displayed an advanced state (accelerated age-related demethylation).

Implications of all the available evidence
Although the importance of maintaining a young epigenetic state is widely recognised, ageing is inevitable. As well as young epigenetic states, accelerated (or old) epigenetic states of appropriate genomic regions are also essential for healthy ageing and exceptional longevity.

functions and regions remain in a young state is unclear. Little evidence exists for epigenetic signatures of the features underlying such longevity, such as a reduced prevalence of age-related diseases and the maintenance of physical and cognitive function. To the best of our knowledge, only Gentilini and colleagues[9] have assessed the potential (indirect) epigenetic signature of healthy centenarian longevity, showing a single differentially methylated tumour-suppressor gene in the offspring of centenarians. With this study, we aimed to explore whether supercentenarian longevity depends only on a young epigenetic state, or also on other epigenetic factors.

Methods

DNA methylation analysis

DNA methylation in non-centenarians

We selected individuals aged 20–80 years, all apparently healthy, from the Tohoku Medical Megabank Community-Based Cohort Study.[10] Our selection criteria were: non-smoker (<100 cigarettes during lifetime); no previous diagnosis of cancer or diabetes; and not obese (BMI <25 kg/m²). We used a self-directed questionnaire to collect data on sex, age, smoking status, and morbidity.

Whole-blood genomic DNA was extracted and processed by bisulfite conversion. Using custom-designed oligonucleotide probe sets specialised in DNA methylation analysis of the Japanese population (appendix 2 p 2), we measured the DNA methylation profiles of 1·4 million genome-wide CpG sites (appendix 2 p 2). On the basis of these DNA methylation profiles, we also estimated the proportions of natural killer cells, B cells, CD4+ T cells, CD8+ T cells, monocytes, and neutrophils (appendix 2 p 2).

See Online for appendix 2

DNA methylation in centenarians and supercentenarians

We collected whole-blood samples from centenarians and supercentenarians in the Tokyo Centenarian Study and the Japanese Semi-supercentenarian Study.[11] Because centenarians and supercentenarians' are quite rare, we used all available samples, enabling us to detect CpG sites with 5% DNA methylation deviation and 90% statistical power (appendix 2 pp 2, 19). We collected blood samples from centenarians and supercentenarians at multiple timepoints upon participants' consent (appendix 2 pp 7, 9). DNA methylation profiles of these specimens were assessed as in non-centenarians. On the basis of their DNA methylation profiles, we further assessed the epigenetic identity of the samples collected from the same individuals to help to ensure that no sample mishandling occurred (appendix 2 p 3).

Outcomes

We defined two outcomes for the epigenetic characterisation of centenarians and supercentenarians: the difference between chronological and epigenetic ages (epigenetic deacceleration), and the deviation in DNA methylation pattern from the age-related findings seen in non-centenarians.

Development of sequence-based epigenetic clocks

The widely available epigenetic clock algorithms require microarray DNA-methylation data from, for example, the Infinium HumanMethylation 450K and MethylationEPIC microarrays, and estimating epigenetic age by fitting our sequence-based data to these clocks was potentially inappropriate because of the bias in the absolute values of DNA methylation level required for epigenetic-age calculation between the platforms. Therefore, we calculated epigenetic age with the available clocks (appendix 2 p 3), and also developed and assessed first-generation epigenetic clocks to predict the chronological age of the non-centenarians on the basis of targeted-bisulfite sequencing data, using published methods.[12,13] For each age group (20–79 years) and sex, we randomly selected 80% of the individuals for epigenetic-clock training, using the rest for testing. To investigate sex differences in epigenetic ageing, we considered three clocks based on female-only training, male-only training, and female and male training. For each training group,

we used elastic net regression to model chronological age, with 8-fold cross-validation, and selected informative CpG sites to predict chronological age. To assess the three epigenetic clocks, we calculated the epigenetic ages of the non-training individuals under each clock and assessed associations with chronological age with Pearson's correlations.

Centenarian and supercentenarian epigenetic age calculation

We used the most accurate of the three epigenetic clocks to calculate centenarian and supercentenarian epigenetic ages, as well as previously published clocks developed from microarray data (appendix 2 p 3). We calculated epigenetic age acceleration (primary outcome) for each person as the residual from a multiple regression line, obtained by regressing epigenetic age on chronological age and the estimated blood-cell counts in non-centenarians (epigenetic age acceleration adjusted for cell type). To remove the bias resulting from multiple sampling of individuals, only samples collected at the oldest age of each individual were included in the calculation.

We further assessed the longitudinal change in epigenetic age in centenarians and supercentenarians, using the blood samples collected at multiple timepoints. For each of these individuals, we did a regression analysis with chronological age as the independent variable and epigenetic age as the dependent variable; the resultant coefficients were considered to be the mean annual change in epigenetic age. To identify epigenetic ageing in centenarians and supercentenarians, we used one-sample *t* tests, with a null hypothesis of zero annual change in epigenetic age.

Characterisation of the centenarian and supercentenarian DNA methylation profile

To characterise the epigenome of centenarians and supercentenarians and the associated biological consequences, we identified which epigenomic regions were maintained young and which were not (secondary outcome). For this purpose, we first screened for CpG sites associated with chronological age using an epigenome-wide association study (EWAS) in non-centenarians. EWAS is generally done by adjusting for cell-type composition, which affects the DNA methylation pattern. However, the cell-type composition was also correlated with chronological age, and adjusting EWAS for such a potential mediator is not appropriate, because doing so can distort the magnitude of the relationship between dependent and explanatory variables. Additionally, lifespan differs between sexes, and possible sex-specific factors exist for longevity. For these reasons, we used a conservative analytical approach. We first did EWAS for each sex using simple linear regression of DNA methylation profiles against chronological age, with Bonferroni correction of significance thresholds; CpG sites that were significantly associated with age in the same direction for both sexes were considered associated with age. To account for the cell-type composition, we further did multiple linear regression of DNA methylation profiles against the proportion of the six blood-cell types. Using ANOVA, we then compared the fit of the simple and multiple regression models. After Bonferroni correction, we excluded CpG sites whose DNA methylation profiles were significantly better explained by the blood cell-type proportion. We considered the remaining significantly positively or negatively age-associated CpG sites in downstream analyses. To characterise the analytical approach combining sex-specific EWAS and ANOVA, we also did non-sex-specific EWAS adjusted for sex and cell-type composition and compared the results.

Using the simple linear regression coefficients in non-centenarians and intercepts for each age-associated CpG, we estimated the expected DNA methylation profiles for the centenarian and supercentenarian CpG sites on the basis of their chronological ages. The expected and observed DNA methylation profiles were then compared using two-sample *t* tests with Bonferroni correction. We selected CpG sites that were significantly hypermethylated or hypomethylated in centenarians and supercentenarians.

From these findings, we defined four CpG categories: first, negatively age-associated in non-centenarians and hypermethylated in centenarians and supercentenarians (termed negHyper); second, positively age-associated in non-centenarians and hypermethylated in centenarians and supercentenarians (termed posHyper); third, positively age-associated in non-centenarians and hypomethylated in centenarians and supercentenarians (termed posHypo); and fourth, negatively age-associated in non-centenarians and hypomethylated in centenarians and supercentenarians (termed negHypo; appendix 2 p 18). The posHyper category remained empty and was therefore excluded.

Annotation and characterisation of CpG sites unique to centenarians and supercentenarians

For the three CpG categories and for all CpG sites in our dataset (background), we did CpG and genic annotations with the annotatr R package,[14] obtaining CpG annotations (CpG islands, shores, shelves, and open sea), genic annotations (including exons, introns, and promoters), and gene annotations. For the posHypo, negHyper, and negHypo categories, we did enrichment and protein–protein interaction network analyses. For enrichment analyses, we used four frameworks, namely Gene Ontology annotations[15] of biological process, cellular component, and molecular function, and Kyoto Encyclopedia of Genes and Genomes pathways[16], through the missMethyl[17] R package, with a slight modification to handle our targeted bisulfite sequencing-derived data. We considered terms and pathways with hypergeometric

	Non-centenarians (n=421)		Centenarians and supercentenarians (n=94)	
	Female (n=231)	Male (n=190)	Female (n=66)	Male (n=28)
Age, years	50 (35–65)	50 (37–66)	108 (107–109)	106 (105–107)
Biochemistry measurements				
Alanine aminotransferase, IU/L	16·16 (7·51)	23·47 (11·33)	11·18 (5·66)	14·54 (6·10)
Aspartate aminotransferase, IU/L	21·51 (6·14)	24·63 (7·06)	21·70 (7·33)	26·43 (7·78)
HDL, mg/dL	70·94 (16·31)	61·60 (16·18)	48·79 (11·01)	51·54 (13·20)
LDL, mg/dL	116·65 (29·44)	114·18 (28·92)	109·72 (27·43)	90·21 (18·29)
HbA_{1c}, %	5·47 (0·32)	5·36 (0·34)	5·54 (0·31)	5·53 (0·30)
Total cholesterol, mg/dL	204·48 (35·60)	195·55 (33·67)	178·42 (34·00)	160·25 (22·54)
Triglycerides, mg/dL	95·66 (51·49)	114·82 (73·66)	99·58 (45·75)	92·50 (54·63)
γ-glutamyl transpeptidase, IU/L	21·11 (19·39)	35·58 (40·01)	15·71 (11·75)	32·64 (40·94)
Questionnaire responses				
Smoking status				
Non-smoker*	231 (100%)	190 (100%)	62 (94%)	27 (96%)
Current smoker	0	0	1 (2%)	1 (4%)
Missing data	0	0	3 (5%)	0
Cancer diagnosis				
Never diagnosed	231 (100%)	190 (100%)	65 (98%)	27 (96%)
Skin cancer	0	0	1 (2%)	0
Prostate cancer	0	0	0	1 (4%)
Diabetes	0	0	0	0
Educational attainment				
Completed elementary or junior high school	37 (16%)	39 (21%)	55 (83%)	19 (68%)
Completed high school	101 (44%)	64 (34%)	5 (8%)	6 (21%)
Completed business college	45 (19%)	33 (17%)	0	0
Completed junior college or technical college	21 (9%)	7 (4%)	0	0
Completed University	21 (9%)	38 (20%)	1 (2%)	3 (11%)
Completed graduate school	2 (<1%)	3 (2%)	0	0
Other	2 (<1%)	4 (2%)	5 (8%)	0
Missing data	2 (<1%)	2 (1%)	0	0
Physical measurements				
BMI, kg/m²	21·14 (2·23)	22·21 (1·90)	20·10 (3·30)	20·09 (3·04)
Diastolic blood pressure, mm Hg	69·93 (10·36)	74·17 (9·76)	75·36 (12·26)	74·25 (15·47)
Systolic blood pressure, mm Hg	118·16 (17·30)	123·61 (16·67)	141·57 (23·08)	143·43 (26·55)

Data are n (%), mean (SD), or median (IQR). Non-centenarians were aged 20–78 years. Centenarians and supercentenarians were aged 101–115 years. *Non-centenarians who smoked fewer than 100 cigarettes in their lifetime were considered non-smokers. For centenarians and supercentenarians, individuals who are not currently smoking were considered non-smokers.

Table: **Baseline characteristics of non-centenarian, centenarian, and supercentenarian participants**

test false discovery rate-adjusted p value below 0·05 to be significantly enriched.

We did the protein–protein interaction network analyses using the STRING web database, specifying *Homo sapiens* as the target organism.[18] The obtained protein–protein interactions were further analysed with the igraph R package[19] to detect network hub proteins. For hub protein detection, we adopted degree centrality, which measures the number of interactions a given protein participates in. Assuming that closely connected proteins in the networks share features and functions, we manually surveyed the literature to characterise each protein cluster.

We additionally assessed the overlaps between CpG categories and certain sets of CpG sites and genes, such as age-related CpG sites, cancer-related genes, tumour-suppressor genes, and immune-related genes (appendix 2 pp 3–4). Because some CpG sites can be linked to neighbouring single-nucleotide polymorphisms, we also did methylation quantitative trait loci analyses, with whole-genome sequencing and whole-genome bisulfite sequencing datasets in the iMETHYL database.

This study was approved by the Ethics Committees of Iwate Medical University (HG-H25-2) and Keio University (2002120). All participants provided written informed consent.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We included 421 non-centenarians (231 [55%] women and 190 [45%] men), aged 20–78 years, recruited between May 20, 2013, and March 31, 2016 (table; figure 1; appendix 2 pp 7, 18), and 94 centenarian and supercentenarians (66 [70%] women and 28 [30%] men) aged 101–115 years, recruited between Jan 20, 2001, and April 17, 2018 (table; appendix 2 pp 7–8).

From the non-centenarians, we obtained DNA methylation profiles for 757 461 CpG sites (100% call rate in both sexes; appendix 2 p 10). The sex-specific principal component analyses (appendix 2 p 20) uncovered no outliers; hence, we used all non-centenarians' data to develop the epigenetic clock. Clock training with data from both men and women produced the highest correlation between epigenetic age and chronological age (appendix 2 pp 11, 21). All three epigenetic clocks produced accurate predictions, with Pearson correlation coefficients between 0·944 and 0·970, similar to those of widely applied microarray-based clocks.[12] By contrast, epigenetic ages calculated on the basis of previously published microarray-based clocks had a poorer correlation with chronological age (appendix 2 p 22). Less than 3% of clock CpG sites of the newly developed epigenetic clock overlapped with existing clocks, whereas

the overlap reached 16% at the gene-level assessment (appendix 2 p 12).

We collected 144 blood samples from 94 centenarians and supercentenarians through single-timepoint or multiple-timepoint blood samplings and yielded DNA methylation profiles for 779 753 CpG sites (100% call rate). In the cladograms based on these DNA methylation profiles and inferred genotypes, specimens from the same individuals formed exclusive clusters (appendix 2 p 23).

Using the epigenetic clock trained with data from 336 participants (183 [54%] women and 153 [46%] men), we calculated centenarian and supercentenarian epigenetic ages (figure 2A). Except for a single centenarian man, all participants had younger epigenetic ages than their chronological ages, with negative epigenetic-age acceleration (figure 2B).

The degree and direction of the longitudinal change in epigenetic age varied across the 26 multiple-sampled individuals, with epigenetic age increasing with chronological age in 21 (81%) of these (appendix 2 p 24). For these 26 participants, the mean annual change in epigenetic age was 0·703 years (SD 0·839), which was significantly different from zero (*t* test p=0·0003). For the non-centenarians, the mean annual change in epigenetic age was 0·851 years (SD 0·001; appendix 2 p 24). Epigenetic ageing estimated from centenarian and supercentenarian longitudinal data was non-significantly slower (p=0·38, two-sample *t* test) than that inferred from non-centenarian cross-sectional data.

Sex-specific EWAS identified 3731 age-associated CpG sites for women and 7525 for men; for women, 1217 [33%] were positively and 2514 [67%] negatively associated and for men, 1071 [14%] were positively and 6454 [86%] were negatively associated (figure 3A–D). CpG sites with lower p values in one sex mostly had lower p values in the other sex (figure 3E), and CpG sites that were significantly associated in both sexes had the same coefficient direction and similar coefficients (figure 3F). Of 2109 (408 [19%] positively and 1701 [81%] negatively) age-associated CpG sites in both sexes, all but three (<1%) CpG sites were reproduced by non-sex-specific EWAS adjusted for sex and cell composition (appendix 2 p 13), confirming that these age-CpG relationships are not affected by cell composition or sex. The reproducibility of the known age-associated CpG sites markedly increased for the combined method of single linear regression and ANOVA (appendix 2 p 25). These CpG sites also showed constant methylation or demethylation along with age, and we found no significant deviation from the trend at any specific age group (appendix 2 p 26). In the downstream analyses, we focused on these 2109 CpG sites.

In centenarians and supercentenarians, 295 (72%) of 408 positively age-associated CpG sites had DNA-methylation profiles that increased as expected as per those of non-centenarians (ie, not significantly different from expected, based on chronological age). The

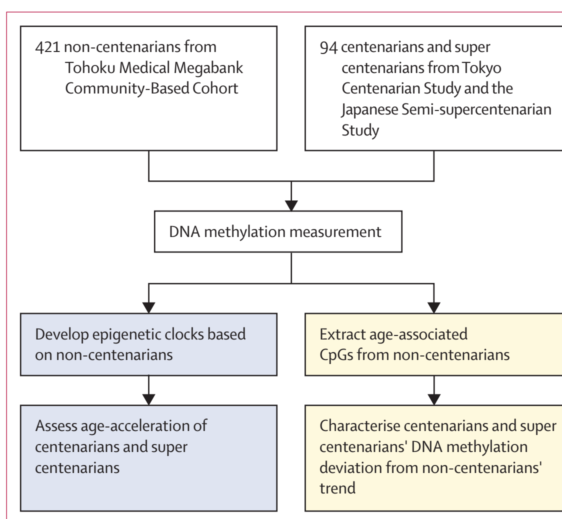

Figure 1: **Study flowchart**
The CpG-level approach is shown in the appendix (p 18).

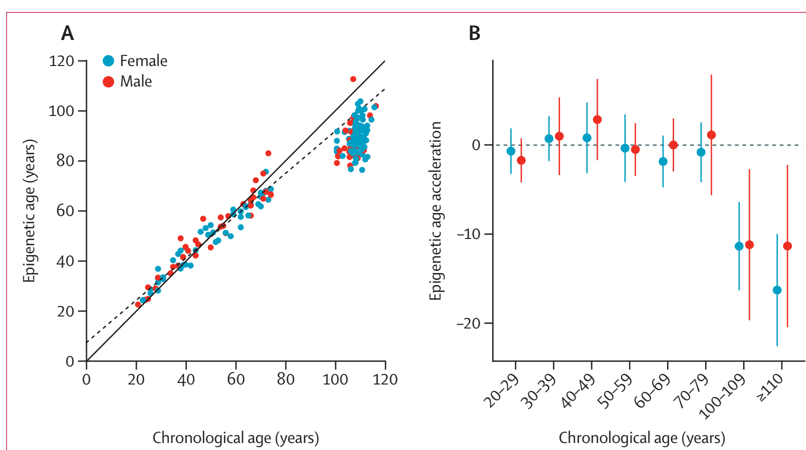

Figure 2: **Epigenetic ages of non-centenarians and centenarians and supercentenarians**
Non-centenarians were aged 20–78 years; centenarians and supercentenarians were aged 101–115 years. (A) Relationship between chronological and epigenetic age. The solid line indicates y=x. The dashed line is the regression line of epigenetic age on chronological age for the non-centenarians. (B) Mean cell-type adjusted epigenetic age acceleration in each age group. Error bars represent SDs. The horizontal dashed line represents an epigenetic age acceleration of zero.

remaining 113 (28%) CpG sites had lower DNA methylation profiles than expected (posHypo group; appendix 2 p 27). By contrast, for 1094 (64%) of 1701 negatively age-associated CpG sites, centenarian and supercentenarian DNA methylation decreased with chronological age, as expected (appendix 2 p 27). Of the remaining CpG sites, centenarian supercentenarian DNA methylation profiles were higher than expected for for 444 (26%; negHyper), and lower than expected for 163 (10%; negHypo; appendix 2 p 27). Thus, for the centenarians and supercentenarians, the 113 posHypo and 444 negHyper CpG sites could be characterised as having a younger DNA-methylation status, whereas the 163 negHypo CpG sites could be characterised as having

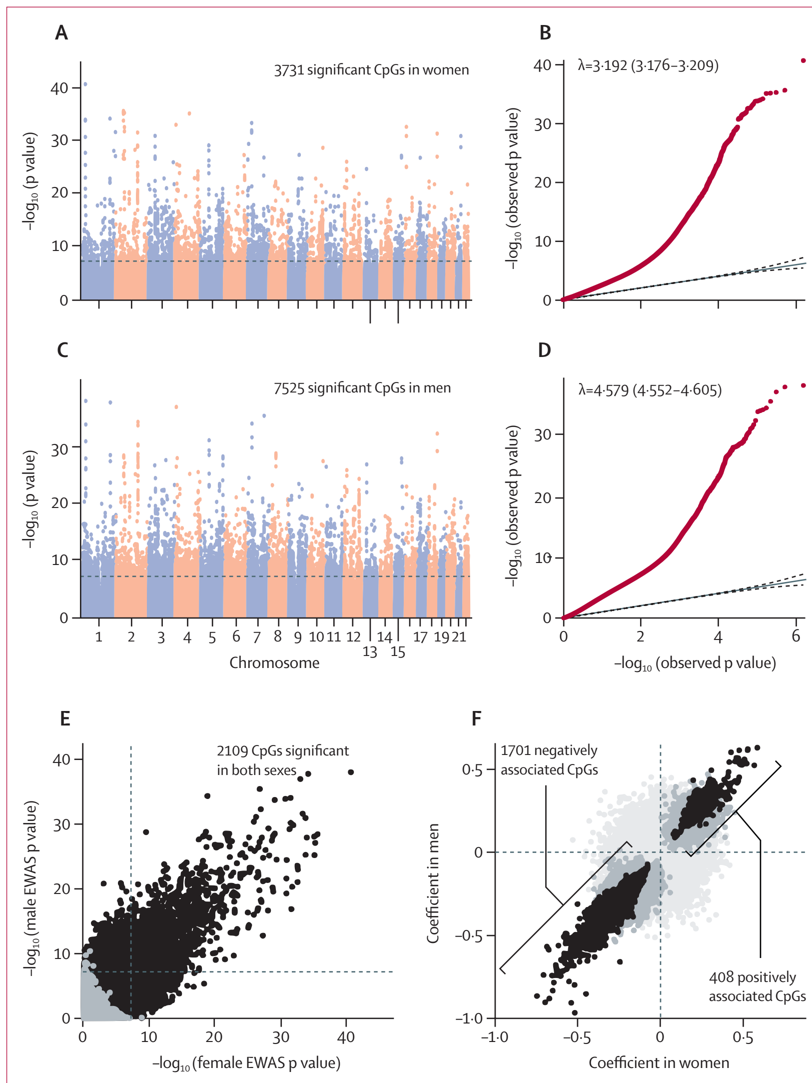

Figure 3: **Age-based EWAS by sex**

(A) Manhattan plot for female non-centenarians (aged 20–78 years). Dashed line indicates Bonferroni-corrected significance threshold. (B) EWAS quantile–quantile plots, showing inflation factor with 95% CIs. (C) Manhattan plot for male non-centenarians. The dashed line indicates Bonferroni-corrected significance threshold. (D) EWAS quantile–quantile plots for male non-centenarians. (E) Comparison of p values by sex. CpG sites with consistent coefficient directions between women and men are shown in black; those with different directions are shown in grey. The dashed lines indicate Bonferroni-corrected significant thresholds. (F) Comparison of EWAS coefficients. CpG sites significantly associated in both sexes are shown in black; those significant in one sex are in dark grey; those significant in neither sex are in light grey. EWAS=epigenome-wide association study.

an advanced (or older) DNA demethylation status. Among the identified posHypo, negHyper, and negHypo CpG sites, less than 10% are covered by the Infinium HumanMethylation450K and MethylationEPIC microarrays; although, from these few CpG sites, previously reported age-related CpG sites listed in the EWAS Atlas were substantially overrepresented (appendix 2 p 28). At the gene-level assessment, several epigenetic clock CpG sites overlapped with posHypo, negHyper, and negHypo CpG sites, and were underrepresented in the negHypo category (appendix 2 p 14).

CpG annotation identified different contexts for positively and negatively age-associated CpG sites: approximately 60% of positively age-associated CpG sites (posHypo) were in CpG islands, whereas 70% of negatively associated CpG sites (negHyper and negHypo) were in open-sea regions (appendix 2 p 27). Genic annotation proportions were also different between the two CpG groups: positively associated CpG sites were more frequently located in exons, and less frequently in intergenic regions, compared with the negatively associated CpG sites.

Enrichment analyses showed that posHypo CpG sites were enriched in cell-adhesion-related groups (appendix 2 p 15); no other enrichment of terms or pathways was observed. Oncogenes were overrepresented among genes with neighbouring negHyper CpG sites (appendix 2 p 28). Conversely, the odds ratios of CpG sites annotated to oncogenes were lower for the posHypo CpG sites (appendix 2 p 28). The odds ratios for the associations between CpG sites and immune-related genes varied depending on the database that we used: using the Calvano et al database generated different odds ratios than did using the others (Immunome and IRIS; appendix 2 p 28). Our meta-analysis showed that immune-related genes were overrepresented for negatively age-associated CpG sites that were further demethylated in centenarians and supercentenarians (the negHypo category; appendix 2 p 28).

Among posHypo, negHyper, and negHypo CpG sites, 6–20% were significantly associated with the neighbouring single-nucleotide polymorphism genotypes (methylation quantitative trait loci CpG sites); the negatively age-associated CpG sites (negHyper and negHypo) were more likely to be affected by genotypes compared with posHypo (appendix 2 p 16). Among these methylation quantitative trait loci, eight negHyper CpG-associated single-nucleotide polymorphisms were reported as genome-wide association study signals (appendix 2 p 17).

In the protein–protein interaction network analysis for negHyper genes, six networks were formed (appendix 2 p 27). The largest network consisted of 51 proteins, with the CD44 antigen having the greatest centrality. This hub protein participates in various functions, of which a cancer-related function is widely shared across neighbouring proteins in the network (Human Protein Atlas database[20]). The hub proteins on the other side of the network (eg, AUTS2, CNTNAP2, ITPR1, and NRXN1), together with their neighbouring proteins, are all implicated in many neuropsychiatric disorders, such as attention-deficit hyperactivity disorder, ataxia, autism spectrum disorder, bipolar disorders, intellectual disabilities, and schizophrenia.[21–24] For the posHypo genes, we found an interaction for only one pair of proteins belonging to the protocadherin beta cluster, which functions in neural cell adhesion (appendix 2 p 27).

Protein–protein interaction analysis of genes with enhanced demethylation in centenarians and

supercentenarians (the negHypo category) showed a network of three proteins, all linked to TGF-β signalling: SMAD7 inhibits TGF-β signalling; ACVR1 activates SMAD7; and TAB2 activates TAK1, which is released from the receptor complex by TGF-β stimulation, and transmits TGF-β signals downstream (appendix 2 p 27). SMAD7 is essential for TGF-β anti-inflammatory activity; by interacting with another protein, SMAD7 can trigger suppression of pro-inflammatory gene expression.[25] Screening of the iMETHYL expression quantitative trait methylation database showed that the two negHypo CpG sites annotated to *SMAD7* are significantly negatively associated with *SMAD7* expression in CD4+ T cells (chr18: 46451196, coefficient –0·0103, p=2·12×10⁻⁶; and 46451216, coefficient–0·0063, p=6·10×10⁻⁶). Therefore, the reduced DNA methylation of these CpG sites in centenarians and supercentenarians is associated with *SMAD7* upregulation.

Discussion

Our findings show that the epigenetic ages of centenarians and supercentenarians were remarkably lower than their chronological ages, consistently with previous findings for Italian semi-supercentenarians,[8] suggesting that their healthy longevity has an epigenetic basis. However, whether these epigenetic ages also reflect biological age has not yet been validated. Whether healthy longevity depends on slowing epigenetic ageing or on having a younger baseline DNA methylation state would be the next subject of interest. For our multiple-sampled centenarians and supercentenarians, the longitudinal changes in epigenetic age showed that their epigenetic ageing was slower than that indirectly inferred from the cross-sectional non-centenarian cohort. Further research comparing the longitudinal epigenetic change of centenarians and supercentenarians with non-centenarians will help to answer this question.

Our study further suggests a link between the specific epigenetic states and exceptional healthy longevity in centenarians and supercentenarians. Some epigenetic signatures in centenarians and supercentenarians were maintained at young states, whereas others were maintained at advanced (or old) states. Young-state DNA-methylation signatures were overrepresented around cancer-related and neuropsychiatric disease-related genes. This finding might provide an epigenetic basis for healthy longevity, associated with avoidance or postponement of age-related disease onset and cognitive decline. Such avoidance of disease onset and cognitive decline is a hallmark of healthy longevity, and our findings identifying specific young epigenetic regions are compelling.

Conversely, CpG sites with accelerated (advanced) demethylation were also detected in centenarians and supercentenarians. Knowledge-based analyses indicated that some of these demethylated CpG sites can affect the activity of TGF-β, a major anti-inflammatory cytokine. Given that many age-related diseases can develop as a consequence of excessive pro-inflammatory responses, anti-inflammatory responses, such as those mediated by TGF-β and other cytokines, are crucial for healthy ageing and longevity.[26] For instance, immunoassays have identified greater TGF-β activity in centenarians than in younger controls.[27] Serum tests on semi-supercentenarians and their offspring identified suppression of chronic inflammation as a potential driver of successful ageing.[28] Although the genomic variation influencing TGF-β plasma concentrations is unknown, our speculation that accelerated demethylation increases TGF-β activity in centenarians and supercentenarians highlights the importance of epigenetic regulation of TGF-β activity in healthy longevity.

This study detected epigenetic signatures that might underlie the hallmarks of exceptional healthy longevity, showcasing the importance of not just younger epigenetic states but also that of advanced states for healthy ageing and longevity, and such advanced states are less likely to be accounted for in epigenetic clocks. However, this study is limited by the fact that we did not include the epigenome of people in aged 79–100 years, who were not available in our cohort. Therefore, the question remains of whether the epigenetic characteristics observed in centenarians and supercentenarians are also manifested in non-centenarian older people (ie, those aged between 80 and 100 years), or if these characteristics are shared only by people who reach exceptionally advanced ages.

Contributors

YK and AS conceived, designed, and directed the study. SK, MN, RO, and AY participated in designing the study. YAr and NH led the recruitment of centenarians and supercentenarians and managed the samples. EA, RO, KO, YAb, HOh, and SU did the laboratorial work. SK, MN, KO, YAb, and NOS analysed the data. SK and AS wrote the original draft. SK, MN, KO, HOh, TH, YS, YO-Y, and AS had access to the data. SK, MN, and AS verified the data. HOk and MS provided the cohort data and commented critically on the study. All authors revised the manuscript and had final responsibility for the decision to submit for publication.

Declaration of interests

SK reports a grant from the Japan Society for the Promotion of Science (JSPS), unrelated to the current work. MN reports salary support from KDDI Corporation, the parent company of KDDI Research. EA reports a grant from JSPS, unrelated to the current work. RO reports a grant from JSPS unrelated to the current work. HOh reports grants from JSPS and the Japan Agency for Medical Research and Development (AMED), unrelated to the current work. SU reports a grant from JSPS, unrelated to the current work. TH reports participation in the board of Genome Analytics Japan, unrelated to of the current work. YS reports a grant from JSPS, unrelated to the current work. YO-Y reports a grant from JSPS, unrelated to the current work. YAr reports grants from JSPS and AMED, unrelated to the current work. AY reports salary support from KDDI. HOk reports involvement as a scientific advisor of SanBio and K Pharma and a grant from AMED, unrelated to of the current work. MS reports grants from JSPS, unrelated to the current work, and from AMED for the current work. YK reports grants from JSPS and AMED, unrelated to the current work, and from Keio University for the current work. AS reports grants from JSPS, AMED, National Cancer Center Japan, IQVIA, and Fujifilm, unrelated to the current work. All other authors declare no competing interests.

Data sharing

The epigenetic clock model and results of EWAS and methylation quantitative trait loci analysis will be placed in a public repository upon acceptance of this manuscript (http://dx.doi.org/10.17632/cg475bbzmf.1).

To protect participant privacy, the individual-level DNA methylation and phenotype data cannot be made publicly available. Please contact the corresponding author for the related documents (eg, study protocol and informed consent form).

Acknowledgments
We thank the participants of this study, including those who provided specimens and data, and all who participated in applications to conduct this study, recruitment, sample management, and data cleaning. This study was supported by the Tohoku Medical Megabank Project (Special Account for the Reconstruction from the Great East Japan Earthquake, AMED under grant number JP15km0105004), KDDI Research, and the Next Generation Research Project, Keio University.

References

1 Scott AJ. The longevity society. *Lancet Healthy Longev* 2021; **2:** e820–27.
2 Takayama M, Hirose N, Arai Y, et al. Morbidity of Tokyo-area centenarians and its relationship to functional status. *J Gerontol A Biol Sci Med Sci* 2007; **62:** 774–82.
3 Arai Y, Inagaki H, Takayama M, et al. Physical independence and mortality at the extreme limit of life span: supercentenarians study in Japan. *J Gerontol A Biol Sci Med Sci* 2014; **69:** 486–94.
4 Schoenhofen EA, Wyszynski DF, Andersen S, et al. Characteristics of 32 supercentenarians. *J Am Geriatr Soc* 2006; **54:** 1237–40.
5 Beker N, Ganz A, Hulsman M, et al. Association of cognitive function trajectories in centenarians with postmortem neuropathology, physical health, and other risk factors for cognitive decline. *JAMA Netw Open* 2021; **4:** e2031654.
6 Liu Z, Zhu Y. Epigenetic clock: a promising mirror of ageing. *Lancet Healthy Longev* 2021; **2:** e304–05.
7 Chiavellini P, Canatelli-Mallat M, Lehmann M, et al. Aging and rejuvenation—a modular epigenome model. *Aging (Albany NY)* 2021; **13:** 4734–46.
8 Horvath S, Pirazzini C, Bacalini MG, et al. Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging* 2015; **7:** 1159–70.
9 Gentilini D, Mari D, Castaldi D, et al. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age* 2013; **35:** 1961–73.
10 Hozawa A, Tanno K, Nakaya N, et al. Study profile of the Tohoku Medical Megabank Community-Based Cohort Study. *J Epidemiol* 2021; **31:** 65–76.
11 Hirata T, Arai Y, Yuasa S, et al. Associations of cardiovascular biomarkers and plasma albumin with exceptional survival to the highest ages. *Nat Commun* 2020; **11:** 3820.
12 Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013; **14:** R115.
13 Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018; **19:** 371–84.
14 Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics* 2017; **33:** 2381–83.
15 Carbon S, Douglass E, Good BM, et al. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* 2021; **49:** D325–34.
16 Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000; **28:** 27–30.
17 Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* 2016; **32:** 286–88.
18 Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019; **47:** D607–13.
19 Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Syst* 2006; **1695:** 1–9.
20 Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science* 2015; **347**.
21 Tromp A, Mowry B, Giacomotto J. Neurexins in autism and schizophrenia–a review of patient mutations, mouse models and potential future directions. *Mol Psychiatry* 2021; **26:** 747–60.
22 Pang W, Yi X, Li L, Liu L, Xiang W, Xiao L. Untangle the multi-facet functions of AUTS2 as an entry point to understand neurodevelopmental disorders. *Front Psychiatry* 2021; **12:** 580433.
23 Rodenas-Cuadrado P, Ho J, Vernes SC. Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet* 2014; **22:** 171–78.
24 Synofzik M, Helbig KL, Harmuth F, et al. De novo *ITPR1* variants are a recurrent cause of early-onset ataxia, acting via loss of channel function. *Eur J Hum Genet* 2018; **26:** 1623–34.
25 Lee YS, Kim JH, Kim ST, et al. Smad7 and Smad6 bind to discrete regions of Pellino-1 via their MH2 domains to mediate TGF-β1-induced negative regulation of IL-1R/TLR signaling. *Biochem Biophys Res Commun* 2010; **393:** 836–43.
26 Santoro A, Bientinesi E, Monti D. Immunosenescence and inflammaging in the aging process: age-related diseases or longevity? *Ageing Res Rev* 2021; **71:** 101422.
27 Carrieri G, Marzi E, Olivieri F, et al. The G/C915 polymorphism of transforming growth factor beta1 is associated with human longevity: a study in Italian centenarians. *Aging Cell* 2004; **3:** 443–48.
28 Arai Y, Martin-Ruiz CM, Takayama M, et al. Inflammation, but not telomere length, predicts successful ageing at extreme old age: a longitudinal study of semi-supercentenarians. *EBioMedicine* 2015; **2:** 1549–58.