CNS Spectrums

www.cambridge.org/cns

Genome-wide DNA methylation profiling in nonagenarians suggests an effect of *PM20D1* in late onset Alzheimer's disease

Carolina Coto-Vílchez[1], José J. Martínez-Magaña[2], Lara Mora-Villalobos[1], Daniel Valerio[4], Alma D. Genis-Mendoza[2], Jeremy M. Silverman[5], Humberto Nicolini[2], Henriette Raventós[1,3] and Gabriela Chavarria-Soley[1,3]*

[1]Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, San José, Costa Rica, [2]Instituto Nacional de Medicina Genómica, Mexico City, México, [3]Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica, [4]Hospital Nacional de Geriatría y Gerontología de Costa Rica, San José, Costa Rica, and [5]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Original Research

Cite this article: Coto-Vílchez C, Martínez-Magaña JJ, Mora-Villalobos L, Valerio D, Genis-Mendoza AD, Silverman JM, Nicolini H, Raventós H, and Chavarria-Soley G (2023). Genome-wide DNA methylation profiling in nonagenarians suggests an effect of *PM20D1* in late onset Alzheimer's disease. *CNS Spectrums* **28**(2), 174–182.
https://doi.org/10.1017/S109285292100105X

Received: 10 August 2021
Accepted: 01 December 2021

Key words:
Late-onset Alzheimer's disease; epigenome; *PM20D1*; quantitative-trait loci

Author for correspondence:
*Gabriela Chavarria-Soley,
Email: gabriela.chavarriasoley@ucr.ac.cr

Abstract

Background. The aim of this study is to identify differentially methylated regions (DMRs) in the genomes of a sample of cognitively healthy individuals and a sample of individuals with LOAD, all of them nonagenarians from Costa Rica.

Methods. In this study, we compared whole blood DNA methylation profiles of 32 individuals: 21 cognitively healthy and 11 with LOAD, using the Infinium MethylationEPIC BeadChip. First, we calculated the epigenetic age of the participants based on Horvath's epigenetic clock. DMRcate and Bumphunter were used to identify DMRs. After in silico and knowledge-based filtering of the DMRs, we performed a methylation quantitative loci (mQTL) analysis (rs708727 and rs960603).

Results. On average, the epigenetic age was 73 years in both groups, which represents a difference of over 20 years between epigenetic and chronological age in both affected and unaffected individuals. Methylation analysis revealed 11 DMRs between groups, which contain six genes and two pseudogenes. These genes are involved in cell cycle regulation, embryogenesis, synthesis of ceramides, and migration of interneurons to the cerebral cortex. One of the six genes is *PM20D1*, for which altered expression has been reported in LOAD. After genotyping previously reported mQTL SNPs for the gene, we found that average methylation in the *PM20D1* DMR differs between genotypes for rs708727, but not for rs960603.

Conclusions. This work supports the possible role of *PM20D1* in protection against AD, by showing differential methylation in blood of affected and unaffected nonagenarians. Our results also support the influence of genetic factors on *PM20D1* methylation levels.

Introduction

Late-onset Alzheimer's disease (LOAD) is the most common cause of dementia in the elderly. It is a complex disorder that results from a combination of genetic and nongenetic risk factors, where the environment plays an important role in its development.[1] Age is the main risk factor for developing LOAD: the disease is present in 1% to 3% of those between ages 60 and 70, 3% to 12% of those between ages 70 and 80, and 25% to 35% of those older than age 85.[2,3] Some risk factors for LOAD are modifiable, that is, they are relevant for prevention. Factors such as smoking, years of education, cognitive stimulation, exercise, and diet have been found to play a role in the disease.[4,5]

Genetic variants in multiple genes have been identified that increase the risk for LOAD.[6,7] Furthermore, there are also genetic variants that may be protective against the development of dementia at advanced ages.[8] The level of protection in each individual depends on the effect of multiple factors such as aging, genetics (protective and risk variants), lifestyle, cardiovascular health, and others. It has been proposed that individuals with a positive balance of protective factors can reach advanced ages without developing dementia,[9] which we will refer to as successful cognitive aging (SCA).

It has been observed that one gene can carry both risk and protective variants for late-onset dementia. This is the case for apolipoprotein E (*APOE*)[10]; the *APOE-ε4* allele increases the risk of LOAD, and the *APOE-ε2* allele protects against it.[11,12] However, some studies have found that the association between the *APOE-ε4* allele and dementia is not present in individuals over 80 years of age.[10,13] These findings could support a survivor effect model for SCA in old age. The protected survivor model proposes that a minority of the general population has a protective factor that mitigates the negative effect of a risk factor on successful cognitive aging for the unprotected

majority. As age increases, the proportion of survivors with protection increases. Therefore, although the association of the risk factor with survival does not change within an individual, the association in the surviving population changes as its age increases.[14]

In the last years, epigenetic modifications of the genome have gained attention in complex diseases such as LOAD, and the identification of epigenetically dysregulated genes has been increasing.[15,16] Some genes, such as ankyrin 1 (*ANK1*), sorbin and SH3-domain-containing 3 (*SORBS3*), and histone deacetylase 2 (*HDAC2*) genes have been reported as dysregulated by independent studies in humans.[15] Epigenetic dysregulation of enhancers in neurons in AD has also been described, and the authors propose a hypothesis where hypomethylation of enhancers could induce the formation and progression of amyloid-beta plaques and neurofibrillary tangle pathology through *BACE1* activation.[17] Another recent whole-genome methylation analysis proposed a role for the immune system and polycomb complex involvement in AD.[16]

The validity of using peripheral blood for epigenetic studies of disorders of the brain (or other tissues) has often been questioned. This is unavoidable in most studies, because of the challenge or impossibility of obtaining brain tissue for analyses. Recent research has shown that, in general, there is a robust correlation in methylation levels between blood and the brain, although this can vary greatly for different genomic regions.[18-21] It has been suggested that interpretation of blood methylation results for brain disorders should focus on CpG sites with a high correlation in DNA methylation across both tissue types, and some tools have been developed for this purpose.[19,20]

While previous work has shown that some risk factors for LOAD could have an age-dependent effect,[10,13] most of the analyses have been performed on individuals under 80 years. Above those ages, the effect of risk factors for LOAD remained underexplored. Similarly, most epigenetic and nonepigenetics studies of SCA have focused on individuals between 65 and 85 years old, and representation of nonagenarians or centenarians has been scarce.[22] It has been proposed that cognitively healthy individuals aged 90 years and above are an optimal population to study genetic protective factors for LOAD.[23] Their first-degree relatives have been observed to maintain intact cognitive function more frequently than relatives of younger nondemented elderly. Also, a high heritability of cognitive functions such as memory has been identified in nonagenarians.[23,24]

Costa Rica is the second country in the American continent with the greatest longevity after Canada.[25] At the age of 90, Costa Rican nonagenarians have a life expectancy of 4.4 years, which is half a year more than any other country in the world.[26] Access to an almost universal health system, lifestyle factors, and family support have been proposed to influence the observed longevity in the country.[27] This study aims to identify differentially methylated regions (DMRs) in the genomes of a sample of cognitively healthy individuals and a sample of individuals with LOAD, all of them over 90 years of age, from the Central Valley of Costa Rica (CVCR).

Methods

Study population

The subjects were recruited in the study "Successful Cognitive Aging and Cardiovascular Risk Factors in the Central Valley of Costa Rica," funded by an NIH Fogarty International Center & National Institute on Aging grant (R21TW009258), and a P01-AG02219 grant funded by the Alzheimer's Association. The project and consent forms were reviewed and approved by the Institutional Review Board of the University of Costa Rica and the Mount Sinai Medical School. The study was explained to each subject and written informed consent was obtained. If the subject was unable to give consent because of cognitive impairment, consent was obtained from the spouse or the primary caretaker relative. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

The sample consisted of 32 nonagenarians, of which 11 had a diagnosis of probable LOAD. For this study, we only included female probands over 90 years old, with between 0 and 9 years of schooling, at least one pregnancy, married or widowed, menopause between 50 and 55 years of age, and without hormone replacement therapy with estrogen.

All subjects were clinically assessed by a medical geriatrician (DV) and a psychologist (LM-V) with a general medical examination, the Clinical Dementia Rating scale (CDR), and the Mini-Mental State Examination (MMSE).[28,29] In the CDR, clinical information is collected from both an informant and the subject. A CDR score of zero indicates no dementia or a recent decline in cognition or functioning, and a score of 3 indicates severe dementia. In the MMSE, a score between 27 and 30 indicates the absence of cognitive impairment, and values below 6 indicate severe dementia. Diagnosis of probable LOAD was defined by the geriatrician based on the clinical history and the CDR and MMSE scores. Patients with a history of stroke or the presence of a disorder other than AD that potentially causes dementia were excluded. In the sample of individuals with probable LOAD we decided to include subjects with a CDR score of 2 (moderate; N = 3) or 3 (severe; N = 8). Individuals with mild cognitive impairment (with CDR scores above zero, but less than 2) were not included in our sample.

Genome-wide methylation analysis

Microarray analysis

Whole blood was collected by peripheral venipuncture from participants and DNA was extracted using the sucrose method.[30] DNA was bisulfite converted with the Zymo Research EZ DNA Methylation Kit (Irvine, CA). We obtained genome-wide methylation profiles using the Infinium HumanMethylationEPIC BeadChip (Illumina, Inc., San Diego, CA). The methylation data were converted to *idat* files in the GenomeStudio software (Illumina, Inc., San Diego, CA).

Idat files were processed with the *minfi* R package (https://www.bioconductor.org/packages/devel/bioc/vignettes/minfi/inst/doc/minfi.html).[31] Quality control was first performed to detect samples that failed to adequately detect DNAm (in our case, all samples passed quality control). Then, we removed probes that could cross-hybridize or overlap with SNPs, which could confound results. Probes with a detection $p > .05$ and probes that failed in more than 50% of the samples were removed. The total number of probes post-filtering was 794,770. We used FunNorm normalization to remove unwanted variation by regressing out variability explained by the control probes present on the array.[32] The proportion of DNAm at a particular CpG site (β values) was ascertained by taking the ratio of the methylated (C) to unmethylated (T) signal, using the formula: β = intensity of the methylated signal/(intensity of the unmethylated signal + intensity of the methylated

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

signal + 100). β values range from 0 (completely unmethylated) to 1 (completely methylated).[33]

Data availability

The data presented in this study are available upon request from the corresponding author; upload to a public repository was not done due to privacy and ethical issues.

Prediction of epigenetic age

The Horvath method in the R package was used to determine DNAm-based age prediction. This method uses a weighted average of DNA methylation at 353 clock CpG sites, which is then transformed to DNAm age using a calibration function (http://dnamage.genetics.ucla.edu).[34] Mean differences between groups were assessed using the Kruskal–Wallis test.

Estimation of cell type proportions

We estimated the cell type proportions in our two groups because whole blood is made up of many different cell types, each with different methylation profiles that can vary in proportion with disease status. These cell types include CD8+ T lymphocytes, CD4+ T lymphocytes, natural killer cells, B cells, monocytes, and granulocytes. For this purpose, we applied Houseman's algorithm using the *minfi* R package.[35] A Mann–Whitney *U* test was used to evaluate the differences between groups for each cell type category.

Identification of differentially methylated probes and DMRs

We applied two different methods to identify differentially methylated probes (DMPs) and DMRs between nonagenarians with and without LOAD: DMRcate and ChAMP-Bumphunter.[36] We assigned a *p*-value cutoff of .05, after false discovery rate (FDR) correction, to determine DMPs and DMRs in both DMRcate and Bumphunter. We also used the ChAMP-Bumphunter R package to do the gene set enrichment analysis based on our DMRs.

Once the DMRs were identified, we used the web application BECon to verify the correlation between the methylation in blood and the methylation in the brain for each one of the CpG sites of the DMRs.[19] This tool uses a DNA methylation database from paired samples of blood and three postmortem brain regions from individuals to show how informative DNA methylation from the blood is for brain DNA methylation.

Gene-level analysis

After reviewing existing literature on the genes included in the DMRs, we decided to focus on the *PM20D1* gene, for which altered expression in AD has previously been reported.[37] Specifically, two SNPs, rs708727 and rs960603 have been reported as expression and methylation QTLs for the gene.[37,38] Sanger sequencing was performed to genotype both SNPs. Average methylation (β values) in the DMR between the genotypes for rs708727 was compared with a Mann–Whitney *U* test (because only two out of the three possible genotypes were observed). For rs960603 all three possible genotypes were observed, and a comparison of average methylation between them was done with a Kruskal-Wallis test. The same comparisons of methylation levels between genotypes for both SNPs were done for each CpG site in the DMR. A stratified comparison by genotype of the methylation level between SCA and LOAD was also performed, using Mann–Whitney *U* tests for each pairwise comparison. The AA genotype for rs960603 was not included in the analysis for statistical reasons (only one individual each for SCA and LOAD was available).

Results

Epigenetic age

We assayed DNAm profiles of 32 females, of which 21 were cognitively healthy and 11 had a probable LOAD diagnosis. The age range is between 90 and 103 years old. The average chronological age of the LOAD group was 95 (SD = 3.36), and the average epigenetic age was 73 (SD = 5.99). In the SCA group, the average chronological age was 93 (SD = 2.77), while the average epigenetic age was 73 (SD = 5.85). Neither the epigenetic age (p-value = .53) nor the difference between epigenetic age and biological age shows differences between groups (p-value = .4200). Nevertheless, when analyzing the sample as a whole, a significant difference was obtained between chronological and epigenetic age (p = .0045). The age acceleration was lower on individuals diagnosed with LOAD compared to those cognitively healthy, but not statistically significant (p = .1972).

Differentially methylated regions

No statistically significant differences were found in the cellular composition between the LOAD and SCA groups, in any of the cell types, so this variable was not included in the analysis (Table 1). We identified several differentially methylated regions in our analysis, but no statistically significant differences at the probe level (DMPs) were detected. After detection of DMRs with the DMRcate and Bumphunter methods, we chose the 11 DMRs that both methods found in common for further analysis (Table 2). Four of the 11 DMRs are hypomethylated in the SCA group compared with the LOAD group, while seven of them are hypermethylated. The mean length of the DMRs was 603 bp, with 197 bp in the shortest region and 1736 bp in the longest. On average, in the 11 DMRs observed, there were 11 CpG sites per region, with a range of 6 to 18 CpG sites. Six out of the 11 DMRs include known genes, and two include pseudogenes. From the eight DMRs associated with genes and pseudogenes, six were located in promoters. The results of the correlation in blood and brain methylation for each of the regions are also shown in Table 2. The DMRs, were enriched on 94 diverse pathways, but the top 3 pathways were: Pilon *KLF1* targets down-regulated (adj. p = 4.44e-09), Martens bound by PML-*RARA* fusion (adj. p = 1.26e-06), and Blalock Alzheimer's Disease up (adj p = 1.26e-06).

Table 1. Cell Type Composition Comparison Between the SCA and LOAD Groups.

Cell Type	Cognitively Healthy	LOAD	*p*-Value
CD8T	0.15	0.16	.63
CD4T	0.08	0.06	.18
Natural killers	0.08	0.06	.28
B cell	0.05	0.04	.19
Monocyte	0.08	0.08	.84
Neutrophil	0.6	0.64	.66

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

Table 2. Differentially Methylated Genomic Regions Between Individuals with SCA and LOAD.

Genomic Coordinates*	DMR Length (bp)	CpG #	Associated Gene	Genomic Region	Mean Δβ	Max Δβ	Methylation Status SCA vs LOAD	p Bump-Hunter	p DMR-CATE	Blood–Brain Correlation
chr1:205818956 - 205819609	653	12	*PM20D1*	Promoter	−0.12	−0.19	Hypomethylated	0.0006	6.6540×10^{-8}	0.6271
chr9:124988720 - 124990456	1736	13	*LHX6*	Intragenic, promoter	0.12	0.18	Hypermethylated	0.0003	2.5250×10^{-10}	−0.0044
chr15:101084980–101085177	197	6	*CERS3*	Promoter	0.10	0.15	Hypermethylated	0.0037	3.3371×10^{-7}	0.5683
chr5:135415693 - 135416029	336	6	*vtRNA2–1*	Intergenic	0.10	0.16	Hypermethylated	0.0004	0.0072	0.6993
chr16:2907819–2908245	426	8	*PRSS22*	Intragenic, promoter	0.04	0.06	Hypermethylated	0.0223	9.3571×10^{-5}	0.1105
chr6:32847548 - 32847845	297	13	*PPP1R2P1*	Intergenic	−0.07	−0.11	Hypomethylated	0.0032	7.4465×10^{-12}	0.0541
chr17:36997420 - 36997740	320	8	*C17orf98*	Promoter	−0.09	−0.15	Hypomethylated	0.0022	2.0186×10^{-20}	0.3842
chr11:67383377 - 67384040	663	8	*DOC2GP*	Promoter	0.10	0.14	Hypermethylated	0.0123	3.3508×10^{-5}	0.1444
chr6:29648225 - 29648756	531	18	–	Intergenic	0.10	0.21	Hypermethylated	0.0003	0.0350	0.7011
chr6:31148332 - 31148748	416	15	–	Intergenic	−0.10	−0.16	Hypomethylated	0.0007	4.5671×10^{-13}	0.1893
chr6:31275148 - 31276212	1064	17	–	Intergenic	0.05	0.13	Hypermethylated	0.0039	0.0020	0.3926

Note: Δβ: (Delta beta; absolute mean or max difference of β values between groups (SCA-LOAD).
*Genome coordinates from Human Genome GRCh37/hg19 Assembly.

After a review of existing literature, we decided to focus on the peptidase M20-domain-containing protein 1 (*PM20D1*) gene. An association of this gene with AD has previously been reported, as well as the existence of SNPs that act as expression and methylation QTLs for the gene.[37,38] The CpG sites included in this DMR are cg17178900, cg26354017, cg14159672, cg14893161, cg07533224, cg12898220, cg05841700, cg11965913, cg07167872, cg24503407, cg16334093, and cg07157834.

mQTL SNPs in PM20D1

We tested whether there is a relationship between the genotypes at mQTL and eQTL SNPs rs708727 and rs960603 and methylation level in the CpG regions of *PM20D1*. For rs708727, average methylation (β values) in the *PM20D1* DMR differs between genotypes, with higher average methylation in heterozygotes (AG) when compared to individuals homozygous for the G allele ($p = 1.4E\text{-}05$, Figure 1). For these SNPs, there were no homozygotes for the A allele in our sample. In the case of rs960603, no significant differences in average methylation levels were found between genotypes ($p = 0.31$, Figure 1). A comparison of methylation levels between genotypes for both SNPs at the 12 individual CpG sites in the DMR is presented in Table 3. The pattern is the same for the DMR as a whole: there is a significant difference in methylation between genotypes in all CpG sites for rs708727, and no difference in any of the sites for rs960603. No significant differences in methylation levels between the SCA and LOAD groups were found within each genotype for either rs708727 or rs960603 ($p > 0.05$ for all pairwise comparisons; Figure 2).

Discussion

Even though it was not the main goal of our study, the availability of whole-genome methylation information allowed us to calculate epigenetic age for the subjects in our sample. DNA methylation levels have been proposed as biomarkers of aging since chronological age correlates with DNA methylation in most human tissues and cell types.[34,39] In addition, this measure has predictive value for all-cause mortality, and estimated from blood-extracted DNA, correlates with measures of cognitive and physical fitness in 70 year-olds.[40,41] Although we found a 20-year or greater difference between chronological and epigenetic age in all our samples, no difference was observed between the SCA and AD groups using Horvath's DNAm age predictors. This is in contrast with the results in brain tissue of Levine et al,[42] who found that postmortem DNAm age in the dorsolateral prefrontal cortex was associated with neuropathological variables and postmortem measures of cognitive decline among individuals with AD.

Nevertheless, our results are valuable in the context of studies dealing with long-lived individuals. As in other studies of such individuals, epigenetic age was lower than chronological age; however, the extent of this difference was markedly larger. McEwen et al studied DNA methylation in a sample from Nicoya,[43] a Costa Rican high longevity region. No difference was found in epigenetic age between Nicoyans and a sample of people of other regions of Costa Rica. However, they reported a difference of −6.9 years between epigenetic age and chronological age when analyzing the whole sample and −12.7 years in centenarians. These results are consistent with the report of an epigenetically younger age in Hispanic populations.[44] Similar results have also been found in other

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

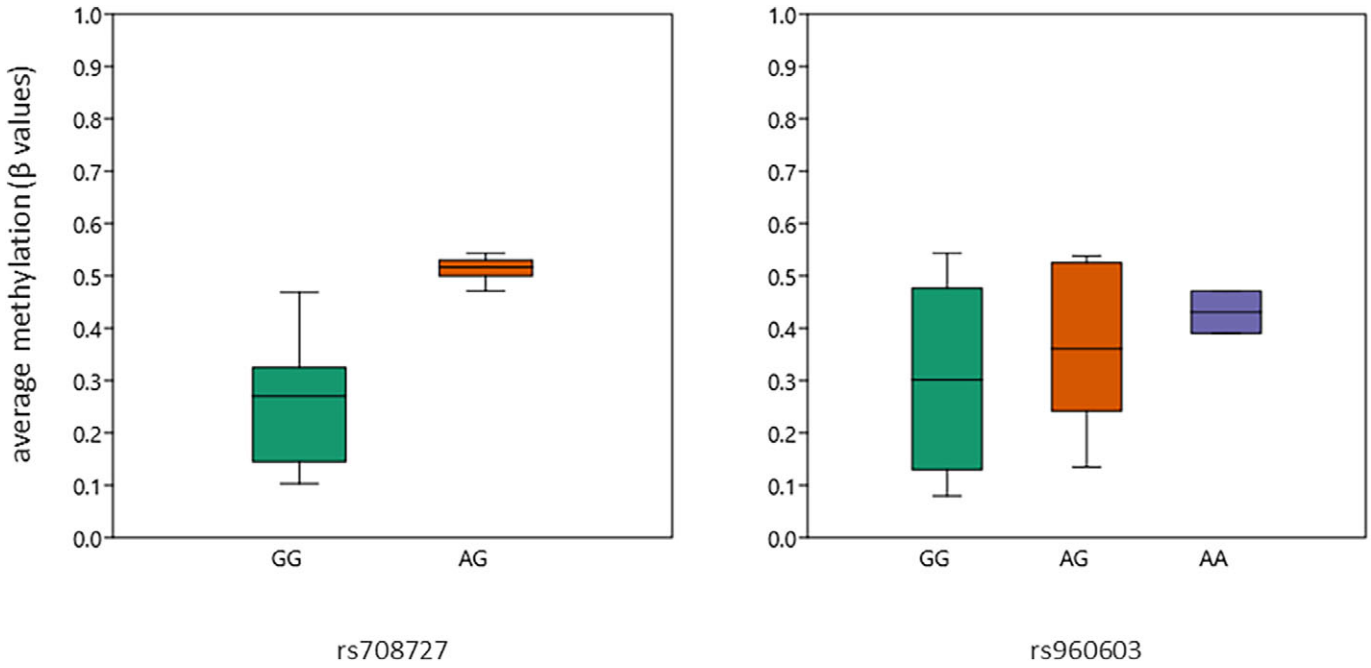

Figure 1. Comparison of average methylation (β values) between genotypes for rs708727 and rs960603.

Table 3. Comparison of the Average Methylation Level (β Values) Between Genotypes for rs708727 and rs960603 for all CpG Sites of the *PM20D1* DMR.

CpG Island	rs708727	rs960603
cg17178900	0.0001	0.05
cg26354017	2.20E-05	0.4
cg14159672	2.70E-05	0.33
cg14893161	1.80E-05	0.36
cg07533224	1.80E-05	0.25
cg12898220	1.80E-05	0.34
cg05841700	3.30E-05	0.44
cg11965913	1.80E-05	0.29
cg07167872	1.40E-05	0.51
cg24503407	1.80E-05	0.4
cg16334093	1.80E-05	0.34
cg07157834	1.40E-05	0.19

Note: Values in the table are *p*-values for the Mann-Whitney *U* test for rs708727, and for the Kruskall-Wallis test for rs960603.

populations. Nonagenarians from Sydney, Australia showed a difference of −9.56 years between epigenetic and chronological age.[45] Furthermore, centenarians from an Italian cohort were 8.6 years younger than their chronological age, while their offspring have a lower epigenetic age than age-matched controls.[46] More research is required to understand the reasons for these differences.

When comparing whole-genome methylation between the SCA and AD groups, we found 11 differentially methylated regions. Some of them have been previously reported to be associated with AD or related to pathways associated with AD. The gene known as peptidase M20-domain-containing protein 1 (*PM20D1*) was found to be hypomethylated in the SCA sample. This gene has been associated with AD, and additional evidence shows a correlation between DNA methylation, RNA expression, and genetic background; it is both methylation and an expression QTL.[37,38,47,48] Recent evidence indicates that *PM20D1* expression might provide a potential cellular defense mechanism against AD.[37] *In vitro* assays showed an increased *PM20D1* expression in neuroblastoma cells treated with reactive oxygen species (ROS) and amyloid-β (Aβ), emulating an AD model. Meanwhile, an analysis of brain tissue in a mouse model of AD, which develops AD-related pathologies with age, such as amyloid plaques, astrogliosis, and learning deficits, showed increased *PM20D1* expression in the frontal cortex at symptomatic stages of the disease in comparison with presymptomatic stages and controls. In addition, manipulation of PM20D1 levels showed that overexpression of *PM20D1* reduced cell death and decreased Aß levels in *vivo* and *in vitro* assays. Additionally, a recent study detected association of a DMR (in peripheral blood) in the PM20D1 gene with both the transition between cognitive-normal to mild cognitive impairment, and the rate of cognitive decline in Alzheimer's disease.[49]

Sánchez-Mut et al. found a significant increase of *PM20D1* repression in AD when compared with nondemented individuals by analyzing DNA methylation and RNA expression. The repression occurs by a CCCTC-binding-factor-mediated chromatin loop that depends on an AD-associated haplotype. Genetic analysis of human brain cortex samples has shown an allele-dependent correlation between the haplotype of two mQTL associated SNPs, rs708727 and rs960603, and *PM20D1* promoter methylation. As expected, *PM20D1* expression was inversely correlated with the methylation level of its promoter.[37] In the present study, we also found a significant change in average methylation at the *PM20D1* DMR according to the genotype for rs708727, while for rs960603 there is no significant effect. Similarly, a strong effect of rs708727 allele dosage on methylation, but not for rs960603, was recently reported by Wang et al.[50] Another recent study by Sanchez-Mut et al[38] found a much stronger correlation with methylation level at *PM20D1* CpG sites for rs708727 than for

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

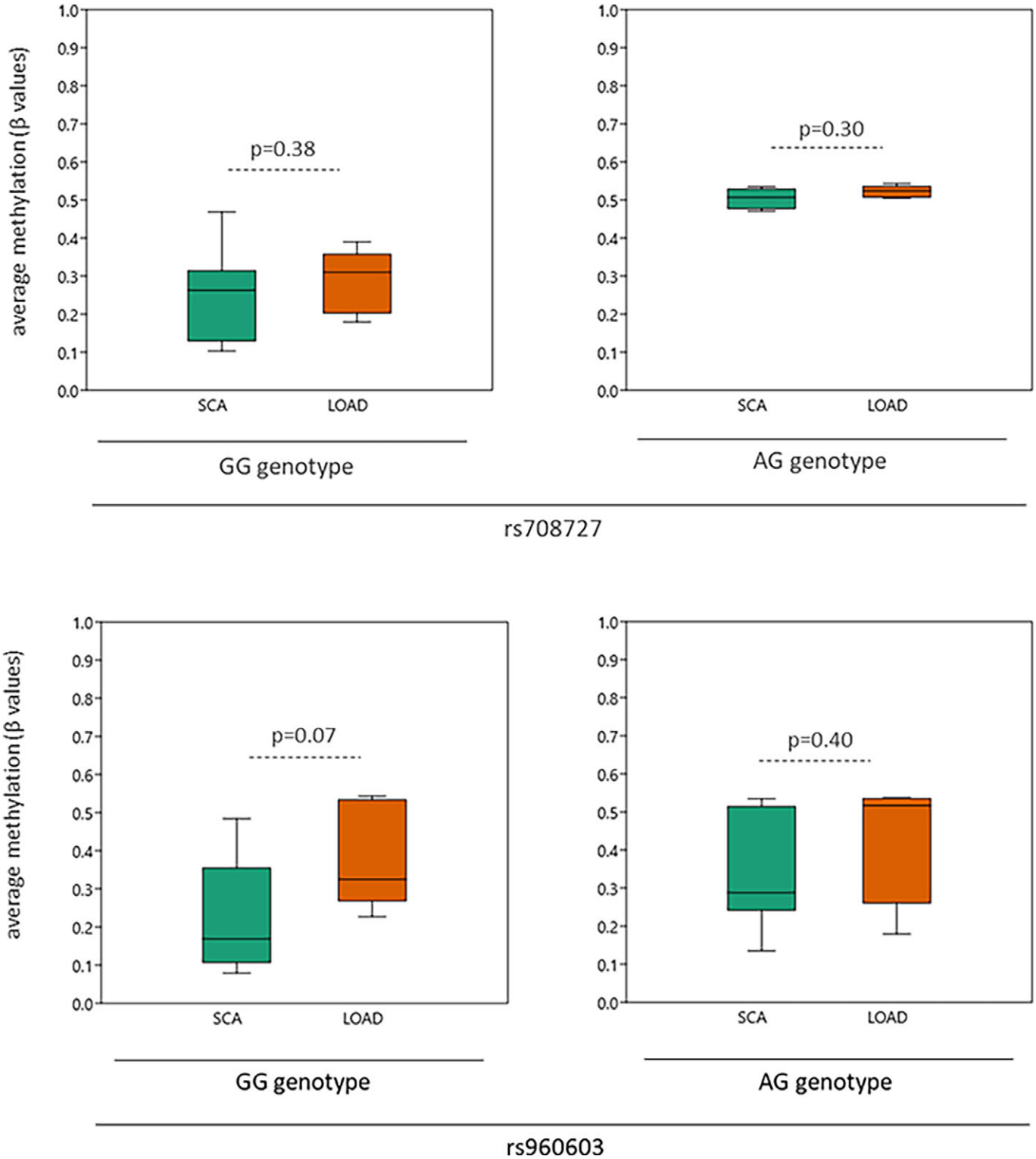

Figure 2. Stratified comparison per genotype of average methylation (β values) between SCA and LOAD for rs708727 and rs960603. The *p*-values for the Mann-Whitney *U* test are shown for each comparison.

rs960603. The minor allele frequencies for these SNPs in the central valley of Costa Rica are 0.41(A) for rs708727, and 0.38 (G) for rs960603 (unpublished data from G. Chavarria-Soley and H. Raventós, based on whole genome sequences of 51 unrelated individuals). In our study, when we performed a stratified comparison per genotype of methylation levels between SCA and LOAD for both SNPs no significant differences were found. However, a tendency toward higher methylation levels in the LOAD sample can be seen for both SNPs, which could be explored with a larger sample size.

A recent longitudinal study of individuals affected with AD has provided evidence in favor of a hypothesis that proposes a change

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

in the methylation status of *PM20D1* (possibly coupled with a change in its expression) throughout the AD pathology.[50] At the initial stages of the disorder of mild cognitive impairment, the gene is hypomethylated in comparison to controls. This hypomethylation could be coupled with a higher expression of the gene, which could play a protective role. In later stages of the disease, the gene is then hypermethylated, as has been reported in several studies.[16,37] Wang et al propose that the turning point for the change in methylation of the gene is at approximately 78–79 years old. Our results fit with this hypothesis, since we observe hypermethylation of the gene in affected individuals, and they are all over 90 years of age and present with moderate or severe dementia. From the point of view of SCA group, the hypomethylation of PM20D1 we detected could play a protective role against the development of LOAD.

Interestingly, *PM20D1* has been associated with obesity and diabetes, which are risk factors for AD.[51-53] *PM20D1* lies within the Parkinson's disease 16 (susceptibility) locus on chromosome 1, which has previously been associated with Parkinson's disease.[54] In addition, *PM20D1* has been reported to be differentially methylated in individuals with obesity and multiple sclerosis patients.[55,56] The protective role of *PM20D1* may be explained by the fact that it has previously been shown to activate mitochondrial uncoupling, which promotes neuronal survival because it contributes to the adaptive responses to bioenergetic and oxidative stressors.

The gene *LHX6* encodes a member of a protein family that contains the LIM domain, a unique cysteine-rich zinc-binding domain. The protein is a transcription factor involved in embryogenesis and in the expression of a subset of genes involved in interneuron migration and development.[57,58] The gene is highly expressed in neural crest-derived mesenchyme cells.[59] In our SCA sample, this region was found to be hypermethylated.

Another one of the differentially methylated genes, *CERS3* is a member of the ceramide synthase family of genes, this region was found to be hypermethylated in our SCA sample. This type of enzyme regulates sphingolipid synthesis by catalyzing the formation of ceramides from the sphingoid base and acyl-CoA substrates. Several lines of evidence suggest that there is a causal link between ceramide or sphingolipids levels and neurodegenerative diseases such as AD.[60] However, *CERS3* is the only one out of the six-ceramide synthases that are not expressed in brain tissue.[61] Its expression has been reported especially in testis and skin.[62,63] In addition, the gene is associated with ichthyosis.[64]

The gene *vtRNA2-1*, also known as *nc886*, encodes a non-coding RNA that represses PKR, a double-stranded RNA-dependent kinase, involved in tumor suppression.[65,66] This gene is often hypermethylated and repressed in cancers.[67-69] Romanelli et al[70] showed that the *vtRNA2-1* region is variably maternally imprinted, namely, it has allele-specific methylation and shows variable levels of methylation levels among tissues. In our sample, this region was found to be hypermethylated in the SCA group.

For the remaining four genes and pseudogenes there is little information regarding their function. The protein phosphatase 1 regulatory inhibitor subunit 2 Pseudogene, *PPP1R2P1*, is a pseudogene, which we found to be hypomethylated in our SCA sample. Evidence shows that *PPP1R2P1* is expressed, but its function remains unknown.[71] The gene *PRSS22* encodes a member of the trypsin family of serine proteases. The enzyme is expressed in the airway epithelial cells in a developmentally regulated manner.[72] *DOC2GP* is a pseudogene expressed mainly in the heart, spleen, and thyroid, while the functionally uncharacterized *C17orf98* is expressed in testis.[73]

Regarding the correlation between blood and brain methylation in our analysis, it was very variable among the DMRs. The lowest correlation was −0.004, and the highest was 0.701. Three out of the six DMRs that contain genes present a correlation above 0.55. The DMR with the highest correlation contains neither a gene nor a pseudogene. A limitation in the determination of the correlations was that the BECon tool does not have information for all the CpG sites of our DMRs, so the average correlation between blood and brain was calculated excluding some sites from the DMR regions. Recently a comparison of methylation levels between blood and four regions of the brain in an independent cohort was performed for CpG sites specifically in *PM20D1*.[50] Correlations were high and significant for all comparisons, with a range of correlation coefficients from 0.857 to 0.976. Therefore, it is safe to assume that the methylation changes we observed in the gene are also occurring in the brain. Current evidence suggests that in many cases blood is an appropriate sample for such studies, which has many implications including the possibility of biomarker detection for early diagnosis.

An important limitation of our study is the small sample size. Nevertheless, our confirmation of the previously reported role of *PM20D1* in AD proves that valid results can be obtained with small samples. Besides, there is increasing recognition of the value of studying ancestrally diverse populations, and Latin American populations have been historically underrepresented in large-scale genomic and epigenomic studies.[74,75]

Conclusions

In conclusion, our work adds evidence to suggest that long-lived individuals have a lower epigenetic age than their chronological age. This work also supports the possible role of *PM20D1* in protection against AD, by showing differential methylation in blood of affected and unaffected individuals in a longer-lived population. We also confirmed the association between rs708727 genotypes and methylation levels in the gene's promoter, which provides further evidence in favor of the influence of genetic factors on *PM20D1* expression (which in turn may influence susceptibility to develop AD). Finally, we found other differentially methylated regions including genes involved in cell cycle regulation, embryogenesis, synthesis of ceramides, and migration of interneurons to the cerebral cortex. These genomic regions might play a role in AD and SCA, and merit further studies.

Acknowledgment. We thank the participants for their collaboration.

Funding Statement. This work received funding support from NIH Fogarty International Center & National Institute on Aging (grant: R21TW009258), Alzheimer's Association, Universidad de Costa Rica (projects A4323 and B8377), and Sistema de Estudios de Posgrado from Universidad de Costa Rica. This study received partial funding from the "Instituto Nacional de Medicina Genomica."

Disclosures. The authors have no conflicts of interest to disclose.

Author Contributions. Conceptualization: C.C.-V., L.M.-V., D.V., J.M.S., H.N., H.R., G.C.-S.; Data curation: C.C.-V., L.M.-V., D.V., A.D.G.-M.; Formal analysis: C.C.-V., J.J.M.-M., D.V., A.D.G.-M., G.C.-S.; Funding acquisition: J.M.S., H.N., H.R.; Investigation: C.C.-V., J.J.M.-M., L.M.-V., D.V., A.D.G.-

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

M., J.M.S., H.N., H.R., G.C.-S.; Methodology: C.C.-V., J.J.M.-M., L.M.-V., D.V., A.D.G.-M., J.M.S., H.N., H.R., G.C.-S.; Project administration: L.M.-V., H.N., H.R., G.C.-S.; Supervision: C.C.-V., J.J.M.-M., G.C.-S.; Validation: J.J.M.-M.; Writing—original draft: C.C.-V., J.J.M.-M., L.M.-V., D.V., J.M.S., H.N., H.R., G.C.-S.; Writing—review and editing: C.C.-V., J.J.M.-M., L.M.-V., D.V., A.D.G.-M., J.M.S., H.N., H.R., G.C.-S.

References

1. Scheltens P, Blennow K, Breteler MMB, *et al.* Alzheimer's disease. *Lancet Lond Engl.* 2016;**388**(10043):505–517. doi:10.1016/S0140-6736(15)01124-1.
2. Farrer LA, Cupples LA, Haines JL, *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA.* 1997;**278**(16):1349–1356.
3. Reitz C, Brayne C, Mayeux R. Epidemiology of Alzheimer disease. *Nat Rev Neurol.* 2011;**7**(3):137–152. doi:10.1038/nrneurol.2011.2.
4. Silva MVF, Loures C de MG, Alves LCV, *et al.* Alzheimer's disease: risk factors and potentially protective measures. *J Biomed Sci.* 2019;**26**(1):33. doi:10.1186/s12929-019-0524-y.
5. Robinson M, Lee BY, Hane FT. Recent progress in Alzheimer's disease research, part 2: genetics and epidemiology. *J Alzheimers Dis JAD.* 2017;**57**(2):317–330. doi:10.3233/JAD-161149.
6. Cuyvers E, Sleegers K. Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. *Lancet Neurol.* 2016;**15**(8):857–868. doi:10.1016/S1474-4422(16)00127-7.
7. Andrews SJ, Fulton-Howard B, Goate A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol.* 2020;**19**(4):326–335. doi:10.1016/S1474-4422(19)30435-1.
8. Suri S, Heise V, Trachtenberg AJ, *et al.* The forgotten APOE allele: a review of the evidence and suggested mechanisms for the protective effect of APOE ε2. *Neurosci Biobehav Rev.* 2013;**37**(10 Pt 2):2878–2886. doi:10.1016/j.neubiorev.2013.10.010.
9. Vemuri P, Knopman DS, Lesnick TG, *et al.* Evaluation of amyloid protective factors and Alzheimer disease neurodegeneration protective factors in elderly individuals. *JAMA Neurol.* 2017;**74**(6):718–726. doi:10.1001/jamaneurol.2017.0244.
10. Scacchi R, De Bernardini L, Mantuano E, *et al.* Apolipoprotein E (APOE) allele frequencies in late-onset sporadic Alzheimer's disease (AD), mixed dementia and vascular dementia: lack of association of epsilon 4 allele with AD in Italian octogenarian patients. *Neurosci Lett.* 1995;**201**(3):231–234. doi:10.1016/0304-3940(95)12190-0.
11. Corder EH, Saunders AM, Strittmatter WJ, *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;**261**(5123):921–923. doi:10.1126/science.8346443.
12. Corder EH, Saunders AM, Risch NJ, *et al.* Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet.* 1994;**7**(2):180–184. doi:10.1038/ng0694-180.
13. Valerio D, Raventos H, Schmeidler J, *et al.* Association of apolipoprotein E-e4 and dementia declines with age. *Am J Geriatr Psychiatry Off J Am Assoc Geriatr Psychiatry.* 2014;**22**(10):957–960. doi:10.1016/j.jagp.2014.03.008.
14. Silverman JM, Schmeidler J. The protected survivor model: using resistant successful cognitive aging to identify protection in the very old. *Med Hypotheses.* 2018;**110**:9–14. doi:10.1016/j.mehy.2017.10.022.
15. Sanchez-Mut JV, Gräff J. Epigenetic alterations in Alzheimer's disease. *Front Behav Neurosci.* 2015;**9**:347. doi:10.3389/fnbeh.2015.00347.
16. Zhang L, Silva TC, Young JI, *et al.* Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nat Commun.* 2020;**11**(1):6114. doi:10.1038/s41467-020-19791-w.
17. Liu L, Lauro BM, Ding L, *et al.* Multiple BACE1 inhibitors abnormally increase the BACE1 protein level in neurons by prolonging its half-life. *Alzheimers Dement J Alzheimers Assoc.* 2019;**15**(9):1183–1194. doi:10.1016/j.jalz.2019.06.3918.
18. Walton E, Hass J, Liu J, *et al.* Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. *Schizophr Bull.* 2016;**42**(2):406–414. doi:10.1093/schbul/sbv074.
19. Edgar RD, Jones MJ, Meaney MJ, *et al.* BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Transl Psychiatry.* 2017;**7**(8):e1187. doi:10.1038/tp.2017.171.
20. Braun PR, Han S, Hing B, *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl Psychiatry.* 2019;**9**(1):47. doi:10.1038/s41398-019-0376-y.
21. Chen J, Zang Z, Braun U, *et al.* Association of a reproducible epigenetic risk profile for schizophrenia with brain methylation and function. *JAMA Psychiatry.* 2020;**77**(6):628–636. doi:10.1001/jamapsychiatry.2019.4792.
22. Depp CA, Jeste DV. Definitions and predictors of successful aging: a comprehensive review of larger quantitative studies. *Am J Geriatr Psychiatry Off J Am Assoc Geriatr Psychiatry.* 2006;**14**(1):6–20. doi:10.1097/01.JGP.0000192501.03069.bc.
23. Silverman JM, Schnaider-Beeri M, Grossman HT, *et al.* A phenotype for genetic studies of successful cognitive aging. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet.* 2008;**147B**(2):167–173. doi:10.1002/ajmg.b.30483.
24. Greenwood TA, Beeri MS, Schmeidler J, *et al.* Heritability of cognitive functions in families of successful cognitive aging probands from the Central Valley of Costa Rica. *J Alzheimers Dis JAD.* 2011;**27**(4):897–907. doi:10.3233/JAD-2011-110782.
25. Rosero-Bixby L, Dow WH. Surprising SES gradients in mortality, health, and biomarkers in a Latin American population of adults. *J Gerontol B Psychol Sci Soc Sci.* 2009;**64**(1):105–117. doi:10.1093/geronb/gbn004.
26. Rosero-Bixby L. The exceptionally high life expectancy of Costa Rican nonagenarians. *Demography.* 2008;**45**(3):673–691. doi:10.1353/dem.0.0011.
27. Rosero-Bixby L, Dow WH. Exploring why Costa Rica outperforms the United States in life expectancy: a tale of two inequality gradients. *Proc Natl Acad Sci U S A.* 2016;**113**(5):1130–1137. doi:10.1073/pnas.1521917112.
28. Hughes CP, Berg L, Danziger WL, *et al.* A new clinical scale for the staging of dementia. *Br J Psychiatry J Ment Sci.* 1982;**140**:566–572. doi:10.1192/bjp.140.6.566.
29. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;**12**(3):189–198. doi:10.1016/0022-3956(75)90026-6.
30. Cheung WY, Hubert N, Landry BS. A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. *PCR Methods Appl.* 1993;**3**(1):69–70. doi:10.1101/gr.3.1.69.
31. Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinforma Oxf Engl.* 2014;**30**(10):1363–1369. doi:10.1093/bioinformatics/btu049.
32. Fortin J-P, Labbe A, Lemire M, *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 2014;**15**(12):503. doi:10.1186/s13059-014-0503-2.
33. Bibikova M, Barnes B, Tsan C, *et al.* High density DNA methylation array with single CpG site resolution. *Genomics.* 2011;**98**(4):288–295. doi:10.1016/j.ygeno.2011.07.007.
34. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;**14**(10):R115. doi:10.1186/gb-2013-14-10-r115.
35. Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;**13**:86. doi:10.1186/1471-2105-13-86.
36. Morris TJ, Butcher LM, Feber A, *et al.* ChAMP: 450k chip analysis methylation pipeline. *Bioinforma Oxf Engl.* 2014;**30**(3):428–430. doi:10.1093/bioinformatics/btt684.
37. Sanchez-Mut JV, Heyn H, Silva BA, *et al.* PM20D1 is a quantitative trait locus associated with Alzheimer's disease. *Nat Med.* 2018;**24**(5):598–603. doi:10.1038/s41591-018-0013-y.
38. Sanchez-Mut JV, Glauser L, Monk D, *et al.* Comprehensive analysis of PM20D1 QTL in Alzheimer's disease. *Clin Epigenetics.* 2020;**12**(1):20. doi:10.1186/s13148-020-0814-y

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press

39. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet.* 2018;**19**(6):371–384. doi:10.1038/s41576-018-0004-3.
40. Marioni RE, Shah S, McRae AF, *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015;**16**:25. doi:10.1186/s13059-015-0584-6.
41. Marioni RE, Shah S, McRae AF, *et al.* The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int J Epidemiol.* 2015;**44**(4):1388–1396. doi:10.1093/ije/dyu277.
42. Levine ME, Lu AT, Bennett DA, *et al.* Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging.* 2015;**7**(12):1198–1211. doi:10.18632/aging.100864.
43. McEwen LM, Morin AM, Edgar RD, *et al.* Differential DNA methylation and lymphocyte proportions in a Costa Rican high longevity region. *Epigenetics Chromatin.* 2017;**10**:21. doi:10.1186/s13072-017-0128-2.
44. Horvath S, Gurven M, Levine ME, *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* 2016;**17**(1):171. doi:10.1186/s13059-016-1030-0.
45. Armstrong NJ, Mather KA, Thalamuthu A, *et al.* Aging, exceptional longevity and comparisons of the Hannum and Horvath epigenetic clocks. *Epigenomics.* 2017;**9**(5):689–700. doi:10.2217/epi-2016-0179.
46. Horvath S, Pirazzini C, Bacalini MG, *et al.* Decreased epigenetic age of PBMCs from Italian semi-supercentenarians and their offspring. *Aging.* 2015;**7**(12):1159–1170. doi:10.18632/aging.100861.
47. Heyn H, Moran S, Hernando-Herraez I, *et al.* DNA methylation contributes to natural human variation. *Genome Res.* 2013;**23**(9):1363–1372. doi:10.1101/gr.154187.112.
48. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* 2015;**348**(6235):648–660. doi:10.1126/science.1262110.
49. Li QS, Vasanthakumar A, Davis JW, *et al.* Association of peripheral blood DNA methylation level with Alzheimer's disease progression. *Clin Epigenetics.* 2021;**13**(1):191. doi:10.1186/s13148-021-01179-2.
50. Wang Q, Chen Y, Readhead B, *et al.* Longitudinal data in peripheral blood confirm that PM20D1 is a quantitative trait locus (QTL) for Alzheimer's disease and implicate its dynamic role in disease progression. *Clin Epigenetics.* 2020;**12**(1):189. doi:10.1186/s13148-020-00984-5.
51. Long JZ, Svensson KJ, Bateman LA, *et al.* The secreted enzyme PM20D1 regulates lipidated amino acid uncouplers of mitochondria. *Cell.* 2016;**166**(2):424–435. doi:10.1016/j.cell.2016.05.071.
52. Larrick JW, Larrick JW, Mendelsohn AR. Uncoupling mitochondrial respiration for diabesity. *Rejuvenation Res.* 2016;**19**(4):337–340. doi:10.1089/rej.2016.1859.
53. Profenno LA, Porsteinsson AP, Faraone SV. Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. *Biol Psychiatry.* 2010;**67**(6):505–512. doi:10.1016/j.biopsych.2009.02.013.
54. Simon-Sanchez J, Schulte C, Bras JM, *et al.* Genome-Wide Association Study reveals genetic risk underlying Parkinson's disease. *Nat Genet.* 2009;**41**(12):1308–1312. doi:10.1038/ng.487.
55. Feinberg AP, Irizarry RA, Fradin D, *et al.* Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Transl Med.* 2010;**2**(49):49ra67. doi:10.1126/scitranslmed.3001262.
56. Maltby VE, Lea RA, Sanders KA, *et al.* Differential methylation at MHC in CD4+ T cells is associated with multiple sclerosis independently of HLA-DRB1. *Clin Epigenetics.* 2017;**9**:71. doi:10.1186/s13148-017-0371-1.
57. Lavdas AA, Grigoriou M, Pachnis V, *et al.* The medial ganglionic eminence gives rise to a population of early neurons in the developing cerebral cortex. *J Neurosci Off J Soc Neurosci.* 1999;**19**(18):7881–7888.
58. Zhang Z, Gutierrez D, Li X, *et al.* The LIM homeodomain transcription factor LHX6: a transcriptional repressor that interacts with pituitary homeobox 2 (PITX2) to regulate odontogenesis. *J Biol Chem.* 2013;**288**(4):2485–2500. doi:10.1074/jbc.M112.402933.
59. Grigoriou M, Tucker AS, Sharpe PT, *et al.* Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM homeodomain encoding genes, suggests a role in mammalian head development. *Dev Camb Engl.* 1998;**125**(11):2063–2074.
60. Ben-David O, Futerman AH. The role of the ceramide acyl chain length in neurodegeneration: involvement of ceramide synthases. *Neuromolecular Med.* 2010;**12**(4):341–350. doi:10.1007/s12017-010-8114-x.
61. Becker I, Wang-Eckhardt L, Yaghootfam A, *et al.* Differential expression of (dihydro)ceramide synthases in mouse brain: oligodendrocyte-specific expression of CerS2/Lass2. *Histochem Cell Biol.* 2008;**129**(2):233–241. doi:10.1007/s00418-007-0344-0.
62. Riebeling C, Allegood JC, Wang E, *et al.* Two mammalian longevity assurance gene (LAG1) family members, trh1 and trh4, regulate dihydroceramide synthesis using different fatty acyl-CoA donors. *J Biol Chem.* 2003;**278**(44):43452–43459. doi:10.1074/jbc.M307104200.
63. Mizutani Y, Kihara A, Igarashi Y. Mammalian Lass6 and its related family members regulate synthesis of specific ceramides. *Biochem J.* 2005;**390**(Pt 1):263–271. doi:10.1042/BJ20050291.
64. Radner FPW, Marrakchi S, Kirchmeier P, *et al.* Mutations in CERS3 cause autosomal recessive congenital ichthyosis in humans. *PLoS Genet.* 2013;**9**(6):e1003536. doi:10.1371/journal.pgen.1003536.
65. Jeon SH, Lee K, Lee KS, *et al.* Characterization of the direct physical interaction of nc886, a cellular non-coding RNA, and PKR. *FEBS Lett.* 2012;**586**(19):3477–3484. doi:10.1016/j.febslet.2012.07.076.
66. Jeon SH, Johnson BH, Lee YS. A tumor surveillance model: a non-coding RNA senses neoplastic cells and its protein partner signals cell death. *Int J Mol Sci.* 2012;**13**(10):13134–13139. doi:10.3390/ijms131013134.
67. Fort RS, Mathó C, Geraldo MV, *et al.* Nc886 is epigenetically repressed in prostate cancer and acts as a tumor suppressor through the inhibition of cell growth. *BMC Cancer.* 2018;**18**(1):127. doi:10.1186/s12885-018-4049-7.
68. Lee H-S, Lee K, Jang H-J, *et al.* Epigenetic silencing of the non-coding RNA nc886 provokes oncogenes during human esophageal tumorigenesis. *Oncotarget.* 2014;**5**(11):3472–3481. doi:10.18632/oncotarget.1927.
69. Lee K-S, Park J-L, Lee K, *et al.* nc886, a non-coding RNA of anti-proliferative role, is suppressed by CpG DNA methylation in human gastric cancer. *Oncotarget.* 2014;**5**(11):3944–3955. doi:10.18632/oncotarget.2047.
70. Romanelli V, Nakabayashi K, Vizoso M, *et al.* Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics.* 2014;**9**(5):783–790. doi:10.4161/epi.28323.
71. Korrodi-Gregório L, Abrantes J, Muller T, *et al.* Not so pseudo: the evolutionary history of protein phosphatase 1 regulatory subunit 2 and related pseudogenes. *BMC Evol Biol.* 2013;**13**:242. doi:10.1186/1471-2148-13-242.
72. Wong GW, Yasuda S, Madhusudhan MS, *et al.* Human tryptase epsilon (PRSS22), a new member of the chromosome 16p13.3 family of human serine proteases expressed in airway epithelial cells. *J Biol Chem.* 2001;**276**(52):49169–49182. doi:10.1074/jbc.M108677200.
73. Fagerberg L, Hallström BM, Oksvold P, *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics MCP.* 2014;**13**(2):397–406. doi:10.1074/mcp.M113.035600.
74. Hindorff LA, Bonham VL, Brody LC, *et al.* Prioritizing diversity in human genomics research. *Nat Rev Genet.* 2018;**19**(3):175–185. doi:10.1038/nrg.2017.89.
75. Peterson RE, Kuchenbaecker K, Walters RK, *et al.* Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 2019;**179**(3):589–603. doi:10.1016/j.cell.2019.08.051.

https://doi.org/10.1017/S109285292100105X Published online by Cambridge University Press