MATERIALS AND METHODS

Study design and populations

We performed a cross-sectional study to determine the association between genome-wide DNA methylation and age-independent cardiovascular risk (CVR). We defined two complementary approaches to identify this CVR, considering it as (a) a continuous trait, using the residuals from the association between CVR and age (residuals approach) and (b) a discrete trait, using the difference between vascular and chronological age (Δage approach). We designed a two-stage EWAS for each approach.

The two populations used in this study have been described in previous EWAS performed in our group.[1] Briefly, in the discovery stage, we used data from the follow-up visit of the REGICOR (REgistre GIroní del COR) cohort in 2009-2013 (n=4980; response rate, 78.4%). This cohort includes participants from towns representing the urban and rural diversity of Girona Province (Spain)[2] who continued to reside in the same towns where they were originally enrolled in 2003-2005. We randomly selected a subsample of 648 individuals, all of European descent. The study was approved by the local ethics committee and meets the principles expressed in the Declaration of Helsinki and the relevant Spanish legislation. All participants provided informed written consent prior to the studies.

In the validation stage, we obtained data from the Framingham Offspring Study through the database of Genotypes and Phenotypes (dbGAP, http://dbgap.ncbi.nlm.nih.gov; project number #9047). We included the exam 8 participants with available DNA methylation data (n=2,568).

REGICOR data were used for cross-sectional assessment of the association between classical vascular risk factors (VRF) and CpGs showing differential methylation in relation to age-independent CVR. Finally, we built an epigenetic risk score and assessed its association with arterial stiffness in the REGICOR cohort, and with incident coronary and cardiovascular events in the Framingham cohort.

Vascular risk factors measurement and estimation of the age-independent cardiovascular risk

The procedures used to measure VRFs and to collect data and samples from both populations have already been described.[1] Briefly, examinations and fasting blood extractions were performed and questionnaires administered by a group of trained nurses in the REGICOR study. Standardized methods and questionnaires were used to collect

sociodemographic, lifestyle, and VRFs information. Further details on the measurement of the biochemical parameters representing the VRFs are extensively described elsewhere.[1] In the Framingham Offspring Study, measurements and data were obtained through dbGaP and correspond to examination 8.

We used a validated risk function (FRESCO) to estimate the CVR and vascular age of both the discovery and validation populations.[3] This function has been developed and validated for the Spanish population, and although the absolute CVR estimation is not valid for the American population, the estimation of vascular age and of the residuals from CVR/age association can be used in any population regardless of any difference in absolute CVR.[4] As the FRESCO function is only valid to estimate CVR of individuals between 35 and 79 years old,[3] we excluded those younger and older than this chronological age range. We also excluded those individuals with no available information for any of the VRFs considered in the risk function. We used both of these REGICOR and Framingham Offspring Cohort populations in the residuals approach. For the Δage approach, we further excluded those individuals whose vascular age was <35 or >79 years.

Assessment of DNA methylation status

DNA extraction and methylation assessment have been fully described in a previous report.[1] In brief, DNA was extracted from whole peripheral blood in the REGICOR cohort, and from buffy coat in the Framingham Offspring Study. Genome-wide DNA methylation was assessed using the Infinium HumanMethylation450 BeadChip (Illumina, CA, USA), according to the standard protocol. This array is based on bisulfite conversion of unmethylated cytosines across the genome, and analyzes over 485,000 CpGs per sample.[5]

Analysis and quality control of the raw data has also been previously described.[1] Methylation status at each CpG site was reported by both M-values and β-values. Equation 1 was used to calculate the M-values, while Equation 2 was applied to estimate the β-values.

$$M_{value} = \log_2\left(\frac{M_i+\alpha}{U_i+\alpha}\right) \qquad \text{Equation 1}$$

$$\beta_{value} = \frac{M_i}{M_i+U_i+\alpha} \qquad \text{Equation 2}$$

Where:

- M_i = intensity of methylated probes,
- U_i = intensity of unmethylated probes, and
- α = constant offset (α=1 and α=100 in the respective equations).

M-values close to 0 mean the CpG site is half-methylated. Positive M-values indicate the presence of more methylated than unmethylated cytosines, while negative M-values denote the opposite ratio. β-values range between 0 (completely unmethylated) and 1 (completely methylated).

Owing to its statistical robustness, we used standardized M-values in the EWASs and the metaanalysis. Moreover, to control for batch effect we standardized the M-values by batch using Equation 3:[1]

$$Z = \frac{x - \bar{x}}{\sqrt{\frac{\sum (x - \bar{x})^2}{(n-1)}}} \quad \text{Equation 3}$$

Where:

- Z = standardized M-value,
- X = M-value for a specific individual,
- $\bar{X}$ = mean of M-value for a specific batch, and
- n = sample size.

For the development of methylation risk scores (MRS) we used β-values, normalized by applying the Dassen method[6], of the CpGs identified as differentially methylated in relation to age-independent CVR.

Information about each CpG location in the genome was obtained using the annotation provided by Illumina (HumanMethylation450 v1.2 Manifest File.csv, www.illumina.com) and Genome Browser.[7]

Subclinical atherosclerosis measurement: arterial stiffness

Two trained technicians performed the ultrasonographic evaluation of the selected subjects utilizing a 10MHz probe (Acuson XP128) and specific acquisition software (Acuson-Siemens; Mountainview, California, United States). Longitudinal images of left and right common carotid arteries were obtained with M-Mode and B-Mode, and stored for offline analysis. Acquired images were analyzed applying a standardized protocol and software.[8] Inter-adventitial systolic (SD) and diastolic diameter (DD) were quantified in the proximal centimeter from the carotid bulb, following international recommendations.[9]

Stiffness of the common carotid artery was estimated by calculating the distensibility coefficient (DC),[10] following equation 4:

$$DC = \frac{2\,\Delta D\,DD + \Delta D^2}{\frac{PP}{7.5}\,DD^2} \quad \text{Equation 4}$$

98 Where:

99 - ΔD = strain or delta diameter (calculated as SD-DD),
100 - DD = diastolic diameter, and
101 - PP= brachial pulse pressure (calculated as the difference between the systolic blood
102 pressure [SBP] and the diastolic blood pressure [DBP]).

103 In Equation 4, DD and ΔD are expressed in meters, whereas PP is expressed in mmHg. The
104 resulting DC is expressed in 10-3*kPa-1. This index represents the stroke change in an
105 artery's cross-sectional area, normalized for the total diastolic cross-sectional area, in
106 response to a stress represented by PP. The lower the DC, the higher the arterial stiffness.

107 Additionally, carotid Pulse Wave Velocity (PWV) was calculated using the Bramwell-Hill
108 equation (equation 5):[10]

109 $$PWV = \sqrt{\frac{1}{DC \cdot \rho}} \quad \text{Equation 5}$$

110 Where:

111 - DC = distensibility coefficient, and
112 - ρ = blood density, assumed as 1050 Kg/m^3.

113

114 **Clinical cardiovascular events assessment**

115 We obtained data from the Framingham Offspring Study through the database of Genotypes
116 and Phenotypes (dbGAP; http://dbgap.ncbi.nlm.nih.gov; project number #9047). We used
117 exam 8 as the baseline visit and the follow-up included events recorded until exam 12.

118

119 **Statistical analysis**

120 Robust multivariable linear regression was used in the EWAS analyses. We considered DNA
121 methylation as the outcome and CVR independent of age as the exposure or independent
122 variables. CVR independent of age was considered as (a) the residuals from CVR/age
123 association, and (b) the difference between vascular and chronological age.

124 To increase the sensitivity of our study, we defined four models for each approach (residuals
125 and Δage). In model 1, standardized M-values were adjusted for three confounders: sex,
126 age, and estimated cell count. In model 3, they were additionally adjusted for smoking status.
127 Models 2 and 4 contained the same covariates as models 1 and 3, respectively, but were
128 further adjusted for surrogate variables to control for unmeasured confounding.[11] Cell count

129 was estimated using Houseman's algorithm implemented in R::minfi,[12,13] while surrogate
130 variables were assessed using the R::sva R package.[14]

131 We validated those CpGs associated with age-independent CVR which exceeded an
132 arbitrary P-value threshold of $1\cdot10^{-5}$ for any of the four models in each approach. Then, we
133 performed a fixed-effects metaanalysis weighted by the inverse of the variance using data
134 from the discovery and validation stages for Model 2 in both approaches. We focused on
135 Model 2 for two main reasons: (1) the smoking status used as a covariate in models 2 and 4
136 is a well-known CVR factor, and (2) models adjusted for surrogate variables are preferred in
137 omics data analysis. We defined as statistically significant any association fulfilling the
138 Bonferroni criteria (P-value<$1.17\cdot10^{-7}$; calculated as 0.05/427,948 probes that passed the
139 quality control).

140 To further study whether the validated CpGs were associated with one or more individual
141 VRFs, or with none of them individually, we assessed this association in the REGICOR and
142 Framingham populations. Considering methylation as the outcome, we used a multivariable
143 linear regression model adjusted for triglycerides, total and HDL cholesterol, systolic and
144 diastolic blood pressures, body mass index (BMI), smoking, diabetes, sex, age, and cell type.
145 In the case of blood pressure and lipid levels, we performed the analyses in subsamples of
146 non-treated individuals.

147 Based on the validated CpGs, we developed a weighted MRS, considering normalized β-
148 values, for each approach in both populations. These four MRS models summarize an
149 individual's epigenetic predisposition to higher age-independent CVR. The weights for each
150 CpG were based on the coefficients of the metaanalysis of normalized β-values from Model 2
151 following Equation 6:

152 $$MRS = \sum_{i=1}^{N} \beta_i^{meta} \cdot \beta_i^{norm} \quad \text{Equation 6}$$

153 Where:

154 - *MRS* = methylation risk score for a specific individual,
155 - *i* = CpG,
156 - *N* = CpG sample size,
157 - β^{meta} = coefficient of the metaanalysis for each CpG,
158 - β^{norm} = normalized β-value for each CpG,

159

160 Since MRSs were estimated in different units in each approach, they were standardized to
161 further test them, according to Equation 7.

162 $$MRS_z = \frac{MRS - \overline{MRS}}{\sqrt{\frac{\Sigma(MRS-\overline{MRS})^2}{(n-1)}}}$$ Equation 7

163 Where:

164 - MRS_Z = standardized methylation risk score,
165 - MRS = methylation risk score for a specific individual,
166 - $\overline{MRS}$ = mean of methylation risk scores for a specific CpG, and
167 - n = population sample size.

168 Finally, we evaluated the association between these standardized MRSs and clinical and
169 subclinical measurements of atherosclerosis. Multivariable linear regression was used in the
170 analysis of the association between MRSs and subclinical parameters (arterial distensibility
171 and pulse wave velocity) in REGICOR. We analyzed the association between MRSs and
172 CVD and CHD incidence in Framingham data, using Cox regression. All analyses were
173 adjusted for age and sex (model 1) and the VRFs considered in the FRESCO risk function
174 (age, sex, total and HDL cholesterol, diabetes, smoking status, SBP, and hypertensive
175 treatment, model 2).

176 All statistical analyses were performed using RStudio software (version 3.1.2).

177

178 **References**

179 1. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, et al. Identification and validation of seven
180 new loci showing differential DNA methylation related to serum lipid profile: an epigenome-
181 wide approach. The REGICOR study. *Hum Mol Genet*. 2016;25:ddw285.

182 2. Grau M, Subirana I, Elosua R, Solanas P, Ramos R, Masiá R, Cordón F, Sala J, Juvinyà
183 D, Cerezo C, Fitó M, Vila J, Covas MI, Marrugat J. Trends in cardiovascular risk factor
184 prevalence (1995-2000-2005) in northeastern Spain. *Eur J Cardiovasc Prev Rehabil*.
185 2007;14:653-659.

186 3. Marrugat J, Subirana I, Ramos R, et al. Derivation and validation of a set of 10-year
187 cardiovascular risk predictive functions in Spain: The FRESCO Study. *Prev Med*.
188 2014;61:66-74.

189 4. Cuende JI, Cuende N, Calaveras-Lagartos J. How to calculate vascular age with the
190 SCORE project scales: a new method of cardiovascular risk evaluation. *Eur Heart J*.
191 2010;31:2351-2358.

192 5. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP,
193 Gunderson KL, Fan J-B, Shen R. High density DNA methylation array with single CpG site

194 resolution. *Genomics*. 2011;98:288-295.

195 6. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach
196 to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293.

197 7. Speir ML, Zweig AS, Rosenbloom KR, et al. The UCSC Genome Browser database: 2016
198 update. *Nucleic Acids Res*. 2016;44:D717-D725.

199 8. de Groot E, van Leuven SI, Duivenvoorden R, Meuwese MC, Akdim F, Bots ML, Kastelein
200 JJ. Measurement of carotid intima–media thickness to assess progression and regression of
201 atherosclerosis. *Nat Clin Pract Cardiovasc Med*. 2008;5:280-288.

202 9. Stein JH, Korcarz CE, Hurst RT, Lonn E, Kendall CB, Mohler ER, Najjar SS, Rembold CM,
203 Post WS, American Society of Echocardiography Carotid Intima-Media Thickness Task
204 Force. Use of Carotid Ultrasound to Identify Subclinical Vascular Disease and Evaluate
205 Cardiovascular Disease Risk: A Consensus Statement from the American Society of
206 Echocardiography Carotid Intima-Media Thickness Task Force Endorsed by the Society for
207 Vascular Medicine. *J Am Soc Echocardiogr*. 2008;21:93-111.

208 10. Van Bortel LM, Duprez D, Starmans-Kool MJ, Safar ME, Giannattasio C, Cockcroft J,
209 Kaiser DR, Thuillez C. Clinical applications of arterial stiffness, Task Force III:
210 recommendations for user procedures. *Am J Hypertens*. 2002;15:445-452.

211 11. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis
212 to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics*.
213 2011;27:1496-1505.

214 12. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH,
215 Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture
216 distribution. *BMC Bioinformatics*. 2012;13:86.

217 13. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry
218 RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium
219 DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.

220 14. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing
221 batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*.
222 2012;28:882-883.