bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Lifestyle and Genetic Factors Modify Parent-of-Origin Effects on the
2 # Human Methylome

3

4 Yanni Zeng[1,2,3]*, Carmen Amador[4], Chenhao Gao[1], Rosie M. Walker[5,6], Stewart W. Morris[5],
5 Archie Campbell[5], Azra Frkatović[7], Rebecca A Madden[8], Mark J. Adams[8], Shuai He[9],
6 Andrew D. Bretherick[4], Caroline Hayward[4], David J. Porteous[5], James F. Wilson[4,10], Kathryn
7 L. Evans[5], Andrew M. McIntosh[8], Pau Navarro[4†], and Chris S. Haley[4,11†]
8
9 1. Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou
10 510080, China.
11 2. Guangdong Province Translational Forensic Medicine Engineering Technology Research Center,
12 Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou 510080, China.
13 3. Guangdong Province Key Laboratory of Brain Function and Disease, Zhongshan School of Medicine,
14 Sun Yat-Sen University, Guangzhou 510080, China.
15 4. MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
16 5. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of
17 Edinburgh, Edinburgh, UK
18 6. Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh
19 BioQuarter, Edinburgh, UK
20 7. Genos Glycoscience Research Laboratory, Borongajska cesta 83h, 10000 Zagreb, Croatia
21 8. Division of Psychiatry, University of Edinburgh, Edinburgh, United Kingdom
22 9. Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China,
23 Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal
24 Carcinoma Diagnosis and Therapy, Guangzhou, 510060, China
25 10. Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK
26 11.Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh,
27 UK
28
29 * Corresponding author
30 zengyn5@mail.sysu.edu.cn
31 †equally contributed senior authors
32 chris.haley@ed.ac.uk, pau.navarro@ed.ac.uk
33

34 Word counts: 2979 (exclude methods), 4470 (include methods)

35

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36 **Abstract**

37 Background: parent-of-origin effects (POE) play important roles in development and
38 complex disease and thus understanding their regulation and associated molecular and
39 phenotypic variation are warranted. Previous studies have mainly focused on the detection of
40 genomic regions or phenotypes regulated by POE. Understanding whether POE may be
41 modified by environmental or genetic exposures is important for understanding of the source
42 of POE-associated variation, but only a few case studies addressing these modifiable POE
43 exist.

44 Methods: in order to understand this high order of POE regulation, we screened 101 genetic
45 and environmental factors such as "predicted mRNA expression levels" of DNA
46 methylation/imprinting machinery genes and early/late lifestyle/environmental exposures.
47 POE-mQTL-modifier interaction models were proposed to test the potential of these factors
48 to modify POE at DNA methylation using data from Generation Scotland: The Scottish
49 Family Health Study(N=2315).

50 Results: a set of vulnerable/modifiable POE-CpGs were identified (modifiable-POE-
51 regulated CpGs, N=3). Four factors, "lifetime smoking status" and "predicted mRNA
52 expression levels" of *TET2*, *SIRT1* and *KDM1A*, were found to significantly modify the POE
53 on the three CpGs in both discovery and replication datasets. Importantly, the POE on one of
54 the CpGs were modified by both genetic and environmental factors. We further identified
55 plasma protein and health-related phenotypes associated with the methylation level of one of
56 the identified CpGs.

57 Conclusions: the modifiable POE identified here revealed an important yet indirect path
58 through which genetic background and environmental exposures introduce their effect on
59 DNA methylation, motivating future comprehensive evaluation of the role of these modifiers
60 in complex diseases.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

61 **Keywords**

62 parent-of-origin effect, imprinting, DNA methylation, interaction(modification) effect,

63 mQTL, DNA methylation machinery genes, smoking, lifestyle, environmental effect

64

65 **Key Messages**

66 • Previous population studies showed that parent-origin-effects(POE) on human

67 methylome can be widespread and affect health-related traits and diseases.

68 • Whether the POE remained stable throughout the life or can be modified by genetic or

69 environmental factors were largely unknown.

70 • By systematically screening 101 genetic and environmental factors in a large

71 cohort(GS:SFHS) we provided the first population-level replicated evidence that

72 those measuring lifestyle (smoking) and predicted expression of DNA methylation- or

73 imprinting- machinery genes are amongst the factors that can modulate the POE of

74 mQTLs for a set of CpG sites.

75 • We found those modifiable-POE-regulated CpGs are also phenotypically relevant –

76 one is associated with the plasma levels of CLEC4C and health-related phenotypes

77 such as HDL levels.

78 • The modifiable POE identified here revealed an important yet indirect path through

79 which genetic background and environmental exposures introduce their effect on

80 DNA methylation and their potential phenotypic consequences. This also provided a

81 paradigm for further studies to explore how environmental and genetic effects can be

82 integrated at methylation level.

83

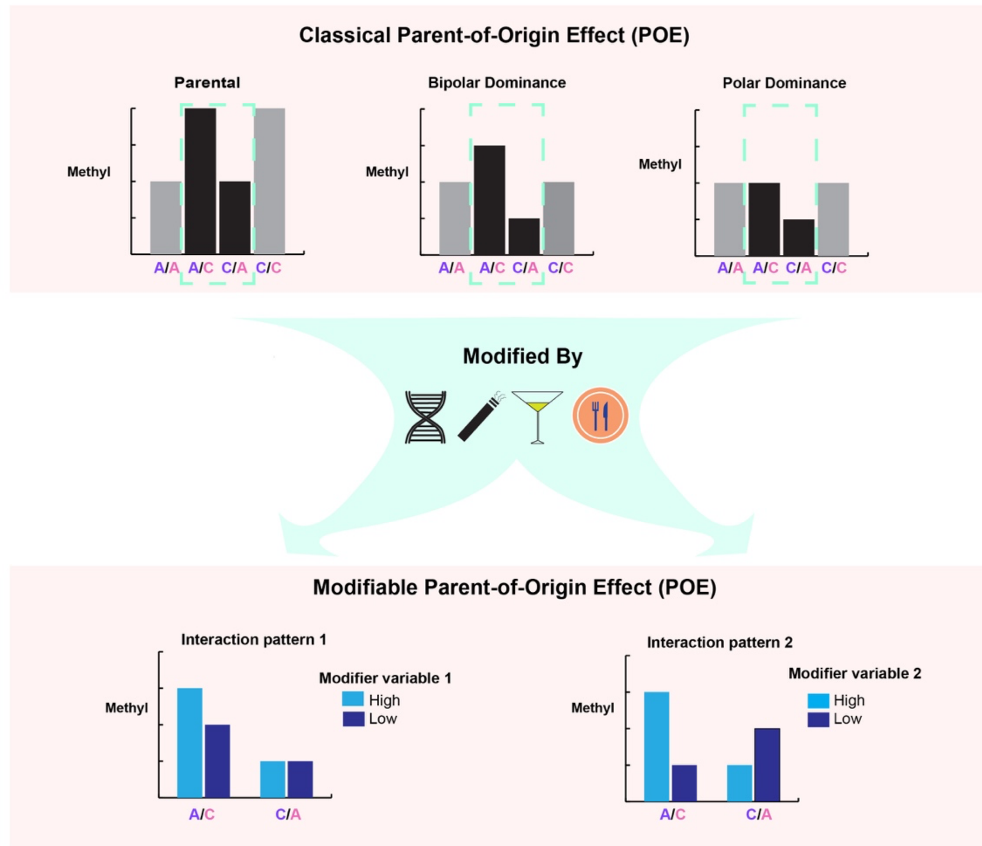
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

84 **Introduction**

85 Illustrating the sources of variation in DNA methylation lays the foundation for
86 understanding epigenetic-based biomarkers for disease risk and progress prediction [1, 2].
87 DNA methylation is known to be influenced by additive and non-additive genetic and
88 environmental factors [3-5]. As a special form of non-additive genetic effects, parent-of-
89 origin effects (POE) on the human methylome manifest as differences in methylation levels
90 between the reciprocal heterozygotes of the mQTL depending on the allelic parent-of-origin
91 (Figure 1)[6]. Through selectively silencing the maternal or paternal allele, genomic
92 imprinting has been considered as the major driving force creating the POE phenomenon [6].
93 We and others have shown that POE-influenced methylation sites are not rare, many are
94 regulated by mQTLs, and that they follow one of the three classical imprinting patterns:
95 parental, bipolar dominance and polar dominance (Figure 1) [4, 7, 8]. Although they only
96 comprise a small proportion of the genome, POE (imprinting)-regulated CpGs and genes
97 have been found to be important for developmental, metabolic and behavioral traits [4, 9].
98
99 Despite their functional importance, the POE patterns in those POE (imprinting)-influenced
100 regions can fluctuate as a consequence of genetic and environmental variation. Previous
101 studies have reported that a large fraction of imprinted regions deviated from mono-allelic
102 expression and that birth phenotypes were associated with the extent of this deviation [10]. A
103 case study on an imprinting influenced long non-coding RNA, lncRNA nc886, found that the
104 imprinting status of this locus is tunable by both genetic variants and environmental factors
105 such as maternal nutrition and maternal age[11]. Importantly, the altered POE has phenotypic
106 consequences: loss of imprinting of nc886 in infants at birth resulted in increased body mass
107 index (BMI) in childhood [11]. For the majority of other POE-influenced regions, however,
108 whether POE are modifiable under certain conditions remains unknown. Here we aim to

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

109 explore modifiable POE, manifesting as the altered methylation difference between
110 reciprocal heterozygotes of the mQTL due to effects from certain genetic or environmental
111 modifiers (Figure 1), which potentially represents an important layer of POE-related
112 regulation requiring systematic examination.
113



114

115 Figure 1 Patterns of classical and modifiable parent-of-origin effect (POE) regulation on DNA methylation. X
116 axis: mQTL genotypes, left purple: paternal allele, right pink: maternal allele. Y axis: methylation level of the
117 regulated CpG. Upper panel: classical POE patterns including parental and complex (dominance) POE patterns.
118 Parental patterns show two levels of methylation depending on the expressed allele and the allelic effect.
119 Complex POE manifests as the two homozygous group having the same methylation level whereas the
120 heterozygous groups are different. Dashed box: difference between methylation level of heterozygous groups of
121 the mQTL is the hallmark of POE. Bottom: scenarios when the POE is modified by genetic or environmental
122 factors, leading to the alteration of POE for different levels of the modifier.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

123

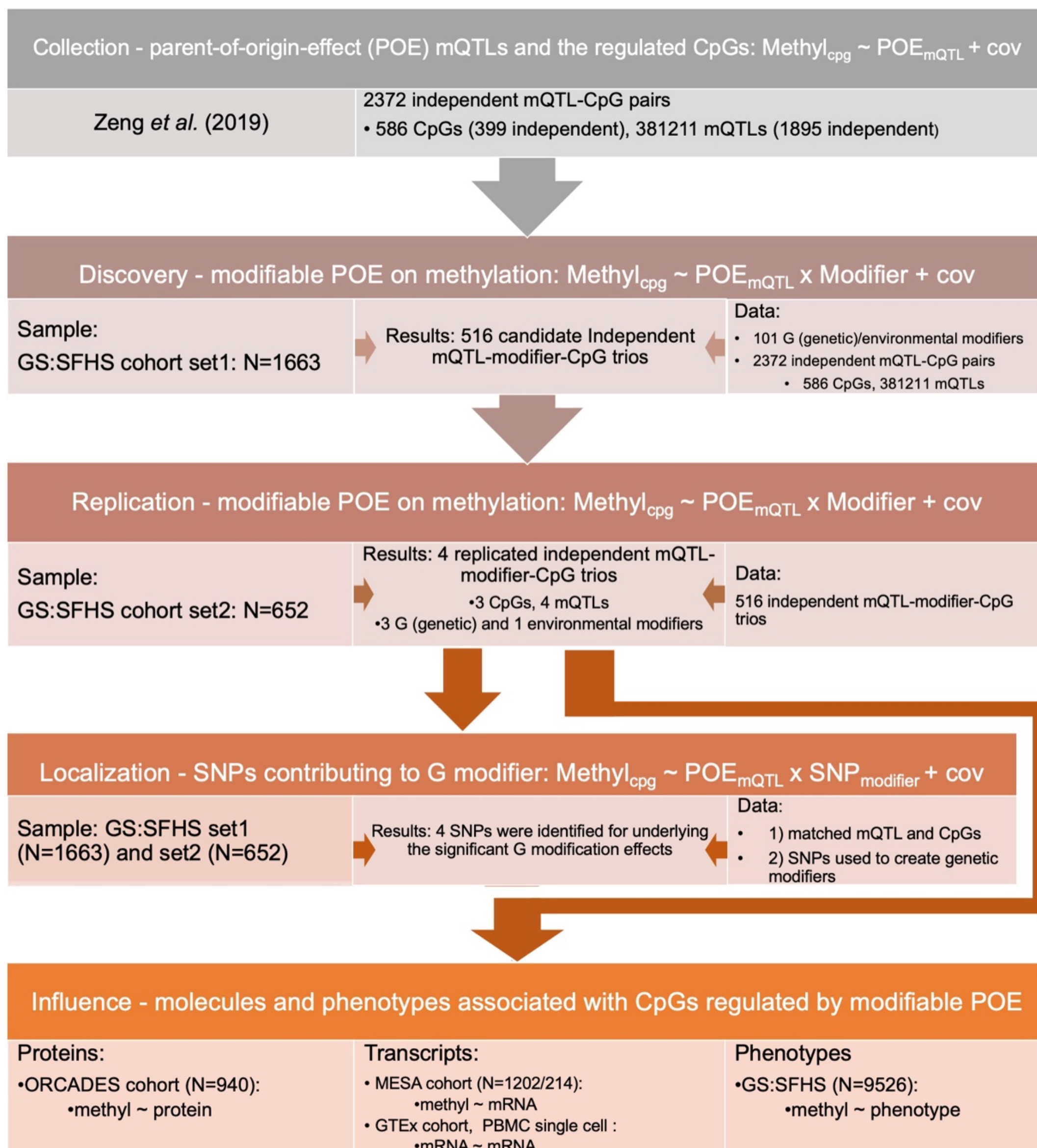
124 To search for modifiable POE on CpGs, a key question is which modifiers may have the
125 potential to regulate the POE. Genomic imprinting, which likely underlies the POE, involves
126 complex and multi-stage DNA methylation reprogramming processes, from the slow erasure
127 of methylation at primordial germ cell stage, to the establishment of imprinted methylation
128 signatures at germ cell stage, followed by pre-implantation maintenance of the imprinted
129 methylation pattern during the global demethylation event, which is subsequently maintained
130 post-implantation [12]. A number of gametic and zygotic genetic factors were found to be
131 involved in these processes, such as those functioning in folate metabolism, the DNA
132 methylation machinery (writers, erasers) and the proteins with which they interact [12].
133 Additionally, imprinting-related processes have also been found to be sensitive to
134 environmental insults, such as the stress induced by assisted reproductive technologies,
135 nutritional deficiency and adverse exposures [12-14]. Given that previous studies of
136 modifiers of genomic imprinting were mostly case studies of individual factors, a systematic
137 and population-wide scanning for genetic and environmental modifiers of POE is essential to
138 fully characterize POE regulation.

139

140 In this study, we used Generation Scotland: The Scottish Family Health Study (GS:SFHS), a
141 large family-based population cohort with extensive environmental and phenotypic records
142 [15, 16], genome-wide genotypes and DNA methylation data [17, 18], to identify genetic and
143 environmental factors that modify the POE on the human methylome (Figure 1). Figure 2
144 illustrates the study design. Based on the 2372 previously identified independent mQTL-
145 CpGs pairs containing mQTLs with parent-of-origin effect (1895 independent mQTLs;
146 381211 SNPs in total) and their regulated CpGs (399 independent CpGs; 586 CpGs in
147 total)[4], we proposed an interaction model which tests for significant interaction effects

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

148 (Figure 1) between each of the 101 candidate environmental/genetic modifiers available in
149 GS:SFHS and the parent-of-origin effect of each mQTL on the corresponding targeted CpG.
150 Significant results from discovery samples (GS:SFHS set1, N=1663) were tested in
151 replication samples (GS:SFHS set2, N=652). Plasma protein levels and health-related
152 phenotypes associated with the modifiable-POE-regulated CpGs were also identified,
153 suggesting phenotypic relevance for this special class of CpGs.
154



155

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

156 Figure 2 Design of the study. Cov: covariates fitted in the model. Zeng *et al.*(2019): the study which reported
157 CpGs regulated by POE and the mQTLs that induce the POE for 586 CpGs (reference 4).
158

159 **Methods**

160 **Population sample**

161 Generation Scotland: the Scottish Family Health Study (GS:SFHS) is a deeply phenotyped
162 population cohort [15, 16] with genome-wide genotypes available for 19994 participants,
163 among which 9526 also have DNA methylation data available[15, 17, 18]. We used
164 GS:SFHS to identify CpGs regulated by modifiable POE, and to explore phenotypes
165 associated with those CpGs.
166

167 ORCADES is a family-based cross-sectional study which recruited 2078 participants
168 between 2005 and 2011 from the Orkney Isles in northern Scotland [19]. Proteomic and DNA
169 methylation data were available for a subset of 940 participants and were used here for
170 association test between methylation sites and plasma protein levels (see below).
171

172 **GS:SFHS cohort: genotypes and inferences of parent-of-origin transmission of alleles in**
173 **offspring**

174 Genome-wide genotypes were generated using the Illumina Human OmniExpressExome -8-
175 v1.0 array [20]. Phasing, imputation and quality control were described previously [4, 21]. In
176 total, 7108491 high-quality imputed common SNPs (MAF $\geq .01$, info score ≥ 0.8) for 19994
177 participants were available for subsequent analyses. Among those individuals, there were
178 7139 offspring with at least one of their parents genotyped in GS:SFHS, which allowed us to
179 successfully infer parent-of-origin allelic inheritance of all imputed common SNPs in 7106
180 offspring with high accuracy [4].
181

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

182 **GS:SFHS cohort: DNA methylation**

183 In GS:SFHS, genome-wide DNA methylation data were produced through a related
184 Stratifying Resilience and Depression Longitudinally (STRADL) project [18]. In 2016, the
185 first wave of methylation data was generated on 5081 participants. These were used as
186 discovery subset. 1663 of these participants also had imputed genotype information with
187 parent-of-origin alleles successfully inferred and were used for the scanning for modifiable
188 POE here. In 2019, another wave of methylation data was generated on an independent
189 subset of 4445 participants. These data were used as replication subset. 652 out of these 4445
190 participants had imputed genotype information with parent-of-origin alleles successfully
191 inferred. Based on a pipeline proposed previously [4], the two datasets were generated,
192 processed and quality controlled in consistent way [22], which was briefly described in text
193 s1.

194

195 **GS:SFHS cohort: Environmental/genetic modification variables**

196 *Environmental modifier variables*

197 The core GS:SFHS cohort has rich collections of environmental variables[15]. Moreover,
198 98% of GS:SFHS participants gave informed consent for data linkage with historic Scottish
199 birth cohorts which contain collections of birth and maternity information(Text s2). In total,
200 we were able to collect 75 environmental variables and used them in downstream analyses. A
201 full list of environmental variables is given in Table s1.

202

203 *Genetic modifier variables*

204 We considered two major sources of genetic modifiers for POE:

205 1) Predicted mRNA expression levels of 17 DNA methylation or imprinting-specific
206 machinery genes imputed by PrediXcan [23].

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

207 2)Nine genetic risk scores for folate metabolism.

208 Details of those genetic modifiers were described in Text s2 and table s1,2.

209

210 **POE-mQTL-modifier interaction models**

211 We applied a POE-mQTL-interaction model to test whether environmental or genetic factors

212 could modify the POE induced by mQTL on CpGs. The model built on a linear regression

213 model that we used to identify POE-specific mQTL-CpG pairs (mQTL with a parent-of-

214 origin effect and the CpG it regulated) in our previous study [4, 24]:

215

216 *model 1 - non-interaction model:*

217 $Methyl_{cpg} = Add_{mQTL} + Dom_{mQTL} + POE_{mQTL} + covariates + error$

218

219 Where for the additive genetic variable (Add_{mQTL}), the dominance genetic variable

220 (Dom_{mQTL}) and the parent-of-origin effect (POE_{mQTL}), genotypes were coded as below:

	AA	Aa	aA	aa
Additive	0	1	1	2
Dominance	0	1	1	0
Parent-of-origin	0	-1	1	0

221

222 In this study, we applied an interaction model that additionally includes a modifier variable,

223 *Mod*, and its interaction with the additive genetic effect and the POE:

224

225 *model2 - interaction model:*

226 $Methyl_{cpg} = Add_{mQTL} + Dom_{mQTL} + POE_{mQTL} + Mod + Add_{mQTL}\ x\ Mod + POE_{mQTL}\ x\ Mod +$

227 $covariates + error$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

228

229 Where the modifier variable (*Mod*) was one of the environmental/genetic variables
230 collected/derived as described in the section "GS:SFHS cohort: Environmental/genetic
231 modification variables". Covariates included age, sex, cell proportions, smoking variables
232 ("pack years" and "lifetime smoking status", which were not included as covariates when
233 they were the tested modifier factor) and principal components (PCs) derived from an
234 OMIC-relationship-matrix (ORM) created in OSCA [25] using all measured DNA
235 methylation sites. To avoid removing genetic signals of interest by fitting ORM-PCs, we only
236 included ORM-PCs among the top 20 ORM-PCs that were not significantly associated with
237 any common SNP(determined through performing GWAS for ORM-PCs). We applied this
238 model and tested the significance of the interaction effect between each mQTL's POE and
239 the modifier variable (POE_{mQTL} *x Mod*) on the corresponding CpG. The mQTL-CpG pairs
240 tested were the 2372 independent POE-specific mQTL-CpG pairs which we reported
241 previously [4]. The interaction between the additive effect and the modifier (*Add x Mod*) was
242 jointly fitted in the model. For simplicity we did not fit an interaction between the dominance
243 effect and the modifier (*Dom x Mod*) in the global scan, but we did this in sensitivity tests for
244 the significant trios we identified. The results indicated only a minor contribution of the
245 Dom_{mQTL} *x Mod* effect (Table s3). Multiple testing correction was performed by a
246 combination of a global permutation test and mQTL-modifier-CpG trio-specific permutation
247 tests for POE_{mQTL} *x Modifier* interaction effect at discovery stage(FDR≤0.05, details see Text
248 s3, Table s4). A successful replication required to both reach statistical significance
249 (FDR≤0.05) and patten consistency (details in Text s3, Table s4-6). Visualization of the
250 results was performed using the R package coMET and ggplot2 [26, 27].

251

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

252 **Identification of proteins and mRNAs associated with CpGs regulated by modifiable**
253 **POE**

254 *Association between DNA methylation and protein levels*

255 ORCADES cohort- DNA methylation and proteomic data:
256 DNA methylation was measured from whole blood samples using Illumina MethylationEPIC
257 Array for 794627 CpG sites in 1052 samples (quality control and pre-correction in Text s4).
258 Abundance of plasma proteins was measured from the fasted EDTA plasma samples for a
259 subset of 1051 participants using Olink Proseek Multiplex cardiometabolic, cell regulation,
260 cardiovascular 2 and 3, developmental, immune response, inflammation 1, metabolism, neuro
261 exploratory, neurology, oncology and organ damage panels. The raw NPX values were used
262 in analysis.
263
264 Association test between DNA methylation and plasma protein levels:
265 A linear mixed model was used to compute methylation residuals after accounting for genetic
266 structure in ORCADES by fitting a random effect represented in the genomic relationship
267 matrix. To assess whether the CpGs (N_{cpg}=3) significantly regulated by interaction effects
268 (POE_{mQTL} *x Mod*) were associated with the abundance of any plasma protein ($N_{protein}$=1092),
269 the association between adjusted methylation of CpGs of interest with each measured protein
270 was tested using a linear regression model in 940 participants where DNA methylation and
271 proteomic data were simultaneously available:
272
273 ***Methyl = protein + cell proportion + age + sex + season of sampling + smoking status +***
274 ***error***
275
276 The Bonferroni method was applied to correct for multiple testing ($N_{correction}$=3x1092=3276).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

277

278 *Correlations between DNA methylation vs mRNA levels:*

279 Transcriptomic and DNA methylation data from human peripheral monocytes and T cells in
280 the Multi-Ethnic Study of Atherosclerosis (MESA) study were downloaded from the NCBI
281 GEO database (Series GSE56047 and GSE56580) [28]. mRNA was measured using the
282 Illumina HumanHT-12 v4 Expression BeadChip, DNA methylation levels were measured
283 using the Illumina HumanMethylation450 BeadChip [28]. Quantile-normalized signal for
284 mRNA (log2 transformed) and DNA methylation data (M-values) were simultaneously
285 available for peripheral monocytes (CD14+) in 1202 participants and for peripheral T cells
286 (CD4+) in 214 participants and were used to calculate the Spearman correlation between
287 DNA methylation and mRNA.

288

289 *Correlations between mRNAs levels:*

290 At the population-level, correlations between mRNA levels of target genes were calculated
291 using the Spearman method on GTEx whole-blood data using the GEPIA portal [29]. At
292 single cell level, a normalized single-cell matrix for 63628 peripheral blood mononuclear cell
293 (PBMC) cells from a healthy donor were obtained from the website http://tisch.comp-
294 genomics.org/gallery/. Feature counts for each cell were normalized by "LogNormalize", a
295 global-scaling method that normalizes the cellular feature expression by dividing the total
296 counts for that cell, multiplied that by a scale factor (10000 by default), followed by a
297 natural-log transformation. Spearman's correlation between mRNA levels of target genes was
298 calculated using cells where normalized expression levels of both genes were larger than
299 zero.

300

301 **Phenome-wide association test for DNA methylation sites regulated by modifiable POE**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

302 We collected 79 phenotypes measured in GS:SFHS (recorded dataset and linked data) to
303 identify phenotypes correlated with DNA methylation levels at CpG sites targeted by
304 modifiable POE. The full list of phenotypes tested can be found in Table s7. The correlation
305 was tested by regressing the adjusted M-values of methylation sites on each phenotype
306 variable, covariates included age, age^2, cell proportion, sex, top 20 ORM-PCs and smoking
307 variables ("pack years" and "lifetime smoking status". These were not included as covariates
308 when they were the tested phenotype). The test was performed in the discovery and
309 replication samples separately and meta-analyzed using a random effect model using the R
310 package metafor [30]. The sample size varied depending on the number of missing samples
311 for each specific phenotype (Table s8). The Bonferroni method was used for multiple testing
312 correction ($N_{correction}$=79*3=237).
313
314 **Results**
315 **DNA methylation sites targeted by modifiable POE**
316 In GS:SFHS, we utilized information on 75 environmental and 26 genetic variables to test if
317 any of them significantly modified the parent-of-origin effects of the mQTLs on methylation
318 level of the targeted CpGs, altering the methylation difference between reciprocal
319 heterozygotes of the mQTLs (Figure 1). The environmental modifiers reflected the
320 environments/events the participants have been exposed to or have experienced, including
321 those measuring baseline non-genetic effects (sex, age), medication, lifestyle, socioeconomic
322 status, birth-related phenotypes (measured before or after participants' birth) and
323 menarche/menopause-related events; the genetic modifiers were genetic factors known to be
324 involved in DNA methylation and imprinting processes, including "predicted whole blood
325 mRNA expression levels" of DNA methylation- or imprinting-specific machinery genes, and
326 polygenic risk scores (PRSs) for folate-related metabolism (details in methods and Table s1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

327

328 In the discovery stage, a global permutation test was combined with mQTL-modifier-CpG
329 trio-specific permutation tests to determine mQTL-modifier-CpG trios for significant
330 POE_{mQTL} *x Modifier* interaction effect on DNA methylation(methods). In the replication
331 stage, a successful replication required both statistical significance and pattern
332 consistency(direction-of-effect) for both the main POE and the POE_{mQTL} *x Modifier*
333 interaction effect in tested trios. In total, four mQTL-modifier-CpG trios reached significance
334 in the discovery sample ($P_{global_permutation(FDR_adjusted)} < 0.05$, $P_{trio_permutation(Bonferroni_adjusted)} <$
335 0.05) and were successfully validated in the replication sample (Table 1, text s2, tables s4-6).
336 These four trios involved one environmental modifier: “lifetime smoking status”; and three
337 genetic modifiers: “predicted mRNA expression levels” of *SIRT1* (a gene that protects
338 methylation at imprinted loci by directly regulating acetylation of DNMT3L), *TET2* (a DNA
339 demethylation gene) and *KDM1A* (a gene involved in removal of methylation and histone
340 H3K4 in imprinted genes). Three CpGs, cg18035618, cg21252175, and cg22592140 were the
341 methylation sites affected by these modification effects (Table 1), with the POE_{mQTL} *x*
342 *Modifier* interaction effect explaining between 1.0% to 2.2% of their methylation variation
343 (Figure 3a, Table s3). These CpGs displayed intermediate methylation levels, when
344 compared with the less-peaked distribution of methylation levels of POE-regulated CpGs not
345 influenced by modifiable effects, or the bi-directionally distributed methylation level of
346 genomic CpGs not influenced by POE (Figure 3b). One of the three CpGs, cg18035618, was
347 simultaneously targeted by both environmental and genetic modification effects (Table 1).
348

349 Table 1 Genetic/environmental modifiers displaying a significant interaction with the POE(mQTL) and the CpGs the interaction effects affect.

350

mQTL-CpG-modifer trio information								Discovery					Replication			
CpG: id	CpG-chr	CpG-POS	mQTL: id	mQTL-chr	mQTL-POS	Modifier	Modifier type	P(dis)	Est(dis)	N(dis)	P (trio-specific -perm)	Adjusted P (trio-specific-perm)	P(rep)	Est(rep)	N(rep)	Adjusted P(rep)
cg22592140	7	130132419	rs142053217	7	130258926	mRNA expression of KDM1A	Genetic	1.35E-06	6.35	1607	1.00E-05	1.09E-02	6.09E-04	6.52	645	2.79E-02
cg21252175	15	64973903	rs111834387	15	64062211	mRNA expression of TET2	Genetic	1.54E-06	3.77	1607	<1e-5	< 0.01	7.76E-11	13.80	647	2.43E-08
cg18035618	20	57415978	rs117020124	20	57347435	Lifetime smoking status	Env	9.94E-07	-0.04	1607	<1e-5	< 0.01	1.42E-03	-0.05	647	4.85E-02
			rs117823351	20	57369773	mRNA expression of SIRT1	Genetic	3.68E-10	-0.24	1607	<1e-5	< 0.01	1.01E-03	-0.22	646	3.95E-02

351

352 CpG-POS/mQTL-POS: genomic position of the CpG and mQTL(hg19). Env: environmental modifier. dis: discovery sample, rep: replication sample. N: sample size.

353 P(dis)/P(rep) and Est(dis)/Est(rep): P values and coefficients for the POE_{mQTL} *x Modifier* effect in the interaction model in discovery/replication samples. P(trio-specific-

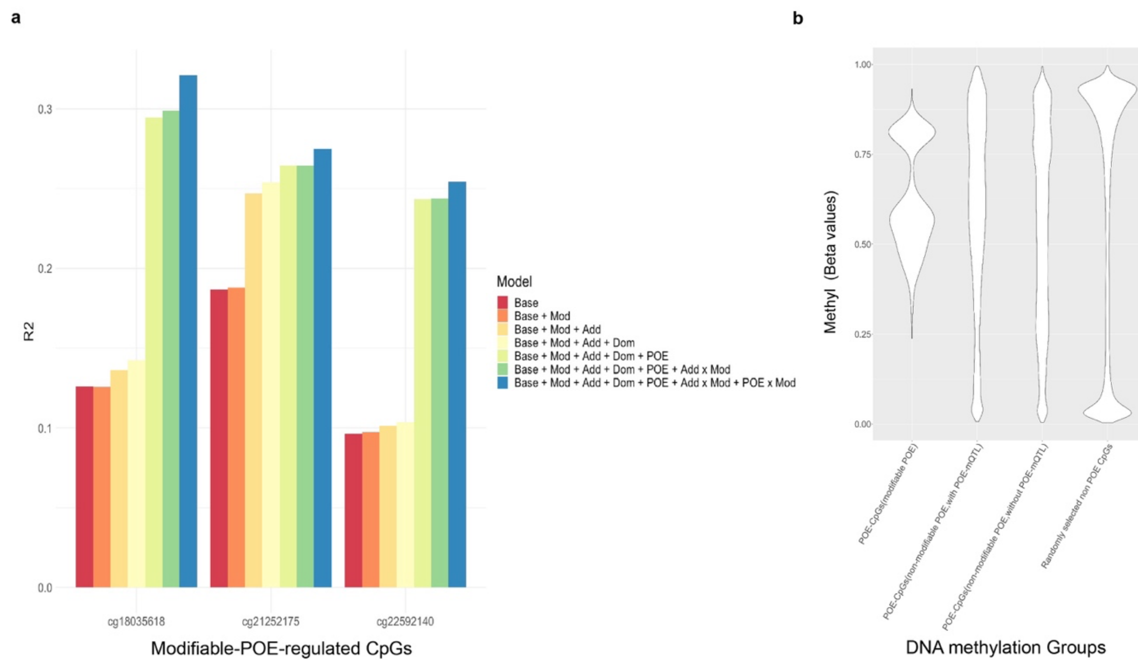
354 perm): P value for the POE_{mQTL} *x Modifier* effect from the trio-specific permutation test. Genomic position of the significant genetic modifiers: *KDM1A*: chr1: 23345941-

355 23410182; *TET2*: chr4: 106067,032-106200973; *SIRT1*: chr10: 69644427-69678147.

356

357

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



358

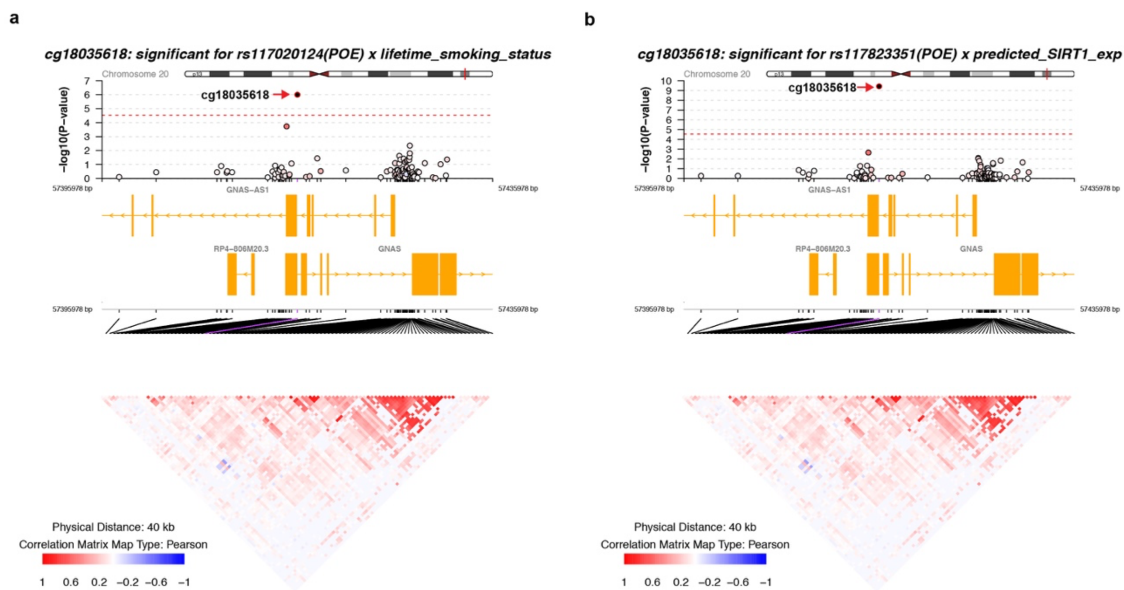
359 Figure 3. CpGs regulated by the modifiable-POE. a. Proportion of methylation variation explained by different
360 models for the three CpGs regulated by modifiable-POE. The "Base model" accounted for age, sex, cell
361 proportion, smoking variables ("pack years" and "lifetime smoking status". These were not included as
362 covariates when they were the tested modifier factor) and principal components of the OMIC (DNA
363 methylation)-relationship-matrix. Mod: modifier; Add: additive genetic effect, Dom: dominance genetic effect.
364 Add x Mod: interaction between additive genetic effect and the modifier; POE x Mod: interaction between
365 parent-of-origin genetic effect and the modifier. b. Distributions of methylation levels of CpGs regulated by
366 modifiable POE (N_{CpG}=3), CpGs regulated by POE from known mQTLs but the POE is not modifiable
367 (N_{CpG}=583), CpGs regulated by POE but without an mQTL identified (the CpGs were reported in ref 11: Zeng
368 *et al.* (2019), N_{CpG}=398) and randomly selected non-POE CpGs from all QCed array probes (N_{CpG}=10,000).
369 Unrelated individuals from the GS:SFHS discovery subset were used to produce the plot.

370

371 cg18035618 (hg19: chromosome 20: 57415978) is located in an intron of the gene *GNAS*
372 (Figure 4). The methylation level of cg18035618 was significantly modulated by a
373 POE_{mQTL} *x Modifier* interaction between its mQTL, rs117020124, and "lifetime smoking
374 status" (P_{dis}=9.94x10^{-7}, P_{rep}=1.42x10^{-3}, figure 4a, table 1), with current smokers displaying
375 larger contrast in methylation levels of cg18035618 between heterozygous groups of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

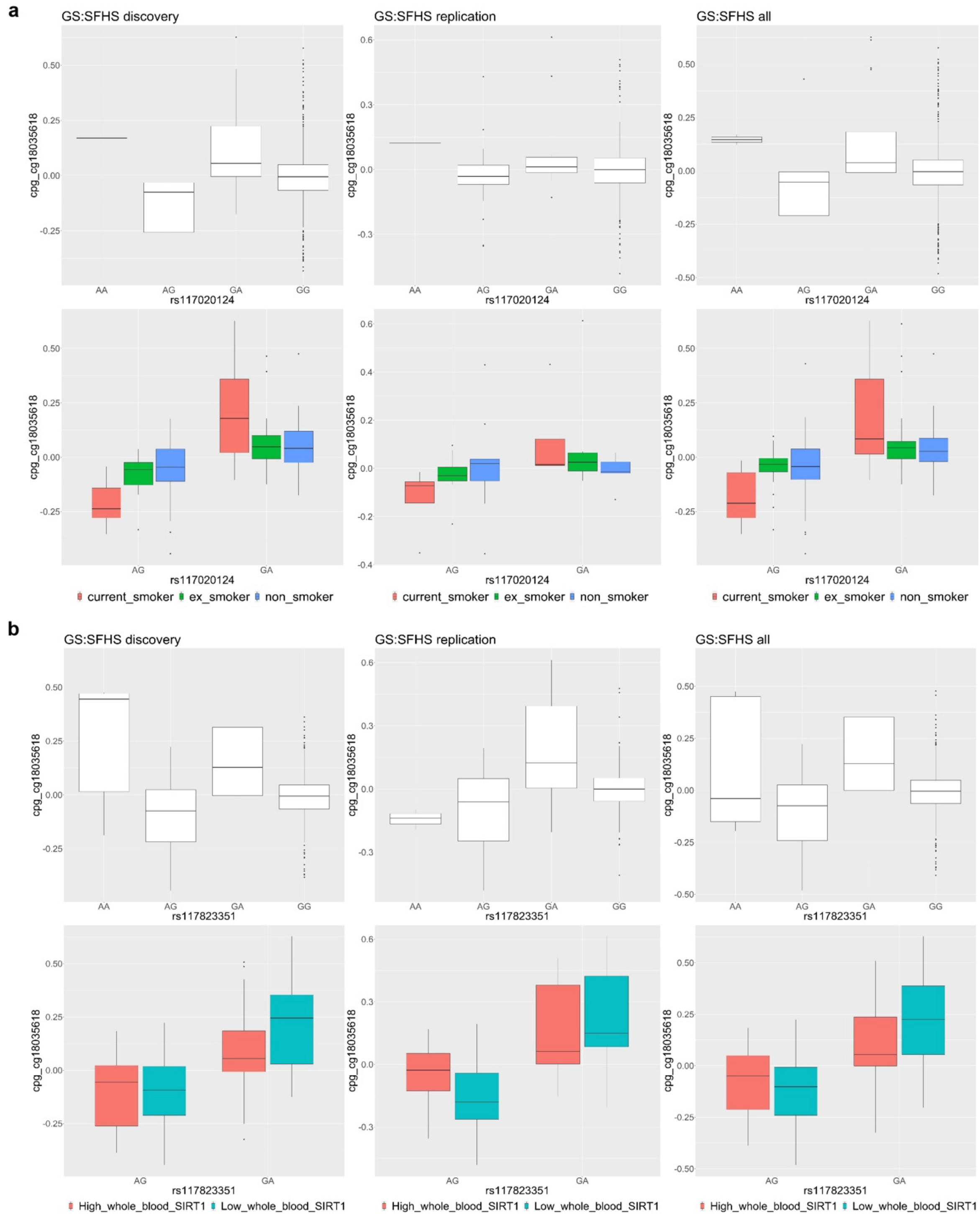
376 mQTL when compared with ex-smokers and never-smokers (Figure 5a). For this CpG a
377 significant POE_{mQTL} *x Modifier* interaction was also detected between another independent
378 mQTL, rs117823351, and "predicted mRNA expression levels of *SIRT1*" (P_{dis}=3.68x10^{-10},
379 P_{rep}=1.01x10^{-3}, figure 4b, table 1. *SIRT1* is located in chromosome 10), with lower *SIRT1*
380 expression corresponding to an increased contrast of methylation levels between
381 heterozygous groups of the mQTL (Figure 5b).
382



383

384 Figure 4 Regional plot of the modifiable-POE affecting cg18035618 and nearby CpGs within a 20kb distance. a.
385 Interaction effect between the POE of cg18035618's mQTL rs117020124 and "lifetime smoking status"; b.
386 Interaction effect between the POE of cg18035618's mQTL rs117823351 and 'predicted mRNA expression
387 levels of *SIRT1*'. Upper panel: –log10 (P-value): minus log10 P-value of the POE_{mQTL} *x Modifier* interaction
388 effect. Dots show nearby measured CpGs located within a 20kb distance from cg18035618, filling colour
389 represents the correlation of methylation levels with cg18035618: red: positive correlation ; blue:negative
390 correlation; white: no significant correlation. . Middle panel: genes located within the 40kb genomic region.
391 Bottom panel: correlation matrix between CpGs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



392

393 Figure 5. Both environmental and genetic factors significantly modified the POE of mQTLs on cg18036618.

394 a.top:cg18036618 was regulated by the POE of the mQTL rs117020124. bottom: the POE of rs117020124 was

395 modified by lifetime smoking status. The contrast in methylation levels of cg18035618 between heterozygotes

396 of the mQTL rs117020124 is largest in current smoker group. b.top:cg18036618 was regulated by the POE of

397 the mQTL rs117823351. bottom: the POE of rs117823351 was modified by “predicted mRNA expression level

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

398 of *SIRT1*". The contrast in methylation levels of cg18035618 between heterozygotes of the mQTL rs117823351
399 is larger in individuals with low SIRT1 expression.
400
401 To rule out the possibility that the sharing of the methylation target (cg18035618) by both
402 genetic and environmental modifiers was due to genetically influenced environmental effects
403 (i.e., *SIRT1* expression influences smoking status) [31], we calculated the correlation between
404 "lifetime smoking status" and "predicted mRNA expression levels of *SIRT1*" and found no
405 significant correlation (R=0, P=0.968).
406
407 cg21252175 (hg19: chromosome15: 64973903) is located in the 3' UTR region of gene
408 *ZNF609* (Figure s1a). For this CpG, a significant POE_{mQTL} *x Modifier* interaction was
409 detected between the mQTL rs111834387 and "predicted mRNA expression levels of *TET2*"
410 (P_{dis}=1.54x10^{-6}, P_{rep}=7.76x10^{-11}. *TET2* is located in chromosome 4), with lower expression
411 of *TET2* reducing the contrast in methylation levels of cg21252175 between heterozygous
412 groups of the mQTL (Table 1, figure s2a). cg22592140, a CpG located in an intron of gene
413 *MEST* (Figure s1b), was significantly regulated by a POE_{mQTL} *x Modifier* interaction between
414 its mQTL rs142053217 and "predicted mRNA expression levels of *KDM1A*" (P_{dis}=1.35x10^{-6},
415 P_{rep}=6.09x10^{-4}. *KDM1A* is located in chromosome 1), with lower expression of *KDM1A*
416 increasing the contrast in methylation levels of cg22592140 between heterozygous groups of
417 the mQTL (Table 1, figure s2b).
418
419 **Localization of regulatory SNPs contributing to the genetic modification effect.**
420 Considering that the three identified genetic modifiers ("predicted mRNA expression levels"
421 of *SIRT1, TET2, KDM1A*) are essentially weighted combinations of allelic scores at multiple
422 regulatory SNPs, we next tested whether the genetically driven modification effects we
423 detected here can be recapitulated by the POE_{mQTL} *x SNP* interactions between mQTLs and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

424 the SNPs used to drive the significant genetic modifiers. Since "predicted mRNA expression"
425 of *SIRT1*, *KDM1A* and *TET2* was derived from two, one and two SNPs respectively by the
426 MASHR method in PrediXcan(Methods)[32], we tested POE_{mQTL} *x SNP* interactions for
427 those 5 SNPs (N_{test}=5). We identified four of the five SNPs significantly interacting with the
428 POE_{mQTL}, that is, regulating the CpGs where the interaction effect was initially detected
429 (Table 2, Figure 6). For example, for cg18035618 we detected a significant interaction effect
430 (P_{dis}=3.69x10^{-9} , P_{rep}=4.33x10^{-3}) between rs932658, a SNP used in the prediction model for
431 *SIRT1* expression, and rs117823351, the mQTL that significantly interacted with "predicted
432 mRNA expression levels of *SIRT1*". When accounting for these significant POE_{mQTL} *x SNP*
433 interaction effects as conditional items in the interaction model for the three genetic-
434 modifiers, the interaction effect at the genetic-modifier-level (POE_{mQTL} *x* $Modifier_{Genetic}$) were
435 reduced to a non-significant level, suggesting the leading role of POE_{mQTL} *x SNP* underlying
436 the significant interaction effect from genetic modifiers we detected here (Table s10).
437

438 Table 2 SNPs significantly interact with POE (POE_{mQTL} x SNP) and the regulated CpGs

439

mQTL-CpG		SNP modifier		Discovery				Replication		
CpG: id	mQTL: id	SNP id	Relation to G modifier	P(dis)	Adjust_P(dis)	Est(dis)	N(dis)	P(rep)	Est(rep)	N(rep)
cg22592140	rs142053217	rs75667410_T	SNP used in predict *KDM1A* exp	1.35E-06	6.75E-06	0.18	1607	6.09E-04	0.18	645
cg21252175	rs111834387	rs11729069_G	SNP used in predicting *TET2* exp	1.05E-08	5.25E-08	0.18	1607	Not enough data*		
		rs7661349_T		3.51E-02	1.76E-01	-0.06	1607	Not significant in Discovery sample		
cg18035618	rs117823351	rs932658_A	SNP used in predict *SIRT1* exp	3.69E-09	1.85E-08	-0.07	1607	4.33E-03	-0.06	646
		rs3740053_G		1.55E-04	7.75E-04	-1.00E-01	1607	6.01E-03	-0.10	646

440

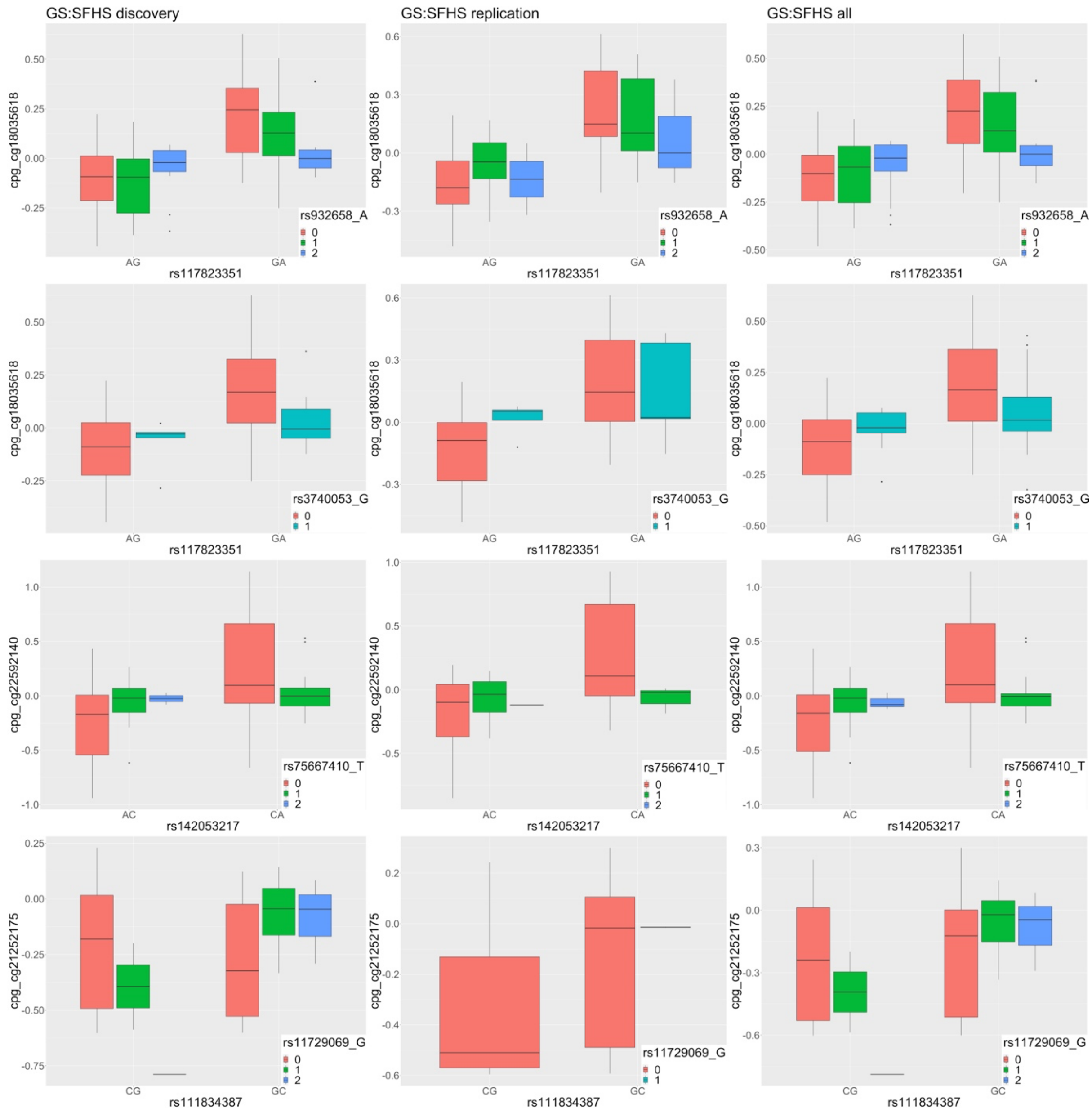
441 dis: discovery sample, rep: replication sample. N: sample size. P(dis)/P(rep) and Est(dis)/Est(rep): P values and coefficients for the POE_{mQTL} x SNP interaction in the

442 interaction model in discovery/replication samples. *, due to the relatively small minor allele frequency (MAF) of rs111834387(MAF=0.01) and the limited sample size of

443 replication sample, there was not enough data for this test in replication samples. Number of individuals within heterozygous groups available for testing *POE(rs111834387)*

444 *x SNP(rs11729069)* interaction effect on cg21252175 in replication samples were shown in Table s9.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



445

446 Figure 6. The three modifiable-POE-targeted CpGs were also significantly regulated by POE_{mQTL} *x SNP*

447 interaction effects between the mQTLs and the SNPs used to drive the genetic modifiers. The contrast in

448 methylation levels of the candidate CpGs in mQTL heterozygotes varied depending on the allelic dosage

449 of the SNP used to derive the corresponding genetic modifier. *due to the limitation of sample size, minor

450 homozygous/heterozygous genotype groups were missing in some tests.

451

452 **Proteins associated with modifiable-POE-regulated methylation sites**

453 Association tests were performed using DNA methylation and proteomic data from

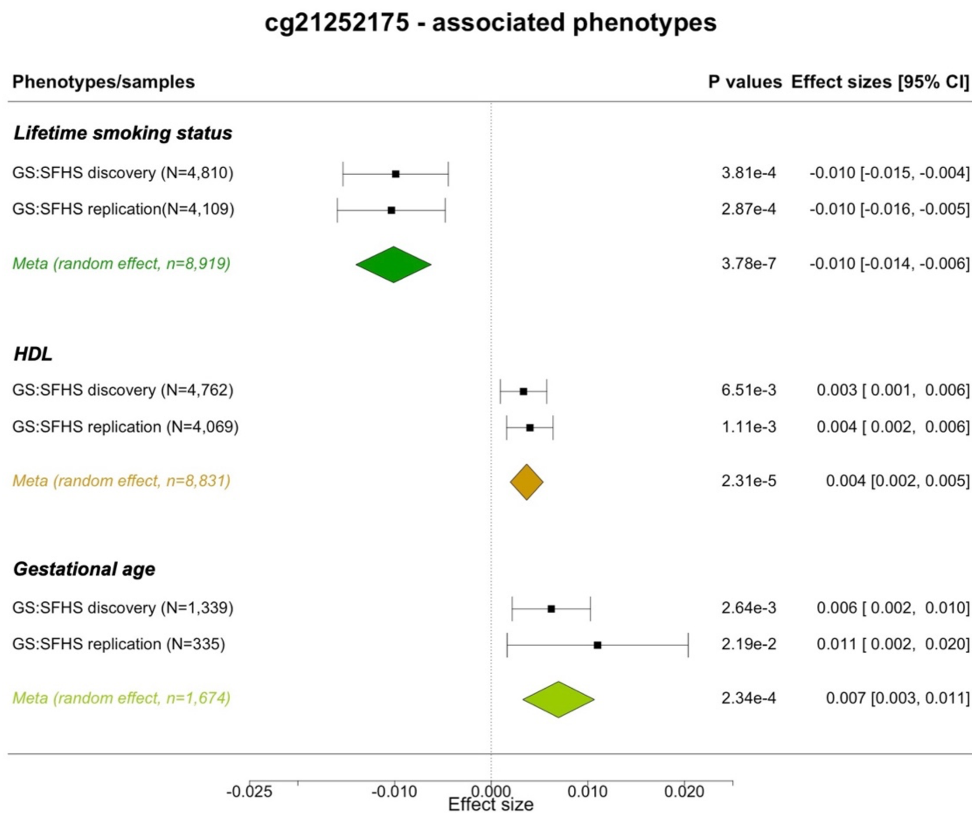
454 ORCADES cohort (N_{sample}=940, $N_{total_protein}$=1092) to identify proteins associated with the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

455 three CpGs regulated by the modifiable POE. Only one CpG-protein pair passed Bonferroni-
456 based multiple testing correction (N_{test}=3,276, full results in Table s11): cg21252175 was
457 significantly associated (in *trans*) with plasma protein level of CLEC4C, a protein from the
458 Olink immune-response panel (Beta=0.049, $P_{adj}=0.002$).
459
460 Since cg21252175 is located in the UTR3 of gene *ZNF609*, we further examined whether the
461 association between cg21252175 and CLEC4C protein levels implied a link between *ZNF609*
462 and *CLEC4C*. Using data from MESA study [28], a significant and positive correlation was
463 detected between methylation levels of cg2125217 and mRNA levels of *ZNF609* in both
464 CD4+ peripheral T cells (R=0.15, *P*=0.03) and CD14+ peripheral monocytes (R=0.15,
465 $P=1.52\times10^{-6}$, figure s3a). Using whole blood data from the GTEx consortium through the
466 GEPIA portal[29, 33], mRNA expression level of *ZNF609* was significantly correlated with
467 mRNA expression levels of *CLEC4C* at population level (R=0.21, $P=1.1\times10^{-4}$, figure s3b).
468 Using a single-cell RNA-seq data of PBMC in an adult donor, mRNA expression levels of
469 *ZNF609* and *CLEC4C* were significantly correlated at the cellular level (R=0.36, *P*=0.0002,
470 figure s3c).
471
472 **Phenotypes associated with modifiable-POE-regulated methylation sites**
473 To explore the association between variation in CpGs targeted by modifiable POE and
474 heath/disease-related phenotypes, we collected 79 phenotypes in GS:SFHS (Table s7).
475 Phenome-wide association tests relating methylation levels to phenotypes were performed for
476 the three identified modifiable-POE-regulated CpGs using the whole GS:SFHS methylation
477 sample (meta-analyzed across discovery (N_{sample}=5081) and replication (N_{sample}=4445)
478 samples; N_{sample}=9526.). After Bonferroni-based multiple testing correction
479 (N_{test}=79x3=237), two CpG-phenotype associations reached phenome-wide significance:

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

480 cg21252175 was both associated with "lifetime smoking status" (P_{adj}= 9.0x10^{-5}) and high-
481 density lipoprotein (HDL) levels (P_{adj} =0.006) (Figure 7, Table s8). Although limited by
482 sample size(Figure 7), cg21252175 was also associated with gestational age (measured as
483 weeks at birth) at an adjusted P ≤ 0.06 level (P_{adj}=0.056). These associations displayed
484 consistent patterns across discovery and replication samples (Figure 7, Table s8).
485



486

487 Figure 7. Forest plot for phenotypes associated with cg21252175. HDL: high-density lipoprotein.
488 Meta: Meta-analysis performed using the random effect model.
489

490 **Discussion**

491 In this study, we reported significant and replicated modification effects from both genetic
492 and environmental variables on the parent-of-origin effect that affects DNA methylation
493 levels at three CpGs. Identified environmental modifiers included "lifetime smoking status";

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

494 identified genetic modifiers included "predicted mRNA expression levels" of several DNA
495 methylation/imprinting machinery genes (*SIRT1, TET2, KDM1A)*. Importantly, we found that
496 both genetic and environmental modifiers were targeting the same CpG (cg18035618). These
497 provided evidence for a special type of CpGs in the human genome regulated by the parent-
498 of-origin-effects that are modulated by genetic or environmental modifiers. We further found
499 that these CpGs are likely to be phenotypically relevant: at the molecular level, DNA
500 methylation level at one the CpG, cg21252175, was associated with protein levels of the
501 immune-response-related protein CLEC4C. At the phenotypic level, this CpG was associated
502 with "lifetime smoking status", HDL levels and gestational age(the latter at the suggestive
503 significance level).
504
505 Statistically, the model we proposed here for detecting modifiable POE is built on a previous
506 POE-mQTL model which we used to localize regulatory mQTLs for the POE-influenced
507 CpGs (model 1 in Methods) [4, 24]. For those CpGs, one of the hallmarks of the POE-
508 regulation was the methylation difference between the two heterozygous mQTL genotype
509 groups (Figure 1)[4, 6]. Here, our new interaction model tests whether that difference
510 remains stable or varies under certain conditions, that is, if the parent-of-origin effects are the
511 same or different across different environments or in different genetic backgrounds.
512 Biologically, this implies the existence of a new and different layer of regulation for DNA
513 methylation: genetic/environmental factors could influence the level of DNA methylation of
514 CpGs, not only through direct effects, but also through interacting with the POE (Figure 1).
515
516 Our results support this hypothesis. Smoking has been widely studied for its direct influence
517 on DNA methylation [34, 35] and its interactions with additive genetic effects on methylation
518 levels [36]. Here, for the first time, our study reported that smoking could also affect DNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

519 methylation variation indirectly through interaction with a non-additive genetic effect, POE.
520 Similarly, variation in DNA methylation and imprinting machinery genes, either in the forms
521 of variable expression or mutation, have been known to directly affect DNA methylation.
522 SIRT1 regulates DNA methylation and protects methylation at imprinted loci by directly
523 regulating acetylation of DNMT3L [12, 37]. TET2 promotes DNA demethylation by
524 converting 5-methyl-cytosine to 5-hydroxymethyl-cytosine and is required for demethylation
525 at imprinted loci in the germline [12, 38]. KDM1A removes methylation of histone H3K4 in
526 imprinted genes, its deficiency is associated with alterations in DNA methylation and
527 expression at imprinted genes [12, 39]. Here, for the first time, we showed that besides
528 known direct effects, these genetic factors also introduce indirect regulatory effects on DNA
529 methylation levels through interactions with POE.
530
531 The detection of interactions between genetic modifiers and POE led us to further identify
532 significant and replicated POE_{mQTL} *x SNP* interaction effects between mQTLs and SNPs used
533 in imputing genetic modifiers. One important feature of the genetic modifier variables we
534 derived here is that they represent the proportion of mRNA variation determined by germline
535 genetic variation, which is constant and stable throughout the life [40]. Our results
536 demonstrated that an individual's genetic background of DNA methylation and imprinting
537 machineries has the potential to modify POE. The localization of the genetic-based
538 modification effect at regulatory SNPs of these DNA methylation and imprinting machinery
539 genes strongly supports this, and importantly, indicated that genetic variation in machinery
540 genes is an important source of epistasis. One of the machinery genes, *SIRT1*, has been well
541 known for its role in mental disorders such as depression[41], but very few studies have
542 examined its role as a modifier for non additive genetic effect such as POE. Our result
543 revealed a new potential path of variation in this gene to introduce molecular differences.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

544

545 The reason why POE can ever be modified is deeply rooted in the unique nature of genomic
546 imprinting: established at an early developmental stage, needing to be protected from global-
547 demethylation and maintained throughout the lifespan [12]. These complex and multi-stage
548 processes have been shown to be vulnerable to environmental fluctuations and involve
549 delicate regulation of multiple gametic and zygotic genetic factors, including *TET2*, *SIRT1*,
550 *KDM1A* as we identified here.[11, 12]. Indeed, the vulnerability of at least a subset of POE-
551 CpGs was revealed here, as the POE_{mQTL} *x Modifier* interaction effect explained a non-
552 negligible proportion of methylation variance (1%-2.2%), and that at least one (cg18035618)
553 of the three CpGs identified was targeted by independent environmental and genetic
554 modifiers. Molecularly, we found one of the modifiable-POE influenced CpG (cg21252175)
555 was associated with CLEC4C, an immune-response transmembrane protein treated as a
556 marker gene for plasmacytoid dendritic cells [42]. Phenotypically, cg21252175 was
557 associated with "lifetime smoking status", HDL levels and gestational age (the latter at the
558 suggestive significance level) (Figure 7) in our analysis, and was previously found to be
559 associated with maternal body mass index/overweight/obesity [43]. These convergent lines
560 indicated that this newly uncovered class of vulnerable POE-CpGs may play an important
561 role in connecting early life stress, variations in genetic background and later life health
562 issues.

563

564 There are limitations in this study. First, mRNA expression levels used here were predicted
565 and only reflect genetically influenced expression variation (and not necessarily all of it).
566 Future studies examining measured mRNA expression will be necessary to account for the
567 modification effect from the environmentally determined fraction of mRNA expression.
568 Second, since the analyses performed here simultaneously require DNA methylation data and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

569 SNP data with information of parent-of-origin of the alleles transmitted to the offspring, the
570 sample size was relatively limited in this study. This is a proof of principle study and larger
571 sample sizes will be needed to uncover more modifiable POE. Finally, the exact time window
572 or developmental stage in which each environmental modifier exerted their influences
573 remains unknown. Cohort data with higher resolution of environmental exposure records, in
574 particular those measuring early life exposures, will be crucial to understand the vulnerable
575 stage or stages for CpGs influenced by modifiable POE.
576

577 **Conclusions**

578 we provided the first population-level evidence for modification effects from multiple
579 genetic and environmental factors on parent-of-origin-effects at the DNA methylation level.
580 A subset of parent-of-origin-effect-influenced CpGs that are vulnerable to modification
581 effects were uncovered, which opens new questions for future profiling of the modification
582 patterns and phenotypic consequences of this class of CpGs.
583

584 **Funding**

585 YZ was supported by the General Program of National Natural Science Foundation of China
586 (81971270) and Sun Yat-sen University Young Teacher Key Cultivate Project. The work of
587 AF was supported by European Union's Horizon 2020 research and innovation programme
588 IMforFUTURE under H2020-MSCA-ITN grant agreement number 721815. ADB would like
589 to acknowledge funding from the Wellcome PhD training fellowship for clinicians
590 (204979/Z/16/Z), the Edinburgh Clinical Academic Track (ECAT) programme. The authors
591 want to acknowledge support from the MRC Human Genetics Unit programme grant,
592 "Quantitative traits in health and disease" (U. MC_UU_00007/10), and grant
593 MC_PC_U127592696. Generation Scotland has received core funding from the Chief

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

594 Scientist Office of the Scottish Government Health Directorates CZD/16/6 and the Scottish
595 Funding Council HR03006. Genotyping of the GS samples was carried out by the Genetics
596 Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and
597 was funded by the UK MRC and the Wellcome Trust (Wellcome Trust Strategic Award
598 "Stratifying Resilience and Depression Longitudinally" (STRADL) Reference
599 104036/Z/14/Z). DNA methylation profiling of the GS:SFHS samples was funded by the
600 Wellcome Trust Strategic Award [10436/Z/14/Z]. The Orkney Complex Disease Study
601 (ORCADES) was supported by the Chief Scientist Office of the Scottish Government
602 (CZB/4/276, CZB/4/710), a Royal Society URF to J.F.W., the MRC Human Genetics Unit
603 quinquennial programme "QTL in Health and Disease", Arthritis Research UK and the
604 European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-
605 018947). MESA and the MESA SHARe projects are conducted and supported by the
606 National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA
607 investigators. Support for MESA is provided by contracts 75N92020D00001,
608 HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160,
609 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006,
610 N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165,
611 N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-
612 TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. The MESA Epigenomics &
613 Transcriptomics Studies were funded by NIH grants 1R01HL101250, 1RF1AG054474,
614 R01HL126477, R01DK101921, and R01HL135009.

615

616 **Ethics approval**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

617 This study was ethically approved by the Tayside Research Ethics Committee (reference
618 05/S1401/89). Participants all gave written consent after having an opportunity to discuss the
619 research and before any data or samples were collected.

620

621 **Data Availability Statement**

622 Generation Scotland data are available from the MRC IGC Institutional Data Access / Ethics
623 Committee for researchers who meet the criteria for access to confidential data. Generation
624 Scotland data are available to researchers on application to the Generation Scotland Access
625 Committee (access@generationscotland.org). The managed access process ensures that
626 approval is granted only to research which comes under the terms of participant consent
627 which does not allow making participant information publicly available.

628

629 **Acknowledgements**

630 We want to acknowledge support from Genetics Core Laboratory at the Wellcome Trust
631 Clinical Research Facility (Edinburgh, Scotland) for genotyping of the GS samples. We are
632 grateful to all the families who took part, the general practitioners, and the Scottish School of
633 Primary Care for their help in recruiting them, and the whole Generation Scotland team,
634 which includes interviewers, computer and laboratory technicians, clerical workers, research
635 scientists, volunteers, managers, receptionists, healthcare assistants and nurses. We would
636 like to acknowledge the invaluable contributions of the research nurses in Orkney, the
637 administrative team in Edinburgh and the people of Orkney. MESA and the MESA SHARe
638 projects want to acknowledge the supported from the National Heart, Lung, and Blood
639 Institute (NHLBI) in collaboration with MESA investigators. YZ wants to acknowledge
640 support from Dr. Lucija Klaric for pre-processing the ORCADES data.

641

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 **Conflict of interest statement**

643 The authors declare that they have no competing interests.

644

645

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Huan, T., et al., *Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease.* Nat Commun, 2019. **10**(1): p. 4267.
2. Barbu, M.C., et al., *Epigenetic prediction of major depressive disorder.* Mol Psychiatry, 2020.
3. Min, J.L., et al., *Genomic and phenomic insights from an atlas of genetic effects on DNA methylation.* medRxiv, 2020.
4. Zeng, Y., et al., *Parent of origin genetic effects on methylation in humans are common and influence complex trait variation.* Nat Commun, 2019. **10**(1): p. 1383.
5. Kessler, N.J., et al., *Establishment of environmentally sensitive DNA methylation states in the very early human embryo.* Science Advances, 2018. **4**.
6. Lawson, H.A., J.M. Cheverud, and J.B. Wolf, *Genomic imprinting and parent-of-origin effects on complex traits.* Nature Reviews Genetics, 2013. **14**(9): p. 608-617.
7. Cuellar Partida, G., et al., *Genome-wide survey of parent-of-origin effects on DNA methylation identifies candidate imprinted loci in humans.* Hum Mol Genet, 2018.
8. Zink, F., et al., *Insights into imprinting from parent-of-origin phased methylomes and transcriptomes.* Nature Genetics, 2018. **50**(11): p. 1542-1552.
9. Plasschaert, R.N. and M.S. Bartolomei, *Genomic imprinting in development, growth, behavior and stem cells.* Development, 2014. **141**(9): p. 1805-13.
10. Vincenz, C., et al., *Loss of Imprinting in Human Placentas Is Widespread, Coordinated, and Predicts Birth Phenotypes.* Mol Biol Evol, 2020. **37**(2): p. 429-441.
11. Carpenter, B.L., et al., *Mother-child transmission of epigenetic information by tunable polymorphic imprinting.* Proc Natl Acad Sci U S A, 2018. **115**(51): p. E11970-E11977.
12. Monk, D., et al., *Genomic imprinting disorders: lessons on how genome, epigenome and environment interact.* Nat Rev Genet, 2019.
13. Cortessis, V.K., et al., *Comprehensive meta-analysis reveals association between multiple imprinting disorders and conception by assisted reproductive technology.* 2018. **35**(6): p. 943-952.
14. Taniguchi, K., T. Kawai, and K. Hata, *Placental Development and Nutritional Environment.* Adv Exp Med Biol, 2018. **1012**: p. 63-73.
15. Smith, B.H., et al., *Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness.* International Journal of Epidemiology, 2013. **42**(3): p. 689-700.
16. Smith, B.H., et al., *Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability.* BMC Med Genet, 2006. **7**: p. 74.
17. Kerr, S.M., et al., *Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study.* BMC Med Genet, 2013. **14**: p. 38.
18. Navrady, L.B., et al., *Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS).* Int J Epidemiol, 2018. **47**(1): p. 13-14g.
19. McQuillan, R., et al., *Runs of homozygosity in European populations.* Am J Hum Genet, 2008. **83**(3): p. 359-72.
20. Gunderson, K.L., *Whole-genome genotyping on bead arrays.* Methods Mol Biol, 2009. **529**: p. 197-213.
21. Amador, C., et al., *Recent genomic heritage in Scotland.* Bmc Genomics, 2015. **16**(1): p. 437.
22. Amador, C., et al., *Omics versus Questionnaires: Using methylation data to assess the environmental impact of smoking on obesity-related traits.* bioRxiv, 2020: p. 2020.10.08.329672.
23. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data.* Nature genetics, 2015. **47**(9): p. 1091-1098.
24. Cuellar Partida, G., et al., *Genome-wide survey of parent-of-origin effects on DNA methylation identifies candidate imprinted loci in humans.* Hum Mol Genet, 2018. **27**(16): p. 2927-2939.
25. Zhang, F., et al., *OSCA: a tool for omic-data-based complex trait analysis.* Genome Biol, 2019. **20**(1): p. 107.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.28.450122; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26. Martin, T.C., et al., *coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns.* BMC bioinformatics, 2015. **16**(1): p. 131.
27. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: springer.
28. Reynolds, L.M., et al., *Age-related variations in the methylome associated with gene expression in human monocytes and T cells.* Nat Commun, 2014. **5**: p. 5366.
29. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* Nucleic Acids Res, 2017. **45**(W1): p. W98-W102.
30. Viechtbauer, W., *Conducting Meta-Analyses in R with the metafor Package.* Journal of Statistical Software, 2010. **36**(3): p. 1-48.
31. Plomin, R., J.C. Loehlin, and J.J.D.p. DeFries, *Genetic and environmental components of" environmental" influences.* 1985. **21**(3): p. 391.
32. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data.* Nat Genet, 2015.
33. Carithers, L.J., et al., *A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project.* Biopreserv Biobank, 2015. **13**(5): p. 311-9.
34. Joehanes, R., et al., *Epigenetic Signatures of Cigarette Smoking.* Circ Cardiovasc Genet, 2016. **9**(5): p. 436-447.
35. You, C., et al., *A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes.* Nat Commun, 2020. **11**(1): p. 4779.
36. Dogan, M.V., S.R.H. Beach, and R.A. Philibert, *Genetically contextual effects of smoking on genome wide DNA methylation.* Am J Med Genet B Neuropsychiatr Genet, 2017. **174**(6): p. 595-607.
37. Heo, J., et al., *Sirt1 Regulates DNA Methylation and Differentiation Potential of Embryonic Stem Cells by Antagonizing Dnmt3l.* Cell Rep, 2017. **18**(8): p. 1930-1945.
38. Wiehle, L., et al., *Tet1 and Tet2 Protect DNA Methylation Canyons against Hypermethylation.* Mol Cell Biol, 2016. **36**(3): p. 452-61.
39. Wasson, J.A., et al., *Maternally provided LSD1/KDM1A enables the maternal-to-zygotic transition and prevents defects that manifest postnatally.* 2016. **5**: p. e08848.
40. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data.* Nat Genet, 2015. **advance online publication**.
41. consortium, C., *Sparse whole-genome sequencing identifies two loci for major depressive disorder.* Nature, 2015. **523**(7562): p. 588-91.
42. Dzionek, A., et al., *BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood.* J Immunol, 2000. **165**(11): p. 6037-46.
43. Sharp, G.C., et al., *Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium.* Human Molecular Genetics, 2017. **26**(20): p. 4067-4085.