bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # Association between long-term air pollution exposure and DNA
2 # methylation: the REGICOR study.

3 Sergi Sayols-Baixeras[1,2,3], Alba Fernández-Sanlés[1,2], Albert Prats[1,4], Isaac Subirana[5,1],
4 Michelle Plusquin[6,7,8], Nino Künzli[9], Jaume Marrugat[10,3], Xavier Basagaña[2,11,5], Roberto
5 Elosua[1,3,12]
6
7 1.-Cardiovascular Epidemiology and Genetics Research Group, IMIM (Hospital del Mar
8 Medical Research Institute), Dr Aiguader 88, 08003 Barcelona, Catalonia, Spain.
9 2.-Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Catalonia, Spain.
10 3.-CIBER Cardiovascular Diseases (CIBERCV), Dr Aiguader 88, 08003 Barcelona,
11 Catalonia, Spain.
12 4.-Preventive Medicine and Public Health Training Unit, Parc de Salut Mar-Universitat
13 Pompeu Fabra-Agència de Salut Pública de Barcelona (UDMPiSP PSMar-UPF-ASPB), Dr
14 Aiguader 88, 08003 Barcelona, Catalonia, Spain.
15 5.-CIBER Epidemiology and Public Health (CIBERESP), Dr Aiguader 88, 08003 Barcelona,
16 Catalonia, Spain.
17 6.-Department of Epidemiology and Biostatics, The School of Public Health, Imperial
18 College London, Medical School Building, St Mary's Hospital, Norfolk Place, London W2
19 1PG, United Kingdom.
20 7.-Medical Research Council-Health Protection Agency Centre for Environment and Health,
21 Imperial College London, St Mary's Campus, Imperial College, Paddington, London W2
22 1PG, United Kingdom.
23 8.-Centre for Environmental Sciences, Hasselt University, Campus Hasselt, Martelarenlaan
24 42, BE3500 Hasselt, Belgium.
25 9.-Swiss Tropical and Public Health Institute, Socinstrasse 57, 4051 Basel, Switzerland.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26 10.-REGICOR Research Group, IMIM (Hospital del Mar Medical Research Institute), DR
27 Aiguader 88, 08003 Barcelona, Catalonia, Spain.
28 11.- ISGlobal (Institute for Global Health), Dr Aiguader 88, 08003 Barcelona, Spain.
29 12.-Faculty of Medicine, University of Vic-Central University of Catalonia (UVic-UCC),
30 Carretera de Roda 70, 08500 Vic, Catalonia, Spain.
31
32 **Author for correspondence:**
33 Roberto Elosua, MD, PhD
34 IMIM, Hospital del Mar Medical Research Institute
35 Dr Aiguader 88, 08003 Barcelona, Catalonia, Spain
36 Telephone: (+34) 933160800
37 Email: relosua@imim.es

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

38 **ABSTRACT**

39 **Background:** Limited evidence suggests that epigenetic mechanisms may partially mediate
40 the adverse effects of air pollution on health. Our aims were to identify new genomic loci
41 showing differential DNA methylation associated with long-term exposure to air pollution
42 and to replicate loci previously identified in other studies.

43 **Methods:** A two-stage epigenome-wide association study was designed: 630 individuals from
44 the REGICOR study were included in the discovery and 454 participants of the EPIC-Italy
45 study in the validation stage. DNA methylation was assessed using the Infinium
46 HumanMethylation450 BeadChip. NOX, NO2, PM10, PM2.5, PMcoarse, traffic intensity and
47 traffic load exposure were measured according to the ESCAPE protocol. A systematic review
48 was undertaken to identify those cytosine-phosphate-guanine (CpGs) associated with air
49 pollution in previous studies and we screened for them in the discovery study.

50 **Results:** In the discovery stage of the epigenome-wide association study, 81 unique CpGs
51 were associated with air pollution (p-value $<10^{-5}$) but none of them were validated in the
52 replication sample. Furthemore, we identified 12 CpGs in the systematic review showing
53 differential methylation with a p-value fulfilling the Bonferroni criteria and 1642 CpGs
54 fulfilling the false discovery rate criteria, all of which were related to $PM_{2.5}$ or NO_2. None of
55 them was replicated in the discovery study, in which the top hits were located in an intergenic
56 region on chromosome 1 (cg10893043, p-value=$6.79 \cdot 10^{-5}$) and in the *PXK* and *ARSA* genes
57 (cg16560256, p-value=$2.23 \cdot 10^{-04}$; cg11953250, p-value=$3.64 \cdot 10^{-04}$).

58 **Conclusions:** Neither new genomic loci associated with long-term air pollution were
59 identified, nor previously identified loci were replicated. Continued efforts to test this
60 potential association are warranted.

61

62 **Keywords:** Air pollution, DNA methylation, Epigenome-wide association study.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

63 **BACKGROUND**

64 Exposure to air pollution remains a global threat with more than 90% of the world's
65 population now exceeding the exposure limits proposed for particulate matter by the World
66 Health Organization (WHO) (1). At the same time, a growing body of evidence consistently
67 supports the adverse health effects of air pollution, which the same WHO report estimates to
68 be related with 3 million premature deaths worldwide each year. However, the mechanisms
69 by which air pollution induces these deleterious effects are not completely understood.

70 Epigenetics encompasses mechanisms that regulate gene expression without changing the
71 DNA sequence, and may contribute to the relation between air pollution and health. The most
72 studied epigenetic mechanism is DNA methylation, which is heritable but can also be
73 modified by life-style and environmental factors. Recently, several studies have analyzed the
74 association between air pollution and DNA methylation using a genome-wide approach, and
75 have reported numerous loci showing differential methylation related to this exposure (2–10).

76 The aims of this study were both to identify new genomic loci showing differential
77 methylation associated with long-term exposure to air pollution in a population-based study in
78 Spain and to replicate loci previously reported in other studies.

79

80 **METHODS**

81 **Identification of new genomic loci showing differential methylation related to long-term**
82 **air pollution exposure (Aim 1)**

83 *Study design and population*

84 We designed a cross-sectional epigenome-wide association study in two stages. We used the
85 REGICOR (REgistre GIroní del COR) cohort as the discovery study and the Italy center of
86 the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy) as the

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

87 replication study, followed by a meta-analysis of the results observed in both studies
88 (REGICOR + EPIC-Italy).
89 As previously described,(11) the REGICOR discovery sample included 648 participants
90 randomly selected from the second wave of the REGICOR study in 2008-2013. The initial
91 survey, performed during 2003-2005, included participants aged between 35 and 79 years, not
92 institutionalized, and residing in Girona province (Catalonia, Spain) (11).
93 The EPIC-Italy replication study included 47,749 individuals in a multicenter prospective
94 cohort recruited during 1993-1998 (12,13). The samples selected for the present study were
95 from two case-control studies on breast cancer (14) and colorectal cancer in Varese and Turin.
96
97 *HumanMethylation450 BeadChip*
98 DNA was extracted using standardized methods from peripheral blood (Puregen TM; Gentra
99 Systems) and buffy coats (QIAGEN QIAsymphony DNA Midi Kit) in the REGICOR and
100 EPIC-Italy studies, respectively. DNA was bisulphite-converted and the epigenome-wide
101 methylation profiles were obtained using the Infinium HumanMethylation450 BeadChip
102 (Illumina) (450K) to assess methylation on 485,577 cytosine-phosphate-guanine (CpGs)
103 throughout the genome, following the Illumina Infinium HD Methylation protocol (15,16).
104 The REGICOR samples were processed in two centers of the Spanish National Genotyping
105 Center: the Center for Genomic Regulation in Barcelona (n=188 samples) and the Centro
106 Nacional de Investigaciones Oncológicas in Madrid (n=460 samples). All the processed
107 batches contained two duplicate samples used as an internal quality control. The EPIC-Italy
108 samples were analyzed at the Human Genetics Foundation in Turin. The same well-defined
109 pipeline was used in both studies to assess the quality control of the methylation data (17).
110 We used the M-value as the main DNA methylation measurement (Equation 1). An M-
111 value=0 indicates that the CpG is half methylated, a positive M-value that the CpG is more

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

112 methylated than unmethylated, and a negative M-value the inverse result. We standardized the
113 M-value by batch (Equation 2) to reduce the batch effect and other potential technical sources
114 of variation.

115 Equation 1: $M_{value} = \log_2\left(\frac{M_i + \alpha}{U_i + \alpha}\right)$ Mi= intensity of methylated probes.
Ui= intensity of unmethylated probes.
α= 1.

116 Equation 2: $Z = \frac{(X - \overline{X})}{\sqrt{\frac{\sum (X - \overline{X})^2}{(n-1)}}}$ X= M-value for a specific individual.
$\overline{X}$= mean of M-value for a specific batch.
n= sample size.

117

118 *Air pollution exposure*

119 Both the REGICOR data and the Turin component of EPIC-Italy contained particulate matter
120 exposure [aerodynamic diameter of <10µm (PM_{10}), <2.5µm ($PM_{2.5}$), and PM_{coarse} (the
121 difference between PM_{10} and $PM_{2.5}$)], nitrogen oxides (NO_x) and nitrogen dioxide (NO_2)
122 measurements. For the Varese component of EPIC-Italy, only NO_x and NO_2 were available.
123 Both studies used the ESCAPE protocol to assess long-term exposure to air pollution (18,19).
124 As previously described, address histories for the past 10 years were collected by
125 questionnaire, and each address was geocoded at the front-door level (20). Using land use
126 regression (LUR) models, 10-year weighted average exposure to NO_X and to nitrogen dioxide
127 NO_2 for each participant were estimated (21). The model's coefficient of determination for
128 NO_2 was 0.63 for REGICOR and 70% for Turin, and for NO_x was 66% and 72%, respectively
129 (20).

130 PM_{10}, $PM_{2.5}$ and PM_{coarse} were also assessed using LUR models. The R^2 of the models for
131 REGICOR and Turin was 71% and 69% for PM_{10}, 51% and 59% for $PM_{2.5}$, and 71% and 58%
132 for PM_{coarse}, respectively (19).

133 We also used traffic proximity markers as surrogates of air pollution exposure in independent
134 analyses. For each address, we calculated the traffic intensity at the nearest street and the

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

135 traffic load (sum of traffic intensity multiplied by length of road segment) for all segments in
136 a 100 meters buffer and derived 10-year average values for each participant.
137
138 *Other covariates*
139 REGICOR's trained team of nurses collected relevant sociodemographic, lifestyle, and
140 cardiovascular risk factors using standardized and validated questionnaires (22,23). Smoking
141 exposure was grouped in four categories: current smoker (smoked ≥1 cigarette/day at the time
142 of the visit, on average, or gave up smoking within the year of the visit); former smoker, 1-5
143 years (gave up smoking up to 5 years before the visit); former smokers >5 years; and never
144 smokers. The EPIC-Italy study collected the same variables, as well as the participating center
145 and patient's diagnostic status.
146 We estimated cell concentration using Houseman algorithm by means of *minfi* R package
147 (24,25). In both cohorts, we also calculated surrogate variables to control for potential
148 residual cofounding, using the *sva* R package (26). These variables identify and remove
149 potential and non-measured sources of variation due to technical and biological confounders.
150
151 *Statistical analysis*
152 We assessed the association between air pollution and DNA methylation using robust linear
153 regression model to reduce the effect of outliers. We used the differing air pollution exposures
154 as independent variables and DNA methylation as the dependent variable. The models were
155 adjusted for age, sex, smoking exposure, and cell composition (Model 1). Moreover, a second
156 model adjusted for surrogate variables was also fitted (Model 2). In the REGICOR discovery
157 study, we selected for validation those CpGs with a p-value of the association below an
158 arbitrary threshold (p-value $<10^{-5}$) for each specific exposure. The results were replicated in
159 the EPIC-Italy study using the same method and models, including also disease status and

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

160 study center as additional covariates. The results for each CpG were then meta-analyzed using
161 a random effects model, applying Bonferroni criteria to declare a result as statistically
162 significant (0.05/427,948 CpGs; p-value<$1.17 \cdot 10^{-07}$).
163

164 **Replication of previously published CpGs associated with air pollution (Aim 2)**

165 We performed a systematic review to identify relevant epigenome-wide association studies
166 indexed in Pubmed (https://www.ncbi.nlm.nih.gov/pubmed/) from its inception to March
167 2018. We used the following search terms strategy: (methylati* [Title/Abstract]) AND
168 (epigenome-wide [Title/Abstract] OR genome-wide [Title/Abstract] OR 450K
169 [Title/Abstract] OR 450 [Title/Abstract] OR HumanMethylation450 [Title/Abstract]) AND
170 ("Air pollution" [Title/Abstract]). The articles identified were manually screened by 1
171 reviewer (SS-B), focusing first on the title and abstract and then on the complete manuscripts
172 to assess their appropriateness for inclusion in the review. The same author extracted the
173 CpGs that were significantly associated with air pollution and replicated in at least one
174 external population. In case of doubts the article was evaluated by a second reviewer (RE) to
175 achieve a consensus. The identified CpGs were then screened in the selected REGICOR
176 cohort.
177

178 *Statistical analysis*

179 The same analysis strategy and methodology described above was followed for the
180 replication. We selected as distinctive the CpGs reported as statistically significant in the
181 original studies, based on both Bonferroni corrected p-values and false discovery rate (FDR)
182 p-values<0.05. We considered as replicated those CpGs that fulfilled both the Bonferroni
183 criteria according to the number of CpGs previously discovered and the FDR p-value.
184

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

185 **RESULTS**

186 **Identification of new genomic loci showing differential methylation related to air**
187 **pollution (Aim 1)**

188 *Discovery Stage*

189 After applying the quality control of the 450K array, we excluded 3 individuals and 57,629
190 CpG probes. Moreover, we removed those individuals without information on air pollution
191 exposure (n=15). Finally, 630 individuals and 427,948 probes were included in the analysis.
192 The main sociodemographic and clinical characteristics and the air pollution exposures of the
193 study participants are shown in Table 1. The Manhattan plots and q-q plots of the associations
194 between air pollution exposures and DNA methylation are shown in Supplementary Figure 1.
195 We identified 81 unique CpGs associated with air pollution exposures with a p-value<10^{-5}. In
196 model 1, 6 CpGs were associated with PM_{10}, 0 related to $PM_{2.5}$, 5 to PM_{coarse}, 2 to NO_X, 4 to
197 NO_2, and 2 to traffic at the nearest street (Supplementary Table 1). In model 2, 7 CpGs were
198 associated with PM_{10}, 18 related to $PM_{2.5}$, 6 to PM_{coarse}, 28 to NO_X, 32 to NO_2, and 9 to traffic
199 at the nearest street (Supplementary Table 2).

200

201 *Validation Stage and Meta-analysis*

202 After applying a similar quality control of the 450K array, we included all 81 CpGs selected
203 for replication and we excluded 5 individuals. Moreover, we removed those individuals
204 without information on confounder variables (n=15) and air pollution exposure (n=61).
205 Finally, 454 individuals were included in the analysis of NO traits (Turin and Varese) and 297
206 in the analysis of PM traits (Turin). The main characteristics and air pollution exposures of
207 the EPIC-Italy participants are shown in Table 1. The associations between air pollution and
208 the selected CpGs in this analysis are shown in Supplementary Table 1 and 2.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

209 The results of the meta-analysis of the REGICOR and the EPIC-Italy studies are shown in
210 Supplementary Table 1 and 2. None of the selected CpGs was validated in the joint analysis.
211
212 **Table 1:** Main characteristics of the participants included in the discovery and validation
213 studies.

	REGICOR N=630	**EPIC-Italy N=454**
Age, years	63.3 (11.7)	54.2 (7.1)
Sex, female, n (%)	323 (51.3)	323 (71)
PM10, µg/m3	29.9 [27.5; 32.7]	46.91 [38.24; 53.23]
PMcoarse, µg/m3	14.1 [13.3; 15.7]	16.75 [10.88; 20.65]
PM2.5, µg/m3	15.0 [13.5; 15.9]	30.95 [26.74; 32.73]
PM2.5abs, µg/m3	2.40 [1.90; 2.91]	3.38 [2.34; 3.64]
NoX, µg/m3	61.7 [42.5; 72.8]	92.83 [34.46; 131.00]
NO2, µg/m3	36.9 [24.3; 42.6]	50.00 [22.26; 67.92]
Traffic near, µg/m3	1725 [862; 7000]	4029 [0; 45012]
Total Cholesterol, mg/dL	209 (36.6)	357 (136.2)
LDL Cholesterol, mg/dL	135 (32.4)	-
HDL Cholesterol, mg/dL	53.1 (12.4)	-
Triglycerides, mg/dL	89.0 [67.0;121]	-
SBP, mmHg	131 (18.2)	135.8 (19.4)
DBP, mmHg	76.0 (9.93)	84.3 (10.7)
Hypertension, n (%)	297 (47.3)	119 (26.2)
Diabetes, n (%)	62 (9.89)	7 (1.54)
Smoking status, n (%)		
Current smokers	105 (16.7)	234 (51.5)
Former 1-5 years	30 (4.76)	-
Former >5 years	159 (25.2)	-
Former	-	111 (24. 4)
Never smokers	336 (53.3)	109 (24.0)
Cholesterol Treatment, n (%)	151 (24.0)	-
Diabetes Treatment, n (%)	43 (6.86)	-
Blood pressure treatment, n (%)	196 (31.3)	-
Breast cancer cases, n (%)	-	82 (18.1)
Colorectal cancer cases, n (%)	-	133 (29.3)
Controls, n (%)	-	239 (52.6)

214
215
216

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

217 *Statistical power*

218 The regression coefficient values (effect size) that we could detect as statistically significant
219 in the meta-analysis, accepting an alpha risk of $1.17 \cdot 10^{-07}$ in a two-sided test and with an 80%
220 power, are shown in Supplementary Table 1 and 2.

221

222 **Replication of the previously published CpGs associated with air pollution (Aim 2)**

223 We initially identified 19 manuscripts (2–10,27–36) based on the search terms and after
224 reading the full manuscripts we finally selected 9 studies (2–10). We selected 12 CpGs
225 showing differential methylation in relation to $PM_{2.5}$ from the only manuscript that corrected
226 for Bonferroni criteria (9). Three of these CpGs could not be analyzed in the REGICOR study
227 as they did not pass the quality control, and none of the others was replicated (Supplementary
228 Table 3).

229 In a secondary analysis, we included 1,642 CpGs from two manuscripts (6,9) fulfilling a FDR
230 p-value<0.05. Among all the selected FDR results, 195 CpGs could not be analyzed in our
231 study as they did not pass the quality control. We did not replicate any of the 1,447 CpGs
232 analyzed (Table 2 and Supplementary Table 3). In the REGICOR study, the top hits were
233 located in an intergenic region on chromosome 1 (cg10893043, p-value=$6.79 \cdot 10^{-5}$) and in the
234 *PXK* and *ARSA* genes (cg16560256, p-value= $2.23 \cdot 10^{-04}$; cg11953250, p-value=$3.64 \cdot 10^{-04}$) for
235 $PM_{2.5}$.

236

237 **Table 2:** Top hits of the replication of previously published Cps associated with NO_2 and
238 PM_{25} exposures.

NO2						
CpG	**Chromosome**	**Position**	**Gene**	**coefficient**	**SE**	**P value**
cg24172570	7	27561178	HIBADH	4.92E-04	1.91E-03	7.97E-01
cg08973675	10	101380289	SLC25A28	-7.67E-04	2.74E-03	7.80E-01
cg12283362	19	5709149	LONP1	-1.88E-04	2.15E-03	9.30E-01
PM25						

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CpG	Chromosome	Position	Gene	coefficient	SE	P value
cg10893043	1	51442760	NA	4.30E-02	1.08E-02	6.79E-05
cg16560256	3	58328580	PXK	-4.39E-02	1.19E-02	2.23E-04
cg05088605	17	79987813	LRRC45	-4.15E-02	1.12E-02	2.15E-04

239

240 **DISCUSSION**

241 This population-based and cross-sectional study did not identify new loci or replicate loci
242 previously identified as showing differential methylation related to long-term exposure to air
243 pollution.

244 The lack of positive results in our study should be interpreted with caution and some
245 methodological issues must be considered. First, our study is underpowered to detect small
246 effect size associations. The statistical power of our study was estimated (Supplementary
247 Table 1 and 2) and is similar to previous studies. Second, we defined strict criteria to consider
248 an association as statistically significant based on the Bonferroni multiple comparisons
249 correction. This approach is more conservative than the FDR p-value used in other studies
250 (6,9). However, in the replication effort we also included those CpGs that were identified
251 using the FDR criteria and we did not replicate any of these CpGs in our sample. Third, the
252 exposure to air pollution and its variability is lower than that observed in other studies,
253 limiting our capability to identify real associations. Fourth, the exposure assessment was
254 estimated using land use regression models. Although this methodology is commonly used
255 and we followed the ESCAPE protocol, (18,19) using validated exposure estimations, (21)
256 some exposure misclassification could still be present, reducing the power to detect the
257 association of interest in our study. Fifth, the replication was carried in a case-control study
258 that may have a different methylation pattern. Finally, we estimated long-term exposure to air
259 pollution whereas other studies have analyzed the association between short-term exposure
260 and DNA methylation; however, in studies that analyzed several time exposures, the longer
261 the exposure the higher the number of loci showing differential methylation (9).

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

262 Despite these limitations that should be considered, we would highlight some of the results
263 observed in this study. In the discovery effort, we identified 7 CpGs associated with PM_{10}, 18
264 related to $PM_{2.5}$, 6 with PM_{coarse}, 28 with NO_X, 32 with NO_2, and 9 related to traffic at the
265 nearest street in the REGICOR study. However, none of them were validated in the EPIC-
266 Italy study. In our effort to replicate previous findings, we identified one locus located in an
267 intergenic region on chromosome 1 (cg10893043, p-value=6.79·10-5) as potentially
268 associated with $PM_{2.5}$. The cg10893043 is close to the *CDKN2C* gene. The protein encoded
269 by this gene is a member of the INK4 family of cyclin-dependent kinase inhibitors and
270 regulates cell growth by controlling cell-cycle G1 progression (37). Some studies have shown
271 that the expression of this gene inhibits the growth of human cells in animal models and have
272 suggested a potential role in tumorigenesis (38).
273 Among the strengths of our study, we would mention the standardized methodology
274 following the ESCAPE protocol that was consistently used to assess air pollution exposure.
275 This methodology was validated in our population and was used to assess long-term air
276 pollution exposure. Moreover, we applied a commonly used methodology to assess DNA
277 methylation at the genome-wide level and a standardized methodology with both a discovery
278 and an independent validation population.
279

280 **CONCLUSIONS**

281 The results of our study are negative as we did not identify any new genomic loci associated
282 with long-term air pollution and we did not replicate any previously identified loci. However,
283 these negative results should be interpreted with caution. New joint efforts, increasing the
284 statistical power of the analysis and the variability of the exposure to air pollution, and
285 considering both short- and long-term exposure, are warranted to assess the potential
286 association between air pollution and DNA methylation.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

287 **LIST OF ABBREVIATIONS**

288 WHO: World Health Organization.

289 REGICOR: REgistre GIroní del COR.

290 EPIC-Italy: The Italy center of European Prospective Investigation into Cancer and Nutrition.

291 450K: Infinium Human Methylation450 BeadChip.

292 CpG: Cytosine-phosphate-guanine.

293 PM_{10}: Particulate matter with an aerodynamic diameter of $<10\mu m$.

294 $PM_{2.5}$: Particulate matter with an aerodynamic diameter of $<2.5\mu m$.

295 PM_{coarse}: The difference between PM_{10} and $PM_{2.5}$.

296 NO_X: Nitrogen oxides.

297 NO_2: Nitrogen dioxide

298 LUR: Land use regression.

299 FDR: False Discovery Rate.

300

301 **DECLARATIONS**

302 **Ethics approval and consent to participate**

303 All participants in both studies (REGICOR and EPIC-Italy) signed an informed consent; the
304 studies were approved by the local ethic committees (PSMAR CEIC- 2012/4729/I) and
305 followed national legislation and the Declaration of Helsinki criteria.

306

307 **Consent for publication**

308 Not applicable.

309

310 **Availability of data and material**

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

311 The datasets used and/or analyzed during the current study are available from the
312 corresponding author [RE] on reasonable request.
313

314 **Competing interests**

315 The authors declare that they have no competing interests.
316

317 **Funding**

318 This work was supported by grants from the Catalan Agència de Gestió d'Ajuts Universitaris
319 de Recerca (2014-SGR-240) and the Spanish Ministry of Economy through the Instituto de
320 Salud Carlos III-FEDER (FIS PI15/00051, PI12/00232, CIBERCV, CIBERESP, Red de
321 Investigación Cardiovascular RD12/0042). S.S-B. was funded by contracts from Instituto de
322 Salud Carlos III-FEDER (IFI14/00007). A.F-S. was funded by the Spanish Ministry of
323 Economy and Competitiveness (BES-2014-069718).
324

325 **Authors' contributions**

326 - Conception or design of the study: SS-B, NK, JM, XB, RE
327 - Acquisition of data for the study: NK, JM, XB, RE
328 - Analysis of data for the manuscript: SS-B, IS, MP, XB
329 - Interpretation of data for the manuscript: SS-B, AF-S, AP, NK, JM, XB, RE
330 - Drafted the manuscript: SS-B, RE
331 - Revised the manuscript critically for important intellectual content: AF-S, AP, IS, MP, NK,
332 JM, XB
333 - All the authors have approved the final version of the manuscript AND agree to be
334 accountable for all aspects of the work in ensuring that questions related to the accuracy or
335 integrity of any part of the work are appropriately investigated and resolved.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

336

337 **Acknowledgments**

338 We thank Elaine M. Lilly, PhD, for revising and editing the English text.

339

340 **REFERENCES:**

341 1. Organization WH. WHO: Air Pollution . 2018. http://www.who.int/airpollution/en/ .
342 Acccessed Feb 2018.
343 2. de F.C. Lichtenfels AJ, van der Plaat DA, de Jong K, van Diemen CC, Postma DS,
344 Nedeljkovic I, et al. Long-term Air Pollution Exposure, Genome-wide DNA
345 Methylation and Lung Function in the LifeLines Cohort Study. Environmental Health
346 Perspectives. 2018;126:027004.
347 3. Plusquin M, Guida F, Polidoro S, Vermeulen R, Raaschou-Nielsen O, Campanella G,
348 et al. DNA methylation and exposure to ambient air pollution in two prospective
349 cohorts. Environment International. 2017;108:127–36.
350 4. Zhong J, Karlsson O, Wang G, Li J, Guo Y, Lin X, et al. B vitamins attenuate the
351 epigenetic effects of ambient fine particles in a pilot human intervention trial.
352 Proceedings of the National Academy of Sciences. 2017;114:3503–8.
353 5. Chi GC, Liu Y, MacDonald JW, Barr RG, Donohue KM, Hensley MD, et al. Long-
354 term outdoor air pollution and DNA methylation in circulating monocytes: results from
355 the Multi-Ethnic Study of Atherosclerosis (MESA). Environmental Health.
356 2016;15:119.
357 6. Gruzieva O, Xu CJ, Breton C V., Annesi-Maesano I, Antó JM, Auffray C, et al.
358 Epigenome-wide meta-analysis of methylation in children related to prenatal NO2air
359 pollution exposure. Environmental Health Perspectives. 2017;125:104–10.
360 7. Goodrich JM, Reddy P, Naidoo RN, Asharam K, Batterman S, Dolinoy DC. Prenatal

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

361 exposures and DNA methylation in newborns: a pilot study in Durban, South Africa.
362 Environ Sci: Processes Impacts. 2016;18:908–17.
363 8. Breton C, Gao L, Yao J, Siegmund K, Lurmann F, Gilliland F. Particulate matter, the
364 newborn methylome, and cardio-respiratory health outcomes in childhood. Environ
365 Epigenet. 2016;2:dvw005.
366 9. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, et al. Genome-wide
367 analysis of DNA methylation and fine particulate matter air pollution in three study
368 populations: KORA F3, KORA F4, and the normative aging study. Environmental
369 Health Perspectives. 2016;124:983–90.
370 10. Jiang R, Jones MJ, Sava F, Kobor MS, Carlsten C. Short-term diesel exhaust inhalation
371 in a controlled human crossover study is associated with changes in DNA methylation
372 of circulating mononuclear cells in asthmatics. Part Fibre Toxicol. 2014;11:71.
373 11. Grau M, Subirana I, Elosua R, Solanas P, Ramos R, Masiá R, et al. Trends in
374 cardiovascular risk factor prevalence (1995-2000-2005) in northeastern Spain.
375 European journal of cardiovascular prevention and rehabilitation : official journal of
376 the European Society of Cardiology, Working Groups on Epidemiology & Prevention
377 and Cardiac Rehabilitation and Exercise Physiology. 2007; 653–9.
378 12. Beulens JWJ, Monninkhof EM, Monique Verschuren WM, van der Schouw YT, Smit
379 J, Ocke MC, et al. Cohort profile: The EPIC-NL study. International Journal of
380 Epidemiology. 2010;39:1170–8.
381 13. Palli D, Berrino F, Vineis P, Tumino R, Panico S, Masala G, et al. A molecular
382 epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and
383 baseline characteristics of participants. Tumori. 2003;89:586–93.
384 14. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, et al.
385 Epigenome-wide association study reveals decreased average methylation levels years

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

386 before breast cancer diagnosis. Clinical Epigenetics. 2015;7:67.

387 15. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA
388 methylation array with single CpG site resolution. Genomics. 2011; 98:288-95.

389 16. Sandoval J, Heyn HA, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al.
390 Validation of a DNA methylation microarray for 450,000 CpG sites in the human
391 genome. Epigenetics. 2011;6:692–702.

392 17. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, Civeira F, Roquer J, Do A, et al.
393 Identification and validation of seven new loci showing differential DNA methylation
394 related to serum lipid profile: an epigenome-wide approach. The REGICOR study.
395 Human Molecular Genetics. 2016;25:4556–65.

396 18. Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, et al.
397 Development of NO2 and NOx land use regression models for estimating air pollution
398 exposure in 36 study areas in Europe - The ESCAPE project. Atmospheric
399 Environment. 2013;72:10–23.

400 19. Eeftens M, Beelen R, De Hoogh K, Bellander T, Cesaroni G, Cirach M, et al.
401 Development of land use regression models for PM2.5, PM 2.5 absorbance, PM10 and
402 PMcoarse in 20 European study areas; Results of the ESCAPE project. Environmental
403 Science and Technology. 2012;46:11195–205.

404 20. Rivera M, Basagaña X, Aguilera I, Foraster M, Agis D, de Groot E, et al. Association
405 between long-term exposure to traffic-related air pollution and subclinical
406 atherosclerosis: The REGICOR study. Environmental Health Perspectives. 2013;
407 121:223–30.

408 21. Basagaña X, Aguilera I, Rivera M, Agis D, Foraster M, Marrugat J, et al. Measurement
409 error in epidemiologic studies of air pollution based on land-use regression models.
410 American Journal of Epidemiology. 2013;178:1342–6.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

411 22. World Health Organisation. Manual of The MONICA Project. 2000.
412 http://www.ktl.fi/publications/monica/manual/index.htm. Acccessed Feb 2018.
413 23. Baena-Díez JM, Alzamora-Sas MT, Grau M, Subirana I, Vila J, Torán P, et al.
414 [Validity of the MONICA cardiovascular questionnaire compared with clinical
415 records]. Gaceta sanitaria. 2010;23:519–25.
416 24. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH,
417 et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC
418 Bioinformatics. 2012; 13:86.
419 25. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et
420 al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of
421 Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9.
422 26. Jeffrey T. Leek and W. Evan Johnson and Hilary S. Parker and Andrew E. Jaffe and
423 John D. Storey. sva: Surrogate Variable Analysis. R package version 3.10.0.
424 27. Jiang C-L, He S-W, Zhang Y-D, Duan H-X, Huang T, Huang Y-C, et al. Air pollution
425 and DNA methylation alterations in lung cancer: A systematic and comparative study.
426 Oncotarget. 2017;8:1369–91.
427 28. Anto JM, Bousquet J, Akdis M, Auffray C, Keil T, Momas I, et al. Mechanisms of the
428 Development of Allergy (MeDALL): Introducing novel concepts in allergy
429 phenotypes. Journal of Allergy and Clinical Immunology. 2017; 139:388–99.
430 29. Mostafavi N, Vlaanderen J, Portengen L, Chadeau-Hyam M, Modig L, Palli D, et al.
431 Associations between genome-wide gene expression and ambient nitrogen oxides
432 (NOx). Epidemiology. 2017; 28:320-8.
433 30. Gref A, Merid SK, Gruzieva O, Ballereau S, Becker A, Bellander T, et al. Genome-
434 wide interaction analysis of air pollution exposure and childhood asthma with
435 functional follow-up. American Journal of Respiratory and Critical Care Medicine.

bioRxiv preprint doi: https://doi.org/10.1101/404483; this version posted September 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

436 2017;195:1373–83.

437 31. Somineni HK, Zhang X, Biagini Myers JM, Kovacic MB, Ulm A, Jurcak N, et al. Ten-
438 eleven translocation 1 (TET1) methylation is associated with childhood asthma and
439 traffic-related air pollution. Journal of Allergy and Clinical Immunology.
440 2016;137:797–805.

441 32. Hong X, Wang X. Epigenetics and Development of Food Allergy (FA) in Early
442 Childhood. Current Allergy and Asthma Reports. 2014; 14:460.

443 33. Janssen BG, Godderis L, Pieters N, Poels K, Kiciński M, Cuypers A, et al. Placental
444 DNA hypomethylation in association with particulate air pollution in early life. Particle
445 and Fibre Toxicology. 2013;10:22.

446 34. Wang T, Garcia JG, Zhang W. Epigenetic Regulation in Particulate Matter-Mediated
447 Cardiopulmonary Toxicities: A Systems Biology Perspective. Curr Pharmacogenomics
448 Person Med. 2012;10:314–21.

449 35. Holloway JW, Savarimuthu Francis S, Fong KM, Yang I a. Genomics and the
450 respiratory effects of air pollution exposure. Respirology. 2012;17:590–600.

451 36. Christensen BC, Marsit CJ. Epigenomics in environmental health. Frontiers in
452 Genetics. 2011;2:84.

453 37. Cánepa ET, Scassa ME, Ceruti JM, Marazita MC, Carcagno AL, Sirkin PF, et al. INK4
454 proteins, a family of mammalian CDK inhibitors with novel biological functions.
455 IUBMB Life. 2007; 59:419–26.

456 38. van Veelen W, Klompmaker R, Gloerich M, van Gasteren CJR, Kalkhoven E, Berger
457 R, et al. P18 is a tumor suppressor gene involved in human medullary thyroid
458 carcinoma and pheochromocytoma development. International journal of cancer.
459 2009;124:339–45.

460