

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Differentially methylated DNA regions in early childhood wheezing: an epigenome-wide study using saliva**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1690516> since 2021-09-02T13:47:18Z

*Published version:*

DOI:10.1111/pai.13023

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

**This is the author's final version of the contribution published as:**

Maja Popovic<sup>1</sup>, Valentina Fiano<sup>1</sup>, Francesca Fasanelli<sup>1</sup>, Morena Trevisan<sup>1</sup>, Chiara Grasso<sup>1</sup>,  
Manuela Bianca Assumma<sup>2</sup>, Anna Gillio-Tos<sup>1</sup>, Silvia Polidoro<sup>2</sup>, Laura De Marco<sup>1</sup>, Franca  
Rusconi<sup>3</sup>, Franco Merletti<sup>1</sup>, Daniela Zugna<sup>1</sup>, Lorenzo Richiardi<sup>1</sup>

**Differentially methylated DNA regions in early childhood wheezing: an epigenome-wide  
study using saliva**

*Pediatr Allergy Immunol.* 2019 May;30(3):305-314.

doi: 10.1111/pai.13023. Epub 2019 Feb 17.

**The publisher's version is available at:**

<http://hdl.handle.net/2318/1690516>

**When citing, please refer to the published version.**

**Link to this full text:**

<http://hdl.handle.net/2318/1690516>

This full text was downloaded from iris-AperTO: <https://iris.unito.it/>

1 **Differentially methylated DNA regions in early childhood wheezing: an epigenome-wide**  
2 **study using saliva**

3 Maja Popovic<sup>1</sup>, Valentina Fiano<sup>1</sup>, Francesca Fasanelli<sup>1</sup>, Morena Trevisan<sup>1</sup>, Chiara Grasso<sup>1</sup>,  
4 Manuela Bianca Assumma<sup>2</sup>, Anna Gillio-Tos<sup>1</sup>, Silvia Polidoro<sup>2</sup>, Laura De Marco<sup>1</sup>, Franca  
5 Rusconi<sup>3</sup>, Franco Merletti<sup>1</sup>, Daniela Zugna<sup>1</sup>, Lorenzo Richiardi<sup>1</sup>

6 **Affiliations**

7 <sup>1</sup> Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin; Centro di  
8 Ricerca in Medicina Sperimentale (CeRMS) and CPO Piemonte, Turin, Italy

9 <sup>2</sup> Italian Institute for Genomic Medicine (IIGM), Turin, Italy.

10 <sup>3</sup> Unit of Epidemiology, ‘Anna Meyer’ Children’s University Hospital, Florence, Italy

11

12 **Running title:** *Epigenome-wide study of infant wheezing*

13 **Corresponding author:**

14 Maja Popovic

15 Department of Medical Sciences,

16 University of Turin

17 Via Santena 7

18 10126 Turin, Italy

19 E-mail: maja.popovic@unito.it

20 Phone: +39(0) 116334628

21 **Word count:** 4351

22 **Table count:** 4

23 **Figure count:** 1

24 **Material in the electronic repository:** Yes

25 **Conflicts of interest**

26 The authors have no potential conflicts of interest to disclose.

27 **Financial support**

28 The NINFEA study was partially funded by the Compagnia San Paolo Foundation. This  
29 research received funding specifically dedicated to the Department of Medical Sciences from  
30 Italian Ministry for Education, University and Research (Ministero dell'Istruzione,  
31 dell'Università e della Ricerca - MIUR) under the programme "Dipartimenti di Eccellenza  
32 2018 – 2022". M.P. was supported by the European Commission through the Erasmus  
33 Mundus for Western Balkans II programme. L.R. received funding from the European  
34 Union's Horizon 2020 research and innovation programme under grant agreement No  
35 733206, LIFE-CYCLE project.

36

37

38

39

40

41

42

43

44

45

46

47 **Abstract**

48 **Background** Epigenetics may play a role in wheezing and asthma development. We aimed to  
49 examine infant saliva DNA methylation in association with early childhood wheezing.

50 **Methods** A case-control study was nested within the NINFEA birth cohort with 68 cases  
51 matched to 68 controls by sex, age (between 6 and 18 months, median: 10.3 months) and  
52 season at saliva sampling. Using a bump hunting region-based approach we examined  
53 associations between saliva methylome measured using Illumina Infinium  
54 HumanMethylation450k array and wheezing between 6 and 18 months of age. We tested our  
55 main findings in independent publicly available datasets of childhood respiratory allergy and  
56 atopic asthma, with DNA methylation measured in different tissues and at different ages.

57 **Results** We identified one wheezing-associated differentially methylated region (DMR)  
58 spanning ten sequential CpG sites in the promoter-regulatory region of *PM20D1* gene (family  
59 wise error rate <0.05). The observed associations were enhanced in children born to atopic  
60 mothers. In the publicly available datasets, hypermethylation in the same region of *PM20D1*  
61 was consistently found at different ages and in all analysed tissues (cord blood, blood, saliva  
62 and nasal epithelia) of children with respiratory allergy/atopic asthma compared with controls.

63 **Conclusion** This study suggests that *PM20D1* hypermethylation is associated with early  
64 childhood wheezing. Directionally consistent epigenetic alteration observed in cord blood and  
65 other tissues at older ages in children with respiratory allergy and atopic asthma provides  
66 suggestive evidence that a long-term epigenetic modification, likely operating from birth, may  
67 be involved in childhood atopic phenotypes.

68 **Key words:** wheezing, asthma, children, epigenome-wide association study, saliva,  
69 epigenetics, infant, EWAS, DMR, PM20D1

## 70 **Introduction**

71 Over the past 20 years the concept of developmental origins of health and disease has become  
72 widely accepted and expanded to many common complex conditions, including, but not  
73 limited to, immunometabolic disorders, such as obesity, asthma, and cardiovascular diseases.<sup>1</sup>  
74 This theory states that intrauterine and early postnatal exposures can induce biological  
75 changes that influence later disease susceptibility. Alteration of the epigenome is one potential  
76 mechanism by which environmental exposures may cause structural and/or functional  
77 changes in cells, tissues, and organs of developing individuals.<sup>2</sup> Several in utero exposures,  
78 such as maternal smoking, maternal body mass index (BMI), and exposure to air pollution  
79 have been associated with alterations in infant DNA methylation, most often measured in cord  
80 blood.<sup>3-5</sup> Most of these exposures are also known to be associated with childhood wheezing  
81 and asthma.<sup>6-8</sup> In addition, postnatal environment, such as air pollution, early microbial  
82 exposures, and child's growth, may increase the risk of childhood wheezing disorders and  
83 asthma through epigenetic modifications.<sup>9-11</sup>

84 Although DNA methylation is the most studied epigenetic mechanism to date, post-  
85 translational histone modifications have also been implicated in the T-cells differentiation and  
86 airway remodelling, contributing to the epigenetic regulation of allergic phenotypes, including  
87 asthma.<sup>12,13,14</sup>

88 While asthma is a diagnostic category typically used from school age onwards, the disease  
89 often initiates much earlier, with wheezing episodes starting from early infancy.<sup>15</sup> Wheezing  
90 in early childhood, especially if accompanied with allergic sensitization or other atopic  
91 conditions, has been shown to be a strong predictor of asthma development and later lung  
92 function,<sup>16</sup> and thus is often studied in the context of large epidemiological studies.

93 Several previous studies reported associations with childhood wheezing and asthma of DNA  
94 methylation changes at a number of single methylation sites and differentially methylated

95 regions (DMRs).<sup>17-24</sup> A large cross-sectional EWAS of childhood asthma conducted within the  
96 MeDALL consortium identified an altered peripheral blood DNA methylation at 35 CpG  
97 sites; most of the associations were with school age asthma, while only one CpG site was  
98 associated with asthma at preschool age.<sup>23</sup> Importantly, all the associations were largely  
99 attenuated, but persisted, following adjustment for eosinophil count. In the same study, DNA  
100 methylation patterns at 14 out of the 35 identified CpG sites were replicated in peripheral  
101 blood of children with asthma from independent cohorts, while none of the identified CpG  
102 sites was associated with asthma when DNA methylation was measured in cord blood at  
103 birth.<sup>23</sup> A recent meta-analysis of epigenome-wide DNA methylation and school-age asthma  
104 identified 9 CpG sites and 35 DMRs in cord blood associated with childhood asthma.<sup>24</sup> These  
105 early DNA methylation markers were, however, not among much larger number of hits  
106 identified in a cross-sectional analysis within the same study where both asthma and blood  
107 DNA methylation were assessed in school-age children and adolescents.<sup>24</sup>

108 It is well established that DNA methylation is tissue-specific, and for diseases for which the  
109 local effect might be more pronounced than their systemic effect, such as asthma, the tissue  
110 selection for DNA methylation measurement is essential. Bronchial tissue is difficult to obtain  
111 in population studies, and most of the studies assessing epigenetic markers of wheezing and  
112 asthma focused on DNA methylation measured in blood samples. Some studies, however,  
113 found asthma methylomic markers in buccal, nasal and airway epithelial cells.<sup>17, 18, 20, 25</sup> As  
114 saliva is a candidate body fluid reflecting pathological changes in the airways during asthma  
115 development,<sup>26</sup> we aimed at investigating the associations between infant saliva methylome  
116 and early childhood wheezing within the Italian birth cohort study NINFEA.

## 117 **Methods**

### 118 *Study population*

119 The NINFEA (Nascita ed INFanzia: gli Effetti dell'Ambiente) study is an Italian web-based  
120 birth cohort that recruited approximately 7500 pregnant women during the period 2005-2016  
121 (<https://www.progettoninfea.it>). Members of the cohort are children of mothers who had  
122 access to the Internet, enough knowledge of Italian to complete online questionnaires, and  
123 volunteered to participate at any time during the pregnancy. At enrolment women completed  
124 the baseline questionnaire, and children are then followed up with five questionnaires  
125 completed by mothers 6 and 18 months after delivery, and when children turn 4, 7 and 10  
126 years of age. At the end of the 6-month questionnaire mothers were invited to donate their and  
127 their child's saliva samples using the Oragene™ DNA self-collection kits (OG-250, DNA  
128 Genotek, Inc., Ottawa, Ontario, Canada). Mothers who did not respond to this initial  
129 invitation were invited again at the end of the 18-month and 4-year questionnaire. The  
130 samples are stored in a biobank at -80°C. Additional information on recruitment, study  
131 advertisement and follow-up is available in the dedicated paper.<sup>27</sup>

132 We conducted a nested case-control study within the subset of the NINFEA children who met  
133 the following criteria: i) singleton child, ii) saliva sample collected between 6 and 18 months  
134 after birth (time window for the outcome assessment), iii) residence in the Province of Turin,  
135 and iv) born to mother who did not have asthma active during the index pregnancy.

136 Wheezing was assessed from the questionnaire completed 18 months after delivery using the  
137 standardized question from the International Study on Asthma and Allergies in Childhood  
138 (ISAAC), and was defined as at least one episode of wheezing or whistling in the chest  
139 occurred between 6 and 18 months of age.

140 At the time of sampling of cases and controls (database version 02.2014), there were 551  
141 children meeting the aforementioned inclusion criteria, of whom 79 had at least one wheezing  
142 episode between 6 and 18 months of age. Of the 79 children with wheezing, seven samples



143 with the lowest DNA content were excluded, and the remaining 72 cases were matched to  
144 controls by sex, season and age at saliva donation keeping a constant 1:1 ratio.

145 The NINFEA study protocol and subsequent amendments have been approved by the Ethical  
146 Committee of the San Giovanni Battista Hospital—CTO/CRF/Maria Adelaide Hospital,  
147 Turin, Italy (approval N.0048362). All women gave their informed consent at recruitment and  
148 when donating saliva samples.

#### 149 *DNA methylation data*

150 The methylation status of over 485,000 probes was measured using the Illumina Infinium  
151 HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA). Details on pre-processing  
152 of samples and data quality control can be found in the **Supporting Information file**  
153 (Methods, *DNA methylation measurement, data pre-processing, and quality control*). DNA  
154 methylation for each probe was expressed as beta value (ratio of methylated probe intensity to  
155 overall intensity, representing 0 to 100% methylation at each probe). After quality control and  
156 probes filtering 136 samples with DNA methylation measured in 421,782 probes remained for  
157 the analyses.

#### 158 *Statistical analyses*

159 To remove variance caused by batch and unmeasured technical effects, and to account for  
160 cell-type heterogeneity we performed surrogate variable analysis using the R package sva.<sup>28,29</sup>  
161 We conducted an exploratory epigenome-wide single-probe analysis using logistic regression  
162 models adjusted for matching variables (child's sex, age and season at saliva sampling) and  
163 for presence of siblings at birth, maternal age at delivery, maternal smoking after delivery as a  
164 proxy for child's exposure to passive smoking and day-care attendance between 6 and 18  
165 months of age. Bonferroni-corrected and Benjamini and Hochberg False-Discovery Rate  
166 (FDR)-corrected p-values were calculated to account for multiple testing.

167 Differentially methylated regions associated with early childhood wheezing were analysed  
168 using the R package *bumphunter*.<sup>30</sup> We clustered probes with a maximum distance of 1kb,  
169 and then with a *bumphunter* function fitted a linear model for each site with wheezing as the  
170 independent variable, adjusting for the aforementioned confounding and matching variables  
171 used in the exploratory epigenome-wide association analyses. The estimated coefficients of  
172 the case-control status were then smoothed within the clusters using running medians. We  
173 performed 1000 random bootstrap iterations to estimate an empirical null distribution. The  
174 regions of interest were chosen based on the cut-off threshold of 5% in the methylation beta  
175 value difference between cases and controls, retaining only regions with at least three probes  
176 within the region and with family-wise error rate (FWER) <0.05.

177 We further tested whether some of in utero exposures confounded the observed associations  
178 by performing additional adjustments for maternal history of atopy (atopic dermatitis and/or  
179 allergic rhinitis; note that mothers with history of asthma were a priori excluded from the  
180 study), maternal depression and/or anxiety and maternal genito-urinary and respiratory  
181 infections (otitis, sinusitis, throat infections, bronchitis and influenza) during the index  
182 pregnancy.

183 Differences between cases and controls in the mean DNA methylation at each CpG site  
184 mapping to a DMR identified by the *bumphunter* procedure were analysed using unpaired t-  
185 test. To explore the role of maternal atopy in the observed associations, these analyses were  
186 additionally stratified by maternal history of atopic disorders (allergic rhinitis and/or atopic  
187 dermatitis).

188 All analyses were performed using the statistical computing software R version 3.5.1.

#### 189 *Analysis of publicly available datasets*

190 To test our main findings in independent samples we identified five datasets from the Gene  
191 Expression Omnibus (GEO) database ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). Details on the search of

192 datasets containing epigenome-wide DNA methylation data and phenotypes related to  
193 childhood wheezing can be found in the **Supporting Information file** (Methods, *Search of*  
194 *publicly available datasets*).

195 In particular, we used three datasets from the project “Epigenetic profiling of children with  
196 respiratory allergy” (GSE110128), which includes children from a Belgian birth cohort with  
197 available cord blood samples, and blood and saliva samples collected at age 11 years.<sup>31</sup> DNA  
198 methylation from cord blood mononuclear cells (CBMCs), peripheral blood mononuclear  
199 cells (PBMCs) and saliva was measured using the Infinium HumanMethylation450 BeadChip  
200 array. This project includes processed DNA methylation data from 485,512 CpG sites for 233  
201 samples: 90 cord blood samples (30 respiratory allergy cases and 60 controls), 90 blood  
202 samples (29 respiratory allergy cases and 61 controls) and 53 saliva samples (33 respiratory  
203 allergy cases and 20 controls). Respiratory allergy was assessed when children were 10 years  
204 old and was defined as IgE sensitization to a mix of airborne allergens and self-report or  
205 doctor’s diagnosis of at least one of the following: asthma, hay fever, other types of rhinitis,  
206 wheezing, or runny nose in the past year and ever.<sup>31</sup>

207 The remaining two datasets (GSE40576 and GSE65163) include partially overlapping  
208 samples of African American or Hispanic with Dominican/Haitian background children from  
209 the Inner-City Asthma Consortium (USA). Atopic asthma was defined as persistent physician-  
210 diagnosed asthma with altered lung function parameters and positive skin prick test to at least  
211 one of a panel of indoor aeroallergens.<sup>19,20</sup> The GSE40576 dataset includes epigenome-wide  
212 DNA methylation measured in PBMCs from 6-12 years old children with atopic asthma  
213 (N=97) and healthy controls (N=97).<sup>19</sup> The GSE65163 dataset involves children with atopic  
214 asthma (N=36) and healthy controls (N=36) aged 9-12 years and epigenome-wide DNA  
215 methylation measured in nasal epithelial cells.<sup>20</sup> Both studies used Infinium  
216 HumanMethylation450 BeadChip array to measure DNA methylation.

217 We selected only probes located within the DMR identified in our study, calculated the  
218 differences in the mean DNA methylation between cases and controls (delta beta [ $\Delta$ beta]),  
219 and formally tested them using the unpaired t-test. We report one-sided p-values because the  
220 alternative hypotheses were determined by the findings from the NINFEA sample. Given the  
221 public nature of the datasets, we were not able to perform data quality checks and to correct  
222 for cell types, technical covariates and other potential confounders.

## 223 **Results**

### 224 *Epigenome-wide DNA methylation analysis of early childhood wheezing*

225 Descriptive statistics of the study population are shown in **Table 1**. Maternal age, educational  
226 level, nationality, and smoking after delivery were balanced between cases and controls.

227 Children with wheezing were more likely to be born to mothers with a history of atopic  
228 disorders ( $p=0.09$ ), to have siblings and to have attended day-care compared with controls  
229 (both  $p$ -values= $0.02$ ). Median age at saliva sampling was 10.3 months (interquartile range  
230 [IQR] 9.4-11.8 months).

231 In an exploratory epigenome-wide association analysis 25,042 (5.9%) probes were associated  
232 with childhood wheezing at the conventional  $p$ -value $<0.05$ , but none remained associated  
233 after adjustment for multiple comparisons (Bonferroni  $p$ -value $<1.19e-07$  and Benjamini-  
234 Hochberg FDR-corrected  $p$ -value $<0.05$ ). The Manhattan plot of the EWAS results is shown in

### 235 **Figure S1.**

236 Of the four DMRs identified by the bump hunting procedure (**Table S1**) only one region was  
237 associated with wheezing with a FWER $<0.05$  (**Table 2**). This DMR remained associated with  
238 childhood wheezing also after adjustment for maternal history of atopy, maternal depression  
239 and anxiety and maternal infections during pregnancy (data not shown).

240 The identified DMR consists of ten probes located on the chromosome 1 and maps to the  
241 promoter region of the *PM20D1* gene. All 10 probes were hypermethylated in cases compared

242 with controls, with seven probes having  $\Delta\text{beta}$  from 4.3% to 9.6% (all p-values<0.05, **Table**  
243 **3**, first column). The largest differences between cases and controls were identified in the  
244 *PM20D1* promoter region and in the region of the first exon.

245 When we stratified analyses on single CpG sites by maternal atopic disorders (**Table 3**) DNA  
246 methylation differences between cases and controls at 7 CpG sites located within the *PM20D1*  
247 promoter region and the region of the first exon were particularly increased in children born to  
248 atopic mothers, with  $\Delta\text{beta}$  ranging from 10% to 20%. The same differences were diluted, but  
249 still followed similar pattern, in children of non-atopic mothers.

#### 250 *Analysis of the publicly available datasets*

251 The difference in the mean DNA methylation between cases and controls in 10 probes located  
252 within the *PM20D1* calculated in the NINFEA sample and in the five public datasets are  
253 reported in **Figure 1** and **Table 4**. The *PM20D1* probes are ordered according to their base  
254 pair position. The seven most strongly associated probes in the NINFEA data were  
255 consistently associated with respiratory allergy in 11-year old children when DNA  
256 methylation was measured in CBMCs and PBMCs (CBMCs:  $\Delta\text{beta}$  8.4%-11.7%; PBMCs:  
257  $\Delta\text{beta}$  10.4%-13.6%; all one-sided p-values<0.05). An analogous direction of the associations,  
258 though lower in magnitude, was observed for DNA methylation measured in saliva of 11-year  
259 old children. It should be noted that the sample size of the saliva public dataset was small  
260 (half of the other two public datasets from the same project) limiting the power of this  
261 analysis.

262 Only two of the 10 probes from the *PM20D1* DMR were available in the nasal epithelia  
263 dataset of persistent atopic asthma in African-American 9-12 years old children, where the  
264 associations were even stronger ( $\Delta\text{beta}$  10%-11%, both one-sided p-values<0.005). Despite  
265 the strong associations observed in nasal samples, the differences in PBMCs of children from  
266 the same study were much smaller, although still in the same direction.

267 The three probes that had  $\Delta\beta$  below 1% (and a  $p$ -value $>0.05$ ) in the NINFEA sample had  
268 the greatest distance from the transcriptional start site and the lowest difference in the mean  
269 DNA methylation between cases and controls also in all public datasets (**Tables 3 and 4**).  
270 As GSE40576 PBMCs dataset was accompanied with paired gene expression data  
271 (GSE40732) we further examined whether differentially methylated sites impact *PM20D1*  
272 expression in healthy children (**Supporting Information file**, Results, *Correlation of*  
273 *differentially methylated sites with expression of PM20D1*). Although the *PM20D1* expression  
274 level was not associated with persistent atopic asthma in this dataset ( $p$ -value=0.382), there  
275 was a negative relationship between DNA methylation in all but one CpG site located within  
276 the DMR and *PM20D1* expression, with Spearman's rho ranging from -0.014 to -0.126 (all  $p$ -  
277 values $>0.05$ , **Supporting Information file, Figure S2**).

## 278 **Discussion**

279 We assessed early life saliva genome-wide patterns of DNA methylation related to wheezing  
280 between 6 and 18 months of age, and identified one DMR related to early childhood wheezing  
281 located in the promoter-regulatory region of *PM20D1* gene. The observed associations were  
282 particularly strong in children born to atopic mothers. Using a public dataset we also found  
283 that DNA methylation of this region, measured at 10 corresponding probes in children's blood  
284 and cord blood, was higher in respiratory allergy cases compared with controls, and that the  
285 same pattern of the association, although lower in magnitude, was present when methylation  
286 was measured in saliva samples of a subset of the same children. Another public dataset  
287 confirmed the observed associations in nasal epithelium from 9-12 years old children with  
288 atopic asthma, while the associations were weaker in PBMCs from a partially overlapping  
289 sample from the same study.

290 Our study is based on DNA methylation measured in early infancy, an approach that differs  
291 from most of the previous studies that measured methylation in cord blood or later in

292 childhood. Our exploratory epigenome-wide single site analysis showed no single CpG site  
293 associated with early wheezing, which is consistent with the findings from the MeDALL  
294 consortium, in which 14 CpG sites identified in the discovery study were replicated in  
295 peripheral blood from children with school-age asthma, while none of the CpG sites was  
296 replicated in analyses focused on early life, where DNA methylation was measured in cord  
297 blood.<sup>23</sup> On the other hand, CpG sites and DMRs identified in cord blood of children with  
298 school-age and adolescent asthma from the PACE consortium<sup>24</sup> do not include *PM20D1*,  
299 which may reflect more heterogeneous outcome definition in the PACE consortium, which is  
300 not strictly related to atopy. Our and the results from the PACE and MeDALL studies suggest  
301 that early life regional, rather than single site, DNA methylation differences are more likely to  
302 be associated with childhood atopic and asthmatic phenotypes. These findings also suggest  
303 that epigenetic markers for wheezing and asthma tend to develop later in childhood as a  
304 consequence of adverse postnatal exposures, or the disease itself, whose effects are not  
305 present at birth and in early infancy.

306 The sample size of our study limited the possibility to detect small-magnitude effect sizes that  
307 can have functional relevance for childhood and later adulthood health. In fact, the magnitude  
308 of the associations for wheezing- and asthma-candidate CpGs identified in several previous  
309 EWAS was also small.<sup>22,23,31</sup> As the power of our exploratory EWAS was relatively low, we  
310 used a more powerful approach that allowed us to identify a wheezing-related differentially  
311 methylated region located in *PM20D1* gene. This gene codes Peptidase M20 Domain  
312 Containing 1 enzyme that regulates production of N-fatty-acyl amino acids, which are a large  
313 family of more than 70 endogenous signalling molecules involved in pain and inflammation  
314 regulation, and their metabolites are important mediators of chronic airway inflammation in  
315 asthma.<sup>32</sup> Differential DNA methylation patterns at this gene have been previously associated  
316 with adult respiratory allergy<sup>33</sup> and maternal asthma.<sup>34</sup> In the study of Langie et al.<sup>33</sup>

317 cg11965913 located in the promoter region of *PM20DI* was hypermethylated in adult  
318 individuals with respiratory allergy compared with controls, with similar differential  
319 methylation profiles in PBMCs and saliva. Furthermore, hypomethylation of *PM20DI* in  
320 blood from 12-month old children was associated with maternal asthma and asthma  
321 medication use, maternal atopic status and serum IgE level.<sup>34</sup> In the NINFEA study, on the  
322 contrary, hypermethylation of *PM20DI* related to wheezing was much stronger in children  
323 born to atopic compared with non-atopic mothers. Although apparently contrasting, findings  
324 by Gunawardhana et al.<sup>34</sup> and our own findings both indicate that maternal atopy likely  
325 modify offspring atopic phenotypes through alterations in *PM20DI* DNA methylation.

326 In our analyses we accounted for the most important determinants of childhood wheezing,  
327 such as maternal age, smoking, child's siblings and day-care attendance. In particular, the  
328 presence of older siblings and day-care attendance, the proxies of early childhood respiratory  
329 tract infections and microbial exposure in general, were strongly positively associated with  
330 wheezing in our sample, but did not confound the observed association.

331 We acknowledge that we did not perform a "real" replication analysis with a well-planned  
332 replication group, and this results in important limitations. First, the phenotypes of the public  
333 datasets were different from the one analysed in our sample. Respiratory allergy in three  
334 public datasets covers several atopic conditions related to sensitization to airborne allergens,  
335 including asthma and wheezing at 10 years of age, while atopic asthma in two other datasets  
336 includes persistent asthma with impaired lung function and allergic sensitization. Although  
337 our phenotype, namely wheezing in early childhood, is a heterogeneous condition, often  
338 related to a relative narrowing of the airways during an acute viral infection,<sup>35</sup> it strongly  
339 predicts later asthma development.<sup>15</sup> As all phenotypes in the public datasets have an atopic  
340 characterization together with data on respiratory conditions, including wheezing and asthma,



341 we considered them as good candidates for replication of DNA methylation patterns related to  
342 wheezing in early childhood. This is further supported by the fact that atopic phenotypes,  
343 especially respiratory allergy and atopic asthma/wheezing, share multiple environmental and  
344 genetic risk factors, as well as immunological features.<sup>36</sup>

345 Second, the age of the outcome assessment in our study was different compared with the  
346 public datasets (18 months vs. 6-12 years). However, identification of wheezing-related  
347 epigenetic markers in early childhood, especially if they overlap with markers of later asthma  
348 or respiratory allergy, would allow identification of high-risk children, improve diagnostic  
349 and therapeutic approaches, and would contribute to the understanding of the involvement of  
350 epigenetics in asthma aetiology.

351 Albeit different traits analysed, the pattern of the association observed in **Figure 1** suggests  
352 that DNA methylation differences related to preschool wheezing and school-age respiratory  
353 allergy and atopic asthma is largely overlapping, which is not likely to be by chance. In fact,  
354 consistently increased DNA methylation within the *PM20D1* region was observed in cases  
355 compared with controls at different ages (at birth, 10 months in the NINFEA study, and 6-12  
356 years in other four public datasets), as well as in different biological materials analysed  
357 (CBMCs, saliva, nasal epithelium and PBMCs). Interestingly, very similar DNA methylation  
358 differences related to childhood respiratory allergy were observed between cord blood and  
359 blood at 11 years of age in the Belgian study, suggesting the stability of DNA methylation  
360 patterns after birth. Although it has been shown that DNA methylation is influenced also by  
361 postnatal environmental exposures, such as early microbial exposures (e.g. DNA methylation  
362 of several asthma-candidate genes was strongly influenced by early farming exposure),<sup>37</sup> our  
363 findings indicate that this is not a general rule. DNA methylation differences in *PM20D1*

364 related to atopic respiratory phenotypes seem to be determined already at birth and stable over  
365 time, suggesting a rather weak impact of postnatal exposures.

366 One might question why *PM20D1* was not among the top genes identified by previous studies  
367 which public datasets we used to replicate our findings.<sup>19,20,31</sup> However, this is not surprising,  
368 as we based our analyses of public datasets on a candidate gene approach compared with the  
369 epigenome-wide approach with multiple comparison adjustments reported by previous  
370 studies.<sup>19,20,31</sup> On the other hand, it should be also noted that despite the smaller sample size,  
371 the study by Langie et al.<sup>31</sup> and by Yang et al.<sup>20</sup> identified many more DMRs compared with  
372 our study, and that these DMRs do not overlap with the one DMR identified in our study. One  
373 possible explanation could be that the methylation-outcome associations were explored at  
374 older ages in the previous two studies which public data was used in our study;<sup>20,31</sup> this  
375 enhanced the possibility to identify a larger number of DMRs, which are also more likely to  
376 be a result either of the disease itself or of the accumulating environmental exposures in the  
377 first 12 years of life.

378 One of the limitations of our study is the fact that the time-sequence between DNA  
379 methylation measurement and wheezing cannot be clearly established as the period of saliva  
380 sample collection (median age 10.3 months) partially overlaps with the period of wheezing  
381 assessment (6-18 months). Although the *PM20D1* DMR could also be a consequence of  
382 wheezing disorders or postnatal environmental influences, similar pattern of association  
383 observed in public cord blood dataset speaks against reverse causation and supports the role  
384 of prenatal factors.

385 Findings from previous studies suggest that the associations observed between DNA  
386 methylation and asthma are largely driven by eosinophils<sup>22,23</sup> that are known to be involved in  
387 airway inflammation present in T-helper-2-type asthmatic phenotypes.<sup>38</sup> However, we used a

388 quite robust method for cell composition adjustment and we believe that the differences  
389 observed in *PM20D1* are not likely to be driven by eosinophil count. This is further supported  
390 by the findings from the publicly available datasets in which the associations were observed  
391 in CBMCs and PBMCs that do not contain eosinophils.

392 The type of tissue used for DNA methylation measurements is the most innovative element of  
393 our study. In the NINFEA study infant samples were collected using saliva sponges at  
394 approximately 10 months of age. This non-invasive method of genomic DNA collection is  
395 becoming an alternative to the blood DNA collection in large-scale studies, especially if  
396 performed on infants and young children. However, there are two aspects that need to be  
397 considered when using saliva as a material for epigenetic studies. First, saliva has a  
398 heterogeneous cell composition that, in addition to leukocytes, contains a large proportion of  
399 epithelial cells that might influence DNA methylation measurements. So far, saliva cell type  
400 composition is poorly characterized, and several reference-free methods have been developed  
401 to correct for cell type heterogeneity in EWAS.<sup>28,29</sup> Using simulated scenarios, a study of  
402 McGregor et al.<sup>29</sup> found that surrogate variable analysis had the best performance; thus we  
403 used this method to account for cell type composition, batch effects and technical residual  
404 confounding. We, therefore, believe that the association observed in our study is not driven by  
405 differential cell type composition. Second, blood and peripheral material, such as saliva,  
406 buccal or nasal epithelium, may not be representative of DNA methylation status at central  
407 organs involved in a disease pathophysiology. However, various studies identified asthma-  
408 and wheezing-related epigenetic marks in several tissues, including nasal and buccal  
409 epithelium, whole peripheral blood, cord blood, as well as in mononuclear cells from cord and  
410 peripheral blood.<sup>17-24,39</sup> While blood may be valuable for studying common immunologic  
411 processes, nasal and saliva epithelium, being probably closer to the methylation status of the  
412 bronchial mucosa than blood or cord blood DNA methylation, are likely better surrogate

413 tissues for studies focused on local inflammation and environmental influences in asthma.  
414 These two tissues are also easily accessible and appear to reflect pathological changes in the  
415 airways during asthma development.<sup>26</sup>

416 In conclusion, we have shown that higher DNA methylation in the promoter-regulatory region  
417 of *PM20D1* from infant saliva was associated with the occurrence of early childhood  
418 wheezing, especially in children of atopic mothers. Directionally consistent epigenetic  
419 alteration observed in cord blood of children with respiratory allergy suggests that this  
420 modification might operate from birth, while the confirmation in blood and nasal epithelia of  
421 school-aged children indicates possibly long-term changes. The *PM20D1* promoter region  
422 was previously implicated in adult respiratory allergy, suggesting its potential role in  
423 aetiology of atopic phenotypes, including asthma.

#### 424 **Acknowledgements**

425 The authors are grateful to all the participants of the NINFEA birth cohort.

#### 426 **References**

- 427 1. Barker DJ. The origins of the developmental origins theory. *J Intern Med.*  
428 2007;**261**:412–417.
- 429 2. Godfrey KM, Lillycrop KA, Burdge GC, Gluckman PD, Hanson MA. Epigenetic  
430 mechanisms and the mismatch concept of the developmental origins of health and  
431 disease. *Pediatr Res.* 2007;**61**:5R-10R.
- 432 3. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA  
433 methylation in newborns and maternal smoking in pregnancy: genome-wide  
434 consortium meta-analysis. *Am J Hum Genet.* 2016;**98**:680-696.

- 435 4. Oelsner KT, Guo Y, To SB, Non AL, Barkin SL. Maternal BMI as a predictor of  
436 methylation of obesity-related genes in saliva samples from preschool-age Hispanic  
437 children at-risk for obesity. *BMC Genomics*. 2017;**18**:57.
- 438 5. Gruzieva O, Xu CY, Breton CV, Annesi-Maesano I, Antó JM, Auffray C, et al.  
439 Epigenome-wide meta-analysis of methylation in children related to prenatal NO<sub>2</sub> air  
440 pollution exposure. *Environ Health Perspect*. 2017;**125**:104–110.
- 441 6. Silvestri M, Franchi S, Pistorio A, Petecchia L, Rusconi F. Smoke exposure,  
442 wheezing, and asthma development: a systematic review and meta-analysis in  
443 unselected birth cohorts. *Pediatr Pulmonol*. 2015;**50**:353-362.
- 444 7. Rusconi F, Popovic M. Maternal obesity and childhood wheezing and asthma.  
445 *Paediatr Respir Rev*. 2017;**22**:66-71.
- 446 8. Hehua Z, Quing C, Shanyan G, Quijun W, Yuhong Z. The impact of prenatal exposure  
447 to air pollution on childhood wheezing and asthma: A systematic review. *Environ Res*.  
448 2017;**159**:519-530.
- 449 9. McErlean P, Favoreto S Jr, Costa FF, et al. Human rhinovirus infection causes  
450 different DNA methylation changes in nasal epithelial cells from healthy and  
451 asthmatic subjects. *BMC Med Genomics*. 2014;7:37.
- 452 10. McDade TW. Early environments and the ecology of inflammation. *Proc Natl Acad*  
453 *Sci U S A*. 2012;109:17281-17288.
- 454 11. Clifford RL, Jones MJ, MacIsaac JL, et al. Inhalation of diesel exhaust and allergen  
455 alters human bronchial epithelium DNA methylation. *J Allergy Clin Immunol*.  
456 2017;139:112-121.
- 457 12. Alaskhar Alhamwe B, Khalaila R, Wolf J, et al. Histone modifications and their role in  
458 epigenetics of atopy and allergic diseases. *Allergy Asthma Clin Immunol*. 2018;14:39.

- 459 13. Potaczek DP, Harb H, Michel S, Alhamwe BA, Renz H, Tost J. Epigenetics and  
460 allergy: from basic mechanisms to clinical applications. *Epigenomics*. 2017;9(4):539-  
461 571.
- 462 14. Harb H, Irvine J, Amarasekera M, et al. The role of PKC $\zeta$  in cord blood T-cell  
463 maturation towards Th1 cytokine profile and its epigenetic regulation by fish oil.  
464 *Biosci Rep*. 2017;37(2).
- 465 15. Kabesch M. Early origins of asthma (and allergy). *Mol Cell Pediatr*. 2016;3:31.
- 466 16. Belgrave DC, Buchan I, Bishop C, Lowe L, Simpson A, Custovic A. Trajectories of  
467 lung function during childhood. *Am J Respir Crit Care Med*. 2014;189:1101-1109.
- 468 17. Baccarelli A, Rusconi F, Bollati V, et al. Nasal cell DNA methylation, inflammation,  
469 lung function and wheezing in children with asthma. *Epigenomics*. 2012;4:91-100.
- 470 18. Murphy TM, Wong CC, Arseneault L, Burrage J, Macdonald R, Hannon E, et al.  
471 Methyloomic markers of persistent childhood asthma: a longitudinal study of asthma-  
472 discordant monozygotic twins. *Clin Epigenetics*. 2015;7:130.
- 473 19. Yang IV, Pedersen BS, Liu A, O'Connor GT, Teach SJ, Kattan M, et al. DNA  
474 methylation and childhood asthma in the inner city. *J Allergy Clin Immunol*.  
475 2015;136:69-80.
- 476 20. Yang IV, Pedersen BS, Liu AH, O'Connor GT, Pillai D, Kattan M, et al. The nasal  
477 methylome and childhood atopic asthma. *J Allergy Clin Immunol*. 2017;139:1478-  
478 1488.
- 479 21. DeVries A, Wlasiuk G, Miller SJ, Bosco A, Stern DA, Lohman IC, et al. Epigenome-  
480 wide analysis links SMAD3 methylation at birth to asthma in children of asthmatic  
481 mothers. *J Allergy Clin Immunol*. 2017;140:534-542.

- 482 22. Arathimos R, Suderman M, Sharp GC, Burrows K, Granell R, Tilling K, et al.  
483 Epigenome-wide association study of asthma and wheeze in childhood and  
484 adolescence. *Clin Epigenetics*. 2017;**9**:112.
- 485 23. Xu CJ, Söderhäll C, Bustamante M, Baiz N, Gruziova O, Gehring U, et al. DNA  
486 methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir*  
487 *Med*. 2018;**6**:379-388.
- 488 24. Reese SE, Xu CJ, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, et al.  
489 Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J Allergy*  
490 *Clin Immunol*. Published Online: December 20, 2018 (doi:  
491 10.1016/j.jaci.2018.11.043).
- 492 25. Stefanowicz D, Hackett TL, Garmaroudi FS, Günther OP, Neumann S, Sutanto EN, et  
493 al. DNA methylation profiles of airway epithelial cells and PBMCs from healthy,  
494 atopic and asthmatic children. *PLoS One*. 2012;**7**:e44213.
- 495 26. Langie SAS, Moisse M, Declerck K, Koppen G, Godderis L, Vanden Berghe W, et al.  
496 Salivary DNA methylation profiling: aspects to consider for biomarker identification.  
497 *Basic Clin Pharmacol Toxicol*. 2017;**121**:93-101.
- 498 27. Firestone R, Cheng S, Pearce N, Douwes J, Merletti F, Pizzi C, et al. Internet-based  
499 birth-cohort studies: is this the future for epidemiology? *JMIR Res Protoc*. 2015;**4**:e71.
- 500 28. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing  
501 batch effects and other unwanted variation in high-throughput experiments.  
502 *Bioinformatics*. 2012;**28**:882–883.
- 503 29. McGregor K, Bernatsky S, Colmegna I, Hudson M, Pastinen T, Labbe A, et al. An  
504 evaluation of methods correcting for cell-type heterogeneity in DNA methylation  
505 studies. *Genome Biol*. 2016;**17**:84.

- 506 30. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting  
507 to identify differentially methylated regions in epigenetic epidemiology studies. *Int J*  
508 *Epidemiol.* 2012;**41**:200-209.
- 509 31. Langie SAS, Moisse M, Szarc Vel Szic K, Van Der Plas E, Koppen G, De Prins S, et  
510 al. GLI2 promoter hypermethylation in saliva of children with a respiratory allergy.  
511 *Clin Epigenetics.* 2018;**10**:50.
- 512 32. Wendell SG, Baffi C, Holguin F. Fatty acids, inflammation, and asthma. *J Allergy*  
513 *Clin Immunol.* 2014;**133**:1255-1264.
- 514 33. Langie SA, Szarc Vel Szic K, Declerck K, Traen S, Koppen G, Van Camp G, et al.  
515 Whole-genome saliva and blood DNA methylation profiling in individuals with a  
516 respiratory allergy. *PLoS One.* 2016;**11**:e0151109.
- 517 34. Gunawardhana LP, Baines KJ, Mattes J, Murphy VE, Simpson JL, Gibson PG.  
518 Differential DNA methylation profiles of infants exposed to maternal asthma during  
519 pregnancy. *Pediatr Pulmonol.* 2014;**49**:852-862.
- 520 35. Lasso-Pirot A, Delgado-Villalta S, Spanier AJ. Early childhood wheezers: identifying  
521 asthma in later life. *J Asthma Allergy.* 2015;**8**:63-73.
- 522 36. Lazic N, Roberts G, Custovic A, Belgrave D, Bishop CM, Winn J, et al. Multiple  
523 atopy phenotypes and their associations with asthma: similar findings from two birth  
524 cohorts. *Allergy.* 2013;**68**(6):764-70.
- 525 37. Michel S, Busato F, Genuneit J, et al. Farm exposure and time trends in early  
526 childhood may influence DNA methylation in genes related to asthma and allergy.  
527 *Allergy.* 2013;**68**(3):355-364.
- 528 38. Foster PS, Maltby S, Rosenberg HF, Tay HL, Hogan SP, Collison AM, et al.  
529 Modeling TH 2 responses and airway inflammation to understand fundamental  
530 mechanisms regulating the pathogenesis of asthma. *Immunol Rev.* 2017;**278**:20-40.



531 39. Harb H, Alashkar Alhamwe B, Garn H, Renz H, Potaczek DP. Recent developments  
532 in epigenetics of pediatric asthma. *Curr Opin Pediatr.* 2016;28:754-763.

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556 **Table 1. Characteristics of the study population**

Characteristics	<b>Cases</b>	<b>Controls</b>	p-value <sup>§</sup>
	N (%) <sup>†</sup> or Mean (SD) <sup>‡</sup>	(N, %) <sup>†</sup> or Mean (SD) <sup>‡</sup>	
Total number	68 (50.0)	68 (50.0)	-
<i>Maternal age (years)</i>	35.3 (3.8)	34.1 (4.4)	0.10
<i>Highly educated mothers<sup>¶</sup></i>	47 (70.2)	51 (75.0)	0.53
<i>Mother born outside Italy</i>	3 (4.4)	3 (4.4)	1.00
<i>Maternal atopic dermatitis and/or allergic rhinitis</i>	21 (31.8)	13 (19.1)	0.09
<i>Maternal depression and/or anxiety</i>	9 (13.6)	10 (14.7)	0.86
<i>Genito-urinary infections during pregnancy</i>	10 (15.6)	19 (28.4)	0.08
<i>Respiratory infections during pregnancy<sup>††</sup></i>	22 (34.4)	18 (26.9)	0.35
<i>≥1 sibling</i>	27 (40.9)	15 (22.4)	0.02
<i>Child's exposure to passive smoking</i>	7 (10.3)	5 (7.4)	0.55
<i>Day-care attendance</i>	45 (66.2)	31 (45.6)	0.02
<i>Child's eczema 0-18 months</i>	10 (14.7)	13 (19.1)	0.49
<b>Matching variables</b>			
<i>Sex</i>			
	Males	38 (55.9)	38 (55.9)
<i>Season of saliva sampling</i>			
	Autumn or winter	32 (47.1)	32 (47.1)
<i>Age at saliva sampling (months)<sup>‡‡</sup></i>	10.3 (9.3-11.7)	10.4 (9.4-	

<sup>†</sup>Total numbers may vary due to missing data

<sup>‡</sup>SD – Standard Deviation

<sup>§</sup>Chi-square test for categorical variables, unpaired t test or Wilcoxon-Mann-Whitney test for continuous variables

<sup>¶</sup>University degree or higher

<sup>††</sup>Otitis and/or sinusitis and/or throat infections and/or bronchitis and/or influenza during pregnancy

<sup>‡‡</sup>Median (Interquartile Range – IQR)

557

558

559

560

561

**Table 2. Differentially methylated region (DMR) associated with early childhood wheezing**

<b>Cluster number</b>	<b>Chromosome</b>	<b>Start Position</b>	<b>End Position</b>	<b>Delta <math>\beta</math><sup>†</sup></b>	<b>Number of probes</b>	<b>p-value</b>	<b>FWER<sup>‡</sup></b>	<b>Nearest gene<sup>§</sup></b>
12362	1	205818484	205819609	0.077	10	0.006	0.028	<i>PM20D1</i>

<sup>†</sup> The average difference in DNA methylation between cases and controls

<sup>‡</sup> Family-wise error rate from 1000 bootstraps

<sup>§</sup> Genome coordinates are given with respect to hg19 genome assembly

**Table 3. Associations of DNA methylation in 10 probes, located within the differentially methylated region in *PM20D1* with childhood wheezing, stratified by maternal history of atopic disorders, ordered by the base pair position.**

Probe	Base pair position <sup>†</sup>	Genic location <sup>‡</sup>	NINFEA (N=136)		Maternal atopic disorders <sup>††</sup>			
					No (N=100)		Yes (N=34)	
			Delta $\beta^{\S}$ (%)	p-value <sup>¶</sup>	Delta $\beta^{\S}$ (%)	p-value <sup>¶</sup>	Delta $\beta^{\S}$ (%)	p-value <sup>¶</sup>
cg07157834	205819609	TSS1500	0.77	0.587	0,99	0,524	1,02	0,724
cg24503407	205819492	TSS1500	4.25	0.047	2,95	0,226	7,77	0,060
cg07167872	205819463	TSS200	5.19	0.018	3,52	0,156	10,58	0,014
cg11965913	205819406	TSS200	8.93	0.009	5,73	0,149	18,70	0,005
cg14893161	205819251	5'UTR;1stExon	9.15	0.009	5,68	0,156	19,27	0,005
cg14159672	205819179	1stExon	9.58	0.009	6,61	0,116	18,07	0,012
cg26354017	205819088	1stExon	9.14	0.004	5,38	0,144	20,33	0,002
cg17178900	205818956	Body	6.36	0.006	4,06	0,129	13,98	0,003
cg06815965	205818668	Body	0.40	0.576	0,54	0,545	0,91	0,351
cg03461704	205818484	Body	0.15	0.809	-0,06	0,938	0,67	0,667

<sup>†</sup> Genome coordinates are given with respect to hg19 genome assembly

<sup>‡</sup> According to UCSC reference gene information

<sup>§</sup> The difference in the mean DNA methylation between cases and controls (%)

<sup>¶</sup> P-values based on a two-sided Student's t-test for the difference in the mean DNA methylation between cases and controls.

<sup>††</sup> Maternal history of allergic rhinitis and/or atopic dermatitis

**Table 4. Associations of DNA methylation in 10 probes, located within the differentially methylated region in *PM20D1*, with wheezing in the NINFEA sample and with childhood respiratory allergy/atopic asthma, in the five publicly available datasets, ordered by the base pair position.**

Probe	NINFEA (wheezing)		GSE110128 (respiratory allergy)				GSE40576 (atopic asthma)		GSE65163 (atopic asthma)			
	Saliva 6-18 months (N=136)		Saliva 11 years (N=53)		Blood <sup>§</sup> 11 years (N=90)		Cord blood <sup>¶</sup> (N=90)		Blood <sup>§</sup> 6-12 years (N=194)		Nasal epithelia 9-12 years (N=72)	
	Delta $\beta^{\dagger}$ (%)	p-value	Delta $\beta^{\dagger}$ (%)	p-value <sup>‡</sup>	Delta $\beta^{\dagger}$ (%)	p-value <sup>‡</sup>	Delta $\beta^{\dagger}$ (%)	p-value <sup>‡</sup>	Delta $\beta^{\dagger}$ (%)	p-value <sup>‡</sup>	Delta $\beta^{\dagger}$ (%)	p-value <sup>‡</sup>
cg07157834	0.77	0.587	-2.22	0.761	6.69	0.004	1.14	0.074	0.98	0.181		NA
cg24503407	4.25	0.047	7.27	0.116	12.06	0.009	8.58	0.031	2.06	0.166		NA
cg07167872	5.19	0.018	9.19	0.066	12.17	0.005	11.66	0.006	3.47	0.062		NA
cg11965913	8.93	0.009	4.80	0.264	13.59	0.009	10.43	0.027	2.53	0.137		NA
cg14893161	9.15	0.009	4.16	0.276	10.44	0.024	8.89	0.037	1.92	0.157		NA
cg14159672	9.58	0.009	5.55	0.215	12.12	0.012	10.42	0.021	3.00	0.094		NA
cg26354017	9.14	0.004	5.61	0.214	11.89	0.012	10.71	0.020	2.94	0.109	10.36	0.005
cg17178900	6.36	0.006	7.81	0.130	11.43	0.019	8.39	0.044	2.42	0.149	10.96	0.001
cg06815965	0.40	0.576	-1.18	0.997	0.73	0.016	0.75	0.051	-0.15	0.771		NA
cg03461704	0.15	0.809	0.15	0.464	0.44	0.362	0.67	0.339	-0.19	0.610		NA

<sup>†</sup> The difference in the mean DNA methylation between cases and controls (%)

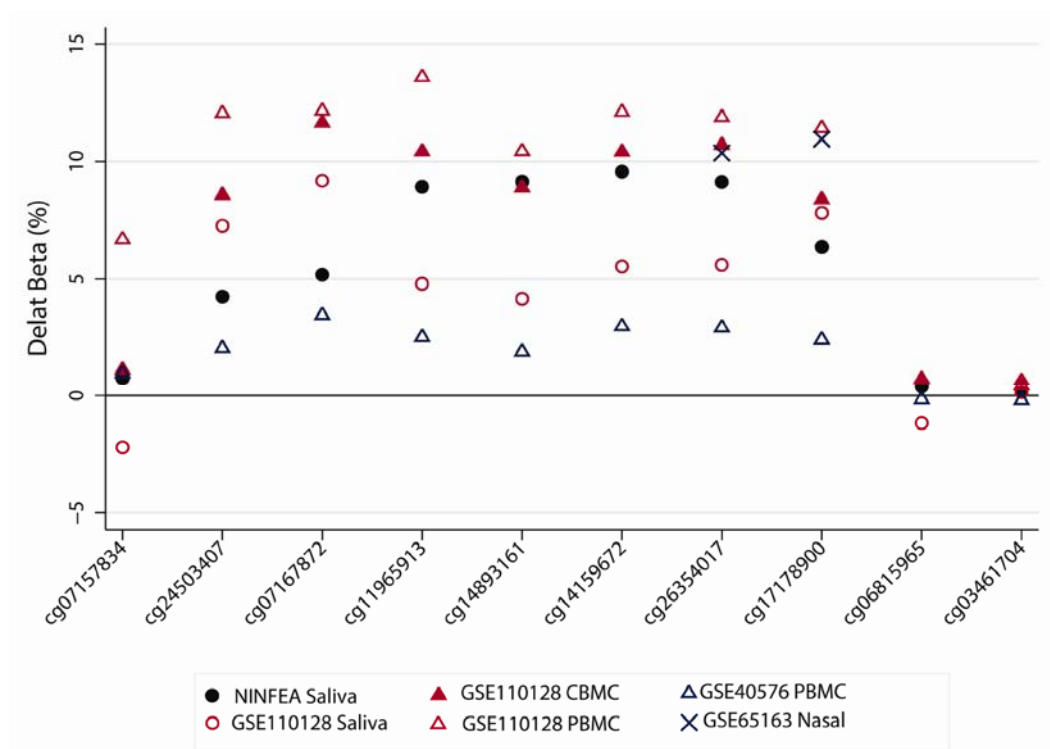
<sup>‡</sup> P-values based on a one-sided Student's t-test for the difference in the mean DNA methylation between cases and controls.

<sup>§</sup> Peripheral blood mononuclear cell (PBMC) fraction

<sup>¶</sup> Cord blood mononuclear cell (CBMC) fraction

**Figure 1. Associations of DNA methylation at 10 CpGs from the *PM20D1* DMR with early childhood wheezing in the NINFEA sample and with childhood respiratory allergy and atopic asthma from the public datasets.** The corresponding CpGs are reported on the x-axis, while the y-axis represents DNA methylation difference between wheezing/respiratory allergy/atopic asthma cases and controls in the NINFEA sample (full black circle), public saliva dataset (hollow red circle), public blood datasets (blue and red hollow triangles), public cord blood dataset (red full triangles) and public nasal epithelia dataset (blue Xs).

**Figure 1**  
 Associations of DNA methylation at 10 CpGs from the *PM20D1* DMR with early childhood wheezing in the NINFEA sample and with childhood respiratory allergy and atopic asthma from the public data sets. The corresponding CpGs are reported on the x-axis, while the y-axis represents DNA methylation difference between wheezing/respiratory allergy/atopic asthma cases and controls in the NINFEA sample (full black circle), public saliva data set (hollow red circle), public blood data sets (blue and red hollow triangles), public cord blood data set (red full triangles) and public nasal epithelia data set (blue Xs)



**Differentially methylated DNA regions in early childhood wheezing: an  
epigenome-wide study using saliva**

**SUPPORTING INFORMATION**

## Methods

### *DNA methylation measurement, data pre-processing, and quality control*

DNA was extracted from 144 saliva sponge samples using an automated protocol and the Oragene™ Purifier Solution (DNA Genotek, Inc., Ottawa, Ontario, Canada). Genomic DNA (500 ng) was bisulphite-converted using the EZ DNA Methylation-Gold™ Kit (Zymo Research, CA, USA) following the manufacturer's protocol.

The methylation status of over 485,000 probes was measured using the Infinium HumanMethylation450 BeadChip array (Illumina, Inc., San Diego, CA). The matched pairs were placed on the same chip to minimize confounding by batch. Data pre-processing steps were performed in R statistical computing software version 3.4.1<sup>1</sup> using the *minfi* package.<sup>2</sup> DNA methylation for each probe was expressed as beta-value (ratio of methylated probe intensity to overall intensity, representing 0 to 100% methylation at each probe).

Three samples with more than 1% of the CpGs with detection p-value>0.01 together with their respective matched pairs were excluded from the study (N=6). An additional matched pair was excluded because of sex mismatch (N=2).

To minimize the unwanted variation within and between samples the preprocessQuantile function was used.<sup>3</sup> This function implements a stratified quantile normalisation procedure that is applied to the methylated and unmethylated signal intensities separately, and takes into account the different probe types.<sup>3</sup> This procedure does not apply background correction, but very small intensities close to zero are thresholded using the *minfi* function `fixMethOutliers`.<sup>2</sup> We filtered out probes that failed in one or more samples based on detection p-value (detection p-value>0.01; N=10,902), probes located on the sex chromosomes (N=10,831), all probes containing a single-nucleotide polymorphism at the CpG interrogation and/or at the single nucleotide extension for any minor allele frequency (using the *minfi* function `dropLociWithSnps`,<sup>2</sup> N=15,654), and polymorphic CpG probes that have been demonstrated



to map to multiple places in the genome<sup>4</sup> (N= 26,343 probes). The final data set included 136 samples with DNA methylation measured in 421,782 probes.

### *Search of publicly available datasets*

On November 15<sup>th</sup> 2018 we searched publicly available datasets containing epigenome-wide DNA methylation data and phenotypes related to early childhood wheezing. In particular, we used the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) using the following field query “child\* [All Fields] AND (“asthma” [All Fields] OR wheez\* [All Fields] OR “allergy” [All Fields] OR atop\* [All Fields]) AND (“Methylation profiling by genome tiling array”[DataSet Type] AND “Homo sapiens”[Organism])”. Of the two available datasets of childhood respiratory allergy and asthma (GSE110128 and GSE85228, respectively), in order to have comparable CpG sites we excluded the dataset GSE85228 because of the different DNA methylation measurement platform compared with our study. The remaining GSE110128 dataset consists of three subsets that include cord blood, and peripheral blood and saliva<sup>5</sup> DNA methylation measured at 11 years of age in children with respiratory allergy and respective controls.

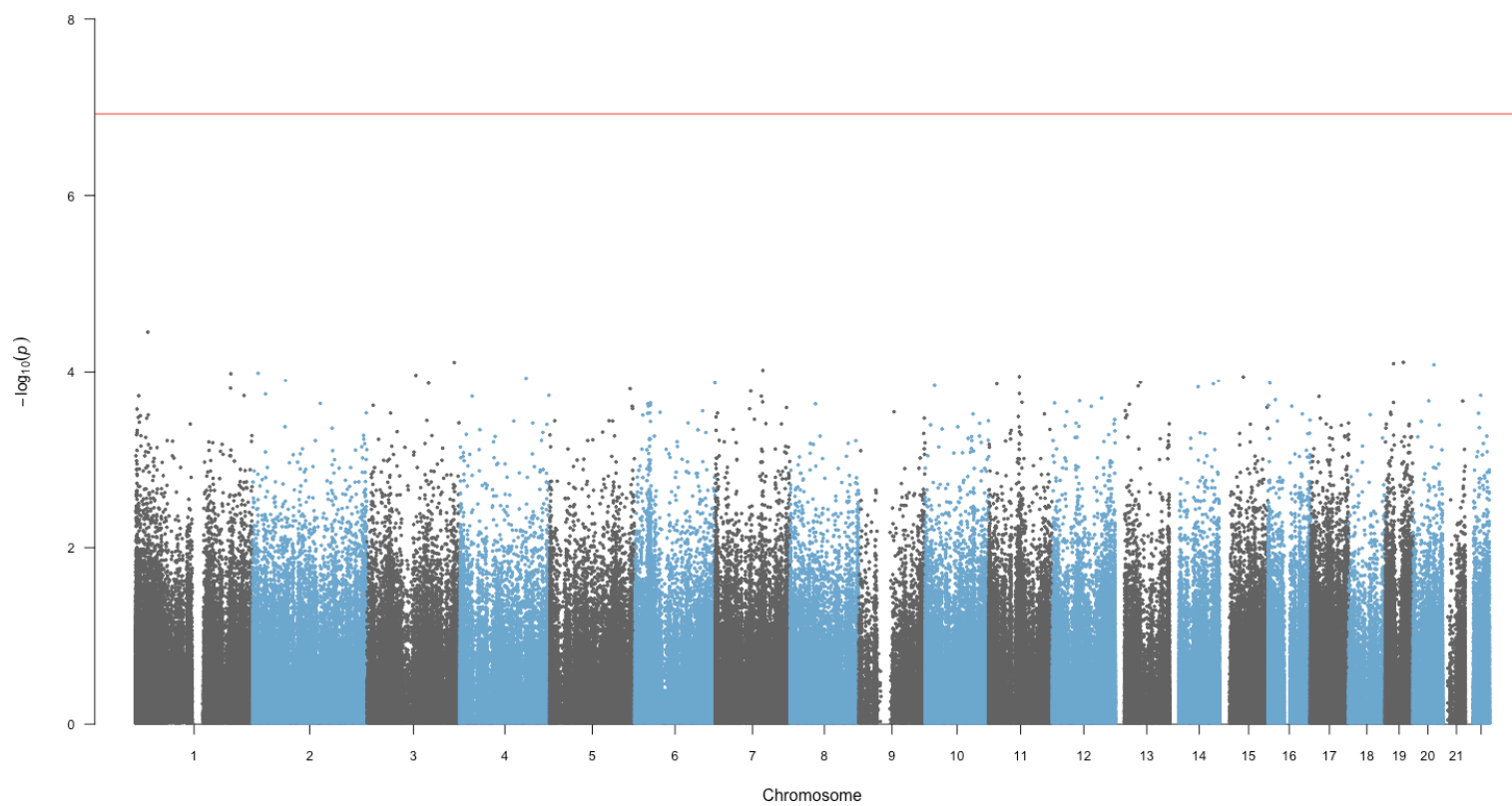
With further free text query (“(asthma OR wheez\* OR allerg\* OR atop\*) AND “DNA methylation” AND child\*”) we identified additional three datasets based on Homo sapiens samples: two nasal epithelia datasets (one of childhood asthma GSE109446, and one of childhood atopic asthma GSE65163) and one peripheral blood dataset of childhood atopic asthma (GSE40576). The dataset GSE109446 was excluded from our analyses, as it is a sibling-control study,<sup>6</sup> which is particularly difficult to study in the context of DNA methylation.

Therefore, to test our main findings in independent samples we used five datasets: i) three GSE110128 datasets of European children (Belgium)<sup>5</sup> with childhood respiratory allergy at 10 years of age and their respective controls consisting of cord blood DNA methylation (N=90), and saliva (N=53) and peripheral blood (N=90) DNA methylation measured at 11 years of age, ii) one GSE65163 dataset (N=72) of African American children, with persistent

atopic asthma at 9-12 years of age and healthy controls, containing epigenome-wide DNA methylation measured from nasal epithelial cells,<sup>7</sup> and iii) GSE40576 dataset (N=194) of African American or Hispanic with Dominican/Haitian background children aged 6-12 years with persistent atopic asthma and healthy controls, with epigenome-wide DNA methylation measured from peripheral blood mononuclear cells (PBMCs).<sup>8</sup> Although not clearly specified in the original studies, the last two datasets include DNA methylation that was most likely measured at the same age as the outcome.<sup>7,8</sup>

## Results

**Figure S1. Manhattan plot of the early childhood wheezing EWAS analysis.** The x-axis is the position of each CpG site by chromosome, and the y-axis is the negative log<sub>10</sub> of the p-value for association of each CpG site with early childhood wheezing. The red line indicates the Bonferroni-corrected epigenome-wide threshold ( $p=1.19e-$



07).

**Table S1. Differentially methylated regions (DMRs) associated with early childhood wheezing at 5% in the methylation beta value difference between cases and controls**

Cluster number	Chromosome	Start Position	End Position	Delta $\beta$ †	Number of probes	p-value	FWER ‡	Nearest gene §	Description	Distance ¶
12362	1	205818484	205819609	0.077	10	0.006	0.028	PM20D1	Overlaps 5'	0
121572	6	30070074	30070109	-0.054	5	0.443	0.838	PPP1R11	Downstream	35142
57141	16	56659552	56659964	0.054	5	0.454	0.845	MT1M	Inside intron	8179
114918	5	102898223	102898729	0.052	5	0.521	0.885	NUDT12	Overlaps 5'	0

† Delta  $\beta$  - the average DNA methylation differences between cases and controls

‡ Family-wise error rate from 1000 bootstraps

§ Genome coordinates are given with respect to hg19 genome assembly

¶ The distance to the 5' end of the nearest transcript (base pairs)

### ***Correlation of differentially methylated sites with expression of PM20D1***

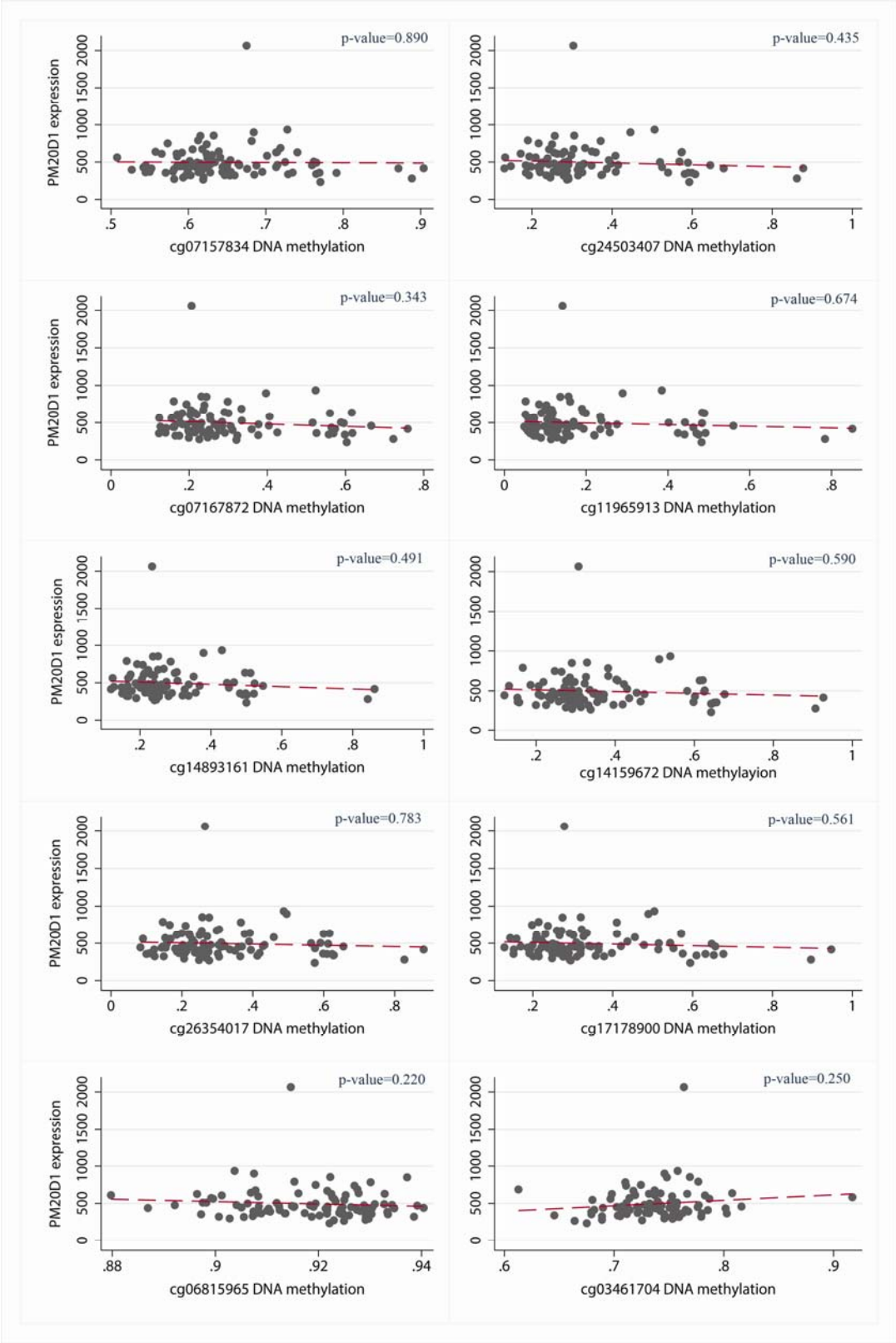
Two of the five public DNA methylation datasets that we used to replicate our differentially methylated region (GSE40576 and GSE65163) have publicly available paired gene expression data. The publicly available gene expression dataset (GSE65204), related to the GSE65163 DNA methylation data, do not contain the expression for the gene under study, and was thus not analyzed in our study.

The project “DNA Methylation Changes and Childhood Asthma in the Inner City” (SuperSeries GSE40736) includes paired DNA methylation and gene expression data from 6-12 years old children with atopic asthma (N=97) and healthy controls (N=97). DNA methylation GSE40576 dataset involves epigenome-wide DNA methylation measured in PBMCs using Infinium HumanMethylation450 BeadChip array, while GSE40732 dataset consists of gene expression data measured using Nimblegen Human Gene Expression array (12x135, Roche Nimblegen, Madison, WI).<sup>7</sup>

To explore to what extent DNA methylation influences *PM20D1* expression, we used only healthy controls (N=97) to relate DNA methylation at the identified region to the *PM20D1* expression using the Spearman correlation coefficients. In the same dataset we also tested the difference in the gene expression between atopic asthma and controls using Wilcoxon-Mann-Whitney test.

There was a negative relationship between DNA methylation in all but one CpG site located within the DMR and *PM20D1* expression levels, with Spearman’s rho ranging from -0.014 to -0,126 (all p-values>0.05, **Figure S2**). Atopic status of children from the same dataset was not associated with the *PM20D1* expression levels (p-value=0.382).

**Figure S2. Correlation between *PM20D1* expression and DNA methylation at 10 CpG sites overlapping the *PM20D1* DMR in healthy children (N=97) from the project “DNA Methylation Changes and Childhood Asthma in the Inner City” (SuperSeries GSE40736). Dashed lines indicate the best-fitting lines for the observed data.**



## References

1. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2017). <https://www.R-project.org/>.
2. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics*. 2014;**30**:1363-1369.
3. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;**4**:325-341.
4. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;**8**:203-209.
5. Langie SAS, Moisse M, Szarc Vel Szic K, Van Der Plas E, Koppen G, De Prins S, et al. GLI2 promoter hypermethylation in saliva of children with a respiratory allergy. *Clin Epigenetics*. 2018;**10**:50.
6. Zhang X, Biagini Myers JM, Burlison JD, Ulm A et al. Nasal DNA methylation is associated with childhood asthma. *Epigenomics* 2018 May;**10**(5):629-641.
7. Yang IV, Pedersen BS, Liu A, O'Connor GT, Teach SJ, Kattan M, et al. DNA methylation and childhood asthma in the inner city. *J Allergy Clin Immunol*. 2015;**136**:69-80.
8. Yang IV, Pedersen BS, Liu AH, O'Connor GT, Pillai D, Kattan M, et al. The nasal methylome and childhood atopic asthma. *J Allergy Clin Immunol*. 2017;**139**:1478-1488.