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

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

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.

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

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

Affiliations

[1] Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin; Centro di Ricerca in Medicina Sperimentale (CeRMS) and CPO Piemonte, Turin, Italy

[2] Italian Institute for Genomic Medicine (IIGM), Turin, Italy.

[3] Unit of Epidemiology, 'Anna Meyer' Children's University Hospital, Florence, Italy

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

Corresponding author:

Maja Popovic

Department of Medical Sciences,

University of Turin

Via Santena 7

10126 Turin, Italy

E-mail: maja.popovic@unito.it

Phone: +39(0) 116334628

Word count: 4351

Table count: 4

Figure count: 1

Material in the electronic repository: Yes

Conflicts of interest

The authors have no potential conflicts of interest to disclose.

Financial support

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

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

Introduction

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

Although DNA methylation is the most studied epigenetic mechanism to date, post-translational histone modifications have also been implicated in the T-cells differentiation and airway remodelling, contributing to the epigenetic regulation of allergic phenotypes, including asthma.[12,13,14]

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

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

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

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

Methods

Study population

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

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

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

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

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

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

DNA methylation data

The methylation status of over 485,000 probes was measured using the Illumina Infinium HumanMethylation450 BeadChip (Illumina, Inc., San Diego, CA). Details on pre-processing of samples and data quality control can be found in the **Supporting Information file** (Methods, *DNA methylation measurement, data pre-processing, and quality control***)**. 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). After quality control and probes filtering 136 samples with DNA methylation measured in 421,782 probes remained for the analyses.

Statistical analyses

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

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

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

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

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

Analysis of publicly available datasets

To test our main findings in independent samples we identified five datasets from the Gene Expression Omnibus (GEO) database (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.[31] 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.[31]
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.[19, 20] 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).[19] 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.[20] Both studies used Infinium
216 HumanMethylation450 BeadChip array to measure DNA methylation.

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

Results

Epigenome-wide DNA methylation analysis of early childhood wheezing

Descriptive statistics of the study population are shown in **Table 1**. Maternal age, educational level, nationality, and smoking after delivery were balanced between cases and controls. Children with wheezing were more likely to be born to mothers with a history of atopic disorders (p=0.09), to have siblings and to have attended day-care compared with controls (both p-values=0.02). Median age at saliva sampling was 10.3 months (interquartile range [IQR] 9.4-11.8 months).

In an exploratory epigenome-wide association analysis 25,042 (5.9%) probes were associated with childhood wheezing at the conventional p-value<0.05, but none remained associated after adjustment for multiple comparisons (Bonferroni p-value<1.19e-07 and Benjamini-Hochberg FDR-corrected p-value<0.05). The Manhattan plot of the EWAS results is shown in **Figure S1**.

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

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

with controls, with seven probes having Δbeta from 4.3% to 9.6% (all p-values<0.05, **Table 3**, first column). The largest differences between cases and controls were identified in the *PM20D1* promoter region and in the region of the first exon.

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

Analysis of the publicly available datasets

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

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

The three probes that had Δbeta below 1% (and a p-value>0.05) in the NINFEA sample had the greatest distance from the transcriptional start site and the lowest difference in the mean DNA methylation between cases and controls also in all public datasets (**Tables 3 and 4**). As GSE40576 PBMCs dataset was accompanied with paired gene expression data (GSE40732) we further examined whether differentially methylated sites impact *PM20D1* expression in healthy children (**Supporting Information file,** Results, *Correlation of differentially methylated sites with expression of PM20D1*). Although the *PM20D1* expression level was not associated with persistent atopic asthma in this datatset (p-value=0.382), there was a negative relationship between DNA methylation in all but one CpG site located within the DMR and *PM20D1* expression, with Spearman's rho ranging from -0.014 to -0,126 (all p-values>0.05, **Supporting Information file, Figure S2**).

Discussion

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

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

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

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

317 cg11965913 located in the promoter region of *PM20D1* 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 *PM20D1* 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.[34] In the NINFEA study, on the
322 contrary, hypermethylation of *PM20D1* 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.[34] and our own findings both indicate that maternal atopy likely
325 modify offspring atopic phenotypes through alterations in *PM20D1* 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,[35] it strongly
339 predicts later asthma development.[15] As all phenotypes in the public datasets have an atopic
340 characterization together with data on respiratory conditions, including wheezing and asthma,

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

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

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

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

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

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

Findings from previous studies suggest that the associations observed between DNA methylation and asthma are largely driven by eosinophils[22,23] that are known to be involved in airway inflammation present in T-helper-2-type asthmatic phenotypes.[38] 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.[28,29] Using simulated scenarios, a study of
402 McGregor et al.[29] 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.[17-24,39] 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

tissues for studies focused on local inflammation and environmental influences in asthma. These two tissues are also easily accessible and appear to reflect pathological changes in the airways during asthma development.[26]

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

Acknowledgements

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

References

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

4. Oelsner KT, Guo Y, To SB, Non AL, Barkin SL. Maternal BMI as a predictor of methylation of obesity-related genes in saliva samples from preschool-age Hispanic children at-risk for obesity. *BMC Genomics*. 2017;**18**:57.
5. Gruzieva O, Xu CY, Breton CV, Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-wide meta-analysis of methylation in children related to prenatal NO2 air pollution exposure. *Environ Health Perspect.* 2017;**125**:104–110.
6. Silvestri M, Franchi S, Pistorio A, Petecchia L, Rusconi F. Smoke exposure, wheezing, and asthma development: a systematic review and meta-analysis in unselected birth cohorts. *Pediatr Pulmonol.* 2015;**50**:353-362.
7. Rusconi F, Popovic M. Maternal obesity and childhood wheezing and asthma. *Paediatr Respir Rev.* 2017;**22**:66-71.
8. Hehua Z, Quing C, Shanyan G, Quijun W, Yuhong Z. The impact of prenatal exposure to air pollution on childhood wheezing and asthma: A systematic review. *Environ Res.* 2017;**159**:519-530.
9. McErlean P, Favoreto S Jr, Costa FF, et al. Human rhinovirus infection causes different DNA methylation changes in nasal epithelial cells from healthy and asthmatic subjects. BMC Med Genomics. 2014;7:37.
10. McDade TW. Early environments and the ecology of inflammation. Proc Natl Acad Sci U S A. 2012;109:17281-17288.
11. Clifford RL, Jones MJ, MacIsaac JL, et al. Inhalation of diesel exhaust and allergen alters human bronchial epithelium DNA methylation. J Allergy Clin Immunol. 2017;139:112-121.
12. Alaskhar Alhamwe B, Khalaila R, Wolf J, et al. Histone modifications and their role in epigenetics of atopy and allergic diseases. Allergy Asthma Clin Immunol. 2018;14:39.

13. Potaczek DP, Harb H, Michel S, Alhamwe BA, Renz H, Tost J. Epigenetics and allergy: from basic mechanisms to clinical applications. Epigenomics. 2017;9(4):539-571.

14. Harb H, Irvine J, Amarasekera M, et al. The role of PKCζ in cord blood T-cell maturation towards Th1 cytokine profile and its epigenetic regulation by fish oil. Biosci Rep. 2017;37(2).

15. Kabesch M. Early origins of asthma (and allergy). *Mol Cell Pediatr.* 2016;**3**:31.

16. Belgrave DC, Buchan I, Bishop C, Lowe L, Simpson A, Custovic A. Trajectories of lung function during childhood. *Am J Respir Crit Care Med*. 2014;**189**:1101-1109.

17. Baccarelli A, Rusconi F, Bollati V, et al. Nasal cell DNA methylation, inflammation, lung function and wheezing in children with asthma. Epigenomics. 2012;4:91-100.

18. Murphy TM, Wong CC, Arseneault L, Burrage J, Macdonald R, Hannon E, et al. Methylomic markers of persistent childhood asthma: a longitudinal study of asthma-discordant monozygotic twins. *Clin Epigenetics.* 2015;**7**:130.

19. 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.

20. 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.

21. DeVries A, Wlasiuk G, Miller SJ, Bosco A, Stern DA, Lohman IC, et al. Epigenome-wide analysis links SMAD3 methylation at birth to asthma in children of asthmatic mothers. *J Allergy Clin Immunol.* 2017;**140**:534-542.

22. Arathimos R, Suderman M, Sharp GC, Burrows K, Granell R, Tilling K, et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. *Clin Epigenetics.* 2017;**9**:112.

23. Xu CJ, Söderhäll C, Bustamante M, Baïz N, Gruzieva O, Gehring U, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med.* 2018;**6**:379-388.

24. Reese SE, Xu CJ, den Dekker HT, Lee MK, Sikdar S, Ruiz-Arenas C, et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. J Allergy Clin Immunol. Published Online: December 20, 2018 (doi: 10.1016/j.jaci.2018.11.043).

25. Stefanowicz D, Hackett TL, Garmaroudi FS, Günther OP, Neumann S, Sutanto EN, et al. DNA methylation profiles of airway epithelial cells and PBMCs from healthy, atopic and asthmatic children. *PLoS One*. 2012;**7**:e44213.

26. Langie SAS, Moisse M, Declerck K, Koppen G, Godderis L, Vanden Berghe W, et al. Salivary DNA methylation profiling: aspects to consider for biomarker identification. *Basic Clin Pharmacol Toxicol*. 2017;**121**:93-101.

27. Firestone R, Cheng S, Pearce N, Douwes J, Merletti F, Pizzi C, et al. Internet-based birth-cohort studies: is this the future for epidemiology? *JMIR Res Protoc*. 2015;**4**:e71.

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

29. McGregor K, Bernatsky S, Colmegna I, Hudson M, Pastinen T, Labbe A, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol*. 2016;**17**:84.

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

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

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

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

[†] Total numbers may vary due to missing data
[‡]SD – Standard Deviation
[§] Chi-square test for categorical variables, unpaired t test or Wilcoxon-Mann-Whitney test for continuous variables
[¶] University degree or higher
[††] Otits and/or sinusitis and/or throat infections and/or bronchitis and/or influenza during pregnancy
[‡‡] Median (Interquartile Range – IQR)

557

558

559

560

561

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

Cluster number	Chromosome	Start Position	End Position	Delta β[†]	Number of probes	p-value	FWER[‡]	Nearest gene[§]
12362	1	205818484	205819609	0.077	10	0.006	0.028	*PM20D1*

[†] The average difference in DNA methylation between cases and controls

[‡] Family-wise error rate from 1000 bootstraps

[§] 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[†]	Genic location[‡]	NINFEA (N=136)		Maternal atopic disorders[††]			
					No (N=100)		Yes (N=34)	
			Delta β[§] (%)	p-value[¶]	Delta β[§] (%)	p-value[¶]	Delta β[§] (%)	p-value[¶]
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

[†] Genome coordinates are given with respect to hg19 genome assembly

[‡] According to UCSC reference gene information

[§] The difference in the mean DNA methylation between cases and controls (%)

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

[††] Maternal history of allergic rhinits 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[§] 11 years (N=90)		Cord blood[¶] (N=90)		Blood[§] 6-12 years (N=194)		Nasal epithelia 9-12 years (N=72)	
	Delta β[†] (%)	p-value	Delta β[†] (%)	p-value[‡]	Delta β[†] (%)	p-value[‡]	Delta β[†] (%)	p-value[‡]	Delta β[†] (%)	p-value[‡]	Delta β[†] (%)	p-value[‡]
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*

[†] The difference in the mean DNA methylation between cases and controls (%)

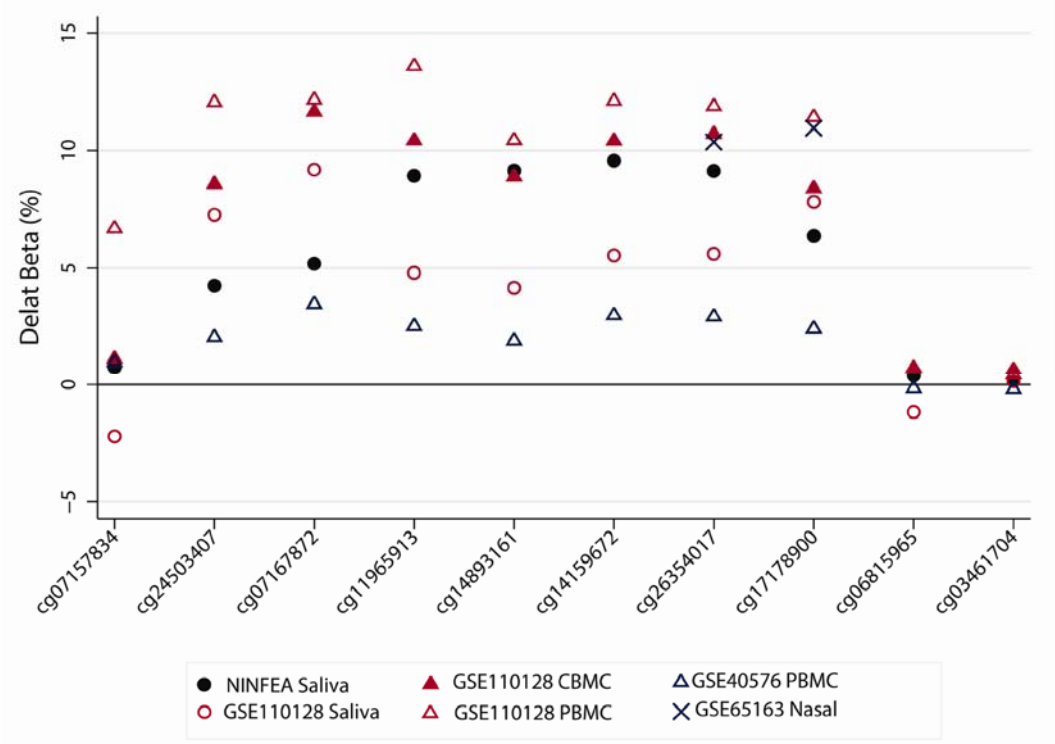
[‡] P-values based on a one-sided Student's t-test for the difference in the mean DNA methylation between cases and controls.

[§] Peripheral blood mononuclear cell (PBMC) fraction

[¶] 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[1] using the *minfi* package.[2] 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.[3] 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.[3] This procedure does not apply background correction, but very small intensities close to zero are thresholded using the *minfi* function fixMethOutliers.[2]

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,[2] N=15,654), and polymorphic CpG probes that have been demonstrated

to map to multiple places in the genome[4] (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 15th 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[5] 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,[6] 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)[5] 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,[7] 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).[8] 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.[7,8]

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 log10 of the p-value for association of each CpG site with early childhood wheezing.The red line indicates the Bonferroni-corrected epigenome-wide treshold (p=1.19e-07).

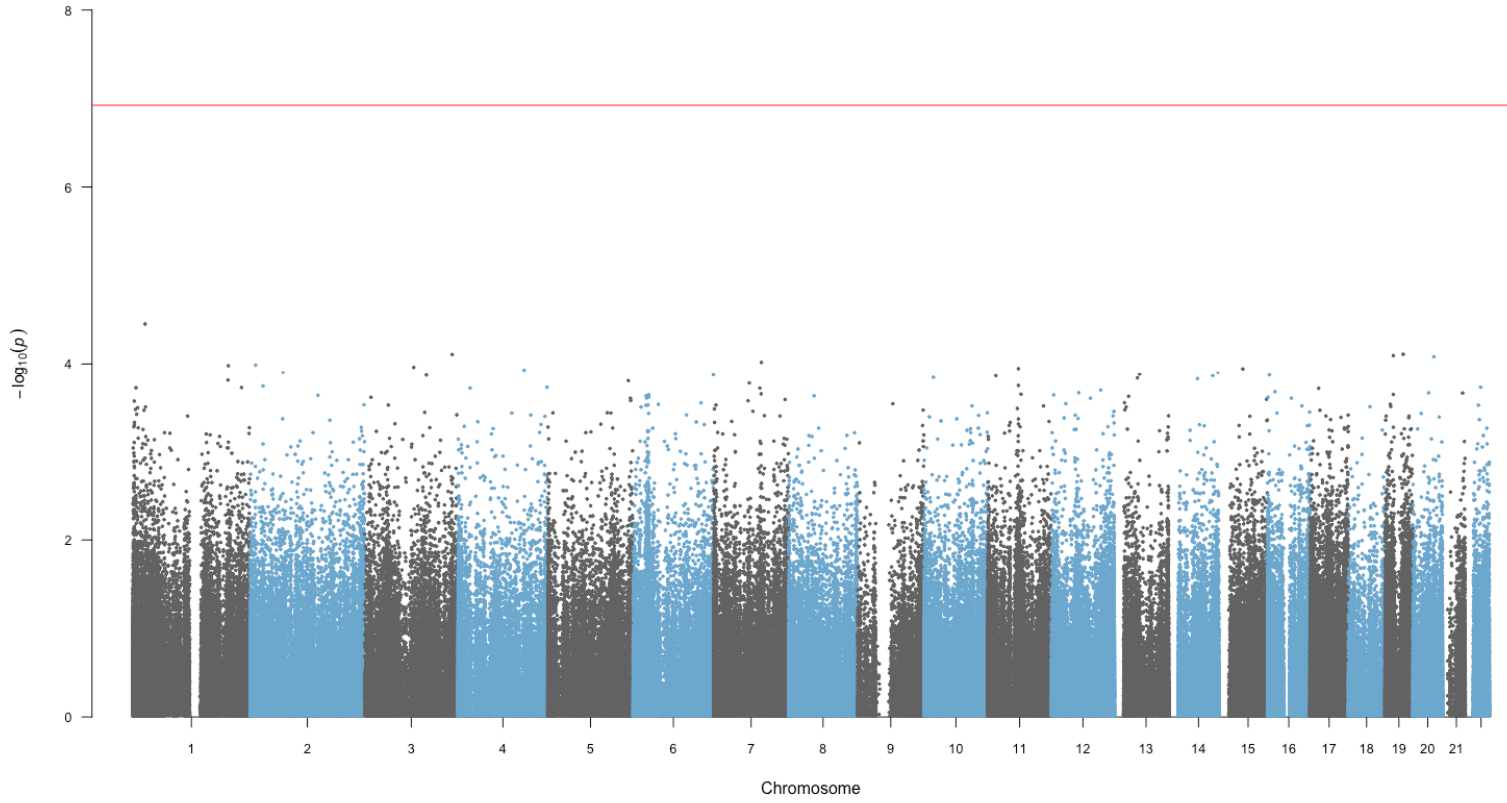


Table S1. Differentially methylated regions (DMRs) associated with early childhood wheezing at 5% in the methylation beta value difference between cases and controls

Cluster number	Chromosome	Start Position	End Position	Delta β [†]	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 β - 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
[d] The distance to the 5' end of the nearest transcript (base pairs)

Correlation of differentially methylated sites with expression of PM20D1

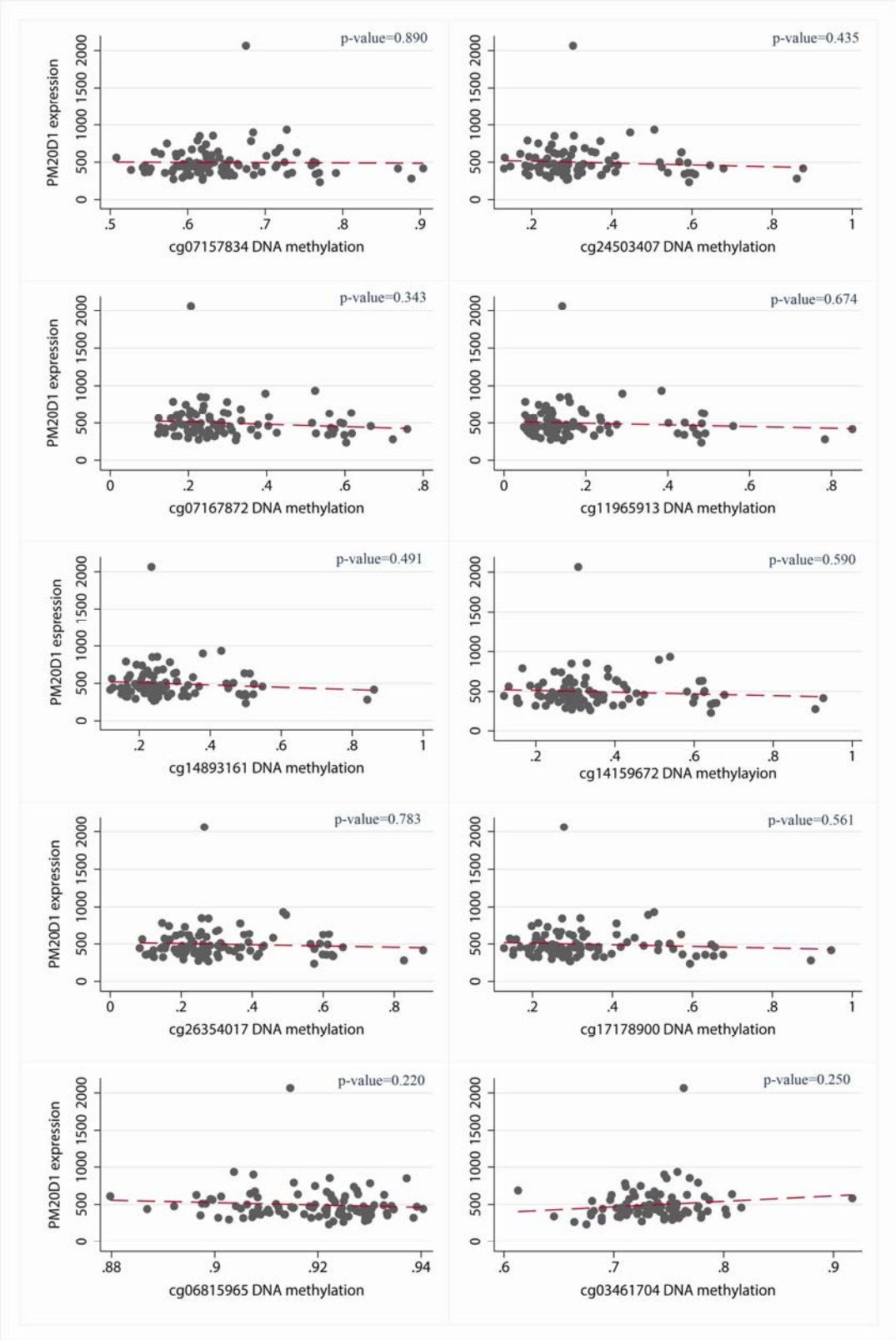
Two of the five public DNA methylation datasets that we used to replicate our differentially methylated region (GSE40576 and GSE65163) have publicly available paired gene expression data. The publicly available gene expression dataset (GSE65204), related to the GSE65163 DNA methylation data, do not contain 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).[7]

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, Burleson 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.