

RESEARCH

Open Access



# Epigenome-wide association study of asthma and wheeze in childhood and adolescence

Ryan Arathimos<sup>1,2\*</sup> , Matthew Suderman<sup>1,2</sup>, Gemma C. Sharp<sup>2,3</sup>, Kimberley Burrows<sup>1,2</sup>, Raquel Granell<sup>1</sup>, Kate Tilling<sup>1,2</sup>, Tom R. Gaunt<sup>1,2</sup>, John Henderson<sup>1</sup>, Susan Ring<sup>1,2</sup>, Rebecca C. Richmond<sup>1,2</sup> and Caroline L. Relton<sup>1,2</sup>

## Abstract

**Background:** Asthma heritability has only been partially explained by genetic variants and is known to be sensitive to environmental factors, implicating epigenetic modifications such as DNA methylation in its pathogenesis.

**Methods:** Using data collected in the Avon Longitudinal Study of Parents and Children (ALSPAC), we assessed associations of asthma and wheeze with DNA methylation at 7.5 and 16.5 years, at over 450,000 CpG sites in DNA from the peripheral blood of approx. 1000 participants. We used Mendelian randomization (MR), a method of causal inference that uses genetic variants as instrumental variables, to infer the direction of association between DNA methylation and asthma.

**Results:** We identified 302 CpGs associated with current asthma status (FDR-adjusted  $P$  value  $< 0.05$ ) and 445 with current wheeze status at 7.5 years, with substantial overlap between the two. Genes annotated to the 302 associated CpGs were enriched for pathways related to movement of cellular/subcellular components, locomotion, interleukin-4 production and eosinophil migration. All associations attenuated when adjusted for eosinophil and neutrophil cell count estimates. At 16.5 years, two sites were associated with current asthma after adjustment for cell counts. The CpGs mapped to the *AP2A2* and *IL5RA* genes, with a  $-2.32$  [95% CI  $-1.47, -3.18$ ] and  $-2.49$  [95% CI  $-1.56, -3.43$ ] difference in percentage methylation in asthma cases respectively. Two-sample bi-directional MR indicated a causal effect of asthma on DNA methylation at several CpG sites at 7.5 years. However, associations did not persist after adjustment for multiple testing. There was no evidence of a causal effect of asthma on DNA methylation at either of the two CpG sites at 16.5 years.

**Conclusion:** The majority of observed associations are driven by higher eosinophil cell counts in asthma cases, acting as an intermediate phenotype, with important implications for future studies of DNA methylation in atopic diseases.

**Keywords:** ALSPAC, Asthma, Wheeze, DNA methylation, Longitudinal, Mendelian randomization

## Background

Asthma continues to be a global burden on health care systems with 8.2% of the US population affected in 2009 [1] and 300 million affected worldwide [2]. The prevalence of paediatric asthma in particular is on the rise along with modern lifestyle factors such as air pollutants and indoor contaminants and allergens thought to be

influencing the increase [3]. Although asthma is known to be strongly heritable [4], genome-wide association studies (GWAS) have only been able to explain a fraction of that heritability [5, 6].

As asthma development and progression are affected by environmental factors, a hypothesised role for epigenetic mechanisms has been proposed [7–9]. There has been particular interest in a role for DNA methylation, one of the most widely studied epigenetic mechanisms which is known to be responsive to environmental exposures [10]. DNA methylation variation has previously been found to be associated with known asthma triggers,

\* Correspondence: ryan.arathimos@bristol.ac.uk

<sup>1</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK

<sup>2</sup>Medical Research Council Integrative Epidemiology Unit, University of Bristol, Oakfield House, Bristol BS8 2BN, UK

Full list of author information is available at the end of the article

including air pollution [11, 12] and smoking [13, 14]. Whereas childhood asthma is often allergic in character and is more common in boys than in girls [15, 16], adult asthma is known to differ in that it is often more severe, non-atopic and leads to a more rapid decline in lung function [16], indicating different biological mechanisms involved in the pathophysiology between asthma in childhood and asthma in late adolescence.

Wheeze is a symptom often observed as a precursor to children that develop asthma [17, 18] and has been indicated to have important prognostic value in early detection of asthma [19]. Whereas wheeze and asthma are related traits, wheezing is a broader and more heterogeneous phenotype than asthma that encompasses a wider variety of respiratory conditions, which may include disorders such as chronic obstructive pulmonary disorder (COPD), viral infection and bronchiolitis [20]. Associations of wheeze with DNA methylation may be expected to capture a wider variety of biological processes that underlie respiratory conditions.

Previous studies have reported an association between peripheral blood DNA methylation and IgE [21, 22], an important mediator of atopic or inflammatory diseases such as asthma. However, whether these associations are also present for the asthma phenotype has not been assessed. Studies that have examined DNA methylation associations with asthma have not specifically determined whether observed associations are due to cell type confounding between samples [23, 24]. Eosinophils, a type of circulating granulocyte, are known to be increased in peripheral blood of some people with allergic conditions, such as certain subtypes of asthma [25, 26]. Eosinophils are involved in the development of characteristic features of asthma, such as airway remodelling and hyperresponsiveness, heightened immune response and initiation of allergic airway inflammation [27, 28]. The increase of eosinophils in some individuals with asthma may lead to 'confounding' or contamination of whole-blood methylation data [29, 30], as the distinct methylation patterns of eosinophils [31] are over-represented in cases.

Associations observed between DNA methylation and disease outcomes, like all observational associations, are vulnerable to confounding, reverse causation and bias. Determining the direction of association is of critical importance to the biological interpretation of such findings [32]. The Mendelian randomization (MR) is an approach that uses genetic variants as instrumental variables (IVs) and has been widely implemented to strengthen causal inference [33, 34]. Any association observed between the IVs and the disease outcome is most likely explained by an unbiased causal effect of the exposure on the outcome, given a certain set of prior assumptions [34]. MR can make use of the strong *cis* effects of genetic variants

on DNA methylation. *Cis*-genetic variants robustly associated with site-specific DNA methylation can be identified and used as causal anchors in such analyses [21, 35]. This approach is particularly useful in the case of DNA methylation, where the causal direction of association between DNA methylation and the disease outcome may be difficult to establish using observational data alone. Furthermore, the application of MR is increasingly facilitated by the availability of GWAS data from large-scale consortia as the methods can be applied using summary level statistics [36].

In the present study, we examine genome-wide DNA methylation in relation to asthma and wheeze in childhood and adolescence in a large contemporary UK-based cohort. We implement a variety of analytical strategies to gain insights into causal relationships of the DNA methylation variation observed.

## Methods

### Study population

The Avon Longitudinal Study of Parents and Children (ALSPAC) recruited 14,541 pregnant women resident in the former county of Avon, UK, with expected dates of delivery 1 April 1991 to 31 December 1992. The initial number of pregnancies was 14,541, for which the mother enrolled in the ALSPAC study and had either returned at least one questionnaire or attended a 'Children in Focus' clinic by 19 July 1999. Of these initial pregnancies, there was a total of 14,062 live births and 13,988 children who were alive at 1 year of age. The phases of enrolment are described in detail in the cohort profile paper [37]. The study website contains details of all the data that is available through a searchable data dictionary <http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>. Written informed consent has been obtained for all ALSPAC participants. Ethical approval was granted from the ALSPAC Law and Ethics Committee and the local Research Ethics Committee in accordance with the guidelines of the Declaration of Helsinki.

### Respiratory phenotypes

Estimates of current asthma/wheeze status and ever asthma/wheeze were derived from questionnaires completed by the mothers of the study children when they were 7.5 and 16.5 years old respectively. Current asthma case status at 7.5 years of age was defined using a positive answer to the question 'Has your child ever been diagnosed with asthma by a doctor?' and a positive answer to one of the following questions: (1) 'Has your child had asthma in the past 12 months?' (2) 'Has your child been on asthma medication the last 12 months?' or (3) 'Has your study child had wheezing or whistling on the chest in the past 12 months?'. Controls were children with negative answers to all the above questions.

Current wheeze at 7.5 years was defined using answers to the question ‘Has your child had wheezing or whistling on the chest in the last 12 months?’. Wheeze is a respiratory condition that may capture broader respiratory pathology that includes but is not limited to asthma. However, not all asthmatics currently wheeze since many are currently taking medication that suppress disease symptoms. The number of individuals with current asthma at 7.5 years is therefore larger than the number of individuals with wheeze at that age due to this lack of symptom manifestation in some asthmatics. Current asthma case status at 16.5 years was defined using the same method but from counterpart questions asked in the child-completed questionnaire. Ever asthma and ever wheeze at 16.5 years were defined using answers to the questions from questionnaires completed at 16.5 years of age by the mothers of the study participants. Answers to the questions ‘Has he/she [your study child] ever had asthma?’ and ‘Has he/she ever had wheezing or whistling in the chest at any time in the past?’ were used to define each measure respectively.

#### **DNA methylation data**

DNA methylation data for a subset of approximately 1000 mother-child pairs from ALSPAC is available as part of the ARIES (Accessible Resource for Integrated Epigenomics Studies) project [38]. ARIES participants were selected based on availability of DNA samples at three time-points for the offspring (neonatal, childhood and adolescence). Mothers included in ARIES were slightly older, more likely to have a non-manual occupation and less likely to have smoked throughout pregnancy when compared to the total ALSPAC study cohort [38]. DNA methylation was measured using the Illumina Infinium HumanMethylation450 BeadChip. DNA methylation at each CpG site is expressed as a beta-value ( $\beta$ ) (ratio of methylated probe intensity to overall intensity representing 0 to 100% methylation at the CpG) based on the proportion of detected CpGs that were methylated at each site on the array [39]. Pre-processing of peripheral blood and cord blood samples was performed as previously described [38]. Details of quality control (QC) of the DNA methylation data can be found in Additional file 1.

#### **Other variables**

Proportions of cell types were estimated from DNA methylation data using the estimateCellCounts function in the minfi R package [40] which is based on the method developed by Houseman et al. [41]. This estimated the proportion of B cells, CD8 T cells, CD4 T cells, granulocytes, eosinophils, neutrophils, NK cells and monocytes at the 7.5-year methylation time point and at the 16.5-year methylation time-point independently. Descriptive statistics

of the derived cell counts can be seen in Additional file 1: Table S1. All epigenome-wide association study (EWAS) analyses were adjusted for three levels of cell-type adjustment. Model 1: unadjusted for cells, model 2: basic cells (where eosinophils and neutrophils were grouped together as granulocytes) and model 3: detailed cells (eosinophils and neutrophils assessed separately).

Atopic status was determined in a subset of ALSPAC children when they were approximately 7.5 years old by skin prick test response to a panel of up to 12 common allergens, including house dust mite, grass pollen and cat. Discrimination of atopic status based on house dust mite, grass pollen and cat has been previously shown to identify over 95% of sensitised children [42], and we were primarily interested in allergic response to aeroallergens as indicators of atopic status which have a plausible biological link to respiratory conditions such as asthma. A positive response was defined as a mean weal diameter of  $> 2$  mm with an absent response to negative control solution, and atopy was defined as a positive response to one or more of house dust mite, grass pollen and cat.

#### **Statistical analyses**

We first examined cross-sectional associations of asthma status at both 7.5 and 16.5 years with estimated cell counts. We used linear regression models in Stata 13 MP2 [43] to determine associations between cell types that were most likely to have a strong effect on results when adjusted for in EWAS models.

#### **Epigenome-wide association studies**

EWAS were carried out in R version 3.0.2. using the CpGassoc package [44] to model either asthma/wheeze status as the exposure and methylation (beta value) at either 7.5 or 16.5 years of age as the outcome. In addition to the cross-sectional EWAS, we also examined two longitudinal EWAS models of current asthma. We did this in order to determine if the asthma status in childhood was longitudinally associated with DNA methylation in adolescence or if DNA methylation in childhood was longitudinally associated with asthma status in adolescence. We specified one model with DNA methylation at 7.5 years (as the exposure) and later life current asthma status at 16.5 years (as the outcome), and one with current asthma at 7.5 years (as the exposure) and DNA methylation at 16.5 years (as the outcome).

EWAS results were then corrected for multiple testing by controlling the expected proportion of false-positives among all discoveries (false discovery rate (FDR)-adjusted  $P$  value  $< 0.05$ ), and type 1 errors were controlled using Bonferroni adjustment. All models were adjusted for sex, maternal smoking during pregnancy, maternal age, parity, maternal education derived from maternal questionnaires and unknown confounders using surrogate

variables (by use of maximum 10 surrogate variables obtained from the *sva* R package [45]), as shown in Additional file 1: Table S2. Probes that were known to contain SNPs or be of low quality [46] were removed. The genomic inflation factor ( $\lambda$ ) and quantile-quantile (Q-Q) plots were used to compare the genome-wide distribution of *P* values with the expected null distribution.

We examined three models: Model 1 was adjusted offspring sex, maternal age, parity, smoking status, maternal education and for unknown confounders using surrogate variables; model 2 adjusted additionally for basic cell counts (which grouped eosinophils and neutrophils under granulocytes); and model 3 adjusted for covariates as above and detailed cell counts. Longitudinal EWAS models were adjusted for estimated cell counts at the relevant DNA methylation time point.

### **Sensitivity analyses**

We performed a sensitivity analysis where individuals with extreme values of derived eosinophil cell counts at 7.5 years were removed using Tukey's method of outlier removal [47, 48]. This was done as some evidence suggest eosinophilic asthma represents a subtype of asthma within a wider spectrum of asthmatic disorders [49, 50]. We used both a stringent and a relaxed version of Tukey's method. For the stringent method, potential outliers in the eosinophil cell count data were removed if their value was more than the upper quartile plus one and a half times the interquartile range (IQR). For the relaxed method, outliers were removed if their value was more than the upper quartile plus three times the IQR. We did not remove values that were less than the lower quartile minus three times the interquartile range as having an eosinophil cell count of zero (or near-zero) is normal for most healthy individuals. We also stratified the outlier removal by sex to determine whether outliers in eosinophil cell counts were more frequent in females or males.

### **Functional analysis**

The Gene Ontology (GO) [51] database in Database for Annotation, Visualisation and Integrated Discovery (DAVID) [52] Bioinformatics Resources 6.8 (beta) was used to examine gene function in potential molecular, cellular and biological processes of loci observed to be associated with asthma at age 7.5 years (adjusted for basic cell counts). The Kyoto Encyclopedia of Genes and Genomes (KEGG) database [53] in DAVID was used to explore whether genes were related to known pathways involved in asthma or eosinophil regulation.

### **Mendelian randomization**

In order to investigate the causal relationships between DNA methylation and asthma, a MR approach was applied. Briefly, MR is a method of investigating causal

relationships by using genetic variants as IVs [33]. The principle of MR is based on Mendel's laws of segregation and independent assortment. Genetic variants that adhere to those laws are distributed independently and randomly with respect to other genetic variants, assuming no linkage disequilibrium (LD) or population stratification. Randomization of these genetic variants ensures that any association observed between the genetic variants and the outcome is most likely explained by an unbiased causal effect of the exposure on the outcome. The reasoning for this is that the MR design is analogous to a randomized control trial (RCT), where study participants are randomly allocated to one or another treatment, avoiding potential confounding between treatment and outcome. The MR approach relies on three key assumptions: that the genetic variants cause the exposure; that there is no path from genetic variants to outcome other than via exposure (i.e. no pleiotropy); and that the genetic variants are not associated with any of the confounders of the exposure-outcome association [54, 55]. A two-sample approach to MR was used which has the additional benefit of only requiring summary level data from an outcome GWAS and an exposure GWAS. The strengths and limitations of this specific technique have been reviewed [56, 57]. A bi-directional approach was implemented in order to assess both the potential causal effect of asthma on DNA methylation and the effect of DNA methylation on asthma. The Two-SampleMR package in R available as part of the MR-Base ([www.mrbase.org](http://www.mrbase.org)) platform [36] was used for all MR analyses. We analysed the causal effect of the asthma SNPs on CpGs from the EWAS of asthma at 7.5 years adjusted for basic cell counts.

For assessing the causal effects of asthma on DNA methylation, asthma SNPs from the GWAS catalogue [58] were used as IVs for asthma and methylation data from the ARIES dataset was used as our second sample (Additional file 1: Figure S2). Specifics on the selection and QC of SNPs from the GWAS catalogue used as IVs can be found in Additional file 1. A combination of methods, including the inverse variance-weighted method, maximum likelihood method, weighted median method and MR-Egger regression, were used as sensitivity analyses. We tested for pleiotropy by assessing the MR-Egger regression intercept with funnel plots/scatter plots and for heterogeneity using Cochran's Q statistic.

In the reverse direction, *cis*-SNPs robustly associated with methylation levels at the specific CpG sites were used as proxy IVs for DNA methylation. For assessing the causal effect of DNA methylation on asthma, we identified *cis*-SNPs associated with EWAS hits in ARIES [59] and assessed the association of these *cis*-SNPs with asthma using summary data from



the GABRIEL consortium GWAS of asthma [6] as the second sample (Additional file 1: Figure S3). The Wald ratio method was used in the DNA methylation to asthma direction as only one *cis*-SNP was selected per CpG due to high linkage disequilibrium (LD) between neighbouring SNPs.

We calculated the effective number of independent tests for both the asthma to methylation and the methylation directions of the MR analysis. Adjustment for multiple testing is required when large numbers of tests are performed in an MR framework. However, CpGs are likely to be co-methylated if they are located in the same gene or close to each other on the chromosome. In order to take into account potential correlation between nearby CpGs, we calculated the effective number of independent tests for the 302 CpGs using matSpDlite [60]. Briefly, the method determines the number of independent tests using spectral decomposition of the correlation matrix of the tested exposures or outcomes. It calculates the equivalent number of independent variables in the correlation matrix by examining the ratio of observed eigenvalue variance to its theoretical maximum. The method calculates a *P* value at which type 1 errors are kept at 5% based on the Veff process described by Li and Ji [61].

We performed power calculations using the mRnd Mendelian randomization power calculator by CNS genomics available online [62]. In the asthma to DNA methylation direction, we had a maximum 6% power to detect a 3% change in methylation. In the DNA methylation to asthma direction, we had a maximum 73% power to detect an OR of 1.1. Details of the power calculation can be found in Additional file 1. Due to the low power to detect causal effects using Mendelian randomization, we present results for all CpGs from the EWAS and discuss in detail results from the top 20 CpGs based on lowest *P* value in the EWAS only, as they demonstrate the most robust association with asthma status. Specific details of the MR process can be found in Additional file 1.

## Results

### Sample characteristics

Descriptive statistics of ARIES individuals included in each EWAS model are shown in Table 1. Individuals included as cases in each model were more likely to have a mother that smoked when compared to controls. Therefore, we adjusted for maternal smoking in all models as well as other potential confounders (maternal education, parity and maternal age).

There was strong evidence of associations between current asthma and estimated cell counts, particularly in childhood (Table 2). Current asthma at 7.5 years demonstrated a strong positive association with estimates of B cells and eosinophils, with a 0.026 [95%CI 0.018, 0.033] increase in the proportion of B cells and a 0.011 [95%CI 0.005, 0.018] increase in the proportion of eosinophils in individuals with asthma respectively. There was a strong negative association with current asthma at the same age for neutrophils, with a 0.042 [95%CI -0.06, -0.022] decrease in proportion of neutrophils in individuals with asthma. However, granulocytes (summed proportion of eosinophils and neutrophils) were not associated with current asthma at the same age. Similarly, current asthma at 16.5 years demonstrated a positive association with estimates of derived eosinophils at 16.5 years with a 0.005 [95%CI 0.002, 0.008] increase in proportion, although with markedly less confidence in the estimate than the associations at 7.5 years. There was no evidence of an association between asthma at either 7.5 or 16.5 years and any of the other cell types. A graphical depiction of the distributions of cell counts by asthma status at both 7.5 and 16.5 years is shown in Fig. 1. The same association with eosinophil cell counts was present for current wheeze at 7.5 years (Additional file 1: Figure S4).

Associations of asthma at 7.5 years with cell counts stratified by atopic status indicate that increased eosinophils were evident in both atopic and non-atopic individuals with asthma. There was a 0.032 [95%CI 0.016, 0.047]

**Table 1** Characteristics of the individuals in each EWAS model

DNA methylation (outcome)	Asthma/wheeze (exposure)	Status	<i>N</i>	Sex [male] <i>N</i> (%)	Maternal age [months] mean (SD)	Maternal smoking [never] <i>N</i> (%)	Parity [nulliparous] <i>N</i> (%)	Maternal education [O-level or lower] <i>N</i> (%)
7.5 years	Current asthma at 7.5 years	Cases	149	88 (59%)	369 (53.4)	124 (83%)	62 (42%)	68 (46%)
		Controls	632	308 (49%)	364 (51.7)	560 (89%)	286 (45%)	306 (48%)
	Current wheeze at 7.5 years	Cases	111	66 (60%)	372 (55)	94 (85%)	43 (39%)	520 (45%)
		Controls	742	366 (49%)	361 (52)	650 (88%)	348 (47%)	362 (49%)
16.5 years	Ever asthma at 16.5 years	Cases	194	102 (53%)	363 (53)	163 (84%)	97 (50%)	87 (45%)
		Controls	554	264 (48%)	364 (51)	503 (91%)	258 (47%)	262 (47%)
	Ever wheeze at 16.5 years	Cases	204	108 (53%)	367 (49)	173 (85%)	102 (50%)	82 (40%)
		Controls	554	264 (48%)	363 (52)	503 (91%)	261 (47%)	271 (49%)
Current asthma at 16.5 years	Cases	184	73 (40%)	362 (51.6)	150 (82%)	89 (48%)	89 (46%)	
	Controls	427	200 (47%)	364 (51.7)	384 (90%)	204 (48%)	198 (46%)	

**Table 2** Associations of estimated cell counts at 7.5 and 16.5 years with current asthma status at 7.5 and 16.5 years respectively

	Current asthma at 7.5 years		Current asthma at 16.5 years	
	Change in proportion <sup>a</sup> [95%CI]	<i>P</i> value	Change in proportion <sup>a</sup> [95%CI]	<i>P</i> value
Monocytes	-0.002 [-0.005, 0.004]	0.695	-0.003 [-0.007, 0.004]	0.534
CD8T	0.002 [-0.008, 0.009]	0.777	-0.004 [-0.013, 0.007]	0.496
CD4T	0.009 [-0.002, 0.018]	0.069	0.009 [0, 0.018]	0.055
B cells	0.011 [0.005, 0.018]	< 0.001	0.005 [-0.003, 0.01]	0.173
NK cells	-0.004 [-0.008, 0.002]	0.159	-0.004 [-0.01, 0.004]	0.306
Granulocytes <sup>b</sup>	-0.016 [-0.032, 0.001]	0.065	-0.005 [-0.023, 0.014]	0.618
Neutrophils	-0.042 [-0.06, -0.022]	< 0.001	-0.009 [-0.029, 0.011]	0.382
Eosinophils	0.026 [0.018, 0.033]	< 0.001	0.005 [0.002, 0.008]	0.008

<sup>a</sup>Change in proportion for individuals with asthma

<sup>b</sup>Granulocytes calculated from the summed proportions of eosinophils and neutrophils

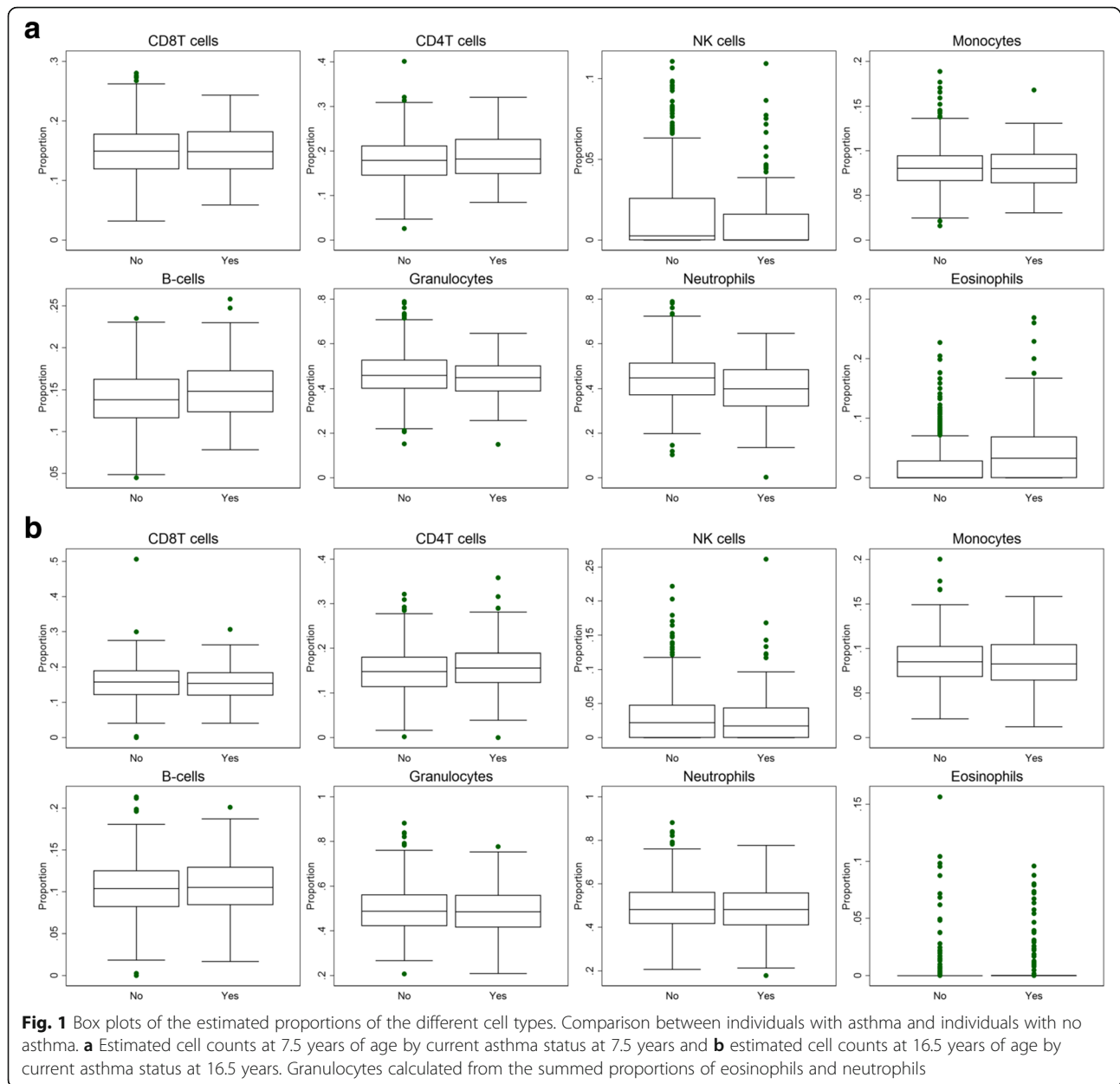
difference in the proportion of eosinophils in atopic individuals with asthma (Additional file 1: Table S4), whereas for non-atopic individuals with asthma, the association was somewhat weaker with a 0.016 [95%CI 0.007, 0.025] difference in the portion of eosinophils. For non-atopic individuals, there was also evidence of an association of asthma with B cells and neutrophils, with a 0.015 [95%CI 0.007, 0.023] and -0.039 [95% CI -0.065, -0.014] difference in proportion respectively.

### Epigenome-wide association analyses

We identified 411 CpG sites associated with current asthma status at 7.5 years (FDR-adjusted *P* value < 0.05) and 611 CpG sites with current wheeze status at 7.5 years. When adjusted for cell counts (eosinophils and neutrophils grouped together as granulocytes), 302 CpG sites remained associated with current asthma and 445 with current wheeze at 7.5 years. Adjusting for cell counts that included neutrophil and eosinophil counts separately (detailed cell counts) attenuated all associations (Table 3). We found evidence of genomic inflation in all unadjusted models that reduced progressively once adjustments for basic cell counts and detailed cell counts were made (Additional file 1: Figure S6). As expected, we found strong associations between DNA methylation and estimated eosinophil cell counts at CpGs associated with current asthma at 7.5 years (Additional file 2: Tables E23 and E24). The top locus associated with current asthma at 7.5 years adjusted for cell counts that did not include eosinophils and neutrophils mapped to *ZFPM1* (cg04983687) with a -5.02 [95% CI -3.5, -6.5] difference in percentage methylation for individuals with asthma. A larger number of sites were associated with current wheeze at age 7.5 years (445 CpG sites), and the top locus mapped to *MFHAS1* (cg1207746) with a -4.60 [95%CI -3.33, -5.86] difference in percentage methylation for individuals with wheeze. There was a high

degree of overlap between current asthma- and current wheeze-associated CpG sites (57.1% of 611 sites associated with current wheeze or 349 CpG sites) in the unadjusted model (55.7% of 445 sites associated with current wheeze or 248 CpG sites in the adjusted model). Current asthma at 16.5 years was associated with 2 CpGs in the unadjusted model that persisted after adjustment for both basic cells and for detailed cells that included eosinophils and neutrophils. The CpGs were annotated to the genes *AP2A2* (cg17676835) and *IL5RA* (cg10159529) with a -2.3 [95% CI -1.47, -3.18] difference and -2.5 [95% CI -1.56, -3.43] difference in percentage methylation, respectively, in the model adjusted for eosinophils and neutrophils (Fig. 2). There were no associations observed between methylation at 16.5 years and ever asthma status at the same age. There were 32 CpGs associated with ever wheeze at 16.5 years once adjusted for basic cells, but all associations attenuated once adjusted for cell counts that included eosinophils and neutrophils. Full results for all probes with a *P* value < 1e-3 can be found in Additional file 2: Tables E1–E15.

In the longitudinal EWAS models, we identified 4 CpGs where DNA methylation at 7.5 years was associated with asthma at 16.5 years (FDR < 0.05). When adjusted for cell counts that included eosinophils and neutrophils, all associations attenuated. In the model unadjusted for cell counts, the top hit mapped to the *IRGC* gene. We identified 28 CpGs where DNA methylation at 16.5 years was associated with asthma at 7.5 years. Following adjustment for detailed cell counts, all associations attenuated. In the unadjusted model, the top hit mapped to the *MACROD1* gene. We observed high inflation in all models that reduced only marginally with adjustment for estimated cell counts (Table 4 and Additional file 1: Figure S7). Full results for all probes with a *P* value < 1e-3 can be found in Additional file 3: Tables F1–F6.



**Sensitivity analyses**

In order to check whether associations of current asthma status at 7.5 years and cell counts were driven by outlier individuals with extreme levels of eosinophil cell counts, two outlier removal methods (Tukey’s method of outlier removal [47], both stringent and relaxed) were applied to the eosinophil cell count data. Additional file 1: Table S5 shows the effect of removal of outliers from the eosinophil cell counts. The association of asthma status with eosinophil cell counts remains even after stringent outlier removal. However, stratifying by sex shows that this appears to be a male phenomenon since there is no

observed association in females ( $P > 0.05$ ), as shown in Additional file 1: Table S6. The persistent association appears to be driven almost entirely by males. Based on this observation, we conducted a sensitivity analysis where we restricted the EWAS sample to just females for current asthma at 7.5 years ( $N = 385$ ). There was one site associated with current asthma in the model unadjusted for cell counts (cg07612991); however, once adjusted for basic cell counts, the association attenuated (Additional file 2: Tables E16, E17 and E18). Power to detect an effect may have been greatly reduced in the model due to halving of the original sample size.

**Table 3** Results of epigenome-wide association studies (EWAS), where methylation is the outcome and current asthma at 7.5 years, current wheeze at 7.5 years, current asthma at 16.5 years, ever asthma at 16.5 years and ever wheeze at 16.5 years are the exposures

Methylation (outcome)	Asthma/wheeze (exposure)	N [cases/controls]	Cell adjustment	FDR hits [Bonferroni]	Lambda
7.5 years	Current asthma at 7.5 years	781 [149/632]	No cells	411 [67]	1.08
			Basic cells	302 [43]	1.01
			Detailed cells	0 [0]	0.97
	Wheeze at 7.5 years	853 [111/742]	No cells	611 [105]	1.14
			Basic cells	445 [78]	1.08
			Detailed cells	0 [0]	1.00
16.5 years	Ever asthma at 16.5 years	748 [194/554]	No cells	0 [0]	1.09
			Basic cells	0 [0]	1.05
			Detailed cells	0 [0]	1.01
	Ever wheeze at 16.5 years	758 [204/554]	No cells	48 [9]	1.02
			Basic cells	32 [9]	1.05
			Detailed cells	0 [0]	1.01
	Current asthma at 16.5 years	611 [184/427]	No cells	2 [2]	1.01
			Basic cells	2 [2]	1.05
			Detailed cells	2 [1]	0.99

All models are adjusted for confounders and technical covariates. Stepwise adjustments for estimated cell counts were made (where basic cells include CD8T, CD4T, NK, monocytes, B cells and granulocytes and detailed cells include eosinophils and neutrophils instead of granulocytes) FDR-adjusted *P* value and Bonferroni hits use a cut-off of <0.05. Lambda based on *P* values of all probes without exclusions

In order to ensure that adjustment for all 7 estimated cell counts was not having unintended effects due to multicollinearity between the estimates, we ran the same EWAS of current asthma at 7.5 years adjusted only for estimated eosinophils. However, we observed the same attenuation of associations as the model adjusted for all 7 estimated cell counts.

**Enrichment analysis/functional analysis/gene ontology**

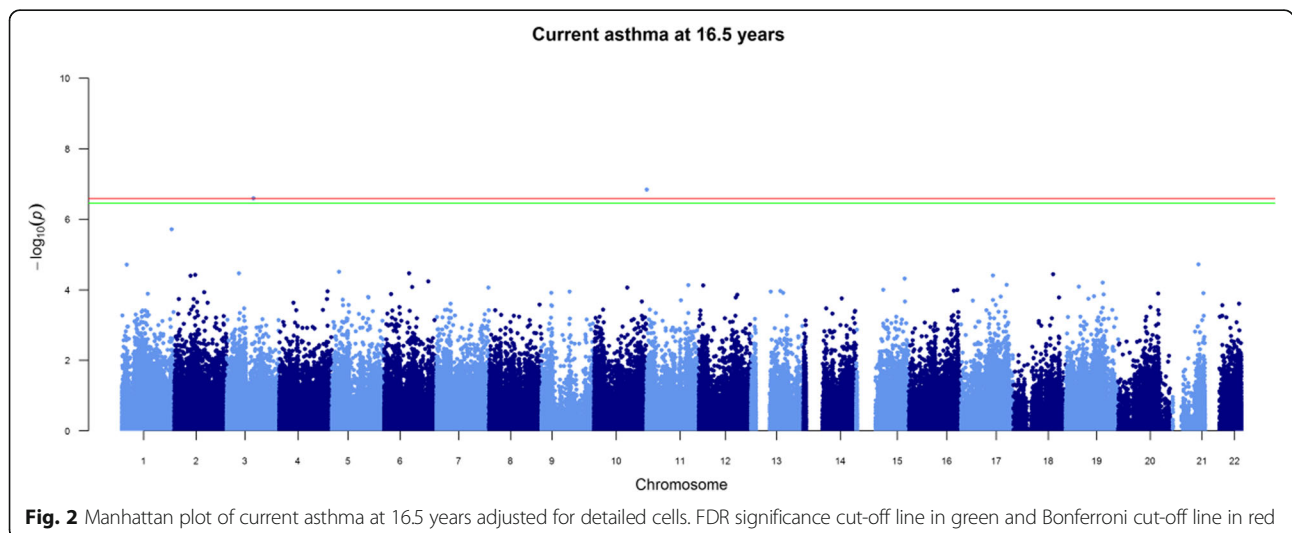
A look up in the GO [51] and KEGG [53] databases showed that the 192 genes annotated to the 302 hit CpGs from the EWAS of current asthma at 7.5 years

adjusted for basic cells were located in pathways related to movement of cellular or subcellular components, locomotion, interleukin-4 production and eosinophil migration. The genes were enriched in the following KEGG pathways: amino/nucleotide sugar metabolism and asthma (Additional file 1: Tables S7 and S8).

**Mendelian randomization**

***Asthma to DNA methylation***

We found some evidence of a causal effect of asthma on DNA methylation at several CpG sites (Additional file 2: Table E19). One CpG was among the top 20 hits from the



**Fig. 2** Manhattan plot of current asthma at 16.5 years adjusted for detailed cells. FDR significance cut-off line in green and Bonferroni cut-off line in red



**Table 4** Results of the longitudinal epigenome-wide association studies (EWAS)

DNA methylation	Asthma definition	N [cases/controls]	Cell adjustment	FDR hits [Bonferroni]	Lambda
7.5 years	Current asthma at 16.5 years	611 [182/429]	No cells	4 [2]	1.11
			Basic cells	9 [2]	1.13
			Detailed cells	0 [0]	1.10
16.5 years	Current asthma at 7.5 years	784 [146/638]	No cells	28 [0]	1.24
			Basic cells	1 [1]	1.16
			Detailed cells	0 [0]	1.14

FDR-adjusted *P* value and Bonferroni hits use a cut-off of <0.05

EWAS of current asthma at 7.5 years, cg12077460 annotated to gene *MFHAS1* with  $-0.022$  difference in methylation in asthma cases ( $P = 0.013$ ) shown in Fig. 3. However, none of the causal effect estimates met the calculated significance threshold of  $P = 3.4 \times 10^{-4}$  calculated based on the number of effective independent tests ( $n = 151$ ).

In the two CpG sites associated with asthma at 16.5 years, there was no evidence of a causal effect of asthma on DNA methylation, at either the *AP2A2* gene (cg17676835) or *IL5RA* (cg10159529) shown in Additional file 2: Table E21.

#### DNA methylation to asthma

In the reverse direction, from DNA methylation to asthma, 31 out of 302 CpGs were found to have *cis*-SNPs that were available in the meta-analysis results of the GABRIEL asthma GWAS. We were therefore only able to test the causal association at these 31 CpGs. There was evidence for a causal association between one CpG (cg11938718) and asthma annotated to the *HPCAL1* gene with a 0.0177 change in log odds of asthma per 1% increase in methylation ( $P = 2.5 \times 10^{-3}$ ) shown in Additional file 2: Table E22. The strength of the effect did not meet the significance threshold calculated based on the number of independent tests ( $n = 21$ , adjusted  $P = 2.4 \times 10^{-3}$ ). The effect of the CpG on asthma was in the same direction as the observational effect from the EWAS (shown in Additional file 2: Table E2).

Neither of the two CpG sites associated with asthma at 16.5 years had available *cis*-SNPs that could be used as IVs. We were therefore unable to test the causal association in the reverse direction.

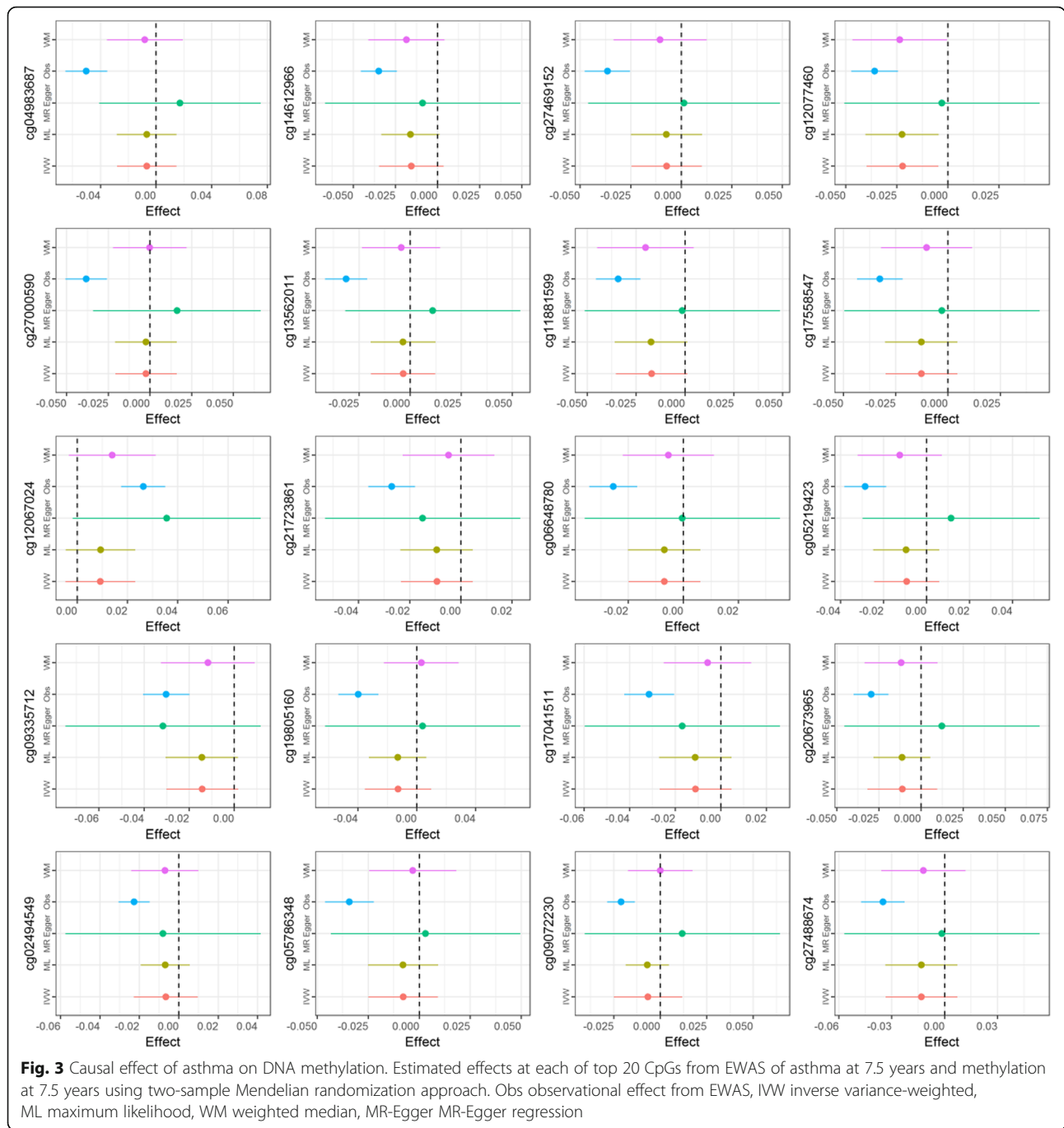
#### Discussion

We have found multiple differentially methylated CpG sites in peripheral blood of asthmatics compared to that of non-asthmatics in childhood. However, adjustment for eosinophils and neutrophils attenuated all associations in childhood. This is in line with our expectations of eosinophils being the driving factor in the detected signals. Eosinophilic inflammation is considered a characteristic of allergic asthma [27, 63], which is more common in childhood than adulthood. As expected, we find a far weaker

association of asthma at 16.5 years with eosinophil cell counts, as adult onset asthma is more often non-atopic, with eosinophilic asthma representing a rarer and more severe endotype of asthma at that age [16]. Differences in the number of CpGs associated with asthma in childhood (at 7.5 years) compared to adolescence (at 16.5 years) are most likely a reflection of the differences in prevalence of allergic asthma between childhood and adolescence. Overall, we found markedly fewer associations at 16.5 years than at 7.5 years with respect to current asthma and adjustments for cell counts at 16.5 years of age did not have the same impact as adjustments at 7.5 years, which is consistent with our hypothesis that eosinophil confounding in adolescence has weaker effects than in childhood. Overall, these observations are consistent with the likelihood that childhood asthma does not affect DNA methylation directly in our sample but does affect estimates of DNA methylation from a mixed cell population due to the association between asthma and the constitution of the cell population.

We found more associations between current wheeze and DNA methylation at 7.5 years (611 hit CpG sites) than for current asthma (411 hit CpG sites) at the same age. However, we observed the same attenuation in effects when adjusting for cell counts that include eosinophils in the current wheeze model. In the unadjusted model of current wheeze, the CpG site with the strong association (cg12077460) mapped to the *MFHAS1* gene, whereas the top association in current asthma mapped to *ZFPM1* (cg04983687). Asthma medications are known to reduce eosinophils in the blood [64], and study participants that report current wheeze may be more likely to be untreated asthma cases than those reporting doctors' diagnosis of asthma. Therefore, current wheeze cases may have a marginally higher proportion of eosinophils than those defined as current asthma which explains the higher number of associations in the wheeze model that attenuate upon adjustments for estimated eosinophils.

In the EWAS of current asthma at 16.5 years, the two CpGs that remained significant after adjustment for detailed cells mapped to the *AP2A2* and *IL5RA* genes. In the EWAS of current asthma at 7.5 years, cg17676835 (*AP2A2*) was not associated with current asthma at



7.5 years when adjusted for granulocytes only (basic cells), whereas cg10159529 (*IL5RA*) was strongly associated in the same direction ( $P = 0.0001$ ) but attenuated when adjusted for cell counts that included eosinophils and neutrophils ( $P = 0.304$ ).

In the longitudinal EWAS, we observed similar attenuation of all associations to the cross-sectional models once adjustments for eosinophils and neutrophils were made. Inflation persisted even after adjustments, indicating residual confounding or unaccounted for variation.

Notably, cg04983687 in the *ZFPM1* gene, which was the top hit in the cross-sectional model of current asthma at 7.5 years, remained nominally associated in both longitudinal models ( $P = 0.008$  for current asthma at 7.5 years and methylation in adolescence, and  $P = 0.01$  for current asthma at 16.5 years and methylation childhood) with the same direction of effect. Several other CpGs from the top 20 hits in the cross-sectional model of current asthma at 7.5 years, such as cg20673965 (*IRGC*) and cg19805160 (*CCDC19*), also demonstrated consistent

direction of effect in both of the longitudinal models despite only weak evidence of an association.

In sensitivity analysis of the sex-stratified EWAS models, we did not identify any single site associated with current asthma at 7.5 years in females indicating that the strong confounding effect seen in the mixed sex EWAS appears to be driven by differences in eosinophils mostly in males. This observation fits with the results of the sex-stratified associations of cell counts and asthma that showed that the strongest association of eosinophils and asthma was in the male subgroup. Several differentially methylated CpG sites in the EWAS of current asthma at 7.5 years were located in genes that have known links to both eosinophil functions and allergic sensitisation, such as cg27469152 in the 3'UTR region of the *EPX* gene. The gene encodes eosinophil peroxidase, an enzyme expressed in eosinophils [65] and released at sites of parasitic infection or allergic stimulation [66, 67]. Both current asthma at 16.5 years and ever wheeze at 16.5 years were negatively associated with methylation levels at *IL5RA*, a gene encoding alpha subunit of the interleukin-5 (IL5) receptor. Decreased expression of IL5 membrane receptor on eosinophils has been previously linked to allergen stimulation and increased IL5 protein levels in the airways [68]. The fact that the association remains significant after adjustment for detailed cell counts in the model of current asthma at 16.5 years may indicate incomplete adjustment for eosinophils. Genetic polymorphisms of the *AP2A2* gene have also previously found to be associated with bronchitis and chronic obstructive pulmonary disease (COPD) [69].

#### Cell type heterogeneity

The increase of eosinophils in some individuals with asthma presents challenges when assessing whole blood DNA methylation [29, 30], as the distinct methylation patterns of eosinophils [31] are over-represented in cases. Correction for white blood cell heterogeneity can be implemented; however, doing so adjusts for known characteristics of the disease outcome (asthma), therefore making it difficult to decipher methylation-related differences that are truly linked to the disease outcome.

Adjusting for cell type heterogeneity that may exist between different samples in an EWAS is of critical importance [29] and multiple different methods have been devised to address the issue, which may be considered as a form of confounding [70], although this categorisation is dependent on the hypothesis being tested. However, when a specific cell type forms an intermediate phenotype with involvement in the pathogenesis and characteristic features of a disease, then the unadjusted signals may be of biological interest. Raised eosinophil levels have been suggested to be an intermediate phenotype or

feature of asthma [71]. In the case of asthma, detected CpG signal from eosinophils may tag potentially pathogenic genes that might be causal to the asthma phenotype, when expressed differentially in blood due to higher numbers of eosinophils. By adjusting for estimated eosinophil counts, that signal is attenuated but may still have biological relevance.

#### Causal analyses

Methods of causal inference such as MR help to define direction of causation in detected associations. Our MR analysis suggested a causal effect of asthma on DNA methylation at several CpG sites associated with asthma at 7.5 years. However, none of the effects survived multiple testing, even after adjusting the *P* value threshold to take into account correlation between the outcomes. Similarly, in the reverse direction, from DNA methylation to asthma, there was no robust evidence for causal effects. When compared to the observational effects from the EWAS, estimates from the MR analysis were weaker, but with consistent direction of effect to the observational evidence. The fact that effect estimates from the MR were so weak in comparison to estimates from the EWAS further confirms lack of power in the MR analysis, as indicated by our power calculation. We used a combination of complementary methods to ascertain these causal effects in MR, including the fixed effect meta-analysis of Wald ratios, the inverse variance-weighted (IVW) method and the MR-Egger regression. The two-sample MR analysis was underpowered due to small sample size used to determine the SNP methylation association, comprising just over 1000 individuals (from the ARIES-ALSPAC sample), and we were not able to determine causality in either direction. Future work should aim to increase power by increasing the sample size in the dataset.

Based on these results, we also attempted a two-sample MR analysis of eosinophils and DNA methylation using previously published SNPs that have shown strong association with eosinophil cell counts in peripheral blood [72]. However, the derived IVs from the published GWAS results were too weak to detect an effect in our sample. Future studies may identify further sequence variants that associate with eosinophil counts and provide more powerful IVs for MR analysis.

The genetic instruments used in the asthma to DNA methylation direction of the MR analysis were relatively weak in comparison to other IVs used in previous MR studies. Asthma GWAS from which the IVs were sourced have identified SNPs that explain very little of the population variance in asthma in comparison to other traits [6, 73]. Furthermore, our strict LD  $R^2$  cut-off threshold of  $<0.1$  reduced the total amount of SNPs used as IVs to just six. In the reverse direction, from

DNA methylation to asthma, we included only one *cis*-SNP for each CpG tested since strong LD was observed between neighbouring SNPs that forced the exclusion of many *cis*-SNPs that would have otherwise been combined to form stronger IVs.

### Comparisons with previous literature

A previously published EWAS of total serum immunoglobulin E (IgE) identified 36 CpG sites associated with differential levels of IgE [21]. IgE is a marker of allergic asthma and, as such, would be expected to be implicated in the biological pathway leading to allergic or inflammatory disease [74]. Whereas IgE is a robust biomarker of inflammatory conditions, it may not accurately proxy asthma as it encompasses a wide variety of atopic or inflammatory conditions. Of the 36 identified CpGs in this previous study, 35 were analysed in our EWAS, with 15 associated with current asthma at 7.5 years (FDR < 0.05), and 13 remaining associated after adjustment for basic cells (FDR < 0.05). Comparisons of results can be seen in Additional file 1: Table S11. This corroboration of findings suggests that eosinophil specific methylation patterns may be directly involved in asthma pathogenesis via raised IgE.

A recent EWAS of IgE in Hispanic children [22] reported associations with DNA methylation in CpGs annotated to the *ZFPM1* (*cg04983687*, *cg08940169*) and *ACOT7* (*cg09249800*, *cg21220721*, *cg11699125*) genes. We observe similar strong associations with CpG sites annotated to the same genes in the unadjusted EWAS of current asthma at 7.5 years (Additional file 2: Tables E1 and E2). However, these associations attenuate following adjustment for eosinophils and neutrophils, unlike the previously reported results. Furthermore, we have excluded all CpGs except *cg04983687* (*ZFPM1*) from the results of our analyses as they have previously been indicated to be low-quality probes due to the effects of repeats, SNPs, INDELS and reduced genome complexity [46]. Differences between results may indicate residual confounding due to eosinophils in previous results or lack of power in our own analyses, due to a lower number of asthma cases compared to controls relative to the previous study. Additionally, we have used exclusively methylation-estimated cell counts to adjust for cell mixture as opposed to cell sorted estimates of the previous study, leading to potentially imprecise adjustments in our own analyses. In a sensitivity analysis, we generated estimated cell counts using the Houseman algorithm [41] and the 200 reference CpGs supplied in the supplementary material of the previous study [22]. Correlations between the estimated cell counts using the estimated reference CpGs and our own Houseman-derived estimates were relatively low. For eosinophils, the correlation was 0.77, for neutrophils 0.96, for NK cells 0.57, for CD4T cells 0.75, for B cells 0.71, for CD8T cells 0.13

and for monocytes 0.76. As a sensitivity analysis using these estimated cell counts we reran the EWAS of asthma at 7.5 years; however, there were no associations between DNA methylation and asthma (FDR < 0.05).

A previously published EWAS of asthma performed in peripheral blood mononuclear cells (PBMCs) found similar associations to our own unadjusted results [23]. The EWAS focused on children with atopic asthma living in the inner city with an African-American, Hispanic or Caribbean background. The study identified differentially methylated regions (DMRs) as opposed to single sites, with many of the resulting associations mapping to genes such as *ACOT7* and *IL13* that were also identified in our own unadjusted analyses. The study used the same algorithm to estimate cell counts; however, EWAS models were only adjusted for residual granulocyte contamination and not eosinophils and neutrophils separately, since the study used PBMC samples (which exclude granulocytes). Differences in samples (PBMCs vs whole blood) make it difficult to compare the study's results to our own findings. As is evident from our analyses, adjusting for granulocytes alone does not appear to be sufficient for countering the confounding effects of eosinophils in EWAS using whole blood; however, EWAS based on DNA methylation from PBMC samples may be less affected by cell-type differences. Similarly, a study of asthma using methylation data from nasal epithelium observed associations between asthma and DNA methylation but did not specifically adjust for eosinophil cell counts [24]. However, the authors in this study did attempt to perform cell sorting which may have negated some of the confounding effects of cell mixtures by ensuring a more homogenous cell sample enriched for epithelial cells. We therefore cannot conclude if confounding by eosinophils affects their results. The authors of both of these studies however do not appear to have discussed confounding effects of differences in eosinophils per se between study cases and controls.

### Strengths and limitations

Strengths of the study include the ability to compare two different time points in the same cohort of individuals, with methylation from childhood and adolescence, forming a series of cross-sectional and longitudinal analyses. Our use of complementary approaches to answering each question, which includes the functional analysis, sensitivity analyses and bi-directional MR (with multiple different testing methods applied), gives greater strength to our conclusions.

Limitations include the low coverage of the HumanMethylation450 Beadchip array, including less than 2% of all CpG sites [75]. Combined with our various probe exclusions, this brings the overall coverage down to just a fraction of the full methylome. Our



analyses rely on the use of methylation data generated from peripheral blood leukocytes (PBLs), which may not be the most appropriate tissue in which to study associations with asthma. Cells from other tissues may be more relevant, such as airway epithelial cells, buccal cells or cells from bronchoalveolar lavage. In addition, our cell counts are estimated from the methylation data using a reference dataset instead of being directly sorted and counted in the original blood samples. Whereas the method has been previously validated [76] and previous EWAS analyses have found strong correlations between estimated and cell sorted measures [22], it may not perform well in predicting proportions of rarer cell types in the blood, such as eosinophils, as they may be subject to less favourable signal-to-noise ratios than high abundance cell types. However, we attempted to counter the potentially poor estimation of cell counts by using surrogate variables in our EWAS analysis, a reference-free method of adjusting for unknown confounding factors. We did not assess to what extent our detected associations were explained by genetic variants, such as SNPs.

Measurement error due to misreporting of asthma diagnosis is another limitation. Since our outcomes are derived from self-report questionnaires from either the mothers of the study children or the study children themselves, they are subject to misreporting of symptoms or previous asthma diagnosis. Any misreporting of asthma/whoeze symptoms is likely to bias results towards the null. However, a validation study linking childhood asthma self-reports in ALSPAC using linked electronic patient records from GP practices shows that parental reports of a doctor's diagnosis agree well with GP-recorded diagnosis [77]. The Mendelian randomization is robust to measurement error in the exposure due to the use of genetic variants but may be subject to other biases in exposure and outcome [78].

Finally, whereas we were able to perform the analyses in a large well-phenotyped birth cohort, we were limited by the number of individuals with available DNA methylation data at each time point and analyses may have been underpowered to detect small effects in the presence of strong confounding by cell type. Future studies should aim to increase sample size by meta-analysis in order to ensure adequate power to detect effects, overcoming the confounding effects of differential cell counts between asthma cases and controls.

## Conclusion

We have shown multiple significant associations between asthma status and peripheral blood DNA methylation in both childhood and adolescence. There is strong evidence to suggest that the observed associations are driven to a

large extent by higher eosinophil counts in asthmatics in childhood. We have shown that the use of whole-blood 450 K data in EWAS of allergic diseases, such as asthma, where eosinophils may be expected to play a contaminating or confounding role in disease cases, makes it difficult to disentangle true associations of the disease with methylation from associations that may be driven by the over-representation of eosinophils. Whereas further work is needed to devise methods of elucidating true associations from those that are due to cell type confounding, this does not fully preclude the biological relevance of the observed differential methylation as raised eosinophils are a constituent part of the asthma phenotype.

## Additional files

**Additional file 1:** Supplementary Tables and Figures. (DOCX 2448 kb)

**Additional file 2:** Supplementary EWAS Tables E. (XLSX 2170 kb)

**Additional file 3:** Supplementary EWAS Tables F.(XLSX 645 kb)

## Abbreviations

ALSPAC: Avon Longitudinal Study of Parents and Children; ARIES: Accessible Resource for Integrated Epigenomics Studies; CNS: Centre for neurogenetics and statistical genomics; COPD: Chronic obstructive pulmonary disorder; DAVID: Database for Annotation Visualisation and Integrated Discovery; DMR: Differentially methylated region; EWAS: Epigenome-wide association study; FDR: False discovery rate; GO: Gene Ontology; GWAS: Genome-wide association study; IgE: Immunoglobulin E; IQR: Interquartile range; IV: Instrumental variable; IWW: Inverse variance-weighted; KEGG: Kyoto Encyclopedia of Genes and Genomes; LD: Linkage disequilibrium; ML: Maximum likelihood; MR: Mendelian randomization; NK: Natural killer cells; OR: Odds ratio; PBL: Peripheral blood leukocyte; PBMC: Peripheral blood mononuclear cells; QC: Quality control; Q-Q: Quantile-quantile plot; RCT: Randomized control trial; SNP: Single-nucleotide polymorphisms; UTR: Untranslated region; WM: Weighted median

## Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We are grateful to Dr. Philip Haycock for his work in curating GWAS data as part of MR base. We are also grateful to Mr. Sam Abbott for his help with ggplot and Mr. Ryan Langdon for his help in implementing Mendelian randomization for eosinophils.

## Funding

This work was supported by the MRC Integrative Epidemiology Unit and the University of Bristol (MC\_UU\_12013\_2, MC\_UU\_12013\_5 and MC\_UU\_12013\_8). The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/1025751/1 and BB/1025263/1). GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The UK Medical Research Council and Wellcome (Grant Ref. 102215/2/13/2) and the University of Bristol provide core support for ALSPAC.

## Availability of data and materials

ALSPAC data management policies do not permit datasets to be made publicly available due to data confidentiality and the potential to identify individual study participants from the data. Data used will be made available following approved request to the ALSPAC Executive (alspac-exec@bristol.ac.uk). The ALSPAC data



management plan is available online: <http://www.bristol.ac.uk/media-library/sites/alspac/migrated/documents/alspac-data-management-plan.pdf>.

#### Authors' contributions

RA, MS, KT, CLR, JH, RG, KB, RR and GCS conceived and designed the analysis. RA, MS and KT analysed the data. RA and CLR wrote the manuscript. RA, MS, KT, CLR, JH, RG, KB, TRG, RR and GCS revised and edited the manuscript. SR and TRG contributed to the data production and management. This publication is the work of the authors and Mr. Ryan Arathimos and will serve as guarantors for the contents of this paper. All authors read and approved the final manuscript.

#### Ethics approval and consent to participate

Written informed consent has been obtained for all ALSPAC participants. The ethical approval was granted from the ALSPAC Law and Ethics Committee and the local Research Ethics Committee in accordance with the guidelines of the Declaration of Helsinki.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>School of Social and Community Medicine, University of Bristol, Bristol, UK. <sup>2</sup>Medical Research Council Integrative Epidemiology Unit, University of Bristol, Oakfield House, Bristol BS8 2BN, UK. <sup>3</sup>School of Oral and Dental Sciences, University of Bristol, Bristol, UK.

Received: 10 May 2017 Accepted: 2 October 2017

Published online: 13 October 2017

#### References

- Winer RA, Qin X, Harrington T, Moorman J, Zahran H. Asthma incidence among children and adults: findings from the behavioral risk factor surveillance system asthma call-back survey—United States, 2006–2008. *J Asthma*. 2012;49(1):16–22. doi: 10.3109/02770903.2011.637594. Epub 2012/01/13. PubMed PMID: 22236442
- Townsend EA, Miller VM, Prakash YS. Sex differences and sex steroids in lung health and disease. *Endocr Rev*. 2012;33(1):1–47.
- Mantzouranis E, Papadopoulou E, Michailidi E. Childhood asthma: recent developments and update. *Curr Opin Pulm Med*. 2014;20(1):8–16. doi: 10.1097/mcp.0000000000000014. Epub 2013/11/19. PubMed PMID: 24240439
- Ober C. The genetics of asthma and allergic disease: a 21(st) century. *Perspective*. 2011;242(1):10–30. doi: 10.1111/j.1600-065X.2011.01029.x. PubMed PMID: 21682736; PubMed Central PMCID: PMC3151648
- Weiss ST, Silverman EK. Pro: genome-wide association studies (GWAS) in asthma. *Am J Respir Crit Care Med*. 2011;184(6):631–3. doi: 10.1164/rccm.201103-0485ED. Epub 2011/09/17. PubMed PMID: 21920925
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med*. 2010;363 doi: 10.1056/NEJMoa0906312.
- Yang IV, Schwartz DA. Epigenetic mechanisms and the development of asthma. *J Allergy Clin Immunol*. 2012;130(6):1243–55. doi: 10.1016/j.jaci.2012.07.052.
- Bégin P, Nadeau KC. Epigenetic regulation of asthma and allergic disease. *Allergy, Asthma Clin Immunol*. 2014;10(1):1–12. doi: 10.1186/1710-1492-10-27.
- Durham AL, Wiegman C, Adcock IM. Epigenetics of asthma. *Biochim Biophys Acta*. 2011;1810(11):1103–9. doi: 10.1016/j.bbagen.2011.03.006. Epub 2011/03/15. PubMed PMID: 21397662.
- Lee KW, Pausova Z. Cigarette smoking and DNA methylation. *Front Genet*. 2013;4 doi: 10.3389/fgene.2013.00132. PubMed PMID: 23882278; PubMed Central PMCID: PMC3713237
- Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, et al. Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. *Environ Health Perspect*. 2016;124(7):983–90. doi: 10.1289/ehp.1509966. Epub 2016/01/06. PubMed PMID: 26731791; PubMed Central PMCID: PMC4937859
- Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, et al. Rapid DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med*. 2009;179. doi: 10.1164/rccm.200807-1097OC.
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet*. 2013;22(5):843–51. doi: 10.1093/hmg/ddt488. Epub 2012/11/24. PubMed PMID: 23175441
- Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*. 2013;8(5):e63812. doi: 10.1371/journal.pone.0063812.
- Arathimos R, Granell R, Henderson J, Relton CL, Tilling K. Sex discordance in asthma and wheeze prevalence in two longitudinal cohorts. *PLoS One*. 2017;12(4):e0176293. doi: 10.1371/journal.pone.0176293.
- de Nijs SB, Venekamp LN, Bel EH. Adult-onset asthma: is it really different? *Eur Respir Rev*. 2013;22(127):44–52. doi: 10.1183/09059180.00007112.
- Martinez FD. The connection between early life wheezing and subsequent asthma: the viral march. *Allergol Immunopathol*. 2009;37(5):249–51. doi: 10.1016/j.aller.2009.06.008. Epub 2009/10/31. PubMed PMID: 19875225
- Martinez FD, Helms PJ. Types of asthma and wheezing. *Eur Respir J Suppl*. 1998;27:3s–8s. Epub 1998/08/12. PubMed PMID: 9699776.
- Castro-Rodríguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med*. 2000;162(4):1403–6. doi: 10.1164/ajrccm.162.4.9912111.
- Martinez JAB. Not all that wheezes is asthma! *J Bras Pneumol*. 2013;39(4):518–20. doi: 10.1590/S1806-37132013000400017. PubMed PMID: PMC4075862
- Liang L, Willis-Owen SAG, Laprise C, Wong KCC, Davies GA, Hudson TJ, et al. An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature*. 2015;520(7549):670–674. doi: 10.1038/nature14125, <http://www.nature.com/nature/journal/v520/n7549/abs/nature14125.html#supplementary-information>
- Chen W, Wang T, Pino-Yanes M, Forno E, Liang L, Yan Q, et al. An epigenome-wide association study of total serum IgE in Hispanic children. *J Allergy Clin Immunol*. 2017; doi: 10.1016/j.jaci.2016.11.030.
- 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(1):69–80. doi: 10.1016/j.jaci.2015.01.025. PubMed PMID: 25769910; PubMed Central PMCID: PMC4494877
- 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*. 2016; doi: 10.1016/j.jaci.2016.07.036.
- Bousquet J, Chané P, Lacoste JY, Barneon G, Ghavanian N, Enander I, et al. Eosinophilic inflammation in asthma. *N Engl J Med*. 1990;323(15):1033–9. doi: 10.1056/nejm199010110113231505. Epub 1990/10/11. PubMed PMID: 2215562
- Walford HH, Doherty TA. Diagnosis and management of eosinophilic asthma: a US perspective. *Journal of Asthma and Allergy*. 2014;7:53–65. doi: 10.2147/jaa.s39119. PubMed PMID: 24748808; PubMed Central PMCID: PMC3990389
- Possa SS, Leick EA, Prado CM, Martins Mí A, Tibério I. Eosinophilic inflammation in allergic asthma. *Front Pharmacol*. 2013;4 doi: 10.3389/fphar.2013.00046. PubMed PMID: 23616768; PubMed Central PMCID: PMC3627984
- Walsh ER, Stokes K, August A. The role of eosinophils in allergic airway inflammation. *Discov Med*. 2010;9(47):357–62. Epub 2010/04/29. PubMed PMID: 20423680
- Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15 doi: 10.1186/gb-2014-15-2-r31.
- Birney E, Smith GD, Grealley JM. Epigenome-wide association studies and the interpretation of disease—omics. *PLoS Genet*. 2016;12(6):e1006105. doi: 10.1371/journal.pgen.1006105.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7(7):e41361. doi: 10.1371/journal.pone.0041361.
- Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol*. 2012;41(1):161–76. doi: 10.1093/ije/dyr233.

33. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22. doi: 10.1093/ije/dyg070.
34. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23 doi: 10.1093/hmg/ddu328.
35. Allard C, Desgagne V, Patenaude J, Lacroix M, Guillemette L, Battista MC, et al. Mendelian randomization supports causality between maternal hyperglycemia and epigenetic regulation of leptin gene in newborns. *Epigenetics.* 2015;10(4):342–51. doi: 10.1080/15592294.2015.1029700. Epub 2015/03/25. PubMed PMID: 25800063; PubMed Central PMCID: PMC4622547
36. Hemani G, Zheng X, Wade KH, Laurin C, Elsworth B, Burgess S, et al. MR-base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv.* 2016. doi: doi: 10.1101/078972.
37. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 2012; doi: 10.1093/ije/dys064.
38. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol.* 2015; doi: 10.1093/ije/dyv072.
39. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics.* 2011;98(4):288–95. doi: 10.1016/j.ygeno.2011.07.007. Epub 2011/08/16. PubMed PMID: 21839163
40. 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(10):1363–9. doi: 10.1093/bioinformatics/btu049. PubMed PMID: 24478339; PubMed Central PMCID: PMC4016708
41. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics.* 2012;13:86. doi: 10.1186/1471-2105-13-86. Epub 2012/05/10. PubMed PMID: 22568884; PubMed Central PMCID: PMC3532182
42. Roberts G, Peckitt C, Northstone K, Strachan D, Lack G, Henderson J, et al. Relationship between aeroallergen and food allergen sensitization in childhood. *Clin Exp Allergy.* 2005;35(7):933–40. doi: 10.1111/j.1365-2222.2005.02280.x. Epub 2005/07/13. PubMed PMID: 16008681
43. StataCorp. Stata statistical software: release 13. College Station: StataCorp LP2013.
44. Barfield RT, Kilaru W, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics.* 2012;28(9): 1280–1. doi: 10.1093/bioinformatics/bts124. PubMed PMID: PMC3577110
45. 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; doi: 10.1093/bioinformatics/bts034.
46. Naeem H, Wong N, Chatterton Z, Hong MK, Pedersen J, Corcoran N, et al. Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics.* 2014;15(1):51. doi: 10.1186/1471-2164-15-51. PubMed PMID
47. Tukey J. *Exploratory Data Analysis.* Addison-Wesley: 1977. p. 43–4.
48. Church RM. How to look at data: a review of John W. Tukey's exploratory data analysis. *J Exp Anal Behav.* 1979;31(3):433–40. doi: 10.1901/jeab.1979.31-433.
49. Douwes J, Gibson P, Pekkanen J, Pearce N. Non-eosinophilic asthma: importance and possible mechanisms. *Thorax.* 2002;57(7):643–8. doi: 10.1136/thorax.57.7.643.
50. Nair P. What is an eosinophilic phenotype of asthma? *J Allergy Clin Immunol.* 132(1):81–3. doi: 10.1016/j.jaci.2013.05.007.
51. Consortium TGO. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 2015;43(D1):D1049–D56. doi: 10.1093/nar/gku1179.
52. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* 2007;35(Web Server issue):W169–75. doi: 10.1093/nar/gkm415. PubMed PMID: 17576678; PubMed Central PMCID: PMC4622547
53. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27–30. PubMed PMID: 10592173; PubMed Central PMCID: PMC4622547
54. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1): R89–98. doi: 10.1093/hmg/ddu328.
55. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol.* 2004;33(1):30–42.
56. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol.* 2015;30(7):543–52. doi: 10.1007/s10654-015-0011-z.
57. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178(7):1177–84. doi: 10.1093/aje/kwt084. Epub 2013/07/19. PubMed PMID: 23863760; PubMed Central PMCID: PMC3783091
58. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(D1):D1001–D6. doi: 10.1093/nar/gkt1229.
59. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016;17(1):1–14. doi: 10.1186/s13059-016-0926-z.
60. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet.* 2004;74(4):765–9. PubMed PMID: PMC1181954.
61. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity.* 2005;95(3):221–7. doi: 10.1038/sj.hdy.6800717. Epub 2005/08/04PubMed PMID: 16077740
62. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42(5):1497–501. doi: 10.1093/ije/dyt179.
63. Wardlaw AJ, Brightling C, Green R, Woltmann G, Pavord I. Eosinophils in asthma and other allergic diseases. *Br Med Bull.* 2000;56(4):985–1003. Epub 2001/05/22. PubMed PMID: 11359633
64. Fulkerson PC, Rothenberg ME. Targeting eosinophils in allergy, inflammation and beyond. *Nat Rev Drug Discov.* 2013;12(2):117–29.
65. Metso T, Venge P, Haahtela T, Peterson CG, Seveus L. Cell specific markers for eosinophils and neutrophils in sputum and bronchoalveolar lavage fluid of patients with respiratory conditions and healthy subjects. *Thorax.* 2002; 57(5):449–51. Epub 2002/04/30. PubMed PMID: 11978925; PubMed Central PMCID: PMC4622547
66. Nair P, Ochkur SI, Protheroe C, Radford K, Efthimiadis A, Lee NA, et al. Eosinophil peroxidase in sputum represents a unique biomarker of airway eosinophilia. *Allergy.* 2013;68(9):1177–84. doi: 10.1111/all.12206. Epub 2013/08/13. PubMed PMID: 23931643; PubMed Central PMCID: PMC3788081
67. Ten RM, Pease LR, McKean DJ, Bell MP, Gleich GJ. Molecular cloning of the human eosinophil peroxidase. Evidence for the existence of a peroxidase multigene family. *J Exp Med.* 1989;169(5):1757–69. doi: 10.1084/jem.169.5.1757.
68. Liu LY, Sedgwick JB, Bates ME, Vrtis RF, Gern JE, Kita H, et al. Decreased expression of membrane IL-5 receptor alpha on human eosinophils: I. Loss of membrane IL-5 receptor alpha on airway eosinophils and increased soluble IL-5 receptor alpha in the airway after allergen challenge. *J Immunol.* 2002;169(11):6452–8. Epub 2002/11/22. PubMed PMID: 12444154.
69. Lee JH, Cho MH, Hersh CP, M-LN MD, Crapo JD, Bakke PS, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res.* 2014;15(1):113. doi: 10.1186/s12931-014-0113-2. PubMed PMID: PMC4190389
70. Lowe R, Rakyansky VK. Correcting for cell-type composition bias in epigenome-wide association studies. *Genome Medicine.* 2014;6(3):1–2. doi: 10.1186/gm540.
71. Holberg CJ, Halonen M, Wright AL, Martinez FD. Familial aggregation and segregation analysis of eosinophil levels. *Am J Respir Crit Care Med.* 1999; 160(5 Pt 1):1604–10. doi: 10.1164/ajrccm.160.5.9807043. Epub 1999/11/11. PubMed PMID: 10556128
72. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadóttir A, Sulem P, Jonsdóttir GM, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet.* 2009;41(3):342–7. [http://www.nature.com/ng/journal/v41/n3/supplinfo/ng.323\\_S1.html](http://www.nature.com/ng/journal/v41/n3/supplinfo/ng.323_S1.html).
73. Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, Lyon HN, et al. Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. *PLoS One.* 2012;7(9) doi: 10.1371/journal.pone.0044008. PubMed PMID: 23028483; PubMed Central PMCID: PMC4622547

74. Platts-Mills TAE. The role of immunoglobulin E in allergy and asthma. *Am J Respir Crit Care Med*. 2001;164(supplement\_1):S1–5. doi: 10.1164/ajrccm.164.supplement\_1.2103024.
75. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, et al. A comparison of the whole genome approach of MeDIP-Seq to the targeted approach of the Infinium HumanMethylation450 BeadChip(®) for methylome profiling. *PLoS One*. 2012;7(11) doi: 10.1371/journal.pone.0050233. PubMed PMID: 23209683; PubMed Central PMCID: PMC3510246
76. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13 doi: 10.1186/1471-2105-13-86.
77. Cornish RP, Henderson J, Boyd AW, Granell R, Van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. *BMJ Open*. 2014;4(4) doi: 10.1136/bmjopen-2014-005345.
78. Pierce BL, VanderWeele TJ. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. *Int J Epidemiol*. 2012;41(5):1383–93. doi: 10.1093/ije/dys141.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

