



Everson, T. M., Zhang, H., Lockett, G. A., Kaushal, A., Forthofer, M., Ewart, S., ... Karmaus, W. (2019). Epigenome-wide association study of asthma- and wheeze characterizes loci within HK1. *Allergy, Asthma and Clinical Immunology*, 15, [43 (2019)]. <https://doi.org/10.1186/s13223-019-0356-z>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1186/s13223-019-0356-z](https://doi.org/10.1186/s13223-019-0356-z)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer Nature at <https://aacijournal.biomedcentral.com/articles/10.1186/s13223-019-0356-z#article-info>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights


This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/pure/about/ebr-terms>

RESEARCH

Open Access



Epigenome-wide association study of asthma and wheeze characterizes loci within *HK1*

Todd M. Everson^{1,2}, Hongmei Zhang³, Gabrielle A. Lockett⁴, Akhilesh Kaushal⁵, Melinda Forthofer^{1,6}, Susan L. Ewart⁷, Kimberley Burrows⁸, Caroline L. Relton⁸, Gemma C. Sharp⁸, A. John Henderson⁹, Veeresh K. Patil¹¹, Faisal I. Rezwan⁴, S. Hasan Arshad^{10,11,12}, John W. Holloway^{4,10} and Wilfried Karmaus^{3*} 

Abstract

Background: To identify novel epigenetic markers of adolescent asthma and replicate findings in an independent cohort, then explore whether such markers are detectable at birth, predictive of early-life wheeze, and associated with gene expression in cord blood.

Methods: We performed epigenome-wide screening with recursive random forest feature selection and internal validation in the IOW birth cohort. We then tested whether we could replicate these findings in the independent cohort ALSPAC and followed-up our top finding with children of the IOW cohort.

Results: We identified 10 CpG sites associated with adolescent asthma at a 5% false discovery rate (IOW, $n = 370$), five of which exhibited evidence of associations in the replication study (ALSPAC, $n = 720$). One site, cg16658191, within *HK1* displayed particularly strong associations after cellular heterogeneity adjustments in both cohorts ($OR_{IOW} = 0.17$, 95% CI 0.04–0.57) ($OR_{ALSPAC} = 0.57$, 95% CI 0.38–0.87). Additionally, higher expression of *HK1* ($OR = 3.81$, 95% CI 1.41–11.77) in cord blood was predictive of wheezing in infancy ($n = 82$).

Conclusion: We identified novel associations between asthma and wheeze with methylation at cg16658191 and the expression of *HK1*, which may serve as markers of, predictors of, and potentially etiologic factors involved in asthma and early life wheeze.

Keywords: ALSPAC, ARIES, Asthma, Expression, *Hexokinase-1*, *HK1*, Infant wheeze, Isle of Wight, Methylation

Background

Asthma is a common chronic respiratory disease particularly among children [1], causing substantial health care costs [2]. Asthma has complex pathophysiology [3], phenotypic variability [4], and is polygenic with a high heritable component, yet GWAS discoveries can only explain a small fraction of asthma variance [5]. Epigenetic mechanisms, which regulate gene expression potential [6], have received attention

in studies of asthma because these mechanisms may be influenced via environmental exposures, particularly exposures that occur in utero [7, 8]. One of the most studied epigenetic mechanism is DNA methylation (DNAm), which is the covalent addition of a methyl group to the DNA at a cytosine residue that is followed by a guanine (CpG site); acting as an important regulator of gene transcription [6].

Recent epigenetic epidemiology studies have associated DNAm in blood with current asthma and/or wheeze in childhood and adolescence [9, 10], in cord blood with childhood asthma [11], and as a possible mediator between maternal and offspring asthma [12]. DNA methylation has also been implicated as a

*Correspondence: Karmaus1@memphis.edu

³ Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health, University of Memphis, 236A Robison Hall, Memphis, TN 38152, USA

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



possible mediator of the relationships between both environmental and genetic factors with asthma. For instance, environmental exposure to air pollution is a risk factor for asthma exacerbations as well as asthma onset [13]. Recent studies have shown that air pollution is associated with differential DNAm of *TET1* [14] and *FOXP3* [15], and that differential methylation of these genes associates with asthma, suggesting that epigenetic regulation has a potential mediating role. Additionally, *GSDMB* and *ORMDL3* are two well-recognized asthma susceptibility genes [16], and recent studies have shown that DNAm may be a mediator between genetic variation and the expression of these genes [17]. These studies provide supportive evidence that epigenetic mechanisms may be involved in the etiology of asthma, potentially as intermediates between recognized risk factors and the development of symptoms. A recent epigenome-wide meta-analysis of multiple European cohorts identified robust associations between asthma and blood DNA methylation throughout childhood (4–8 years of age), which retained strong associations with asthma status among isolated eosinophils and these epigenetic signatures were indicative of eosinophil and cytotoxic cell activation [18]. The above studies highlight the mounting evidence that differential epigenetic regulation of specific genes contributes to asthma etiology, and that epigenome-wide approaches have led to the identification of novel asthma-associated epigenomic loci. Performing additional EWAS in independent populations with different ages and different asthmatic phenotypes can improve our understanding of which loci are informative across multiple populations, how these epigenetic variations relate to asthma throughout the life course, and whether their methylation levels correlate with specific phenotypic characteristics, such as inflammation or lung function.

In the current study, we performed an EWAS using both a standard CpG-by-CpG approach as well as an innovative feature selection method to identify novel epigenetic markers of prevalent asthma in 18 years olds, investigated if the identified loci were predictive of early-life wheeze, and whether DNA methylation at these loci were related to gene expression. We first conducted an exploratory epigenome-wide screening study of DNAm in whole blood within the Isle of Wight (IOW) birth cohort, followed by a replication study within the Avon Longitudinal Study of Parents and Children (ALSPAC). Then, with data from the offspring of the IOW birth cohort participants, we tested whether the same associations exist between cord blood DNAm and wheeze without upper respiratory viral infection (cold) within the 1st year of life, followed by testing for

associations between gene expression with DNAm and with infant wheeze.

Methods

The Isle of Wight birth cohort

The Isle of Wight (IOW) birth cohort is an unselected birth cohort of children born between January 1, 1989 and February 28, 1990 in Isle of Wight, UK. Details about the birth cohort have been described in detail elsewhere [19]. After exclusion of adoptions and prenatal deaths, 1456 children were enrolled and followed-up through to 18 years of age ($n=1313$; 90.2% retention). At each follow-up, participants were evaluated for manifestations of allergic disease and administered detailed questionnaires, including study specific questions, as well as questions derived from the International Study of Asthma and Allergies in Childhood (ISAAC), the most extensive international study of asthma, which led to the development and validation of questions about asthma and wheeze symptoms [20]. Ethical approval was obtained from National Research Ethics Service, NRES Committee South Central—Southampton B for the 18-year follow-up (06/Q1701/34) and NRES Committee South Central—Hampshire B (09/H0504/129) for the follow-up of IOW participants' offspring; written informed consent was provided by the infants' parents.

At the 18-year follow-up a subset of participants ($n=370$) were selected to take part in an epigenetic screening; this sample is referred to as the 'IOW F1 sample' herein. The primary outcome for this study was current asthma defined as having an asthma diagnosis and self-reported wheeze and/or use of asthma medications in the previous 12 months. Those attending the 18 years follow-up in person also performed spirometry and fractional-exhaled nitric oxide (FeNO), which were in accordance with American Thoracic Society (ATS) guidelines [21, 22], as well as allergen sensitization via skin prick tests (SPTs). Lung function assessments were performed using a Koko Spirometer and software with a desktop portable device (PDS Instrumentation, Louisville, USA). FeNO measurements (Niox mino, Aerocrine AB, Solna, Sweden) were obtained prior to spirometric assessments. Atopy was defined as having at least one positive SPT among 11 allergens (cows' milk, hens' egg, peanut, cod, house dust mite, cat, dog, *Alternaria alternata*, *Cladosporium herbarium*, grass pollen mix, and tree pollen mix). DNA was extracted from peripheral blood collected at the 18-year follow-up using a salting out procedure.

The IOW offspring (IOW F2 sample) are being enrolled in the IOW 3rd Generation study through ongoing recruitment since 2010. To date, 390 newborns have been enrolled; cord blood samples were collected at birth and

have been processed on 111 newborns for DNAm and 82 newborns for gene expression. Questionnaires about allergy and wheeze symptoms were administered to the parents at follow-up visits 3, 6, and 12 months after birth. The primary dependent variable for this sample was parent reported wheeze occurring when the infant had no symptoms of a cold. We also investigated any reported wheeze as an alternate outcome.

The ALSPAC cohort

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large, prospective cohort study based in the South West of England. In total, 14,541 pregnant women resident in Avon, UK with expected delivery dates between 1st April 1991 and 31st December 1992 were initially enrolled; 13,988 children were alive at 1 year [23, 24]. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees; written informed consent was provided by all participants. Self-completed questionnaires were administered during pregnancy and then at regular intervals. Current asthma status was obtained around the age of 17 years, defined as a reported doctor's diagnosis of asthma in addition to reported wheezing, asthma or the use of asthma medication in the previous 12 months. 5036 adolescents had complete phenotype data of which, 720 also had DNA methylation data from whole blood collected at an average age of 17 years. Genome-wide methylation measurements were conducted at the University of Bristol as part of Accessible Resource for Integrated Epigenomic Studies (ARIES) project (<http://www.ariesepigenomics.org.uk>) [25]. For the purposes of this study, multiple births and children of non-white ethnicity were excluded due to small numbers.

DNA methylation arrays

In the IOW F1 sample (18 years of age), IOW F2 sample (cord blood), and ALSPAC sample (17 years of age) DNAm was assessed genome-wide using the Illumina Infinium[®] HumanMethylation450k BeadChip (Illumina, Inc., CA, USA). The details of data processing steps are provided in Additional file 1: Method S1. Briefly, quality control and preprocessing methods for both cohorts included background correction, probe-type standardization, batch effect adjustments, and exclusion of potentially problematic probes. Methylation levels were calculated as beta (β) values, which can be interpreted as percent methylation. Because β -values can suffer from severe heteroscedasticity, M -values were calculated via $\log_2(\beta/(1-\beta))$ which better approximate a normal distribution. Cellular heterogeneity of blood samples was assessed by estimating the proportions of

CD8⁺ T-cells, CD4⁺ T-cells, natural killer cells, B-cells, monocytes, eosinophils and other granulocytes [26, 27] via the `estimateCellCounts` function in R. These proportions were included in our regression models as potential confounders of the relationship between DNA methylation and current asthma.

Gene expression array

At birth, IOW F2 cord blood samples were collected into PAXgene Bone Marrow RNA Tubes and RNA extracted using PAXgene RNA kits (PreAnalytiX GmbH, Switzerland). RNA integrity was verified with the Agilent 2100 Bioanalyzer system. Genome-wide mRNA expression was assessed via one color (Cy3) experiments with the Agilent (Agilent Technologies, Santa Clara, CA) SurePrint G3 Human Gene Expression 8 × 60 k v2 microarray kits. Array content was sourced from RefSeq, Ensembl, UniGene, and GenBank databases and provides full coverage of the human transcriptome in 50,599 biological features (including replicate probes and control probes). The oligos were 60mer in length and each transcript was tagged at least once and some had multiple tagging oligos for genes with documented splice variants. Data QC indices and analyses were performed with Agilent GeneSpring software. These data were then percent shift normalized and \log_2 -transformed.

Statistical analyses (discovery—IOW F1)

We randomly divided the IOW F1 (18 years of age) sample into two independent sub-samples. The stage-1 data ($n_{S1}=91$) were used for random forest (RF) feature selection because RFs rely on few statistical assumptions, are efficient with high-dimensional data, are robust to outliers and noise, and produce measures of variable importance [28, 29]. This feature selection technique was utilized in a recent epigenetic study of atopy that yielded many replicable loci [30]. The RF algorithm has a tendency to produce a predictor that is overfit to the supplied data; however, in our study RF was applied to select features based on variable importance rather than prediction. In addition, the RF algorithm was only applied to a subset of the IOW data to further diminish the possibility of overfitting, allowing us to examine the associations between DNAm and asthma in a statistically independent dataset. The stage-2 sample ($n_{S2}=279$) was larger to retain greater power, which was necessary for hypothesis testing and multiple testing adjustment.

Recursive RF feature selection was implemented on the stage-1 sample ($n_{S1}=91$) to select the CpGs most informative for asthma. We utilized balanced sampling, tested 10% of predictors per node ($mtry=0.10$) and grew forests with 7500 trees ($ntree=7500$). We implemented the RF recursively: (1) ran the RF algorithm on all available

predictors (248,336 CpGs) via the randomForest package in R, (2) extracted out-of-bag (OOB) misclassification rates and variable importance measures (VIMs), (3) sorted the predictors by their VIMs, (4) excluded half of the predictors with the smallest VIMs, and (5) repeated the sequence until the asthma-specific misclassification levelled off. Predictors from the final iteration were selected for stage-2 analyses.

M-values for the selected CpGs were tested for their associations with asthma status with logistic regression after trimming potential strong outliers identified with adjusted boxplots. We generated false discovery rate (FDR) adjusted p-values [31] via the q-value package in R; CpGs within a 5% FDR (q-values < 0.05) were considered 'discovered' and were candidates for the replication study.

Finally, we also performed a traditional EWAS regressing the beta-values for each individual loci on asthma status in unadjusted linear models, and models adjusted for sex, CD4⁺ T-cells, CD8⁺ T-cells, monocytes, eosinophils, natural killer, and granulocytes. Models that produced association within a 5% FDR (q-values < 0.05) were considered statistically significant.

Statistical analyses (independent replication—ALSPAC)

Candidate CpGs were tested for their associations with asthma in the independent cohort, ALSPAC (N=720). To assess consistency of associations, results from the ALSPAC cohort were compared to results from the full IOW F1 sample (N=370) using two logistic regression models for each CpG site: a crude model between M-values and asthma, and a second model adjusting for sex and estimated cell-type proportions of CD8⁺ T-cells, CD4⁺ T-cells, natural killer cells, B-cells, monocytes, eosinophils and other granulocytes, which were estimated from the methylation array data [26, 27]. ALSPAC also included batch variables (Additional file 1: Method S2), to adjust for technical variations across the DNAM arrays. Statistical significance was determined at α of 0.05.

Statistical analyses (functional validation—IOW F2)

Among the successfully replicated loci, wheeze without cold and any wheeze were modeled with logistic regression. This included all newborns for which at least one infant follow-up visit had been completed (n=111 for DNAM models and n=82 for expression models). Cord blood proportions of CD8⁺ T-cells, CD4⁺ T-cells, natural killer cells, B-cells, monocytes, granulocytes, and nucleated red blood cells (nRBCs) were estimated via the estimateCellCounts function [26] using a cord blood reference panel [32]. We adjusted for season of birth, infant sex, and cell-type proportions. Statistical significance was determined at α of 0.05.

Results

Sample characteristics and study flow chart

A flowchart of all analyses is provided in Fig. 1. The subjects in the IOW F1 discovery sample were all 18 years old, predominantly female (66.2%) and 13.9% (n=51) of participants were asthmatic. Asthmatics were more likely to be atopic (66.0% vs 29.5%), have lower FEV₁/FVC Ratio (means: 0.83 vs. 0.88), greater FeNO (medians: 21.0 vs 14.0) and have higher proportions of B-cells (0.046 vs 0.039) and eosinophils (0.045 vs 0.021) (Table 1). The average age of subjects in the ALSPAC sample was 17 years old; 16.7% of the ALSPAC sample had asthma and 56.3% of participants were female. The IOW stage-1 and stage-2 samples, utilized for feature selection and internal validation respectively, had similar distributions of all covariates (Additional file 2: Table S1).

Discovery phase (stage-1 feature selection)

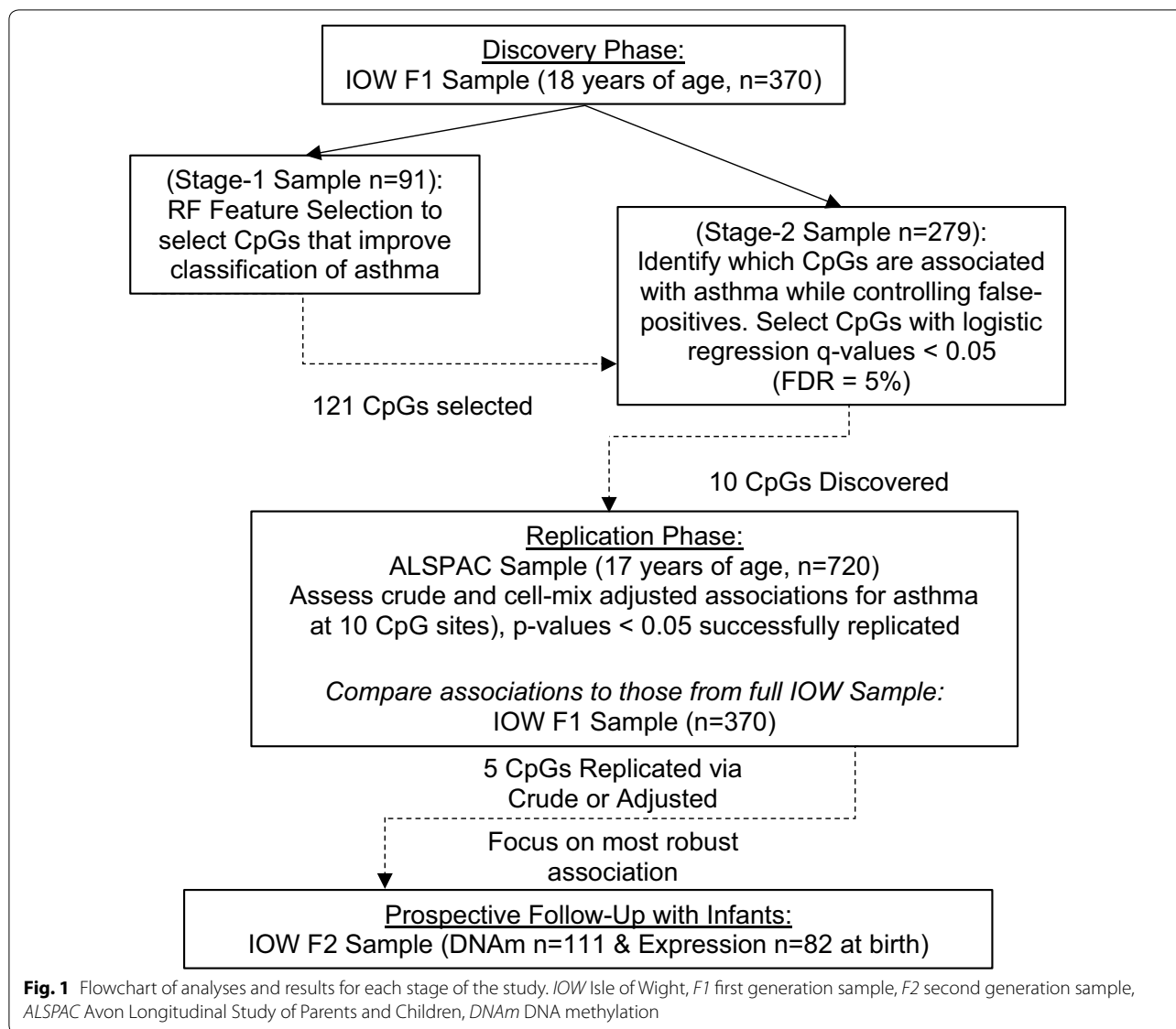
Recursive RF feature selection was implemented on the stage-1 sub-sample (n_{s1}=91), with a starting set of 248,336 CpG sites. The asthma-specific misclassification rates levelled off at the 12th iteration of the recursive RF algorithm, meaning that further reductions in the number of features would result in loss of information about asthma-associated loci. Thus, the 121 features (CpG sites) included in the 12th iteration (Additional file 3: Figure S1) were selected for stage-2 analysis.

Discovery phase (stage-2 logistic regression)

The stage-2 analysis was performed in an independent sub-sample (n_{s2}=279), to test the associations between DNA methylation and asthma at the 121 selected CpG sites with logistic regression. Of the 121 CpGs, 10 were associated with asthma at a 5% FDR (q-values < 0.05) (Additional file 2: Table S2). For all 10 sites, lower methylation was associated with greater odds of asthma. Adjustment for cellular heterogeneity substantially attenuated many of the parameter estimates and none of the adjusted models retained 5% FDR-significant q-values. However, the parameter estimates for the top five hits were mostly unperturbed and retained at least nominally significant p-values (< 0.05); in the case of cg16658191 and cg25578728, the magnitude of the associations became stronger after cell-mixture adjustment.

Replication analysis in ALSPAC

We then aimed to see whether the associations observed in IOW could be replicated in an independent cohort, ALSPAC. To compare associations between the IOW and ALSPAC cohorts, we produced odds ratios (ORs) and 95% confidence intervals (CIs) using the pooled IOW samples from stage-1 and stage-2 (IOW F1 n=370) (Table 2) and



ALSPAC (n=720) (Table 3) for the 10 FDR-significant CpG sites. All 10 CpGs exhibited the same direction of association, while 3 of these associations were statistically significant after cell-mix adjustment (p-values < 0.05) (Table 3; Additional file 2: Table S3): cg04359558 (*LITAF*), cg13753183 (*APTX*), and cg16658191 (*HK1*). These CpGs have been annotated with genomic information and function (Table 4). Differences in the distributions of cell-types are presented in Additional file 2: Table S4).

Adjusting for estimated cell mixtures attenuated most ORs, and led to some discordance between IOW and ALSPAC, with only cg04359558 and cg16658191 exhibiting significant associations with asthma in both cohorts after cell-type adjustments. Only our top-hit (cg16658191) was significantly associated with asthma in all models across both cohorts.

Some of the tested CpGs were observed to have moderate-to-strong Spearman correlations (cg06866208, cg07948085, cg09241885, cg11310939, cg13753183, cg16658191) with the proportions of estimated eosinophils (range of rho values: -0.51 to -0.59, p-values < 0.0001) and were also moderately correlated with each other (range of rho values: 0.26 to 0.49, p-values < 0.0001) (Additional file 3: Figure S2), suggesting that methylation levels at these CpGs may be partial markers of eosinophils.

Given the inconsistent confounding effects of cell-type, we considered cg16658191 within the *hexokinase-1* (*HK1*) gene as the finding with the most consistent evidence for an association with asthma and carried this CpG forward for cross-sectional analyses with allergy, inflammation and lung-function, as well as prospective analyses with infant respiratory outcomes.

Table 1 Baseline characteristics among those with and without asthma in the IOW F1 sample (18 years of age)

Categorical factors	Asthmatics (n = 51)		Non-asthmatics (n = 319)	
	n	Percent	n	Percent
Female sex	35	68.6	210	65.8
Atopy	33	66.0	93	29.5
Continuous factors	n	Mean (SD)	n	Mean (SD)
FVC	49	4.50 (0.874)	307	4.48 (0.857)
FEV ₁	49	3.74 (0.750)	307	3.94 (0.719)
FEV ₁ /FVC Ratio	49	0.83 (0.088)	307	0.88 (0.064)
FeNO ^a	47	21.00 (60.00)	300	14.0 (10.50)
Estimated cell proportions	n	Mean (SD)	n	Mean (SD)
CD8 ⁺ T-Cells	51	0.101 (0.082)	319	0.098 (0.054)
CD4 ⁺ T-Cells	51	0.121 (0.052)	319	0.110 (0.047)
Natural Killer Cells	51	0.055 (0.061)	319	0.070 (0.062)
B-Cells	51	0.046 (0.023)	319	0.039 (0.024)
Monocytes	51	0.072 (0.024)	319	0.074 (0.024)
Eosinophils	51	0.045 (0.045)	319	0.021 (0.025)
Neutrophils	51	0.558 (0.118)	319	0.589 (0.114)

FEV₁ forced expiratory volume in 1 s, FVC forced vital capacity, FeNO fractional-exhaled nitric oxide

^a Due to heavy positive skew of FeNO, medians (interquartile ranges) are presented here

Table 2 Crude and adjusted associations between M-values and asthma in the IOW F1 (18 years of age, n = 370) sample via logistic regression

CpG annotations		DNAm distribution		Crude model		Adjusted model ^a	
CpG ID	Gene	Median	(5th, 95th pct.)	OR	CI	OR	CI
cg00100703	<i>UNC45B</i>	0.93	(0.89, 0.95)	0.12	(0.05, 0.28)	0.18	(0.06, 0.50)
cg01069468	<i>SYNGAP1</i>	0.88	(0.84, 0.91)	0.12	(0.04, 0.31)	0.39	(0.08, 1.72)
cg04359558	<i>LITAF</i>	0.93	(0.87, 0.96)	0.23	(0.12, 0.42)	0.34	(0.17, 0.68)
cg06866208	–	0.81	(0.73, 0.84)	0.14	(0.05, 0.34)	0.40	(0.09, 1.72)
cg07948085	–	0.89	(0.84, 0.93)	0.16	(0.07, 0.34)	0.26	(0.09, 0.74)
cg09241885	<i>C20orf118</i>	0.89	(0.80, 0.93)	0.29	(0.16, 0.49)	0.41	(0.17, 0.99)
cg11310939	<i>MARCH3</i>	0.91	(0.86, 0.92)	0.19	(0.09, 0.40)	0.45	(0.15, 1.44)
cg13753183	<i>APTX</i>	0.90	(0.86, 0.93)	0.11	(0.04, 0.30)	0.30	(0.06, 1.38)
cg16658191	<i>HK1</i>	0.93	(0.89, 0.95)	0.12	(0.05, 0.27)	0.17	(0.04, 0.57)
cg25578728	<i>CHD7</i>	0.90	(0.85, 0.94)	0.24	(0.11, 0.46)	0.27	(0.11, 0.60)

OR odds ratio, CI 95% confidence intervals, IOW Isle of Wight, pct. percentile

^a Adjusted for cell mixture and sex

HK1 DNA methylation is associated with allergy, inflammation and lung function

We found that DNAm at cg16658191 was lower among those with atopy (T-test: *HK1* p-value < 0.001) and had an inverse non-linear association with logFeNO (rho = -0.22, p-value < 0.0001), suggesting that it is involved in allergic sensitization and airway inflammation. Additionally, those with lower DNAm at this locus tended to have lower FEV₁/FVC (rho = 0.10, p-value = 0.057)

and FEF_{25-75%} (rho = 0.095, p-value = 0.075) though these correlations were not statistically significant (Fig. 2).

Prospective follow-up for HK1 associations in infants

We then performed follow-up analyses for associations of our top locus, cg16658191, with wheeze during infancy and variations in gene-expression in the IOW F2 sample. Infants with lower levels of cord blood DNAm at cg16658191 had greater odds of wheeze without cold within the 1st year of

Table 3 Replication of crude and adjusted associations between M-values and asthma in the ALSPAC (17 years of age, n = 720) sample via logistic regression

CpG annotations		Crude model		Adjusted model ^a	
CpG ID	Gene	OR	CI	OR	CI
cg00100703	<i>UNC45B</i>	0.71	(0.52, 0.96)	0.77	(0.52, 1.15)
cg01046943	<i>NUP210</i>	0.93	(0.71, 1.20)	0.83	(0.62, 1.11)
cg04359558	<i>LITAF</i>	0.75	(0.56, 1.00)	0.61	(0.40, 0.94)
cg06866208	–	1.01	(0.70, 1.44)	0.86	(0.38, 1.97)
cg07948085	–	0.63	(0.43, 0.91)	0.70	(0.45, 1.06)
cg09241885	<i>C20orf118</i>	0.85	(0.64, 1.13)	0.75	(0.51, 1.10)
cg11310939	<i>MARCH3</i>	0.78	(0.60, 1.02)	0.90	(0.65, 1.24)
cg13753183	<i>APTX</i>	0.76	(0.58, 1.00)	0.67	(0.46, 0.98)
cg16658191	<i>HK1</i>	0.51	(0.37, 0.72)	0.57	(0.38, 0.87)
cg25578728	<i>CHD7</i>	0.73	(0.53, 1.02)	0.78	(0.50, 1.24)

OR odds ratio, CI 95% confidence intervals, ALSPAC Avon Longitudinal Study of Parents and Children, pct. percentile

^a Adjusted for cell mixture, sex, and batch variables

life (Table 5), though adjustments for estimated cell-types confounded this association, particularly due to a strong correlation with nRBCs ($\rho = -0.84$, p -value < 0.0001) and moderate correlation with granulocytes ($\rho = 0.59$, p -value < 0.0001). Additionally, DNAm at cg16658191 was inversely associated with the expression of *HK1* ($\rho = -0.22$, p -value = 0.039) and increased expression of *HK1* was associated with increased odds of wheeze without cold and odds of any wheeze, during the 1st year of life. Interestingly, these associations became stronger after adjusting for cellular heterogeneity (Table 5).

Discovery of epigenomic loci associated with asthma (traditional EWAS approach)

Finally, we examined DNAm-asthma associations using a standard EWAS approach, regressing methylation levels for all CpGs on asthma status in unadjusted models and models adjusted for sex, CD4⁺ T-cells, CD8⁺ T-cells,

monocytes, eosinophils, natural killer, and granulocytes. In the unadjusted models, 148 CpGs were significantly associated (FDR 5%) with asthma status. However, adjusting for sex and cell mixture resulted in attenuation of most of these results and none of the adjusted models produced FDR-significant associations. We compared the results from our models, unadjusted (Additional file 2: Table S5) and adjusted (Additional file 2: Table S6) that yielded p -values < 0.001 , to the results from a prior EWAS in ALSPAC for current asthma and current wheeze at ages 7.5 and 16.5 years that yielded p -values < 0.001 [10]. Of the 674 CpGs that were associated with asthma (p -value < 0.001) in IOW prior to cell-type adjustment, 20 CpGs yielded p -values < 0.001 for all four models in ALSPAC (Additional file 2: Table S7). However, only one CpG yielded a p -value < 0.001 in the IOW and a p -value < 0.001 in ALSPAC when adjusting for cell mixture, and that was cg16658191. We also compared our results to the asthma-associated CpGs identified in a meta-analysis of children between the ages 4 and 8 years old [18] at the 11 (out of 14) sites that passed QC in our study. Although all 11 sites yielded nominally significant inverse associations with asthma in IOW (Additional file 2: Table S8) only cg10142874 retained even a nominal association with asthma in IOW after cell-mix adjustment (p -value = 0.013).

Discussion

We performed an epigenome-wide association study of current asthma in the IOW cohort utilizing two statistical approaches and a replication analysis in an independent population. We identified that lower DNAm at cg16658191 within the 1st exon of *HK1* as a marker of current asthma. This CpG was identified via random forest feature selection and confirmed using standard EWAS, and was replicated within an independent cohort (ALSPAC). We then produced similar associations between DNAm of cg16658191 and the expression of *HK1* in cord blood with infant wheeze in the children of the IOW cohort. We also observed functional evidence

Table 4 Annotations and biological functions of genes associated with CpG sites associated with asthma in the replication study via either the adjusted or unadjusted models

CpG ID	Region	Gene ID	Gene name	Function
cg04359558	Body	<i>LITAF</i>	<i>Lipopolysaccharide-Induced TNF-α Factor</i>	DNA binding-protein that promotes inflammatory cytokine expression; involved in apoptotic signaling and inhibition of proliferation [46, 47]
cg13753183	Body; 1st Exon; 5'UTR	<i>APTX</i>	<i>Aprataxin</i>	Involved in DNA repair; mutations have been associated with ataxia-ocular apraxia [49]
cg16658191	Body; 1st Exon	<i>HK1</i>	<i>Hexokinase-1</i>	Involved in glucose metabolism [34] and inhibition of apoptotic signaling [40]

Ch Chromosome, UTR untranslated region

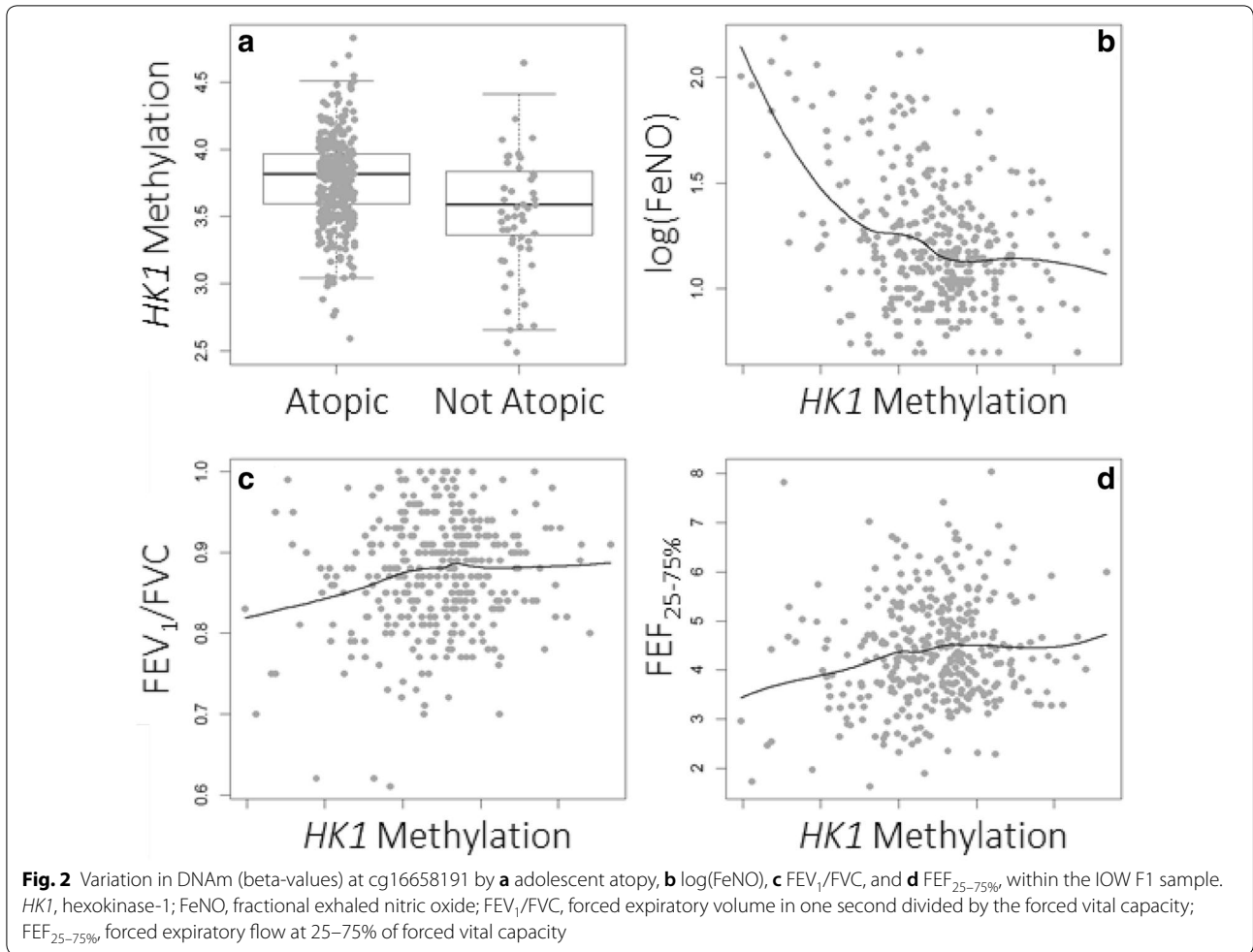


Table 5 Associations between cord blood DNAm at cg16658191 and expression of *HK1* with infant wheeze without cold and any wheeze within the IOW F2 sample

Exposure	Outcomes					
	Wheeze without cough/cold			Any wheeze		
	OR	p-value	95% CI	OR	p-value	95% CI
cg16658191 (n = 111)						
Crude model	0.23	0.0097	(0.07–0.68)	0.58	0.23	(0.23–1.41)
Adjusted model 1 ^a	0.21	0.0062	(0.06–0.61)	0.57	0.23	(0.22–1.40)
Adjusted model 2 ^b	0.37	0.30	(0.05–2.35)	2.42	0.27	(0.52–12.18)
<i>HK1</i> (A_23_P217958) (n = 82)						
Crude model	3.14	0.0183	(1.26–8.62)	3.44	0.0045	(1.54–8.59)
Adjusted model 1 ^a	3.62	0.015	(1.34–10.95)	3.22	0.0075	(1.42–8.08)
Adjusted model 2 ^b	7.51	0.0085	(1.89–40.50)	4.89	0.0087	(1.63–18.15)

OR, Odds Ratio; CI, 95% confidence interval; CB, cord blood; *HK1*, hexokinase-1; w/o, without

^a Adjusted for infant sex and season of birth

^b Adjusted for infant sex, season of birth, and cell mixture

of *HK1*'s involvement in infant wheeze using gene expression data that exhibited the expected associations with infant wheeze, given that promoter and first exon methylation are most commonly associated with repression of gene expression [6, 33]. DNAm at cg16658191, which is within the body and/or first exon, was inversely associated with *HK1* expression and we showed that higher expression of *HK1* was predictive of wheezing without a cold during infancy. The *HK1* gene resides in 10q22.1 and encodes a protein that is integral in the first step of glycolysis [34] and in apoptotic resistance [35]. The consistency of these associations across different ages, with different respiratory outcomes, and utilizing both DNAm and gene expression as predictors, suggests that this gene may play an important role in the predisposition for wheezing and/or asthma.

We found that many of our RF-identified hits, including cg16658191, were inversely correlated with eosinophil counts in adult blood, similar to what was observed by Arathimos et al. [10]. However, confounding by cell-mixture may not be limited to eosinophil proportions. For instance, because of its crucial role in glucose metabolism, *HK1* is highly expressed by erythrocytes [36]. This is consistent with the strong inverse correlation we observed between cg16658191 and estimated nRBC proportions in cord blood, which could indicate prematurity, restricted growth, or pregnancy complications [37]. Additionally, premature and low birth weight neonates are predisposed to early-life respiratory morbidity [38, 39]. However, adjustments for weeks of gestation did not appreciably alter our results (data not shown).

Though DNAm at cg16658191 may, in part, be a marker of high eosinophil counts in adult blood and nRBCs in cord blood, we found that, in addition to the relationship between DNAm and asthma, *HK1* expression was strongly and significantly associated with infant wheeze even after cell-type adjustments. These findings suggest a role for DNAm of the *HK1* gene in asthma and wheeze etiology that is independent of cell-type proportions, possibly through differential epigenetic regulation within a subset of asthma-associated cell-types. However, it is difficult to disentangle such relationships in studies that utilize tissues composed of mixed cell populations, such as blood. *HK1* is involved in apoptotic resistance via binding to and stabilizing the mitochondrial membrane, whereas the dissociation of *HK1* from the membrane makes those cells more susceptible to apoptosis [40]. Up-regulation of *HK1* resulting in increased apoptotic resistance has been observed in cancerous cells [41] and

HIV-1 infected macrophages [42]. Apoptotic-resistant pro-inflammatory cells are known to lead to prolonged inflammation [43] and apoptosis appears to be delayed in neutrophils [44] and T-lymphocytes [45] of asthmatics. This provides a possible mechanism through which *HK1* epigenetic regulation and expression by immune cells may be involved in asthma and wheeze etiology.

We also examined relationships between DNAm and asthma using a more traditional EWAS approach, regressing the methylation beta-values for each loci on asthma status, and compared our findings to two recent EWAS, one performed by ALSPAC [10] and a meta-analysis of childhood asthma from multiple European cohorts [18]. This approach identified 148 CpGs that were significantly associated (5% FDR) with asthma prior to cell-type adjustment, and no FDR-significant findings after adjustment. We also found very little consistency in observed associations between our study and the ALSPAC study after adjusting for cell mixture. Only cg16658191 demonstrated an association with current asthma and wheeze in the fully adjusted models ($p < 0.001$) in both the IOW and ALSPAC. When comparing our traditional EWAS results to a meta-analysis of childhood asthma, only cg10142874 from that meta-analysis yielded an even nominally-significant association with asthma in IOW after cell-mixture adjustment.

Strengths of our study included the use of multiple samples to discover and replicate our findings, supported by gene expression studies, and the use of a validated tool, the ISAAC core questionnaire, to define current asthma status. However, it is also important to recognize this study's limitations. One limitation is that detection of CpGs in the IOW birth cohort and the replication study in ALSPAC investigated concurrent associations. Hence, reverse causation in which asthma may result in differential methylation of *HK1* cannot be excluded. Differences between the two cohorts, confounding by cell-mixture, errors in cell-mixture estimates, and asthma heterogeneity may have limited replicability of more loci after additional adjustments. The discovery and replication samples were similar in sex-distribution, prevalence of asthma, and age, but differed in estimated cell-type distributions (Additional file 2: Table S4). Additionally, though the estimated cell-proportions are imperfect, we utilized the gold standard for predicting cell mixtures from DNAm arrays [27] and comprehensively evaluated the impact of cellular heterogeneity on our findings. There is also the possibility of residual confounding, perhaps by genotype. Numerous SNPs have

been implicated as asthma susceptibility loci, and some SNPs have been shown to influence the methylation status of CpG sites. We cannot rule out the possibility that our findings are markers of upstream genetic effects on both DNA methylation and asthma susceptibility. It is also important to point out that the relationships we observed between DNAm and expression of *HK1* in cord blood with infant wheeze, cannot be directly extrapolated to asthma. It is unclear whether DNAm patterns of *HK1* in cord blood are informative for the later development of asthma, although our findings provide evidence that lower DNAm and increased expression of *HK1* in cord blood are associated with wheezing in the 1st year of life. Finally, asthma is a heterogeneous condition, in which different phenotypes may arise via different underlying physiological mechanisms [3]. We showed that *HK1* DNAm levels were also strongly associated with atopy and FeNO, which may indicate that the regulation of this gene is particularly important in allergic-asthma. This also raises the possibility that some of our other discovered, but not replicated, loci may be associated with specific asthma-phenotypes. If the prevalences of these phenotypes differ between IOW and ALSPAC, this may have contributed to the discordant results. Interestingly, some of the CpGs with discordant results between the two cohorts were within genes or genomic regions that have previously been associated with asthma or are involved in apoptotic signaling, like *HK1*. For instance, cg04359558 is within the body of *LITAF*, a gene that encodes a DNA-binding protein that promotes the expression of TNF- α and other cytokines known to be involved in pro-inflammatory and apoptotic signaling [46, 47]. *UNC45B*, annotated to cg00100703, lies within the asthma susceptibility region 17q12–21 [48], though this particular gene has not previously been linked to asthma.

Conclusions

In summary, we discovered a novel epigenetic association with adolescent asthma at cg16658191 within *HK1*, whose DNAm and expression levels in cord blood were also associated with infant wheeze without cold. In addition, the association of cg16658191 with asthma was replicated in an independent cohort. However, we also found that our findings may be affected, at least in part, by heterogeneous cell-mixtures. Further research is required to determine whether these observed associations are reproducible in other populations, particularly with different racial and ethnic characteristics, and whether some of these loci are differentially regulated between those with and without asthma in specific cell-type populations such as eosinophils.

Additional files

Additional file 1: Method S1. Cohort-Specific DNA-M preprocessing steps. **Methods S2.** SVA to account for technical variations in ALSPAC.

Additional file 2: This file includes supplemental Tables (S1–S8)—

Table S1. Comparison of cell-proportions and lung function variables across the Stage-1 (n = 91) and Stage-2 (n = 279) samples from the IOW F1 Sample. **Table S2.** Parameter estimates from logistic regression models performed in the Stage-2 sample (ns2 = 279), regressing current asthma status on DNA methylation M-values for all CpGs selected from the Stage-1 analysis. **Table S3.** Parameter estimates, standard errors, and p-values to compare the IOW F1 results to the ALSPAC replication results for CpGs that yielded an association within a 5% FDR in the Stage-2 analysis. **Table S4.** Comparing adjustment covariates between the IOW F1 sample and the ALSPAC sample; the within cohort comparisons are testing for differences in these variables between those with and without asthma. **Table S5.** Results from linear asthma EWAS in IOW F1 (n = 370), in which methylation beta values were regressed on asthma status, unadjusted for possible confounders. **Table S6.** Results from linear asthma EWAS in IOW F1 (n = 370), in which methylation beta values were regressed on asthma status, adjusted for sex, CD4T cells, CD8T cells, Monocytes, Natural Killers, Eosinophils, and other Granulocytes. **Table S7.** CpGs that yielded unadjusted association (p < 0.001) with current asthma in IOW F1 (unadjusted models) that were also identified as having unadjusted associations with current asthma and wheeze at age 7.5 and 16.5 years old in the previously published ALSPAC EWAS. **Table S8.** Associations between asthma and DNA methylation within the IOW cohort (unadjusted models) at the 14 CpGs that were identified as being asthma associated in a meta analysis of European children.

Additional file 3: This file includes supplemental figures (S1, S2)—**Figure S1.** Tracking of the misclassification rates (y-axis) across iterations (x-axis) of the recursive RF feature selection. **Figure S2.** Correlations (Spearman) between estimated cell-proportions in blood and DNAm M-values for the 10 CpGs that were identified as candidates for the replication study. Statistically significant correlations are designated at p-values < 0.05 (*), p-values < 0.001 (**), and p-values < 0.001 (***).

Abbreviations

ALSPAC: Avon Longitudinal Study of Parents and Children; DNAm: DNA methylation; HK1: hexokinase-1; IOW: Isle of Wight; OOB: out of bag; RF: random forest; VIM: variable importance measure.

Acknowledgements

This publication is the work of the authors, who will serve as guarantors for the contents of this paper. The authors gratefully acknowledge the cooperation of the children and parents who participated in the IOW birth cohort and appreciate the hard work of Mrs. Sharon Matthews and the Isle of Wight research team in collecting data and Nikki Graham for technical support. We are extremely grateful to all the families who took part in the ALSPAC 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.

Authors' contributions

WK, JWH & SHA coordinated the study, SLE generated gene expression data for IOW, TME & HZ designed the analysis plan and TME carried out the statistical analysis for IOW. AK, FIR, & GAL performed data processing, QA, and QC. MF, SLE & VKP, assisted in interpreting results for the IOW. KB performed statistical analyses for ALSPAC. AJH, CLR, GCS & KB, assisted in interpreting ALSPAC results. TME, HZ, WK, JWH and KB, were major contributors in writing the manuscript. All authors revised the final manuscript. All authors read and approved the final manuscript.

Funding

The IOW portion of this work was supported by the National Institute of Health under award numbers R01 AI091905 and R01HL132321 (PI: Wilfried Karmaus) and R01AI121226 (MPI: Hongmei Zhang, John Holloway). The 10-year follow-up of this study was funded by the National Asthma Campaign,

UK (Grant No 364) and the 18-year follow-up by NIH/NHLBI R01 HL082925-01 (PI: S. Hasan Arshad). We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the IOW methylation data. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1). This work was supported by the Medical Research Council Integrative Epidemiology Unit and the University of Bristol (MC_UU_12013_2).

Availability of data and materials

The minimal data sets analyzed in the current study are available from the corresponding author upon reasonable request. For access to the full Isle of Wight Cohort data please see: <http://www.allergyresearch.org.uk/studies/birth-cohort/-cohort-data-use>; for access to ALSPAC data please see: <http://www.bristol.ac.uk/alspac/researchers/access/>.

Ethics approval and consent to participate

Ethical approval was obtained from National Research Ethics Service, NRES Committee South Central—Southampton B for the 18-year follow-up (06/Q1701/34) and NRES Committee South Central—Hampshire B (09/H0504/129) for the follow-up of IOW participants' offspring; written informed consent was provided by the infants' parents. Ethical approval for the ALSPAC study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees; written informed consent was provided by all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they are no competing interests.

Author details

¹ Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 915 Greene Street, Columbia, SC 29208, USA. ² Present Address: Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Rd, Atlanta, GA 30322, USA. ³ Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health, University of Memphis, 236A Robison Hall, Memphis, TN 38152, USA. ⁴ Human Development and Health, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton SO16 6YD, UK. ⁵ Center for Precision and Environmental Health, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA. ⁶ Present Address: Department of Public Health Sciences at the College of Health and Human Services, University of North Carolina Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA. ⁷ Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, MI, USA. ⁸ MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK. ⁹ Avon Longitudinal Study of Parents and Children, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK. ¹⁰ Clinical and Experimental Sciences Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton SO16 6YD, UK. ¹¹ The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Newport, Isle of Wight, UK. ¹² NIHR Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton, UK.

Received: 16 May 2018 Accepted: 12 July 2019

Published online: 24 July 2019

References

- Martinez FD, Vercelli D. Asthma. *Lancet*. 2013;382:1360–72.
- Akinbami LJ, Moorman JE, Liu X. Asthma Prevalence, Health Care Use, and Mortality: United States, 2005–2009. *National Health Statistics Reports*. Jhyattsville, MD; 2011.
- Lötvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. *J Allergy Clin Immunol*. 2011;127(2):355–60.
- DeVries A, Vercelli D. Early predictors of asthma and allergy in children: the role of epigenetics. *Curr Opin Allergy Clin Immunol*. 2015;15(5):435–9.
- Durham AL, Wiegman C, Adcock IM. Epigenetics of asthma. *Biochim Biophys Acta*. 2011;1810(11):1103–9.
- Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13(7):484–92.
- Ho S. Environmental epigenetics of asthma: an update. *J Allergy Clin Immunol*. 2010;126(3):453–65.
- Yang IV, Tomfohr J, Singh J, Foss CM, Marshall HE, Que LG, et al. The clinical and environmental determinants of airway transcriptional profiles in allergic asthma. *Am J Respir Crit Care Med*. 2012;185(6):620–7.
- 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(1):69–80. <https://doi.org/10.1016/j.jaci.2015.01.025>.
- Arathimos R, Suderman M, Sharp GC, Burrows K, Graneli R, Tilling K, et al. Epigenome-wide association study of asthma and wheeze in childhood and adolescence. *Clin Epigenetics*. 2017;9:112.
- Barton SJ, Ngo S, Costello P, Garratt E, El-Heis S, Antoun E, et al. DNA methylation of Th2 lineage determination genes at birth is associated with allergic outcomes in childhood. *Clin Exp Allergy*. 2017;47(12):1599–608.
- 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(2):534–42.
- Guarnieri M, Balmes JR. Outdoor air pollution and asthma. *Lancet*. 2014;383(9928):1581–92.
- Somineni HK, Zhang K, Biagini Myers JM, Kovacic MB, Ulm A, Jurcak N, et al. TET1 methylation is associated with childhood asthma traffic-related air pollution. *J Allergy Clin Immunol*. 2016;137(3):797–805.
- Prunicki M, Stell L, Dinakarandian D, de Planell-Saguer M, Lucas RW, Hammond SK, et al. Exposure to NO₂, CO, and PM_{2.5} is linked to regional DNA methylation differences in asthma. *Clin Epigenetics*. 2018;10(1):2.
- Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Larivière M, Moussette S, et al. Allele-specific chromatin remodeling in the ZBP2/GSDMB/ormdl3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet*. 2009;85(3):377–93.
- Kothari PH, Qiu W, Croteau-Chonka DC, Martinez FD, Liu AH, Lemanske RF, et al. The role of local CpG DNA methylation in mediating the 17q21 asthma-susceptibility GSDMB/ORMDL3 expression quantitative trait locus. *J Allergy Clin Immunol*. 2018;141(6):2282–2286.e6.
- Xu CJ, Söderhäll C, Bustamante M, Baiz N, Gruzieva O, Gehring U, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med*. 2018;6:379–88.
- Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, et al. Cohort profile: the Isle of Wight whole population birth cohort (IOWBC). *Int J Epidemiol*. 2018;47(4):1043–1044i.
- Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International study of asthma and allergies in childhood (ISAAC): rationale and methods. *Eur Respir J*. 1995;8(3):483–91.
- Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319–38.
- Scott M, Raza A, Karmaus W, Mitchell F, Grundy J, Kurukulaaratchy RJ, et al. Influence of atopy and asthma on exhaled nitric oxide in an unselected birth cohort study. *Thorax*. 2010;65:258–63.
- Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Smith GD, et al. Cohort profile: the avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42:97–110.
- 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. *Eur J Epidemiol*. 2013;42:111–27.
- 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;44:1181–90.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén SE, 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.
27. 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.
 28. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
 29. Goldstein BA, Hubbard AE, Cutler A, Barcellos LF. An application of Random Forests to a genome-wide association dataset: Methodological considerations & new findings. *BMC Genet*. 2010;11:49.
 30. Everson TM, Lyons G, Zhang H, Soto-Ramírez N, Lockett GA, Patil V, et al. DNA methylation loci associated with atopy and high serum IgE: a genome-wide application of recursive Random Forest feature selection. *Genome Med*. 2015;7:89.
 31. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Stat Soc Ser B Stat Methodol*. 2004;66(1):187–205.
 32. Bakulski KM, Feinberg JL, Andrews SV, Yang J, Mckenney S, Witter F, et al. DNA methylation of cord blood cell types: applications for mixed cell birth studies. *Epigenetics*. 2016;11(5):354–62.
 33. Brenet F, Moh M, Funk P, Feisterstein E, Viale AJ, Socci ND, et al. DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE*. 2011;6(1):e14524.
 34. John S, Weiss JN, Ribalet B. Subcellular localization of hexokinases i and ii directs the metabolic fate of glucose. *PLoS ONE*. 2011;6(3):e17674.
 35. Abu-Hamad S, Zaid H, Israelson A, Nahon E, Shoshan-Barmatz V. Hexokinase-I protection against apoptotic cell death is mediated via interaction with the voltage-dependent anion channel-1. *J Biol Chem*. 2008;283(19):13482–90.
 36. Van Wijk R, Van Solinge WW. The energy-less red blood cell is lost: erythrocyte enzyme abnormalities of glycolysis. *Blood*. 2005;106(13):4034–42.
 37. Hermansen MC. Nucleated red blood cells in the fetus and newborn. *Arch Dis Child Fetal Neonatal Ed*. 2001;84(3):F211–5.
 38. Rusconi F, Galassi C, Forastiere F, Bellasio M, De Sario M, Ciccone G, et al. Maternal complications and procedures in pregnancy and at birth and wheezing phenotypes in children. *Am J Respir Crit Care*. 2007;175:16–21.
 39. Edwards MO, Kotecha SJ, Lowe J, Richards L, Watkins WJ, Kotecha S. Management of prematurity-associated wheeze and its association with atopy. *PLoS ONE*. 2016;11(5):e0155695.
 40. Schindler A, Foley E. Hexokinase 1 blocks apoptotic signals at the mitochondria. *Cell Signal*. 2013;25(12):2685–92.
 41. Kroemer G, Pouyssegur J. Review tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell*. 2008;13:472–82.
 42. Sen S, Kaminiski R, Deshmane S, Langford D, Kahlili K, Amini S, et al. Role of hexokinase-1 in the survival of HIV-1- infected macrophages. *Cell Cycle*. 2015;14(7):980–9.
 43. Luo HR, Loison F. Constitutive neutrophil apoptosis: mechanisms and regulation. *Am J Hematol*. 2008;83(4):288–95.
 44. Yang EJ, Choi E, Ko J, Kim D, Lee J-S, Kim IS. Differential effect of CCL2 on constitutive neutrophil apoptosis between normal and asthmatic subjects. *J Cell Physiol*. 2011;227:2567–77.
 45. Potapinska O, Demkow U. T lymphocyte apoptosis in asthma. *Eur J Med Res*. 2009;14:192–5.
 46. Min J, Zhang W, Gu Y, Hong L, Yao L, Li F, et al. CIDE-3 interacts with lipopolysaccharide-induced tumor necrosis factor, and overexpression increases apoptosis in hepatocellular carcinoma. *Med Oncol*. 2011;28:S219–2227.
 47. Tang X, Molina M, Amar S. p53 short peptide (p53pep164) regulates lipopolysaccharide-induced tumor necrosis factor-alpha factor/cytokine expression. *Cancer Res*. 2007;67(3):1308–16.
 48. Naumova AK, Al Tuwaijri A, Morin A, Vaillancout VT, Madore A-M, Berlivot S, et al. Sex- and age-dependent DNA methylation at the 17q12-q21 locus associated with childhood asthma. *Hum Genet*. 2013;132:811–22.
 49. Sano Y, Date H, Igarashi S, Onodera O, Oyake M, Takahashi T, et al. Aprataxin, the causative protein for EAOH is a nuclear protein with a potential role as a DNA repair protein. *Am Neurol Assoc*. 2004;55:241–9.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

