Check for updates

Epigenetic signatures of asthma: a comprehensive study of DNA methylation and clinical markers

Austin J. Van Asselt[1,2*], Jeffrey J. Beck[1], Casey T. Finnicum[1], Brandon N. Johnson[1], Noah Kallsen[1], Sarah Viet[1], Patricia Huizenga[1], Lannie Ligthart[2,3], Jouke-Jan Hottenga[2,3], René Pool[2,3], Anke H. Maitland-van der Zee[3,4], S. J. Vijverberg[3,4], Eco de Geus[2,3], Dorret I. Boomsma[2,3,5,6], Erik A. Ehli[1] and Jenny van Dongen[2,3,5]

Abstract

Background Asthma, a complex respiratory disease, presents with inflammatory symptoms in the lungs, blood, and other tissues. We investigated the relationship between DNA methylation and 35 clinical markers of asthma.

Methods The Illumina Infinium EPIC v1 methylation array was used to evaluate 742,442 CpGs in whole blood from 319 participants from 94 families. They were part of the Netherlands Twin Register from families with at least one member suffering from severe asthma. Repeat blood samples were taken after 10 years from 182 individuals. Principal component analysis on the clinical asthma markers yielded ten principal components (PCs) that explained 92.8% of the total variance. We performed epigenome-wide association studies (EWAS) for each of the ten PCs correcting for familial structure and other covariates.

Results 221 unique CpGs reached genome-wide significance at timepoint 1 after Bonferroni correction. PC7, which correlated with loadings of eosinophil counts and immunoglobulin levels, accounted for the majority of associations (204). Enrichment analysis via the EWAS Atlas identified 190 of these CpGs to be previously identified in EWASs of asthma and asthma-related traits. Proximity assessment to previously identified SNPs associated with asthma identified 17 unique SNPs within 1 MB of two of the 221 CpGs. EWAS in 182 individuals with epigenetic data at a second timepoint identified 49 significant CpGs. EWAS Atlas enrichment analysis indicated that 4 of the 49 were previously associated with asthma or asthma-related traits. Comparing the estimates of all the significant associations identified across the two time points yielded a correlation of 0.81.

Conclusion We identified 270 unique CpGs that were associated with PC scores generated from 35 clinical markers of asthma, either cross-sectionally or 10 years later. A strong correlation was present between effect sizes at the 2 timepoints. Most associations were identified for PC7, which captured blood eosinophil counts and immunoglobulin levels and many of these CpGs have previous associations in earlier studies of asthma and asthma-related traits. The results point to a robust DNA methylation profile as a new, stable biomarker for asthma.

Keywords Methylation, Asthma, Epigenetics, Epigenome-wide association, Microarrays

*Correspondence:
Austin J. Van Asselt
Austin.vanasselt@avera.org
Full list of author information is available at the end of the article

BMC

Introduction

Asthma affects approximately 262 million individuals worldwide and poses a significant health burden resulting in over 450,000 deaths annually [1, 2]. Often, the challenges associated with treating and diagnosing asthma stem from the complex nature of the disease [3] due to its multi-level heterogeneity with various clinical presentations, treatment responses, and disease trajectories [4]. Clinically, diagnosing asthma is done through the collection of several different clinical measurements and symptoms including a medical history assessment, a symptom and physical assessment, lung function tests, bronchial challenge tests, response to bronchodilators, and allergy testing [5]. These different diagnostic criteria, along with different individual phenotype characteristics, such as age of onset, have led to the identification of subtypes of asthma [6–8] and different endotypes with varying molecular underpinnings of the disease [6].

Several large-scale genomic studies have elucidated associations in multiple immune and regulatory genes for these endotypes, which may form the basis for polygenic scores [9]. A GWAS of asthma by Demenais et al. in 2018 included over 140,000 individuals and identified five novel, independent associations and confirmed several others that had previously been identified [10]. A study performed by Ferreira et al. in 2019, which included over 300,000 individuals, identified 123 single nucleotide polymorphisms (SNPs) significantly associated with childhood asthma, 56 SNPs associated with adult-onset asthma, and 37 associated with both [11]. Collectively, Ferreira et. al. found 28 novel, independent associations between the two tested age groups [11]. Pividori et al. (2019) also investigated genetic associations to childhood-onset and adult-onset asthma [12] and identified 23 loci associated with childhood-onset asthma, one locus associated with adult-onset asthma, and 37 loci shared between the two groups [12].

Many of the associations that have been identified appear to be strongly linked to immune-related mechanisms, several of which are responsible for driving other inflammatory illnesses [10, 11, 13]. Zhu et al. identified strong genetic correlations between asthma and allergic diseases and significant overlap in specific loci significantly associated with both asthma and allergic illness [13]. Many of these loci reside in open chromatin areas of immune cells, further implying their importance in immune system function [14]. These genetic associations have also been shown to be tied to other illnesses such as certain mental health disorders like ADHD, major depressive disorder, and anxiety, indicating that their effects and influence may be broader than previously thought [15]. Though these genetic associations have been extensively studied and corroborated, the amount

of variance explained in relation to asthma appears limited (3.5–5.1%) [10, 11]. These outcomes reaffirm both the notion that asthma is polygenic and that larger studies are needed, and that it is influenced by environmental stimuli, which has led researchers to explore additional fields of study such as epigenetics and, more specifically, DNA methylation [16–18].

DNA methylation, as a molecular mechanism in humans, involves the addition of a methyl group to cytosine nucleotides, often in response to certain environmental queues [19–21]. The addition and removal of DNA methylation, specifically in regulatory regions of DNA, can lead to changes in gene expression and downstream cellular and tissue function [20, 21]. This entanglement with the external environment provides a basis for investigating the association of DNA methylation and asthma. The first studies of asthma and DNA methylation were conducted on candidate genes through techniques such as bisulfite pyrosequencing, which yielded moderate success highlighting CpG sites near *IFNγ* as potential mediators of asthma [22, 23]. The candidate gene studies were followed by epigenome-wide association studies (EWAS) of asthma in different age cohorts, asthma subtypes, and tissue types [9], with most studies focusing on childhood, allergic asthma in whole blood and nasal epithelial samples [16, 17]. Studies of childhood asthma in nasal epithelial samples have identified several CpGs in both immune regulatory pathways and pathways of basic cell function [24–26]. Furthermore, a small number have shown to replicate in EWASs of whole blood, indicating some pan-tissue effects, though this effect appears to be limited [9]. Studies in adults using whole blood samples found CpGs in genes relating to general inflammation, but the results show more variability in the CpGs that have been identified [27]. Though the majority of these results point toward immune-mediated pathways, the complete biological dynamic between these sites and their influence on asthmatic phenotypes has yet to be investigated [28]. Additionally, little is known regarding which measurable clinical markers associated with asthma may be contributing to these significantly associated CpGs, an area we explore here.

More generally, the continued identification of genetic and epigenetic variants influencing asthma is valuable for personalized treatment approaches. Here, we examined the genome-wide, CpG-specific methylation association with 35 different clinical markers of asthma from whole blood samples measured on the Illumina EPIC v1 methylation array and assessed the proximity of significantly associated CpGs to SNPs previously associated with asthma. We included individuals from families enrolled in the Netherlands Twin Register where at least one member of a nuclear family

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 3 of 15

(consisting of 4 or 5 individuals) was diagnosed with asthma. Blood samples were collected at two time-points, with 319 individuals having a sample collected at the first timepoint in the 1990's and 182 individuals a second timepoint approximately 10 years later in the 2000's [29]. Clinical markers were assessed at time-point 1, while epigenetic data was generated on DNA from blood specimens for both timepoint 1 and time-point 2.

Results

Clinical marker summary and data reduction

An initial set of 39 asthma-related clinical markers were measured. Markers that exhibited no variation between individuals were excluded (4 out of 39). Markers included measurements of lung function/capacity, skin prick test results measuring the reaction to common allergens, and others previously shown to be associated with asthma. Table 1 provides descriptions of the remaining 35 clinical markers. We performed an imputation step via the R package mice v3.16.0 (Multivariate Imputation by

Table 1 Descriptions/definitions for each clinical marker that was measured at T1

| Marker	Description
Alt.alte	Alternaria alternata, source: skin prick test
Asp.fum	Aspergillus fumigatus, source: skin prick test
Atopy	Allergy: positive skin test, and/or increased specific IgE, calculated
Bronch_hyperreactivity1	Bronchial hyperreactivity, source: questionnaire
Bronch_hyperreactivity2	Bronchial hyperreactivity (measured by methacholine threshold test), the cut-off point is at a pc20 of 20 mg/ml, source: lung function test
Cand.al	Candida albicans, source: skin prick test
Cat	Reaction to cat, source: skin prick test
Cats_IgE	Specific IgE for cats in E per ml, source: laboratory
Clad.h	Cladosporium herbarum, source: skin prick test
Coughing	Coughing complaints, source: questionnaire
Dog	Reaction to dog, source: skin prick test
Dustmites_IgE	Specific IgE for dust mites in E per ml, source: laboratory
Dysp	Dyspnea, source: questionnaire
Dyspn	Dyspnea at night, source: questionnaire
Eosin_granulocyt	Number of eosinophil granulocytes×10 per L, source: laboratory
Feathers	Reaction to feathers (dove, canary, parakeet, parrot), source: skin prick test
Forc_exp_vol	Forced expiratory volume for 1 s (L per sec), source: lung function test
Grass_IgE	Specific IgE for grasses in E per ml, source: laboratory
Grasses	Reaction to specific grasses (perennial rye grass, orchard grass, timothy grass) source: skin prick test
Hairs	Reaction to hair of other animals (guinea pig, rabbit), source: skin prick test
Horse	Reaction to horse, source: skin prick test
Immunoglobulin	Immunoglobulin E in E per ml, source: laboratory
Mites1	Reaction to mites, source: skin prick test
Mites2	Reaction to mites in specific places: hay mite, copra mite, flour mite, source: skin prick test
Pc10m	Pc10 methacholine (10% decrease relative to baseline of methacholine), calculated
Pc20m	Pc20 methacholine (mg/ml), a measure of bronchial hyperreactivity, one of the hallmarks of asthma, source: lung function test
Peak_flow	Peak flow (L/ sec), source: lung function test
Reversibility	Reversibility (better lung function after ventolin), source: lung function test
Skin_reaction_degr	Degree of skin test reaction, source: skin prick test
Tiffenau_index	Tiffenau Index (FEVI per VC) to measure bronchus obstruction, source: lung function test
Trees1	Reaction to trees: alder, birch, hazel, source: skin prick test
Trees2	Reaction to trees: elder, oak, elm, poplar, source: skin prick test
Vital_capacity	Vital capacity (liters), source: lung function test
Weeds	Reaction to weeds: mugwort, plantain, source: skin prick test
Wheezing	Wheezing, source: questionnaire

Chained Equations) to supplement 12 of the 35 markers that had missing values. Distribution plots of these markers before and after imputation can be found in Supplemental Fig. 1. We then performed principal component analysis on the 35 variables to reduce the data dimensionally using the r function robPCA (rospca version 1.0.4), which is specifically suited for non-normally distributed data [30]. The distributions of the principal component scores can be found in Supplemental Fig. 2.

PCA investigating the multidimensional structure of the clinical marker data showed the first principal component to account for 35.27% of the variance. Principal component (PC) 1 was largely associated with methacholine challenge values, other measurements of lung capacity/functionality, and lung ailments such as coughing and wheezing. PCs 2, 3, and 4 accounted for 19.54%, 12.46%, and 9.75% of the variance, respectively, and were also mostly associated with functional lung measurements.

PCs 5 and 6 accounted for 8.39% and 6.05% of the variance, respectively, and showed strong association with the FEV1/FVC ratio, also called modified Tiffeneau-Pinelli index. PC7 accounted for 4.28% of the variance and showed very strong associations to eosinophil counts and immunoglobulin levels. PCs 8, 9 and 10 accounted for 2.20%, 1.22%, and 0.84% of the variance and each associated with a variety of allergen response measurements (Fig. 1). Supplemental Table 1 contains the raw eigen values and percentage of variance explained for each PC.

Timepoint 1 EWAS via 10 generated PCs

Epigenome-wide association studies of each of the ten principal components yielded multiple CpGs that reached genome-wide significance. In total, we identified 222 significantly associated CpGs (following a Bonferroni correction, $\alpha = 0.05/742,442$). Of the 222, one CpG,
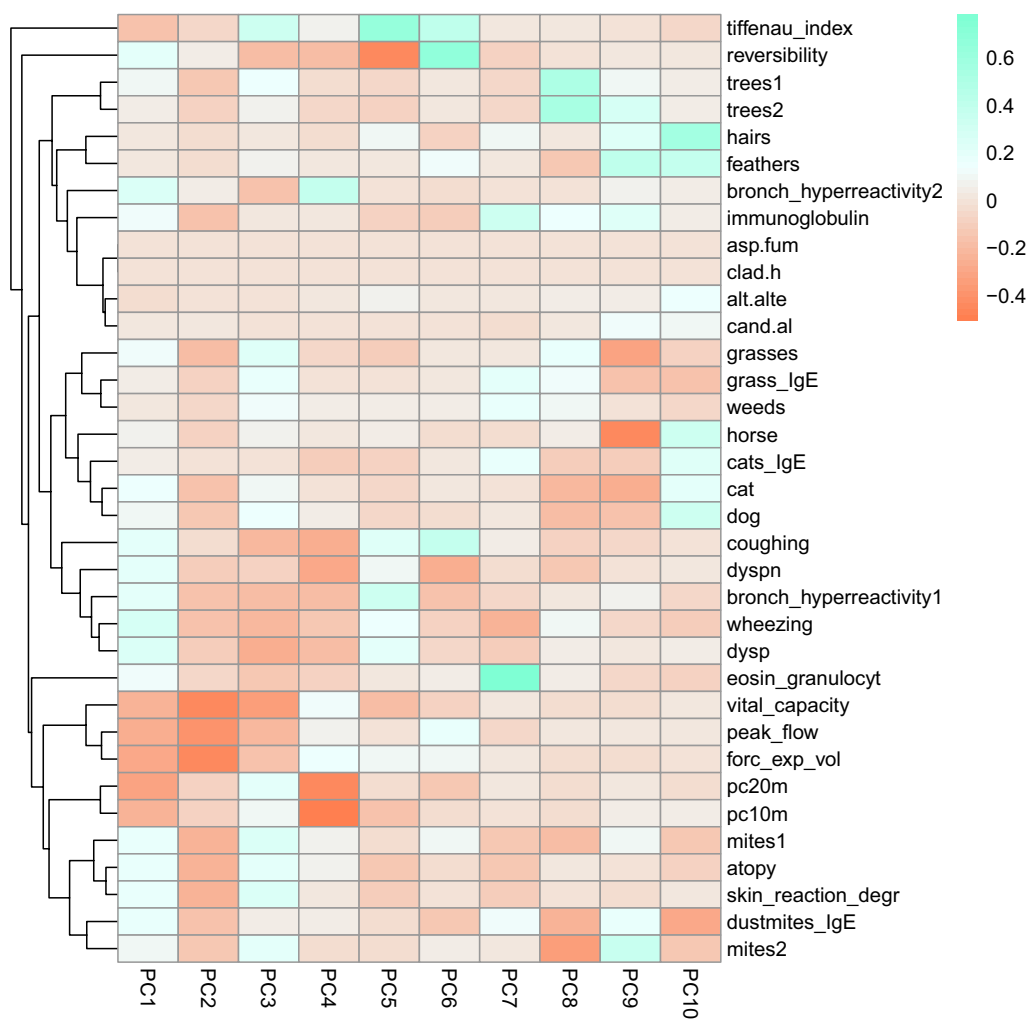
Fig. 1 Heatmap showing correlations between the clinical asthma markers and principal components generated from the asthma marker data. An explanation of the variables is provided in Table 1

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 5 of 15

cg07329820, was associated with two PCs (PCs 9 and 10), resulting in a total of 221 unique CpGs that were significantly associated. Table 2 provides a description of the covariates for the sample population. Table 3 summarizes the number of CpGs identified in each EWAS and Supplemental Fig. 3 shows the results of each individual EWAS via Manhattan plots. Inflation values for each EWAS can be found in Supplemental Table 2. A complete list of all the CpGs significantly associated at T1 can be found in Supplemental Table 3, and complete summary statistics for each of the EWASs can be found in Supplemental Table 4.

By far, the largest number of CpGs were associated with PC7 which mostly captures variance from eosinophil counts and immunoglobulin levels. These 204 CpGs are mainly composed of CpGs with a high average methylation level and almost exclusively show negative effect sizes, such that higher scores on PC7 correlate with lower methylation levels. The PC10 EWAS identified associations with 5 CpGs, followed by the PC2 EWAS with 4 significantly associated CpGs. EWASs for PCs 3, 5, and 9 showed significant associations with 2 CpGs. EWASs for PCs 1, 4, and 8 were significantly associated with 1 CpG, while the EWAS for PC6 was the only association study that did not identify any significantly associated CpGs.

Erichment analysis via the online EWAS Atlas (https://ngdc.cncb.ac.cn/ewas/atlas/index, accessed June 11, 2024) for the 221 unique CpGs was performed. This

Table 2 Sample population covariate frequencies

Age		
Mean	Min	Max
35.8	13	77
Smoke status		
Never	Previous	Current
176	60	83
Sex		
Variable	Freq	Percent (%)
Female	165	51.7
Male	154	48.3
Medication usage		
Variable	Freq	Percent (%)
General lung medication	48	15.0
Betamimetica	37	11.6
Inhaled corticosteroids	26	8.2

analysis showed that 71 (32.1%) and 173 (78.3%) of these CpGs had previously been identified to associate with allergic asthma and fractional exhaled nitric oxide (FENO), respectively (shown in Fig. 2). Further, asthma-related phenotypes such as childhood asthma, allergic sensitization, atopy, serum IgE levels, and others were shown to be previously associated with CpGs identified in our analysis. In total, 190 of the 221 CpGs (86.0%)
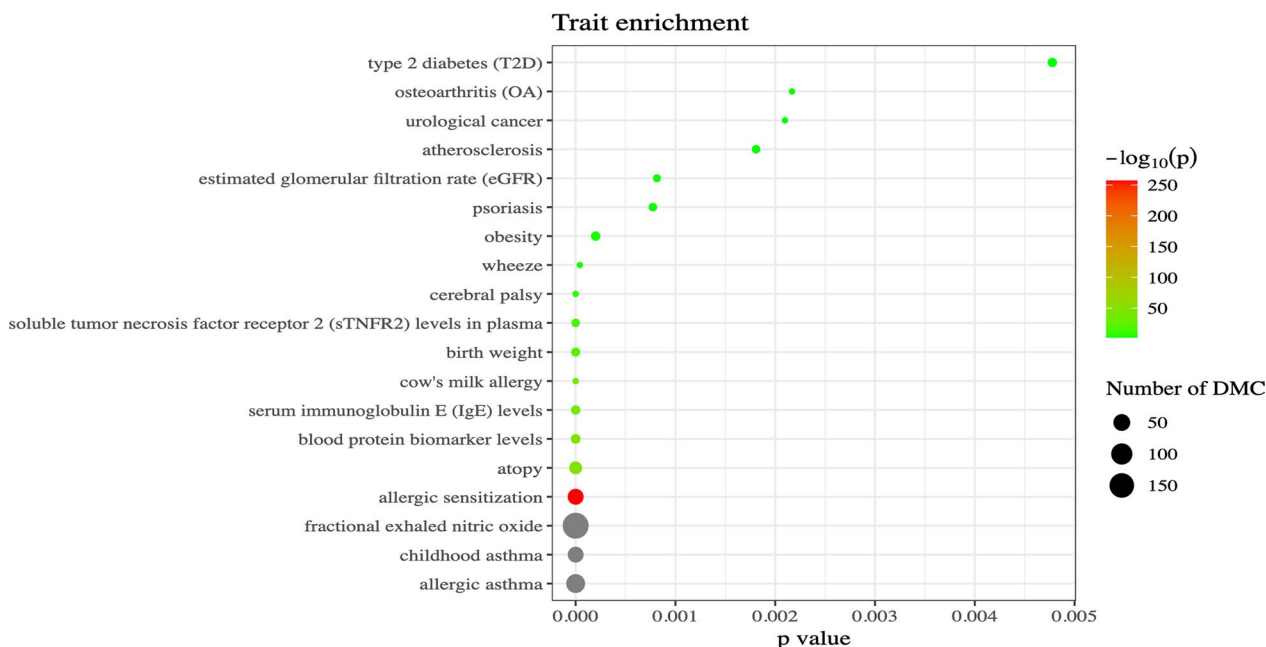
Fig. 2 EWAS atlas enrichment analysis of the 221 unique CpGs identified at TP1 across the 10 EWASs. The plot shows the number of previously identified Differentially Methylated CpGs (DMC) that are present within the CpGs identified in our analysis. The EWAS Atlas job can be found via this job ID: ea5abe0e33f28b27ab3948e2ccd4044c

Van Asselt *et al. Clinical Epigenetics*　　(2024) 16:151

Page 6 of 15

Table 3 Genome-wide significant CpGs identified in each EWAS for time point 1 and 2

Timepoint 1

PC	Genome-wide significant CpGs	Samples
PC1	1	319
PC2	4	319
PC3	2	319
PC4	1	319
PC5	2	319
PC6	0	319
PC7	204	315
PC8	1	319
PC9	2	319
PC10	5	319
Total	222	

Timepoint 2

PC	Genome-wide significant CpGs	Samples
PC1	5	182
PC2	6	182
PC3	1	182
PC4	0	182
PC5	3	182
PC6	1	182
PC7	8	181
PC8	12	182
PC9	2	182
PC10	11	182
Total	49	

had a previously identified association with asthma or some asthma marker, leaving 31 novel, unique associations. The enrichment analysis also highlighted significant overlap with the FoxO signaling pathway (7 CpGs previously associated). Enrichment analysis restricting to the 204 CpGs associated with PC7 recapitulated these findings (Supplemental Fig. 4), showing that all of the previously associated CpGs were identified via this PC7 EWAS. We then performed an additional query of the eFORGE online database (https://eforge.altiusinstitute.org/, accessed July 2nd, 2024) with the 204 CpGs associated with PC7 to identify overlap with cell-type specific DNA regulatory elements. This identified significant associations with regulatory elements in multiple tissue sources, with the greatest number of significant findings (31 in total) associating with various subpopulations of white blood cells (Supplemental Fig. 5).

Overlap with previously identified SNPs
An investigation of proximity with the 221 unique significantly associated CpGs was performed to identify SNPs that have been previously identified by Demenais et al.

to associate with asthma ($p < 5 \times 10^{-8}$, $N = 892$ SNPs). This assessment was performed by searching for SNPs within 1 MB upstream and downstream of each CpG. We identified a total of 17 unique SNPs (1.9% of all asthma-associated SNPs) to reside within 1 MB of a CpG identified at time point 1. These 17 unique SNPs involved only two unique CpGs (cg02046836, cg23515090). A summary of these SNPs within proximity to each CpG is in Supplemental Table 5. Additionally, a query of the BBMRI methylation Quantitative Trait Loci (mQTL) database with these two CpGs was performed, which showed that the asthma GWAS SNPs are not mQTLs for cg02046836, cg23515090, and that these CpGs are also not associated with any other SNPs [31].

Timepoint 2 EWAS
Of the initial 319 individuals who were included at the first timepoint, 182 individuals had a second whole blood sample collected approximately 10 years later. The DNA from these blood samples were also profiled on Illumina EPIC v1 methylation array. No detailed asthma phenotyping was performed at follow-up. We repeated the

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 7 of 15

EWASs for the ten PCs with DNA methylation data at T2. The EWASs of the ten PCs with DNA methylation data from samples collected ten years after asthma phenotyping identified 49 unique CpGs at epigenome-wide significance. While none of the CpGs identified at T1 were significant at T2, three CpGs identified at T2 were within 100 kb of a significant CpG identified at timepoint 1 (Supplemental Table 8). Furthermore, we assessed the correlation (Pearson correlation) of the estimates for all 271 significant associations (222 at T1, 49 at T2) across the two time points, and we identified a strong correlation ($r = 0.82$). Table 3 outlines the number of CpGs identified in each time point 2 EWAS and Supplemental Fig. 6 contains the Manhattan plots for these EWASs. Supplemental Table 6 contains all of the 49 significantly associated CpGs identified at T2, and Supplemental Table 7 states the inflation values for each of these 10 EWASs.

Discussion

Asthma is a complex chronic disease that is influenced by multiple factors including genetics and epigenetics. Here, we have identified 221 unique CpGs that significantly associate with PC scores generated from an advanced, comprehensive panel of 35 clinical markers of asthma. Of these, 190 have previously been found to associate with asthma and asthma-related phenotypes, while the other 31 CpGs are novel findings. Two of these CpGs reside within close proximity (1 MB) to 17 SNPs previously identified in a large GWAS of asthma. We then performed a longitudinal assessment in methylation data from whole blood samples collected approximately 10 years and identified 49 unique significantly associated CpGs and obtained a correlation of the estimates across the two timepoints of 0.82 from the 271 (222 T1 + 49 T2) significant associations.

Historically, asthma as a disease has been challenging to identify the root cause of and to treat effectively [32]. The disease heterogeneity and mixed involvement of environmental stimuli and the genome have largely contributed to this difficulty. Based on GWA studies, certain loci within the genome are associated with most forms of asthma, while specific subtypes of asthma are associated with distinct genetic loci [33, 34]. Further, genetic pleiotropy is most certainly at play within asthma, and other complex diseases, leading to increased complexity when attempting to determine genetic causality. While the genetic involvement of this heterogeneity has started to become elucidated, much of our understanding about variability in phenotype remains to be discovered.

We utilized measured clinical markers of asthma in young adults and their middle-aged parents from families with a history of asthma where at least one individual within the family had been diagnosed with asthma. For each individual, an extensive number of clinical markers related to asthma were captured. This comprehensive dataset, while valuable in its depth, is too large -given the sample size- to glean the valuable nuances through individual assessments of each specific marker. Additionally, some markers that were collected lacked a normal distribution, which is challenging in a traditional EWAS approach. For these reasons, we applied PCA to reduce the dimensionality of the 35 different variables, into 10 PCs with each successive PC capturing a lesser and different portion of the overall variance within the data set, making comparisons of it with a rich DNA methylation dataset much more feasible. Importantly, the PCs maintained a much more normal distribution, making them ideal for EWAS. This method of phenotype data reduction and handling has been previously implemented in a GWAS of rice variants, where GWASs were successfully performed using the PC scores from multiple phenotype variables [35]. To our knowledge, this is the first instance of implementing it for epigenome-wide association studies.

Through our EWASs of each of the 10 PCs, we identified 222 CpGs to be significantly associated with one of the PCs derived from the 35 clinical asthma markers. Of these 222 associations, only a single CpG was found to be replicated across multiple principal components, leaving 221 unique CpGs to be significantly associated. The single overlapping CpG, cg07329820, was identified with PCs 9 and 10. These two PCs align with asthma markers that are very similar (largely allergen response measures), which could explain why they identified the same CpG.

A breakdown of the identified CpGs from each PC shows some intriguing details. PC7 had the greatest number of CpG associations with 204. While this PC accounts for only 4.3% of the total clinical marker variance, it almost exclusively represents the variance associated with blood eosinophil counts and blood immunoglobulin levels. This large number of associations could be indicative of the eosinophilic inflammatory response typically seen in allergic asthma. Previous literature has shown that methylation experiments in whole blood may capture more of the Th2-mediated immune response due to the nature of the cell population, making it the ideal sample source for identifying this type of association [36–38]. It is critical to note that because the proportions of cells themselves in a sample can drastically influence the measurement of methylation, we made sure to incorporate cell proportion estimates as a covariate in our analyses. Without this inclusion, it could be argued that the findings identified here are only due to differences in cellular composition. With this correction in mind, it may be the case that the differential methylation identified here is something global, across multiple

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 8 of 15

immune cell types, either in response to or to promote the eosinophilic inflammatory pathway. A query of the EWAS Atlas of these 204 CpGs showed that, almost exclusively, the CpGs we identified have a previous association with asthma and measurements of asthma (Supplemental Fig. 4). In total 190 of the 204 CpGs had some form of previous association (93.1%), leaving 14 novel associations. Since the majority of overlap of our findings with previous EWASs of asthma, asthma phenotypes, and symptoms are captured within this EWAS of PC7, we postulate that many of the previous findings generated in studies utilizing whole blood samples could be identifying associations with the underlying influences of increased eosinophil counts/activation and increased immunoglobulin levels.

An additional investigation of these 204 CpGs via the eFORGE online database identified multiple significantly associated tissue-specific regulatory elements (Supplemental Fig. 5). The greatest number of associations were from subpopulations of white blood cells in enhancers and weak transcriptional activators. This is to be expected considering our results were generated from samples of whole blood. Many of the subpopulations identified here are lymphocytes, potentially indicating that the methylation profiles of these adaptive immune cells may be altered due to the changes in eosinophil counts and immunoglobulin levels. The tissue group with the next largest number of significant associations was the digestive tract, with six significant associations to various subpopulations of cells. These associations could hint at the conserved role that all mucosal and epithelial barriers play in pathogen protection and the body's response to the external environment. It has also been previously shown that respiratory viral infections can cause dramatic changes on both the presentation of asthma symptoms as well as DNA methylation [39]. In a previous study, Zhu et al. identified 33 differentially methylated CpGs (DMCs) associated with bronchiolitis severity [39]. An assessment for any overlap between these DMCs and our significantly associated CpGs did not identify any overlap, though it could be possible that pathogen exposure could still be playing a role in the DNA methylation signature associated with asthma.

A KEGG pathway enrichment analysis (conducted via the EWAS Atlas) of the 221 unique CpGs highlighted the FoxO signaling pathway, showing that 7 of the CpGs (all found within the PC7 EWAS) have been previously identified to associate with this pathway. This pathway is of particular interest due its significant role in regulating apoptosis, glucose metabolism, oxidative stress, and longevity, which has led to findings associating its dysregulation with other illnesses such as Alzheimer's disease, type 2 diabetes, and cancer [40, 41]. It has been

previously shown that FoxO1 modulates IL-9 generating Th9 cells [42]. Further, overexpression of IL-9 in the lungs can result in eosinophilic inflammatory infiltration and mucus secretion [42, 43]. Additionally, FoxO3 has been found to be expressed in airway epithelium playing a critical role in controlling the innate immune response to airway infections [43]. It could be, based on these intertangled findings of DNA methylation with eosinophilic presence and FoxO signaling, that epigenetic modulation of FoxO signaling plays a role in its contribution to asthma and airway inflammation.

Principal component 1 identified a single significantly associated CpG, cg18182148. While this principal component does capture the largest amount of variance (35.3%), the variance it captures is largely from one specific subset of measurements, most notably those measuring general lung functionality such as vital capacity, peak flow, forced expiratory volume, and methacholine challenge response. This is continued with PC2, PC3, and PC4 which were significantly associated with 4, 2, and 1 CpGs, respectively. These PCs, like PC1, capture a significant amount of variance from these same lung function measurements. This may indicate that these measurements of lung function, though important in characterizing symptoms of asthma, may have limited identifiable associations with DNA methylation. The specific CpG significantly associated with PC1, cg18182148, is the only CpG of these identified here related to immune regulation. It has not been previously shown to associate with asthma phenotypes, though it has been previously identified in studies of various cancers of the prostate, liver, and colon (via the EWAS Atlas) [44–47]. These associations are likely due to its placement within the transcription start site (TSS) of GFI-1, a strong oncogenic and hematopoietic regulator. GFI-1 plays a critical role in lymphoid differentiation, which could be the reason for its association here due to the large influence the immune system can play on lung functionality and, more generally, asthma as a whole [48]. While analyzing a whole blood sample, as done in this study, is a reasonable approach for investigating immune-mediated conditions like asthma, examining the relationship between lung function measurements and DNA methylation from lung epithelial tissue may reveal additional insights not captured here. This could offer a deeper understanding of the local cellular environment influencing specific lung function differences.

Principal components 5 and 6 capture variance associated with bronchial hyperreactivity, coughing, reversibility (of lung function following the administering of ventolin), and Tiffenau Index. PC5 was significantly associated with 2 CpGs, while PC6 did not have any significant associations. The two CpGs identified here,

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 9 of 15

cg07454584 and cg18561513, have not been previously shown to have any associations with asthma or other immune-mediated illnesses (via the EWAS Atlas). Like the measurements captured in PCs 1–4, these asthma markers, while important for asthma diagnosis and the understanding of symptoms, may not have large quantities of associations with DNA methylation measured via whole blood.

Principal components 8, 9, and 10 contain much of the variance from the numerous allergen tests included in this study (containing both skin prick test results and serum specific IgE measurements). These PCs were significantly associated with one, two, and 5 CpGs, respectively. Of the 8 CpGs identified here, all but one has not been previously shown to associate with chronic illnesses or immune-mediated diseases. Cg14161241, which was significantly associated with PC10, is the lone exception with previous associations with obesity and type 2 diabetes [49]. Like many of the other markers represented via PCs 1–6, the effect on DNA methylation measurable via a whole blood sample is likely limited with these markers.

Due to the well-documented genetic contributions to asthma, we next compared our findings within the methylome to those previously discovered in the genome via large GWASs [10, 50]. Based on the findings of Demenais et al., we searched for significantly associated SNPs within 1 MB of the 221 unique CpGs that we identified. Of the significantly associated SNPs ($p < 5 \times 10^{-8}$, $N = 892$ SNPs), we found 17 (1.9%) to be within 1 MB of 2 of the 221 CpGs that we identified (cg02046836, cg23515090). Interestingly, neither of these two CpGs have been previously identified as an mQTL [31]. The proximity of these SNPs to the identified CpGs could be caused from multiple factors outside of known standard mQTL associations. It could be that these CpGs do have some genetic influence acting on their methylation status that has not yet been characterized. Additionally, it is possible that the SNPs and CpGs that are in close proximity to one another is simply a random occurrence, and their presence and downstream effects are independent of one another. To summarize, we found limited overlap of our EWAS signal with top loci from a previously performed GWAS of asthma and that these significant CpGs nearby these GWAS SNPs were not directly affected by mQTLs, which suggests that our EWAS largely captures independent epigenetic signal at loci that also harbor small amounts of genetic variants that influence genetic susceptibility to asthma.

Thus far, we have shown significant associations with clinical markers of asthma from a cross-sectional timepoint that corroborates other previous findings from similar studies. However, an increase in our understanding of asthma may lie in investigating the disease to some

longitudinal capacity. Two previous studies in the Isle of Wight Birth Cohort (IOWBC) were conducted in 2022, which investigated the association of DNA methylation with asthma acquisition across adolescence and adulthood [51, 52]. In each of these studies, several CpGs, with some residing in immune regulatory genes, were identified to be associated with asthma acquisition [51, 52]. The findings from these two studies highlight promising insight into the potential longitudinal persistence of DNA methylation signatures associated with asthma.

After our initial findings, with the availability of an additional sample collected after 10 years for many of the individuals, we investigated the longitudinal persistence of the DNA methylation signature that was identified to associate with asthma markers at the first timepoint. This analysis was based on the same PC data from the original 35 clinical markers measured at baseline. In total, these 10 EWASs identified a total of 49 significantly associated CpGs, without replication across multiple PCs. When looking within a 100 kb window of proximity, we found that 3 CpGs identified at timepoint 2 (found via PCs 2, 3, and 7) were within 100 kb of 4 CpGs identified at T1 (all found via PC7). A query of the EWAS Atlas, however, using these 49 CpGs identified at T2 showed that 3 and 1 CpGs have been previously associated with FENO and allergic sensitization, respectively (Fig. 3). This highlights that there may be some longitudinal persistence of an asthmatic epigenetic signature. We estimated, when comparing the estimates of all 271 significant associations (222 + 49) across the two timepoints, a strong correlation of 0.82 adding to the notion that some long-term persistence of the asthma-associated methylation signature is occurring. It would have been of value if our study of asthma clinical markers had been reassessed along with the longitudinal blood sampling, as changes in asthma prevalence and presentation could lead to changes in DNA methylation, which could be better captured in future studies. Additional studies investigating the long-term lung functionality following asthma diagnosis, even in cases of remission, could benefit from investigating DNA methylation longitudinally in a more nuanced, complete manner.

Our study does have some limitations that should be considered. First, all individuals included in this study are from the Netherlands and of European descent. Asthma prevalence can differ greatly depending on the local environment of the individual and their ethnicity. Specifically, the environment of the Netherlands and their clinical diagnostic criteria could influence an individual's asthmatic status. Including a diverse population could improve the generality of the findings to other groups of individuals. It is well documented that research involving non-European populations is lacking greatly,
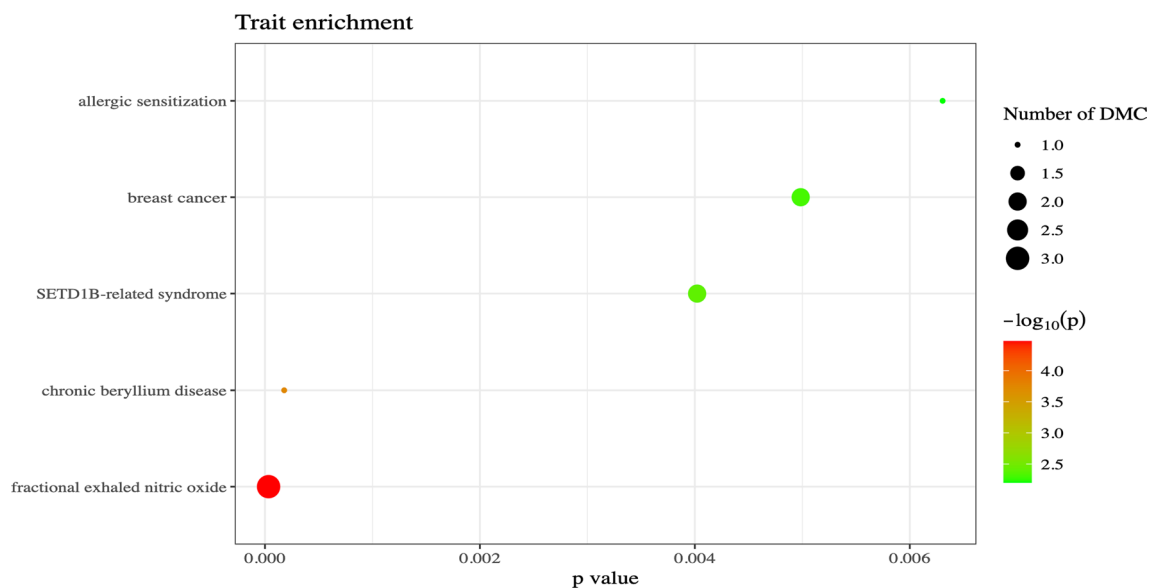
Fig. 3 EWAS atlas enrichment analysis of the 49 unique CpGs identified at TP2 across the 10 EWASs. The plot shows the number of previously identified Differentially Methylated CpGs (DMC) that are present within the CpGs identified in our analysis. The EWAS Atlas job can be found via this job ID: a3e03507f39d7af4c7a82d0b0725308e

and we strongly advocate for research within these populations to expand the generality and applicability of research findings to these populations of individuals. The results generated from the second time point contained a lower number of individuals, who took part in the NTR biobank study and did not have clinical marker measurements, which could have shed some insight on the persistence of the asthma symptoms and severity at that time. Future studies investigating asthma longitudinally could benefit greatly from having longitudinal DNA methylation and clinical marker data. Overall, increasing the number of individuals at both timepoints using multiple individuals from varying backgrounds would be the most optimal strategy and offer the most comprehensive assessment of asthma.

Conclusion

Asthma, along with many other chronic, immune-mediated diseases, is challenging to diagnose, classify, and treat. Here, we contribute a comprehensive assessment of the relationship between DNA methylation and a collection of clinical asthma markers. We identified 270 unique, significant CpGs over two timepoints to be associated with principal component scores based on clinical asthma markers. The large majority of these CpGs from timepoint 1 were associated with PC7, which represented eosinophil counts and blood immunoglobulin levels and 190 were previously associated with asthma, asthma markers, and other allergic illnesses. We also investigated the proximity of the 221 identified unique CpGs at

timepoint 1 to previously identified SNPs that associate with asthma. We found 17 unique SNPs that were previously identified to be associated with asthma within 1 MB of two of the 221 CpGs reported here demonstrating limited proximal overlap with these known genetic associations. Over approximately 10 years we showed that the correlation of the estimates across the two timepoints from the 271 significant associations was strong at 0.82, pointing to a persistent methylation signature associated with asthma. Thus, these results highlight a robust, persistent epigenetic signal in whole blood associated with asthma and, more specifically, a selection of clinical asthma markers.

Methods

Participants

The study participants and the enrollment process have been previously described [53–55]. Participants came from twin families (young adult twins and their parents) from the Netherlands [55]. To be eligible for the study, at least one member of the family had to report to have been diagnosed with asthma, indicated by self-report in the Netherlands Twin Register (NTR) survey 1 or 2, collected in the early 1990s (please see Supplementary Tables 9 and 10). During a visit to the Vrije Universiteit Medical Center (VUMC), blood samples and over 40 biological and clinical measures of asthma were collected from 425 persons [56]. A complete list of the clinical markers that were collected can be found in Table 1. A subset of 233 persons participated in the NTR Biobank

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 11 of 15

project between 2004 and 2008 in which blood and buccal samples were collected [29]. DNA was extracted at the time of sample collection or immediately prior to the assessment of methylation [57]. All samples were simultaneously measured on the Illumina Infinium EPIC v1 methylation array. In total, 375 of the 425 individuals had a sample remaining for analysis, and 341 of these had good quality Illumina blood EPIC Array data from at least the first time point, and 232 individuals had blood Illumina EPIC Array data from two time points.

The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03–180).

Sample collection and DNA extraction

The procedures of whole blood sample collection have been previously described [29, 53, 58]. The genomic material analyzed in this study have been extracted and assessed for DNA quantity, quality, purity, and individual identity. The DNA analyzed in this study was extracted from whole blood samples using the Zymo Quick DNA mini-prep Kit [57]. Genomic material was quantified via the Invitrogen Qubit Broad-Range Fluorescent Assay, and sample purity was assessed using standard absorbance metrics via a SpectraMax microplate reader.

Genotyping

The genotyping process has been previously described [54, 59]. Briefly, genotype data were generated via the Illumina Infinium Global Screening Array. All sample data quality was assessed vigorously, and the refined genotype data was used to confirm familial relationships and sample sex. In total, 20 individuals were omitted for analysis due to mismatched familial identities.

DNA methylation assessment

DNA bisulfite conversion was performed utilizing the Zymo EZ-96 DNA Methylation Kit [60]. DNA methylation was assessed using the Illumina Infinium EPIC v1 DNA Methylation Array on all samples at the Avera Genetics Laboratory [61]. The samples were fully randomized across arrays.

DNA methylation data quality control

Quality control has been described in detail previously [54]. In brief, DNA methylation data quality was performed using two bioinformatic tools. First, DNA methylation data were preliminarily assessed using the Illumina GenomeStudio 1.0 software. Second,

methylation data quality control and normalization were conducted using a pipeline developed by the Biobank-based Integrative Omics Study (BIOS) Consortium in R (version 4.4.0). Sample quality was first assessed using the R package MethylAid (v1.38.0) to omit any outliers in sample performance (default thresholds) [62]. Array probe filtering and functional normalization were performed using the R package DNAmArray (v2.2.0) [63]. The R package omicsPrint (v.1.24.0) was utilized to identify sample genotypes based on methylation probes to verify sample relationships gathered from genotype data via the GSA [64]. The functions getSex from DNAmArray and the R package meffil (v1.3.8) were used to identify and confirm sample sex based on X chromosome methylation pattern [65]. The following probe filters were applied: Probes were set to missing (NA) in a sample if they had an intensity value of exactly zero, detection P-value > 0.01, or bead count < 3. DNAmArray will also remove any probes that show a success rate below 0.95 across all samples. In total, this filtering process reduced our number of individuals to 341 (from the original 375). Finally, previous studies performed by Zhou et al. in 2017 identified polymorphic and cross-reactive probes that are included on Illumina methylation array platforms [66]. Generally, they recommend an omitting of approximately 100,000 probes on the EPIC methylation array. This removal process was carried out using the probemasking() function via DNAmArray. Following these steps, a total of 759,263 methylation sites were included in this study out of the total 865,859 possible CpGs. Only autosomal methylation sites were considered for downstream analyses, which left a total of 742,442 CpGs included in the final analyses.

Covariates

Several covariates were utilized in our epigenome-wide association studies. It has been shown that many of these can produce dramatic effects on DNA methylation, so it is critical to correct for them. For each sample at each timepoint, covariate data was collected at the same time. We included age, sex, array row, sample plate (dummy-coded), smoking status, asthma/lung medication use, and cellular proportions as covariates. For smoking status, individuals were classified into three groups (codes as 0 = Never smoker, 1 = former smoker, 2 = current smoker). Cellular proportions (natural killer cells, monocytes, B cells, CD4 + T cells, CD8 + T cells) for each of the whole blood samples were estimated via the methylation data from that sample. We calculated cellular proportions for all samples using the IDOL (Identifying Optimal Libraries) whole blood reference library, which is specifically optimized for estimating cellular composition in whole blood samples [67, 68]. The r script utilized for this

process can be located at https://github.com/AVA-27/Asthma_EWAS. The usage of specific and general asthma medications (inhaled corticosteroids (0 = no, 1 = yes), betamimetica (0 = no, 1 = yes), general lung medication (0 = no, 1 = yes)) was recorded for each participant at the first time point, and all three medications were included as covariates for the T1 EWAS. The information gathered here was used at both time points. Due to limited variability in medication use at the second timepoint, only a single "general lung medication use" variable was used as a covariate for T2.

Methylation data annotation

Genomic annotations were gathered from the EPIC manifest file that is provided by Illumina (MethylationEPIC_v-1-0_B5.csv): locations of CpG islands, ENCODE DNase I Hypersensitive sites (DHSs), ENCODE transcription factor binding sites (TFBSs), open chromatin, FANTOM4 enhancers and FANTOM5 enhancers, etc. Additionally, the gene name associated with (or nearest gene if not located within one) each significantly associated CpG was gathered using the cpgInfo() function via the R package DNAmArray. For CpGs not located within a gene, the closest gene was recorded using the UCSC Genome Browser. These updated gene associations are within the "Gene_consolidated" column of supplemental Tables 3 and 4. Genome build 37 coordinates were utilized for all the analyses.

Asthma phenotyping

Of the total list of available clinical markers, exclusions for markers were made if there was no variation among participants. In total, 35 clinical markers were selected to be included in the principal component analysis. If an individual themselves had more than 6 missing values they were excluded from the analysis as well, which resulted in removing 4 individuals leaving a total of 319 to include. 10 of the 35 markers contained missing values following these exclusion processes. For these missing values, the R package mice (v3.16.0) was utilized to impute the missing data. The default settings were utilized, 5 imputations per marker, predictive mean matching (pmm) for continuous traits, logistic regression (logreg) for categorical traits. Distribution plots for these markers prior to and after imputation can be found in Supplemental Fig. 1.

Principal component analysis

We applied principal component analysis (PCA) to reduce the number of comparisons in lieu of investigating each clinical marker individually for a couple critical reasons. Many of the clinical markers had a non-normal distribution and statistical outliers, which can make analysis

via EWAS difficult. Rather than choosing to omit any markers (in a potentially biased manner), PCA can concatenate these markers based on the variance that they explain in a non-biased manner. For these reasons, PCA was an optimal choice for both data reduction and normalization, and the process utilized for PCA is explained below. Following imputation, the data was normalized using the scale function in R, which performs a z-score transformation for each marker. We then performed PCA including 10 principal components (PCs) on our 35 clinical markers that were included. PCA was performed using the robpca() r function via the R package rospca (v1.0.4), which is suited for non-normally distributed data [30, 69]. A scree plot of the eigen values for each of the 10 PCs can be found in Supplemental Fig. 7. Outliers for each PC were identified via the R package ewaff (v0.0.2) using the ewaff.handle.outliers function with the following settings: method set to iqr, iqr range set to 3. Outliers were found in PCs 7–10. Principal component 7 contained four outliers, which were removed prior to further analyses ($n = 315$). Principal components 8–10 each contained several outliers. Due to the nature of the clinical marker data and the way the PCA captured the variance, these PCs largely captured variance from markers recording allergen responses, which resulted in individuals who had strong allergen responses having extreme PC scores. Due to the biological significance of these people, we decided to keep these individuals in the subsequent analyses by dichotomizing the group based on their outlier status (1 = outliers, 0 = not outlier), rather than excluding them. These new dichotomized values for PCs 8, 9, and 10 were then used for the downstream analyses.

Frequencies for these PCs can be found in Supplemental Table 11.

EWAS

Ten epigenome-wide association studies (EWAS) were then performed via the PC scores for each individual for each of these 10 PCs using DNA methylation beta value as the outcome and the PC scores for each individual as the predictor. Due to the related nature of the samples utilized here, each EWAS was conducted via the r package gee (v4.13–25) using a generalized estimating equations (GEE) model that corrected for the correlation structure in families. The following settings were utilized: Gaussian link function, 100 iterations, and the "exchangeable" option to correct for the familial structure. Additionally, each EWAS included age, sex, methylation array row, bisulfite sample plate (dummy-coded), smoking status (codes as 0 = Never smoker, 1 = former smoker, 2 = current smoker), estimated cellular proportions (natural killer cells, monocytes, B cells, CD4 + T cells, CD8 + T cells), and lung

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 13 of 15

medication status via questionnaire (inhaled corticosteroids, betamimetica, general lung medication) as covariates. For each EWAS, a standard Bonferroni correction ($\alpha = 0.05/$# of CpGs (742,442)) was applied to account for multiple testing to identify significantly associated CpGs.

Enrichment analysis was performed in the EWAS Atlas with the EWAS Toolkit, which compares against previously known associations from previous studies investigating DNA methylation. We also conducted an enrichment analysis via the online tool eFORGE, which is a database of known CpG associations to cell-type specific regulatory elements. This analysis was performed using the "Consolidated Roadmap Epigenomics – Chromatin – All 15-state marks" data set and the standard settings. Next, we cross-analyzed our significantly associated CpGs with significantly associated SNPs ($p < 5 \times 10^{-8}$, $N = 892$ SNPs) identified by Demenais et al. [10] in a large meta-analysis, utilizing leave-one-out GWAS summary statistics without NTR participants. As has been previously utilized in other studies, a 1 MB range around each CpG was utilized as a cutoff point to identify SNPs within close proximity [70]. Finally, we investigated the CpGs that were found to be within close proximity to a known significantly associated SNP for any previously identified mQTL association by querying the BBMRI mQTL database (based on a previous study by Bonder et. al. [31]).

Of the 319 individuals included in timepoint 1, 182 had a follow-up blood sample with methylation data available for subsequent analyses at a later time point. For this second timepoint, an additional 10 EWASs were performed using the same PC scores generated from the timepoint 1 clinical marker data to assess longitudinal persistence of the DNA methylation signature associated with these asthma markers. The EWASs were performed with largely the same models as at timepoint 1 samples (stated above) but utilizing covariate data from time point 2. Due to the reduction of individuals and a lack of unique medication covariate combinations, asthma medication status was reduced to a Bayesian variable grouping individuals on the basis of using any one of the three listed medications. This single covariate capturing asthma medication status was then used for these timepoint 2 EWASs. An additional longitudinal comparison was made looking for CpGs in timepoint 1 that were within 100 KB (upstream and downstream) of a CpG identified in timepoint 2. Similar to timepoint 1, following this the timepoint 2 CpGs were grouped and a query of the EWAS Atlas was performed looking for previously identified associations.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13148-024-01765-0.

Additional file1 (CSV 35 KB)

Additional file2 (CSV 9 KB)

Additional file3 (docx 5,925 KB)

Author contributions

Conceptualization of the project was performed by A.J.V.A., E.d.G., D.I.B., E.A.E. and J.v.D. Formal analysis was performed by A.J.V.A., J.J.B., C.T.F., B.N.J., and J.v.D. Methylation and sample data curation and validation was completed by A.J.V.A, N.K., S.V., and P.H. Preliminary data analysis, genotype data validation, and sample validation was performed by J.-J.H., R.P., and L.L. Asthma data preparation and validation was completed by L.L., A.H.M.Z., and S.J.V. Writing of the original draft was performed by A.J.V.A. and J.v.D. Review and editing of the draft was performed by A.J.V.A., J.J.B., C.T.F., E.d.G., D.I.B., E.A.E. and J.v.D. Project administration and supervision was performed by E.d.G., D.I.B., E.A.E. and J.v.D. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

All NTR data can be requested from bona fide researchers (https://ntr-data-request.psy.vu.nl/). Because of the consent given by the study participants, the data cannot be made publicly available. The pipeline for DNA methylation–array analysis developed by the Biobank-based Integrative Omics Study (BIOS) consortium is available at https://molepi.github.io/DNAmArray_workflow/ (https://doi.org/https://doi.org/10.5281/zenodo.3355292). The scripts utilized for the imputation, PCA, scaling, and EWAS processes can be found at https://github.com/AVA-27/Asthma_EWAS.

Declarations

Ethics approval and consent to participate

All individuals within this study provided written consent for their participation. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance-FWA00017598; IRB/institute codes, NTR 03–180).

Consent for publication

All authors consent to the publication of this manuscript and the accompanying material in its entirety.

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 14 of 15

Author details
[1]Avera McKennan Hospital and University Health Center, Sioux Falls, SD, USA. [2]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [3]Amsterdam Public Health Research Institute, Amsterdam, The Netherlands. [4]Department Pulmonary Medicine, Amsterdam University Medical Center, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands. [5]Amsterdam Reproduction and Development (AR&D) Research Institute, Amsterdam, The Netherlands. [6]Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

References
1. Asthma. 2023 [cited 2024 02/26/2024]; Available from: https://www.who.int/news-room/fact-sheets/detail/asthma#:~:text=Inflammation%20and%20narrowing%20of%20the,and%20caused%20455%20000%20deaths
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet. 2020;396(10258):1204–22. https://doi.org/10.1016/S0140-6736(20)30925-9.
3. Jones TL, Neville DM, Chauhan AJ. Diagnosis and treatment of severe asthma: a phenotype-based approach. Clin Med (Lond). 2018;18(Suppl 2):s36–40.
4. Svenningsen S, Nair P. Asthma endotypes and an overview of targeted therapy for asthma. Front Med. 2017;4:158.
5. Gans MD, Gavrilova T. Understanding the immunology of asthma: pathophysiology, biomarkers, and treatments for asthma endotypes. Paediatr Respir Rev. 2020;36:118–27.
6. Kuruvilla ME, Lee FE, Lee GB. Understanding asthma phenotypes, endotypes, and mechanisms of disease. Clin Rev Allergy Immunol. 2019;56(2):219–33.
7. Lötvall J, 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.
8. Kaur R, Chupp G. Phenotypes and endotypes of adult asthma: moving toward precision medicine. J Allergy Clin Immunol. 2019;144(1):1–12.
9. Ntontsi P, et al. Genetics and epigenetics in asthma. Int J Mol Sci. 2021;22(5):2412.
10. Demenais F, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet. 2018;50(1):42–53.
11. Ferreira MAR, et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. Am J Hum Genet. 2019;104(4):665–84.
12. Pividori M, et al. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med. 2019;7(6):509–22.
13. Zhu Z, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat Genet. 2018;50(6):857–64.
14. Han Y, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. Nat Commun. 2020;11(1):1776.
15. Zhu Z, et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. Eur Respir J. 2019;54(6):1901507.
16. Yang IV, et al. DNA methylation and childhood asthma in the inner city. J Allergy Clin Immunol. 2015;136(1):69–80.
17. Xu CJ, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. Lancet Respir Med. 2018;6(5):379–88.
18. Sheikhpour M, et al. A review of epigenetic changes in asthma: methylation and acetylation. Clin Epigenetics. 2021;13(1):65.
19. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–20.
20. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23–38.
21. Jin Z, Liu Y. DNA methylation in human diseases. Genes Dis. 2018;5(1):1–8.
22. Torrone D, et al. Reproducibility and intraindividual variation over days in buccal cell DNA methylation of two asthma genes, interferon γ (IFNγ) and inducible nitric oxide synthase (iNOS). Clin Epigenetics. 2012;4(1):3.
23. Lovinsky-Desir S, et al. DNA methylation of the allergy regulatory gene interferon gamma varies by age, sex, and tissue type in asthmatics. Clin Epigenetics. 2014;6(1):9.
24. Cardenas A, et al. The nasal methylome as a biomarker of asthma and airway inflammation in children. Nat Commun. 2019;10(1):3095.
25. Yang IV, et al. The nasal methylome and childhood atopic asthma. J Allergy Clin Immunol. 2017;139(5):1478–88.
26. Forno E, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. Lancet Respir Med. 2019;7(4):336–46.
27. Jeong A, et al. DNA methylation in inflammatory pathways modifies the association between BMI and adult-onset non-atopic asthma. Int J Environ Res Public Health. 2019;16(4):600.
28. Bae DJ, et al. Epigenetic changes in asthma: role of DNA CpG methylation. Tuberc Respir Dis (Seoul). 2020;83(1):1–13.
29. Willemsen G, et al. The Netherlands twin register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet. 2010;13(3):231–45.
30. Hubert M, Engelen S. Robust PCA and classification in biosciences. Bioinformatics. 2004;20(11):1728–36.
31. Bonder MJ, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017;49(1):131–8.
32. North ML, Ellis AK. The role of epigenetics in the developmental origins of allergic disease. Ann Allergy Asthma Immunol. 2011;106(5):355–61.
33. Zhu Z, et al. Investigating asthma heterogeneity through shared and distinct genetics: insights from genome-wide cross-trait analysis. J Allergy Clin Immunol. 2021;147(3):796–807.
34. Schoettler N, et al. Advances in asthma and allergic disease genetics: is bigger always better? J Allergy Clin Immunol. 2019;144(6):1495–506.
35. Yano K et al. GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. In: Proceeding National Academy Science U S A, 2019;116(42): 21262-21267.
36. Georas SN, et al. T-helper cell type-2 regulation in allergic disease. Eur Respir J. 2005;26(6):1119–37.
37. Bosnjak B, et al. Treatment of allergic asthma: modulation of Th2 cells and their responses. Respir Res. 2011;12(1):114.
38. Barnes PJ. Pathophysiology of allergic inflammation. Immunol Rev. 2011;242(1):31–50.
39. Zhu Z, et al. Epigenome-wide association analysis of infant bronchiolitis severity: a multicenter prospective cohort study. Nat Commun. 2023;14(1):5495.
40. Du S, Zheng H. Role of FoxO transcription factors in aging and age-related metabolic and neurodegenerative diseases. Cell Biosci. 2021;11(1):188.
41. Farhan M, et al. FOXO signaling pathways as therapeutic targets in cancer. Int J Biol Sci. 2017;13(7):815–27.
42. Malik S, Awasthi A. Transcriptional control of Th9 cells: role of Foxo1 in interleukin-9 induction. Front Immunol. 2018;9:995.
43. Kim ME, Kim DH, Lee JS. FoxO transcription factors: applicability as a novel immune cell regulators and therapeutic targets in oxidative stress-related diseases. Int J Mol Sci. 2022;23(19):11877.
44. Tang Y, et al. Promoter DNA methylation analysis reveals a combined diagnosis of CpG-based biomarker for prostate cancer. Oncotarget. 2017;8(35):58199–209.
45. Shen J, et al. Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using infinium humanmethylation 450 beadchips. Epigenetics. 2013;8(1):34–43.
46. Sun XJ, et al. An integrated analysis of genome-wide DNA methylation and gene expression data in hepatocellular carcinoma. FEBS Open Bio. 2018;8(7):1093–103.

Van Asselt *et al. Clinical Epigenetics*   (2024) 16:151

Page 15 of 15

47. Zhu L, et al. Genome-wide DNA methylation profiling of primary colorectal laterally spreading tumors identifies disease-specific epimutations on common pathways. Int J Cancer. 2018;143(10):2488–98.

48. van der Meer LT, Jansen JH, van der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. Leukemia. 2010;24(11):1834–43.

49. Kirchner H, et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Mol Metab. 2016;5(3):171–83.

50. Moffatt MF, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363(13):1211–21.

51. Rathod A, et al. DNA methylation and asthma acquisition during adolescence and post-adolescence, an epigenome-wide longitudinal study. J Pers Med. 2022;12(2):202.

52. Li L, et al. Newborn DNA methylation and asthma acquisition across adolescence and early adulthood. Clin Exp Allergy. 2022;52(5):658–69.

53. Wu T, et al. Genetic and environmental influences on objective intermediate asthma phenotypes in Dutch twins. Eur Respir J. 2010;36(2):261–8.

54. Van Asselt AJ, et al. Genome-wide DNA methylation profiles in whole-blood and buccal samples-cross-sectional, longitudinal, and across platforms. Int J Mol Sci. 2023;24(19):14640.

55. Los H, Postmus PE, Boomsma DI. Asthma genetics and intermediate phenotypes: a review from twin studies. Twin Res. 2001;4(2):81–93.

56. Bechger TM, Boomsma DI, Koning H. A limited dependent variable model for heritability estimation with non-random ascertained samples. Behav Genet. 2002;32(2):145–51.

57. Quick-DNA Miniprep Kit Instruction Manual, in Online, Z. Research, Editor. 2020, Zymo Research: https://www.zymoresearch.com/collections/quick-dna-kits/products/quick-dna-miniprep.

58. van Dongen J, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun. 2016;7:11115.

59. Van Asselt AJ, Ehli EA. Whole-genome genotyping using DNA microarrays for population genetics. In: Eyster KM, editor. Estrogen receptors methods and protocols. New York, NY: Springer US; 2022. p. 269–87.

60. Research Z, EZ-96 DNA Methylation™ Kit, in Online, Z. Research, Editor. 2021.

61. Infinium HD Assay Methylation Protocol Guide. [Protocol] 2015 [cited 2020; 15019519 v01:[Available from: https://www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html.

62. van Iterson M, et al. MethylAid: visual and interactive quality control of large illumina 450k datasets. Bioinformatics. 2014;30(23):3435–7.

63. Fortin J-P, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15(11):503.

64. van Iterson M, et al. omicsPrint: detection of data linkage errors in multiple omics studies. Bioinformatics. 2018;34(12):2142–3.

65. Min JL, et al. Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics. 2018;34(23):3983–9.

66. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45(4): e22.

67. Salas LA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the illumina humanmethylationepic beadarray. Genome Biol. 2018;19(1):64.

68. Koestler DC, et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinf. 2016;17(1):120.

69. Hubert M, Rousseeuw P, Branden K. ROBPCA: a new approach to robust principal component analysis. Technometrics. 2005;47:64–79.

70. Odintsova VV, et al. DNA methylation in peripheral tissues and left-handedness. Sci Rep. 2022;12(1):5606.

Publisher's Note