- 1 Epigenetic Signatures of Asthma: A Comprehensive Study of DNA Methylation and Clinical
- 2 Markers
- 3 Authors: Austin J. Van Asselt^{1,2}, Jeffrey J. Beck¹, Brandon N. Johnson¹, Casey T. Finnicum¹,
- 4 Noah Kallsen¹, Sarah Viet¹, Patricia Huizenga¹, Lannie Ligthart^{2,3}, Jouke-Jan Hottenga^{2,3}, René
- 5 Pool^{2,3}, A.H Maitland-van der Zee^{3,4}, S.J. Vijverberg^{3,4}, Eco de Geus^{2,3}, Dorret I. Boomsma^{2,3,5,6},
- 6 Erik A. Ehli¹, Jenny van Dongen^{2,3,5}
- 7 Affiliations:
- 8 1. Avera McKennan Hospital & University Health Center, Sioux Falls, SD, USA
- 9 2. Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands
- 10 3. Amsterdam Public Health research institute, Amsterdam, the Netherlands
- 11 4. Amsterdam Medical Center, Dept. Pulmonary Medicine, Amsterdam UMC, University of
- 12 Amsterdam, Amsterdam, the Netherlands
- 13 5. Amsterdam Reproduction and Development (AR&D) research institute, Amsterdam, the
- 14 Netherlands
- 15 6. Complex Trait Genetics, Center for Neurogenomics and Cognitive Research,
- 16 Vrije Universiteit Amsterdam, The Netherlands
- 17 Keywords: Methylation, Asthma, Epigenetics, Epigenome-wide association, Microarrays
- 18
- 19 Abstract:
- 20 Background: Asthma, a complex respiratory disease, presents with inflammatory symptoms in
- 21 the lungs, blood, and other tissues. We investigated the relationship between DNA methylation
- 22 and 35 clinical markers of asthma. The Illumina Infinium EPIC v1 methylation array was used to
- evaluate 742,442 CpGs in whole blood samples from 319 participants. They were part of the

24 Netherlands Twin Register from families with at least one member suffering from severe asthma. 25 Repeat blood samples were taken after 10 years from 182 of these individuals. Principal 26 component analysis (PCA) on the clinical markers yielded ten principal components (PCs) that 27 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. 28 29 Results: 221 unique CpGs reached genome-wide significance at timepoint 1 (T1) after Bonferroni correction. PC7 accounted for the majority of associations (204), which correlated 30 31 with loadings of eosinophil counts and immunoglobulin levels. Enrichment analysis via the 32 EWAS Atlas identified 190 of these CpGs to be previously identified in EWASs of asthma and 33 asthma-related traits. Proximity assessment to previously identified SNPs associated with asthma 34 identified 17 unique SNPs within 1 MB of two of the 221 CpGs. EWAS in 182 individuals with 35 epigenetic data at a second timepoint (T2) identified 49 significant CpGs. EWAS Atlas enrichment analysis indicated that 4 of the 49 were previously associated with asthma or asthma-36 37 related traits. Comparing the estimates of all the significant associations identified across the two time points (271 in total) yielded a correlation of 0.81. 38 39 Conclusion: We identified 270 unique CpGs that were associated with PC scores generated from 40 35 clinical markers of asthma, either cross-sectionally or 10 years later. A strong correlation was 41 present between effect sizes at the 2 timepoints. Most associations were identified for PC7, 42 which captured blood eosinophil counts and immunoglobulin levels and many of these CpGs 43 have previous associations in earlier studies of asthma and asthma-related traits. The results point 44 to using this robust DNA methylation profile as a new, stable biomarker for asthma.

45 Introduction:

Asthma affects approximately 262 million individuals worldwide and poses a significant 46 47 health burden resulting in over 450,000 deaths annually [1, 2]. Often, the challenges associated 48 with treating and diagnosing asthma stem from the complex nature of the disease [3] due to its 49 multi-level heterogeneity with various clinical presentations, treatment responses, and disease 50 trajectories [4]. Clinically, diagnosing asthma is done through the collection of several different 51 clinical measurements and symptoms including a medical history assessment, a symptom and 52 physical assessment, lung function tests, bronchial challenge tests, response to bronchodilators, 53 and allergy testing [5]. These different diagnostic criteria, along with different individual 54 phenotype characteristics, such as age of onset, have led to the identification of subtypes of 55 asthma [6-8] and different endotypes with varying molecular underpinnings of the disease [6]. 56 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]. Two of 57 58 the largest genome-wide association studies (GWAS) of asthma were performed by Demenais et 59 al. and Ferreira et al. The GWAS of asthma by Demenais et al. in 2018 included over 140,000 60 individuals and identified five novel, independent associations and confirmed several others that 61 had previously been identified [10]. The study performed by Ferreira et al. in 2019, included 62 over 300,000 individuals, identified 123 single nucleotide polymorphisms (SNPs) significantly 63 associated with childhood asthma, 56 SNPs associated with adult-onset asthma, and 37 to be 64 associated with both [11]. Collectively, Ferreira et. al. found 28 novel, independent associations 65 between the two tested age groups [11]. Many of the associations that have been identified 66 appear to be strongly linked to immune-related mechanisms [10, 11]. Though significant genetic 67 associations were identified in both instances, the amount of total variance in relation to asthma

68 they explained was limited (3.5-5.1%) [10, 11]. These outcomes reaffirm the notion that asthma, 69 as a disease, is heavily influenced by environmental stimuli, which has led researchers to explore additional fields of study such as epigenetics and, more specifically, DNA methylation [12-14]. 70 71 DNA methylation, as a molecular mechanism in humans, involves the addition of a 72 methyl group to cytosine nucleotides, often in response to certain environmental queues [15-17]. 73 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 [16, 17]. This 74 75 entanglement with the external environment provides a basis for investigating the association of 76 DNA methylation and asthma. The first studies of asthma and DNA methylation were conducted 77 on candidate genes through techniques such as bisulfite pyrosequencing, which yielded moderate 78 success highlighting CpG sites near *IFNy* as potential mediators of asthma [18, 19]. The 79 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 80 81 childhood, allergic asthma in whole blood and nasal epithelial samples [12, 13]. Studies of 82 childhood asthma in nasal epithelial samples have identified several CpGs in both immune 83 regulatory pathways and pathways of basic cell function [20-22]. Furthermore, a small number 84 have shown to replicate in EWASs of whole blood, indicating some pan-tissue effects, though 85 this effect appears to be limited [9]. Studies in adults found CpGs in genes relating to general 86 inflammation, but the results show more variability in the CpGs that have been identified [23]. 87 Though the majority of these results point towards immune-mediated pathways, the complete 88 biological dynamic between these sites and their influence on asthmatic phenotypes has yet to be 89 investigated [24]. Additionally, little is known regarding which measurable clinical markers

90 associated with asthma may be contributing to these significantly associated CpGs, an area we91 explore here.

92 More generally, the continued identification of genetic and epigenetic variants 93 influencing asthma is valuable for personalized treatment approaches. Here, we examined the 94 genome-wide, CpG-specific methylation association with 35 different clinical markers of asthma 95 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 96 97 included individuals from families enrolled in the Netherlands Twin Register where at least one 98 member of a nuclear family (consisting of 4 or 5 individuals) was diagnosed with asthma. Blood 99 samples were collected at two timepoints, with 341 individuals having a sample collected at the 100 first timepoint in the 1990's and 233 individuals a second timepoint approximately 10 years later 101 in the 2000's [25]. Clinical markers were assessed at timepoint 1, while epigenetic data was 102 generated on DNA from blood specimens for both timepoint 1 and timepoint 2.

103 **Results:**

104 Clinical marker summary and data reduction.

105 An initial set of 39 asthma-related clinical markers were measured. Markers that 106 exhibited no variation between individuals were excluded (4 out of 39). Markers included 107 measurements of lung function/capacity, skin prick test results measuring the reaction to 108 common allergens, and others previously shown to be associated with asthma. Table 1 provides 109 descriptions of the remaining 35 clinical markers. We performed an imputation step via the R 110 package mice (v3.16.0) to supplement 12 of the 35 markers that had missing values. Distribution 111 plots of these markers before and after imputation can be found in Supplemental Figure 1. We 112 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
- 114 non-normally distributed data [26]. The distributions of the principal component scores can be

115 found in Supplemental Figure 2.

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 x 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 second (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

116

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

118 PCA investigating the multidimensional structure of the clinical marker data showed the 119 first principal component to account for 35.27% of the variance. Principal component (PC) 1 was 120 largely associated with methacholine challenge values, other measurements of lung 121 capacity/functionality, and lung ailments such as coughing and wheezing. PCs 2, 3, and 4 122 accounted for 19.54%, 12.46%, and 9.75% of the variance, respectively, and were also mostly 123 associated with functional lung measurements. PCs 5 and 6 accounted for 8.39% and 6.05% of 124 the variance, respectively, and showed strong association with the FEV1/FVC ratio, also called 125 modified Tiffeneau-Pinelli index. PC7 accounted for 4.28% of the variance and showed very 126 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 127 128 response measurements (Figure 1). Supplemental Table 1 contains the raw eigen values and 129 percentage of variance explained for each PC.



130



134 Timepoint 1 EWAS via 10 Generated PCs

135 Epigenome-wide association studies of each of the ten principal components yielded

- 136 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,
- 138 cg07329820, was associated with two PCs (PCs 9 and 10), resulting in a total of 221 unique
- 139 CpGs that were significantly associated. Table 2 provides a summary of the covariate frequencies

140 for the sample population. Table 3 summarizes the number of CpGs identified in each EWAS and

141 Supplemental Figure 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.

	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%
Μ	ledication Usage	e
Variable	Freq.	Percent
General Lung Medication	48	15.0%
Betamimetica	37	11.6%
Inhaled Corticosteroids	26	8.2%

145 **Table 2.** Sample population covariate frequencies.

	Timepoint 1	Timepoint 2				
РС	Genome-wide Significant CpGs	Samples	РС	Genome-wide Significant CpGs	Samples	
PC1	1	319	PC1	5	182	
PC2	4	319	PC2	6	182	
PC3	2	319	PC3	1	182	
PC4	1	319	PC4	0	182	
PC5	2	319	PC5	3	182	
PC6	0	319	PC6	1	182	
PC7	204	315	PC7	8	181	
PC8	1	319	PC8	12	182	
PC9	2	319	PC9	2	182	
PC10	5	319	PC10	11	182	
Total	222		Total	49		

Table 3. A summary of the genome-wide significant CpGs identified in each EWAS for timepoint 1 and timepoint 2.

148	By far, the largest number of CpGs were associated with PC7 (which mostly captures
149	variance from eosinophil counts and immunoglobulin levels). These 204 CpGs, largely, are
150	composed of CpGs with a high average methylation level and almost exclusively show negative
151	effect sizes, such that higher scores on PC7 correlate with lower methylation levels. The PC10
152	EWAS identified associations with 5 CpGs, followed by the PC2 EWAS with 4 significantly
153	associated CpGs. EWASs for PCs 3, 5, and 9 showed significant associations with 2 CpGs.
154	EWASs for PCs 1, 4, and 8 were significantly associated with 1 CpG, while the EWAS for PC6
155	was the only association study that did not identify any significantly associated CpGs.
156	An enrichment analysis via the online EWAS Atlas
157	(https://ngdc.cncb.ac.cn/ewas/atlas/index, accessed June 11, 2024) for the 221 unique CpGs was
158	performed. This analysis identified that 71 (32.1%) and 173 (78.3%) of these CpGs had
159	previously been identified to associate with allergic asthma and fractional exhaled nitric oxide
160	(FENO), respectively (shown in Figure 2). Further, asthma related phenotypes such as childhood
161	asthma, allergic sensitization, atopy, serum IgE levels, and others were shown to be previously
162	associated with CpGs identified in our analysis. In total, 190 of the 221 CpGs (86.0%) were
163	shown to have a previously identified association with asthma or some asthma marker, leaving
164	31 novel, unique associations. This enrichment analysis also highlighted significant overlap with
165	the FoxO signalling pathway (7 CpGs previously associated). Enrichment analysis restricting to
166	the 204 CpGs associated with PC7 recapitulated these findings (Supplemental Figure 4),
167	showing that all of the previously associated CpGs were identified via this PC7 EWAS. We then
168	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-

170 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 Figure 5).



173

Figure 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.

178 Analysis of overlap with previously identified SNPs

An investigation of proximity with the 221 unique significantly associated CpGs was

180 performed to identify SNPs that have been previously identified by Demenais et al. to associate

- 181 with asthma (p< 5 x 10^{-8} , N= 892 SNPs). This assessment was performed by searching for SNPs
- 182 within 1 MB upstream and downstream of each CpG. We identified a total of 17 unique SNPs
- 183 (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 [27].

189 Timepoint 2 EWAS

190 Of the initial 319 individuals who were included at the first timepoint, 182 individuals 191 had a second whole blood sample collected approximately 10 years later. The DNA from these 192 blood samples were also profiled on Illumina EPIC v1 methylation arrays (but no detailed 193 asthma phenotyping was performed at follow-up). We repeated the EWASs for the ten PCs with 194 DNA methylation data at T2. The EWASs of the ten PCs with DNA methylation data from 195 samples collected ten years after asthma phenotyping identified 49 unique CpGs at epigenome-196 wide significance. While none of the CpGs identified at T1 were significant at T2, three CpGs 197 identified at T2 were within 100 kb of a significant CpG identified at timepoint 1 (Supplemental 198 Table 8). Furthermore, we assessed the correlation (Pearson's) of the estimates for all 271 199 significant associations (222 at T1, 49 at T2) across the two time points, and we identified a 200 strong correlation (r = 0.82). Table 3 outlines the number of CpGs identified in each time point 2 201 EWAS and Supplemental Figure 6 contains the Manhattan plots for these EWASs. Supplemental 202 Table 6 contains all of the 49 significantly associated CpGs identified at T2, and Supplemental 203 Table 7 states the inflation values for each of these 10 EWASs.



Figure 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.

209 Discussion:

204

210 Asthma is a complex chronic disease that has been shown to be influenced by multiple 211 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 35 212 clinical markers of asthma. Of these, 190 have previously been identified to associate with 213 214 asthma and asthma-related phenotypes, while the other 31 CpGs are novel findings. Further, two 215 of these CpGs reside within close proximity (1 MB) to 17 SNPs previously identified in a large 216 GWAS of asthma (source). We then performed a longitudinal assessment using methylation data 217 from whole blood samples collected approximately 10 years and identified 49 unique

218	significantly associated CpGs. Finally, we yielded a correlation of the estimates across the two
219	timepoints of 0.82 from the 271 (222 T1 + 49 T2) significant associations.

220 Historically, asthma as a disease has been revealed to be extremely challenging to 221 identify the root cause of and treat effectively [28]. The disease heterogeneity and mixed 222 involvement of environmental stimuli and the genome have largely contributed to this difficulty. 223 To avoid the complications that can come with this, we utilized measured clinical markers of 224 asthma in young adults and their middle-aged parents from families with a history of asthma 225 burden where at least one individual within the family had been diagnosed with asthma in some 226 capacity. For each individual, an extensive number of clinical markers related to asthma were 227 captured. This comprehensive dataset, while immensely valuable with its depth, is too 228 overwhelming to glean the valuable nuances through individual assessments of each specific 229 marker. Additionally, some markers that were collected lacked a normal distribution, which 230 could reduce their performance in a traditional EWAS. For this reason, we utilized PCA to 231 simplify the 35 different variables, and their accompanying variance, into 10 PCs with each 232 successive PC capturing a lesser and different portion of the overall variance within the data set. 233 This method reduced the total number of variables to investigate from 35 to 10, making 234 comparisons of it with a rich DNA methylation dataset much more feasible. Additionally, the 235 PCs maintained a much more normal distribution, making them ideal for EWAS. This method of 236 phenotype data reduction and handling has been previously implemented in a GWAS of rice 237 variants, where GWASs were successfully performed using the PC scores from multiple 238 phenotype variables [29]. To our knowledge, this is the first instance of implementing it for 239 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.

247 A breakdown of the identified CpGs from each PC shows some intriguing details. PC7 248 had the greatest number of CpG associations with 204. While this PC accounts for only 4.3% of 249 the total clinical marker variance, it almost exclusively represents the variance associated with 250 blood eosinophil counts and blood immunoglobulin levels. This large number of associations 251 could be indicative of the eosinophilic inflammatory response typically seen in allergic asthma. 252 Previous literature has shown that methylation experiments in whole blood may capture more of 253 the Th2-mediated immune response due to the nature of the cell population, making it the ideal 254 sample source for identifying this type of association [30-32]. It is critical to note that because 255 the proportions of cells themselves in a sample can drastically influence the measurement of 256 methylation, we made sure to incorporate cell proportion estimates as a covariate in our analyses. 257 Without this inclusion, it could be argued that the findings identified here are only due to 258 differences in cellular composition. With this correction in mind, it may be the case that the 259 differential methylation identified here is something global, across multiple immune cell types, 260 either in response to or to promote the eosinophilic inflammatory pathway. A query of the EWAS 261 Atlas of these 204 CpGs showed that, almost exclusively, the CpGs we identified have a 262 previous association with asthma and measurements of asthma (Supplemental Figure 4). In total

190 of the 204 CpGs were shown to have 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.

269 An additional investigation into these 204 CpGs via the eFORGE online database 270 identified multiple significantly associated tissue-specific regulatory elements (Supplemental 271 Figure 5). The greatest number of associations were from subpopulations of white blood cells in 272 enhancers and weak transcriptional activators. This is to be expected considering our results were 273 generated from samples of whole blood. Many of the subpopulations identified here are 274 lymphocytes, potentially indicating that the methylation profiles of these adaptive immune cells 275 may be altered due to the changes in eosinophil counts and immunoglobulin levels. The tissue 276 group with the next largest number of significant associations was the digestive tract, with six 277 significant associations to various subpopulations of cells. These associations could hint at the 278 conserved role that all mucosal and epithelial barriers play in pathogen protection and the body's 279 response to the external environment.

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 [33,

34]. It has been previously shown that FoxO1 modulates IL-9 generating Th9 cells [35]. Further,
overexpression of IL-9 in the lungs can result eosinophilic inflammatory infiltration and mucus
secretion [35, 36]. Additionally, FoxO3 has been found to be expressed in airway epithelium
playing a critical role in controlling the innate immune response to airway infections [36]. It
could be the case, 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.

293 Principal component 1 identified a single significantly associated CpG, cg18182148. 294 While this principal component does capture the largest amount of variance (35.3%), the 295 variance it captures is largely from one specific subset of measurements, most notably those 296 measuring general lung functionality such as vital capacity, peak flow, forced expiratory volume, 297 and methacholine challenge response. This is continued with PC2, PC3, and PC4 which were 298 significantly associated with 4, 2, and 1 CpGs, respectively. These PCs, like PC1, capture a 299 significant amount of variance from these same lung function measurements. This may indicate 300 that these measurements of lung function, though important in characterizing symptoms of 301 asthma, may have limited identifiable associations with DNA methylation. The specific CpG 302 significantly associated with PC1, cg18182148, is the only CpG of these identified here related 303 to immune regulation. It has not been previously shown to associate with asthma phenotypes, 304 though it has been previously identified in studies of various cancers of the prostate, liver, and 305 colon (via the EWAS Atlas) [37-40]. These associations are likely due to its placement within the 306 transcription start site (TSS) of GFI-1, a strong oncogenic and hematopoietic regulator. GFI-1 307 plays a critical role in lymphoid differentiation, which could be the reason for its association here 308 due to the large influence the immune system can play on lung functionality and, more generally,

309 asthma as a whole [41]. While the use of a whole blood sample, as done here, is logical when 310 investigating an immune-mediated illness such as asthma, an investigation of these lung function 311 measurements in association with DNA methylation measured via a lung epithelial sample may 312 highlight additional findings not seen here and provide more information on the local cellular 313 environment responsible for the specific differences in lung function. 314 Principal components 5 and 6 capture variance associated with bronchial hyperreactivity, 315 coughing, reversibility (of lung function following the administering of ventolin), and Tiffenau 316 Index. PC5 was significantly associated with 2 CpGs, while PC6 did not have any significant 317 associations. The two CpGs identified here, cg07454584 and cg18561513, have not been 318 previously shown to have any associations with asthma or other immune-mediated illnesses (via 319 the EWAS Atlas). Like the measurements captured in PCs 1-4, these asthma markers, while 320 important for asthma diagnosis and the understanding of symptoms, may not have large 321 quantities of associations with DNA methylation measured via whole blood. 322 Principal components 8, 9, and 10 contain much of the variance from the numerous 323 allergen tests included in this study (containing both skin prick test results and serum specific 324 IgE measurements). These PCs were significantly associated with one, two, and 5 CpGs, 325 respectively. Of the 8 CpGs identified here, all but one has not been previously shown to 326 associate with chronic illnesses or immune-mediated diseases. Cg14161241, which was 327 significantly associated with PC10, is the lone exception with previous associations with obesity 328 and type 2 diabetes [42]. Like many of the other markers represented via PCs 1-6, the effect on 329 DNA methylation measurable via a whole blood sample is likely limited with these markers. 330 Due to the well-documented genetic contributions to asthma, we next wanted to compare 331 our findings within the methylome to those previously discovered in the genome via large

332 GWASs [10, 43]. Utilizing the findings of Demenais et al., we searched for significantly 333 associated SNPs within 1 MB of the 221 unique CpGs that we identified. Of the significantly associated SNPs (p< 5 x 10^{-8} , N= 892 SNPs), we found 17 (1.9%) to be within 1 MB of 2 of the 334 335 221 CpGs that we identified (cg02046836, cg23515090). Interestingly, neither of these two 336 CpGs have been previously identified as an mQTL [27]. The proximity of these SNPs to the 337 identified CpGs could be caused from multiple factors outside of known standard mQTL 338 associations. It could be the case that these CpGs do have some genetic influence acting on their 339 methylation status that has not yet been characterized. Additionally, it is certainly possible that 340 the SNPs and CpGs being within close proximity to one another is simply a random occurrence, 341 and their presence and downstream effects are independent of one another. To summarize, we 342 found limited overlap of our EWAS signal with top loci from a previously performed GWAS of 343 asthma and that these significant CpGs nearby these GWAS SNPs were not directly affected by 344 mQTLs, which suggests that our EWAS largely captures independent epigenetic signal at loci 345 that also harbor small amounts of genetic variants that influence genetic susceptibility to asthma. 346 Thus far, we have shown significant associations with clinical markers of asthma from a 347 singular, cross-sectional timepoint that corroborates other previous findings from similar studies. 348 However, a potential increase in our understanding of asthma may lie in investigating the disease 349 to some longitudinal capacity. Two previous studies in the Isle of Wight Birth Cohort (IOWBC) 350 were conducted in 2022, which investigated the association of DNA methylation with asthma 351 acquisition across adolescence and adulthood [44, 45]. In each of these studies, several CpGs, 352 with some residing in immune regulatory genes, were identified to be associated with asthma 353 acquisition [44, 45]. The findings from these two studies highlight promising insight into the 354 potential longitudinal persistence of DNA methylation signatures associated with asthma, though

additional studies are needed to investigate this persistence later in life (adulthood), in additionalcohorts, and how they relate to specific measurements of asthma.

357 After our initial findings, with the availability of an additional sample collected 358 significantly later for many of the individuals, we were interested in investigating the 359 longitudinal persistence of the DNA methylation signature that was identified to associate with 360 asthma markers at the first timepoint. This analysis was based on the same PC data from the 361 original 35 clinical markers measured at baseline and DNA methylation at follow-up from a sample collected ~10 years later. The results from these additional 10 EWASs showed very little 362 363 overlap with the findings from the initial assessment. In total, these 10 EWASs identified a total 364 of 49 significantly associated CpGs (of which none were replicated across multiple PCs). No 365 CpGs were replicated across the two timepoints, however, when looking within a 100 kb window 366 of proximity, we found that 3 CpGs identified at timepoint 2 (found via PCs 2, 3, and 7) were 367 within 100 kb of 4 CpGs identified at T1 (all found via PC7). The low direct overlap between the two timepoints could be due to the methylation signature associating with asthma diminishing 368 369 over time, potentially as individuals' asthma status and severity changes. A query of the EWAS 370 Atlas, however, using these 49 CpGs identified at T2 showed that 3 and 1 CpGs have been 371 previously associated with FENO and allergic sensitization, respectively (Figure 3). This 372 highlights that there may be some longitudinal persistence of an asthmatic epigenetic signature. 373 Further, we identified, when comparing the estimates of all 271 significant associations (222 + 374 49) across the two timepoints, a strong correlation of 0.82 adding to the notion that some long-375 term persistence of the asthma associated methylation signature may be occurring. This 376 consistency of the EWAS estimates indicates that replication of specific CpGs could occur with 377 an increased sample size. 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 380 381 included in this study are of European descent and from the Netherlands. Asthma prevalence can 382 differ greatly depending on the local environment of the individual and their ethnicity. Including 383 a diverse population could improve the generality of the findings to other groups of individuals. Additionally, the results generated from the second time point contained a significantly reduced 384 385 number of individuals due to sample availability, which could contribute to the differences 386 reported. We also did not have clinical marker measurements available from the second 387 timepoint, which could have shed some insight on the persistence of the asthma symptoms and 388 severity at that time. Overall, increasing the number of individuals at both timepoints using 389 multiple individuals from varying backgrounds would be the most optimal strategy and offer the 390 most comprehensive assessment of asthma.

391 Conclusion:

392 Asthma, along with many other chronic, immune-mediated diseases, is challenging to 393 diagnose, classify, and treat. Here, we contribute a comprehensive assessment of the relationship 394 between DNA methylation and a collection of clinical asthma markers. We identified 270 unique, 395 significant CpGs to be associated with principal component scores based on clinical asthma 396 markers. The large majority of these CpGs were associated with PC7, which represented 397 eosinophil counts and blood immunoglobulin levels and 190 were previously associated with 398 asthma, asthma markers, and other allergic illnesses. We also investigated the proximity of these 399 221 identified unique CpGs to previously identified SNPs that associate with asthma. We found 400 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.

407 Methods:

408 Participants

409 The study participants and the enrollment process have been previously described [46-410 48]. Participants came from twin families (young adult twins and their parents) from the 411 Netherlands [48]. To be eligible for the study, at least one member of the family had to report to 412 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). 413 414 During a visit to the Vrije Universiteit Medical Center (VUMC), blood samples and over 40 415 biological and clinical measures of asthma were collected from 425 persons [49]. A complete list 416 of the clinical markers that were collected can be found in Table 1. A subset of 233 persons 417 participated in the NTR Biobank project between 2004 and 2008 in which blood and buccal 418 samples were collected [25]. DNA was extracted at the time of sample collection or immediately 419 prior to the assessment of methylation [50]. All samples were simultaneously measured on the 420 Illumina Infinium EPIC v1 methylation array. In total, 375 of the 425 individuals had a sample 421 remaining for analysis, and 341 of these had good quality Illumina blood EPIC Array data from 422 at least the first time point, and 232 individuals had blood Illumina EPIC Array data from two 423 time points.

424 The study was approved by the Central Ethics Committee on Research Involving Human 425 Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board 426 certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under 427 Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03-180). 428 Sample collection and DNA extraction 429 The procedures of whole blood sample collection have been previously described [25, 46, 430 51]. 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 431 432 from whole blood samples using the Zymo Quick DNA mini-prep Kit [50]. Genomic material 433 was quantified via the Invitrogen Qubit Broad-Range Fluorescent Assay, and sample purity was 434 assessed using standard absorbance metrics via a SpectraMax microplate reader. 435 Genotyping The genotyping process has been previously described [47, 52]. Briefly, genotype data 436 437 was generated via the Illumina Infinium Global Screening Array. All sample data quality was 438 assessed vigorously, and the refined genotype data was used to confirm familial relationships and 439 sample sex. In total, 20 individuals were omitted for analysis due to mismatched familial 440 identities. 441 DNA Methylation Assessment

442 DNA bisulfite conversion was performed utilizing the Zymo EZ-96 DNA Methylation
443 Kit [53]. DNA methylation was assessed using the Illumina Infinium EPIC v1 DNA Methylation
444 Array on all samples at the Avera Genetics Laboratory [54]. The samples were fully randomized
445 across arrays.

446 DNA Methylation Data Quality Control

447 Quality control has been described in detail previously [47]. In brief, DNA methylation data quality was performed using two bioinformatic tools. First, DNA methylation data was 448 449 preliminarily assessed using the Illumina GenomeStudio 1.0 software. Second, methylation data 450 quality control and normalization were conducted using a pipeline developed by the Biobank-451 based Integrative Omics Study (BIOS) Consortium in R (version 4.4.0). Sample quality was first 452 assessed using the R package MethylAid (v1.38.0) to omit any outliers in sample performance 453 (default thresholds) [55]. Array probe filtering and functional normalization were performed 454 using the R package DNAmArray (v2.2.0) [56]. The R package omicsPrint (v.1.24.0) was 455 utilized to identify sample genotypes based on methylation probes to verify sample relationships 456 gathered from genotype data via the GSA [57]. The functions getSex from DNAmArray and the 457 R package meffil (v1.3.8) were used to identify and confirm sample sex based on X chromosome 458 methylation pattern [58]. The following probe filters were applied: Probes were set to missing 459 (NA) in a sample if they had an intensity value of exactly zero, detection P-value > 0.01, or bead 460 count < 3. DNAmArray will also remove any probes that show a success rate below 0.95 across 461 all samples. In total, this filtering process reduced our number of individuals to 341 (from the 462 original 375). Finally, previous studies performed by Zhou et al. in 2017 identified polymorphic 463 and cross-reactive probes that are included on Illumina methylation array platforms [59]. 464 Generally, they recommend an omitting of approximately 100,000 probes on the EPIC 465 methylation array. This removal process was carried out using the probemasking() function via 466 DNAmArray. Following these steps, a total of 759,263 methylation sites were included in this 467 study out of the total 865,859 possible CpGs. Only autosomal methylation sites were considered 468 for downstream analyses, which left a total of 742,442 CpGs included in the final analyses.

469 Finally, we calculated cellular proportions for all samples using the IDOL whole blood reference470 library [60].

471 Covariates

472 Several covariates were utilized in our epigenome-wide association studies. For each 473 sample at each timepoint, covariate data was collected at the same time. We included age, sex, 474 array row, sample plate (dummy-coded), smoking status, asthma/lung medication use, and 475 cellular proportions. For smoking status, individuals were classified into three groups (codes as 0 = Never smoker, 1 = former smoker, 2 = current smoker). Cellular proportions (natural killer 476 477 cells, monocytes, B cells, CD4+ T cells, CD8+T cells) for each of the whole blood samples were 478 estimated via the methylation data from that sample. The usage of specific and general asthma 479 medications (inhaled corticosteroids (0=no, 1=yes), betamimetica (0=no, 1=yes), general lung 480 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 481 482 used at both time points. Due to limited variability in medication use at the second timepoint, 483 only a single "general lung medication use" variable was used as a covariate for T2. 484 Methylation Data Annotation 485 Genomic annotations were gathered from the EPIC manifest file that is provided by 486 Illumina (MethylationEPIC v-1-0 B5.csv): locations of CpG islands, ENCODE DNase I 487 Hypersensitive sites (DHSs), ENCODE transcription factor binding sites (TFBSs), open 488 chromatin, FANTOM4 enhancers and FANTOM5 enhancers, etc. Genome build 37 coordinates 489 were utilized for all the analyses.

490 Asthma phenotyping

491	Of the total list of available clinical markers, exclusions for markers were made if there
492	was no variation amongst participants. In total, 35 clinical markers were selected to be included
493	in the principal component analysis. If an individual themselves had more than 6 missing values
494	they were excluded from the analysis as well, which resulted in removing 4 individuals leaving a
495	total of 319 to include. 10 of the 35 markers contained missing values following these exclusion
496	processes. For these missing values, the R package mice (v3.16.0) was utilized to impute the
497	missing data. The default settings were utilized, 5 imputations per marker, predictive mean
498	matching (pmm) for continuous traits, logistic regression (logreg) for categorical traits.
499	Distribution plots for these markers prior to and after imputation can be found in Supplemental
500	Figure 1.
501	Principal component analysis
502	Following imputation, the data was normalized using the scale function in R, which
503	performs a z-score transformation for each marker. We then performed principal component
504	analysis (PCA) including 10 principal components (PCs) on our 35 clinical markers that were
505	included. PCA was performed using the robpca() r function via the R package rospca (v1.0.4),
506	which is suited for non-normally distributed data [26, 61]. A scree plot of the eigen values for
507	each of the 10 PCs can be found in Supplemental Figure 7. Outliers for each PC were identified
508	via the R package ewaff (v0.0.2) using the ewaff.handle.outliers function with the following
509	settings: method set to iqr, iqr range set to 3. Outliers were found in PCs 7-10. Principal
510	component 7 contained four outliers, which were removed prior to further analyses ($n = 315$).
511	Principal components 8-10 each contained several outliers. Due to the nature of the clinical
512	marker data and the way the PCA captured the variance, these PCs largely captured variance
513	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,

515 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

517 dichotomized values for PCs 8, 9, and 10 were then used for the downstream analyses.

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

519 EWAS

520 Ten epigenome-wide association studies (EWAS) were then performed via the PC scores 521 for each individual for each of these 10 PCs using DNA methylation beta value as the outcome 522 and the PC scores for each individual as the predictor. Due to the related nature of the samples 523 utilized here, each EWAS was conducted via the r package gee (v4.13-25) using a generalized 524 estimating equations (GEE) model that corrected for the correlation structure in families. The 525 following settings were utilized: Gaussian link function, 100 iterations, and the "exchangeable" 526 option to correct for the familial structure. Additionally, each EWAS included age, sex, 527 methylation array row, bisulfite sample plate (dummy-coded), smoking status (codes as 0 =528 Never smoker, 1 = former smoker, 2 = current smoker), estimated cellular proportions (natural 529 killer cells, monocytes, B cells, CD4+ T cells, CD8+T cells), and lung medication status via 530 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 531 532 account for multiple testing to identify significantly associated CpGs. 533 Enrichment analysis was performed in the EWAS Atlas with the EWAS Toolkit, which 534 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 535

536 database of known CpG associations to cell-type specific regulatory elements. This analysis was

537	performed using the "Consolidated Roadmap Epigenomics - Chromatin - All 15-state marks"
538	data set and the standard settings. Next, we cross-analyzed our significantly associated CpGs
539	with significantly associated SNPs (p< 5 x 10^{-8} , N= 892 SNPs) identified by Demenais et al in a
540	large meta-analysis [10], utilizing leave-one-out GWAS summary statistics without NTR
541	participants. As has been previously utilized in other studies, a 1 MB range around each CpG
542	was utilized as a cutoff point to identify SNPs within close proximity [62]. Finally, we
543	investigated the CpGs that were found to be within close proximity to a known significantly
544	associated SNP for any previously identified mQTL association by querying the BBMRI mQTL
545	database (based on a previous study by Bonder et. Al.) [27].
546	Of the 319 individuals included in timepoint 1, 182 had a follow-up blood sample with
547	methylation data available for subsequent analyses at a later time point. For this second
548	timepoint, an additional 10 EWASs were performed using the same PC scores generated from the
549	timepoint 1 clinical marker data to assess longitudinal persistence of the DNA methylation
550	signature associated with these asthma markers. The EWASs were performed with largely the
551	same models as at timepoint 1 samples (stated above) but utilizing covariate data from time point
552	2. Due to the reduction of individuals and a lack of unique medication covariate combinations,
553	asthma medication status was reduced to a Bayesian variable grouping individuals on the basis of
554	using any one of the three listed medications. This single covariate capturing asthma medication
555	status was then used for these timepoint 2 EWASs. An additional longitudinal comparison was
556	made looking for CpGs in timepoint 1 that were within 100 KB (upstream and downstream) of a
557	CpG identified in timepoint 2. Similar to timepoint 1, following this the timepoint 2 CpGs were
558	grouped and a query of the EWAS Atlas was performed looking for previously identified
559	associations.

560 Funding:

- 561 We acknowledge funding from the Amsterdam Public Health Institute (personalized medicine
- 562 innovation grant); the Avera Institute, Sioux Falls (USA); the National Institutes of Health (NIH
- 563 R01 HD042157-01A1, MH081802, and Grand Opportunity grants: 1RC2 MH089951 and 1RC2
- 564 MH089995); the Netherlands Organization for Scientific Research (NWO): Netherlands Twin
- 565 Registry Repository: researching the interplay between genome and environment (NWO-Groot
- 566 480-15-001/674); Biobanking and Biomolecular Research Infrastructure (BBMRI–NL, NWO
- 567 184.033.111); the BBRMI-NL-financed BIOS Consortium (NWO 184.021.007), Geno-
- 568 type/phenotype database for behavior genetic and genetic epidemiological studies (ZonMw
- 569 Middelgroot 911-09-032); and large-scale infrastructures X-Omics (184.034.019). DIB
- 570 acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635). AJVA
- acknowledges the University of South Dakota Wesley H. Parke Research Award.

572 Data Availability Statement:

- 573 All NTR data can be requested from bona fide researchers (https://ntr-data-request.psy.vu.nl/).
- 574 Because of the consent given by the study participants, the data cannot be made publicly
- 575 available. The pipeline for DNA methylation–array analysis developed by the Biobank-based

576 Integrative Omics Study (BIOS) consortium is available at

577 https://molepi.github.io/DNAmArray workflow/ (https://doi.org/10.5281/zenodo.3355292).

578 Acknowledgments:

- 579 We would like to acknowledge all the participants in the Netherlands Twin Register for their
- 580 willingness to participate in this study. We would also like to acknowledge the BIOS
- 581 Consortium. A list of the contributors to this consortium and other additional details can be found
- 582 at https://www.bbmri.nl/bios-consortium.

583 Conflicts of Interest:

584 The authors declare no conflict of interest.

585 Supplementary Tables and Figures:

586

РС	Eigen Value	% Variance	Cumulative
			<u>%</u> 0
PC1	4.4832	35.27%	35.27%
PC2	2.4844	19.54%	54.81%
PC3	1.5834	12.46%	67.26%
PC4	1.2396	9.75%	77.01%
PC5	1.0671	8.39%	85.41%
PC6	0.7696	6.05%	91.46%
PC7	0.5435	4.28%	95.74%
PC8	0.2794	2.20%	97.94%
PC9	0.1555	1.22%	99.16%
PC10	0.1071	0.84%	100.00%
Total	12.7125		

587

588 Supplementary Table 1. Table showing PC eigen values and % of total variance captured.589

PC	Inflation
PC1	0.999737038
PC2	0.894699702
PC3	0.861547269
PC4	1.034053431
PC5	1.021476296
PC6	0.749147555
PC7	0.805103148
PC8	1.003658406
PC9	0.843756081
PC10	0.848921965

590

Supplementary Table 2. EWAS results for time point 1 with inflation values.

592

593 Supplementary Table 3. Independent attachment with a complete list of significant CpGs for all
594 10 timepoint 1 EWASs.

595

596 Supplementary Table 4. Independent attachment with complete summary statistics for each of597 the 10 EWASs.

598														
snp_chr	snp_pos	ref	alt	reffrq	info	rs	snp_pval	effalt	Ν	cpg	cpg_chr	cpg_mapinfo	cpg_estimate	cpg_pval
chr11	76270683	А	G	0.645075	1	11:76270683	1.49E-10	0.08471681	67341	cg02046836	11	75550064	-0.01168842	3.16E-11
chr11	76299194	G	Т	0.512605	1	11:76299194	2.90E-15	0.1050031	67341	cg02046836	11	75550064	-0.01168842	3.16E-11
chr11	76301316	С	Т	0.617355	1	11:76301316	3.23E-12	0.09592278	67341	cg02046836	11	75550064	-0.01168842	3.16E-11
chr11	76301375	С	Т	0.617005	1	11:76301375	8.40E-12	0.0961933	67341	cg02046836	11	75550064	-0.01168842	3.16E-11
chr15	67441997	Т	С	0.698821	1	15:67441997	3.03E-14	0.1094348	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67442596	С	Т	0.763772	1	15:67442596	8.81E-16	0.117968	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67444747	С	Т	0.505252	1	15:67444747	1.73E-12	0.08952885	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67446785	G	А	0.503151	1	15:67446785	9.76E-13	- 0.08880695	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67447452	С	Т	0.503093	1	15:67447452	1.56E-12	- 0.08800783	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67448899	А	G	0.697946	1	15:67448899	5.58E-14	0.1036299	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67449660	А	G	0.697946	1	15:67449660	2.47E-14	0.1034242	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67450305	А	G	0.762722	1	15:67450305	2.85E-15	0.1157078	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67450893	А	G	0.697946	1	15:67450893	5.83E-14	0.1035278	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67458152	G	А	0.700222	1	15:67458152	2.04E-13	0.1004536	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67464013	А	G	0.720472	1	15:67464013	5.28E-12	0.09549112	67341	cg23515090	15	68092384	- 0.007250445	5.83E-09
chr15	67467541	Т	С	0.510621	1	15:67467541	1.88E-11	0.08491053	67341	cg23515090	15	68092384	0.007250445	5.83E-09
chr15	67468285	А	G	0.740721	1	15:67468285	3.76E-14	0.1190431	67341	cg23515090	15	68092384	0.007250445	5.83E-09

599

600 Supplementary Table 5. Summary table of the GWAS SNPs identified by Demenais et. al. to

reside within 1 MB of a CpG identified at timepoint 1.

602

603

Supplementary Table 6. Independent attachment with a complete list of significant CpGs for all
 10 timepoint 2 EWASs.

606

PC	Inflation
PC1	1.027943459
PC2	0.979550307
PC3	0.999929682
PC4	1.000268796
PC5	0.775029681
PC6	0.949333881
PC7	0.952948417
PC8	1.004404304
PC9	0.999975322
PC10	1.02109527

607

608 **Supplementary Table 7.** EWAS results for time point 2 with inflation values.

609 610

Genome_Build	TP1_IlmnID	TP1_CHR	TP1_MAPINFO	TP1_estimate	TP1_pval	TP1_PC	TP2_IlmnID	TP2_CHR	TP2_MAPINFO	TP2_estimate	TP2_pval	TP2_PC
37	cg01043901	13	43061678	-0.0061855	1.23E-08	PC7	cg04217496	13	43119835	0.003371794	4.51E-08	PC7
37	cg04983687	16	88558223	-0.0269112	1.64E-12	PC7	cg09247061	16	88571844	0.005722604	2.14E-08	PC2
37	cg08940169	16	88540241	-0.0130795	2.54E-09	PC7	cg09247061	16	88571844	0.005722604	2.14E-08	PC2
37	cg12077754	2	75089669	-0.0128193	2.36E-08	PC7	cg19359099	2	75060770	-0.000276636	1.97E-09	PC3

611

- 612 Supplementary Table 8. The CpGs from timepoint 1 that reside within 100 kb of a CpG
- 613 identified at timepoint 2.
- 614

	Survey 1	Survey 2
No	207 (65.1%)	219 (61.7%)
Yes	111 (34.9%)	136 (38.3%)
Total	318	355
Response Rate (%)	74.8%	83.5%

615

616 Supplementary Table 9. Questionnaire responses for two surveys asking about asthma status

- 617 prior to study enrollment for the initial 425 individuals.
- 618

Asthma Survey 1 and 2 Overlap					
		Survey 2		Total	
		no	yes		
Survey 1	no	151	21	172	
	yes	13	86	99	
Total		164	107	271	

619

620 Supplementary Table 10. Comparison of survey responses from individuals who responded to

621 both surveys.

622

РС	Non-outliers	Outliers
PC8	265	54
PC9	266	53
PC10	272	47

623

- 624 Supplementary Table 11. Frequencies of outliers for principal components 8, 9, and 10. Outlier
- 625 status was determined using thresholds of $Q1 (3 \times IQR)$ and $Q3 + (3 \times IQR)$.



628

Supplementary Figure 1. Distribution plots of the imputed clinical markers of asthma prior to (blue) and after imputation (yellow).



631 632 Supplementary Figure 2. Distribution plots of the individual principal component scores for





634 635 Supplementary Figure 3. EWAS results calculated the 10 different sets of PC scores from the first timepoint. The red line indicates the significance threshold set via a Bonferroni correction. 636 637





640 Supplementary Figure 4. EWAS atlas enrichment analysis of the 204 CpGs identified at TP1

- 641 from PC7. The EWAS Atlas job can be found via this job ID:
- 642 2a37ab12d2785eaa0d564e062cff5f96.
- 643





Supplementary Figure 6. EWAS results calculated the 10 different sets of PC scores from the

- second timepoint. The red line indicates the significance threshold set via a Bonferronicorrection.
- 652



653 654

655 Supplementary Figure 7. Scree plot showing the raw eigen values for each of the 10 principal

656 components included.

657

660 661		References
662 663	1.	Asthma. 2023 [cited 2024 02/26/2024]; Available from: <u>https://www.who.int/news-</u> room/fact-
664		sheets/detail/asthma#:~:text=Inflammation%20and%20narrowing%20of%20the.and%20
665		caused%20455%20000%20deaths.
666	2.	Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a
667		systematic analysis for the Global Burden of Disease Study 2019. Lancet, 2020.
668		396 (10258): p. 1204-1222.
669	3.	Jones, T.L., D.M. Neville, and A.J. Chauhan, <i>Diagnosis and treatment of severe asthma:</i>
670		a phenotype-based approach. Clin Med (Lond), 2018. 18(Suppl 2): p. s36-s40.
671	4.	Svenningsen, S. and P. Nair, Asthma Endotypes and an Overview of Targeted Therapy for
672	_	Asthma. Frontiers in Medicine, 2017. 4.
673	5.	Gans, M.D. and T. Gavrilova, Understanding the immunology of asthma:
674		Pathophysiology, biomarkers, and treatments for asthma endotypes. Paediatr Respir Rev,
675	(2020. 36: p. 118-12/.
6/6	6.	Kuruvilla, M.E., F.E. Lee, and G.B. Lee, Understanding Asthma Phenotypes, Endotypes,
677	7	and Mechanisms of Disease. Clin Rev Allergy Immunol, 2019. 56 (2): p. 219-233.
678	1.	Lotvall, J., et al., Asinma enabypes: a new approach to classification of disease entities
679	0	Wunin the astrima synarome. J Allergy Clin Infinunol, 2011. 127(2): p. 555-60.
080 691	0.	Rauf, R. and G. Chupp, Phenolypes and endolypes of dault dsinma. Moving loward
601	0	Ntentai D et al. Consting and Enjoynetics in Asthurg. Int I Mal Sci. 2021. 22(5)
602	9. 10	Nonisi, F., et al., Genetics and Epigenetics in Astrimu. Int J Moi Sci, 2021. 22(5).
684	10.	colocalize with immune-cell enhancer marks. Nat Genet, 2018. 50 (1): p. 42-53.
685	11.	Ferreira, M.A.R., et al., Genetic Architectures of Childhood- and Adult-Onset Asthma Are
686		Partly Distinct. Am J Hum Genet, 2019. 104(4): p. 665-684.
687	12.	Yang, I.V., et al., DNA methylation and childhood asthma in the inner city. J Allergy Clin
688		Immunol, 2015. 136 (1): p. 69-80.
689	13.	Xu, C.J., et al., DNA methylation in childhood asthma: an epigenome-wide meta-
690		analysis. Lancet Respir Med, 2018. 6(5): p. 379-388.
691	14.	Sheikhpour, M., et al., A review of epigenetic changes in asthma: methylation and
692		acetylation. Clinical Epigenetics, 2021. 13(1): p. 65.
693	15.	Smith, Z.D. and A. Meissner, DNA methylation: roles in mammalian development. Nat
694		Rev Genet, 2013. 14(3): p. 204-20.
695	16.	Moore, L.D., T. Le, and G. Fan, DNA Methylation and Its Basic Function.
696	. –	Neuropsychopharmacology, 2013. 38 (1): p. 23-38.
697	17.	Jin, Z. and Y. Liu, <i>DNA methylation in human diseases</i> . Genes Dis, 2018. 5 (1): p. 1-8.
698	18.	Torrone, D., et al., <i>Reproducibility and intraindividual variation over days in buccal cell</i>
699		DNA methylation of two asthma genes, interferon γ (IFN γ) and inducible nitric oxide
700	10	synthase (iNOS). Clin Epigenetics, 2012. 4(1): p. 3.
/01	19.	Lovinsky-Desir, S., et al., DNA methylation of the allergy regulatory gene interferon
702		gamma varies by age, sex, and tissue type in asthmatics. Clin Epigenetics, 2014. 6(1): p.
/03	20	
/04	20.	Cardenas, A., et al., <i>The nasal methylome as a biomarker of asthma and airway</i>
/05		inflammation in children. Nat Commun, 2019. 10(1): p. 3095.

706	21.	Yang, I.V., et al., <i>The nasal methylome and childhood atopic asthma</i> . J Allergy Clin
/0/		Immunol, 2017. 139 (5): p. 14/8-1488.
708	22.	Forno, E., et al., <i>DNA methylation in nasal epithelium, atopy, and atopic asthma in</i>
709	~~	children: a genome-wide study. Lancet Respir Med, 2019. 7(4): p. 336-346.
/10	23.	Jeong, A., et al., DNA Methylation in Inflammatory Pathways Modifies the Association
711		between BMI and Adult-Onset Non-Atopic Asthma. Int J Environ Res Public Health,
712		2019. 16(4).
713	24.	Bae, D.J., et al., <i>Epigenetic Changes in Asthma: Role of DNA CpG Methylation</i> . Tuberc
714		Respir Dis (Seoul), 2020. 83 (1): p. 1-13.
715	25.	Willemsen, G., et al., The Netherlands Twin Register Biobank: A Resource for Genetic
716		<i>Epidemiological Studies</i> . Twin Research and Human Genetics, 2010. 13 (3): p. 231-245.
717	26.	Hubert, M. and S. Engelen, Robust PCA and classification in biosciences.
718		Bioinformatics, 2004. 20(11): p. 1728-36.
719	27.	Bonder, M.J., et al., Disease variants alter transcription factor levels and methylation of
720		their binding sites. Nat Genet, 2017. 49(1): p. 131-138.
721	28.	North, M.L. and A.K. Ellis, The role of epigenetics in the developmental origins of
722		allergic disease. Ann Allergy Asthma Immunol, 2011. 106(5): p. 355-61; quiz 362.
723	29.	Yano, K., et al., GWAS with principal component analysis identifies a gene
724		comprehensively controlling rice architecture. Proc Natl Acad Sci U S A, 2019. 116(42):
725		p. 21262-21267.
726	30.	Georas, S.N., et al., <i>T-helper cell type-2 regulation in allergic disease</i> . European
727		Respiratory Journal, 2005. 26(6): p. 1119-1137.
728	31.	Bosnjak, B., et al., Treatment of allergic asthma: modulation of Th2 cells and their
729		responses. Respir Res, 2011. 12(1): p. 114.
730	32.	Barnes, P.J., Pathophysiology of allergic inflammation. Immunol Rev, 2011. 242(1): p.
731		31-50.
732	33.	Du, S. and H. Zheng, Role of FoxO transcription factors in aging and age-related
733		metabolic and neurodegenerative diseases. Cell & Bioscience, 2021. 11(1): p. 188.
734	34.	Farhan, M., et al., FOXO Signaling Pathways as Therapeutic Targets in Cancer. Int J Biol
735		Sci, 2017. 13 (7): p. 815-827.
736	35.	Malik, S. and A. Awasthi, Transcriptional Control of Th9 Cells: Role of Foxo1 in
737		Interleukin-9 Induction. Front Immunol, 2018. 9: p. 995.
738	36.	Kim, M.E., D.H. Kim, and J.S. Lee, FoxO Transcription Factors: Applicability as a
739		Novel Immune Cell Regulators and Therapeutic Targets in Oxidative Stress-Related
740		Diseases. Int J Mol Sci, 2022. 23(19).
741	37.	Tang, Y., et al., Promoter DNA methylation analysis reveals a combined diagnosis of
742		CpG-based biomarker for prostate cancer. Oncotarget, 2017. 8(35): p. 58199-58209.
743	38.	Shen, J., et al., Exploring genome-wide DNA methylation profiles altered in
744		hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips.
745		Epigenetics, 2013. 8(1): p. 34-43.
746	39.	Sun, X.J., et al., An integrated analysis of genome-wide DNA methylation and gene
747		expression data in hepatocellular carcinoma. FEBS Open Bio, 2018. 8(7): p. 1093-1103.
748	40.	Zhu, L., et al., Genome-wide DNA methylation profiling of primary colorectal laterally
749		spreading tumors identifies disease-specific epimutations on common pathways. Int J
750		Cancer, 2018. 143 (10): p. 2488-2498.

751 van der Meer, L.T., J.H. Jansen, and B.A. van der Reijden, Gfil and Gfilb: key regulators 41. 752 of hematopoiesis. Leukemia, 2010. 24(11): p. 1834-1843. 753 Kirchner, H., et al., Altered DNA methylation of glycolytic and lipogenic genes in liver 42. 754 from obese and type 2 diabetic patients. Mol Metab, 2016. 5(3): p. 171-183. 755 43. Moffatt, M.F., et al., A large-scale, consortium-based genomewide association study of 756 asthma. N Engl J Med, 2010. 363(13): p. 1211-1221. 757 44. 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). 758 759 Li, L., et al., Newborn DNA methylation and asthma acquisition across adolescence and 45. early adulthood. Clin Exp Allergy, 2022. 52(5): p. 658-669. 760 761 46. Wu, T., et al., Genetic and environmental influences on objective intermediate asthma 762 phenotypes in Dutch twins. Eur Respir J, 2010. 36(2): p. 261-8. 763 Van Asselt, A.J., et al., Genome-Wide DNA Methylation Profiles in Whole-Blood and 47. 764 Buccal Samples-Cross-Sectional, Longitudinal, and across Platforms. Int J Mol Sci, 765 2023. 24(19). 766 48. Los, H., P.E. Postmus, and D.I. Boomsma, Asthma genetics and intermediate phenotypes: a review from twin studies. Twin Res, 2001. 4(2): p. 81-93. 767 49. Bechger, T.M., D.I. Boomsma, and H. Koning, A Limited Dependent Variable Model for 768 769 Heritability Estimation with Non-Random Ascertained Samples. Behavior Genetics, 770 2002. **32**(2): p. 145-151. 771 Ouick-DNA Miniprep Kit Instruction Manual, in Online, Z. Research, Editor. 2020, Zymo 50. 772 Research: https://www.zymoresearch.com/collections/quick-dna-kits/products/quick-dna-773 miniprep. 774 51. van Dongen, J., et al., Genetic and environmental influences interact with age and sex in 775 shaping the human methylome. Nat Commun, 2016. 7: p. 11115. 776 Van Asselt, A.J. and E.A. Ehli, Whole-Genome Genotyping Using DNA Microarravs for 52. 777 Population Genetics, in Estrogen Receptors: Methods and Protocols, K.M. Eyster, Editor. 778 2022, Springer US: New York, NY. p. 269-287. Research, Z., EZ-96 DNA Methylation™ Kit, in Online, Z. Research, Editor. 2021. 779 53. 780 54. Infinium HD Assav Methylation Protocol Guide. [Protocol] 2015 [cited 2020: 15019519] v01: [Available from: https://www.illumina.com/products/by-type/microarray-781 782 kits/infinium-methylation-epic.html. 783 55. van Iterson, M., et al., MethylAid: visual and interactive quality control of large Illumina 784 450k datasets. Bioinformatics, 2014. 30(23): p. 3435-7. 785 Fortin, J.-P., et al., Functional normalization of 450k methylation array data improves 56. 786 replication in large cancer studies. Genome Biology, 2014. 15(11): p. 503. 787 57. van Iterson, M., et al., omicsPrint: detection of data linkage errors in multiple omics 788 studies. Bioinformatics, 2018. 34(12): p. 2142-2143. Min, J.L., et al., Meffil: efficient normalization and analysis of very large DNA 789 58. 790 methylation datasets. Bioinformatics, 2018. 34(23): p. 3983-3989. 791 59. Zhou, W., P.W. Laird, and H. Shen, Comprehensive characterization, annotation and 792 innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res, 2017. 793 45(4): p. e22. 794 60. Salas, L.A., et al., An optimized library for reference-based deconvolution of whole-blood 795 biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome 796 Biol, 2018. 19(1): p. 64.

- 797 61. Hubert, M., P. Rousseeuw, and K. Branden, *ROBPCA: A new approach to robust*
- *principal component analysis*. Technometrics, 2005. **47**: p. 64-79.
- 799 62. Odintsova, V.V., et al., DNA methylation in peripheral tissues and left-handedness. Sci
 800 Rep, 2022. 12(1): p. 5606.