

# Epigenome-wide association study identifies DNA methylation markers for asthma remission in blood and nasal epithelium

Cancan Qi<sup>1</sup>, Judith Vonk<sup>2</sup>, Diana A. van der Plaat<sup>3</sup>, Maartje Nieuwenhuis<sup>4</sup>, Nicole Dijk<sup>5</sup>, Dylan Aïssi<sup>6</sup>, Valérie SIROUX<sup>6</sup>, H. Boezen<sup>2</sup>, Chengjian Xu<sup>7</sup>, and Gerard Koppelman<sup>1</sup>

<sup>1</sup>University Medical Center Groningen

<sup>2</sup>University of Groningen, University Medical Center Groningen

<sup>3</sup>University of Groningen, University Medical Center Groningen

<sup>4</sup>University of Groningen, University Medical Center Groningen, GRIAC Research Institute

<sup>5</sup>University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital

<sup>6</sup>Université de Grenoble

<sup>7</sup>Research Group of Bioinformatics and Computational Genomics, CiiM, Centre for individualized infection medicine, a joint venture between Hannover Medical School and the Helmholtz Centre for Infection Research

June 9, 2020

## Abstract

Background: Asthma is a chronic respiratory disease which is not curable, yet some patients experience spontaneous remission. We hypothesized that epigenetic mechanisms may be involved in asthma remission. Methods: Clinical remission (ClinR) was defined as the absence of asthma symptoms and medication for at least 12 months, and complete remission (ComR) was defined as ClinR with normal lung function and absence of airway hyperresponsiveness. We analyzed differential DNA methylation of ClinR and ComR comparing to persistent asthma (PersA) in whole blood samples (n=72) and nasal brushing samples (n=97) in a longitudinal cohort of well characterized asthma patients. Significant findings of whole blood DNA methylation were tested for replication in two independent cohorts, Lifelines and EGEA. Results: We identified differentially methylated CpG sites associated with ClinR (7 CpG sites) and ComR (129 CpG sites) in whole blood. One CpG (cg13378519, Chr1) associated with ClinR was replicated and annotated to PEX11 (Peroxisomal Biogenesis Factor 11 Beta). The whole blood DNA methylation levels of this CpG were also different between ClinR and healthy subjects. One ComR-associated CpG (cg24788483, Chr10) that annotated to TCF7L2 (Transcription Factor 7 Like 2) was replicated and associated with expression of TCF7L2 gene. One out of seven ClinR-associated CpG sites and 8 out of 129 ComR-associated CpG sites identified from whole blood samples showed nominal significance ( $P < 0.05$ ) and the same direction of effect in nasal brushes. Conclusion: We identified DNA methylation markers possibly associated with clinical and complete asthma remission in nasal brushes and whole blood.

**KEYWORDS:** asthma remission, DNA methylation, whole blood, nasal brushes

## Introduction

Asthma is a chronic airway disorder that affects more than 300 million people around the world. Asthma cannot be completely cured with the available treatment up to now. However, some asthma patients may grow out of this disease, which is called asthma remission.

The remission rate of asthma varies from 35% to 60% according to different studies, and asthma remission is more common in patients with childhood onset asthma<sup>1</sup>. There are two types of asthma remission, one

is defined by absence of asthma symptoms and medication for at least one year, which is called “clinical remission” (ClinR). Some ClinR subjects may still have airway hyperresponsiveness (AHR) or a low lung function. Therefore, “complete remission” (ComR) was put forward<sup>2</sup>. In addition to criteria of ClinR, ComR has to meet additional criteria of normal lung function and absence of AHR. This is a more rare phenomenon, that takes place in 5%-22% of asthmatics<sup>3</sup>.

Asthma remission is associated with both genetic and environmental factors. Environmental factors including breast feeding and having pets in childhood were reported to be positively associated with asthma remission<sup>4</sup>. A recent genome-wide association study (GWAS) identified one SNP (single nucleotide polymorphism) that was associated with ClinR and three SNPs with ComR. One of these SNPs was also associated with expression of known asthma genes including *ILRL1* and *IL13* in lung tissue<sup>5</sup>.

Epigenetic mechanisms such as DNA methylation may help to build a link between genetic factors and the environment. DNA methylation can be regulated by SNPs, and can reflect environmental exposures, ageing, cell type constitution and activation<sup>6</sup>. DNA methylation refers to the addition of a methyl-group to cytosine, resulting in a CpG site, which may relate to the regulation of gene expression<sup>7</sup>. Epigenome-wide association studies (EWAS) have provided insights into the development of asthma and its remission. CpG sites associated with asthma in both blood and nasal epithelial cells have been identified, such as CpG sites located in *DICER1*, *STX3*, and *LIPIN1* in blood cells, and *CDHR3*, *FBXL7* and *NTRK1* in nasal epithelial cells<sup>8,9</sup>. Vermeulen *et al.* identified 4 CpG sites and 42 regions that were differentially methylated between remission and persistent asthma (PersA) in bronchial biopsies, and top CpG sites were annotated to genes including *ACKR2* and *DGKQ* by gene expression<sup>10</sup>.

Although this latter paper provided a proof of concept of the relation between asthma remission and DNA methylation, bronchial biopsies are not easy to obtain for further studies. DNA methylation in whole blood and nasal epithelium is a good proxy for bronchial epithelium and can therefore also help to understand the mechanism of asthma remission. Here, we hypothesize that epigenetic mechanisms may be involved in asthma remission, reflected in different DNA methylation patterns in whole blood and nasal epithelium. To test this hypothesis, we performed an EWAS of whole blood and nasal DNA in subjects with PersA, ClinR and ComR. We investigated a longitudinal cohort in which asthma was initially carefully defined and the remission status was assessed during follow-up (median 39 years). We subsequently replicated the whole blood DNA results in two independent cohorts, and also verified the top results from whole blood DNA in cells obtained by nasal brushing.

## Methods

A full description of methods is provided in the online supplement.

### Study populations

Subjects included in this study were from long-term follow-up studies in the University Medical Center Groningen (UMCG). This study consists of two sources of subjects, 1) subjects from the third visit (2013-2014) of a longitudinal study described previously by Carpaij *et al.*<sup>4</sup>, and 2) additional asthma remission subjects that were followed up from previous genetic studies<sup>11</sup> and were re-invited in 2013-2014 to take nasal brushes and blood samples, during which the remission status was assessed again. The participants included in this cohort were different from those of the cohort described by Vermeulen *et al.*<sup>10</sup> from our research institute. All participants were diagnosed with asthma at baseline with a doctor diagnosis of asthma and AHR. Then, subjects had at least one follow-up measurement in which their asthma status was evaluated by questionnaires and in most patients additionally with spirometry and an AHR test<sup>5</sup>. The medical ethical board of the UMCG approved the studies and all participants gave written informed consent.

We replicated findings in whole blood in two independent cohorts: the Lifelines population-based cohort in The Netherlands (where we could replicate our results on ClinR using Illumina 450 K array) and the Epidemiological study on the Genetics and Environment of Asthma (EGEA) cohort in France, a case control and family study on asthma, where we could replicate our results on ClinR and ComR with methylC-capture

sequencing<sup>12</sup>. No cohort was available to replicate our nasal findings. Extensive information on these cohorts is described in the supplemental materials.

## Phenotype Definition

The presence of PersA, ClinR and ComR was determined at the most recent visit. ClinR was defined according to the following criteria: 1) No use of any asthma medication, and 2) no symptoms (asthma attacks and/ or wheezing) in the past year. ComR was defined as ClinR combined with 3) no AHR (PC<sub>20</sub> (provocative concentration causing a 20% fall in FEV<sub>1</sub> (forced expiratory volume in 1 s)) methacholine ? 39.3 mg·mL<sup>-1</sup>), and/or 4) FEV<sub>1</sub>% predicted pre-bronchodilator > 80%. PersA was defined as the presence of asthma symptom and/or the use of asthma medication. Detailed phenotype definitions of the replication cohorts are described in the supplemental materials.

## DNA methylation measurements and statistical analyses

DNA was extracted from 72 whole blood and 103 nasal brushing samples. Genome-wide DNA methylation was determined using Illumina Infinium HumanMethylation450 BeadChips. After quality control, 72 whole blood samples and 97 nasal epithelium samples (of in total 103 unique subjects) with 436,824 CpG sites remained for following steps.

We used robust linear regression to determine the differential methylation between persistent asthma and asthma remission: 1) PersA versus ClinR, and 2) PersA versus ComR, in whole blood and nasal brushing samples, with adjustment for covariates that are known to affect DNA methylation (age, sex, smoking status, pack years, and batch). For whole blood samples, we performed adjustment on the percentage of monocytes, B cells, NK cells, CD4+ T cells, CD8+ T cells, neutrophils and eosinophils, which were predicted by the Houseman<sup>13</sup> algorithm using minfi<sup>14</sup> package. For nasal brushing samples, we applied the R package sva<sup>15</sup> to estimate significant surrogate variables (SVs), representing unknown latent factors that capture heterogeneity in data, such as cell type composition. Differentially methylated regions (DMR) were identified using comb-p v0.48<sup>16</sup>.

## Replication and meta-analyses

Genome-wide significant CpG sites that passed Bonferroni correction ( $P < 1.14 \times 10^{-7}$ , which is  $0.05 / 436,824$ ) in whole blood DNA were selected for replication in two independent cohorts, Lifelines and EGEA. Lifelines did not include an assessment of AHR, so only included a ClinR phenotype available for replication. The weighted Z-score method was used to meta-analyze the results of discovery and replication cohorts, considering EGEA used methylC-capture sequencing method, which was different from discovery and Lifelines (450K array), to investigate DNA methylation. CpG sites that passed the epigenome-wide significance threshold of  $P < 1.14 \times 10^{-7}$  (Bonferroni correction) in the meta-analysis of all studies were considered to be replicated. Finally, top sites identified in whole blood were verified in nasal brushes.

## Annotation and functional relevance

Significant CpG sites were annotated by GREAT v3.0.0<sup>17</sup>. We correlated the significant CpG sites we identified to the expression level of gene nearby (with a region of +/- 250kb) by *cis* expression quantitative trait DNA methylation (*cis*-eQTM) analysis. Blood eQTM was assessed in the BIOS consortium dataset<sup>18</sup>. Matched nasal RNA sequencing data and DNA methylation data was used to perform eQTM analysis for significant nasal CpG sites and these results were additionally replicated in the larger collection of nasal brushes obtained in the PIAMA study<sup>19</sup>. The genes that were identified in eQTM analysis were used for pathway analysis by ConsensusPathDB<sup>20</sup>. To check if the results of replicated CpG sites were affected by inhaled corticosteroids (ICS), we stratified asthma subjects by ICS usage, and compared the DNA-methylation in the remission group to asthma with ICS and asthma without ICS group, respectively, by t-test.

## Results

### Subject characteristics

Subject characteristics of the discovery and replication cohorts are shown in Table 1. Regarding the discovery cohort, the median [range] duration of follow-up of the entire population was 39 [4-49] years. Most of the subjects (n=76) were from the study described by Carpaij *et al.*<sup>4</sup> at the third visit; the remaining (n=27) were remission subjects added to this study to further enrich remission cases. Of the 103 subjects included in this study, 54 (52.4% of total subjects) had ClinR, and 20 of these had ComR (19.4% of total subjects). At baseline, subjects in remission later had higher FEV<sub>1</sub> and FVC level compared to PersA (Figure S1). Within the whole dataset, 72 subjects (44 PersA, 28 ClinR and 10 ComR) had whole blood methylation data and 97 (44 PersA, 53 ClinR and 19 ComR) had nasal methylation data, with 66 subjects providing both (Figure 1).

### Differential methylation in the discovery cohort of ClinR

Whole blood DNA methylation levels at seven individual CpG sites and 60 differentially methylated regions (DMR) were significantly associated with ClinR (Figure 2a, Table 2, Table S1). The Q-Q plot (Figure S2a) and the inflation factor  $\lambda$  of 1.064 indicated no obvious inflation of the results. Three out of the seven CpG sites were significantly correlated with gene-expression level in *cis* (FDR<0.05), resulting in four CpG-gene pairs that showed negative correlation (Table S2). These four eQTM genes were enriched in two pathways: G Protein Signaling Pathways and Hemostasis (P<0.01, Table S3)

We also identified four epigenome-wide significant CpG sites and 67 DMRs that associated with ClinR in nasal brushes (Table S4-S5). The Q-Q plot (Figure S3a) and the  $\lambda$  of 1.125 indicated no obvious inflation of the results. We verified the CpG sites identified from whole blood samples in nasal brushes, and found that one CpG (cg15404785) associated with ClinR showed nominal significance (P<0.05) and the same direction of effect in nasal brushes (Table S6). Four ClinR-associated CpG sites were correlated to expression level of four genes in *cis* (nominal significant P<0.05) in nasal brushes, with one pair (cg07673230-*RPL30* (Ribosomal Protein L30)) also being nominal significant and in the same direction in the PIAMA dataset (Table S7).

### Differential methylation in the discovery cohort of ComR

We identified 129 individual CpG sites and 233 regions that were significantly associated with ComR in whole blood DNA (Figure 2b, Table 3, Table S8-S9). Q-Q plot is shown in Figure S2b and  $\lambda$  was 1.178. The eQTM analysis identified 99 CpG-gene pairs for 45 out of the 129 CpG sites (FDR<0.05, Table S2). These eQTM genes were enriched in eight pathways (P <0.01, Table S3), and the top pathways were Activation of SMO (Smoothen), Platelet Adhesion to exposed collagen, and Cilium Assembly.

In nasal brushes, we identified 21 CpG sites and 119 DMRs that were associated with ComR (Table S10-S11). Q-Q plot is shown in Figure S3b and  $\lambda$  was 1.318. Eight of the ComR-associated CpG sites identified from whole blood also showed nominal significance in nasal brushes with the same direction of effect (Table S12). Seven ComR-associated CpG sites were correlated to expression level of nine genes in *cis* (nominal significant P<0.05) in nasal brushes, and four pairs also showed nominal significant and in the same direction in PIAMA dataset (Table S7).

### Replication study

We replicated the top findings in whole blood in two independent cohorts. After replication and meta-analysis, one CpG associated with ClinR and another one CpG associated with ComR were replicated (meta P<1.14×10<sup>-7</sup>). The ClinR-associated CpG (cg13378519) was located on Chr 1 and close to gene *PEX11B* (Peroxisomal Biogenesis Factor 11 Beta). This CpG was lower methylated in ClinR subjects compared to PersA (Figure 3a). A similar trend was also found in nasal brushes although the association was not statistically significant (Figure 3b). The methylation levels at this CpG also differed between ClinR and healthy subjects in Lifelines cohort (Figure 3c), indicating that DNA methylation status of asthma remission subjects is not equal to that of healthy subjects. The ComR-associated CpG (cg24788483) was located on chr10 and close to gene *HABP2* (Hyaluronan Binding Protein 2) and *TCF7L2* (Transcription Factor 7 Like 2). This CpG was lower methylated in ComR subjects compared to PersA in whole blood, but showed no obvious difference in nasal brushes (Figure S4). In whole blood *cis*-eQTM analysis, this CpG was significantly

negatively correlated to the expression level of *TCF7L2* gene ( $P=9.12 \times 10^{-7}$ ).

To evaluate potential SNP effects within the probe for the two replicated CpG sites, the  $\beta$  value distributions were visually assessed in the discovery cohorts (Figure S5), and no bimodal distribution was detected indicating no underlying SNP effect within the probe. The regional co-methylation plots of genes annotated to replicated CpG sites are shown in the Figure S6, which showed that cg13378519 is located in the promoter region of *PEX11B* gene, and cg24788483 is located in heterochromatin region of *TCF7L2* gene. We did not identify any SNP that was associated with the two CpG sites using an online database<sup>18</sup>.

We verified if the results of the two replicated CpG sites were affected by inhaled corticosteroids (ICS), known to affect DNA methylation at multiple sites in the genome (Figure S7). The boxplots showed that the methylation levels in the remission groups (both ClinR and ComR) were significantly lower than both persistent asthma with ICS usage group and persistent asthma without ICS usage group, which indicated that the results were not affected by the use of ICS.

We also compared our results from the discovery cohort with the published DNA methylation study of asthma remission in bronchial biopsies (Vermeulen *et al*<sup>10</sup>). One CpG (cg13525448), located close to *LBX1* (Ladybird Homeobox 1) gene and *TLX1* (T Cell Leukemia Homeobox 1) gene, in their top list of remission vs persistent asthma reached nominal significance ( $P < 0.05$ ) and showed the same direction of effect in our results of ComR in whole blood (Table S13). Two genes of their DMRs associated with ClinR, at the gene *PTCHD3* (patched domain-containing protein 3) and *LOC100507389*, were also present in our results of DMRs associated with ClinR in nasal brushes (Table S4).

## Discussion

In this study, we identified one CpG site associated with ClinR and a different CpG site associated with ComR from whole blood, that were replicated in independent cohorts. We also revealed that 25 CpG sites were associated with asthma remission phenotypes in nasal brushes. In the nasal dataset, two differentially methylated regions were previously observed in another DNA methylation study of asthma remission using bronchial biopsies. This is the first epigenome-wide association study of asthma remission in both whole blood and nasal epithelial cells. This study is important since understanding asthma remission may provide new leads for future asthma treatments, and epigenetic studies may help to reveal these cellular mechanisms.

The two replicated CpG sites were annotated to genes which were not known to be involved in asthma (remission) before. ClinR-associated CpG cg13378519 was located in the promoter region of *PEX11B* gene. *PEX11B* encodes a protein of the PEX11 family that participates in proliferation of peroxisomes<sup>21</sup>. Overexpression of the *PEX11B* gene in human cells induces peroxisome proliferation<sup>22,23</sup>. Our results could indicate a possible role of peroxisomal proliferation as being involved in asthma remission. ComR-associated CpG cg24788483 was located in the *TCF7L2* gene and also correlated with the expression level of this gene. The *TCF7L2* gene product is a transcription factor that plays a role in the Wnt signaling pathway. SNPs in *TCF7L2* were previously reported to be strongly associated with type-2 diabetes and this gene plays a role in the pancreatic insulin secretory response to incretins<sup>24, 25</sup>. Previously, the risk of asthma was reported to be higher in type 2 diabetes patients<sup>26</sup> and it has been suggested that treatment targeting insulin resistance may have a positive effect on asthma patients<sup>27</sup>. However, it has never been studied whether this treatment could contribute to asthma remission.

Although remission subjects do not have asthma symptoms any more, they still may be different from healthy people. Airway abnormalities such as basement membrane thickening still exist in clinical and complete remission subjects<sup>28</sup>. Vermeulen *et al.* reported that the DNA methylation profile in remission subjects is different from that in healthy subjects<sup>10</sup>. Our data in whole blood for our significantly replicated CpG confirmed this observation. The methylation levels of the ClinR associated CpG (cg13378519) were different between ClinR and healthy subjects, which indicated that the methylation status of subjects with asthma remission may not simply return to that of healthy people.

It is yet unknown if mechanisms of remission include local, tissue specific airway (epithelial) effects, immune

cell effects, or both. DNA methylation is tissue and cell-type specific. Matched data from whole blood and nasal brush samples enable us to compare DNA methylation profiles associated with asthma remission in two different tissues, as proxies of both immunological and airway specific mechanisms. Although we identified several interesting CpG sites related to remission in nasal epithelium, our results of nasal brushes still need further replication. In previous studies, several signals of DNA methylation were significantly associated with asthma in both whole blood and nasal epithelium<sup>9,29</sup>. In our study, no significant enriched cross-tissue effects of DNA methylation in asthma remission were identified, which we interpret that cells contributing to the differentially methylation patterns are not present in both compartments.

DNA methylation might be related to the regulation of gene expression, and eQTM analysis may help to understand the function of CpG sites. Among the eQTM genes that correlated with ClinR-associated CpG sites in whole blood, the protein encoded by *PRKCH* (Protein Kinase C Eta) plays a key role in epithelial tight junction regulation which is important in maintaining the integrity and function of the airway epithelial barrier<sup>30</sup>; *PDE1B* (Phosphodiesterase 1B) encoded a protein belonging to phosphodiesterases (PDEs) family, and various PDE inhibitors showed anti-inflammatory, anti-remodelling and bronchodilator effect and are potential treatment of asthma<sup>31</sup>. 99 eQTM genes in whole blood were enriched in eight pathways, among which three pathways were related to airway epithelial function: activation of SMO (Smoothen), cilium assembly and focal adhesion. The activation of SMO activity in bronchial epithelia enhanced the allergen-induced goblet cell metaplasia, which is defined as a reversible transformation of airway epithelial cells to mucous cells such as goblet cells and may occur in asthma<sup>32</sup>. Airway cilia are important for clearance of inhaled particles and pathogens. One study showed asthma patients had less ciliated cells in airway samples than healthy people<sup>33</sup>, and changes in ciliary function may be relevant for the development of asthma in children<sup>34</sup>. Focal adhesion between the cell membrane and matrix are essential elements of airway smooth muscle cells migration which may play an important role during the airway remodeling of persistent asthma<sup>35</sup>. Our results suggest that asthma remission associated CpG sites reflect a role for airway epithelial barrier and airway remodeling in asthma remission.

When comparing our results with those of Vermeulen *et al*<sup>10</sup> in bronchial biopsies, one CpG (cg13525448) in their top list of remission vs persistent asthma reached nominal significance ( $P < 0.05$ ) in our data and showed the same direction in our results of ComR in whole blood (Table S13). This CpG site is annotated to *LBX1* gene and *TLX1* gene by position. Besides, among their DMRs associated with ClinR, regions at the gene *PTCHD3* and *LOC100507389* also showed up in our results of ClinR in nasal brushes. SNPs in *PTCHD3* gene were previously associated with asthma in African American children<sup>36</sup>.

There are strengths and limitations in this study. To the best of our knowledge, this is the only cohort worldwide to study asthma remission in nasal brushes, which enable us to compare the DNA methylation profile in relation to asthma remission in whole blood and nasal epithelium. However, because of the uniqueness of this data, we could not replicate our results in nasal brushes in another independent cohort. Besides, we had matched DNA methylation and gene expression data in nasal brushes, which help us to get a better understanding of the function of CpG sites associated with asthma remission in nasal brushes. Regarding the limitations of this study, firstly, this study was performed in a relatively small sample with limited number of ComR cases, and the results were replicated in EGEA which is also with few ComR cases, which lead to low power of replication of ComR results. Although we studied over 450,000 DNA methylation sites, this only represents 1.6 % of the human DNA methylome. Thirdly, we studied mixed contributions of cells in whole blood and nasal brushes; and although we corrected for cell types using established methods, potentially stronger results may be anticipated when studying pure cell types, as was previously shown for DNA methylation sites associated with asthma in purified blood eosinophils<sup>8</sup>. Finally, we used a cross sectional analysis at follow up to investigate asthma remission. Future studies, with prospective designs, should be performed to investigate if these DNA methylation sites can predict future asthma remission.

In conclusion, we identified replicable DNA methylation signals associated with clinical and complete asthma remission, which may play a role in peroxisome proliferation and Wnt signaling pathway. This could help in identifying the underlying mechanisms of asthma

remission and conversely also of the chronicity of the disease.

## References

1. Upham JW, James AL. Remission of asthma: The next therapeutic frontier? *Pharmacology & Therapeutics* . 2011;130(1):38-45. doi:10.1016/j.pharmthera.2011.01.002
2. Vonk JM, Postma DS, Boezen HM, et al. Childhood factors associated with asthma remission after 30 year follow up. *Thorax* . 2004;59(11):925-929. doi:10.1136/thx.2003.016246
3. Carpaij OA, Burgess JK, Kerstjens HAM, Nawijn MC, van den Berge M. A review on the pathophysiology of asthma remission. *Pharmacol Ther* . 2019;201:8-24. doi:10.1016/j.pharmthera.2019.05.002
4. Carpaij OA, Nieuwenhuis MAE, Koppelman GH, van den Berge M, Postma DS, Vonk JM. Childhood factors associated with complete and clinical asthma remission at 25 and 49 years. *European Respiratory Journal* . 2017;49(6):1601974. doi:10.1183/13993003.01974-2016
5. Vonk JM, Nieuwenhuis MAE, Dijk FN, et al. Novel genes and insights in complete asthma remission: A genome-wide association study on clinical and complete asthma remission. *Clin Exp Allergy* . 2018;48(10):1286-1296. doi:10.1111/cea.13181
6. Qi C, Xu C-J, Koppelman GH. The role of epigenetics in the development of childhood asthma. *Expert Review of Clinical Immunology* . 2019;15(12):1287-1302. doi:10.1080/1744666X.2020.1686977
7. Gibney ER, Nolan CM. Epigenetics and gene expression. *Heredity* . 2010;105(1):4-13. doi:10.1038/hdy.2010.54
8. Xu C-J, Söderhäll C, Bustamante M, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med* . 2018;6(5):379-388. doi:10.1016/S2213-2600(18)30052-3
9. Forno E, Wang T, Qi C, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. *Lancet Respir Med* . 2019;7(4):336-346. doi:10.1016/S2213-2600(18)30466-1
10. Vermeulen CJ, Xu C-J, Vonk JM, et al. Differential DNA methylation in bronchial biopsies between persistent asthma and asthma in remission. *Eur Respir J* . Published online November 14, 2019. doi:10.1183/13993003.01280-2019
11. Nieuwenhuis MA, Siedlinski M, van den Berge M, et al. Combining genomewide association study and lung eQTL analysis provides evidence for novel genes associated with asthma. *Allergy* . 2016;71(12):1712-1720. doi:10.1111/all.12990
12. Allum F, Shao X, Guénard F, et al. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun* . 2015;6:7211. doi:10.1038/ncomms8211
13. Houseman E, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* . 2012;13(1):86. doi:10.1186/1471-2105-13-86
14. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* . 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
15. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* . 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034
16. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* . 2012;28(22):2986-2988. doi:10.1093/bioinformatics/bts545

17. McLean CY, Bristor D, Hiller M, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* . 2010;28(5):495-501. doi:10.1038/nbt.1630
18. the BIOS Consortium, Bonder MJ, Luijk R, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* . 2017;49(1):131-138. doi:10.1038/ng.3721
19. Qi C, Jiang Y, Yang IV, et al. Nasal DNA methylation profiling of asthma and rhinitis. *J Allergy Clin Immunol* . Published online January 14, 2020. doi:10.1016/j.jaci.2019.12.911
20. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nature Protocols* . 2016;11(10):1889-1907. doi:10.1038/nprot.2016.117
21. Koch J, Pranjic K, Huber A, et al. PEX11 family members are membrane elongation factors that coordinate peroxisome proliferation and maintenance. *Journal of Cell Science* . 2010;123(19):3389-3400. doi:10.1242/jcs.064907
22. Schrader M, Reuber BE, Morrell JC, et al. Expression of PEX11beta mediates peroxisome proliferation in the absence of extracellular stimuli. *J Biol Chem* . 1998;273(45):29607-29614. doi:10.1074/jbc.273.45.29607
23. Ebberink MS, Koster J, Visser G, et al. A novel defect of peroxisome division due to a homozygous non-sense mutation in the PEX11 $\beta$  gene. *J Med Genet* . 2012;49(5):307-313. doi:10.1136/jmedgenet-2012-100778
24. Grant SFA, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* . 2006;38(3):320-323. doi:10.1038/ng1732
25. Pearson ER. Translating TCF7L2: from gene to function. *Diabetologia* . 2009;52(7):1227-1230. doi:10.1007/s00125-009-1356-1
26. Thomsen SF, Duffy DL, Kyvik KO, Skytthe A, Backer V. Risk of asthma in adult twins with type 2 diabetes and increased body mass index: Type 2 diabetes, obesity and asthma in twins. *Allergy* . 2011;66(4):562-568. doi:10.1111/j.1398-9995.2010.02504.x
27. Carpaij OA, van den Berge M. The asthma-obesity relationship: underlying mechanisms and treatment implications. *Current Opinion in Pulmonary Medicine* . 2018;24(1):42-49. doi:10.1097/MCP.0000000000000446
28. Broekema M, Timens W, Vonk JM, et al. Persisting Remodeling and Less Airway Wall Eosinophil Activation in Complete Remission of Asthma. *American Journal of Respiratory and Critical Care Medicine* . 2011;183(3):310-316. doi:10.1164/rccm.201003-0494OC
29. Cardenas A, Sordillo JE, Rifas-Shiman SL, et al. The nasal methylome as a biomarker of asthma and airway inflammation in children. *Nat Commun* . 2019;10(1):3095. doi:10.1038/s41467-019-11058-3
30. Suzuki T, Elias BC, Seth A, et al. PKC  $\epsilon$  regulates occludin phosphorylation and epithelial tight junction integrity. *Proc Natl Acad Sci USA* . 2009;106(1):61-66. doi:10.1073/pnas.0802741106
31. Zuo H, Cattani-Cavaliere I, Musheshe N, Nikolaev VO, Schmidt M. Phosphodiesterases as therapeutic targets for respiratory diseases. *Pharmacol Ther* . 2019;197:225-242. doi:10.1016/j.pharmthera.2019.02.002
32. Xu C, Zou C, Hussain M, et al. High expression of Sonic hedgehog in allergic airway epithelia contributes to goblet cell metaplasia. *Mucosal Immunol* . 2018;11(5):1306-1315. doi:10.1038/s41385-018-0033-4
33. Vieira Braga FA, Kar G, Berg M, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* . Published online June 17, 2019. doi:10.1038/s41591-019-0468-5
34. Dizier M-H, Nadif R, Margaritte-Jeannin P, et al. Interaction between the *DNAH9* gene and early smoke exposure in bronchial hyperresponsiveness. *Eur Respir J* . 2016;47(4):1072-1081. doi:10.1183/13993003.00849-2015



35. Gerthoffer WT. Migration of Airway Smooth Muscle Cells. *Proceedings of the American Thoracic Society* . 2008;5(1):97-105. doi:10.1513/pats.200704-051VS
36. White MJ, Risse-Adams O, Goddard P, et al. Novel genetic risk factors for asthma in African American children: Precision Medicine and the SAGE II Study. *Immunogenetics* . 2016;68(6-7):391-400. doi:10.1007/s00251-016-0914-1

Table 1. Characteristics of study participants in the discovery and replication cohorts

	Discovery cohort PersA (n=49)	Discovery cohort ClinR (n=54)	Discovery cohort ComR (n=20)	Lifelines PersA (n=99)	Lifelines ClinR (n=25)	Lifelines Healthy (n=636)	EGEA PersA (n=106)	EGEA ClinR (n=15)	EGEA Healthy (n=10)
<b>Characteristics at last visit</b>									
<b>Age</b>	49.8	50.5	50.5	46.9	43.2	46.3	52.6	49.2	30.1
<b>at</b>	(48.5;	(48.3;	(48.4;	(40.7;50.7)	(35.7;50.3)	(39.2;52.9)	(36.7;	(36.7;	(28.1;
<b>remis-</b>	51.3)	57.6)	59.8)				61.4)	57.6)	59.8)
<b>sion</b>									
<b>sta-</b>									
<b>tus,</b>									
<b>years</b>									
<b>Male,</b>	31	37	15	58	10	384	49	7	1
<b>n (%)</b>	(63.3%)	(68.5%)	(75.0%)	(58.6%)	(40.0%)	(60.4%)	(46.7%)	(46.7%)	(33.3%)
<b>Duration</b>	39	39 (37;	39				11.3	11.6	11.1
<b>of</b>	(38;40.5)	42)	(37;41)				(10.7;	(10.7;	(10.7;
<b>follow-</b>							12.1)	12.2)	12.1)
<b>up,</b>									
<b>years</b>									
<b>Current-</b>	4	8	2	27	14	207	10	2	1
<b>smoking,</b>	(8.2%)	(14.8%)	(10.0%)	(27.3%)	(56.0%)	(32.5%)	(9.5%)	(13.3%)	(33.3%)
<b>n (%)</b>									
<b>Ex-</b>	13	18	7	0	0	0	35	5	0
<b>smoking,</b>	(26.5%)	(33.3%)	(35.0%)	(0.0%)	(0.0%)	(0.0%)	(33.3%)	(33.3%)	
<b>n (%)</b>									
<b>Never-</b>	32	28	11	72	11	429	60	8	2
<b>smoking,</b>	(65.3%)	(51.9%)	(55.0%)	(72.7%)	(44.0%)	(67.5%)	(57.1%)	(53.3%)	(66.7%)
<b>n (%)</b>									
<b>ICS, n</b>	31	0	0	48	0	0	72	0	0
<b>(%)</b>	(63.3%)			(48.5%)	(0.0%)	(0.0%)	(70.6%)		
<b>FEV1%</b>	80.6	<b>90.2</b>	<b>101.8</b>	81.2	87.1	101.5	87.0	95.2	97.1
<b>pred</b>	(71.1;	<b>(82.9;</b>	<b>(85.9;</b>	(70.9;91.6)	(74.0;93.4)	(96.1;108.7)	(71.7;	(87.5;	(90.1;
	91.9)	<b>104.7)</b>	<b>110.0)</b>				98.4)	107.2)	121.1)
<b>FVC%</b>	94.2	<b>101.7</b>	<b>106.6</b>	94.1	100.3	104.0	98.5	95.2	96.1
<b>pred</b>	(88.4;	<b>(95.2;</b>	<b>(101.2;</b>	(85.8;102.2)	(90.4;109.2)	(97.5;111.4)	(84.8;	(86.9;	(88.4;
	103.7)	<b>111.5)</b>	<b>115.9)</b>				108.9)	106.1)	119.1)
<b>Characteristics at base- line</b>									

<b>Age, years</b>	10.0 (9.0; 11.0)	11.0 (9.0; 17.5)	11.5 (9.8; 15.8)				40.6 (25.7; 48.7)	37.0 (26.0; 46.0)	18.0 (18.0; 46.0)
<b>FEV1% pred</b>	76.6 (67.4; 84.0)	77.1 (67.5; 83.8)	78.1 (73.5; 83.7)				81.7 (66.2; 95.9)	90.6 (82.0; 103.0)	82.0 (80.0; 120.0)
<b>FVC% pred</b>	88.8 (84.0; 96.1)	88.3 (80.9; 95.3)	88.9 (79.8; 96.5)				95.5 (83.7; 104.3)	86.4 (81.9; 100.6)	84.0 (70.0; 110.0)
<b>Start asthma before age 16, n (%)</b>	48 (98.0%)	50 (92.6%)	19 (95.0%)	59 (62.1%)	18 (81.8%)		52 (49.5%)	8 (53.3%)	2 (60.0%)

Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.	Continuous data are pre- sented as median (25th per- centile; 75th per- centile), cate- gory data are pre- sented as num- ber (per- cent- age). Bold values repre- sent signifi- cant differ- ences com- pared with persis- tent asthma with P<0.05 in the discov- ery cohort.
--	--	--	--	--	--	--	--	--	--

Table 2. Seven genome-wide significant CpG sites (Pvalue < 1.14×10<sup>-7</sup>) of ClinR in relation to methylation in whole blood

---

## CpG

**cg24788483+**

cg15223066

cg26909813

cg02341571

cg00269245

cg24833566

cg25881850

cg24730612

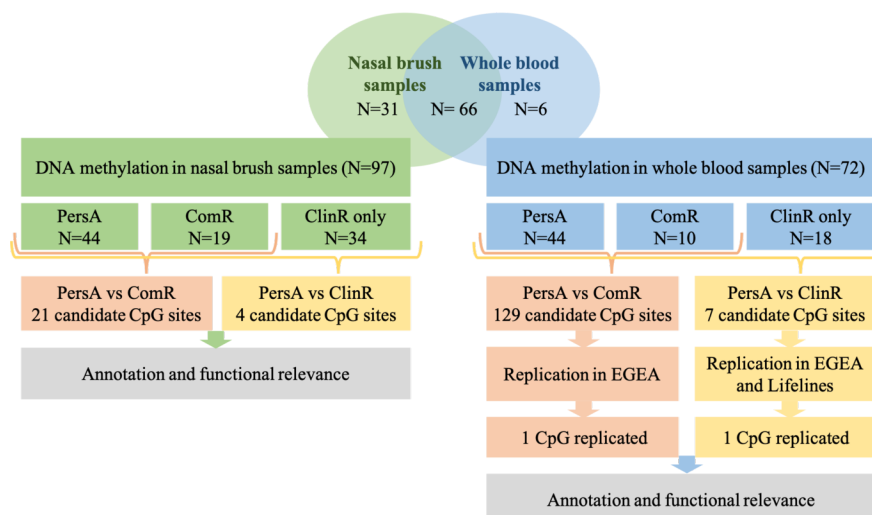
cg23250019

+ the CpG site was replicated with meta P value  $< 1.14 \times 10^{-7}$ . \* N (persA/ComR), N varies from one CpG to another because

---

## Figure Legends

**Figure 1. Study design.** In the discovery panel, 103 samples were assessed in this study, 72 subjects had whole blood DNA methylation data, 97 subjects had nasal DNA methylation data and 66 subjects have both. Epigenome-wide association studies were performed on clinical remission (ClinR) and complete remission (ComR) respectively in both whole blood and nasal samples. Significant CpG sites identified from whole blood were further replicated in two independent cohorts.



**Figure 2. Manhattan plot of association between clinical remission (a)/ complete remission (b) and DNA methylation in whole blood in discovery cohort.** In total, 436,824 CpGs were tested. The red line represents the genome-wide significance threshold (Bonferroni correction,  $P < 1.14 \times 10^{-7}$ ). Highlighted sites represent replicate CpG sites.

#### Hosted file

image2.emf available at <https://authorea.com/users/331635/articles/458240-epigenome-wide-association-study-identifies-dna-methylation-markers-for-asthma-remission-in-blood-and-nasal-epithelium>

**Figure 3. Boxplot illustrating DNA methylation levels of cg13378519 in persistent asthma, clinical remission and complete remission subjects.** (a) DNA methylation levels in whole blood in discovery cohort; (b) DNA methylation levels in nasal brushes in discovery cohort; (c) DNA methylation levels in whole blood in replication cohort (Lifelines).

#### Hosted file

image3.emf available at <https://authorea.com/users/331635/articles/458240-epigenome-wide-association-study-identifies-dna-methylation-markers-for-asthma-remission-in-blood-and-nasal-epithelium>

