Genetic and epigenetic regulation of YKL-40 in childhood

Stefano Guerra¹⁴, MD, PhD; Erik Melén⁵, MD, PhD; Jordi Sunyer¹³, MD, PhD; Cheng-Jian Xu⁷⁻⁹, PhD; Iris Lavi¹⁻³, PhD; Marta Benet¹⁻³, BStat; Mariona Bustamante¹, M, PhD; Anne-Elie Carsin¹⁻³, MS; Carlota Dobaño¹¹, PhD; Mònica Guxens¹⁻³, MD, PhD; Christina Tischer¹⁻³, PhD; Martine Vrijheid¹⁻³, PhD; Inger Kull¹³, RN, PhD; Anna Bergström⁵, PhD; Ashish Kumar⁵⁻¹⁵, MSc; Cilla Söderhäll¹⁶, PhD; Ulrike Gehring¹⁷, PhD; Dorieke J. Dijkstra¹⁸, M. Sc.; Pieter van der Vlies⁹, BSc; Magnus Wickman⁵, MD, PhD; Jean Bousquet¹⁹, MD, PhD; Dirkje S Postma⁷⁻⁸, MD, PhD; Josep M Anto¹⁻³, MD, PhD; and Gerard H Koppelman⁸,²⁻²⁰, MD, PhD

1. ISGlobal, Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
2. Universitat Pompeu Fabra, Barcelona, Spain
3. CIBER Epidemiología y Salud Pública (CIBERESP), Spain
4. Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ, USA
5. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
6. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
7. University of Groningen, University Medical Center Groningen, Department of Pulmonology, The Netherlands
8. Groningen Research Institute for Asthma and COPD, The Netherlands
9. University of Groningen, University Medical Center Groningen, Department of Genetics, The Netherlands
10. Centre for Genomic Regulation (CRG), the Barcelona Institute of Science and Technology, Barcelona, Spain
11. ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clinic, Universitat de Barcelona, Barcelona
12. Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Centre–Sophia Children’s Hospital, Rotterdam, The Netherlands
13. Sachs’ Children’s Hospital, Södersjukhuset, Stockholm, Sweden; Department of Clinical Science and Education, Karolinska Institutet at Södersjukhuset, Stockholm, Sweden
14. Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
15. University of Basel, Basel, Switzerland
16. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge; and Department of Women’s and Children’s Health, Karolinska Institutet, Stockholm, Sweden
17. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
18. University of Groningen, University Medical Center Groningen, Department of Obstetrics and Gynecology, The Netherlands
19. University Hospital Montpellier; and Respiratory and Environmental Epidemiology Team, INSERM 1018, CESP Centre, Villejuif, France
20. University of Groningen, University Medical Center Groningen, Beatrix Children’s Hospital, Department of Pediatric Pulmonology and Pediatric Allergology, The Netherlands

Corresponding author:
Stefano Guerra, MD, PhD
ISGlobal CREAL, Doctor Aiguader 88, 08003 Barcelona, Spain; and
Asthma and Airway Disease Research Center, 1501 N Campbell Ave, Tucson, AZ 85724, USA
Email: stefano@email.arizona.edu

This study was supported by grant agreement number 261357 (Mechanisms of the Development of ALLergy - MeDALL) from the EU Seventh Framework Programme.

Additional funding for individual cohorts was provided by:
- For INMA: Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; PI041436; PI081151 incl. FEDER funds; MS13/00054), Generalitat de Catalunya-CIRIT 1999SGR 00241, Fundació La marató de TV3 (090430)
- For BAMSE: The Swedish Heart-Lung Foundation, The Swedish Research Council, Freemason Child House Foundation in Stockholm, Stockholm County Council (ALF), Swedish Foundation for Strategic Research (SSF, Rbc08-0027, EpiGene project) and the Strategic Research Programme (SFO) in Epidemiology at Karolinska Institutet
- For PIAMA: The Netherlands Organization for Health Research and Development; The Netherlands Organization for Scientific Research; The Netherlands Lung Foundation (AF 4.1.14.001); The Netherlands Ministry of Spatial Planning, Housing, and the Environment; and The Netherlands Ministry of Health, Welfare, and Sport.
Abstract

Background: Circulating levels of the chitinase-like protein YKL-40 are influenced by genetic variation in its encoding gene (CHI3L1) and are increased in several diseases, including asthma. Epigenetic regulation of circulating YKL-40 early in life is unknown.

Objective: To determine (1) whether methylation levels at CHI3L1 CpG sites mediate the association of CHI3L1 single nucleotide polymorphisms (SNPs) with YKL-40 levels in the blood; and (2) whether these biomarkers (CHI3L1 SNPs, methylation profiles, and YKL-40 levels) are associated with asthma in early childhood.

Methods: We used data from up to 2405 participants from the INMA, BAMSE, and PIAMA birth cohorts. Associations between 68 CHI3L1 SNPs, methylation levels at 14 CHI3L1 CpG sites in whole blood DNA, and circulating YKL-40 levels at 4 years of age were tested using correlation analysis, multivariable regression, and mediation analysis. Each of these biomarkers was also tested for association with asthma at 4 years of age using multivariable logistic regression.

Results: YKL-40 levels were significantly associated with seven SNPs and with methylation at five CpG sites. Consistent associations between these seven SNPs (particularly rs10399931 and rs4950928) and five CpG sites were observed. Alleles linked to lower YKL-40 levels were associated with higher methylation levels. Participants with high YKL-40 levels (defined as the highest YKL-40 tertile) had increased odds for asthma as compared with subjects with low YKL-40 levels [meta-analyzed adjusted odds ratio (adjOR): 1.90, 1.08-3.36]. In contrast, neither SNPs nor methylation levels at CpG sites in CHI3L1 were associated with asthma.

Conclusions: The effects of CHI3L1 genetic variation on circulating YKL-40 are partly mediated by methylation profiles. In our study, YKL-40 levels, but not CHI3L1 SNPs or methylation levels, were associated with childhood asthma.
Clinical Implications

Methylation levels at CHI3L1 CpG sites mediate part of the effects of CHI3L1 genetic variation on circulating levels of YKL-40, but they are not associated with childhood asthma.

Capsule Summary

Circulating YKL-40 and variation in its encoding gene CHI3L1 have been associated with asthma. We found that methylation levels at CHI3L1 CpG sites partly mediated CHI3L1 genetic effects on circulating YKL-40, although they were not associated with childhood asthma.

Key words

YKL-40; CHI3L1; asthma; epigenetics; DNA methylation; genetics

Abbreviations

adjOR: adjusted odds ratio
BMI: body mass index
CHI3L1: chitinase-3-like-1
GWAS: Genome-Wide Association Study
LD: linkage disequilibrium
MAF: minor allele frequency
meQTL: methylation quantitative trait locus
PCA: principal component analysis
pQTL: protein quantitative trait locus
SNP: single nucleotide polymorphism
Introduction

YKL-40, a chitinase-like protein, is upregulated in asthma, cancer, and other diseases characterized by inflammation and tissue remodeling. In adults, YKL-40 levels are increased in the blood and lungs of people with asthma and correlate with lung function deficits, disease severity, and persistence. However, to date, findings on YKL-40 and asthma in childhood have been conflicting.

The mechanisms linking YKL-40 with asthma remain to be determined, although in vitro and animal studies support its role in Th2 adaptive immune responses. The association between genetic variation in chitinase-3-like-1 (CHI3L1, the gene encoding YKL-40) and asthma susceptibility/severity suggests a possibly causal link. A group of four single nucleotide polymorphisms (SNPs) tagged by rs4950928 were first associated with asthma in three of four tested populations, with the major alleles conferring increased risk. In the same study, serum YKL-40 levels were found to be a highly heritable quantitative trait in the general population and to be directly associated with the same C allele at rs4950928 that was associated with asthma risk. Subsequent studies showed that another SNP, rs10399931, which is in strong linkage disequilibrium (LD) with rs4950928, had similar, if not stronger, effects on gene expression, plasma YKL-40 levels, and asthma. However, whereas the association between CHI3L1 genetic variation and YKL-40 levels has been conclusively established, the relation of CHI3L1 variation to asthma remains controversial because other reports, including a large study of over 6500 Danish adults, failed to replicate the aforementioned genetic associations with asthma.

The identification of the mechanisms by which genetic variation in CHI3L1 regulates YKL-40 levels may have important implications for the understanding of the potential impact of this gene on human disease. Epigenetic regulation of gene expression is one of the possible mechanisms by which genetic variation can affect protein levels and disease susceptibility, including childhood respiratory diseases. DNA sequence variants across the genome have been shown to have cis- (and to less extent trans-) effects on methylation levels at specific CpG sites. Yet, to date no study has addressed DNA methylation as a possible...
intermediary mechanism of the relation between CHI3L1 genetic variation and upregulated YKL-40 protein levels. The primary goal of the present study was to determine whether methylation levels at CHI3L1 CpG sites mediate the association of CHI3L1 genetic variation to YKL-40 levels in blood. In secondary analyses, we also sought to assess whether the CHI3L1 genotype and methylation levels that regulate YKL-40 levels are associated with asthma in early childhood.

Methods
Study populations and design. This study was part of the Mechanisms of the Development of ALLergy (MeDALL) project\textsuperscript{31}, which included analyses on YKL-40 as an \textit{a priori} biomarker candidate. The design and available data for the present study are summarized in Figure 1. Primary analyses (Figure 1a) included molecular and phenotypic data from 433 participants who were 4 years of age from the Spanish INfancia y Medio Ambiente\textsuperscript{32} (INMA, N=203) and the Swedish Barn/Children, Allergy, Milieu, Stockholm, Epidemiological survey\textsuperscript{33} (BAMSE, N=230) birth cohorts. These children were selected from their original cohorts for epigenetic and YKL-40 studies in MeDALL based on the following sampling strategy: among participants who provided paired DNA samples at two time points (birth and 4 years for INMA, 4 and 8 years for BAMSE), we selected children who had one or more of three diagnoses (asthma, eczema, and allergic rhinitis) and a similar number of randomly selected control subjects (none of the three diagnoses, nested case-control design). Of them, 172 INMA and 78 BAMSE participants also had available Genome-Wide Association Study (GWAS) data.

In secondary analyses of association with asthma (Figure 1b), in addition to INMA and BAMSE, we also included 4-year-old participants from the Prevention and Incidence of Asthma and Mite Allergy (PIAMA) study to increase statistical power\textsuperscript{34, 35}. No serum YKL-40 levels were available for PIAMA, but epigenetic studies were completed on 193 PIAMA participants who were selected based on the same sampling strategy described previously.
For analyses of genetic association, we used data from all participants from the three
cohorts who had both GWAS data and asthma information (N = 336 for INMA, 467 for
BAMSE, and 1602 for PIAMA; total N=2405).

A detailed description of the three birth cohorts and additional information on molecular
assays are provided in the online repository. For all cohorts, parents gave written informed
consent, and the local ethics review boards approved the performed studies and procedures.

Definition of asthma. In MeDALL, an asthma definition[^36] was used that included positive
answers to at least two of the three following questions: (1) “Has your child ever been
diagnosed by a doctor as having asthma?”; (2) “Has your child taken any medicines for
asthma (including inhalers, nebulizers, tablets, or liquid medicines) or for breathing
difficulties (chest tightness, shortness of breath) in the last 12 months?”; and (3) “Has your
child had wheezing or whistling in their chest at any time in the last 12 months?”; or “Has
your child had breathing difficulties (chest tightness, shortness of breath) in the last 12
months?”. This definition was used for BAMSE and PIAMA. Because information on doctor-
diagnosed asthma was not available in INMA, for this cohort the asthma definition included a
positive response to the two remaining questions.

Molecular data.

**DNA methylation assays at CHI3L1 CpG sites.** Epigenome-wide analysis scans of paired
whole peripheral blood DNA methylation samples at birth and age 4 (INMA), and at age 4 and
8 (BAMSE and PIAMA) were generated using the Illumina Infinium HumanMethylation 450
BeadChip assays. All samples were quality checked as described in the online supplement.
The Illumina 450K included 485577 assays. During processing, the probes on sex
chromosomes, the probes that mapped on multi-loci, 65 SNP assays, and the probes
containing SNPs at the target CpG sites with a minor allele frequency (MAF>10%) were
excluded. This resulted in a total of 439306 CpG sites. Among them, 14 CpG sites were
annotated to the CHI3L1 gene and analyzed in the present study (Table E1). Two of them
(cg17014757 and cg03625911) have probes that include known SNPs with MAF greater than 5% in European populations, as reported by Chen et al. The probe of cg17014757 includes rs10399931, and the probe of cg03625911 includes rs7515776, which according to the CEU panel of the 1000 Genomes project, is in nearly complete LD with rs10399805. These two CpG sites were included in statistical analyses, but associations of their methylation levels with overlapped SNPs or with other SNPs in LD with the overlapped SNPs (rs10399931 and rs4950928 for cg17014757; and rs2886117, rs10399805, and rs7542294 for cg03625911) were identified in tables to caution interpretation. Methylation levels (beta values, β) at a given CpG site were derived from the ratio of the methylated probe intensity and overall intensity (sum of methylated and unmethylated probe intensities). Thus, β is equal to $\frac{M}{U + M + \alpha}$, where M is intensity of the methylated probe, U is the intensity of the unmethylated probe, and α is the constant offset with the default value of 100. The intensity has been corrected for type I and type II probe differences and normalized by using the “dasen” method in the watermelon R package. Results from analyses on methylation levels were also confirmed after adjustment for blood cell-type composition as predicted by the Houseman algorithm, and, in a subset of 116 INMA participants at year 4, after adjustment for available differential cell counts obtained directly from microscopic inspection of blood smears. The methylation signal of the top five CpG sites obtained by the Illumina microarray was also validated by pyrosequencing in whole blood DNA from 96 subjects participating in the PIAMA study (see online repository, Table E11, and Figure E4).

YKL-40 measurements. Circulating levels of YKL-40 were measured in serum (INMA) and plasma (BAMSE) samples at age 4 years using a commercially available enzyme-linked immunosorbent assay kit (Quantikine Human CHI3L1 immunoassay by R&D, Abingdon, UK).

GWAS. Genome-wide genotyping had been previously completed for the three cohorts, INMA, BAMSE, and PIAMA, using various platforms (see online repository). For the present study, in the INMA cohort we tested 68 SNPs that were genotyped from the genomic region surrounding (+/- 50kb) CHI3L1 (hg19: chr1:203,098,059-203,205,922) using the
HumanOmni1-Quad Beadchip (Illumina, San Diego, CA, USA). SNPs that were found not to be in Hardy-Weinberg equilibrium using exact tests\textsuperscript{40} were filtered (see additional methods for GWAS assays in the online repository). SNPs that were significantly associated with YKL-40 in INMA were also tested in BAMSE and PIAMA.

Analytical approach and statistical analysis.

For the primary goal, we analyzed interrelationships between multi-level biomarkers (SNPs, methylation at CpG sites, and YKL-40 levels) in INMA and BAMSE according to a stepwise approach (Figure 1a). Because results of these analyses did not significantly differ between cases and controls, they are presented with no stratification by disease. YKL-40 levels were log-transformed and then standardized within each cohort by subtracting the mean and dividing the result by the standard deviation as done in previous multi-cohort studies\textsuperscript{41}. Discovery analyses were completed in INMA with a conservative Bonferroni correction, and replication analyses in BAMSE with a one-tailed $\alpha = 0.05$ (i.e., only associations with the same direction of effect between the two cohorts were tested for significance). First, associations between \textit{CHI3L1} SNPs and YKL-40 levels were studied. In INMA, 68 SNPs were tested in linear regressions predicting YKL-40 protein levels according to additive genetic models. Because INMA and BAMSE used different GWAS platforms, in BAMSE we used genotyped or imputed SNPs for replication of INMA results as appropriate. Second, methylation levels were tested for correlation with YKL-40 levels using Spearman correlation coefficients. Third, the SNPs and CpG sites that were found to be related to YKL-40 levels were tested for association with each other using robust regressions to reduce the impact of potential outlier observations (see online repository for additional information).

To determine what SNPs and CpG sites were independently related to YKL-40 protein levels, multivariable regression models with backward stepwise variable selection were used. A final mediation analysis was completed in INMA using the R package “mediation”\textsuperscript{42} to estimate the effects by the aforementioned independent SNPs on YKL-40 that were mediated by methylation. For this analysis, we used both methylation levels at the CpG site.
(cg07423149) that was identified by the above backward stepwise selection as well as the first principal component obtained by a principal component analysis (PCA) on methylation levels at all the five CpG sites that were associated with YKL-40.

For our secondary analyses (Figure 1b), we performed separate multiple logistic regression models testing the association of SNPs, CpG methylation levels, and YKL-40 levels with asthma with adjustment for sex, age, and body mass index (BMI) at 4 years, as these demographic factors may affect methylation and YKL-40 levels as well as asthma risk. To test for non-linearity of effects, YKL-40 levels were also used as tertiles, and asthma risks were compared between subjects having medium and high YKL-40 levels and subjects having low YKL-40 levels.

**Results**

Table 1 shows characteristics of children included in the primary and secondary analyses.

Table E2 compares demographic characteristics of participants from the three cohorts that were or were not included in the present study.

**Primary Analyses: Relation between CHI3L1 SNPs, methylation, and YKL-40 levels**

*SNPs and YKL-40 protein levels*

Among the 68 tested SNPs, after Bonferroni correction, seven SNPs were found to be significantly associated with YKL-40 levels in INMA (Table E3 for complete analysis, Table 2 for significant associations). Overall, minor alleles were associated with higher YKL-40 levels, with the exceptions of rs10399931 and rs4950928. These seven SNPs explained individually up to 16% of the variability in YKL-40 levels ($R^2$ between 0.08 and 0.16), and they did not belong to a single block of LD. The LD matrix of CHI3L1 SNPs in INMA is shown in Figure E1.
In BAMSE, associations with YKL-40 levels were replicated for five (rs2886117, rs10399931, rs10399805, rs4950928, and rs7542294) of the seven SNPs that provided significant signals in INMA (Table E4).

- **DNA methylation and YKL-40 protein levels**

Descriptive statistics of the methylation levels at the 14 CpG sites are shown in Table E5, their distributions in Figure E2, and their correlation matrices in Figures 2 and E4. The first seven CpG sites covering the 5’ region to the first exon of the gene correlated with each other. This correlation pattern was found at all ages: birth (INMA), age 4 years (INMA and BAMSE), and age 8 years (BAMSE) (Figures 2 and E4).

Correlations between methylation levels at CpG sites and YKL-40 levels for INMA and BAMSE are shown in Table 3. In INMA, serum YKL-40 levels at age 4 years correlated significantly with methylation levels at age 4 years at eight CpG sites after Bonferroni correction (p values ranging from $3 \times 10^{-16}$ to 0.003). All correlations were negative, i.e., the higher the methylation, the lower is the YKL-40 level. Inverse correlations between methylation levels and YKL-40 at age 4 years were replicated in BAMSE for five of these CpG sites (cg13134650, cg07423149, cg17014757, cg14085262, and cg03625911) (Table 3, Figure 2). In addition, methylation levels at these five CpG sites as measured at birth in INMA and at age 8 years in BAMSE also correlated with YKL-40 levels at age 4 years (Table 3, Figure E3). These CpG associations with YKL-40 levels were confirmed after adjustment for blood cell composition (data not shown).

At each of these five CpG sites, methylation levels obtained by the microarray and those obtained by pyrosequencing were strongly correlated (Spearman correlation coefficients between 0.60 and 0.92; see Figure E4).

- **SNPs and DNA methylation**

Next, we tested associations between the seven SNPs and the five CpG sites that were found to be related to YKL-40 levels in the aforementioned analyses (see Figure E5 for a map of the CpG genomic locations). In INMA, 4-year methylation levels at four of five CpG sites were significantly associated with all seven SNPs after Bonferroni correction (Table 4,
upper part), with alleles linked to lower YKL-40 being associated with higher methylation levels. These genetic associations were remarkably similar when methylation levels from cord blood were analyzed (Table 4, lower part). Based on the effect estimates and the percent variability in methylation levels that they explained, the SNPs rs10399931 and rs4950928, which are in strong LD, showed the strongest associations for all five CpG sites. These two SNPs were also associated with methylation levels at all five CpG sites in BAMSE both at age 4 and 8 years (Table E6).

All the aforementioned five CpG sites showed moderate-to-strong correlation of methylation levels between birth and 4 years in INMA (Spearman correlation coefficients 0.35-0.85), and between 4 and 8 years in BAMSE (0.44-0.88) (Figure E6), in line with the possibility of a consistent genetic control of their methylation levels from birth to age 8 years. These five CpG sites also showed significant associations with CHI3L1 SNPs and significant correlations between their methylation levels at 4 and 8 years of age in the PIAMA study (Table E7 and Figure E6c).

- **Multivariable and mediation analyses**

Given the aforementioned associations, we conducted multivariable analyses to identify independent effects by SNPs and by CpG sites on YKL-40. Among the seven SNPs, in INMA the final backward stepwise regression model included rs4950928 and rs7542294, whereas the final model in BAMSE included rs10399931 and rs10399805. Of note, rs4950928 and rs7542294 are in strong LD with rs10399931 and rs10399805, respectively (Figure E1), indicating that in both cohorts genetic influences on YKL-40 may be driven by these two groups of SNPs in LD. Among the five CpG sites, final stepwise models predicting YKL-40 included only cg07423149 in INMA and only cg17014757 in BAMSE, but it should be noted that methylation levels at the five CpG sites were strongly correlated with each other both in INMA (Figure 2, Spearman correlation coefficients 0.64-0.87) and BAMSE (Figure E3, 0.51-0.83), making it difficult to determine whether a single CpG site or rather a global regional methylation profile was driving the association with YKL-40. To evaluate the components of methylation profiles across these five highly correlated CpG sites, we
conducted a PCA and found that the first component explained up to 79% and up to 75% of variance in INMA and BAMSE, respectively. To evaluate to what extent methylation profiles mediated the relation of \textit{CHI3L1} variation to YKL-40 levels, we completed mediation analyses in INMA (the largest cohort with complete data). In these analyses, we tested the two SNPs (rs4950928 and rs7542294) and the CpG site (cg07423149) identified by the multivariable analyses plus the first component identified from the PCA. In these analyses, cg07423149 mediated 62% of the rs4950928 effects and 43% of the rs7542294 effects on YKL-40 (Table E8). The corresponding percentages for the overall methylation score were only slightly higher (66% and 46%, respectively). These results indicate that \textit{CHI3L1} methylation profiles may mediate a substantial proportion of the effects of \textit{CHI3L1} genetic variation on YKL-40 levels.

**Secondary Analyses: Associations of 3-level biomarkers (SNPs, methylation, and YKL-40) with asthma (Figure 1b)** We also tested all the markers (i.e., variation at seven SNPs, methylation at five CpG sites at age 4 years, and YKL-40 levels at age 4 years) for cross-sectional associations with asthma using the maximum number of available genetic and epigenetic samples in INMA, BAMSE, and PIAMA. Numbers and characteristics of participants included in these analyses are shown in the lower part of Table 1. Table 5 shows results of association analyses with asthma for each of the biomarkers. Neither the SNPs nor the CpG sites showed consistent associations with asthma in INMA, BAMSE, and PIAMA. The lack of association between methylation at any of the CpG sites and asthma was also confirmed after further adjustment for blood cell composition (data not shown). In contrast, YKL-40 levels were associated, albeit weakly, with asthma (Table 5). In meta-analyses, each standard deviation (SD) increase in YKL-40 levels was associated with a borderline 22% increase in the odds for asthma (p=0.08), and participants in the highest YKL-40 tertile had 90% increased odds for asthma compared with subjects in the lowest tertile (p=0.03).
In additional analyses, we confirmed the lack of association of genetic and epigenetic markers with asthma when methylation levels at the other nine CpG sites in \textit{CHI3L1} were analyzed (Table E9) and when SNPs and methylation levels were tested for associations with the comorbidity cluster (asthma, rhinitis, and eczema) that was recently described\textsuperscript{43} in these cohorts (Table E10).

Discussion

This is the first study to integrate genetic and epigenetic regulation of YKL-40 in childhood. We found consistent interrelationships between \textit{CHI3L1} genetic variants, methylation levels at several \textit{CHI3L1} CpG sites, and circulating YKL-40 levels. These associations were replicated in multiple independent cohorts. Taken together with our mediation analysis, these data indicate that methylation levels mediate part of the known effects of genetic variation on YKL-40 levels. We also observed an association of YKL-40 levels, but not \textit{CHI3L1} SNPs or methylation levels, with asthma.

YKL-40 has been proposed as a potential biomarker for a broad range of diseases\textsuperscript{1}. Elevated levels of YKL-40 have been consistently found in the blood and airways of adults with asthma, particularly those with severe and persistent disease\textsuperscript{2, 3, 5-10}. The underlying mechanisms of these associations remain largely to be determined. \textit{In vitro} studies have shown that human bronchial epithelial cells express \textit{CHI3L1} and secrete YKL-40 in response to mechanical stress similar to that experienced during bronchoconstriction\textsuperscript{44}. Furthermore, YKL-40 levels were increased in broncho-alveolar lavage from people with asthma on segmental allergen challenge\textsuperscript{45, 46}. In support of a possible causal involvement of this gene in asthma, \textit{Chi3l1} null mice showed decreased aeroallergen-induced Th2 inflammatory responses in their lungs compared with wild type animals\textsuperscript{10}.

Whereas the link between YKL-40 and asthma has been consistently reported in adults, results from asthma studies in children have been conflicting. Serum YKL-40 levels at birth and in the first 5 years of life were not significantly associated with asthma at age 6 years in
two studies, although positive trends were observed in both cases\textsuperscript{11, 47}. Serum YKL-40 levels were increased in children with therapy-resistant asthma\textsuperscript{12}, but no association between YKL-40 and asthma severity was found in a subsequent study of 61 asthmatic children\textsuperscript{13}. In our larger study, we found children with asthma to have elevated circulating levels of YKL-40, although this association was relatively weak. This may be due to the fact that we studied population-based epidemiological cohorts, in which the prevalence of severe asthma is expected to be quite low.

Circulating levels of YKL-40 have been previously described to be under strong genetic control. In multivariable analyses, we found SNPs from two groups of LD to be independently associated with serum YKL-40. Among them, the two SNPs rs10399931 and rs4950928 fell within a single LD block and were identified as the strongest protein quantitative trait loci (pQTL) in our study, consistent with previous reports\textsuperscript{11, 48}. They are located in proximity of the transcription start site and, because they are in strong LD, it is difficult to dissect their independent effects on YKL-40 levels and/or disease risk. Gene reporter assays for \textit{CHI3L1} promoter haplotypes indicated that both SNPs contributed significant cis-regulatory effects on gene expression in Jurkat cells and that the magnitude of these effects was the strongest for rs10399931\textsuperscript{16}, which was also the SNP with the strongest effects on methylation levels in our study.

Findings from our study indicate that the effects of \textit{CHI3L1} SNPs on circulating YKL-40 may be mediated by methylation levels at \textit{CHI3L1} CpG sites. Several SNPs were associated with methylation levels at five CpG sites located within 1kb of the transcription start site that were, in turn, negatively correlated with YKL-40 levels. It should be noted that the probe for cg17014757 includes rs10399931 and the probe for cg03625911 includes rs7515776. Therefore, the strong associations that were found between methylation levels at these CpG sites and the corresponding SNPs (or other SNPs in LD with them) may be due to allele-specific differences in probe hybridization\textsuperscript{37}. However, similar genetic associations, albeit smaller in magnitude than those observed for cg17014757, were found for CpG sites (i.e., cg13134650, cg07423149, and cg14085262) whose probes do not include known SNPs.
Mediation analyses that were completed both on cg07423149 and on the methylation principal component supported SNP→CpG→YKL-40 causal models. Our findings are in line with a growing body of evidence that points to significant effects of DNA sequence variants across the genome on methylation levels at nearby and distal CpG sites\textsuperscript{23, 25-30}. Previous studies have shown that these methylation-quantitative trait loci (meQTL) can in turn affect gene expression\textsuperscript{25, 49} and are enriched in motifs for DNA-binding factors and DNaseI hypersensitivity regions\textsuperscript{29, 50}. MeQTLs have been shown to be enriched for disease risk variants\textsuperscript{29} and they may ultimately influence disease risk through their effects on methylation patterns\textsuperscript{23, 28}. Similarly, \textit{CHI3L1} meQTLs could influence asthma risk by affecting methylation, gene expression and, in turn, YKL-40 levels. However, in contrast with this scenario, in our study neither \textit{CHI3L1} SNPs nor methylation at \textit{CHI3L1} CpG sites were associated with asthma, although YKL-40 levels tended to be higher in children with asthma. These results have two alternative explanations: either they represent a true negative finding, which would argue against a causal role of \textit{CHI3L1} variants in childhood asthma, or they are due to possible methodological factors that should be taken into account. First, asthma, particularly in the pre-school years, is characterized by a large phenotypic heterogeneity that may impact our ability to capture phenotype-specific genetic associations. Along the same lines, the lack of genetic associations may be explained by \textit{CHI3L1} influencing asthma risk through interactions with environmental factors or through effects on asthma phenotypes that were not included or were under-represented in our study (e.g., adult or severe asthma). It is also possible that our genetic association analyses were under-powered to detect a true signal, although in that case the magnitude of the \textit{CHI3L1} genetic effects would be expected to be relatively small.

Our study has some limitations. Although we did not find any statistical evidence that interrelationships between \textit{CHI3L1} genetic variation, methylation, and YKL-40 levels differed between cases and controls, these stratified analyses had limited sample size, and potential differences by disease status could not be determined conclusively. The methylation array that we used is primarily designed for genomic discovery. Thus, information from a large
proportion of CHI3L1 CpG sites and, in turn, analyses on regional methylation profiles could not be included in our study and will need to be addressed in future studies. Finally, we acknowledge that, by using whole blood for methylation studies, we may have missed contributions of methylation profiles from other tissues (e.g., from the airways), and our results may have been impacted by blood cell composition, particularly in analyses that used cord blood. However, it should be noted that we were able to replicate all the associations between methylation at the five CpG sites and serum YKL-40 after adjustment for blood cell composition, both as estimated by the Houseman method and, in a subset of INMA participants, as directly assessed from blood smears.

In conclusion, in multiple independent cohorts, we found genetic variation in the CHI3L1 gene to be related to both methylation levels in nearby CpG sites and to circulating YKL-40 levels. Our findings indicate that CHI3L1 genetic variation may affect circulating YKL-40 levels by regulating its gene methylation profiles.

Acknowledgments

Pyrosequencing was performed at the University Medical Center Groningen, in the Laboratory of Obstetrics, under the supervision of Dr. Torsten Plösch. We thank Dr. Xavier Basagaña for his helpful advice on mediation analysis.
References


Figure Legends

**Figure 1 title.** Study design for the primary (Figure 1a) and secondary (Figure 1b) analyses

**Figure 1 footnotes.**
G = CHI3L1 genetic variation
M = methylation levels at CHI3L1 CpG sites
P = circulating YKL-40 levels
SNP = single nucleotide polymorphism

**Figure 2 title.** Figures show p values for association with YKL-40 levels (top panel) and correlation matrices (bottom panel) for methylation levels at the 14 CHI3L1 CpG sites in INMA at 4 years of age (Figure 2a) and in BAMSE at 4 years of age (Figure 2b).

**Figure 2 footnotes.** The top panel of each figure shows the –log10 (p value) for the association between YLK40 levels and DNA methylation at each CpG site (the color of the symbol is the color of the co-methylation pattern between that specific CpG and the reference CpG site indicated in black [cg07423149]). The bottom panel of each figure shows the correlation between methylation levels at CpG sites as indicated in the legend. In addition, several annotation tracks are shown (ENSEMBL gene annotation in orange; SNPs from GWAS studies (rs4950928) in black; and DNAse hypersensitivity regions in blue).

See reference52
Table 1. Numbers and basic characteristics of participants included in this study.

<table>
<thead>
<tr>
<th>Data available for primary analyses*</th>
<th>INMA</th>
<th>BAMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>203</td>
<td>230</td>
</tr>
<tr>
<td>Male sex: N (%)</td>
<td>102 (50%)</td>
<td>126 (55%)</td>
</tr>
<tr>
<td>Age in months at the 4-year survey: mean (SD)</td>
<td>53.2 (2.3)</td>
<td>47.9 (1.7)</td>
</tr>
<tr>
<td>Body mass index at 4 years: mean (SD)</td>
<td>16.2 (1.5)</td>
<td>16.2 (1.3)</td>
</tr>
<tr>
<td>Asthma at 4 years: N (%)</td>
<td>33 (16%)</td>
<td>88 (38%)</td>
</tr>
<tr>
<td>YKL-40 levels at 4 years: N</td>
<td>203</td>
<td>230</td>
</tr>
<tr>
<td>Methylation data at 4 years: N</td>
<td>203</td>
<td>230</td>
</tr>
<tr>
<td>Methylation data at birth: N</td>
<td>189</td>
<td>-</td>
</tr>
<tr>
<td>Methylation data at 8 years: N</td>
<td>-</td>
<td>225</td>
</tr>
<tr>
<td>GWAS data: N</td>
<td>172</td>
<td>78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data available for association analyses with asthma – secondary analyses**</th>
<th>INMA</th>
<th>BAMSE</th>
<th>PIAMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS data: N asthma / N total</td>
<td>58 / 336</td>
<td>171 / 467</td>
<td>69 / 1602</td>
</tr>
<tr>
<td>Male sex: N (%)</td>
<td>179 (53%)</td>
<td>260 (56%)</td>
<td>800 (50%)</td>
</tr>
<tr>
<td>Age in months at the 4-year survey: mean (SD)</td>
<td>51.4 (3.3)</td>
<td>48.1 (1.9)</td>
<td>48.3 (1.1)</td>
</tr>
<tr>
<td>Methylation data at 4 years: N asthma / N tot</td>
<td>33 / 203</td>
<td>88 / 229</td>
<td>13 / 193</td>
</tr>
<tr>
<td>YKL-40 levels at 4 years: N asthma / N tot</td>
<td>33 / 203</td>
<td>88 / 229</td>
<td>N/A***</td>
</tr>
</tbody>
</table>

* Children with complete data on CHI3L1 methylation and YKL-40 levels at age 4 years.

** Children with complete data on asthma, covariates, and at least one of the three biomarkers at age 4 years (i.e., CHI3L1 SNPs, CHI3L1 methylation, or YKL-40 levels).

*** Not available.
SD = standard deviation.

GWAS = genome-wide association study.
Table 2. Associations between SNPs in the CHI3L1 genomic region and standardized protein levels of YKL-40 at age 4 years in the INMA cohort (N=172). Only SNPs significant after Bonferroni correction are shown (data for all SNPs are shown in Table E3).

<table>
<thead>
<tr>
<th>SNP**</th>
<th>Location chr1 (hg19)</th>
<th>Effect allele</th>
<th>Other allele</th>
<th>Effect allele frequency</th>
<th>HWE P value*</th>
<th>Linear regression coefficient^</th>
<th>(95% CI)</th>
<th>P value</th>
<th>R^2 ***</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2886117</td>
<td>203168881</td>
<td>A</td>
<td>G</td>
<td>15.7</td>
<td>0.77</td>
<td>0.61</td>
<td>(0.326-0.894)</td>
<td>4*10^-5</td>
<td>0.10</td>
</tr>
<tr>
<td>rs10399931</td>
<td>203156080</td>
<td>T</td>
<td>C</td>
<td>19.2</td>
<td>0.33</td>
<td>-0.68</td>
<td>(-0.941,-0.420)</td>
<td>7*10^-7</td>
<td>0.14</td>
</tr>
<tr>
<td>rs10399805</td>
<td>203155998</td>
<td>A</td>
<td>G</td>
<td>15.4</td>
<td>0.77</td>
<td>0.62</td>
<td>(0.331-0.901)</td>
<td>3*10^-5</td>
<td>0.10</td>
</tr>
<tr>
<td>rs4950928</td>
<td>203155882</td>
<td>G</td>
<td>C</td>
<td>16.0</td>
<td>0.08</td>
<td>-0.82</td>
<td>(-1.108,-0.536)</td>
<td>6*10^-4</td>
<td>0.16</td>
</tr>
<tr>
<td>rs7542294</td>
<td>203151176</td>
<td>A</td>
<td>G</td>
<td>16.9</td>
<td>0.42</td>
<td>0.64</td>
<td>(0.360-0.914)</td>
<td>1*10^-5</td>
<td>0.11</td>
</tr>
<tr>
<td>rs2791718</td>
<td>203141424</td>
<td>A</td>
<td>C</td>
<td>12.8</td>
<td>0.32</td>
<td>0.61</td>
<td>(0.286-0.924)</td>
<td>2*10^-4</td>
<td>0.08</td>
</tr>
<tr>
<td>rs10920576</td>
<td>203129179</td>
<td>T</td>
<td>C</td>
<td>12.8</td>
<td>0.32</td>
<td>0.72</td>
<td>(0.406-1034)</td>
<td>1*10^-4</td>
<td>0.11</td>
</tr>
</tbody>
</table>

In bold: associations significant after Bonferroni correction.

^Coefficient from additive models predicting standardized YKL-40 levels.

*Hardy-Weinberg equilibrium P value.

**SNPs are shown according to position in the gene.

*** Percent variability in YKL-40 explained by the SNP.

SNP = single nucleotide polymorphism.
HWE = Hardy-Weinberg equilibrium.
95% CI = 95% confidence interval.
### Table 3. Spearman correlation coefficients between methylation levels at CpG sites at various ages and YKL-40 levels in INMA and BAMSE.

<table>
<thead>
<tr>
<th>CpG sites</th>
<th>INMA methylation at 4y</th>
<th></th>
<th>INMA methylation at birth</th>
<th></th>
<th>BAMSE methylation at 8y</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 203</td>
<td>N = 230</td>
<td>N = 189</td>
<td>N = 225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cg13134650</td>
<td>203156765</td>
<td>-0.36 $1 \times 10^{-7}$</td>
<td>-0.13 0.06</td>
<td>-0.25 $5 \times 10^{-4}$</td>
<td>-0.22 $7 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>cg19081101</td>
<td>203156626</td>
<td>-0.35 $3 \times 10^{-7}$</td>
<td>-0.06 0.38</td>
<td>-0.15 0.04</td>
<td>0.01 0.90</td>
<td></td>
</tr>
<tr>
<td>cg02097014</td>
<td>203156375</td>
<td>-0.21 0.003</td>
<td>-0.11 0.10</td>
<td>-0.09 0.24</td>
<td>-0.12 0.08</td>
<td></td>
</tr>
<tr>
<td>cg07423149</td>
<td>203156247</td>
<td>-0.53 $3 \times 10^{-8}$</td>
<td>-0.27 $2 \times 10^{-6}$</td>
<td>-0.30 $3 \times 10^{-3}$</td>
<td>-0.26 $6 \times 10^{-3}$</td>
<td></td>
</tr>
<tr>
<td>cg17014757^</td>
<td>203156098</td>
<td>-0.44 $7 \times 10^{-11}$</td>
<td>-0.34 $2 \times 10^{-7}$</td>
<td>-0.31 $2 \times 10^{-3}$</td>
<td>-0.35 $4 \times 10^{-3}$</td>
<td></td>
</tr>
<tr>
<td>cg14085282</td>
<td>203155939</td>
<td>-0.44 $3 \times 10^{-8}$</td>
<td>-0.21 0.001</td>
<td>-0.24 $7 \times 10^{-4}$</td>
<td>-0.22 $7 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>cg03625911^^</td>
<td>203155738</td>
<td>-0.42 $4 \times 10^{-10}$</td>
<td>-0.18 0.006</td>
<td>-0.21 0.004</td>
<td>-0.19 $7 \times 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>cg17000774</td>
<td>203154457</td>
<td>-0.04 0.57</td>
<td>0.00 0.97</td>
<td>-0.04 0.62</td>
<td>0.03 0.65</td>
<td></td>
</tr>
<tr>
<td>cg11196333</td>
<td>203154371</td>
<td>-0.04 0.62</td>
<td>-0.14 0.04</td>
<td>-0.08 0.26</td>
<td>-0.16 0.01</td>
<td></td>
</tr>
<tr>
<td>cg05526099</td>
<td>203152296</td>
<td>0.00 0.95</td>
<td>-0.02 0.80</td>
<td>-0.05 0.51</td>
<td>0.02 0.77</td>
<td></td>
</tr>
<tr>
<td>cg04361579</td>
<td>203152047</td>
<td>0.00 0.95</td>
<td>-0.16 0.01</td>
<td>-0.06 0.44</td>
<td>0.00 0.99</td>
<td></td>
</tr>
<tr>
<td>cg20707774</td>
<td>203149828</td>
<td>-0.10 0.15</td>
<td>-0.20 0.003</td>
<td>-0.04 0.56</td>
<td>-0.11 0.09</td>
<td></td>
</tr>
<tr>
<td>cg14165900</td>
<td>203148901</td>
<td>-0.02 0.74</td>
<td>0.05 0.44</td>
<td>0.04 0.61</td>
<td>0.00 0.97</td>
<td></td>
</tr>
<tr>
<td>cg15490070</td>
<td>203148132</td>
<td>-0.33 $2 \times 10^{-6}$</td>
<td>-0.05 0.47</td>
<td>-0.13 0.08</td>
<td>-0.05 0.49</td>
<td></td>
</tr>
</tbody>
</table>

In bold: correlations significant both in the discovery population (INMA 4 yr) after Bonferroni correction and in the replication population (BAMSE 4 yr, 1-sided α=0.05).

Associations in bold confirmed after adjustment for blood cell composition.
^ probe overlaps known SNP with MAF > 10% (rs10399931).

^^ probe overlaps known SNP with MAF > 10% (rs7515776).
Table 4. Associations between the seven SNPs and methylation levels at the five CpG sites that were significantly related to YKL-40 levels in the INMA cohort. N=172 (methylation levels at 4 years) in the upper panel and N=161 (methylation levels at birth) in the lower panel.

<table>
<thead>
<tr>
<th>YEAR 4</th>
<th>cg13134650</th>
<th>cg07423149</th>
<th>cg17014757*</th>
<th>cg14085262</th>
<th>cg03625911**</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP (effect allele)</td>
<td>Regression coefficient*</td>
<td>95% CI</td>
<td>P value</td>
<td>R^2***</td>
<td>Regression coefficient*</td>
</tr>
<tr>
<td>rs2886117 (A)</td>
<td>-1.01</td>
<td>-1.83, -0.19</td>
<td>0.02</td>
<td>0.03</td>
<td>-3.59</td>
</tr>
<tr>
<td>rs10399951 (T)</td>
<td>2.42</td>
<td>1.74, 3.09</td>
<td>0.20</td>
<td>0.03</td>
<td>6.52</td>
</tr>
<tr>
<td>rs10399885 (A)</td>
<td>-1.04</td>
<td>-1.86, -0.22</td>
<td>0.01</td>
<td>0.03</td>
<td>-3.76</td>
</tr>
<tr>
<td>rs4950928 (G)</td>
<td>2.49</td>
<td>1.71, 3.26</td>
<td>2*10^-8</td>
<td>0.17</td>
<td>6.66</td>
</tr>
<tr>
<td>rs7542294 (A)</td>
<td>-1.22</td>
<td>-2.01, -0.42</td>
<td>0.003</td>
<td>0.05</td>
<td>-3.54</td>
</tr>
<tr>
<td>rs2791178 (A)</td>
<td>-1.09</td>
<td>-2.01, -0.17</td>
<td>0.02</td>
<td>0.03</td>
<td>-3.86</td>
</tr>
<tr>
<td>rs10920776 (T)</td>
<td>-1.23</td>
<td>-2.14, -0.31</td>
<td>0.009</td>
<td>0.04</td>
<td>-4.02</td>
</tr>
<tr>
<td>SNP (effect allele)</td>
<td>BIRTH (cg13134650) Regression coefficient</td>
<td>95% CI</td>
<td>P value</td>
<td>R^2</td>
<td>BIRTH (cg07423149) Regression coefficient</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>-----</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>rs2886117 (A)</td>
<td>-0.55</td>
<td>-1.43</td>
<td>0.33</td>
<td>0.22</td>
<td>-1.88</td>
</tr>
<tr>
<td>rs10399931 (T)</td>
<td>2.02</td>
<td>1.25</td>
<td>2.79</td>
<td>7*10^-4</td>
<td>0.12</td>
</tr>
<tr>
<td>rs10399805 (A)</td>
<td>-0.59</td>
<td>-1.47</td>
<td>0.30</td>
<td>0.19</td>
<td>-1.91</td>
</tr>
<tr>
<td>rs4650928 (G)</td>
<td>1.80</td>
<td>0.91</td>
<td>2.68</td>
<td>9*10^-4</td>
<td>0.08</td>
</tr>
<tr>
<td>rs7542294 (A)</td>
<td>-0.66</td>
<td>-1.53</td>
<td>0.21</td>
<td>0.13</td>
<td>-1.97</td>
</tr>
<tr>
<td>rs2791718 (A)</td>
<td>-0.97</td>
<td>-1.94</td>
<td>0.01</td>
<td>0.05</td>
<td>-2.10</td>
</tr>
<tr>
<td>rs10920576 (T)</td>
<td>-1.17</td>
<td>-2.14</td>
<td>0.09</td>
<td>0.02</td>
<td>-2.17</td>
</tr>
</tbody>
</table>

^Coefficient from additive models predicting % methylation levels.

* probe overlaps known SNP with MAF > 10% (rs10399931).

** probe overlaps known SNP with MAF > 10% (rs7515776, which is in nearly complete LD with rs10399805).

^^ association with a SNP overlapped by the CpG probe (or with a SNP in strong LD [r^2>0.7] with the overlapped SNP).

*** Percent variability in methylation levels explained by the SNP.

In bold: associations significant after Bonferroni correction.

SNP = single nucleotide polymorphism.

95% CI = 95% confidence interval.
Table 5. Associations of CHI3L1 SNPs, CHI3L1 methylation levels, and YKL-40 levels with asthma in the study cohorts. Complete biomarker data were available for INMA and BAMSE. No YKL-40 levels were available for PIAMA. All models adjusted for sex, age, and BMI.

<table>
<thead>
<tr>
<th>SNPs (effect allele)</th>
<th>INMA</th>
<th>BAMSE</th>
<th>PIAMA</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>adjOR (95% CI)</td>
<td>P value</td>
<td>adjOR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>rs2886117 (A)</td>
<td>1.30 (0.774-2.174)</td>
<td>0.32</td>
<td>0.85 (0.552-1.294)</td>
<td>0.44</td>
</tr>
<tr>
<td>rs10399931 (T)*%</td>
<td>0.55 (0.305-0.992)</td>
<td>0.05</td>
<td>1.27 (0.919-1.748)</td>
<td>0.15</td>
</tr>
<tr>
<td>rs10399805 (A)</td>
<td>1.40 (0.829-2.354)</td>
<td>0.21</td>
<td>0.75 (0.482-1.178)</td>
<td>0.21</td>
</tr>
<tr>
<td>rs4950928 (G)*%</td>
<td>0.50 (0.250-0.986)</td>
<td>0.05</td>
<td>1.44 (1.014-2.042)</td>
<td>0.04</td>
</tr>
<tr>
<td>rs7542294 (A)</td>
<td>1.12 (0.658-1.909)</td>
<td>0.67</td>
<td>0.81 (0.540-1.219)</td>
<td>0.31</td>
</tr>
<tr>
<td>rs2791718 (A)*%</td>
<td>0.66 (0.331-1.295)</td>
<td>0.22</td>
<td>0.68 (0.422-1.094)</td>
<td>0.11</td>
</tr>
<tr>
<td>rs10920576 (T)</td>
<td>0.65 (0.338-1.253)</td>
<td>0.20</td>
<td>0.67 (0.435-1.046)</td>
<td>0.08</td>
</tr>
</tbody>
</table>
### Methylation levels

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Methylation (N)</th>
<th>Effect (95% CI)</th>
<th>p-value</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg13134650&lt;sup&gt;+&lt;/sup&gt;</td>
<td>625&lt;sup&gt;^^&lt;/sup&gt;</td>
<td>0.95 (0.837-1.081)</td>
<td>0.44</td>
<td>1.07 (0.966-1.187)</td>
</tr>
<tr>
<td>cg07423149&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.01 (0.945-1.075)</td>
<td>0.81</td>
<td>0.99 (0.951-1.029)</td>
<td>0.59</td>
</tr>
<tr>
<td>cg17014757&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.98 (0.941-1.021)</td>
<td>0.33</td>
<td>1.01 (0.985-1.036)</td>
<td>0.42</td>
</tr>
<tr>
<td>cg14085262&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.02 (0.952-1.090)</td>
<td>0.59</td>
<td>0.96 (0.926-1.004)</td>
<td>0.08</td>
</tr>
<tr>
<td>cg03625911&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.04 (0.951-1.128)</td>
<td>0.42</td>
<td>0.98 (0.932-1.026)</td>
<td>0.37</td>
</tr>
</tbody>
</table>

### Standardized YKL-40 levels

<table>
<thead>
<tr>
<th>Effect for 1-SD increase in YLK-40</th>
<th>N = 432&lt;sup&gt;^^^&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect for 1-SD increase in YLK-40</td>
<td>1.27 (0.861-1.860)</td>
</tr>
</tbody>
</table>

#### Medium vs Low YKL40 tertile

<table>
<thead>
<tr>
<th>Effect for 1-SD increase in YLK-40</th>
<th>N = 432&lt;sup&gt;^^^&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect for 1-SD increase in YLK-40</td>
<td>2.62 (0.927-7.377)</td>
</tr>
</tbody>
</table>

#### High vs Low YKL40 tertile

<table>
<thead>
<tr>
<th>Effect for 1-SD increase in YLK-40</th>
<th>N = 432&lt;sup&gt;^^^&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect for 1-SD increase in YLK-40</td>
<td>2.36 (0.820-6.773)</td>
</tr>
</tbody>
</table>

<sup>+</sup> N asthma / N total: INMA (58 / 336); BAMSE (171 / 467); PIAMA (69 / 1602)

<sup>^^</sup> N asthma / N total: INMA (33 / 203); BAMSE (88 / 229); PIAMA (13 / 193).
^^^ N asthma / N total: INMA (33 / 203); BAMSE (88 / 229)

*imputed in BAMSE (minimal quality of imputation >0.77).

%imputed in PIAMA (minimal quality of imputation >0.969).

* OR for CpG sites express effects for 1% increase in their methylation levels.

N/A: not available.

SNP = single nucleotide polymorphism.

adjOR = adjusted odds ratio.

95% CI = 95% confidence interval.
**PRIMARY ANALYSES**
To determine whether *CHI3L1* methylation levels mediate the association of *CHI3L1* SNPs to YKL-40 levels

**AVAILABLE DATA**

- **Birth**
  - *CHI3L1* methylation
    - INMA (N=189)

- **Year 4**
  - *CHI3L1* methylation Circulating YKL-40 levels
    - INMA / BAMSE (N=433)

- **Year 8**
  - *CHI3L1* methylation
    - BAMSE (N=225)

**CHI3L1 SNPs**
INMA / BAMSE (N=250)

**ANALYTICAL APPROACH**

Diagram showing the relationships:

1. G → M → P
2. G → M → P
3. G → M → P

Question mark (?) indicating further analysis or hypothesis.
SECONDARY ANALYSES
To assess whether \textit{CHI3L1} SNPs and methylation levels that regulate YKL-40 are associated with asthma in early childhood

AVAILABLE DATA

\textit{CHI3L1} SNPs
INMA (336) / BAMSE (467) / PIAMA (1602)
(tot N=2405)

\textit{CHI3L1} methylation at year 4
INMA (203) / BAMSE (229) / PIAMA (193)
(tot N=625)

Circulating YKL-40 levels at year 4
INMA (203) / BAMSE (229)
(tot N=432)

\textbf{Analytical Approach}

\begin{tikzpicture}
  \node (A) at (0,0) {G};
  \node (B) at (2,0) {M};
  \node (C) at (4,0) {P};
  \node (D) at (2,-2) {?};
  \node (E) at (2,-4) {Asthma at year 4};

  \draw[->, thick, blue] (A) -- (B);
  \draw[->, thick, blue] (B) -- (C);
  \draw[->, thick, blue] (A) -- (D);
  \draw[->, thick, blue] (B) -- (D);
  \draw[->, thick, blue] (C) -- (D);

  \node [draw=black, fill=black!10] at (2,-4) {Asthma at year 4};
\end{tikzpicture}