1 **Genetic and epigenetic regulation of YKL-40 in childhood**

2 Stefano Guerra[1-4], MD, PhD; Erik Melén[5], MD, PhD; Jordi Sunyer[1-3,6], MD, PhD; Cheng-Jian
3 Xu[7-9], PhD; Iris Lavi[1-3], PhD; Marta Benet[1-3], BStat; Mariona Bustamante[1-3,10], PhD; Anne-Elie
4 Carsin[1-3,6], MS; Carlota Dobaño[11], PhD; Mònica Guxens[1-3,12], MD, PhD; Christina Tischer[1-3],
5 PhD; Martine Vrijheid[1-3], PhD; Inger Kull[13], RN, PhD; Anna Bergström[5], PhD; Ashish
6 Kumar[5,14-15], MSc; Cilla Söderhäll[16], PhD; Ulrike Gehring[17], PhD; Dorieke J. Dijkstra[18], M.
7 Sc.; Pieter van der Vlies[9], BSc; Magnus Wickman[5], MD, PhD; Jean Bousquet[19], MD, PhD;
8 Dirkje S Postma[7-8], MD, PhD; Josep M Anto[1-3,6], MD, PhD; and Gerard H Koppelman[8,20], MD,
9 PhD

10

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

31 18. University of Groningen, University Medical Center Groningen, Department of Obstetrics and Gynecology,
32 The Netherlands
33 19. University Hospital Montpellier; and Respiratory and Environmental Epidemiology Team, INSERM 1018,
34 CESP Centre, Villejuif, France
35 20. University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Department of
36 Pediatric Pulmonology and Pediatric Allergology, The Netherlands
37
38 *Corresponding author:*
39 Stefano Guerra, MD, PhD
40 ISGlobal CREAL, Doctor Aiguader 88, 08003 Barcelona, Spain; and
41 Asthma and Airway Disease Research Center, 1501 N Campbell Ave, Tucson, AZ 85724,
42 USA
43 Email: stefano@email.arizona.edu
44
45 This study was supported by grant agreement number 261357 (Mechanisms of the
46 Development of ALLergy - MeDALL) from the EU Seventh Framework Programme.
47
48 Additional funding for individual cohorts was provided by:
49 - For INMA: Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; PI041436;
50 PI081151 incl. FEDER funds; MS13/00054), Generalitat de Catalunya-CIRIT 1999SGR
51 00241, Fundació La marató de TV3 (090430)
52 - For BAMSE: The Swedish Heart-Lung Foundation, The Swedish Research Council,
53 Freemason Child House Foundation in Stockholm, Stockholm County Council (ALF),
54 Swedish Foundation for Strategic Research (SSF, RBc08-0027, EpiGene project) and the
55 Strategic Research Programme (SFO) in Epidemiology at Karolinska Institutet
56 - For PIAMA: The Netherlands Organization for Health Research and Development; The
57 Netherlands Organization for Scientific Research; The Netherlands Lung Foundation (AF
58 4.1.14.001); The Netherlands Ministry of Spatial Planning, Housing, and the Environment;
59 and The Netherlands Ministry of Health, Welfare, and Sport.

60 **Abstract**

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

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

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

74 *Results:* YKL-40 levels were significantly associated with seven SNPs and with methylation
75 at five CpG sites. Consistent associations between these seven SNPs (particularly
76 rs10399931 and rs4950928) and five CpG sites were observed. Alleles linked to lower YKL-
77 40 levels were associated with higher methylation levels. Participants with high YKL-40
78 levels (defined as the highest YKL-40 tertile) had increased odds for asthma as compared
79 with subjects with low YKL-40 levels [meta-analyzed adjusted odds ratio (adjOR): 1.90, 1.08-
80 3.36]. In contrast, neither SNPs nor methylation levels at CpG sites in *CHI3L1* were
81 associated with asthma.

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

85

86 **Clinical Implications**

87 Methylation levels at *CHI3L1* CpG sites mediate part of the effects of *CHI3L1* genetic

88 variation on circulating levels of YKL-40, but they are not associated with childhood asthma.

89

90 **Capsule Summary**

91 Circulating YKL-40 and variation in its encoding gene *CHI3L1* have been associated with

92 asthma. We found that methylation levels at *CHI3L1* CpG sites partly mediated *CHI3L1*

93 genetic effects on circulating YKL-40, although they were not associated with childhood

94 asthma.

95

96 **Key words**

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

98

99 **Abbreviations**

100 adjOR: adjusted odds ratio

101 BMI: body mass index

102 *CHI3L1: chitinase-3-like-1*

103 GWAS: Genome-Wide Association Study

104 LD: linkage disequilibrium

105 MAF: minor allele frequency

106 meQTL: methylation quantitative trait locus

107 PCA: principal component analysis

108 pQTL: protein quantitative trait locus

109 SNP: single nucleotide polymorphism

110 **Introduction**

111 YKL-40, a chitinase-like protein, is upregulated in asthma, cancer, and other diseases
112 characterized by inflammation and tissue remodeling[1]. In adults, YKL-40 levels are
113 increased in the blood and lungs of people with asthma[2-4] and correlate with lung function
114 deficits, disease severity, and persistence[2-10]. However, to date, findings on YKL-40 and
115 asthma in childhood have been conflicting[11-13].

116 The mechanisms linking YKL-40 with asthma remain to be determined, although *in vitro* and
117 animal studies support its role in Th2 adaptive immune responses[14]. The association
118 between genetic variation in *chitinase-3-like-1* (*CHI3L1*, the gene encoding YKL-40) and
119 asthma susceptibility/severity suggests a possibly causal link[11, 15-18]. A group of four single
120 nucleotide polymorphisms (SNPs) tagged by rs4950928 were first associated with asthma in
121 three of four tested populations, with the major alleles conferring increased risk[11]. In the
122 same study, serum YKL-40 levels were found to be a highly heritable quantitative trait in the
123 general population and to be directly associated with the same C allele at rs4950928 that
124 was associated with asthma risk. Subsequent studies showed that another SNP,
125 rs10399931, which is in strong linkage disequilibrium (LD) with rs4950928, had similar, if not
126 stronger, effects on gene expression[16], plasma YKL-40 levels[19], and asthma[16]. However,
127 whereas the association between *CHI3L1* genetic variation and YKL-40 levels has been
128 conclusively established, the relation of *CHI3L1* variation to asthma remains controversial
129 because other reports[20-22], including a large study of over 6500 Danish adults[20], failed to
130 replicate the aforementioned genetic associations with asthma.

131 The identification of the mechanisms by which genetic variation in *CHI3L1* regulates YKL-40
132 levels may have important implications for the understanding of the potential impact of this
133 gene on human disease. Epigenetic regulation of gene expression is one of the possible
134 mechanisms by which genetic variation can affect protein levels and disease susceptibility[23],
135 including childhood respiratory diseases[24]. DNA sequence variants across the genome have
136 been shown to have *cis*- (and to less extent *trans*-) effects on methylation levels at specific
137 CpG sites[23, 25-30]. Yet, to date no study has addressed DNA methylation as a possible

138 intermediary mechanism of the relation between *CHI3L1* genetic variation and upregulated
139 YKL-40 protein levels.
140 The primary goal of the present study was to determine whether methylation levels at
141 *CHI3L1* CpG sites mediate the association of *CHI3L1* genetic variation to YKL-40 levels in
142 blood. In secondary analyses, we also sought to assess whether the *CHI3L1* genotype and
143 methylation levels that regulate YKL-40 levels are associated with asthma in early childhood.
144
145
146 **Methods**
147 Study populations and design. This study was part of the Mechanisms of the Development of
148 ALLergy (MeDALL) project[31], which included analyses on YKL-40 as an *a priori* biomarker
149 candidate. The design and available data for the present study are summarized in Figure 1.
150 Primary analyses (Figure 1a) included molecular and phenotypic data from 433 participants
151 who were 4 years of age from the Spanish INfancia y Medio Ambiente[32] (INMA, N=203) and
152 the Swedish Barn/Children, Allergy, Milieu, Stockholm, Epidemiological survey[33] (BAMSE,
153 N=230) birth cohorts. These children were selected from their original cohorts for epigenetic
154 and YKL-40 studies in MeDALL based on the following sampling strategy: among
155 participants who provided paired DNA samples at two time points (birth and 4 years for
156 INMA, 4 and 8 years for BAMSE), we selected children who had one or more of three
157 diagnoses (asthma, eczema, and allergic rhinitis) and a similar number of randomly selected
158 control subjects (none of the three diagnoses, nested case-control design). Of them, 172
159 INMA and 78 BAMSE participants also had available Genome-Wide Association Study
160 (GWAS) data.
161 In secondary analyses of association with asthma (Figure 1b). in addition to INMA and
162 BAMSE, we also included 4-year-old participants from the Prevention and Incidence of
163 Asthma and Mite Allergy (PIAMA) study to increase statistical power[34, 35]. No serum YKL-40
164 levels were available for PIAMA, but epigenetic studies were completed on 193 PIAMA
165 participants who were selected based on the same sampling strategy described previously.

166 For analyses of genetic association, we used data from all participants from the three
167 cohorts who had both GWAS data and asthma information (N = 336 for INMA, 467 for
168 BAMSE, and 1602 for PIAMA; total N=2405).
169 A detailed description of the three birth cohorts and additional information on molecular
170 assays are provided in the online repository. For all cohorts, parents gave written informed
171 consent, and the local ethics review boards approved the performed studies and procedures.
172
173 Definition of asthma. In MeDALL, an asthma definition[36] was used that included positive
174 answers to at least two of the three following questions: (1) "Has your child ever been
175 diagnosed by a doctor as having asthma?"; (2) "Has your child taken any medicines for
176 asthma (including inhalers, nebulizers, tablets, or liquid medicines) or for breathing
177 difficulties (chest tightness, shortness of breath) in the last 12 months?"; and (3) "Has your
178 child had wheezing or whistling in their chest at any time in the last 12 months?"; or "Has
179 your child had breathing difficulties (chest tightness, shortness of breath) in the last 12
180 months?". This definition was used for BAMSE and PIAMA. Because information on doctor-
181 diagnosed asthma was not available in INMA, for this cohort the asthma definition included a
182 positive response to the two remaining questions.
183
184 Molecular data.
185 *DNA methylation* assays *at CHI3L1 CpG sites*. Epigenome-wide analysis scans of paired
186 whole peripheral blood DNA methylation samples at birth and age 4 (INMA), and at age 4 and
187 8 (BAMSE and PIAMA) were generated using the Illumina Infinium HumanMethylation 450
188 BeadChip assays. All samples were quality checked as described in the online supplement.
189 The Illumina 450K included 485577 assays. During processing, the probes on sex
190 chromosomes, the probes that mapped on multi-loci, 65 SNP assays, and the probes
191 containing SNPs at the target CpG sites with a minor allele frequency (MAF>10%) were
192 excluded. This resulted in a total of 439306 CpG sites. Among them, 14 CpG sites were
193 annotated to the *CHI3L1* gene and analyzed in the present study (Table E1). Two of them

194 (cg17014757 and cg03625911) have probes that include known SNPs with MAF greater than
195 5% in European populations, as reported by Chen et al[37]. The probe of cg17014757 includes
196 rs10399931, and the probe of cg03625911 includes rs7515776, which according to the CEU
197 panel of the 1000 Genomes project, is in nearly complete LD with rs10399805. These two
198 CpG sites were included in statistical analyses, but associations of their methylation levels
199 with overlapped SNPs or with other SNPs in LD with the overlapped SNPs (rs10399931 and
200 rs4950928 for cg17014757; and rs2886117, rs10399805, and rs7542294 for cg03625911)
201 were identified in tables to caution interpretation. Methylation levels (beta values, β) at a given
202 CpG site were derived from the ratio of the methylated probe intensity and overall intensity
203 (sum of methylated and unmethylated probe intensities). Thus, β is equal to $M/(U + M + \alpha)$,
204 where M is intensity of the methylated probe, U is the intensity of the unmethylated probe, and
205 α is the constant offset with the default value of 100. The intensity has been corrected for type
206 I and type II probe differences and normalized by using the "dasen" method in the wateRmelon
207 R package[38]. Results from analyses on methylation levels were also confirmed after
208 adjustment for blood cell-type composition as predicted by the Houseman algorithm[39], and, in
209 a subset of 116 INMA participants at year 4, after adjustment for available differential cell
210 counts obtained directly from microscopic inspection of blood smears. The methylation signal
211 of the top five CpG sites obtained by the Illumina microarray was also validated by
212 pyrosequencing in whole blood DNA from 96 subjects participating in the PIAMA study (see
213 online repository, Table E11, and Figure E4).
214 *YKL-40 measurements*. Circulating levels of YKL-40 were measured in serum (INMA) and
215 plasma (BAMSE) samples at age 4 years using a commercially available enzyme-linked
216 immunosorbent assay kit (Quantikine Human CHI3L1 immunoassay by R&D, Abingdon,
217 UK).
218 *GWAS*. Genome-wide genotyping had been previously completed for the three cohorts,
219 INMA, BAMSE, and PIAMA, using various platforms (see online repository). For the present
220 study, in the INMA cohort we tested 68 SNPs that were genotyped from the genomic region
221 surrounding (+/- 50kb) *CHI3L1* (hg19: chr1:203,098,059-203,205,922) using the

222 HumanOmni1-Quad Beadchip (Illumina, San Diego, CA, USA). SNPs that were found not to
223 be in Hardy-Weinberg equilibrium using exact tests[40] were filtered (see additional methods
224 for GWAS assays in the online repository). SNPs that were significantly associated with
225 YKL-40 in INMA were also tested in BAMSE and PIAMA.
226
227 Analytical approach and statistical analysis.
228 For the primary goal, we analyzed interrelationships between multi-level biomarkers (SNPs,
229 methylation at CpG sites, and YKL-40 levels) in INMA and BAMSE according to a stepwise
230 approach (Figure 1a). Because results of these analyses did not significantly differ between
231 cases and controls, they are presented with no stratification by disease. YKL-40 levels were
232 log-transformed and then standardized within each cohort by subtracting the mean and
233 dividing the result by the standard deviation as done in previous multi-cohort studies[41].
234 Discovery analyses were completed in INMA with a conservative Bonferroni correction, and
235 replication analyses in BAMSE with a one-tailed $\alpha = 0.05$ (i.e., only associations with the
236 same direction of effect between the two cohorts were tested for significance). First,
237 associations between *CHI3L1* SNPs and YKL-40 levels were studied. In INMA, 68 SNPs
238 were tested in linear regressions predicting YKL-40 protein levels according to additive
239 genetic models. Because INMA and BAMSE used different GWAS platforms, in BAMSE we
240 used genotyped or imputed SNPs for replication of INMA results as appropriate. Second,
241 methylation levels were tested for correlation with YKL-40 levels using Spearman correlation
242 coefficients. Third, the SNPs and CpG sites that were found to be related to YKL-40 levels
243 were tested for association with each other using robust regressions to reduce the impact of
244 potential outlier observations (see online repository for additional information).
245 To determine what SNPs and CpG sites were independently related to YKL-40 protein
246 levels, multivariable regression models with backward stepwise variable selection were
247 used. A final mediation analysis was completed in INMA using the R package "mediation"[42]
248 to estimate the effects by the aforementioned independent SNPs on YKL-40 that were
249 mediated by methylation. For this analysis, we used both methylation levels at the CpG site

250 (cg07423149) that was identified by the above backward stepwise selection as well as the
251 first principal component obtained by a principal component analysis (PCA) on methylation
252 levels at all the five CpG sites that were associated with YKL-40.
253 For our secondary analyses (Figure 1b), we performed separate multiple logistic regression
254 models testing the association of SNPs, CpG methylation levels, and YKL-40 levels with
255 asthma with adjustment for sex, age, and body mass index (BMI) at 4 years, as these
256 demographic factors may affect methylation and YKL-40 levels as well as asthma risk. To
257 test for non-linearity of effects, YKL-40 levels were also used as tertiles, and asthma risks
258 were compared between subjects having medium and high YKL-40 levels and subjects
259 having low YKL-40 levels.
260
261
262 **Results**
263 Table 1 shows characteristics of children included in the primary and secondary analyses.
264 Table E2 compares demographic characteristics of participants from the three cohorts that
265 were or were not included in the present study.
266
267 ***Primary Analyses: Relation between CHI3L1 SNPs, methylation, and YKL-40 levels***
268 ***(Figure 1a)***
269 - ***SNPs and YKL-40 protein levels***
270 Among the 68 tested SNPs, after Bonferroni correction, seven SNPs were found to be
271 significantly associated with YKL-40 levels in INMA (Table E3 for complete analysis, Table 2
272 for significant associations). Overall, minor alleles were associated with higher YKL-40
273 levels, with the exceptions of rs10399931 and rs4950928. These seven SNPs explained
274 individually up to 16% of the variability in YKL-40 levels (R^2 between 0.08 and 0.16), and
275 they did not belong to a single block of LD. The LD matrix of *CHI3L1* SNPs in INMA is shown
276 in Figure E1.

277 In BAMSE, associations with YKL-40 levels were replicated for five (rs2886117, rs10399931,
278 rs10399805, rs4950928, and rs7542294) of the seven SNPs that provided significant signals
279 in INMA (Table E4).

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

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

286 Correlations between methylation levels at CpG sites and YKL-40 levels for INMA and
287 BAMSE are shown in Table 3. In INMA, serum YKL-40 levels at age 4 years correlated
288 significantly with methylation levels at age 4 years at eight CpG sites after Bonferroni
289 correction (p values ranging from $3x10^{-16}$ to 0.003). All correlations were negative, i.e., the
290 higher the methylation, the lower is the YKL-40 level. Inverse correlations between
291 methylation levels and YKL-40 at age 4 years were replicated in BAMSE for five of these
292 CpG sites (cg13134650, cg07423149, cg17014757, cg14085262, and cg03625911) (Table
293 3, Figure 2). In addition, methylation levels at these five CpG sites as measured at birth in
294 INMA and at age 8 years in BAMSE also correlated with YKL-40 levels at age 4 years (Table
295 3, Figure E3). These CpG associations with YKL-40 levels were confirmed after adjustment
296 for blood cell composition (data not shown).

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

300 - ***SNPs and DNA methylation***

301 Next, we tested associations between the seven SNPs and the five CpG sites that were
302 found to be related to YKL-40 levels in the aforementioned analyses (see Figure E5 for a
303 map of the CpG genomic locations). In INMA, 4-year methylation levels at four of five CpG
304 sites were significantly associated with all seven SNPs after Bonferroni correction (Table 4,

305 upper part), with alleles linked to lower YKL-40 being associated with higher methylation
306 levels. These genetic associations were remarkably similar when methylation levels from
307 cord blood were analyzed (Table 4, lower part). Based on the effect estimates and the
308 percent variability in methylation levels that they explained, the SNPs rs10399931 and
309 rs4950928, which are in strong LD, showed the strongest associations for all five CpG sites.
310 These two SNPs were also associated with methylation levels at all five CpG sites in
311 BAMSE both at age 4 and 8 years (Table E6).
312 All the aforementioned five CpG sites showed moderate-to-strong correlation of methylation
313 levels between birth and 4 years in INMA (Spearman correlation coefficients 0.35-0.85), and
314 between 4 and 8 years in BAMSE (0.44-0.88) (Figure E6), in line with the possibility of a
315 consistent genetic control of their methylation levels from birth to age 8 years.
316 These five CpG sites also showed significant associations with *CHI3L1* SNPs and significant
317 correlations between their methylation levels at 4 and 8 years of age in the PIAMA study
318 (Table E7 and Figure E6c).
319 - ***Multivariable and mediation analyses***
320 Given the aforementioned associations, we conducted multivariable analyses to identify
321 independent effects by SNPs and by CpG sites on YKL-40. Among the seven SNPs, in
322 INMA the final backward stepwise regression model included rs4950928 and rs7542294,
323 whereas the final model in BAMSE included rs10399931 and rs10399805. Of note,
324 rs4950928 and rs7542294 are in strong LD with rs10399931 and rs10399805, respectively
325 (Figure E1), indicating that in both cohorts genetic influences on YKL-40 may be driven by
326 these two groups of SNPs in LD. Among the five CpG sites, final stepwise models predicting
327 YKL-40 included only cg07423149 in INMA and only cg17014757 in BAMSE, but it should
328 be noted that methylation levels at the five CpG sites were strongly correlated with each
329 other both in INMA (Figure 2, Spearman correlation coefficients 0.64-0.87) and BAMSE
330 (Figure E3, 0.51-0.83), making it difficult to determine whether a single CpG site or rather a
331 global regional methylation profile was driving the association with YKL-40. To evaluate the
332 components of methylation profiles across these five highly correlated CpG sites, we

333 conducted a PCA and found that the first component explained up to 79% and up to 75% of
334 variance in INMA and BAMSE, respectively.
335 To evaluate to what extent methylation profiles mediated the relation of *CHI3L1* variation to
336 YKL-40 levels, we completed mediation analyses in INMA (the largest cohort with complete
337 data). In these analyses, we tested the two SNPs (rs4950928 and rs7542294) and the CpG
338 site (cg07423149) identified by the multivariable analyses plus the first component identified
339 from the PCA. In these analyses, cg07423149 mediated 62% of the rs4950928 effects and
340 43% of the rs7542294 effects on YKL-40 (Table E8). The corresponding percentages for the
341 overall methylation score were only slightly higher (66% and 46%, respectively). These
342 results indicate that *CHI3L1* methylation profiles may mediate a substantial proportion of the
343 effects of *CHI3L1* genetic variation on YKL-40 levels.
344
345 ***Secondary Analyses: Associations of 3-level biomarkers (SNPs, methylation, and***
346 ***YKL-40) with asthma (Figure 1b)***
347 We also tested all the markers (i.e., variation at seven SNPs, methylation at five CpG sites at
348 age 4 years, and YKL-40 levels at age 4 years) for cross-sectional associations with asthma
349 using the maximum number of available genetic and epigenetic samples in INMA, BAMSE,
350 and PIAMA. Numbers and characteristics of participants included in these analyses are
351 shown in the lower part of Table 1. Table 5 shows results of association analyses with
352 asthma for each of the biomarkers. Neither the SNPs nor the CpG sites showed consistent
353 associations with asthma in INMA, BAMSE, and PIAMA. The lack of association between
354 methylation at any of the CpG sites and asthma was also confirmed after further adjustment
355 for blood cell composition (data not shown). In contrast, YKL-40 levels were associated,
356 albeit weakly, with asthma (Table 5). In meta-analyses, each standard deviation (SD)
357 increase in YKL-40 levels was associated with a borderline 22% increase in the odds for
358 asthma (p=0.08), and participants in the highest YKL-40 tertile had 90% increased odds for
359 asthma compared with subjects in the lowest tertile (p=0.03).

360 In additional analyses, we confirmed the lack of association of genetic and epigenetic
361 markers with asthma when methylation levels at the other nine CpG sites in *CHI3L1* were
362 analyzed (Table E9) and when SNPs and methylation levels were tested for associations
363 with the comorbidity cluster (asthma, rhinitis, and eczema) that was recently described[43] in
364 these cohorts (Table E10).
365
366
367 **Discussion**
368 This is the first study to integrate genetic and epigenetic regulation of YKL-40 in childhood.
369 We found consistent interrelationships between *CHI3L1* genetic variants, methylation levels
370 at several *CHI3L1* CpG sites, and circulating YKL-40 levels. These associations were
371 replicated in multiple independent cohorts. Taken together with our mediation analysis, these
372 data indicate that methylation levels mediate part of the known effects of genetic variation on
373 YKL-40 levels. We also observed an association of YKL-40 levels, but not *CHI3L1* SNPs or
374 methylation levels, with asthma.
375 YKL-40 has been proposed as a potential biomarker for a broad range of diseases[1].
376 Elevated levels of YKL-40 have been consistently found in the blood and airways of adults
377 with asthma, particularly those with severe and persistent disease[2, 3, 5-10]. The underlying
378 mechanisms of these associations remain largely to be determined. *In vitro* studies have
379 shown that human bronchial epithelial cells express *CHI3L1* and secrete YKL-40 in response
380 to mechanical stress similar to that experienced during bronchoconstriction[44]. Furthermore,
381 YKL-40 levels were increased in broncho-alveolar lavage from people with asthma on
382 segmental allergen challenge[45, 46]. In support of a possible causal involvement of this gene in
383 asthma, *Chi3l1* null mice showed decreased aeroallergen-induced Th2 inflammatory
384 responses in their lungs compared with wild type animals[10].
385 Whereas the link between YKL-40 and asthma has been consistently reported in adults,
386 results from asthma studies in children have been conflicting. Serum YKL-40 levels at birth
387 and in the first 5 years of life were not significantly associated with asthma at age 6 years in

388 two studies, although positive trends were observed in both cases[11, 47]. Serum YKL-40
389 levels were increased in children with therapy-resistant asthma[12], but no association
390 between YKL-40 and asthma severity was found in a subsequent study of 61 asthmatic
391 children[13]. In our larger study, we found children with asthma to have elevated circulating
392 levels of YKL-40, although this association was relatively weak. This may be due to the fact
393 that we studied population-based epidemiological cohorts, in which the prevalence of severe
394 asthma is expected to be quite low.
395 Circulating levels of YKL-40 have been previously described to be under strong genetic
396 control. In multivariable analyses, we found SNPs from two groups of LD to be
397 independently associated with serum YKL-40. Among them, the two SNPs rs10399931 and
398 rs4950928 fell within a single LD block and were identified as the strongest protein
399 quantitative trait loci (pQTL) in our study, consistent with previous reports[11, 48]. They are
400 located in proximity of the transcription start site and, because they are in strong LD, it is
401 difficult to dissect their independent effects on YKL-40 levels and/or disease risk. Gene
402 reporter assays for *CHI3L1* promoter haplotypes indicated that both SNPs contributed
403 significant cis-regulatory effects on gene expression in Jurkat cells and that the magnitude of
404 these effects was the strongest for rs10399931[16], which was also the SNP with the strongest
405 effects on methylation levels in our study.
406 Findings from our study indicate that the effects of *CHI3L1* SNPs on circulating YKL-40 may
407 be mediated by methylation levels at *CHI3L1* CpG sites. Several SNPs were associated with
408 methylation levels at five CpG sites located within 1kb of the transcription start site that were,
409 in turn, negatively correlated with YKL-40 levels. It should be noted that the probe for
410 cg17014757 includes rs10399931 and the probe for cg03625911 includes rs7515776.
411 Therefore, the strong associations that were found between methylation levels at these CpG
412 sites and the corresponding SNPs (or other SNPs in LD with them) may be due to allele-
413 specific differences in probe hybridization[37]. However, similar genetic associations, albeit
414 smaller in magnitude than those observed for cg17014757, were found for CpG sites (i.e.,
415 cg13134650, cg07423149, and cg14085262) whose probes do not include known SNPs.

416 Mediation analyses that were completed both on cg07423149 and on the methylation
417 principal component supported SNP→CpG→YKL-40 causal models.
418 Our findings are in line with a growing body of evidence that points to significant effects of
419 DNA sequence variants across the genome on methylation levels at nearby and distal CpG
420 sites[23, 25-30]. Previous studies have shown that these methylation-quantitative trait loci
421 (meQTL) can in turn affect gene expression[25, 49] and are enriched in motifs for DNA-binding
422 factors and DNaseI hypersensitivity regions[29, 50]. MeQTLs have been shown to be enriched
423 for disease risk variants[29] and they may ultimately influence disease risk through their effects
424 on methylation patterns[23, 28]. Similarly, *CHI3L1* meQTLs could influence asthma risk by
425 affecting methylation, gene expression and, in turn, YKL-40 levels. However, in contrast with
426 this scenario, in our study neither *CHI3L1* SNPs nor methylation at *CHI3L1* CpG sites were
427 associated with asthma, although YKL-40 levels tended to be higher in children with asthma.
428 These results have two alternative explanations: either they represent a true negative
429 finding, which would argue against a causal role of *CHI3L1* variants in childhood asthma, or
430 they are due to possible methodological factors that should be taken into account. First,
431 asthma, particularly in the pre-school years, is characterized by a large phenotypic
432 heterogeneity that may impact our ability to capture phenotype-specific genetic associations.
433 Along the same lines, the lack of genetic associations may be explained by *CHI3L1*
434 influencing asthma risk through interactions with environmental factors or through effects on
435 asthma phenotypes that were not included or were under-represented in our study (e.g.,
436 adult or severe asthma). It is also possible that our genetic association analyses were under-
437 powered to detect a true signal, although in that case the magnitude of the *CHI3L1* genetic
438 effects would be expected to be relatively small.
439 Our study has some limitations. Although we did not find any statistical evidence that
440 interrelationships between *CHI3L1* genetic variation, methylation, and YKL-40 levels differed
441 between cases and controls, these stratified analyses had limited sample size, and potential
442 differences by disease status could not be determined conclusively. The methylation array
443 that we used is primarily designed for genomic discovery. Thus, information from a large

444 proportion of *CHI3L1* CpG sites and, in turn, analyses on regional methylation profiles could
445 not be included in our study and will need to be addressed in future studies. Finally, we
446 acknowledge that, by using whole blood for methylation studies, we may have missed
447 contributions of methylation profiles from other tissues (e.g., from the airways), and our
448 results may have been impacted by blood cell composition, particularly in analyses that used
449 cord blood[51]. However, it should be noted that we were able to replicate all the associations
450 between methylation at the five CpG sites and serum YKL-40 after adjustment for blood cell
451 composition, both as estimated by the Houseman method and, in a subset of INMA
452 participants, as directly assessed from blood smears.
453 In conclusion, in multiple independent cohorts, we found genetic variation in the *CHI3L1*
454 gene to be related to both methylation levels in nearby CpG sites and to circulating YKL-40
455 levels. Our findings indicate that *CHI3L1* genetic variation may affect circulating YKL-40
456 levels by regulating its gene methylation profiles.
457
458
459 **Acknowledgments**
460 Pyrosequencing was performed at the University Medical Center Groningen, in the
461 Laboratory of Obstetrics, under the supervision of Dr. Torsten Plösch.
462 We thank Dr. Xavier Basagaña for his helpful advice on mediation analysis.

References

1. Lee CG, Da Silva CA, Dela Cruz CS, Ahangari F, Ma B, Kang MJ, et al. Role of chitin and chitinase/chitinase-like proteins in inflammation, tissue remodeling, and injury. Annu Rev Physiol 2011; 73:479-501.
2. Chupp GL, Lee CG, Jarjour N, Shim YM, Holm CT, He S, et al. A chitinase-like protein in the lung and circulation of patients with severe asthma. N Engl J Med 2007; 357:2016-27.
3. Tang H, Fang Z, Sun Y, Li B, Shi Z, Chen J, et al. YKL-40 in asthmatic patients, and its correlations with exacerbation, eosinophils and immunoglobulin E. Eur Respir J 2010; 35:757-60.
4. James AJ, Reinius LE, Verhoek M, Gomes A, Kupczyk M, Hammar U, et al. Increased YKL-40 and Chitotriosidase in Asthma and Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2016; 193:131-42.
5. Otsuka K, Matsumoto H, Niimi A, Muro S, Ito I, Takeda T, et al. Sputum YKL-40 levels and pathophysiology of asthma and chronic obstructive pulmonary disease. Respiration 2012; 83:507-19.
6. Guerra S, Halonen M, Sherrill DL, Venker C, Spangenberg A, Carsin AE, et al. The relation of circulating YKL-40 to levels and decline of lung function in adult life. Respir Med 2013; 107:1923-30.
7. Lai T, Chen M, Deng Z, L Y, Wu D, Li D, et al. YKL-40 is correlated with FEV1 and the asthma control test (ACT) in asthmatic patients: influence of treatment. BMC Pulm Med 2015; 15:1.
8. Duru S, Yuce G, Ulasli SS, Erdem M, Kizilgun M, Kara F, et al. The relationship between serum YKL-40 levels and severity of asthma. Iran J Allergy Asthma Immunol 2013; 12:247-53.
9. Saba M, Sharif MR, Akbari H, Nikoueinejad H, Ramazani Jolfaii M. YKL-40 in Asthma and its correlation with different clinical parameters. Iran J Allergy Asthma Immunol 2014; 13:271-7.

10. Ahangari F, Sood A, Ma B, Takyar S, Schuyler M, Qualls C, et al. Chitinase 3-like-1 Regulates both Visceral Fat Accumulation and Asthma-like Th2 Inflammation. Am J Respir Crit Care Med 2015.
11. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, et al. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. N Engl J Med 2008; 358:1682-91.
12. Konradsen JR, James A, Nordlund B, Reinius LE, Soderhall C, Melen E, et al. The chitinase-like protein YKL-40: a possible biomarker of inflammation and airway remodeling in severe pediatric asthma. J Allergy Clin Immunol 2013; 132:328-35 e5.
13. Santos CB, Davidson J, Covar RA, Spahn JD. The chitinase-like protein YKL-40 is not a useful biomarker for severe persistent asthma in children. Ann Allergy Asthma Immunol 2014; 113:263-6.
14. Lee CG, Hartl D, Lee GR, Koller B, Matsuura H, Da Silva CA, et al. Role of breast regression protein 39 (BRP-39)/chitinase 3-like-1 in Th2 and IL-13-induced tissue responses and apoptosis. J Exp Med 2009; 206:1149-66.
15. Cunningham J, Basu K, Tavendale R, Palmer CN, Smith H, Mukhopadhyay S. The CHI3L1 rs4950928 polymorphism is associated with asthma-related hospital admissions in children and young adults. Ann Allergy Asthma Immunol 2011; 106:381-6.
16. Verlaan DJ, Ouimet M, Adoue V, Sirois-Gagnon D, Lariviere M, Ge B, et al. Promoter polymorphisms in CHI3L1 are associated with asthma. J Allergy Clin Immunol 2012; 130:533-5.
17. Gomez JL, Crisafi GM, Holm CT, Meyers DA, Hawkins GA, Bleecker ER, et al. Genetic variation in chitinase 3-like 1 (CHI3L1) contributes to asthma severity and airway expression of YKL-40. J Allergy Clin Immunol 2015.
18. Tsai Y, Ko Y, Huang M, Lin M, Wu C, Wang C, et al. CHI3L1 polymorphisms associate with asthma in a Taiwanese population. BMC Med Genet 2014; 15:86.

19. Kjaergaard AD, Johansen JS, Nordestgaard BG, Bojesen SE. Genetic variants in CHI3L1 influencing YKL-40 levels: resequencing 900 individuals and genotyping 9000 individuals from the general population. J Med Genet 2013; 50:831-7.
20. Rathcke CN, Holmkvist J, Husmoen LL, Hansen T, Pedersen O, Vestergaard H, et al. Association of polymorphisms of the CHI3L1 gene with asthma and atopy: a populations-based study of 6514 Danish adults. PLoS One 2009; 4:e6106.
21. Wu AC, Lasky-Su J, Rogers CA, Klanderman BJ, Litonjua A. Polymorphisms of chitinases are not associated with asthma. J Allergy Clin Immunol 2010; 125:754-7, 7 e1-7 e2.
22. Sohn MH, Lee JH, Kim KW, Kim SW, Lee SH, Kim KE, et al. Genetic variation in the promoter region of chitinase 3-like 1 is associated with atopy. Am J Respir Crit Care Med 2009; 179:449-56.
23. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011; 12:529-41.
24. Gruzieva O, Merid SK, Melen E. An update on epigenetics and childhood respiratory diseases. Paediatr Respir Rev 2014; 15:348-54.
25. Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. Am J Hum Genet 2010; 86:411-9.
26. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 2008; 40:904-8.
27. Hellman A, Chess A. Extensive sequence-influenced DNA methylation polymorphism in the human genome. Epigenetics Chromatin 2010; 3:11.
28. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013; 31:142-7.

29. Shi J, Marconett CN, Duan J, Hyland PL, Li P, Wang Z, et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nat Commun 2014; 5:3365.

30. Lemire M, Zaidi SH, Ban M, Ge B, Aissi D, Germain M, et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. Nat Commun 2015; 6:6326.

31. Bousquet J, Anto J, Auffray C, Akdis M, Cambon-Thomsen A, Keil T, et al. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. Allergy 2011; 66:596-604.

32. Ribas-Fito N, Ramon R, Ballester F, Grimalt J, Marco A, Olea N, et al. Child health and the environment: the INMA Spanish Study. Paediatr Perinat Epidemiol 2006; 20:403-10.

33. Kull I, Melen E, Alm J, Hallberg J, Svartengren M, van Hage M, et al. Breast-feeding in relation to asthma, lung function, and sensitization in young schoolchildren. J Allergy Clin Immunol 2010; 125:1013-9.

34. Brunekreef B, Smit J, de Jongste J, Neijens H, Gerritsen J, Postma D, et al. The prevention and incidence of asthma and mite allergy (PIAMA) birth cohort study: design and first results. Pediatr Allergy Immunol 2002; 13 Suppl 15:55-60.

35. Wijga AH, Kerkhof M, Gehring U, de Jongste JC, Postma DS, Aalberse RC, et al. Cohort profile: The Prevention and Incidence of Asthma and Mite Allergy (PIAMA) birth cohort. Int J Epidemiol 2013.

36. Pinart M, Benet M, Annesi-Maesano I, von Berg A, Berdel D, Carlsen KC, et al. Comorbidity of eczema, rhinitis, and asthma in IgE-sensitised and non-IgE-sensitised children in MeDALL: a population-based cohort study. Lancet Respir Med 2014; 2:131-40.

37. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013; 8:203-9.

38. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013; 14:293.
39. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012; 13:86.
40. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005; 76:887-93.
41. Guerra S, Halonen M, Vasquez MM, Spangenberg A, Stern DA, Morgan WJ, et al. Relation between circulating CC16 concentrations, lung function, and development of chronic obstructive pulmonary disease across the lifespan: a prospective study. Lancet Respir Med 2015; 3:613-20.
42. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. Journal of Statistical Software 2014; 59.
43. Garcia-Aymerich J, Benet M, Saeys Y, Pinart M, Basagana X, Smit HA, et al. Phenotyping asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic comorbidity cluster. Allergy 2015; 70:973-84.
44. Park JA, Drazen JM, Tschumperlin DJ. The chitinase-like protein YKL-40 is secreted by airway epithelial cells at base line and in response to compressive mechanical stress. J Biol Chem 2010; 285:29817-25.
45. Kuepper M, Bratke K, Virchow JC. Chitinase-like protein and asthma. N Engl J Med 2008; 358:1073-5; author reply 5.
46. Gavala ML, Kelly EA, Esnault S, Kukreja S, Evans MD, Bertics PJ, et al. Segmental allergen challenge enhances chitinase activity and levels of CCL18 in mild atopic asthma. Clin Exp Allergy 2013; 43:187-97.
47. Usemann J, Frey U, Mack I, Schmidt A, Gorlanova O, Roosli M, et al. CHI3L1 polymorphisms, cord blood YKL-40 levels and later asthma development. BMC Pulm Med 2016; 16:81.

48. Kruit A, Grutters JC, Ruven HJ, van Moorsel CC, van den Bosch JM. A CHI3L1 gene polymorphism is associated with serum levels of YKL-40, a novel sarcoidosis marker. Respir Med 2007; 101:1563-71.

49. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol 2011; 12:R10.

50. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 2011; 43:1091-7.

51. Yousefi P, Huen K, Quach H, Motwani G, Hubbard A, Eskenazi B, et al. Estimation of blood cellular heterogeneity in newborns and children for epigenome-wide association studies. Environ Mol Mutagen 2015; 56:751-8.

52. Martin TC, Yet I, Tsai PC, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. BMC Bioinformatics 2015; 16:131.

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 reference[52]

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

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

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

* 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).

SNP**	Location chr1 (hg19)	Effect allele	Other allele	Effect allele frequency	HWE P value*	Linear regression coefficient^	(95% CI)	P value	R^{2} ***
rs2886117	203168881	A	G	15.7	0.77	0.61	(0.326-0.894)	**$4*10^{-5}$**	0.10
rs10399931	203156080	T	C	19.2	0.33	-0.68	(-0.941, -0.420)	**$7*10^{-7}$**	0.14
rs10399805	203155998	A	G	15.4	0.77	0.62	(0.331-0.901)	**$3*10^{-5}$**	0.10
rs4950928	203155882	G	C	16.0	0.08	-0.82	(-1.108, -0.536)	**$6*10^{-8}$**	0.16
rs7542294	203151176	A	G	16.9	0.42	0.64	(0.360-0.914)	**$1*10^{-5}$**	0.11
rs2791718	203141424	A	C	12.8	0.32	0.61	(0.286-0.924)	**$2*10^{-4}$**	0.08
rs10920576	203129179	T	C	12.8	0.32	0.72	(0.406-1034)	**$1*10^{-5}$**	0.11

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.

		INMA methylation at 4y		BAMSE methylation at 4y		INMA methylation at birth		BAMSE methylation at 8y	
		N = 203		N = 230		N = 189		N = 225	
CpG sites	Position chr1 (hg19)	Spearman corr with YKL-40 at 4y	P value	Spearman corr with YKL-40 at 4y	P value	Spearman corr with YKL-40 at 4y	P value	Spearman corr with YKL-40 at 4y	P value
cg13134650	**203156765**	**-0.36**	**$1*10^{-7}$**	**-0.13**	**0.06**	**-0.25**	**$5*10^{-4}$**	**-0.22**	**$7*10^{-4}$**
cg19081101	203156626	-0.35	$3*10^{-7}$	-0.06	0.38	-0.15	0.04	0.01	0.90
cg02097014	203156375	-0.21	0.003	-0.11	0.10	-0.09	0.24	-0.12	0.08
cg07423149	**203156247**	**-0.53**	**$3*10^{-16}$**	**-0.27**	**$2*10^{-5}$**	**-0.30**	**$3*10^{-5}$**	**-0.26**	**$6*10^{-5}$**
cg17014757^	**203156098**	**-0.44**	**$7*10^{-11}$**	**-0.34**	**$2*10^{-7}$**	**-0.31**	**$2*10^{-5}$**	**-0.35**	**$4*10^{-8}$**
cg14085262	**203155939**	**-0.44**	**$3*10^{-11}$**	**-0.21**	**0.001**	**-0.24**	**$7*10^{-4}$**	**-0.22**	**$7*10^{-4}$**
cg03625911^^	**203155738**	**-0.42**	**$4*10^{-10}$**	**-0.18**	**0.006**	**-0.21**	**0.004**	**-0.19**	**0.005**
cg17000774	203154457	-0.04	0.57	0.00	0.97	-0.04	0.62	0.03	0.65
cg11196333	203154371	-0.04	0.62	-0.14	0.04	-0.08	0.26	-0.16	0.01
cg05526099	203152296	0.00	0.95	-0.02	0.80	-0.05	0.51	0.02	0.77
cg04361579	203152047	0.00	0.95	-0.16	0.01	-0.06	0.44	0.00	0.99
cg20707774	203149828	-0.10	0.15	-0.20	0.003	-0.04	0.56	-0.11	0.09
cg14165900	203148901	-0.02	0.74	0.05	0.44	0.04	0.61	0.00	0.97
cg15490070	203148132	-0.33	$2*10^{-6}$	-0.05	0.47	-0.13	0.08	-0.05	0.49

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.

YEAR 4	cg13134650				cg07423149				cg17014757*				cg14085262				cg03625911**			
SNP (effect allele)	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}
rs2886117 (A)	-1.01	-1.83, -0.19	0.02	0.03	-3.59	-5.34, -1.84	**$8*10^{-5}$**	0.09	-6.56	-9.41, -3.71	**$1*10^{-5}$**	0.11	-4.26	-5.93, -2.60	**$1*10^{-6}$**	0.13	-2,76^^	-4.09, -1.43	**$6*10^{-5}$**	0.08
rs10399931 (T)	2.42	1.74, 3.09	**$4*10^{-11}$**	0.20	6.52	5.13, 7.90	**$7*10^{-17}$**	0.31	15,55^^	14.28, 16.81	**$3*10^{-57}$**	0.70	5.25	3.73, 6.77	**$1*10^{-10}$**	0.20	4.09	2.94, 5.23	**$4*10^{-11}$**	0.21
rs10399805 (A)	-1.04	-1.86, -0.22	0.01	0.03	-3.76	-5.50, -2.01	**$3*10^{-5}$**	0.09	-6.56	-9.41, -3.71	**$1*10^{-5}$**	0.11	-4.26	-5.93, -2.58	**$1*10^{-6}$**	0.13	-2,75^^	-4.08, -1.41	**$7*10^{-5}$**	0.08
rs4950928 (G)	2.49	1.71, 3.26	**$2*10^{-9}$**	0.17	6.66	5.06, 8.25	**$4*10^{-14}$**	0.26	16,16^^	14.56, 17.75	**$2*10^{-46}$**	0.58	4.82	3.07, 6.57	**$2*10^{-7}$**	0.14	4.02	2.70, 5.35	**$1*10^{-8}$**	0.17
rs7542294 (A)	-1.22	-2.01, -0.42	**0.003**	0.05	-3.54	-5.25, -1.83	**$7*10^{-5}$**	0.09	-5.00	-7.89, -2.11	**$8*10^{-4}$**	0.07	-3.96	-5.61, -2.31	**$5*10^{-6}$**	0.11	-2,78^^	-4.09, -1.48	**$4*10^{-5}$**	0.09
rs2791718 (A)	-1.09	-2.01, -0.17	0.02	0.03	-3.86	-5.79, -1.92	**$1*10^{-4}$**	0.08	-5.73	-8.98, -2.47	**$7*10^{-4}$**	0.07	-3.46	-5.38, -1.55	**$5*10^{-4}$**	0.07	-2.54	-4.04, -1.04	**0.001**	0.06
rs10920576 (T)	-1.23	-2.14, -0.31	0.009	0.04	-4.02	-5.94, -2.10	**$6*10^{-5}$**	0.09	-5.79	-9.04, -2.54	**$6*10^{-4}$**	0.07	-3.59	-5.51, -1.68	**$3*10^{-4}$**	0.07	-2.60	-4.09, -1.10	**$7*10^{-4}$**	0.06

BIRTH	cg13134650				cg07423149				cg17014757*				cg14085262				cg03625911**			
SNP (effect allele)	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}	Regression coefficient^	95% CI	P value	R^{2***}
rs2886117 (A)	-0.55	-1.43, 0.33	0.22	0.01	-1.88	-3.19, -0.57	**0.005**	0.04	-5.28	-8.04, -2.52	**$2*10^{-4}$**	0.09	-3.31	-4.85, - 1.77	**$4*10^{-5}$**	0.09	-1,05^^	-2.18, 0.09	0.07	0.02
rs10399931 (T)	2.02	1.25, 2.79	**$7*10^{-7}$**	0.12	4.75	3.68, 5.81	**$2*10^{-15}$**	0.29	14,77^^	13.53, 16.01	**$1*10^{-53}$**	0.71	3.92	2.49, 5.35	**$2*10^{-7}$**	0.14	3.01	2.03, 3.99	**$9*10^{-9}$**	0.17
rs10399805 (A)	-0.59	-1.47, 0.30	0.19	0.01	-1.91	-3.22, -0.60	**0.005**	0.04	-5.24	-8.02, -2.46	**$3*10^{-4}$**	0.08	-3.46	-5.00, -1.93	**$2*10^{-5}$**	0.10	-1,16^^	-2.30, -0.03	0.05	0.02
rs4950928 (G)	1.80	0.91, 2.68	**$9*10^{-5}$**	0.08	4.61	3.42, 5.79	**$2*10^{-12}$**	0.23	14,72^^	13.08, 16.36	**$2*10^{-39}$**	0.56	3.55	1.93, 5.16	**$2*10^{-5}$**	0.10	3.00	1.90, 4.09	**$2*10^{-7}$**	0.14
rs7542294 (A)	-0.66	-1.53, 0.21	0.13	0.01	-1.97	-3.25, -0.70	**0.003**	0.05	-3.91	-6.72, -1.10	**0.007**	0.05	-3.17	-4.69, -1.64	**$6*10^{-5}$**	0.09	-1,33^^	-2.44, -0.22	0.02	0.03
rs2791718 (A)	-0.97	-1.94, 0.01	0.05	0.02	-2.10	-3,55, -0.65	**0.005**	0.05	-5.22	-8.34, -2.10	**0.001**	0.07	-3.39	-5.11, -1.67	**$1*10^{-4}$**	0.08	-1.49	-2.75, -0.22	0.02	0.03
rs10920576 (T)	-1.17	-2.14, -0.19	0.02	0.03	-2.17	-3.61, -0.72	**0.004**	0.05	-5.64	-8.74, -2.54	**$4*10^{-4}$**	0.08	-3.83	-5.52, - 2.13	**$2*10^{-5}$**	0.11	-1.77	-3.03, - 0.51	**0.006**	0.04

^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 [r2>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.

	INMA			**BAMSE**			**PIAMA**				**Meta-analysis**		
	adjOR	(95% CI)	P value	adjOR	(95% CI)	P value	adjOR	(95% CI)	P value		adjOR	(95% CI)	P value
SNPs (effect allele) **N = 2405 ^**													
rs2886117 (A)	1.30	(0.774-2.174)	0.32	0.85	(0.552-1.294)	0.44	1.30	(0.811-2.080)	0.28		**1.09**	**(0.835-1.431)**	**0.516**
rs10399931 (T)[*%]	0.55	(0.305-0.992)	0.05	1.27	(0.919-1.748)	0.15	0.91	(0.590-1.393)	0.66		**1.00**	**(0.792-1.269)**	**0.985**
rs10399805 (A)	1.40	(0.829-2.354)	0.21	0.75	(0.482-1.178)	0.21	1.12	(0.693-1.825)	0.64		**1.02**	**(0.776-1.352)**	**0.867**
rs4950928 (G)[*%]	0.50	(0.250-0.986)	0.05	1.44	(1.014-2.042)	0.04	0.74	(0.457-1.193)	0.22		**1.01**	**(0.778-1.314)**	**0.933**
rs7542294 (A)	1.12	(0.658-1.909)	0.67	0.81	(0.540-1.219)	0.31	1.18	(0.754-1.855)	0.47		**1.00**	**(0.767-1.297)**	**0.987**
rs2791718 (A)[*%]	0.66	(0.331-1.295)	0.22	0.68	(0.422-1.094)	0.11	1.42	(0.873-2.313)	0.16		**0.90**	**(0.664-1.221)**	**0.499**
rs10920576 (T)	0.65	(0.338-1.253)	0.20	0.67	(0.435-1.046)	0.08	1.41	(0.886-2.252)	0.15		**0.89**	**(0.665-1.181)**	**0.411**

Methylation levels N = 625 ^^													
cg13134650"	0.95	(0.837-1.081)	0.44	1.07	(0.966-1.187)	0.19	1.22	(0.967-1.550)	0.09		**1.04**	**(0.965-1.123)**	**0.297**
cg07423149"	1.01	(0.945-1.075)	0.81	0.99	(0.951-1.029)	0.59	1.06	(0.981-1.150)	0.14		**1.00**	**(0.974-1.036)**	**0.796**
cg17014757"	0.98	(0.941-1.021)	0.33	1.01	(0.985-1.036)	0.42	1.05	(0.995-1.116)	0.07		**1.01**	**(0.988-1.028)**	**0.446**
cg14085262"	1.02	(0.952-1.090)	0.59	0.96	(0.926-1.004)	0.08	1.07	(0.987-1.162)	0.10		**0.99**	**(0.960-1.024)**	**0.603**
cg03625911"	1.04	(0.951-1.128)	0.42	0.98	(0.932-1.026)	0.37	1.09	(0.992-1.200)	0.07		**1.01**	**(0.969-1.047)**	**0.714**
Standardized YKL-40 levels N = 432 ^^^													
Effect for 1-SD increase in YLK-40	1.27	(0.861-1.860)	0.23	1.21	(0.913-1.591)	0.19	N/A				**1.23**	**(0.978-1.534)**	**0.077**
Medium vs Low YKL40 tertile	2.62	(0.927-7.377)	0.07	1.50	(0.761-2.96)	0.24	N/A				**1.77**	**(1.004-3.130)**	**0.048**
High vs Low YKL40 tertile	2.36	(0.820-6.773)	0.11	1.75	(0.889-3.427)	0.11	N/A				**1.90**	**(1.078-3.364)**	**0.026**

^ N asthma / N total: INMA (58 / 336); BAMSE (171 / 467); PIAMA (69 / 1602)

^^ 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.

Figure 2a. INMA 4 yrs

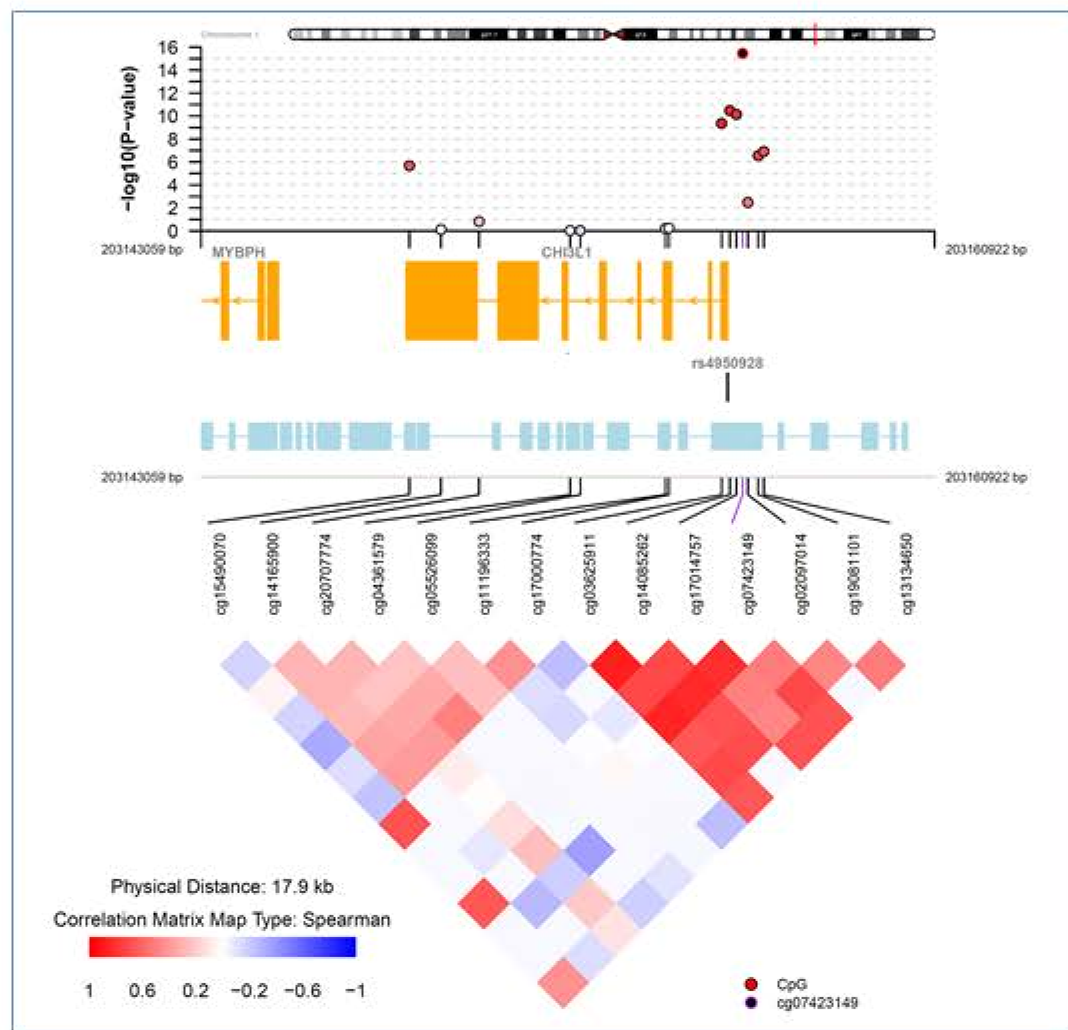
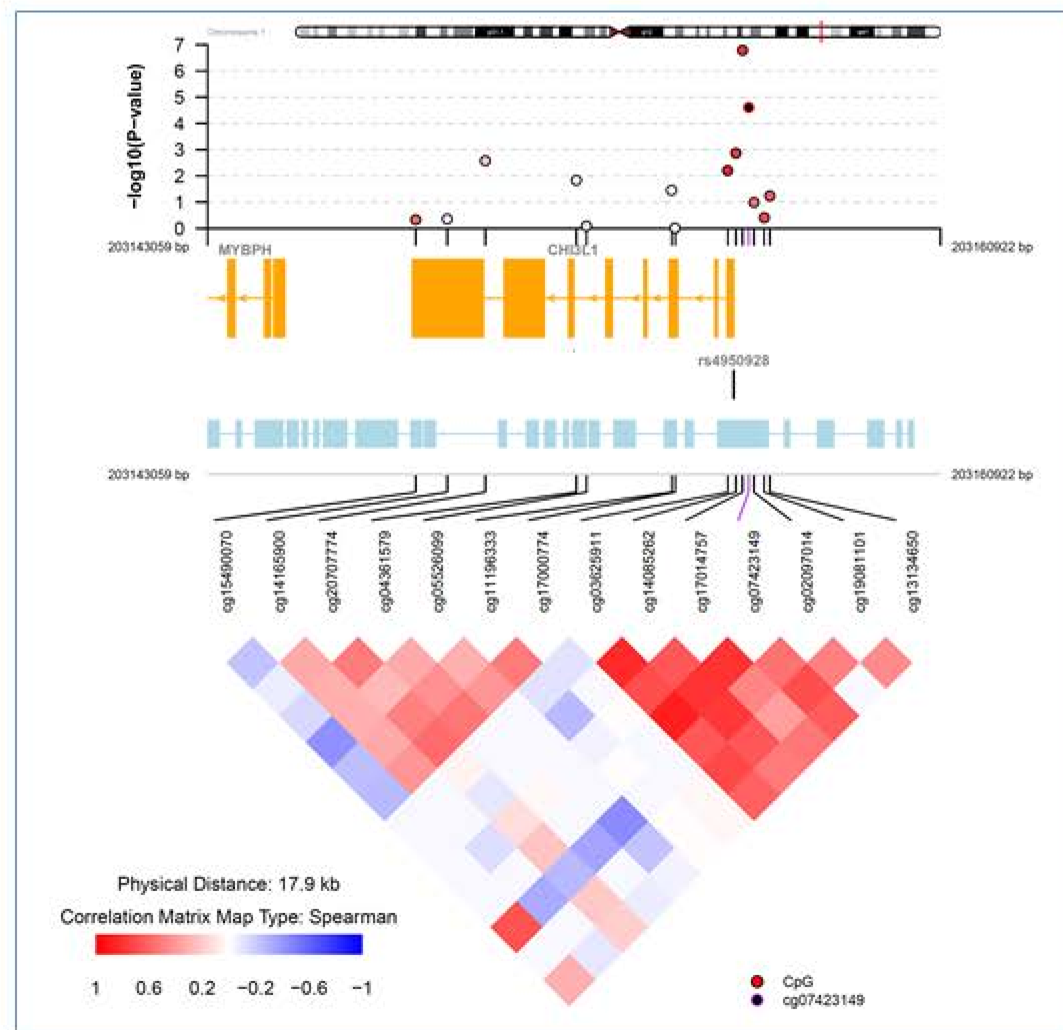


Figure 2b. BAMSE 4 yrs



PRIMARY ANALYSES

To determine whether *CHI3L1* methylation levels mediate the association of *CHI3L1* SNPs to YKL-40 levels

AVAILABLE DATA

CHI3L1 methylation

INMA (N=189)

***CHI3L1* methylation**

Circulating YKL-40 levels

INMA / BAMSE (N=433)

CHI3L1 methylation

BAMSE (N=225)

Birth

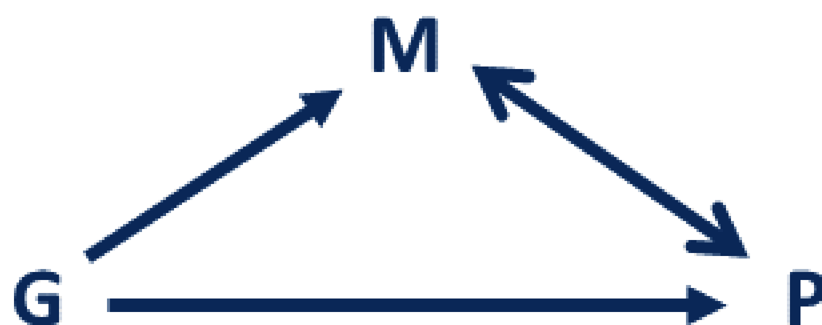
year 4

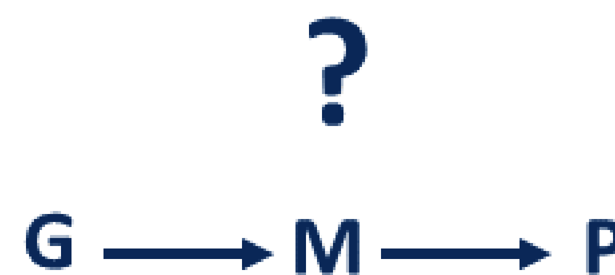
Year 8

***CHI3L1* SNPs**

INMA / BAMSE (N=250)

ANALYTICAL APPROACH





SECONDARY ANALYSES

To assess whether *CHI3L1* SNPs and methylation levels that regulate YKL-40 are associated with asthma in early childhood

AVAILABLE DATA

***CHI3L1* SNPs**

INMA (336) / BAMSE (467) / PIAMA (1602)

(tot N=2405)

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)

Asthma at year 4

INMA / BAMSE / PIAMA

ANALYTICAL APPROACH