

## **Identification of a new locus at 16q12 associated with time-to-asthma onset**

Chloé Sarnowski, Pierre-Emmanuel Sugier, Raquel Granell, Debbie Jarvis, Marie-Hélène Dizier, Markus Ege, Medea Imboden, Catherine Laprise, Elza Khusnutdinova, Maxim Freidin, et al.

► **To cite this version:**

Chloé Sarnowski, Pierre-Emmanuel Sugier, Raquel Granell, Debbie Jarvis, Marie-Hélène Dizier, et al.. Identification of a new locus at 16q12 associated with time-to-asthma onset. *Journal of Allergy and Clinical Immunology*, Elsevier, 2016, 138 (4), pp.1071-1080. <10.1016/j.jaci.2016.03.018>. <inserm-01321840>

**HAL Id: inserm-01321840**

**<http://www.hal.inserm.fr/inserm-01321840>**

Submitted on 26 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Identification of a new locus at 16q12 associated with time-to-asthma onset**

2

3 Chloé Sarnowski<sup>1,2</sup>, PhD, Pierre-Emmanuel Sugier<sup>1,2</sup>, MSc, Raquel Granell<sup>3</sup>, PhD, Debbie  
4 Jarvis<sup>4,5</sup>, MD, Marie-Hélène Dizier<sup>1,2</sup>, PhD, Markus Ege<sup>6,7</sup>, MD, Medea Imboden<sup>8,9</sup>, PhD,  
5 Catherine Laprise<sup>10</sup>, PhD, Elza K. Khusnutdinova<sup>11,12</sup>, PhD, Maxim B. Freidin<sup>13</sup>, PhD,  
6 William O.C. Cookson<sup>14</sup>, MD, DPhil, Miriam Moffatt<sup>14</sup>, DPhil, Mark Lathrop<sup>15</sup>, PhD, Valérie  
7 Siroux<sup>16,17,18</sup>, PhD, Ludmila M. Ogorodova<sup>19</sup>, MD, PhD, Alexandra S. Karunas<sup>11,12</sup>, MD,  
8 PhD, Alan James<sup>20</sup>, MD, Nicole M. Probst-Hensch<sup>8,9</sup>, PhD, Erika von Mutius<sup>6,7</sup>, MD, Isabelle  
9 Pin<sup>16,17,21</sup>, MD, PhD, Manolis Kogevinas<sup>22,23,24,25</sup>, MD, PhD, A. John Henderson<sup>3</sup>, MD,  
10 Florence Demenais<sup>1,2</sup>, MD, Emmanuelle Bouzigon<sup>1,2</sup>, MD, PhD

11

12 1. Inserm, UMR-946, F-75010, Paris, France

13 2. Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, F-  
14 75007, Paris, France

15 3. School of Social and Community Medicine, University of Bristol, UK

16 4. Respiratory Epidemiology, Occupational Medicine and Public Health, National Heart and  
17 Lung Institute, Imperial College, London, UK

18 5. MRC-PHE Centre for Environment & Health, London, UK

19 6. Dr von Hauner Children's Hospital, Ludwig Maximilian University, Munich, Germany

20 7. Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research,  
21 Munich, Germany

22 8. Swiss Tropical and Public Health Institute, Basel, Switzerland

23 9. University of Basel, Basel, Switzerland

24 10. Département des sciences fondamentales, Université du Québec à Chicoutimi, Saguenay,  
25 QC, Canada

- 26 11. Institute of Biochemistry and Genetics, Ufa Scientific Centre, Russian Academy of  
27 Sciences, Ufa, Russian Federation;
- 28 12. Bashkir State University, Department of Genetics and Fundamental Medicine, Ufa,  
29 Russian Federation
- 30 13. Research Institute for Medical Genetics, Tomsk, Russia;
- 31 14. National Heart Lung Institute, Imperial College London, London, UK
- 32 15. McGill University and Génome Québec Innovation Centre, Montréal, Canada
- 33 16. Univ. Grenoble Alpes, IAB, Team of Environmental Epidemiology applied to  
34 Reproduction and Respiratory Health, F-38000 Grenoble, France
- 35 17. Inserm, IAB, Team of Environmental Epidemiology applied to Reproduction and  
36 Respiratory Health, F-38000 Grenoble, France
- 37 18. CHU de Grenoble, IAB, Team of Environmental Epidemiology applied to Reproduction  
38 and Respiratory Health, F-38000 Grenoble, France
- 39 19. Siberian State Medical University, Tomsk, Russia
- 40 20. Busselton Population Medical Research Institute, Department of Pulmonary Physiology  
41 and Sleep Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia; School of  
42 Population Health, University of Western Australia, Crawley, Western Australia
- 43 21. CHU de Grenoble, Pediatrics, Grenoble, France
- 44 22. Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
- 45 23. CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
- 46 24. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
- 47 25. Universitat Pompeu Fabra, Barcelona, Spain
- 48
- 49 **Corresponding author:**
- 50 Emmanuelle Bouzigon, MD, PhD

51 UMR-946, INSERM / Université Paris-Diderot  
52 Institut de Génétique Moléculaire  
53 27 rue Juliette Dodu  
54 75010 Paris  
55 Tel: 33 (0) 1 53 72 50 23  
56 Fax: 33 (0) 1 53 72 50 49  
57 Email: [emmanuelle.bouzigon@inserm.fr](mailto:emmanuelle.bouzigon@inserm.fr)

58

59 **Disclosure of potential conflict of interest:**

60 Erika von Mutius has acted as consultant to GlaxoSmithKline, Novartis, Astellas Pharma  
61 Europe Ltd., ALK Abelló, and Vifor Pharma. She is Associate Editor at the Journal of Allergy  
62 and Clinical Immunology. Erika von Mutius and Markus Ege declared that their institution  
63 received research grants from the European Commission and the European Research Council.  
64 A. John Henderson declared that his institution received money from the Medical Research  
65 Council and the Wellcome Trust. Alexandra S. Karunas declared her institution and herself  
66 received money from the European Commission and the Russian Foundation for Basic  
67 Research. Elza K. Khusnutdinova declared her institution and herself received money from  
68 the European Commission. Isabelle Pin declared payment for lectures and  
69 travel/accommodations/meeting expenses from GSK, Novartis. Valérie Siroux has acted as  
70 consultant to Edimark Santé and TEVA. All the other authors declare no conflict of interest.

71

72 **Sources of support:**

73 This work was supported by French National Agency for Research (ANR-CES-2009, ANR-  
74 11-BSV1-027-GWIS-AM), Région Ile-de-France (DIM-SEnT grant), ‘Fonds de dotation  
75 *Recherche en Santé Respiratoire*’, Russian Foundation for Basic Research (grants: №13-04-

76 01397 and 01-04-48213a), Healthway and the Departments of Science and of Health of the  
77 Government of Western Australia, the UK Medical Research Council, the Wellcome Trust  
78 (grant: 102215/2/13/2), the University of Bristol and the Swiss National Science Foundation  
79 (current grants no 33CS30-148470/1). The Canada Research Chair held by C Laprise and the  
80 funding supports from Canadian Institutes of Health Research (CIHR) enabled the  
81 maintenance and continuation of the SLSJ asthma study. C. Laprise is the director of the  
82 Asthma Strategic Group of the Respiratory Health Network of the Fonds de la recherche en  
83 santé du Québec (FRSQ) and member of Allergen network. Genotyping was supported by  
84 grants from the European Commission (No. LSHB-CT-2006-018996-GABRIEL) and the  
85 Wellcome Trust (WT084703MA).

86

87 **Manuscript word count:** 3,581

88 **Abstract word count:** 247

89

90 **ABSTRACT**

91 **Background:** Asthma is a heterogeneous disease in which age-of-onset plays an important  
92 role.

93 **Objective:** We sought to identify the genetic variants associated with time-to-asthma onset.

94 **Methods:** We conducted a large-scale meta-analysis of nine genome-wide association studies  
95 of time-to-asthma onset (total of 5,462 asthmatics with a broad range of age-of-asthma onset  
96 and 8,424 controls of European ancestry) performed using survival analysis techniques.

97 **Results:** We detected five regions associated with time-to-asthma onset at genome-wide  
98 significant level ( $P < 5 \times 10^{-8}$ ). We evidenced a new locus in 16q12 region (near *cylindromatosis*  
99 *turban tumor syndrome* gene (*CYLD*)) and confirmed four asthma risk regions: 2q12  
100 (*IL1RL1*), 6p21 (*HLA-DQA1*), 9p24 (*IL33*) and 17q12-q21 (*ZPBP2-GSDMA*). Conditional  
101 analyses identified two distinct signals at 9p24 (both upstream of *IL33*) and at 17q12-q21  
102 (near *ZPBP2* and within *GSDMA*). These seven distinct loci explained together 6.0% of the  
103 variance in time-to-asthma onset. In addition, we showed that genetic variants at 9p24 and  
104 17q12-q21 were strongly associated with an earlier onset of childhood asthma ( $P \leq 0.002$ )  
105 whereas 16q12 SNP was associated with a later asthma onset ( $P = 0.04$ ). A high burden of  
106 disease risk alleles at these loci was associated with earlier age-of-asthma onset (4 years  
107 *versus* 9-12 years,  $P = 10^{-4}$ ).

108 **Conclusion:** The new susceptibility region for time-to-asthma onset at 16q12 harbors variants  
109 that correlate with the expression of *CYLD* and *NOD2* (*nucleotide-binding oligomerization*  
110 *domain 2*), two strong candidates for asthma. This study demonstrates that incorporating the  
111 variability of age-of-asthma onset in asthma modeling is a helpful approach in the search for  
112 disease susceptibility genes.

113

114

115 **Key Messages:**

- 116 • 16q12 genetic variants are associated with time-to-asthma onset and correlate with  
117 *CYLD* and *NOD2* expressions.
- 118 • Genetic variants at 9p24 (upstream of *IL33*) and 17q12-q21 (nearby *ZPBP2* and within  
119 *GSDMA*) are associated with an earlier asthma-onset whereas variants at 16q12 are  
120 associated with a later asthma onset.
- 121 • Taking into account the variability of age-of-asthma onset in disease modeling can  
122 increase the power of identifying new genes involved in asthma pathophysiology.

123

124 **Capsule summary:**

125 This large-scale meta-analysis of nine genome-wide association studies identified 16q12  
126 genetic variants associated with time-to-asthma onset that correlate with *CYLD* and *NOD2*  
127 expressions, two strong candidate genes implicated in inflammation.

128

129 **Keywords:** Asthma, age-of-onset, genetics, genome-wide association study, survival analysis,  
130 conditional analysis, *CYLD*, *NOD2*

131

132 **Abbreviations:**

133 *CYLD*: Cylindromatosis (turban tumor syndrome)

134 eQTL: Expression quantitative trait locus

135 *GSDMA*: Gasdermin A

136 GWAS: Genome-wide association study

137 *HLA-DQA1*: Major histocompatibility complex, class II, DQ alpha 1

138 HR: Hazard ratio

139 *IL1RL1*: Interleukin 1 receptor-like 1

- 140 *IL33*: Interleukin 33
- 141 LCL: Lymphoblastoid cell line
- 142 LD: Linkage Disequilibrium
- 143 *NFkB1*: Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
- 144 *NOD2*: Nucleotide-binding oligomerization domain containing 2
- 145 QC: Quality control
- 146 SNP: Single nucleotide polymorphism
- 147 TAO: time-to-asthma onset
- 148 *ZPBP2*: Zona pellucida binding protein 2
- 149



150 **INTRODUCTION**

151 The prevalence of asthma has dramatically increased over the past decades in high-income  
152 countries, affecting 5–16% of people worldwide.<sup>1</sup> It is the most common chronic disease  
153 among children and a decrease in the age-of-asthma onset has been recently documented.<sup>2</sup>

154 Asthma is a complex and heterogeneous disease with variable clinical expression over the life  
155 span.<sup>1</sup> It is now well recognized that asthma is not a single disease but rather a collection of  
156 different phenotypes which may represent different manifestations of a common underlying  
157 pathological process or may be separate disease entities.<sup>3</sup> One of the simplest characteristic  
158 that can be used to differentiate disease phenotypes is the age at onset.<sup>4,5</sup> Indeed, asthma  
159 displays different characteristics according to the lifetime period at which it occurs.<sup>6</sup> Early age  
160 of onset is more frequently associated with a family history of asthma, allergy sensitization  
161 and clinical response to triggers, whereas late-onset disease is associated with eosinophilic  
162 inflammation and obesity, is more common in women, and is generally less allergic.<sup>3</sup>

163 The risk of developing asthma has a strong genetic component, with estimated heritability  
164 ranging from 35 to 95%.<sup>7</sup> Genome-wide association studies (GWASs) have been successful in  
165 identifying more than 20 loci associated with asthma.<sup>8</sup> However, the genetic factors identified  
166 to date account only for a small part of the genetic component of the disease.<sup>1</sup> This hidden  
167 heritability might be linked to the phenotypic heterogeneity of asthma.<sup>9</sup> The vast majority of  
168 GWASs conducted until now have analyzed asthma as a binary phenotype. A few genetic  
169 studies have considered more specific definition of asthma incorporating the age of disease  
170 onset. A genome-wide linkage screen conducted for time-to-asthma onset in French families  
171 revealed two regions, 1p31 and 5q13, potentially linked to this phenotype.<sup>10</sup> A single GWAS  
172 has been performed on age-of-asthma onset in asthmatic children and led to the identification  
173 of two loci, not found by the previous asthma GWASs; these loci on chromosomes 3p26 and  
174 11q24 were associated with an earlier onset of childhood asthma.<sup>11</sup> Moreover, the effect of

175 17q12-q21 genetic variants, identified by the first GWAS of asthma,<sup>12</sup> was found to be  
176 restricted to early-onset asthma.<sup>13,14</sup>

177 Instead of stratifying the data according to age-of-onset of disease using an arbitrary  
178 threshold, one can integrate the age-of-onset in modeling asthma risk, by using survival  
179 analysis methodology applied to both asthmatic and non-asthmatic subjects. The goal of the  
180 present study was to identify the genetic determinants underlying time-to-asthma onset in a  
181 large meta-analysis of 5,462 asthmatics and 8,424 controls from nine independent European-  
182 ancestry populations.

183

## 184 **METHODS**

### 185 *Populations*

186 We studied 13,886 individuals of European ancestry from nine independent studies (one birth  
187 cohort, five population-based and three family studies) which were part of the GABRIEL  
188 European consortium on the genetics of asthma.<sup>14</sup> A brief description of these studies with  
189 appropriate references is provided in the online repository and in Table E1. All of these  
190 studies had the age of asthma onset and imputed genetic data available.

191 For all studies, ethical approval was obtained from the appropriate institutional ethic  
192 committees and all individuals or child's legal guardians provided written informed consent.

193

### 194 *Time-to-asthma onset definition*

195 Definition of asthma was based on report of doctor's diagnosis and/or on standardized  
196 questionnaires (see online repository). To model time-to-asthma onset (TAO), we used age of  
197 onset or age at first wheeze for individuals who developed asthma, while in individuals who  
198 were free of disease upon examinations, we used age at last examination.

199

### 200 *Genotyping*

201 Genotyping, single nucleotide polymorphism (SNP) imputation process and quality control  
202 (QC) criteria (for individuals and SNPs) for each study are described in Table E1 in the online  
203 repository. All datasets were genotyped at Centre National de Génotypage (CNG, Evry,  
204 France), as part of the European GABRIEL asthma consortium.<sup>14</sup> Quality control and  
205 imputations were performed independently for each study. Genome-wide imputations were  
206 conducted using MACH 1.0 software,<sup>15</sup> with reference haplotype panels from HapMap 2.  
207 SNPs with imputation quality score (Rs<sub>q</sub>)  $\geq 0.5$  and minor allele frequency  $\geq 1\%$  were kept  
208 for analysis. Then, to further investigate the regions associated with time-to-asthma onset at

209 the genome-wide significant level, we used imputed data from the 1000 Genomes Project and  
210 applied the same SNP QC criteria.

211

### 212 *Statistical analysis and strategy of analysis*

213 After the study-specific QC, a total of 13,886 individuals from the nine cohorts were included  
214 in the present study. In each dataset, association between time-to-asthma onset and individual  
215 SNPs was investigated under an additive genetic model using a Cox proportional-hazards  
216 regression model adjusted for sex and the first four principal components to account for  
217 population structure. A robust sandwich estimation of the variance<sup>16</sup> was used in family data  
218 to take into account familial dependencies. Moreover, due to the complex sampling design of  
219 the GABRIELA study, survey regression techniques were used for this study to estimate  
220 robust standard errors ('svy' command in Stata). Proportional hazard assumptions for the  
221 main SNP effect were tested and never rejected. Genome-wide association studies of time-to-  
222 asthma onset were first conducted in each of the nine datasets separately, and then combined  
223 through meta-analysis in order to increase power and to obtain more robust findings. Meta-  
224 analyzed hazard-ratios (HRs) and 95% confidence intervals (CIs) were calculated using a  
225 fixed-effect (inverse variance) model. To assess heterogeneity of SNP effect across studies,  
226 the Cochran's Q statistic was calculated. If heterogeneity was evidenced, a random-effect  
227 model was fitted. All analyses were performed using Stata® version 13.1 (STATA Corp.,  
228 College Station, Texas, USA). After the meta-analysis, we only kept SNPs for which at least  
229 66% of the studies ( $\geq 6$  studies among the 9 studies) participated to the summary statistics in  
230 the meta-analysis, to reduce the rate of false-positive findings. The meta-analysis results were  
231 obtained for a total of 2,387,926 SNPs. We used the classical threshold of  $P\text{-value} \leq 5 \times 10^{-8}$  to  
232 declare a meta-analyzed SNP effect as genome-wide significant.

233

234 ***Conditional analysis to uncover distinct signals at TAO-associated loci***

235 To identify distinct TAO-associated SNPs in each region harboring genome-wide significant  
236 signals, we re-analyzed separately these regions in each of the nine studies. For that purpose,  
237 we added the region top SNP into the primary Cox model as a covariate and tested the effect  
238 of each other SNP of that region. Then, the results were meta-analyzed using the same  
239 strategy as the primary GWASs. If a secondary signal was detected in a region, a second run  
240 of conditional analyses was performed to check for a third distinct signal in that region.  
241 Length of explored regions was based on regional association plots and ranged from 200 kb to  
242 500 kb depending on recombination hotspots.

243

244 ***eQTL analysis and functional annotations***

245 We queried whether significant SNPs (or their proxies) associated with time-to-asthma onset  
246 at  $P \leq 5 \times 10^{-8}$  and potentially secondary signals from conditional analysis were expression  
247 quantitative trait loci (eQTLs). We used existing eQTL databases in multiple tissues  
248 (especially blood and lung) for populations of European ancestry (see online repository).<sup>17-23</sup>  
249 Functional annotations of significant SNPs (or their proxies) were obtained using  
250 ENcyclopedia Of DNA Elements (ENCODE) data<sup>24</sup> provided by the HaploReg tool.<sup>25</sup>

251

252 ***Relationship of TAO-associated loci with age-of-asthma onset***

253 In a first step, we investigated in asthmatics whether each of the SNPs associated with TAO  
254 were also associated with age-of-asthma onset using a non-parametric rank test followed by a  
255 non-parametric equality-of-medians test. In a second step, we assessed the cumulative effect  
256 of risk alleles of SNPs found associated with the age-of-asthma onset at step 1. For that  
257 purpose, we used either the number of risk alleles or the quintiles of a polygenic score  
258 distribution. The polygenic risk score is the weighted sum of the number of age of asthma

259 onset associated alleles with weight being the log of adjusted hazard ratio estimated in  
260 asthmatics only. The associations were tested in eight studies for which we had access to raw  
261 data (all datasets except ALSPAC) using a cox proportional hazard model adjusted on sex and  
262 principal components.  
263

264 **RESULTS**

265 *Description of populations*

266 A total of 13,886 subjects were included in the present study (5,462 asthmatics and 8,424  
267 non-asthmatics). Asthmatics had a mean age-of-asthma onset of 12.5 years (ranging from 0.5  
268 to 75 years, Figure E1), a mean age of 26.8 years at examination (mean per study ranging  
269 from 9.1 to 51.3 years) and 52.6% were males. Non-asthmatics had a mean age of 32.4 years  
270 at examination (mean per study ranging from 8.9 to 55.8 years) and 49% were males (Table  
271 E1).

272

273 *Genetic variants associated with time-to-asthma onset*

274 The Manhattan and the Q-Q plots of the meta-analysis of time-to-asthma onset GWAS results  
275 are shown in Figure 1 and Figure E2 in the online repository respectively. A total of 155  
276 SNPs were associated with time-to-asthma onset at the genome-wide significance level of  
277  $P < 5 \times 10^{-8}$ . These SNPs clustered into five distinct chromosomal regions (Table I), that  
278 included a new risk locus on 16q12 (nearby *CYLD* (cylindromatosis turban tumor syndrome),  
279 1 SNP) and four established risk loci for asthma: 2q12 (*IL1RL1-IL18R1*, 7 SNPs), 6p21  
280 (nearby *HLA-DQA1*, 1 SNP), 9p24 (flanking *IL33*, 25 SNPs) and 17q12-q21 (121 SNPs  
281 spanning 389 kb, and with the main signal located near *ZPBP2*). The regional association  
282 plots for these genome-wide associated loci are shown in Figures 2 and E3, and the forest  
283 plots for the top signal in each region are shown in Figure E4. Three additional loci were  
284 associated with time-to-asthma onset at a suggestive significance threshold ( $5 \times 10^{-8} < P < 10^{-6}$ ;  
285 Table I): *MAP4K4* (2q11-q12), *RORA* (15q22) and *IL4R* (16p12-p11).

286 To determine whether any of the five TAO loci harbored additional association signals, we  
287 performed conditional association analysis in each region. For this analysis, a threshold of  
288  $P \leq 2.1 \times 10^{-5}$  was used to declare significance, corresponding to a Bonferroni threshold for

289 2,382 independent tests. These analyses evidenced two secondary signals (Table II and Figure  
290 E5 in the online repository): 1) rs413382 in 9p24 region at 73 kb of *IL33* ( $P=9.7\times 10^{-6}$  after  
291 conditioning on top SNP and  $P=5.9\times 10^{-8}$  in the primary meta-analysis) and 2) rs3859192 in  
292 17q12-q21 region within *GSDMA* ( $P=4.0\times 10^{-6}$  after conditioning on top SNP and  $P=1.5\times 10^{-13}$   
293 in the primary meta-analysis). In contrast, at 2q12, 6p21 and 16q12 regions, the inclusion of  
294 the most significant time-to-asthma onset GWAS SNP as a covariate in association analysis  
295 resulted in nearly complete reduction of the association signal in these regions, suggesting  
296 that there was no evidence for a second distinct genetic factor in these regions.

297 To obtain a denser map of the new TAO 16q12 locus, we repeated association analyses using  
298 1000 Genomes Project imputed SNPs. These analyses strengthened our original finding with  
299 additional signals ( $3.8\times 10^{-8}\leq P\leq 2.6\times 10^{-7}$ ) located in an intergenic region encompassing the  
300 lead SNP rs1861760 (Table E2 and Figure E6). These SNPs were in moderate to high LD  
301 with rs1861760 ( $0.71\leq r^2\leq 0.81$ ) and thus did not represent independent signals from that top  
302 hit. Similar analyses conducted in the four other TAO-associated regions also supported our  
303 original findings and did not evidence any additional independent signal in these regions.

304 Overall, the seven distinct SNPs (five top SNPs and two secondary SNPs) associated with  
305 TAO showed low heterogeneity between studies ( $P>0.11$ ) and explained together 6.0% of the  
306 variance in time-to-asthma onset.

307

### 308 ***Functional annotations and effect on gene expression***

309 To provide some insights into the potential molecular mechanisms underlying the TAO-  
310 associated variants, we queried whether the five top SNPs and two secondary signals (and  
311 their proxies) were 1) tagging potentially deleterious SNPs, 2) located in regulatory elements,  
312 and 3) reported to influence the expression of one or more of nearby genes (eQTLs at  
313  $P<5\times 10^{-5}$ ). We focused on the new TAO risk locus at 16q12 region. Functional annotations



314 for the remaining six loci are presented in the online repository and eQTL data are presented  
315 in Table III (main results) and in Table E3 in the online repository.

316 The 16q12 TAO-associated variants are located in an intergenic region delimited by two  
317 recombination hotspots on each side, near *CYLD* (22 kb downstream). The rs1861760 maps to  
318 FOXJ1 and SOX binding sites. This SNP and/or its proxies correlate with the expression of  
319 *CYLD* in both blood and human lung tissues and of *NOD2* (*nucleotide-binding*  
320 *oligomerization domain 2*) in blood<sup>20</sup> (Table III for main results and Table E3 in the online  
321 repository).<sup>17,20</sup>

322

### 323 ***Relationship between TAO-associated variants and age-of-asthma onset***

324 To investigate whether TAO-associated SNPs influence the age-of-asthma onset, we  
325 compared, in asthmatics, the distribution of age-of-asthma onset according to the number of  
326 risk alleles at each of the seven main and secondary TAO-associated SNPs (Figure 3).  
327 Asthmatic subjects carrying one or two copies of risk allele at 17q12-q21 SNPs (rs9901146  
328 and rs3859192) or at 9p24 rs928413 had a younger age-of-asthma onset than non-carriers  
329 (median 6-8 years *versus* 10 years ( $P \leq 6 \times 10^{-4}$ ) and median 6-8 years *versus* 9 years ( $P = 0.002$ ),  
330 respectively), whereas those having at least one copy of rs1861760 risk-allele at 16q12 had a  
331 later age-of-asthma onset than non-carriers (median 10 years *versus* 8 years;  $P = 0.04$ ). No  
332 significant difference was found for the other three SNPs. We evidenced that an increased  
333 number of risk alleles at these four SNPs was associated with a younger age-of-asthma onset  
334 (median 12 years for carrying one risk allele to 4 years for carrying 6-8 risk alleles;  $P = 10^{-4}$ ).  
335 Finally, we detected a strong association between the age-of-asthma onset and the polygenic  
336 risk score (from median 10 years in first quintile to 6 years in last quintile;  $P = 4 \times 10^{-4}$ ).

337

### 338 ***Comparison of time-to-asthma onset GWAS results with previous asthma GWASs***

339 To investigate the impact of taking into account the age-of-asthma onset in disease modeling  
340 through survival analysis, we explored whether the top TAO SNPs were associated with  
341 asthma modeled as a binary trait in the nine cohorts included in the present study (Table E4,  
342 see online repository). We also investigated GABRIEL top SNPs in TAO meta-analysis  
343 (Table E4).<sup>14</sup> We observed a strong decrease of heterogeneity of SNP effect across studies in  
344 TAO analysis ( $P_{\text{Het}} \geq 0.11$ ) compared with asthma binary trait analyzed in the same datasets  
345 ( $P_{\text{Het}} \geq 0.004$ ) as well as in all GABRIEL datasets ( $P_{\text{Het}} \geq 0.0009$ ), especially in 9p24 and 17q12-  
346 q21 regions. The association signals were always more significant in TAO analysis as  
347 compared with the binary trait analysis in the same datasets. This increase of significance  
348 level was very high: 100-fold for 2q12 and 16q12, and  $10^4$  to  $10^6$ -fold for 9p24 and 17q12-  
349 q21. In fact, the asthma binary trait analysis only detected two loci (*HLA* and *GSDMA*) at the  
350 genome-wide significance level out of the seven TAO-associated loci. Conversely, the present  
351 TAO analysis identified at the genome-wide significance level four of the six main published  
352 GABRIEL regions,<sup>14</sup> and even at higher significance for 9p24 and 17q12-q21 regions (100 to  
353  $10^4$ -fold) as compared to GABRIEL significance levels. The two remaining GABRIEL loci  
354 not detected by our TAO analysis were those with the weaker effects (OR=1.12 for rs744910  
355 in 15q22 and rs2284033 in 22q13) in GABRIEL meta-analysis.<sup>14</sup>  
356 Finally, we evaluated whether previously reported susceptibility loci for asthma<sup>26</sup> were  
357 associated with time-to-asthma onset in our meta-analysis (see online repository, Table E5).  
358 Among the 21 loci detected in European populations, 12 were replicated at 5% in our TAO  
359 meta-analysis with the same direction of effects. Among the nine non-replicated signals, three  
360 SNPs (or some proxies) were not available in our data, and the remaining six loci had been  
361 reported for specific phenotypes: asthma exacerbation, age-of-asthma onset *per se* in  
362 asthmatic children only (quantitative trait) or childhood asthma (binary trait).<sup>11,27,28</sup>

363

364 **DISCUSSION**

365 By taking into account the age-of-asthma onset in asthma association analysis, we identified  
366 in this large meta-analysis including both asthmatic and non-asthmatic individuals, adults and  
367 children, a new susceptibility locus at 16q12 associated with time-to-asthma onset and  
368 confirmed the involvement of six other distinct loci belonging to four regions in asthma  
369 pathogenesis (2q12, 6p21, 9p24 and 17q12-q21). Genetic variants at 9p24 and 17q12-q21  
370 were strongly associated with an earlier onset of childhood asthma whereas the 16q12 lead  
371 SNP was associated with a risk of later onset asthma.

372 The most significant 16q12 genetic variant (rs1861760) is located nearby *CYLD* and *NOD2*  
373 and also maps to a binding site of FOXJ1, a transcription factor associated with allergic  
374 rhinitis.<sup>29</sup> Genetic variants located in a 130-kb region around rs1861760 were reported  
375 associated with immune-related diseases: inflammatory bowel diseases (Crohn's disease) and  
376 leprosy.<sup>30-35</sup> Interestingly, haplotype reconstruction (Figure 4) showed that the TAO  
377 rs1861760-A risk allele was always associated with SNP alleles that conferred a decreased  
378 risk of Crohn's disease (rs17221417-C, rs5743289-C and rs2076756-A located in *NOD2* and  
379 rs12324931-A located in *CYLD*)<sup>30-32,35,36</sup> and of leprosy (rs16948876-G located in intergenic  
380 region at 2 kb from rs1861760).<sup>33</sup> Indeed, GWASs revealed common genetic susceptibility  
381 loci for asthma and other immune-related disorders, suggesting shared molecular pathways  
382 involved in their etiology; however opposite alleles appear to be at risk.<sup>37</sup> Interestingly, an  
383 opposite effect of rs1861760-A allele is also observed at the gene expression level. Thus,  
384 TAO risk-allele at rs1861760 correlated with both the expression of *CYLD* and *NOD2* in  
385 blood but with an opposite effect.<sup>20</sup> However, this TAO risk allele was only associated with  
386 increased *CYLD* expression in lung tissue.<sup>17</sup> *CYLD* encodes a deubiquitinating enzyme that  
387 regulates diverse physiological processes including immune response and inflammation.<sup>38</sup>  
388 *CYLD* mainly acts as a negative regulator of nuclear factor- $\kappa$ B (*NFkB1*) to protect the host

389 from an over-reactive inflammatory response.<sup>38</sup> Conversely *NOD2*, which plays an important  
390 role in the innate immune response to intracellular bacterial lipopolysaccharides (LPS),  
391 activates the *NFkB1* pathway.<sup>39</sup> NFkB1 is a pleiotropic transcription factor that acts as a key  
392 regulator of immune and inflammatory genes, and activation of the NFkB1 pathway has been  
393 implicated in airway inflammation and asthma.<sup>40,41</sup> Moreover, FOXP1 transcription factor that  
394 binds to the genomic region encompassing 16q12 TAO-associated SNP (rs1861760) was  
395 described to inhibit NFkB1 activity.<sup>42</sup> Recently, *CYLD* has been shown to regulate lung  
396 fibrosis in mice by inhibiting transforming growth factor- $\beta$ -signaling (TGF- $\beta$ ) through a  
397 decrease of SMAD3 protein stability.<sup>43</sup> Of interest, *SMAD3* has been reported to be associated  
398 with asthma by previous GWAS.<sup>14</sup>

399 Defining the phenotype is an important consideration because phenotypic heterogeneity can  
400 reduce power of GWAS.<sup>44</sup> In the present analyses, we studied the variability of time-to-  
401 asthma onset in both asthmatics and non-asthmatics based on survival analysis methods. The  
402 information used for such analysis was the age-of-onset in asthmatics and the age at last  
403 examination or death in non-asthmatics. In such a model, unaffected subjects represent  
404 censored observations as they are still at risk for disease, being perhaps too young to exhibit  
405 the trait. This approach, which allowed combining the age-of-asthma onset together with the  
406 disease status (affected/unaffected) led to decrease the genetic heterogeneity across studies  
407 and increase the power to detect association signals (upon  $10^6$ -fold increase as compared to  
408 the disease status only analysis. More specifically, increased evidence of association was  
409 observed in regions where age-of-asthma onset explained at least in part the genetic  
410 heterogeneity such as the 17q12-q21 locus for which a restricted SNP effect to a particular  
411 group of age-of-onset (early childhood onset asthma) was demonstrated.<sup>13</sup> Moreover, this  
412 analysis led to the identification of a new locus at 16q12 near *CYLD* and of an additional  
413 signal in the 9p24 region. These results support the hypothesis that a better consideration of

414 the phenotypic heterogeneity of asthma may help to disentangle the genetic heterogeneity of  
415 asthma.

416 Our study included both children and adult asthmatic subjects. Age-of-disease onset may be  
417 subject to recall bias, especially among individuals who are the furthest from the time of the  
418 first symptoms (e.g. adults who suffered of asthma in childhood) because it is often defined in  
419 a retrospective manner. However, high accuracy of the self-reported year of asthma onset by  
420 adult subjects has been shown by two independent studies among which the European  
421 Community Respiratory Health Survey, which was part of the present study.<sup>45,46</sup> Erroneous  
422 recall of the age-of-asthma onset is unlikely to have significantly affected the results since we  
423 observed little genetic heterogeneity across studies (e.g. childhood onset asthma reported by  
424 either adults or children).

425 It was suggested that some genetic variants might influence asthma in an age-specific manner.  
426 Among TAO-associated SNPs, we confirmed the association of 17q12-q21 SNPs with an  
427 early age-of-asthma onset<sup>13,14</sup> and evidenced for the first time that the top 9p24 genetic variant  
428 near *IL33* was also associated with early childhood-onset asthma (median age-of-onset of 6-8  
429 years in risk-allele carriers). Indeed, in GABRIEL meta-analysis, 9p24 SNPs were more  
430 strongly associated with early-onset asthma (before 16 years) than late-onset asthma (after 16  
431 years), but this difference was not significant.<sup>14</sup> Conversely, genetic variants at the new  
432 susceptibility locus, 16q12, conferred a risk of later onset asthma (median age-of-onset of 10  
433 years in risk-allele carriers). Moreover, we evidenced that a high burden of disease risk alleles  
434 at these loci is associated with earlier age-of-asthma onset (4 years *versus* 9-12 years). This  
435 difference in asthma onset may reflect difference in patterns of onset of disease.<sup>47</sup> Indeed, we  
436 evidenced in the GABRIELA study that persistent early wheezing subjects carried more risk  
437 alleles than transient early wheezing subjects, and we confirmed previous association between  
438 persistent early wheezing and 9p24 and 17q12-q21 loci (data not shown). The 17q12-q21

439 genetic variants were reported associated with persistent childhood wheeze phenotype  
440 whereas 9p24 were mostly associated with intermediate onset wheeze but also with persistent  
441 early wheeze.<sup>48,49</sup> Moreover, 17q12-q21 SNPs were associated with fractional exhaled Nitric  
442 Oxide (FeNO) levels in children but not in adults, childhood severe asthma and allergic  
443 rhinitis, and 9p24 SNPs were associated with childhood severe asthma, asthma-plus-rhinitis,  
444 atopic asthma, allergy and eosinophil counts.<sup>50-55</sup>

445 In summary, we identified five regions harboring seven distinct signals that were associated  
446 with time-to-asthma onset, among which the 16q12 region that is reported for the first time.  
447 Several lines of evidence suggest that *CYLD* and *NOD2* located in that region are strong  
448 candidate genes for asthma. This study demonstrates that incorporating the variability of age-  
449 of-asthma onset in disease modeling is a useful strategy to uncover new disease genes.

450

451 **ACKNOWLEDGMENTS**

452 **EGEA:** The authors thank all those who participated to the setting of the study and on the  
453 various aspects of the examinations involved: interviewers, technicians for lung function  
454 testing and skin prick tests, blood sampling, IgE determinations, coders, those involved in  
455 quality control, data and sample management and all those who supervised the study in all  
456 EGEA centers. The authors are grateful to the three CIC-Inserm of Necker, Grenoble and  
457 Marseille who supported the EGEA study and in which subjects were examined. They are  
458 also grateful to the biobanks in Lille (CIC-Inserm) and at Annemasse (Etablissement français  
459 du sang) where EGEA biological samples are stored. We thank the Epidemiological Study on  
460 Genetics and Environment of Asthma (EGEA) cooperative group members as follows.

461 **Coordination:** V Siroux (epidemiology, PI since 2013); F Demenais (genetics); I Pin (clinical  
462 aspects); R Nadif (biology). F Kauffmann (PI 1992-2012); **Respiratory epidemiology:**  
463 Inserm U 700, Paris M Korobaëff (Egea1), F Neukirch (Egea1); Inserm U 707, Paris: I  
464 Annesi-Maesano (Egea1-2); Inserm CESP/U 1018, Villejuif: F Kauffmann, N Le Moual, R  
465 Nadif, MP Oryszczyn (Egea1-2), R. Varraso; Inserm U 823, Grenoble: V Siroux. **Genetics:**  
466 Inserm U 393, Paris: J Feingold; Inserm U 946, Paris: E Bouzigon, F Demenais, MH Dizier;  
467 CNG, Evry: I Gut (now CNAG, Barcelone, Spain), M Lathrop (now Univ McGill, Montreal,  
468 Canada). **Clinical centers:** Grenoble: I Pin, C Pison; Lyon: D Ecochard (Egea1), F Gormand,  
469 Y Pacheco; Marseille: D Charpin (Egea1), D Vervloet (Egea1-2); Montpellier: J Bousquet;  
470 Paris Cochin: A Lockhart (Egea1), R Matran (now in Lille); Paris Necker: E Paty (Egea1-2),  
471 P Scheinmann (Egea1-2); Paris-Trousseau: A Grimfeld (Egea1-2), J Just. **Data and quality**  
472 **management:** Inserm ex-U155 (Egea1): J Hochez; Inserm CESP/U 1018, Villejuif: N Le  
473 Moual, Inserm ex-U780: C Ravault (Egea1-2); Inserm ex-U794: N Chateigner; Grenoble: J  
474 Ferran (Egea1-2)

475 **SLSJ**: We thank all participants included in the SLSJ asthma familial collection. Catherine  
476 Laprise built, coordinates and manages the SLSJ study. Drs. Paul Bégin and Charles Morin  
477 confirmed the respiratory diagnosis for the adults and children respectively. We also thank the  
478 laboratory technician (Nadia Mior, Denise Morin), research professional (Anne-Marie  
479 Madore) and nurses (from the ECOGENE-21 clinical research group). Catherine Laprise is  
480 the Canada Research Chair in Environment and Genetics of Respiratory Disorders and  
481 Allergy, Director of the Asthma Strategic Group of the Respiratory Health Network (RHN) of  
482 Fonds de la recherche en santé du Québec (FRSQ) and researcher of the AllerGen NCE.

483 **ALSPAC**: We are extremely grateful to all the families who took part in this study, the  
484 midwives for their help in recruiting them, and the whole ALSPAC team, which includes  
485 interviewers, computer and laboratory technicians, clerical workers, research scientists,  
486 volunteers, managers, receptionists and nurses.

487 **UFA**: We thank the staff members of Department of Pediatrics and Department of  
488 Propaedeutics of internal diseases of Bashkir Medical State University.

489 **SAPALDIA**: The study could not have been done without the help of the study participants,  
490 technical and administrative support and the medical teams and field workers at the local  
491 study sites. **Study directorate:** NM Probst-Hensch (PI; e/g); T Rochat (p), C Schindler (s), N  
492 Künzli (e/exp), JM Gaspoz (c). **Scientific team:** JC Barthélémy (c), W Berger (g), R  
493 Bettschart (p), A Bircher (a), C Brombach (n), PO Bridevaux (p), L Burdet (p), Felber  
494 Dietrich D (e), M Frey (p), U Frey (pd), MW Gerbase (p), D Gold (e), E de Groot (c), W  
495 Karrer (p), F Kronenberg (g), B Martin (pa), A Mehta (e), D Miedinger (o), M Pons (p), F  
496 Roche (c), T Rothe (p), P Schmid-Grendelmeyer (a), D Stolz (p), A Schmidt-Trucksäss (pa),  
497 J Schwartz (e), A Turk (p), A von Eckardstein (cc), E Zemp Stutz (e). **Scientific team at**  
498 **coordinating centers:** M Adam (e), I Aguilera (exp), S Brunner (s), D Carballo (c), S  
499 Caviezel (pa), I Curjuric (e), A Di Pascale (s), J Dratva (e), R Ducret (s), E Dupuis Lozeron



500 (s), M Eeftens (exp), I Eze (e), E Fischer (g), M Foraster (e), M Germond (s), L Grize (s), S  
501 Hansen (e), A Hensel (s), M Imboden (g), A Ineichen (exp), A Jeong (g), D Keidel (s), A  
502 Kumar (g), N Maire (s), A Mehta (e), R Meier (exp), E Schaffner (s), T Schikowski (e), M  
503 Tsai (exp). (a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp)  
504 exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational  
505 health, (p) pneumology, (pa) physical activity, (pd) pediatrics, (s) statistics.

506

507 **REFERENCES**

- 508 1. Martinez FD, Vercelli D. Asthma. *Lancet* 2013; 382:1360-72.
- 509 2. Radhakrishnan DK, Dell SD, Guttmann A, Shariff SZ, Liu K, To T. Trends in the age of  
510 diagnosis of childhood asthma. *J Allergy Clin Immunol* 2014; 134:1057-62 e5.
- 511 3. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat*  
512 *Med* 2012; 18:716-25.
- 513 4. Siroux V, Basagana X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, et al. Identifying  
514 adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011; 38:310-7.
- 515 5. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of  
516 asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J*  
517 *Respir Crit Care Med* 2010; 181:315-23.
- 518 6. Szeffler SJ, Chmiel JF, Fitzpatrick AM, Giacoia G, Green TP, Jackson DJ, et al. Asthma  
519 across the ages: knowledge gaps in childhood asthma. *J Allergy Clin Immunol* 2014;  
520 133:3-13; quiz 4.
- 521 7. Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective.  
522 *Immunol Rev* 2011; 242:10-30.
- 523 8. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI  
524 GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;  
525 42:D1001-6.
- 526 9. Siroux V, Gonzalez JR, Bouzigon E, Curjuric I, Boudier A, Imboden M, et al. Genetic  
527 heterogeneity of asthma phenotypes identified by a clustering approach. *Eur Respir J*  
528 2014; 43:439-52.
- 529 10. Bouzigon E, Ulgen A, Dizier MH, Siroux V, Lathrop M, Kauffmann F, et al. Evidence  
530 for a pleiotropic QTL on chromosome 5q13 influencing both time to asthma onset and  
531 asthma score in French EGEA families. *Hum Genet* 2007; 121:711-9.

- 532 11. Forno E, Lasky-Su J, Himes B, Howrylak J, Ramsey C, Brehm J, et al. Genome-wide  
533 association study of the age of onset of childhood asthma. *J Allergy Clin Immunol* 2012;  
534 130:83-90 e4.
- 535 12. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants  
536 regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 2007;  
537 448:470-3.
- 538 13. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of  
539 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008;  
540 359:1985-94.
- 541 14. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-  
542 scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;  
543 363:1211-21.
- 544 15. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum*  
545 *Genet* 2009; 10:387-406.
- 546 16. Williams RL. A note on robust variance estimation for cluster-correlated data. *Biometrics*  
547 2000; 56:645-6.
- 548 17. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to  
549 help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8:e1003029.
- 550 18. Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, et al. A cross-  
551 platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid  
552 cell lines. *Genome Res* 2013; 23:716-26.
- 553 19. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al.  
554 Transcriptome genetics using second generation sequencing in a Caucasian population.  
555 *Nature* 2010; 464:773-7.

- 556 20. Westra HJ, Peters MJ, Esko T, Yaghoobkar H, Schurmann C, Kettunen J, et al.  
557 Systematic identification of trans eQTLs as putative drivers of known disease  
558 associations. *Nat Genet* 2013; 45:1238-43.
- 559 21. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. Population  
560 genomics of human gene expression. *Nat Genet* 2007; 39:1217-24.
- 561 22. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-  
562 resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS*  
563 *Genet* 2008; 4:e1000214.
- 564 23. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; 45:580-5.
- 565 24. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated  
566 encyclopedia of DNA elements in the human genome. *Nature* 2012; 489:57-74.
- 567 25. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation,  
568 and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids*  
569 *Res* 2012; 40:D930-4.
- 570 26. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al.  
571 Potential etiologic and functional implications of genome-wide association loci for  
572 human diseases and traits. *Proc Natl Acad Sci U S A* 2009; 106:9362-7.
- 573 27. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, et al.  
574 Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene.  
575 *Am J Hum Genet* 2009; 84:581-93.
- 576 28. McGeachie MJ, Wu AC, Tse SM, Clemmer GL, Sordillo J, Himes BE, et al. CTNNA3  
577 and SEMA3D: Promising loci for asthma exacerbation identified through multiple  
578 genome-wide association studies. *J Allergy Clin Immunol* 2015.

- 579 29. Li CS, Chae SC, Lee JH, Zhang Q, Chung HT. Identification of single nucleotide  
580 polymorphisms in FOXP1 and their association with allergic rhinitis. *J Hum Genet* 2006;  
581 51:292-7.
- 582 30. Kenny EE, Pe'er I, Karban A, Ozelius L, Mitchell AA, Ng SM, et al. A genome-wide  
583 scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS Genet*  
584 2012; 8:e1002559.
- 585 31. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of  
586 seven common diseases and 3,000 shared controls. *Nature* 2007; 447:661-78.
- 587 32. Cleynen I, Vazeille E, Artieda M, Verspaget HW, Szczypiorska M, Bringer MA, et al.  
588 Genetic and microbial factors modulating the ubiquitin proteasome system in  
589 inflammatory bowel disease. *Gut* 2014; 63:1265-74.
- 590 33. Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, et al. Identification of two new loci at  
591 IL23R and RAB32 that influence susceptibility to leprosy. *Nat Genet* 2011; 43:1247-51.
- 592 34. Liu H, Irwanto A, Fu X, Yu G, Yu Y, Sun Y, et al. Discovery of six new susceptibility  
593 loci and analysis of pleiotropic effects in leprosy. *Nat Genet* 2015; 47:267-71.
- 594 35. Kugathasan S, Baldassano RN, Bradfield JP, Sleiman PM, Imielinski M, Guthery SL, et  
595 al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel  
596 disease. *Nat Genet* 2008; 40:1211-5.
- 597 36. Elding H, Lau W, Swallow DM, Maniatis N. Dissecting the genetics of complex  
598 inheritance: linkage disequilibrium mapping provides insight into Crohn disease. *Am J*  
599 *Hum Genet* 2011; 89:798-805.
- 600 37. Li X, Ampleford EJ, Howard TD, Moore WC, Torgerson DG, Li H, et al. Genome-wide  
601 association studies of asthma indicate opposite immunopathogenesis direction from  
602 autoimmune diseases. *J Allergy Clin Immunol* 2012; 130:861-8 e7.

- 603 38. Sun SC. CYLD: a tumor suppressor deubiquitinase regulating NF-kappaB activation and  
604 diverse biological processes. *Cell Death Differ* 2010; 17:25-34.
- 605 39. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G. Nod2, a Nod1/Apaf-1  
606 family member that is restricted to monocytes and activates NF-kappaB. *J Biol Chem*  
607 2001; 276:4812-8.
- 608 40. Poynter ME, Cloots R, van Woerkom T, Butnor KJ, Vacek P, Taatjes DJ, et al. NF-kappa  
609 B activation in airways modulates allergic inflammation but not hyperresponsiveness. *J*  
610 *Immunol* 2004; 173:7003-9.
- 611 41. Kurakula K, Vos M, Logiantara A, Roelofs JJ, Nieuwenhuis MA, Koppelman GH, et al.  
612 Nuclear Receptor Nur77 Attenuates Airway Inflammation in Mice by Suppressing NF-  
613 kappaB Activity in Lung Epithelial Cells. *J Immunol* 2015.
- 614 42. Lin L, Spoor MS, Gerth AJ, Brody SL, Peng SL. Modulation of Th1 activation and  
615 inflammation by the NF-kappaB repressor Foxj1. *Science* 2004; 303:1017-20.
- 616 43. Lim JH, Jono H, Komatsu K, Woo CH, Lee J, Miyata M, et al. CYLD negatively  
617 regulates transforming growth factor-beta-signalling via deubiquitinating Akt. *Nat*  
618 *Commun* 2012; 3:771.
- 619 44. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide  
620 association signals. *Nat Rev Genet* 2009; 10:318-29.
- 621 45. Pattaro C, Locatelli F, Sunyer J, de Marco R. Using the age at onset may increase the  
622 reliability of longitudinal asthma assessment. *J Clin Epidemiol* 2007; 60:704-11.
- 623 46. Toren K, Palmqvist M, Lowhagen O, Balder B, Tunsater A. Self-reported asthma was  
624 biased in relation to disease severity while reported year of asthma onset was accurate. *J*  
625 *Clin Epidemiol* 2006; 59:90-3.
- 626 47. Dijk FN, de Jongste JC, Postma DS, Koppelman GH. Genetics of onset of asthma. *Curr*  
627 *Opin Allergy Clin Immunol* 2013; 13:193-202.

- 628 48. Granell R, Henderson AJ, Timpson N, St Pourcain B, Kemp JP, Ring SM, et al.  
629 Examination of the relationship between variation at 17q21 and childhood wheeze  
630 phenotypes. *J Allergy Clin Immunol* 2013; 131:685-94.
- 631 49. Savenije OE, Mahachie John JM, Granell R, Kerkhof M, Dijk FN, de Jongste JC, et al.  
632 Association of IL33-IL-1 receptor-like 1 (IL1RL1) pathway polymorphisms with  
633 wheezing phenotypes and asthma in childhood. *J Allergy Clin Immunol* 2014; 134:170-7.
- 634 50. van der Valk RJ, Duijts L, Timpson NJ, Salam MT, Standl M, Curtin JA, et al. Fraction  
635 of exhaled nitric oxide values in childhood are associated with 17q11.2-q12 and 17q12-  
636 q21 variants. *J Allergy Clin Immunol* 2014; 134:46-55.
- 637 51. Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, et al.  
638 A genome-wide association study identifies CDHR3 as a susceptibility locus for early  
639 childhood asthma with severe exacerbations. *Nat Genet* 2014; 46:51-5.
- 640 52. Fuertes E, Soderhall C, Acevedo N, Becker A, Brauer M, Chan-Yeung M, et al.  
641 Associations between the 17q21 region and allergic rhinitis in 5 birth cohorts. *J Allergy*  
642 *Clin Immunol* 2015; 135:573-6.
- 643 53. Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide  
644 association analysis identifies 11 risk variants associated with the asthma with hay fever  
645 phenotype. *J Allergy Clin Immunol* 2014; 133:1564-71.
- 646 54. Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, et al. A genome-  
647 wide association meta-analysis of self-reported allergy identifies shared and allergy-  
648 specific susceptibility loci. *Nat Genet* 2013; 45:907-11.
- 649 55. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, et  
650 al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial  
651 infarction. *Nat Genet* 2009; 41:342-7.

- 652 56. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom:  
653 regional visualization of genome-wide association scan results. *Bioinformatics* 2010;  
654 26:2336-7.
- 655 57. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and  
656 haplotype maps. *Bioinformatics* 2005; 21:263-5.
- 657
- 658



659 Table I. Top SNPs in main loci associated with time-to-asthma onset at genome-wide ( $P \leq 5 \times 10^{-8}$ ) and suggestive significance levels ( $5 \times 10^{-8} < P < 10^{-6}$ )  
660  $^{\circ} < P < 10^{-6}$ )

Chromosome	Marker	Position *	Nearest gene(s) (kb distance)	Effect/ Ref Alleles †	Effect Freq	Time to asthma onset - N=13,886		
						Hazard Ratio [95% CI]	P ‡	P-Het **
<b>Loci with genome-wide significance - <math>P \leq 5 \times 10^{-8}</math></b>								
2q12	rs10208293	102966310	<i>IL1RL1</i> ***	G/A	0.73	1.14 [1.08-1.19]	$3.1 \times 10^{-8}$	0.26
6p21	rs9272346	32604372	<i>HLA-DQA1</i> (0.8)	A/G	0.59	1.13 [1.08-1.17]	$1.6 \times 10^{-8}$	0.12
9p24	rs928413	6213387	<i>IL33</i> (2)	G/A	0.25	1.19 [1.13-1.25]	$6.5 \times 10^{-16}$	0.15
16q12	rs1861760	50857693	<i>CYLD</i> (22)	A/C	0.04	1.28 [1.17-1.40]	$4.2 \times 10^{-8}$	0.11
17q12-q21	rs9901146	38043343	<i>ZBP2</i> (9) <i>GSDMB</i> (17)	G/A	0.51	1.18 [1.13-1.22]	$1.9 \times 10^{-16}$	0.17
<b>Suggestive loci - <math>5 \times 10^{-8} &lt; P &lt; 10^{-6}</math></b>								
2q11-q12	rs12468899	102426140	<i>MAP4K4</i> ***	G/A	0.69	1.12 [1.09-1.16]	$1.7 \times 10^{-7}$	0.89
15q22	rs11071559	61069988	<i>RORA</i> ***	C/T	0.85	1.16 [1.10-1.24]	$8.3 \times 10^{-7}$	0.96
16p12-p11	rs1805013	27373980	<i>ILAR</i> ***	T/C	0.05	1.22 [1.13-1.32]	$8.0 \times 10^{-7}$	0.37

661 \*Position in base pairs (bp) – build 37.3 NCBI

662 †For the calculation of the hazard ratios, effect alleles were designated as risk alleles. Effect Freq denotes frequency of the effect allele, CI  
663 confidence interval, and Ref reference allele.

664 ‡P-value obtained from single-SNP Cox model for time-to-asthma onset adjusted for sex and principal components (fixed-effect model when  
665 there was no significant evidence of heterogeneity or random-effect model otherwise)

666 \*\*P-Het reflects P-value of Cochran's Q statistic across studies

667 \*\*\*SNP is located within reported gene

668

669 Table II. Secondary signals associated with time-to-asthma onset after stepwise conditional analysis in 9p24 and 17q12-q21 regions. This table  
 670 contains, for these two regions, the top time-to-asthma onset SNP in bold (rs928413 and rs9901146 respectively) and the most significant SNP in  
 671 the conditional analysis, after fitting the lead SNP in the region.

Chr	Marker	Nearest gene (kb distance)	Position *	Effect/Ref Alleles <sup>†</sup>	Effect Freq	Single SNP analysis			Fitted SNP(s)		
						Hazard Ratio [95% CI]	P <sup>‡</sup>	P-Het <sup>**</sup>	Hazard Ratio [95% CI]	P <sup>‡</sup>	P-Het <sup>**</sup>
<b>9p24 region</b>						<b>rs928413</b>					
9	rs413382	<i>IL33</i> (73)	6142948	A/C	0.80	1.15 [1.08-1.22]	5.9x10 <sup>-8</sup>	0.84	1.13 [1.06-1.20]	9.7x10 <sup>-6</sup>	0.80
9	<b>rs928413</b>	<i>IL33</i> (2)	6213387	G/A	0.25	1.19 [1.13-1.25]	6.5x10 <sup>-16</sup>	0.15	-	-	-
<b>17q12-q21 region</b>						<b>rs9901146</b>					
17	<b>rs9901146</b>	<i>ZPBP2</i> (9)	38043343	G/A	0.51	1.18 [1.13-1.22]	1.9x10 <sup>-16</sup>	0.17	-	-	-
17	rs3859192	<i>GSDMA</i> <sup>***</sup>	38128648	T/C	0.48	1.16 [1.12-1.21]	1.5x10 <sup>-13</sup>	0.90	1.11 [1.06-1.15]	4.0x10 <sup>-6</sup>	0.74

672 \*Position: Position in base pairs (bp) - build 37.3 NCBI

673 †For the calculation of the hazard ratio, effect alleles were designated as risk alleles. Effect Freq denotes frequency of the effect allele, CI

674 confidence interval, and Ref reference allele

675 †P-values are obtained from Cox model of time-to-asthma onset adjusted for sex and principal components

676 \*\*P-Het reflects P-value of Cochran's Q statistic across studies

677 \*\*\*SNP is located within reported gene

678

679 Table III. Main cis-eQTLs results for the top SNPs in genome-wide associated regions from the meta-analysis of time-to-asthma onset. We  
 680 focused on eQTLs measured in blood, lymphoblastoid cell lines (LCLs) and lung tissue.

Locus	SNP*	Alleles (Ref/Effect)	Gene(s)	Range of P-values	Tissue	Source <sup>††</sup>
2q12	<b>rs10208293</b>	G/A	<i>IL18RAP, IL18R1</i>	$2.5 \times 10^{-13}$ - $9.8 \times 10^{-198}$	Blood, LCLs	Blood eQTLs, eQTL Browser
6p21	<b>rs9272346</b>	G/A	<i>HLA-DQA1/DQA2/DQAS1/ DQB1/DQB2,HLA-DRA/DRB1/ DRB5/DRB6, TAP2</i>	$1.3 \times 10^{-6}$ - $2.1 \times 10^{-121}$	LCLs, Lung, Blood	eQTL_Chicago,GTEEx, Blood eQTLs
16q12	<b>rs1861760</b>	C/A	<i>NOD2</i>	$3.6 \times 10^{-11}$	Blood	Blood eQTLs
	rs5743266 <sup>†</sup>		<i>CYLD, NOD2</i>	$5.0 \times 10^{-9}$ - $3.2 \times 10^{-120}$	Blood	Blood eQTLs
	(D'=1, r <sup>2</sup> =0.02)					
	rs7205760 <sup>†</sup>		<i>CYLD, NOD2</i>	$2.8 \times 10^{-6}$ - $4.0 \times 10^{-15}$	Lung, Blood	Lung eQTLs, Blood eQTLs
	(D'=1, r <sup>2</sup> =0.005)					
17q12-q21	<b>rs9901146</b>	A/G	<i>GSDMB, ORMDL3</i>	$3.8 \times 10^{-6}$ - $9.8 \times 10^{-198}$	Blood, LCLs	Blood eQTLs, GTEEx, eQTL Browser, eQTL_Chicago
	<b>rs3859192</b>	C/T	<i>GSDMA, GSDMB, ORMDL3</i>	$1.1 \times 10^{-7}$ - $2.5 \times 10^{-12}$	Lung, LCLs	GTEEx, eQTL Browser

681 \* Top genome-wide significant SNPs in time-to-asthma onset meta-analysis and secondary associations identified by conditional analyses are  
 682 indicated in bold.

683 †Haplotype reconstruction was done using Haploview; the effect allele of the top SNP (A-rs1861760) is always transmitted with the effect allele  
684 of its proxy (G-rs5743266 and G-rs7205760).

685 ††Interrogated databases: *eQTL Browser* (LCLs of British subjects with asthma or eczema),<sup>18</sup> *Blood eQTL Browser* (non-transformed peripheral  
686 blood sample),<sup>20</sup> *Lung eQTLs* (lung),<sup>17</sup> *GTEX eQTL Browser v4* (several tissues among which blood and lung),<sup>23</sup> and *eQTL Chicago Browser*  
687 (LCLs).<sup>19,21,22</sup>

688 **FIGURE LEGENDS**

689 **Figure 1.** Manhattan plot showing the association P-values of the genome-wide association  
690 results for time-to-asthma onset from the meta-analysis. The  $-\log_{10}$  of the P-value for each of  
691 2,400,368 SNPs (y-axis) is plotted against the genomic position (x-axis). The solid red line  
692 indicates the genome-wide significance threshold of  $P=5 \times 10^{-8}$ .

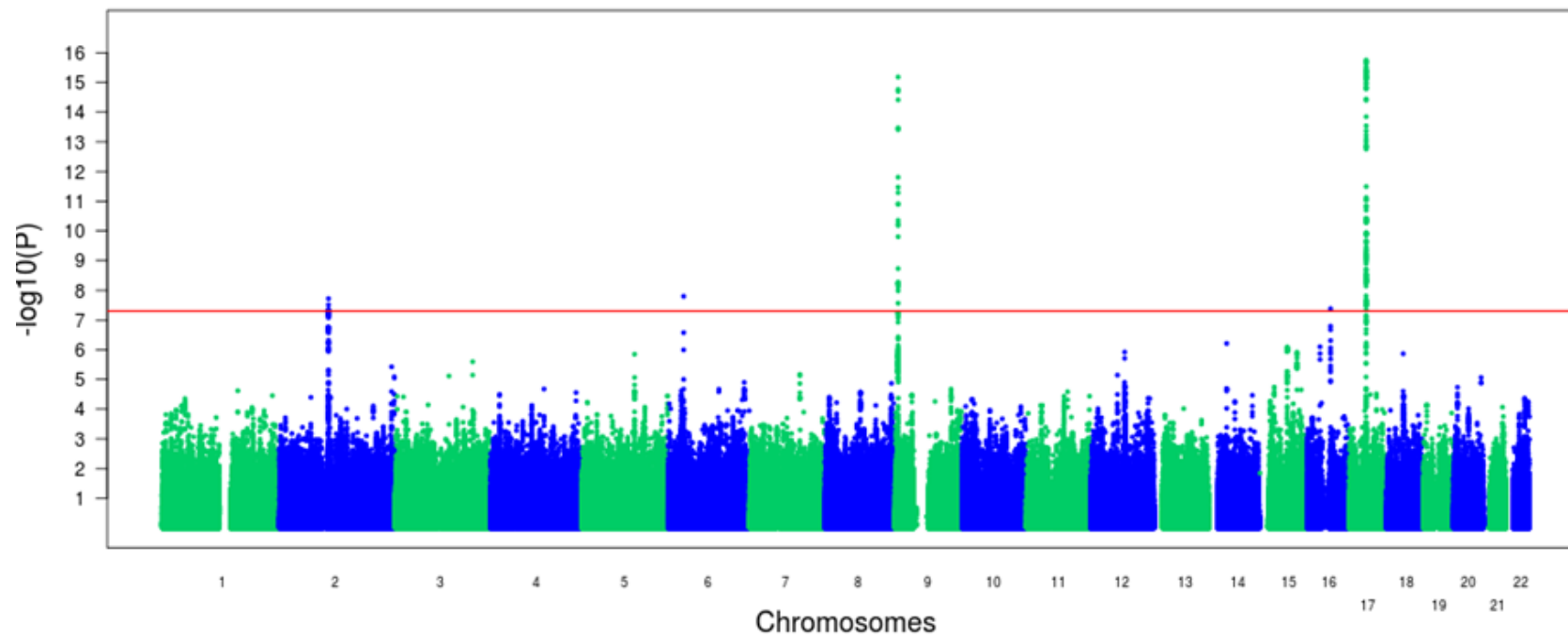
693 **Figure 2.** Regional association plot of the 16q12 region using Locuzoom software.<sup>56</sup> SNPs are  
694 plotted with their P-values ( $-\log_{10}$  values, left y-axis) as a function of genomic position (x-  
695 axis). Estimated recombination rates (right y-axis) taken from 1000G (EUR) are plotted to  
696 reflect the local LD structure around the top associated SNP (purple circle) and correlated  
697 proxies (according to a blue to red scale from  $r^2=0$  to 1).

698 **Figure 3.** Relationship between time-to-asthma onset associated SNPs and age-of-asthma  
699 onset. Part A) Association between age-of-asthma onset and genotypes at individual locus;  
700 Part B) Median of age-of-asthma onset as a function of the number of individual's risk allele  
701 burden; Part C) Median of age-of-asthma onset by quintile of genetic risk score.

702 **Figure 4.** Map of 16q12 region (build 37.3 position: 50,723,355-50,860,722) and haplotype  
703 reconstruction for SNPs found associated with inflammatory bowel disease (among which  
704 Crohn's disease, blue), leprosy (green) or asthma (red) or with the expression of *CYLD* or  
705 *NOD2* (black). Linkage disequilibrium plot was obtained using Hapmap2 CEU reference  
706 sample from Haploview<sup>57</sup> (values and colors reflect  $r^2$  and  $D'$  respectively). The 16q12 top  
707 SNP (rs1861760) associated with time-to-asthma onset is indicated in bold.

708

709 **Figure 1.**

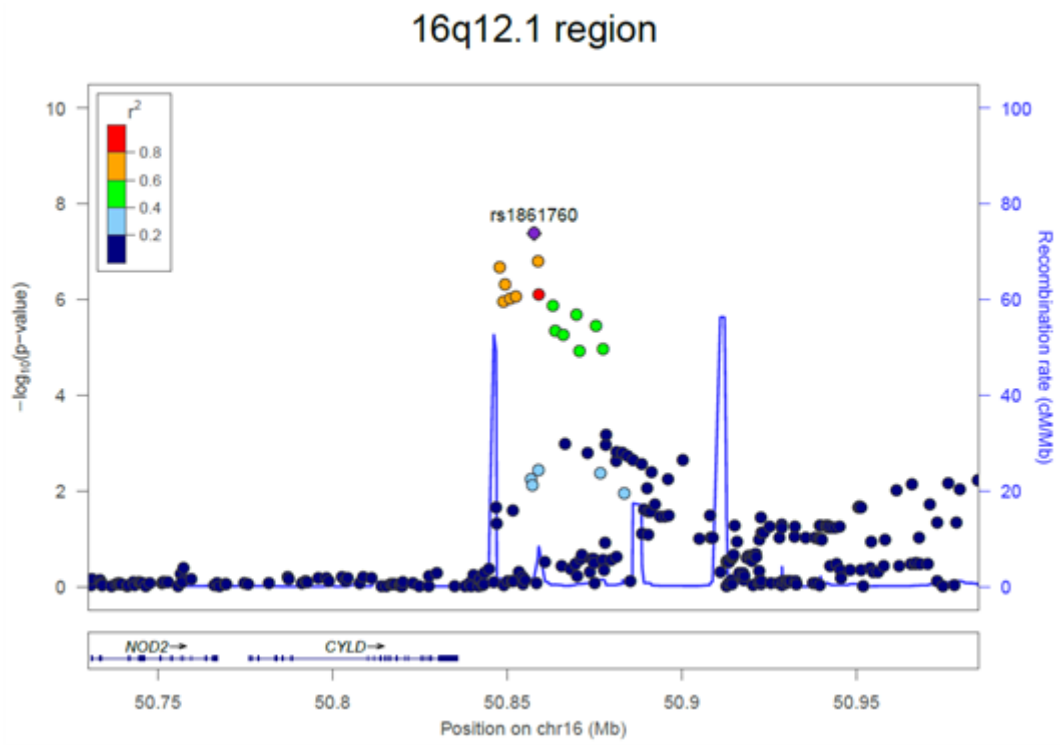


710

711



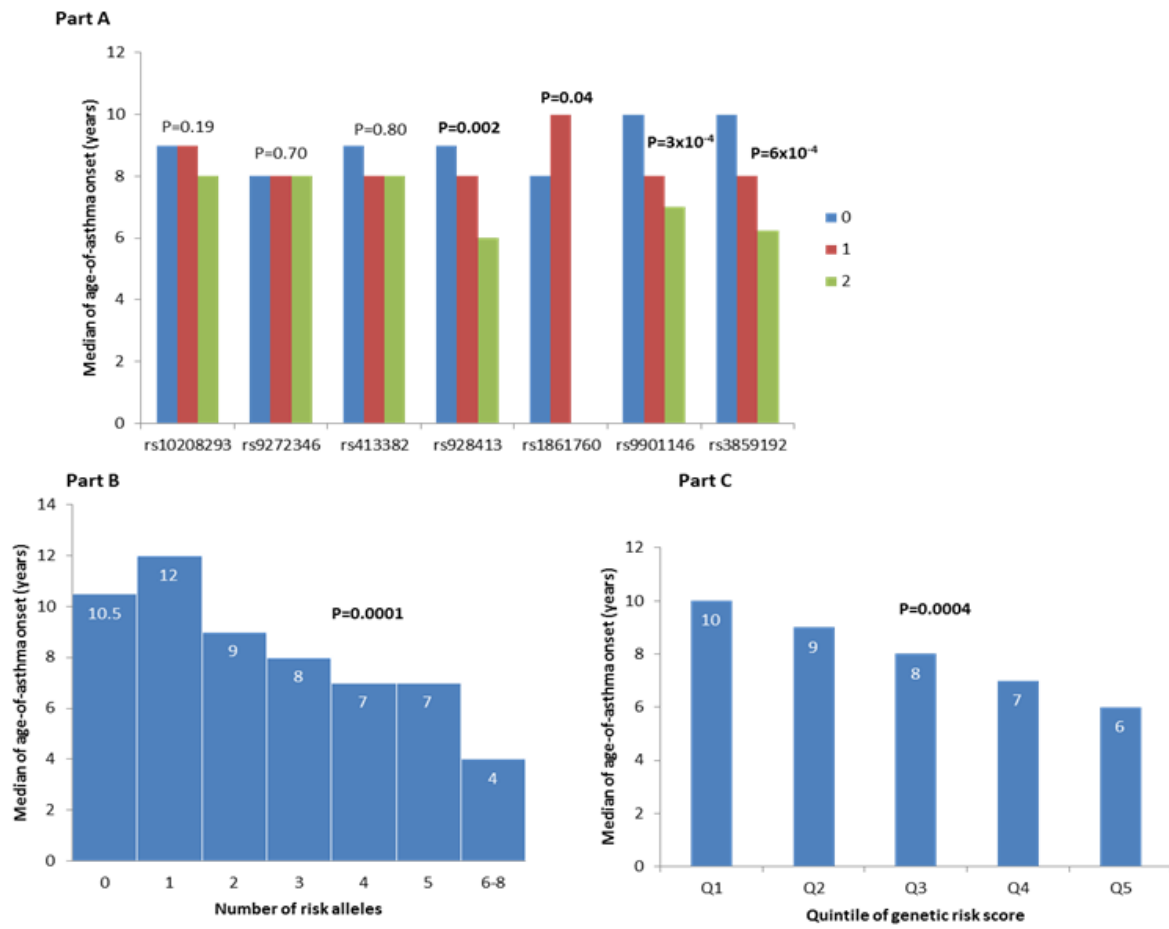
712 **Figure 2.**



713

714

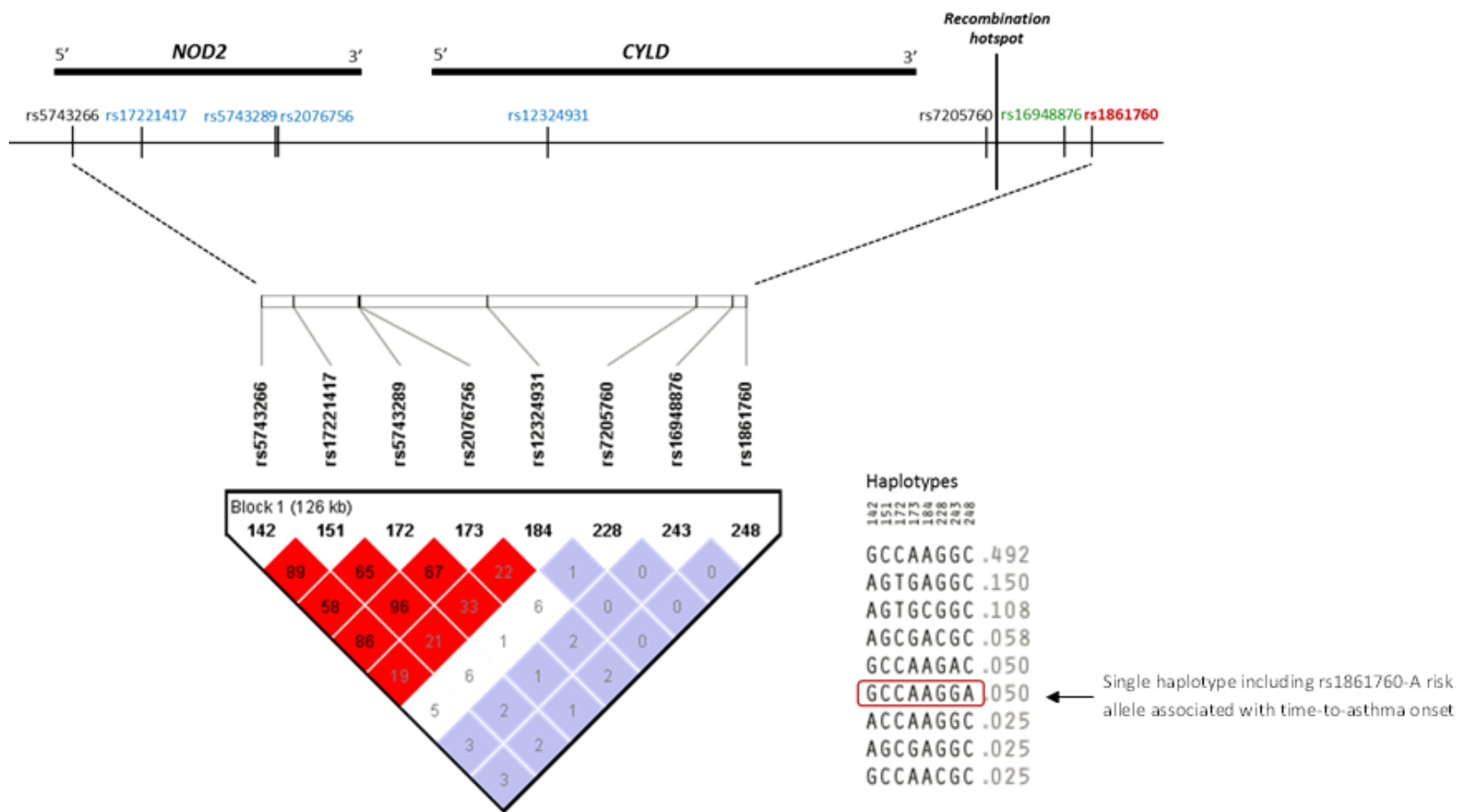
715 **Figure 3.**



716

717

718 **Figure 4.**



719