ERJ Express. Published on December 5, 2013 as doi: 10.1183/09031936.00032713

Genetic heterogeneity of asthma phenotypes identified by a clustering approach

Valérie Siroux[1,2]*, Juan R González[3,4,5]*, Emmanuelle Bouzigon [6,7], Ivan Curjuric [8,9], Anne Boudier [1,2], Medea Imboden [8,9], Josep Maria Anto [3,5,10,11], Ivo Gut [12,13] , Deborah Jarvis [14], Mark Lathrop [7,12], Ernst Reidar Omenaas [15,16], Isabelle Pin [1,2,17], Mathias Wjst [18,19], Florence Demenais [6,7], Nicole Probst-Hensch [8,9], Manolis Kogevinas [3,5,11,20], Francine Kauffmann [21,22]

* equal first

1 Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Inserm, U823, Grenoble, France

2 Univ Joseph Fourier, Grenoble, France

3 Centre for Research in Environmental Epidemiology (CREAL), Barcelona;

4 Department of Mathematics, Autonomous University of Barcelona;

5 CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

6 Inserm, UMRS-946, F-75010 Paris, France

7 Univ Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, F-75007, Paris, France

8 Swiss Tropical and Public Health Institute, Basel, Switzerland

9 University of Basel, Basel, Switzerland

10 Universitat Pompeu Fabra. Departament de Ciències Experimentals i de la Salut, Barcelona;

11 IMIM (Hospital del Mar Medical Research Institute), Barcelona

12 CEA-Centre National de Genotypage, 2 rue Gaston Cremieux, 91057 Evry France

13 Centro Nacional de Analisis Genomico, C/Baldiri Reixac 4, 08028 Barcelona, Spain.

14 Respiratory Epidemiology and Public Health, Imperial College, and MRC-HPA Centre for Environment and Health, London, United Kingdom

15 Bergen Respiratory Research Group, Institute of Medicine, University of Bergen, Bergen, Norway

16 Center for Clinical Research, Haukeland University Hospital, Bergen, Norway

17 Department of pediatrics, CHU Grenoble, France

18 Comprehensive Pneumology Center (CPC), Helmholtz Zentrum Muenchen, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

19 Institute of Medical Statistics and Epidemiology, Klinikum rechts der Isar der TU Muenchen, Munchen, Germany

20 National School of Public Health, Athens, Greece

21 Inserm, U1018, CESP Centre for research in Epidemiology and Population Health, Respiratory and environmental epidemiology Team, Villejuif, France

22 Université Paris Sud, UMRS 1018, Villejuif, France

Corresponding author:

Valérie Siroux

Centre de Recherche INSERM/UJF U823

Institut Albert Bonniot

BP 170

38042 Grenoble Cedex 9

Tel: 33-(0)4-76-54-95-56

Fax: 33-(0)4-76-54-94-13

Mail: valerie.siroux@ujf-grenoble.fr

Funding support:

The EGEA study was supported, in part, by grants from: Merck Sharp and Dohme; the Hospital Program of Clinical Research (Paris, France); National Research Agency (ANR) Health Environment, Health-Work Program; ANR Biological Collections for Health Program; ANR Public Health Research Program ; French Agency of Health Safety; Environment and Work; the Scientific Committee of the Medico-Technology Support at Home (Agir à dom) and Isere Committee Against Respiratory Diseases.

ECRHS: The coordination of ECRHS II was supported by the European Commission, as part of their Quality of Life programme. The following bodies funded the local studies in ECRHS II: Albacete: Fondo de Investigaciones Santarias (FIS) (grants 97/0035-01, 99/0034-01, *and* 99/0034-02), Hospital Universitario de Albacete, Consejeria de Sanidad; Barcelona: SEPAR, Public Health Service (grant R01 HL62633-01), Fondo de Investigaciones Santarias (FIS) (grants 97/0035-01, 99/0034-01, and 99/0034-02) CIRIT (grant 1999SGR 00241) Red Respira ISCII; CIBER Epidemiologia y Salud Publica (CIBERESP), Spain Basel: Swiss National Science Foundation, Swiss Federal Office for Education & Science, Swiss National Accident Insurance Fund (SUVA), USC NIEHS Center grant 5P30 ES07048; Bergen: Norwegian Research Council, Norwegian Asthma & Allergy Association (NAAF), Glaxo Wellcome AS, Norway Research Fund; Erfurt: GSF-National Research Centre for Environment & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code FR 1526/1-1); Galdakao: Basque Health Dept; Grenoble: Programme Hospitalier de Recherche Clinique-

DRC de Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de Grenoble, Ministere de l'Emploi et de la Solidarite, Direction Generale de la Sante, Comite des Maladies Respiratoires de l'Isere; Hamburg: GSF-National Reasearch Centre for Environment & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code MA 711/4-1); Ipswich and Norwich: Asthma UK (formerly known as National Asthma Campaign); Huelva: Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02); Oviedo: Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02); Paris: Ministere de l'Emploi et de la Solidarite, Direction Generale de la Sante, UCB-Pharma (France), Aventis (France), Glaxo France, Programme Hospitalier de Recherche Clinique-DRC de Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de Grenoble; Tartu: Estonian Science Foundation; Umea: Swedish Heart Lung Foundation, Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy Foundation, Swedish Cancer & Allergy Foundation; Uppsala: Swedish Heart Lung Foundation, Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy Foundation, Swedish Cancer & Allergy Foundation; Financial support for ECRHS I for centers in ECRHS II was provided by: Ministere de la Sante, Glaxo France, Insitut Pneumologique d'Aquitaine, Contrat de Plan Etat-Region Languedoc-Rousillon, CNMATS, CNMRT (90MR/10, 91AF/6), Ministre delegue de la sante, RNSP, France; GSF, and the Bundesminister f€ur Forschung und Technologie, Bonn, Germany; Norwegian Research Council project no. 101422/310; Ministero Sanidad y Consumo FIS (grants #91/0016060/00E-05E and #93/0393), and grants from Hospital General de Albacete, Hospital General Juan Ramon Jimenenz, Consejeria de Sanidad Principado de Asturias, Spain; The Swedish Medical Research Council, the Swedish Heart Lung Foundation, the Swedish Association against Asthma and Allergy; Swiss National Science Foundation grant 4026-28099; National Asthma Campaign, British Lung Foundation,

Department of Health, South Thames Regional Health Authority, UK. A.R. was supported by the Department of Health, UK and the European Commission as part of GABRIEL contract number 018996 under the Integrated Program LSH-2004-1.2.5-1.

The Swiss Study on Air Pollution and Lung and Heart Diseases in Adults (SAPALDIA) was supported by the Swiss National Science Foundation (grants no 33CS30_134276/1, 33CSCO-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 3247-065896, 3100-059302, 3200-052720, 3200-042532, 4026-028099, 3233-054996, PDFMP3-123171), the Federal Office for Forest, Environment and Landscape, the Federal Office of Public Health, the Federal Office of Roads and Transport, the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, Valais, Zurich, the Swiss Lung League, the canton's Lung League of Basel Stadt / Basel Landschaft, Geneva, Ticino, Valais and Zurich, Schweizerische Unfallversicherungsanstalt (SUVA), Freiwillige Akademische Gesellschaft, UBS Wealth Foundation, Talecris Biotherapeutics GmbH, and Abbott Diagnostics.

Running Title: Genetics of asthma phenotypes

Word count: 3379

ABSTRACT

The aim was to identify genetic variants associated with refined asthma phenotypes enabling to take into account multiple features of the disease.

Latent class analysis (LCA) was applied in 3001 adults ever having asthma recruited in the frame of three epidemiological surveys (ECRHS, SAPALDIA, EGEA). Fourteen personal and phenotypic characteristics from questionnaires and clinical examination were used. A genome wide association study (GWAS) was conducted for each LCA-derived asthma phenotype, compared to subjects without asthma (n=3474).

The LCA identified four adult asthma phenotypes, mainly characterized by the activity of the disease, the age of asthma onset and the atopic status. Associations of genome wide significance (<1.25x10-7) were observed between "active adult-onset non-allergic asthma" and rs9851461 flanking *CD200* (3q13.2) and between "Inactive/mild non-allergic asthma" and rs2579931 flanking *GRIK2* (6q16.3). Borderline significant results ($2.5x10^{-7}<p<8.2x10^{-7}$) were observed between three SNPs in *ALCAM* region (3q13.11) and "active adult-onset non-allergic asthma". These results were consistent across studies. Fifteen SNPs identified in previous GWAS for asthma have been replicated with at least one asthma phenotype, most of them with the "active allergic asthma" phenotype.

Our results provide evidence that a better understanding of the asthma phenotypic heterogeneity helps to disentangle the genetic heterogeneity of asthma.

Key words

latent class analysis, epidemiology, genetics

INTRODUCTION

Recent genetic studies, including meta-analyses of large-scale genome-wide association studies (GWAS), have successfully identified several genetic loci that influence asthma susceptibility, providing a better understanding of the pathogenesis of this complex disorder [1, 2]. However, only a small proportion of heritability can be explained by the previously identified asthma associated SNPs [3, 4]. The missing heritability could partly reside in the phenotypic heterogeneity of asthma not taken into account in genetic studies.

Asthma is a heterogeneous disease constituted by overlapping separate syndromes probably with different, but yet undefined, etiologies and natural histories [5]. Childhood and adult onset asthma are among the most commonly accepted phenotypes. Interestingly, 17q21 genetic variants were specifically associated with childhood onset asthma in the French Epidemiological Study on the Genetics of Asthma, bronchial hyperresponsiveness and atopy (EGEA), a result further confirmed by a large GWAS conducted within the Gabriel consortium [1, 6]. Such results provided first evidence for a genetic heterogeneity of asthma phenotypes. Unsupervised models aimed to identify homogeneous subgroups of subjects have been applied to unravel the phenotypic heterogeneity of asthma [7]. In adult asthma, these studies led to the identification of asthma phenotypes that exhibited differences in clinical response to treatment, in clinical, physiologic and inflammatory parameters and in health-related quality of life [5, 8-10]. Although such refinement of asthma characterization may shed light on asthma genetics, to date, no genetic association studies have been conducted on asthma phenotypes defined by clustering approach.

We aimed to identify genetic variants associated with cluster-derived asthma phenotypes in a large set of subjects recruited in three large epidemiological studies, ECRHS (the European Community Respiratory Health Survey), SAPALDIA (the Swiss study on Air Pollution and Lung Disease in Adults) and EGEA (ESE Consortium).

METHODS

More information is provided in the online supplementary material.

Study populations

The ECRHS study is a European population-based study of young adults with a 8-year follow-up (ECRHSI (1991-1993), n=18356; ECRHSII (1999-2002), n=10933) [11]. The SAPALDIA study is a cohort study in the Swiss population initiated in 1991 (SAPALDIA 1, n=9 651) with a follow-up assessment in 2002 (SAPALDIA 2, n=8 047) [12]. The EGEA is a French case-control and family-based study with a 12 year follow-up investigation (EGEA1 (1991-1995) n=2047; EGEA2 (2003-2007), n= 1601) [13]. Similar protocols, questionnaires and clinical examination were used in the three studies.

Cluster analysis in adult subjects with ever asthma

In ECRHSII and SAPALDIA2, subjects with asthma answered positively to “Have you ever had asthma?”. In EGEA2 asthma was defined by a positive answer to “Have you ever had attacks of breathlessness at rest with wheezing?” or “Have you ever had asthma attacks?” or being recruited as an asthma case in chest clinics.

We first performed a latent class analysis (LCA) in 3001 adults who ever had asthma (ECRHSII, n=1895; SAPALDIA2, n=465; EGEA2, n=641), irrespective of the availability of genotypes, to define asthma phenotypes. Fourteen variables covering personal characteristics (age and sex), asthma characteristics (age at asthma onset, asthma exacerbations), respiratory symptoms over the past 12 months, allergic characteristics, lung function and bronchial hyperresponsiveness have been considered in the LCA model. Asthma treatment was not included because of the lack of detailed information in the SAPALDIA survey, but sensitivity

analyses conducted in ECRHS and EGEA showed that the model with treatment leads to similar clusters.

Genotypic data

Then we conducted genetic analyses on subjects with genotypic information: 1689 subjects with asthma and 3452 controls without asthma. Genotyped data were available for almost the whole EGEA population. In SAPALDIA and ECRHS the sample included in the genetic analysis represents a nested asthma case-control sample from the cohort (all subjects with asthma at baseline or follow-up with DNA and a random sample of controls).

The subjects were genotyped in the framework of the European GABRIEL consortium. Genotyping was carried out using the Illumina Human610 quad array at the French national genotyping centre (CNG). After quality control of genotyping, as previously described, the number of SNPS analysed was 499 138 [14]. The 39 candidate genetic loci included were those identified in previously published GWAS for asthma [1, 2, 15-24]. If the reported SNP was not genotyped in our data, the closest proxy (among SNPs in strong LD, assessed using a web-based tool (SNAP) with r^2 in the CEU panel of HapMap project) was used [25].

Strategy of analysis

First, we aimed at identifying distinct adult asthma phenotypes by applying LCA, a latent variable model that serves to cluster subjects into classes, as previously used in ECRHS and EGEA [10]. Models with different numbers of latent classes were compared using the Bayesian Information Criterion (BIC) and when BIC were on similar magnitude on the phenotypes prevalence (to avoid low-prevalent phenotype in the prospects of GWAS analysis). Each subject was assigned to the latent class for which he had the highest membership probability. To better characterize the phenotypes observed, smoking, treatment

(in ECRHS and EGEA) and blood eosinophil and neutrophil counts (in EGEA) were compared between LCA-derived phenotypes.

Then, in order to identify genetic variants associated with specific asthma phenotypes, a genome-wide association analyses of each LCA-derived asthma phenotype compared to non-asthma controls were conducted. Genetic associations under a genetic additive model were assessed using logistic regression model using robust sandwich estimation of the variance to model clustering of family genotypes, with adjustment for sex, study (ECRHS, SAPALDIA, EGEA), and informative principal components for within-Europe diversity (snpMatrix R package). The quantile-quantile plots are provided (Figure E1). Lambda values for phenotype A, B, C and D were 1.10, 1.07, 1.11 and 0.95 respectively. On the basis of the effective number of independent markers for the adjustment of multiple testing, p-value $< 1.25 \times 10^{-7}$ was considered as significant in the GWAS [26]. We also reported all SNPs indicating association signals, defined with two consecutive p-value $<10^{-6}$. We further investigated regions of about 20kb upstream and downstream of these loci using the imputed genomic data (estimated by the MACH software http://www.sph.umich.edu/csg/abecasis/MACH and the HapMap2 Release 22 CEU reference sample). Regional association plot for each region was performed using locuszoom (http://csg.sph.umich.edu/locuszoom/). Non-parametric Fisher tests are presented additionally to account for the effect of low MAF in some SNPs (Table E2). To statistically compare the SNP effects across phenotypes, a test for heterogeneity was conducted using a multinomial regression model described in Morris et al [27]. To better interpret our GWAS findings on specific asthma phenotypes in light of the largest asthma GWAS conducted so far on a simple asthma outcome, we provided for each SNP identified in our GWAS the association observed in the meta-analysis in the Gabriel study, after exclusion of the three ESE studies. As a sensitivity analysis, we conducted the GWAS analysis using LCA probabilities (continuous outcomes), to address the robustness of our results to the

outcomes definition. We also investigated the contribution of genetic loci identified for asthma by previous GWAS. In this candidate loci analysis, p values <0.01 were considered significant since only a priori defined candidate genetic loci are tested [28].

RESULTS

The population under study includes 3001 subjects with asthma (mean age 42.9 years, 44% men) and 3452 subjects without asthma (mean age 46.2 years, 48.9% of men) (table 1). Half of the population reported childhood onset asthma (<16 years). Current asthma treatment was reported by 45.2% of the population. In ECRHII and EGEA2 19% reported daily ICS use at the time of the survey. The population participating in the genetic analysis comprised older subjects, more men, more asthmatics with early childhood onset asthma and bronchial hyper-responsiveness (BHR) compared to the non-included population. The description of the asthmatic population for each survey is presented in table E1.

Asthma phenotypes identified by latent class analysis

The 4-class model was retained (figure E2 and online supplementary material). The mean highest posterior probability was high (82%), varying from 80% for phenotype D to 83% for phenotype A and B, indicating that participants were assigned to classes with a fairly high probability. Phenotype A (18%), labeled “Inactive/mild non-allergic asthma” was characterized by individuals with no or few asthma symptoms at the time of examination, low allergic disorders and BHR and high FEV1 (Table 2, figure 1). Phenotype B (37%), labeled “Inactive/mild allergic asthma” was characterized by individuals with no or few symptoms at the time of examination, but presenting atopy and allergic disorders. Phenotype C (27%), labeled “Active allergic asthma”, was composed of younger individuals with childhood-onset asthma, atopy, asthma symptoms as well as BHR at examination. Phenotype D (18%), labeled “Active adult-onset non-allergic asthma”, was characterized by subjects reported adult-onset asthma and asthma symptoms at examination, while few had atopy. Subjects belonging to phenotype D had more often FEV1 <80%predicted as compared to the other groups. In the EGEA dataset, this last phenotype was significantly associated with higher blood neutrophil

counts (neutrophil count/mm3 geometric means were 3801, 3584, 3956 and 4626 for phenotypes A, B, C and D respectively, $p<0.0001$).

The current smoking frequency did not vary strongly across phenotypes (from 22.2% to 25.8%), however subjects included in phenotypes B and C, characterized by a younger age as compared to phenotypes A and D, were more often never-smokers (Table E2). Subjects belonging to phenotypes A and B less often used asthma treatment in the past 3 months as compared to others (Table E2).

The total overall agreement between the latent classes identified and a simple classification defined by the atopy status and asthma attacks in the past 12 months (two highly discriminative variables in the LCA and often collected in epidemiological surveys) was 82.4%, but varied widely between phenotypes (from 92.8% for phenotype A to 61.1% for phenotype D).

GWAS results

Manhattan plots of association results for each asthma phenotype are presented in figure 2. We detected two genome-wide significant associations between "active adult-onset non allergic asthma" (phenotype D) and rs9851461 on chromosome 3, flanking *CD200* ($p=9.4x10^{-9}$), and between "Inactive/mild non allergic asthma" (phenotype A) and rs2579931 on chromosome 6 in *GRIK2* ($p=2.7x10^{-9}$) (Table 3). On the criteria of two consecutive SNPs at $p<10^{-6}$ four chromosomal locations were detected. Pairwise linkage disequilibrium (LD) measures between SNPs of each of these genes showing multiple signals are presented in Table E4. Of the 11 SNPs, three belonging or flanking *ALCAM* gene located on chromosome 3 were associated with phenotype D "active adult-onset non-allergic asthma" (rs9842772, rs9288812, rs1051124 with p-value equal to $2.5x10^{-7}$, $6.6x10^{-7}$ and $8.2\ x10^{-7}$ respectively). The two latter SNPS were not in strong LD with rs9842772 ($r^2=0.39$). Association signals were detected between phenotype "Inactive/mild non-allergic asthma" (phenotype A) and

four SNPs in *LOC401410* on chromosome 7 (rs10264996, rs10259042, rs10230811, rs17162196), two SNPs in *LRRC6* on chromosome 8 (rs7834760, rs13272108) and two SNPs in *SBF2* on chromosome 11 (rs4576815, rs7938647) with p-value ranging from $2.5x10^{-7}$ to $9.6x10^{-7}$. All SNPs detected at $p<10^{-6}$, exhibited statistically significant heterogeneity of effects observed across the asthma phenotypes ($p<0.002$). GWAS on the other two asthma phenotypes did not provide prominent association signals (table 3). None of the SNPs identified in our GWAS showed any trend for association with asthma in the Gabriel meta-analysis after exclusion of the ESE studies ($p>0.50$, table 3. GWAS analyses using the LCA probabilities (continuous outcomes) led to same conclusions (Table E5).

The six loci detected at $p<10^{-6}$ in the present GWAS were investigated using imputed data (figure E3) and the analyses were conducted in the pooled sample as well as separately in each study (forest plots, figure E4). Association signals were consistent across studies for all SNPs, except for rs9851461 flanking *CD200* and phenotype "active adult-onset non-allergic asthma" showing stronger association in EGEA.

The two SNPs flanking *ALCAM* (rs9288812, rs1051124) were significantly related to blood neutrophil counts, the alleles associated with a greater risk for "active adult-onset asthma" (phenotype D) being also associated with an increased level of neutrophils (n=533, adjusted p-values=0.01). No association was detected between these SNPs and blood neutrophils among subjects without asthma and between blood neutrophils and the two other SNPs reported for phenotype D.

Replication of SNPs identified in previous asthma GWAS

Thirteen SNPs located in *IL1RL1, IL18R1, DPP10, TSLP, RAD50-IL13, HLA-DQ, IL33, RORA, ORMDL3/GSDMB*, and *IL12 RB* were replicated ($p<0.01$) with the "Active allergic asthma" (phenotype C), and heterogeneity of association across phenotypes was significant (p for heterogeneity ≤ 0.01) for 8 SNPs located in *IL1RL1, IL18R1, DPP10, TSLP, HLA-DQ*

and *IL33* (Table 4). Six SNPs located in *IL1RL1*, *HLA-DQ*, *IL33* and *SMAD3* were significantly associated with the "Inactive/mild allergic asthma" (phenotye B). A single SNP in *SRP9* (rs4653433), was associated with "active adult-onset non-allergic asthma" (phenotype D) (p -heterogeneity=0.003). None of the 39 SNPs was associated with "Inactive/mild non-allergic asthma" (phenotype A).

DISCUSSION

Applying latent class analysis, a model-based clustering approach, in a large sample of well-characterized subjects with asthma led to the identification of four asthma phenotypes, mainly characterized by the disease activity, the allergic status and the age of asthma onset. Beside these main characteristics, sex, FEV_1 and BHR also played a role in the classification. GWAS on each asthma phenotype revealed a gene of potential interest in "active adult-onset non-allergic asthma", *ALCAM* (activated leukocyte cell adhesion molecule) with evidence of heterogeneity of SNP effect across asthma phenotypes. All replication of asthma SNPs identified by previous GWAS (located in *IL1RL1*, *SMAD3*, RORA, *ORMDL3/GSDMB*, *DPP10*, *TSLP*, *RAD50-IL13*, *HLA-DQ*, *IL33* and *IL12RB*) were observed with the "allergic asthma" phenotypes, except one belonging to *SRP9* gene found to be associated with the "active adult-onset non-allergic asthma". Taken together, our results support the hypothesis that a better understanding of the asthma phenotypic heterogeneity may help to disentangle the genetic heterogeneity of asthma.

One strength of the study relates to the large sample of well-characterized adults with asthma, recruited in three epidemiological settings using standardized protocols and clinical examination, allowing identification of subgroups of subjects with shared characteristics of multiple disease features. As the results of the study designs, mostly population-based, the population includes both persistent and remittent asthma and the prevalence of severe asthma is low in this population. Interestingly, the asthma phenotypes identified by LCA conducted on the pooled EGEA2, ECRHSII and SAPALDIAII population with asthma were highly consistent with phenotypes previously identified in ECRHSII and EGEA2 separately [10]. As previously discussed [5, 10] similarities with previous cluster-derives adult asthma phenotypes [8, 9] regards the identification of a group of subjects with early-onset atopic asthma and groups of subjects with benign (mild) asthma. Further, our phenotype D mainly

characterized by adult-onset non-atopic asthma shows similarity with phenotype 5 described by Moore et al (groups showing higher airflow limitation and exacerbation rate compared to the other phenotypes) [8]. Nevertheless, the phenotypes were defined at one time point and further work is needed in the context of longitudinal data to also account for disease expression variability over time.

Although the phenotypic heterogeneity has been considered as a major limitation in understanding the genetic determinants of asthma, few studies have examined to what extent a better phenotypic resolution leads to identify new genetic determinants. [3] Moffat *et al* previously applied this approach with a single asthma characteristic, age at asthma onset (childhood *versus* adult onset)[1]. Li *et al* performed a GWAS on a population of well characterized patients with severe or difficult-to-treat asthma and identified the RAD50-IL13 region and the HLA DR/DQ region.[17] These studies used stratification on a limited number of traditional phenotypic traits and thus may have had little opportunity to unravel new associations. Our study is the first to perform genetic analysis in a large population based sample in which asthma phenotypes were obtained in an unsupervised manner by means of LCA.

The lack of formal replication of the genetic association signals detected by this GWAS approach is a limitation of our study. Nevertheless, the findings were supported by the association patterns observed within the region and the consistency of the association across studies Replication is particularly challenging here as there are very few large adult studies with similar detailed phenotypic information that enable generation of the phenotypes we have used. The low agreement between the LCA-derived asthma phenotypes and a simple classification based on atopy and the presence of asthma attacks in the past 12 months for phenotype D “active adult-onset non allergic asthma” indicates that these two characteristics are not sufficient to suitably define this phenotype of major interest given our GWAS results.

Overall our GWAS association findings with specific asthma phenotypes have to be interpreted cautiously.

The sample size may be seen as a limitation of the study. However, this power consideration has to be discussed taking into account the improved phenotypic characterization accounting for the disease heterogeneity [29]. Large consortiums on asthma genetics have been set up based on "poor" asthma phenotype definition, which in the context of a highly heterogeneous disease, may explain part of the missing heritability as some genetic effects might be *diluted* as the results of phenotype miss-classification [3]. The approach used in the present analysis, favoring the phenotypic characterization over the sample size, could therefore provide complementary insights to large asthma genetic studies. This is in part supported by the lack of overlap between our GWAS results and the Gabriel results. One limitation on the approach used lies on the difficulty to directly compare findings across studies since cluster-based phenotypes differ between studies.

ALCAM "activated leukocyte cell adhesion molecule", a member of the immunoglobulin superfamily, is a good asthma candidate gene. The *ALCAM* gene encodes the CD166 antigen and was originally identified as a transmembrane receptor that is involved in T-cell activation and may play a role in the binding of T- and B-cells to activated leukocytes. Altered expression of ALCAM has been associated with differentiation state and progression in many tumors [30]. More interestingly, *ALCAM* was identified as a common gene in three inflammatory diseases, Crohn's disease, rheumatoid arthritis and type 1 diabetes [31]. Furthermore, *ALCAM* interacts with *ADAM17* (a disintegrin and metalloproteinase 17) which is implicated in immune cell development and function and has been shown to play a role in epidermal barrier [32, 33]. Interestingly, a further SNP in the *CD200* gene (CD200 molecule), located in the same genomic region as *ALCAM* and encoding a protein also belonging to the immunoglobulin superfamily, was exclusively associated with the "active adult-onset asthma"

phenotype. Nevertheless this result in CD200 should be interpreted cautiously because of the low minor allele frequency of the identified SNP (0.06). To date, there is no strong biological evidence to support the role of the 4 genes showing association signal with "Inactive/mild non-allergic asthma" (*GRIK2* (glutamate receptor, ionotropic, kainate 2, 6q16.3) involved in neurophysiologic processes, *LOC401410* (7q34), *LRRC6* (leucine rich repeat containing 6, 8q24.22) possibly involved in spermatocytogenesis and *SBF2* (SET-binding factor, *11p15.4)* possibly involved in biological processes related to bone and muscle growth [34]). The lack of hits with phenotype C might be explained by a limited statistical power.

None of the variants identified in previous asthma GWAS meet the GWAS significance level in the present study. This might be explained by a lack of power of our study to identify shared genetic variants between phenotypes as compared to previous asthma GWAS which considered the whole group of subjects with asthma. Most of the replications (using the 0.01 threshold) were observed for the "active allergic asthma" phenotype, including SNPs in or near *IL1RL1, IL18R1, DPP10, RAD50-IL13, HLA-DQ, IL33, RORA, ORMDL3/GSDMB*, and *IL12 RB*. This may have occurred because childhood onset asthma is more prone to be associated with allergic phenotypes, and has been more intensively investigated in previous asthma GWAS. Our results are consistent with a specific role of SNPs in *ORMDL3/GSDMB* with childhood-onset asthma (stronger Odds Ratio observed with phenotype C) although the p value for heterogeneity did not indicate marked differences between the described phenotypes (possibly because of our smaller sample size). Our results indicate a specific role of SNP rs1837253 in *TSLP*, a gene involved in the Th2 immune processes in the airways of subjects with asthma, with "active asthma" as supported by the heterogeneity in allelic odds ratios between phenotypes. This suggests that phenotypic heterogeneity may partly explain the genetic heterogeneity previously identified by the GABRIEL study for this SNP.

Our results support the hypothesis that a better understanding of the asthma phenotypic heterogeneity helps to disentangle the genetic heterogeneity of asthma. The etiology of asthma may be clarified, for its genetic and environmental components, by considering specific asthma phenotypes.

AKNOWLEDGMENTS

EGEA

We thank the Epidemiological Study on Genetics and Environment of Asthma (EGEA) cooperative group members as follows. **Coordination** : F Kauffmann; V Siroux (epidemiology) ; F Demenais (genetics); I Pin (clinical aspects) ; R Nadif (biology). **Respiratory epidemiology** : Inserm U 700, Paris M Korobaeff (Egea1), F Neukirch (Egea1); Inserm U 707, Paris : I Annesi-Maesano ; Inserm CESP/U 1018, Villejuif : F Kauffmann, N Le Moual, R Nadif, MP Oryszczyn ; Inserm U 823, Grenoble : V Siroux. **Genetics** : Inserm U 393, Paris : J Feingold ; Inserm U 946, Paris : E Bouzigon , F Demenais, MH Dizier ; CNG, Evry : I Gut (now CNAG), M Lathrop (now CEPH/Mc Gill). **Clinical centers :** Grenoble : I Pin, C Pison; Lyon : D Ecochard (Egea1), F Gormand, Y Pacheco ; Marseille : D Charpin (Egea1), D Vervloet ; Montpellier : J Bousquet ; Paris Cochin : A Lockhart (Egea1), R Matran (now in Lille) ; Paris Necker : E Paty, P Scheinmann ; Paris-Trousseau : A Grimfeld, J Just. **Data and quality management** : Inserm ex-U155 (Egea1) : J Hochez ; Inserm CESP/U 1018, Villejuif : N Le Moual, Inserm ex-U780 : C Ravault ; Inserm ex-U794 : N Chateigner ; Grenoble : J Ferran

The authors thank all those who participated in the setting of the study and on the various aspects of the examinations involved: interviewers, technicians for lung function testing, coders, those involved in quality control and data management, and all those who supervised the study in all centers. The authors are grateful to the 3 CIC-Inserm of Necker, Grenoble and Marseille who supported the study and in which subjects were examined. They are indebted to all the subjects who participated, without whom that study would not have been possible.

ECRHS

We thank the the European Community Respiratory Health Survey study team, a collaboration of European research groups, many of whom also agreed to provide blood samples for genotying as part of the GABRIEL initiative. Investigators in the collaborating centers are Debbie Jarvis, Matthias Wjst, Manolis Kogevinas, Rain Jogi, Christer Janson, Karl Franklin, Ernst Omenaas, Benedicte Leynaert, Isabelle Pin, Joachim Heinrich, Nino Kuenzli, Nicole M. Probst-Hensch, Josep M. Anto, Jordi Sunyer, Jose-Antonio Maldonado, Jesus Martinez-Moratalla, Isabel Urritia, and Felix Payo.

SAPALDIA

We thank the SAPALDIA study team members:

Study directorate: T Rochat (p), JM Gaspoz (c), N Künzli (e/exp), NM Probst Hensch (e/g), C Schindler (s).

Scientific team: JC Barthélémy (c), W Berger (g), R Bettschart (p), A Bircher (a), G Bolognini (p), O Brändli (p), C Brombach (n), M Brutsche (p), L Burdet (p), M Frey (p), U Frey (pd), MW Gerbase (p), D Gold (e/c/p), E de Groot (c), W Karrer (p), R Keller (p), B Knöpfli (p), B Martin (pa), D Miedinger (o), U Neu (exp), L Nicod (p), M Pons (p), F Roche (c), T Rothe (p), E Russi (p), P Schmid-Grendelmeyer (a), A Schmidt-Trucksäss (pa), A Turk (p), J Schwartz (e), D. Stolz (p), P Straehl (exp), JM Tschopp (p), A von Eckardstein (cc), E Zemp Stutz (e).

Scientific team at coordinating centers: M Adam (e/g), E Boes (g), PO Bridevaux (p), D Carballo (c), E Corradi (e), I Curjuric (e), J Dratva (e), A Di Pasquale (s), L Grize (s), D Keidel (s), S Kriemler (pa), A Kumar (g), M Imboden (g), N Maire (s), A Mehta (e), F Meier (e), H Phuleria (exp), E Schaffner (s), GA Thun (g) A Ineichen (exp), M Ragettli (e), M Ritter (exp), T Schikowski (e), G Stern (pd), M Tarantino (s), M Tsai (e), M Wanner (pa)

(a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp) exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational health, (p) pneumology, (pa) physical activity, (pd) pediatrics, (s) statistics

The study could not have been done without the help of the study participants, technical and administrative support and the medical teams and field workers at the local study sites. Local fieldworkers : Aarau: S Brun, G Giger, M Sperisen, M Stahel, Basel: C Bürli, C Dahler, N Oertli, I Harreh, F Karrer, G Novicic, N Wyttenbacher, Davos: A Saner, P Senn, R Winzeler, Geneva: F Bonfils, B Blicharz, C Landolt, J Rochat, Lugano: S Boccia, E Gehrig, MT Mandia, G Solari, B Viscardi, Montana: AP Bieri, C Darioly, M Maire, Payerne: F Ding, P Danieli A Vonnez, Wald: D Bodmer, E Hochstrasser, R Kunz, C Meier, J Rakic, U Schafroth, A Walder.

Administrative staff: C Gabriel, R Gutknecht.

EGEA, SAPPALDIA, and ECRHS were part of the **GABRIEL Consortium**, a European 6th Framework Research project on asthma genetics, which allowed us to obtain the genotype information used in this analysis.

References

1. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WO. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211-1221.

2. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes BE, Levin AM, Mathias RA, Hancock DB, Baurley JW, Eng C, Stern DA, Celedon JC, Rafaels N, Capurso D, Conti DV, Roth LA, Soto-Quiros M, Togias A, Li X, Myers RA, Romieu I, Van Den Berg DJ, Hu D, Hansel NN, Hernandez RD, Israel E, Salam MT, Galanter J, Avila PC, Avila L, Rodriquez-Santana JR, Chapela R, Rodriguez-Cintron W, Diette GB, Adkinson NF, Abel RA, Ross KD, Shi M, Faruque MU, Dunston GM, Watson HR, Mantese VJ, Ezurum SC, Liang L, Ruczinski I, Ford JG, Huntsman S, Chung KF, Vora H, Calhoun WJ, Castro M, Sienra-Monge JJ, del Rio-Navarro B, Deichmann KA, Heinzmann A, Wenzel SE, Busse WW, Gern JE, Lemanske RF, Jr., Beaty TH, Bleecker ER, Raby BA, Meyers DA, London SJ, Gilliland FD, Burchard EG, Martinez FD, Weiss ST, Williams LK, Barnes KC, Ober C, Nicolae DL. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011; 43: 887-892.

3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-753.

4. Wjst M, Sargurupremraj M, Arnold M. Genome-wide association studies in asthma: what they really told us about pathogenesis. *Curr Opin Allergy Clin Immunol* 2013; 13: 112-118.

5. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012; 18: 716-725.

6. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, Chateigner N, Gormand F, Just J, Le MN, Scheinmann P, Siroux V, Vervloet D, Zelenika D, Pin I, Kauffmann F, Lathrop M, Demenais F. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008; 359: 1985-1994.

7. Siroux V, Garcia-Aymerich J. The investigation of asthma phenotypes. *Curr Opin Allergy Clin Immunol* 2011; 11: 393-399.

8. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino JR, Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER. Identification of Asthma Phenotypes using Cluster Analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315-323.

9. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218-224.

10. Siroux V, Basagana X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, Slama R, Jarvis D, Anto J, Kauffmann F, Sunyer J. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011; 38: 310-317.

11. The European Community Respiratory Health Survey II steering committee. The European Community Respiratory Health Survey II. *Eur Respir J* 2002; 20: 1071-1079.

12. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, Bayer-Oglesby L, Baum F, Brandli O, Brutsche M, Downs SH, Keidel D, Gerbase MW, Imboden M, Keller R, Knopfli B, Kunzli N, Nicod L, Pons M, Staedele P, Tschopp JM,

Zellweger JP, Leuenberger P. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. *Soz Praventivmed* 2005; 50: 245-263.

13. Kauffmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, Bousquet J, Neukirch F, Annesi I, Oryszczyn MP, Lathrop M, Demenais F, Lockhart A, Feingold J. Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. *Am J Respir Crit Care Med* 1997; 156: S123-S129.

14. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, Wilk JB, Vonk JM, Thun GA, Siroux V, Nadif R, Monier F, Gonzalez JR, Wjst M, Heinrich J, Loehr LR, Franceschini N, North KE, Altmuller J, Koppelman GH, Guerra S, Kronenberg F, Lathrop M, Moffatt MF, O'Connor GT, Strachan DP, Postma DS, London SJ, Schindler C, Kogevinas M, Kauffmann F, Jarvis DL, Demenais F, Probst-Hensch NM. Genome-wide association study of lung function decline in adults with and without asthma. *J Allergy Clin Immunol* 2012; 129: 1218-1228.

15. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von BA, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SA, Wong KC, Illig T, Vogelberg C, Weiland SK, von ME, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WO. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; 448: 470-473.

16. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, Wang K, Rafaels NM, Michel S, Bonnelykke K, Zhang H, Kim CE, Frackelton EC, Glessner JT, Hou C, Otieno FG, Santa E, Thomas K, Smith RM, Glaberson WR, Garris M, Chiavacci RM, Beaty TH, Ruczinski I, Orange JM, Allen J, Spergel JM, Grundmeier R, Mathias RA, Christie JD, von Mutius E, Cookson WO, Kabesch M, Moffatt MF, Grunstein MM, Barnes KC,

Devoto M, Magnusson M, Li H, Grant SF, Bisgaard H, Hakonarson H. Variants of DENND1B associated with asthma in children. *N Engl J Med* 2010; 362: 36-44.

17. Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA, Bleecker ER. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 2010; 125: 328-335.

18. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, Williams C, Hui J, Beilby J, Warrington NM, James A, Palmer LJ, Koppelman GH, Heinzmann A, Krueger M, Boezen HM, Wheatley A, Altmuller J, Shin HD, Uh ST, Cheong HS, Jonsdottir B, Gislason D, Park CS, Rasmussen LM, Porsbjerg C, Hansen JW, Backer V, Werge T, Janson C, Jonsson UB, Ng MC, Chan J, So WY, Ma R, Shah SH, Granger CB, Quyyumi AA, Levey AI, Vaccarino V, Reilly MP, Rader DJ, Williams MJ, van Rij AM, Jones GT, Trabetti E, Malerba G, Pignatti PF, Boner A, Pescollderungg L, Girelli D, Olivieri O, Martinelli N, Ludviksson BR, Ludviksdottir D, Eyjolfsson GI, Arnar D, Thorgeirsson G, Deichmann K, Thompson PJ, Wjst M, Hall IP, Postma DS, Gislason T, Gulcher J, Kong A, Jonsdottir I, Thorsteinsdottir U, Stefansson K. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009; 41: 342-347.

19. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SA, Klanderman B, Lasky-Su J, Lazarus R, Murphy AJ, Soto-Quiros ME, Avila L, Beaty T, Mathias RA, Ruczinski I, Barnes KC, Celedon JC, Cookson WO, Gauderman WJ, Gilliland FD, Hakonarson H, Lange C, Moffatt MF, O'Connor GT, Raby BA, Silverman EK, Weiss ST. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* 2009; 84: 581-593.

20. Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, Vergara C, Tsai YJ, Yang M, Campbell M, Foster C, Gao P, Togias A, Hansel NN, Diette G, Adkinson NF, Liu MC, Faruque M,

Dunston GM, Watson HR, Bracken MB, Hoh J, Maul P, Maul T, Jedlicka AE, Murray T, Hetmanski JB, Ashworth R, Ongaco CM, Hetrick KN, Doheny KF, Pugh EW, Rotimi CN, Ford J, Eng C, Burchard EG, Sleiman PM, Hakonarson H, Forno E, Raby BA, Weiss ST, Scott AF, Kabesch M, Liang L, Abecasis G, Moffatt MF, Cookson WO, Ruczinski I, Beaty TH, Barnes KC. A genome-wide association study on African-ancestry populations for asthma. *J Allergy Clin Immunol* 2009; 125: 336-346.

21. Hancock DB, Romieu I, Shi M, Sienra-Monge JJ, Wu H, Chiu GY, Li H, del Rio-Navarro BE, Willis-Owens SA, Weiss ST, Raby BA, Gao H, Eng C, Chapela R, Burchard EG, Tang H, Sullivan PF, London SJ. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in mexican children. *PLoS Genet* 2009; 5: e1000623.

22. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, Kurosaka F, Akasawa A, Yoshihara S, Kanno N, Yamada Y, Shimojo N, Kohno Y, Suzuki Y, Kang MJ, Kwon JW, Hong SJ, Inoue K, Goto Y, Yamashita F, Asada T, Hirose H, Saito I, Fujieda S, Hizawa N, Sakamoto T, Masuko H, Nakamura Y, Nomura I, Tamari M, Arinami T, Yoshida T, Saito H, Matsumoto K. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* 2011; 7: e1002170.

23. Ober C, Cox NJ, Abney M, Di Rienzo A, Lander ES, Changyaleket B, Gidley H, Kurtz B, Lee J, Nance M, Pettersson A, Prescott J, Richardson A, Schlenker E, Summerhill E, Willadsen S, Parry R. Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. *Hum Mol Genet* 1998; 7: 1393-1398.

24. Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S, Nyholt DR, Jenkins M, Hayden C, Willemsen G, Ang W, Kuokkanen M, Beilby J, Cheah F, de Geus EJC, Ramasamy A, Vedantam S, Salomaa V, Madden PA, Heath AC, Hopper JL, Visscher PM, Musk B, Leeder SR, Jarvelin M-R, Pennell C, Boomsma DI, Hirschhorn JN,

Walters H, Martin NG, James A, Jones G, Abramson MJ, Robertson CF, Dharmage SC, Brown MA, Montgomery GW, Thompson PJ. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 2011; 378: 1006-1014.

25. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008; 24: 2938-2939.

26. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 2012; 131: 747-756.

27. Morris AP, Lindgren CM, Zeggini E, Timpson NJ, Frayling TM, Hattersley AT, McCarthy MI. A powerful approach to sub-phenotype analysis in population-based genetic association studies. *Genet Epidemiol* 2010; 34: 335-343.

28. Ege MJ, Strachan DP, Cookson WO, Moffatt MF, Gut I, Lathrop M, Kabesch M, Genuneit J, Buchele G, Sozanska B, Boznanski A, Cullinan P, Horak E, Bieli C, Braun-Fahrlander C, Heederik D, von Mutius E. Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol* 2011; 127: 138-144, 144 e131-134.

29. Bennett SN, Caporaso N, Fitzpatrick AL, Agrawal A, Barnes K, Boyd HA, Cornelis MC, Hansel NN, Heiss G, Heit JA, Kang JH, Kittner SJ, Kraft P, Lowe W, Marazita ML, Monroe KR, Pasquale LR, Ramos EM, van Dam RM, Udren J, Williams K. Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. *Genet Epidemiol* 2011; 35: 159-173.

30. Weidle UH, Eggle D, Klostermann S, Swart GW. ALCAM/CD166: cancer-related issues. *Cancer Genomics Proteomics* 2010; 7: 231-243.

31. Eleftherohorinou H, Wright V, Hoggart C, Hartikainen AL, Jarvelin MR, Balding D, Coin L, Levin M. Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. *PLoS ONE* 2009; 4: e8068.

32. Micciche F, Da Riva L, Fabbi M, Pilotti S, Mondellini P, Ferrini S, Canevari S, Pierotti MA, Bongarzone I. Activated leukocyte cell adhesion molecule expression and shedding in thyroid tumors. *PLoS ONE* 2011; 6: e17141.

33. Murthy A, Shao YW, Narala SR, Molyneux SD, Zuniga-Pflucker JC, Khokha R. Notch activation by the metalloproteinase ADAM17 regulates myeloproliferation and atopic barrier immunity by suppressing epithelial cytokine synthesis. *Immunity* 2012; 36: 105-119.

34. Lei SF, Tan LJ, Liu XG, Wang L, Yan H, Guo YF, Liu YZ, Xiong DH, Li J, Yang TL, Chen XD, Guo Y, Deng FY, Zhang YP, Zhu XZ, Levy S, Papasian CJ, Hamilton JJ, Recker RR, Deng HW. Genome-wide association study identifies two novel loci containing FLNB and SBF2 genes underlying stature variation. *Hum Mol Genet* 2009; 18: 1661-1669.

35. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, Radford S, Parry RR, Heinzmann A, Deichmann KA, Lester LA, Gern JE, Lemanske RF, Jr., Nicolae DL, Elias JA, Chupp GL. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med* 2008; 358: 1682-1691.

FIGURE LEGENDS

Figure 1 : Summary of the asthma phenotypes identified using the LCA. The phenotypes are plotted according to the characteristics playing a major role in the classification.

The overlaps between the clusters are proportional to the estimated membership probabilities. As an example, subjects assigned to phenotype B had a mean posterior probability to belong to phenotype A, C and D of 7%, 7% and 2% respectively.

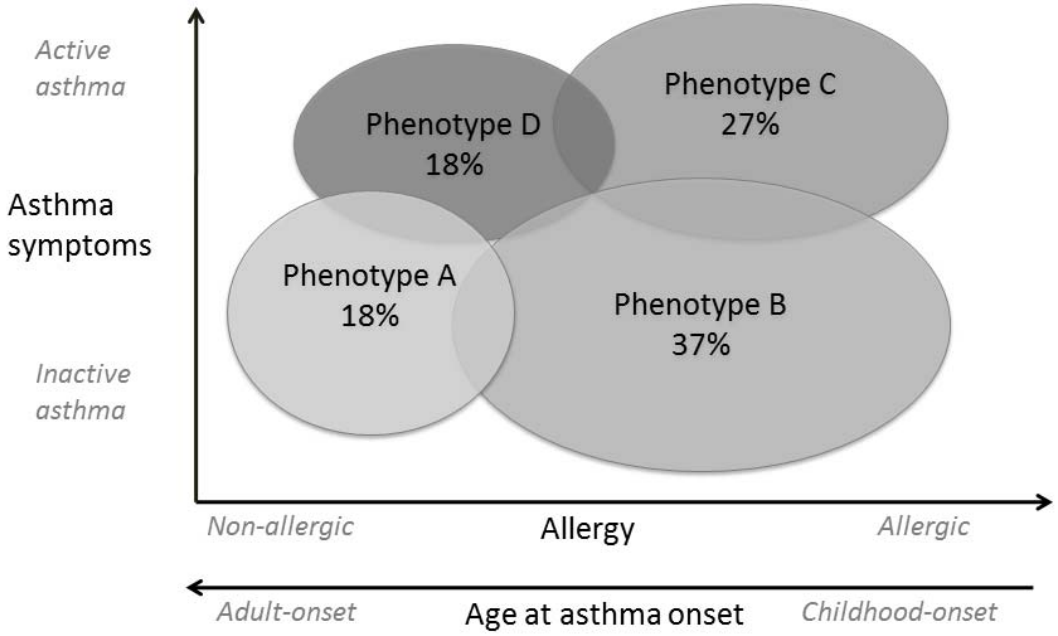
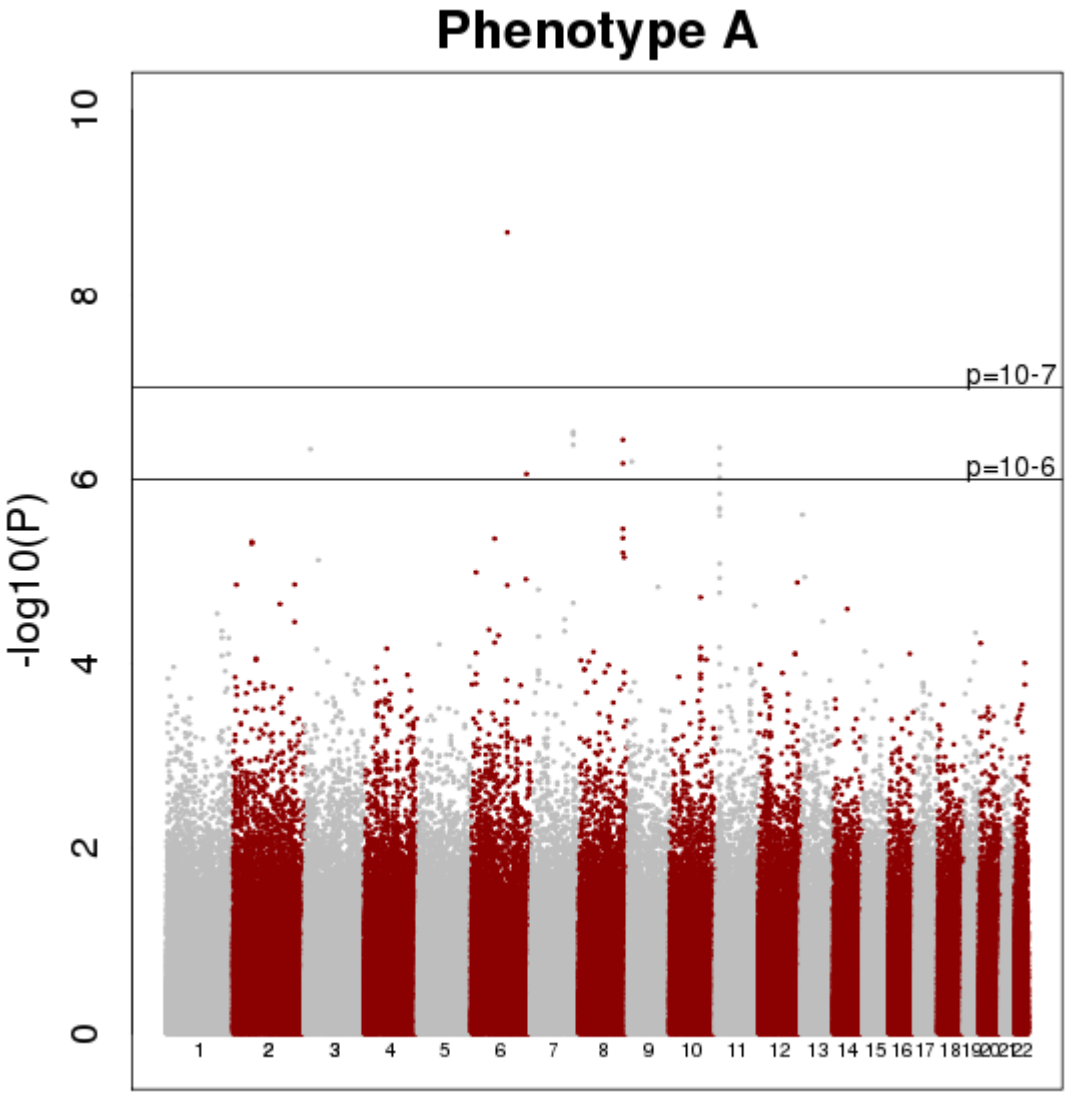
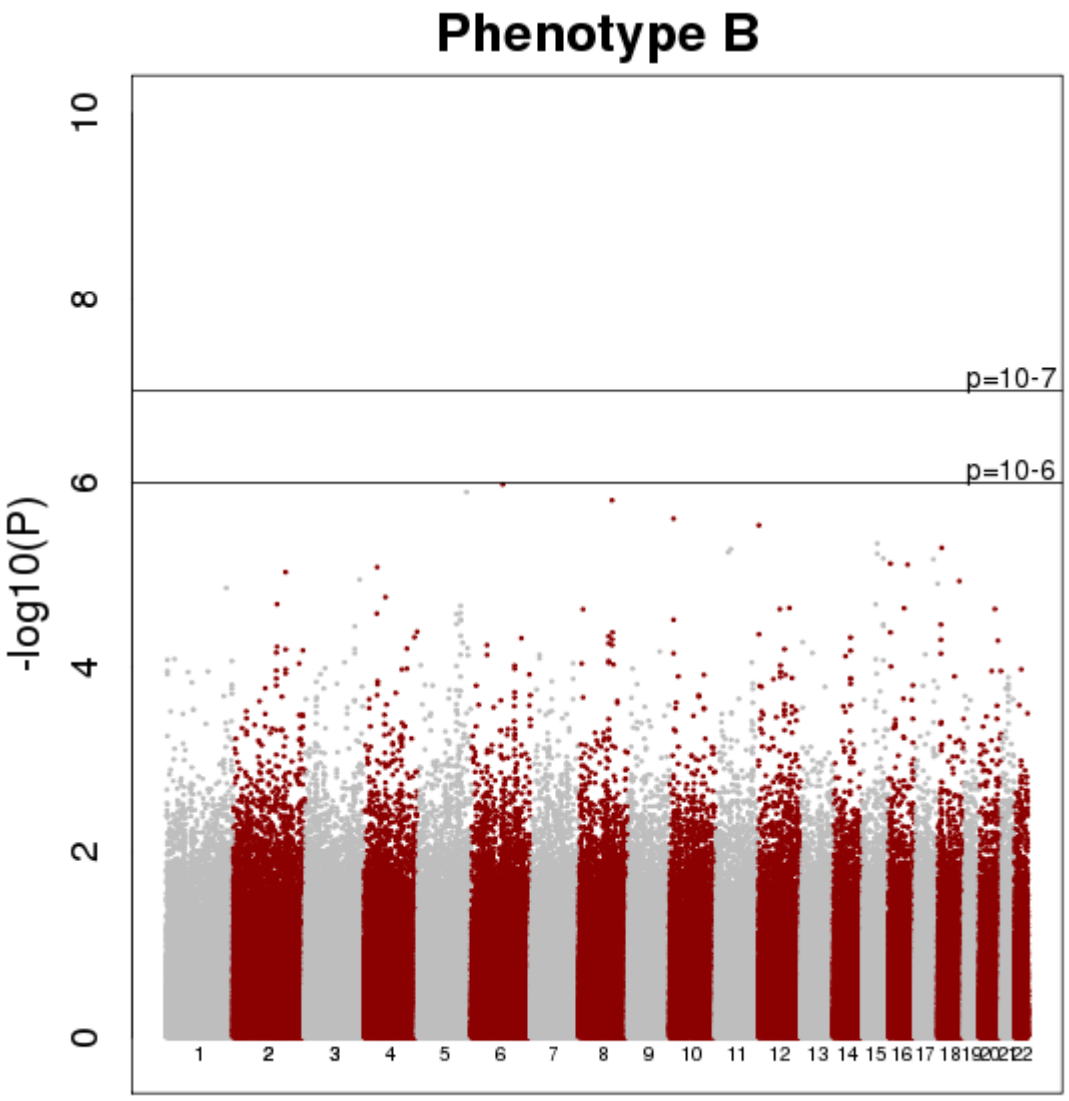
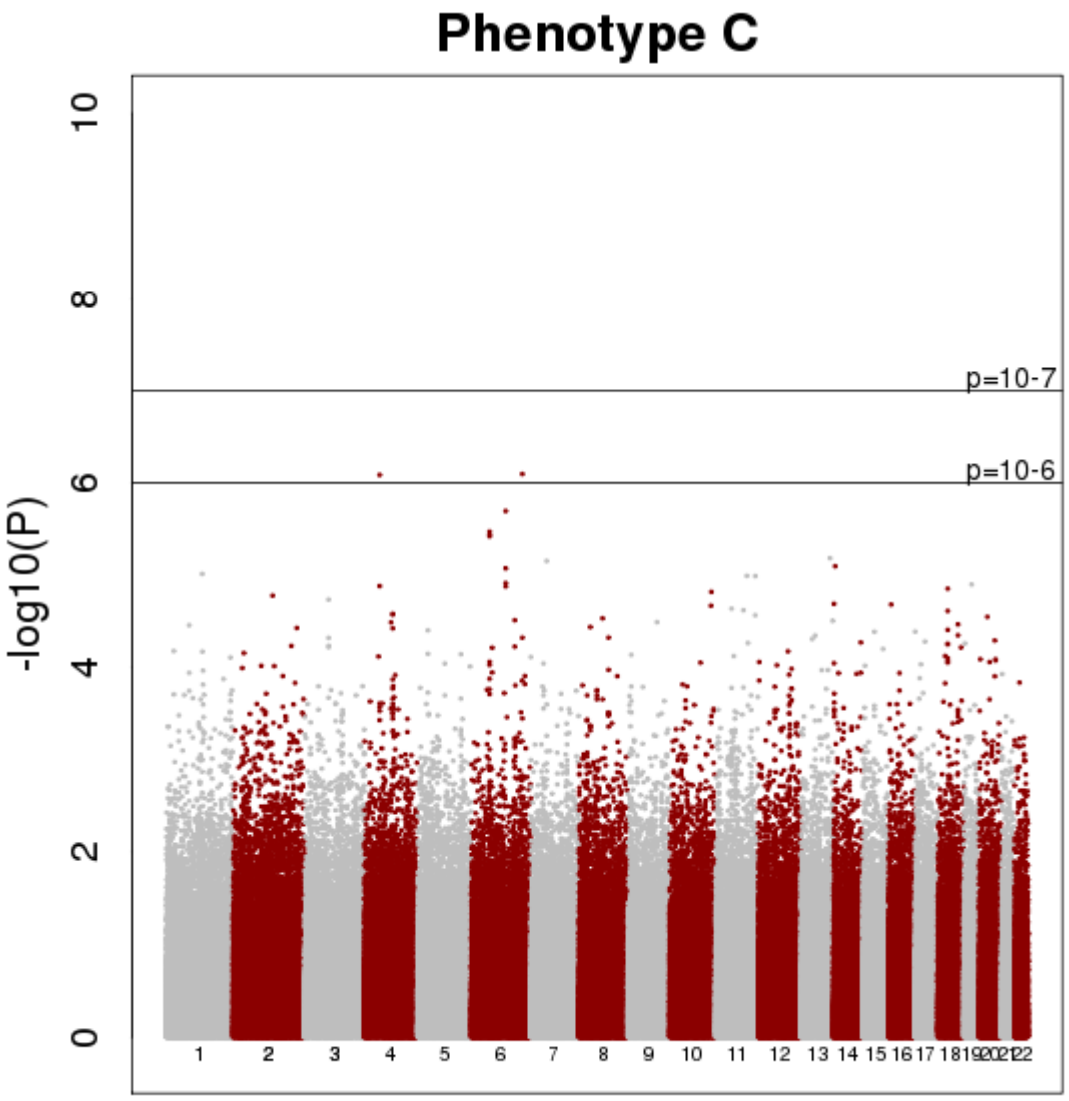


Figure 2: Manhattan plots of association results for each asthma phenotype, A) phenotype A; B) phenotype B, C) Phenotype C and D) phenotype D




Phenotype B
-log10(P)
0
2
4
6
8
10
p=10-7
p=10-6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22


Phenotype C
-log10(P)
0
2
4
6
8
10
p=10-7
p=10-6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

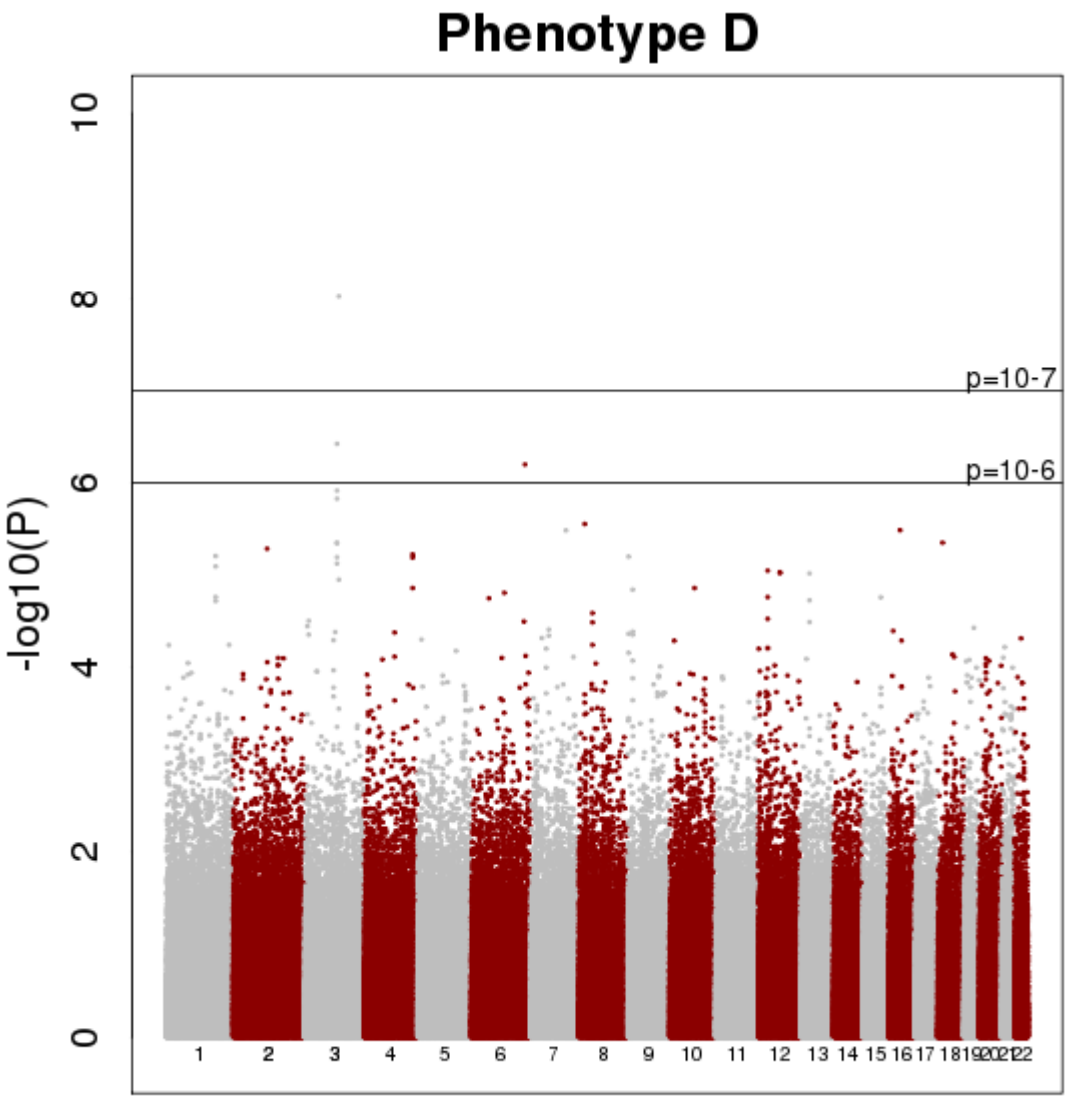

Phenotype D
-log10(P)
0
2
4
6
8
10
p=10-7
p=10-6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Table 1: Description of the population included in the present analysis

	Asthmatics included in the LCA	Asthmatics with GWAS data	Asthmatics without GWAS data
n	3001	1689	1312
Age, mean ±sd	42.9±11.1	43.8±13.1	41.8±7.7*
Sex, men, %	44.3	46.2	41.8*
Age of asthma onset			
≤4 years	19.0	21.0	16.4*
]4-16] years	30.7	29.4	32.3
>16 years	50.3	49.6	51.3
Asthma attack in the past 12 months, %	39.7	37.8	42.1*
Atopy**, %	65.5	65.5	65.4
FEV1 < 80% predicted, %	13.9	14.6	12.9
BHR, PD20≤ 1mg	44.6	42.3	48.5*

*p value comparing subjects with and without GWAS data <0.05

**assessed with skin prick tests or specific IgE

Table 2: Characteristics of the population and probability of individuals presenting the characteristics given membership in each of the 4 phenotypes identified by the latent class analysis

	Frequency of each variable in the whole sample	Phenotype A *Inactive/mild non-allergic asthma*	Phenotype B *Inactive-mild allergic asthma*	Phenotype C *Active Allergic asthma; More often childhood onset asthma and BHR*	Phenotype D *Active adult-onset non allergic asthma; More women*
Subjects, total (ESE),% (n)	100 (3001)	18 (554)	37 (1100)	27 (810)	18 (537)
By study : ECRHS, % (n)	100 (1895)	18 (328)	34 (648)	28 (533)	20 (386)
SAPALDIA, % (n)	100 (465)	31 (143)	38 (176)	13 (62)	18 (84)
EGEA, % (n)	100 (641)	13 (83)	43 (276)	34 (215)	10 (67)
Age ≥40 years	0.59	0.76	0.53	0.42	0.81
Sex, men	0.44	0.33	0.54	0.48	0.29
Age of asthma onset					
≤4 years	0.19	0.17	0.21	0.28	0.03
]4-16] years	0.31	0.18	0.40	0.41	0.10
>16 years	0.50	0.65	0.39	0.31	0.87
Woken by coughing, past 12 months	0.43	0.37	0.25	0.50	0.75
Asthma symptom score, past 12 months: 0 symptom	0.26	0.48	0.47	0.00	0.01
1 or 2 symptoms	0.41	0.48	0.50	0.28	0.33
≥3 symptoms	0.33	0.04	0.03	0.72	0.66
Chronic cough or phlegm	0.20	0.15	0.09	0.23	0.45
Asthma attack, past 12 months	0.40	0.06	0.09	0.81	0.75
Exacerbation*, past 12 months	0.11	0.02	0.02	0.19	0.23
Eczema	0.55	0.51	0.56	0.63	0.45
Rhinitis	0.64	0.32	0.74	0.80	0.54
Atopy**	0.65	0.01	0.98	0.98	0.18
IgE ≥100 IU/ml	0.48	0.09	0.59	0.77	0.24
FEV1 <80% predicted	0.14	0.08	0.09	0.17	0.25
BHR, PD20≤ 1mg	0.45	0.19	0.39	0.72	0.47

* Asthma exacerbation was defined as either the hospitalization for asthma or the use of oral steroids used in the past 12 months

** Atopy was assessed with skin prick test or specific IgE

Table 3: Results from the GWAS analysis

Chr	Gene	Rs number	Position	Ref/alt all.*	Alt all. freq	Phenotype A (n=331) vs controls (n=3452)		Phenotype B (n=618) vs controls (n=3452)		Phenotype C (n=453) vs controls (n=3452)		Phenotype D (n=287) vs controls (n=3452)		P for heterogeneity***	Association with asthma in the Gabriel study after exclusion of ESE studies (Moffatt et al. 2010)	
						OR [95%CI]	P	OR [95%CI]	P	OR [95%CI]	P	OR [95%CI]	p		OR [95%C]	p
3	ALCAM	rs9842772	106735892	G/A	0.86	1.01 [0.81;1.28]	0.98	1.09 [0.91;1.29]	0.41	1.08 [0.89;1.33]	0.42	1.98 [1.45;2.71]	2.5e-07	3.3e-04	*0.98 [0.92;1.04]*	0.50
3	*ALCAM***	*rs9288812*	*106783827*	*A/G*	*0.80*	*0.99 [0.82;1.19]*	*0.71*	*1.06 [0.91;1.23]*	*0.71*	*1.02 [0.86;1.21]*	*0.91*	*1.63 [1.3;2.05]*	*8.2e-07*	*0.005*	*1.00 [0.94;1.05]*	*0.88*
3	*ALCAM***	*rs10511245*	*106783905*	*G/A*	*0.80*	*0.99 [0.82;1.2]*	*0.74*	*1.05 [0.9;1.21]*	*0.81*	*1.03 [0.86;1.22]*	*0.86*	*1.64 [1.31;2.06]*	*6.6e-07*	*0.004*	*1.00 [0.95;1.05]*	*0.94*
3	CD200**	rs9851461	113593771	C/T	0.94	1.05 [0.75;1.45]	0.86	0.88 [0.69;1.12]	0.32	0.93 [0.7;1.24]	0.71	**2.95 [1.7;5.13]**	**9.4e-09**	0.002	0.99 [0.90;1.09]	0.89
6	GRIK2	rs2579931	101961354	A/G	0.90	**2.28 [1.59;3.27]**	**2.7e-09**	1 [0.82;1.22]	0.92	1.19 [0.93;1.53]	0.18	1.02 [0.76;1.37]	0.89	3.4e-04	0.99 [0.92;1.08]	0.84
7	LOC401410**	rs10264996	140731438	A/G	0.95	2.71 [1.49;4.93]	4.3e-07	1.08 [0.83;1.41]	0.68	1.02 [0.76;1.37]	0.83	1.36 [0.91;2.04]	0.06	2.2e-04	1.00 [0.91;1.10]	0.98
7	LOC401410**	rs10259042	140734967	A/G	0.95	2.73 [1.5;4.96]	3.4e-07	1.11 [0.85;1.45]	0.56	0.97 [0.73;1.29]	0.58	1.32 [0.89;1.96]	0.08	1.7e-04	0.99 [0.88;1.10]	0.82
7	LOC401410**	rs10230811	140735196	T/C	0.95	2.73 [1.5;4.96]	3.1e-07	1.11 [0.85;1.45]	0.57	0.99 [0.74;1.32]	0.68	1.32 [0.89;1.97]	0.08	1.9e-04	*1.00 [0.91;1.10]*	0.97
7	LOC401410**	rs17162196	140739352	T/C	0.95	2.73 [1.5;4.96]	3.1e-07	1.12 [0.86;1.46]	0.53	0.98 [0.73;1.3]	0.63	1.38 [0.92;2.07]	0.05	1.4e-04	*1.00 [0.91;1.10]*	0.97
8	*LRRC6*	*rs7834760*	*133714562*	*C/T*	*0.90*	*1.97 [1.39;2.79]*	*6.5e-07*	*1.19 [0.96;1.47]*	*0.08*	*1.04 [0.83;1.31]*	*0.76*	*1.09 [0.83;1.44]*	*0.52*	*1.3e-05*	1.01 [0.93;1.11]	*0.78*
8	LRRC6	rs13272108	133736255	G/A	0.92	2.11 [1.43;3.12]	3.7e-07	1.15 [0.91;1.44]	0.22	0.9 [0.72;1.14]	0.30	0.98 [0.74;1.31]	0.96	5.4e-05	*1.02 [0.92;1.12]*	0.74
11	*SBF2*	*rs4576815*	*9943804*	*C/T*	*0.71*	*1.52 [1.27;1.82]*	*9.6e-07*	*1.02 [0.89;1.16]*	*0.77*	*1.03 [0.88;1.2]*	*0.76*	*0.91 [0.75;1.11]*	*0.37*	*4.9e-05*	0.99 [0.92;1.06]	*0.72*
11	*SBF2*	*rs7938647*	*10017999*	*C/T*	*0.72*	*1.54 [1.28;1.85]*	*6.1e-07*	*1.02 [0.89;1.17]*	*0.74*	*1.01 [0.87;1.18]*	*0.94*	*0.9 [0.74;1.1]*	*0.35*	*4.1e-05*	0.99 [0.92;1.06]	*0.78*

All SNPs showing one p-value < 1.25 x 10e-7 or two consecutive p-values <1.0e-06 with at least one asthma phenotype were reported in this table.
Figures in bold showed significant association (p-value < 1.25e-07)
*reference *versus* alternative allele
*not localized in a gene from dbsnp, but nearby the gene
*** P value for the heterogeneity of the association observed between the asthma phenotypes, assessed with using the multinomial model, as described in Morris et al. [27]

Table 4: Genetic association with candidate SNPs for asthma.

Initial Study	chr	Gene	SNP previously reported	SNP used*	LD (r^2)	Position	Ref/alt All **	Alt. all freq	Phenotype A OR [95%CI]	p	Phenotype B OR [95%CI]	p	Phenotype C OR [95%CI]	p	Phenotype D OR [95%CI]	p	Het. P***
Torgerson [2]	1	CRCT1	rs4845783	rs1053590	0.73	150755052	T/C	0.63	1.01 [0.85;1.20]	0.98	1.03 [0.91;1.17]	0.60	0.86 [0.74;0.99]	0.06	1.13 [0.94;1.35]	0.22	0.34
Ferreira [24]	1	IL6R	rs4129267			152692888	T/C	0.58	0.97 [0.83;1.13]	0.64	0.88 [0.78;0.99]	0.06	0.90 [0.78;1.04]	0.25	0.99 [0.83;1.18]	0.76	0.29
Sleiman [16]	1	DENND1B	rs2786098			195592531	T/G	0.77	0.89 [0.73;1.07]	0.29	0.95 [0.82;1.09]	0.61	1.00 [0.85;1.18]	0.99	1.10 [0.89;1.36]	0.43	0.61
Ober [35]	1	CHI3L1	rs880633	rs762625	0.17	201419424	T/C	0.82	0.85 [0.69 ; 1.04]	0.15	1.03 [0.88;1.21]	0.83	1.04 [0.85;1.28]	0.76	0.98 [0.78;1.22]	0.92	0.29
Torgerson [2]	1	SRP9	rs4653433			224041154	A/G	0.56	0.87 [0.74;1.02]	0.09	0.95 [0.84;1.08]	0.57	1.01 [0.87;1.16]	0.79	**0.75 [0.64;0.89]**	**8.7e-04**	**0.003**
Torgerson [2]	2	IL1RL1	rs10173081	rs13431828	1	102321085	T/C	0.86	1.25 [0.98;1.61]	0.04	**1.37 [1.14;1.65]**	**1.6e-04**	**1.46 [1.16;1.84]**	**4.4e-04**	1.16 [0.90;1.5]	0.23	**4.6e-04**
Gudbjartsson [18]	2	IL1RL1	rs1420101	rs17026974	0.85	102318792	A/G	0.73	0.95 [0.80;1.13]	0.47	0.93 [0.81;1.07]	0.26	0.86 [0.74;1.00]	0.07	1.02 [0.85;1.24]	0.82	0.71
Moffatt [1]	2	IL18R1	rs3771166 (p=6.2e-06)ŧ			102352654	A/G	0.63	1.16 [0.98;1.36]	0.05	1.14 [1.01;1.29]	0.04	**1.31 [1.13;1.52]**	**5.2e-04**	1.04 [0.87;1.24]	0.61	**0.001**
Mathias [20]	2	DPP10	rs1435879			115209357	G/A	0.88	0.98 [0.77;1.26]	0.95	0.84 [0.70;1.00]	0.05	**0.76 [0.62;0.93]**	**0.01**	0.99 [0.75;1.30]	0.97	**0.01**
Torgerson [2]	3	RTP2	rs2017908	rs7616923	1	188903913	T/G	0.87	1.05 [0.83;1.33]	0.58	0.99 [0.82;1.18]	0.89	0.99 [0.80;1.23]	0.87	0.95 [0.73;1.24]	0.74	0.32
Torgerson [2]	4	EPHA5	rs11735820	rs7697951	1	66169736	C/T	0.70	1 [0.85;1.19]	0.86	0.92 [0.81;1.04]	0.13	1.09 [0.94;1.27]	0.19	0.91 [0.76;1.08]	0.29	0.26
Himes [19]	5	PDE4D	rs1588265			59405551	G/A	0.70	1.02 [0.86;1.22]	0.94	0.95 [0.83;1.08]	0.40	0.88 [0.76;1.03]	0.14	1.00 [0.83;1.21]	0.99	0.20
Torgerson [2]	5	TSLP	rs1837253			110429771	T/C	0.76	0.86 [0.72;1.03]	0.12	1.15 [1.00;1,33]	0.07	**1.29 [1.10;1.52]**	**0.002**	1.25 [1.02;1.53]	0.02	**2.0e-04**
Gudbjartsson [18]	5	WDR36	rs2416257			110463389	T/C	0.87	0.83 [0.66;1.04]	0.13	1.12 [0.92;1.35]	0.24	1.22 [0.98;1.51]	0.08	1.12 [0.87;1.45]	0.34	**0.006**
Moffatt [1]	5	SLC22A5	rs2073643 (p=2.4e-06) ŧ			131751187	T/C	0.57	0.92 [0.79;1.08]	0.32	0.91 [0.80;1.03]	0.08	0.90 [0.78;1.04]	0.11	1.03 [0.87;1.23]	0.65	0.15
Li [17]	5	RAD50-IL13	rs2897443			131957493	T/G	0.80	1.04 [0.85;1.26]	0.78	0.96 [0.83;1.12]	0.57	**0.77 [0.66;0.91]**	**0.006**	0.96 [0.78;1.17]	0.67	0.22
Moffatt [1]	5	IL13	rs1295686 (p=1.8e-06) ŧ			132023742	T/C	0.80	1.10 [0.89;1.36]	0.50	0.99 [0.85;1.16]	0.99	0.87 [0.74;1.03]	0.18	1.00 [0.81;1.24]	0.89	0.10
Torgerson [2]	5	GALNT10	rs10064618			153752482	A/G	0.61	1.00 [0.85;1.18]	0.91	1.04 [0.92;1.17]	0.66	0.97 [0.84;1.11]	0.52	1.05 [0.88;1.25]	0.51	0.26
Mathias [20]	5	ADRA1B	rs10515807			159297576	A/G	0.80	1.17 [0.95;1.43]	0.15	1.01 [0.87;1.17]	0.83	1.10 [0.92;1.31]	0.23	1.17 [0.95;1.45]	0.16	0.28
Moffatt [1]	6	HLA-DQ	rs9273349 (p=4.4e-08) ŧ			32733847	T/C	0.59	1.15 [0.97;1.36]	0.12	**1.28 [1.13;1.45]**	**7.3e-05**	**1.33 [1.14;1.54]**	**1.7e-04**	1.19 [0.99;1.43]	0.06	**1.3e-05**

Noguchi [22]	6	HLADP	rs987870			33150858	G/A	0.83	1.16 [0.93;1.46]	0.13	0.87 [0.74;1.02]	0.18	0.79 [0.66;0.94]	0.03	0.9 [0.72;1.13]	0.33	**0.01**
Torgerson [2]	8	FBXO43	rs2453626			101207073	T/C	0.55	1.02 [0.87;1.2]	0.59	0.95 [0.84;1.07]	0.41	1.10 [0.95;1.28]	0.23	1.14 [0.96;1.35]	0.13	0.27
Moffatt [1]	9	IL33	rs1342326 (p=1.2e-07) ŧ			6180076	C/A	0.81	0.83 [0.69;1.01]	0.07	**0.78 [0.68;0.90]**	**0.003**	**0.76 [0.64;0.90]**	**0.006**	0.91 [0.73;1.13]	0.36	**0.005**
Torgerson [2]	9	IL33	rs2381416	rs928413	1	6203387	G/A	0.71	0.81 [0.68;0.97]	0.02	**0.78 [0.68;0.89]**	**7.2e-04**	**0.73 [0.63;0.84]**	**1.6e-04**	0.89 [0.74;1.08]	0.22	**8.1e-05**
Gudbjartsson [18]	9	IL33	rs3939286			6200099	T/C	0.72	0.81 [0.68;0.96]	0.02	**0.79 [0.69;0.91]**	**0.002**	**0.74 [0.64;0.86]**	**3.1e-04**	0.90 [0.74;1.08]	0.23	**2.6e-04**
Hancock [21]	9	TLE4	rs2378383			81229182	G/A	0.89	1.00 [0.77;1.30]	0.94	0.89 [0.74;1.08]	0.16	1.11 [0.88;1.39]	0.52	0.88 [0.68;1.14]	0.44	0.66
Ferreira [24]	11	LRRC32_GARP	rs7130588			75948331	G/A	0.64	0.99 [0.84;1.16]	0.89	0.88 [0.78;1.00]	0.08	0.89 [0.76;1.03]	0.16	1.00 [0.84;1.20]	0.94	0.08
Torgerson [2]	11	C11orf71	rs11214966	rs12223585	0.83	113740900	T/C	0.94	1.31 [0.93;1.85]	0.11	1.28 [0.99;1.66]	0.05	1.04 [0.78;1.38]	0.80	1.01 [0.72;1.41]	0.87	0.62
Torgerson [2]	12	RASSF8	rs16929496	rs10842635	0.96	25924553	A/G	0.78	1.04 [0.85;1.26]	0.52	1.11 [0.95;1.29]	0.16	1.16 [0.97;1.39]	0.17	1.19 [0.96;1.49]	0.08	0.41
Moffatt [1]	15	RORA	rs11071559 (p=5.9e-06) ŧ			58857280	T/C	0.87	0.98 [0.77;1.25]	0.87	1.22 [1.00;1.49]	0.03	**1.48 [1.17;1.86]**	**3.4e-04**	1.26 [0.95;1.65]	0.08	0.02
Moffatt [1]	15	SMAD3	rs744910 (p=3.4e-07) ŧ			65233839	G/A	0.50	0.98 [0.84;1.15]	0.71	**0.83 [0.74;0.94]**	**0.002**	0.85 [0.74;0.97]	0.03	0.82 [0.69;0.97]	0.02	**0.002**
Torgerson [2]	17	AURKB	rs9891949	rs7503353	0.43	8048704	G/T	0.50	1.01 [0.87;1.18]	0.90	1.00 [0.89;1.12]	0.90	0.98 [0.85;1.12]	0.78	1.01 [0.85;1.19]	0.99	0.92
Moffatt [1]	17	GSDMB	rs2305480 (p=1.2e-07) ŧ			35315722	A/G	0.56	1.00 [0.86;1.18]	0.86	1.12 [0.99;1.26]	0.11	**1.26 [1.09;1.46]**	**0.004**	1.14 [0.95;1.36]	0.12	0.03
Moffatt [15]	17	ORMDL3/ GSDMB	rs7216389			35323475	C/T	0.52	1.00 [0.85;1.18]	0.97	1.09 [0.97;1.23]	0.25	**1.25 [1.08;1.44]**	**0.006**	1.13 [0.95;1.35]	0.12	0.05
Moffatt [1]	17	GSDMA	rs3894194 (p=2.0e-08) ŧ			35375519	A/G	0.53	0.95 [0.80;1.12]	0.49	0.94 [0.83;1.06]	0.46	0.85 [0.74;0.98]	0.06	0.86 [0.72;1.03]	0.06	0.32
Mathias [20]	17	GBNA13	rs3972219			60448995	G/A	0.95	1.27 [0.87;1.85]	0.30	1.07 [0.82;1.40]	0.72	1.06 [0.78;1.44]	0.69	1.00 [0.70;1.42]	0.99	0.72
Torgerson [2]	19	C19orf2	rs335016	rs34707	1	35068558	T/G	0.75	0.87 [0.73;1.05]	0.18	0.93 [0.81;1.07]	0.37	0.96 [0.82;1.13]	0.68	1.24 [1.01;1.53]	0.03	0.40
Mathias [20]	20	PRNP	rs6052761			4605017	C/T	0.89	1.05 [0.81;1.36]	0.61	0.79 [0.65;0.95]	0.02	0.85 [0.69;1.05]	0.12	1.25 [0.92;1.68]	0.12	0.04
Moffatt [1]	22	IL12RB	rs2284033 (p=5.5e-07) ŧ			35863980	A/G	0.57	1.00 [0.85;1.17]	0.94	1.04 [0.92;1.17]	0.51	**1.21 [1.05;1.40]**	**0.01**	1.16 [0.98;1.37]	0.08	0.06

* The SNP with the strongest LD with the reported SNP in the literature was used if the SNP reported in the literature was not available in the genotyped data.

** reference versus alternative allele

*** P value for the heterogeneity of the association observed between the asthma phenotypes, assessed with using the multinomial model, as described in Morris et al. [27] ;
Figures in bold show the associations with p value ≤0.001

ŧ Because the GABRIEL paper (refered as Moffatt [1] in this table) includes participants from ECRHS, SAPALDIA and EGEA, we provide here the association from the GABRIEL study after excluding these three studies in the meta-analysis