

Genome-Wide Association Studies of Asthma in Population-Based Cohorts Confirm Known and Suggested Loci and Identify an Additional Association near HLA

Adaikalavan Ramasamy^{1,2,3,9}, Mikko Kuokkanen^{4,9}, Sailaja Vedantam^{5,6,9}, Zofia K. Gajdos^{5,6,9}, Alexsander Couto Alves², Helen N. Lyon^{5,6}, Manuel A. R. Ferreira⁷, David P. Strachan⁸, Jing Hua Zhao⁹, Michael J. Abramson¹⁰, Matthew A. Brown¹¹, Lachlan Coin², Shyamali C. Dharmage¹², David L. Duffy⁷, Tari Haahtela¹³, Andrew C. Heath¹⁴, Christer Janson¹⁵, Mika Kähönen¹⁶, Kay-Tee Khaw¹⁷, Jaana Laitinen¹⁸, Peter Le Souef¹⁹, Terho Lehtimäki²⁰, Australian Asthma Genetics Consortium collaborators[†], Pamela A. F. Madden¹⁴, Guy B. Marks²¹, Nicholas G. Martin⁷, Melanie C. Matheson¹², Cameron D. Palmer^{5,6}, Aarno Palotie^{22,23,24,25}, Anneli Pouta^{26,27}, Colin F. Robertson²⁸, Jorma Viikari²⁹, Elisabeth Widen²³, Matthias Wjst³⁰, Deborah L. Jarvis^{1,31}, Grant W. Montgomery⁷, Philip J. Thompson³², Nick Wareham⁹, Johan Eriksson^{33,34,35,36}, Pekka Jousilahti⁴, Tarja Laitinen^{37,38}, Juha Pekkanen^{39,40}, Olli T. Raitakari^{41,42}, George T. O'Connor^{43,44}, Veikko Salomaa^{4,9}, Marjo-Riitta Jarvelin^{2,31,33,45,46,9}, Joel N. Hirschhorn^{5,6,47,9}

1 Respiratory Epidemiology and Public Health, Imperial College London, London, United Kingdom, **2** Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom, **3** Department of Medical and Molecular Genetics, King's College London, London, United Kingdom, **4** Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **5** Divisions of Genetics and Endocrinology, Children's Hospital, Boston, Massachusetts, United States of America, **6** Broad Institute, Cambridge, Massachusetts, United States of America, **7** The Queensland Institute of Medical Research, Brisbane, Australia, **8** Division of Community Health Sciences, St George's, University of London, London, United Kingdom, **9** MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom, **10** Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia, **11** University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Australia, **12** Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Australia, **13** Skin and Allergy Hospital, Helsinki University Hospital, Helsinki, Finland, **14** Washington University School of Medicine, St. Louis, Missouri, United States of America, **15** Department of Medical Sciences: Respiratory Medicine and Allergology, Uppsala University, Uppsala, Sweden, **16** Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland, **17** Clinical Gerontology Unit, Addenbrooke's Hospital, Cambridge, United Kingdom, **18** Finnish Institute of Occupational Health, Oulu, Finland, **19** School of Paediatrics and Child Health, Princess Margaret Hospital for Children, Perth, Australia, **20** Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere, Finland, **21** Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia, **22** Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **23** Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland, **24** Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **25** Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland, **26** Department of Children, Young People and Families, National Institute for Health and Welfare, Helsinki, Finland, **27** Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, Oulu, Finland, **28** Respiratory Medicine, Murdoch Children's Research Institute, Melbourne, Australia, **29** Department of Medicine, University of Turku, Turku, Finland, **30** Helmholtz Zentrum München German Research Center for Environmental Health, Munich-Neuherberg, Germany, **31** MRC Health Protection Agency (HPA) Centre for Environment and Health, Imperial College London, London, United Kingdom, **32** Lung Institute of Western Australia and Centre for Asthma, Allergy and Respiratory Research, University of Western Australia, Perth, Australia, **33** National Institute for Health and Welfare, Helsinki, Finland, **34** Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland, **35** Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland, **36** Folkhälsan Research Center, Helsinki, Finland, **37** Department of Pulmonary Diseases and Clinical Allergology, Turku University Hospital, Turku, Finland, **38** University of Turku, Turku, Finland, **39** Department of Environmental Health, National Institute for Health and Welfare (THL), Kuopio, Finland, **40** Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland, **41** Research Centre of Applied and Preventive Medicine, University of Turku, Turku, Finland, **42** Department of Clinical Physiology, Turku University Hospital, Turku, Finland, **43** Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **44** The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **45** Institute of Health Sciences, University of Oulu, Oulu, Finland, **46** Biocenter Oulu, University of Oulu, Oulu, Finland, **47** Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Rationale: Asthma has substantial morbidity and mortality and a strong genetic component, but identification of genetic risk factors is limited by availability of suitable studies.

Objectives: To test if population-based cohorts with self-reported physician-diagnosed asthma and genome-wide association (GWA) data could be used to validate known associations with asthma and identify novel associations.

Methods: The APCAT (Analysis in Population-based Cohorts of Asthma Traits) consortium consists of 1,716 individuals with asthma and 16,888 healthy controls from six European-descent population-based cohorts. We examined associations in APCAT of thirteen variants previously reported as genome-wide significant ($P < 5 \times 10^{-8}$) and three variants reported as suggestive ($P < 5 \times 10^{-7}$). We also searched for novel associations in APCAT (Stage 1) and followed-up the most promising variants in 4,035 asthmatics and 11,251 healthy controls (Stage 2). Finally, we conducted the first genome-wide screen for interactions with smoking or hay fever.

Main Results: We observed association in the same direction for all thirteen previously reported variants and nominally replicated ten of them. One variant that was previously suggestive, rs11071559 in *RORA*, now reaches genome-wide significance when combined with our data ($P=2.4\times 10^{-9}$). We also identified two genome-wide significant associations: rs13408661 near *IL1RL1/IL18R1* ($P_{\text{Stage1+Stage2}}=1.1\times 10^{-9}$), which is correlated with a variant recently shown to be associated with asthma (rs3771180), and rs9268516 in the *HLA* region ($P_{\text{Stage1+Stage2}}=1.1\times 10^{-8}$), which appears to be independent of previously reported associations in this locus. Finally, we found no strong evidence for gene-environment interactions with smoking or hay fever status.

Conclusions: Population-based cohorts with simple asthma phenotypes represent a valuable and largely untapped resource for genetic studies of asthma.

Citation: Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, et al. (2012) Genome-Wide Association Studies of Asthma in Population-Based Cohorts Confirm Known and Suggested Loci and Identify an Additional Association near *HLA*. *PLoS ONE* 7(9): e44008. doi:10.1371/journal.pone.0044008

Editor: John R.B. Perry, Peninsula College of Medicine and Dentistry, United Kingdom

Received: April 24, 2012; **Accepted:** July 27, 2012; **Published:** September 28, 2012

Copyright: © 2012 Ramasamy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Please see Supplementary Methods S1 for the details of the many charities, governmental bodies and scientific funding organisations that supported the study recruitment, phenotyping, DNA collection and genotyping for the studies involved in the discovery stage (Stage1) and replication stage (Stage2). The personal research funding supports are listed as follows: AR was supported through the European Commission (through project GABRIEL – contract #018996 under the Integrated Program LSH-2004-1.2.5-1) and the Department of Health, UK. ACA was funded by the European Commission, Framework 7 (grant #223367). The National Health and Medical Research Council, Australia (NHMRC) supported MARF through a project grant (#613627), MAB via a Principal Research Fellowship (#455836) and SCD, DLD, NGM, MCM and GWM via a fellowship scheme. AP acknowledges funding from The Academy of Finland Center of Excellence in Complex Disease Genetics (#213506 and 129680), the Wellcome Trust (#089062), the European Community's Framework 7 Programme and the ENGAGE Consortium (#201413). The Academy of Finland supported EV (#129287, #134839) and VS (#129494, #139635). JE was supported through Academy of Finland, Samfundet Folkhälsan, Finnish Diabetes Research Foundation, Finska Läkaresällskapet, Finnish Foundation for Cardiovascular Research; Yrjö Jahnsson Foundation and Foundation Liv och Hälsa. TL was supported by Foundation of the Finnish Anti-Tuberculosis and Allergy Associations. M-RJ has received financial support from the Academy of Finland, Biocenter Oulu, University of Oulu and National Heart Lung and Blood Institute (NHLBI) – National Institutes of Health (#5R01HL087679-02 through the STAMPEED program 1RL1MH083268-01). JNH was supported by a grant from the American Asthma Foundation for analysis of Framingham SHARe data for association with asthma. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: joelh@broadinstitute.org (JNH); m.jarvelin@imperial.ac.uk (M-RJ); veikko.salomaa@thl.fi (VS)

† These authors contributed equally to this work.

‡ Collaborators of the Asthma Genetics Consortium are listed in Supplementary Methods S1.

Introduction

Asthma, characterized by episodic breathlessness, chest tightness, coughing and wheezing, is estimated to affect 300 million people worldwide [1] and is associated with morbidity and economic costs that are comparable with other common chronic diseases [2,3]. As with many common diseases, asthma risk is determined by both genetic and environmental factors, and estimates of heritability (the proportion of variability in risk within a population due to inherited factors) range from 35 to 90% in twin studies [4,5,6]. However, only a handful of genetic variants have thus far been validated as associated with asthma risk at stringent levels of statistical significance.

The first genome-wide association (GWA) study for asthma [7] was conducted in 2007, with 994 child-onset asthmatics and 1,243 non-asthmatics. This study implicated a locus near *ORMDL3* – a gene not previously suspected to have a role in asthma susceptibility. Although the initial association did not reach a widely used threshold for genomewide significance in discovery samples ($P<5\times 10^{-8}$), the association was subsequently replicated, including numerous independent studies, particularly in European-derived and Hispanic populations [8,9,10,11]. Subsequent GWA studies of asthma identified variants in *PDE4D* [12] and *DENND1B* [13] associated with asthma at genome-wide significant levels, although the proposed variant in *DENND1B* showed considerable heterogeneity of associations between different populations. In a GWA study of severe asthma, two other loci narrowly missed genome-wide significance [14]: one in the *HLA* region, and one near *RAD50*, (a locus previously shown to be

associated with total IgE levels) [15]. A GWA study of eosinophil counts with follow-up testing in asthma case-control studies identified variants near *IL1RL1/IL18R1* and *IL33* as being associated with asthma [16]. More recently, a meta-analysis of 23 GWA studies by the GABRIEL consortium [17] studied 10,365 physician-diagnosed asthmatics and 16,110 participants without asthma. They observed associations with asthma, particularly childhood-onset asthma, at multiple loci (*ORMDL3/GSDMB*, *IL1RL1/IL18R1*, *HLA-DQ*, *IL33*, *SMAD3*, and *IL2RB*). Several additional loci in this study narrowly missed the genome-wide significance threshold. Most recently, GWA studies in largely non-European ancestries [18,19] confirmed some of the loci discovered in Europeans, and identified additional associations at *PYHIN1*, *USP38-GAB1*, an intergenic region of 10p14, and a gene-rich region of 12q13. A very recent GWA study of severe asthma in Europeans found strong evidence for two previously established loci (*ORMDL3/GSDMB* and *IL1RL1/IL18R1*) in patients with severe asthma but did not identify any novel loci [20].

Many of the above findings have come from GWA studies on samples specifically established for the investigation of allergy and asthma, often with rich phenotypic information on asthma and related diseases. While such detailed phenotypic information is extremely valuable, collection of such samples is resource-intensive, potentially limiting the sample size. As for all other polygenic traits, sample sizes of current GWA studies of asthma are a limiting factor in the search for genetic risk factors [21]. Expanding GWA studies to include additional cohorts that may

not have such detailed phenotypic information could increase the power to detect associations.

One route to increasing power may be found in the numerous population-based cohorts with existing genome-wide genotype data that also have basic information on doctor-diagnosed asthma, but have not been comprehensively analyzed for associations with asthma. The idea of reanalyzing existing GWA data is similar to the approach adopted by consortia studying quantitative traits that are routinely measured in many studies, such as height and weight [22,23]. However, unlike anthropometric measures, asthma is a disease where diagnostic criteria may not always be consistent [24,25,26]. As such, it is uncertain whether the rather minimal phenotype of a self-report of doctor-diagnosed asthma is sufficient to be useful in genetic studies of asthma.

In this paper, we test whether a self-report of doctor-diagnosed asthma in population-based cohort could be used to replicate findings previously reported by other asthma GWA studies or identify novel genetic associations with asthma. To achieve this aim, we formed the Analysis in Population-based Cohorts for Asthma Traits (APCAT) genetics consortium, which currently includes 18,604 adults of European ancestry (1,716 cases and 16,888 controls) from six population-based cohorts with GWA data. To study further the top hits emerging from the meta-analyses of APCAT (*Stage 1*), we utilized *in-silico* replication data (*Stage 2*) from 15,286 adults of European ancestry (4,035 cases and 11,251 controls). We were able both to replicate known signals and to identify new associations, indicating that genetic studies of asthma in population-based cohorts are likely to be useful complements to more focused studies of asthma.

Results

The APCAT genetics consortium includes 1,716 individuals with asthma diagnosed (ever) by a physician, and 16,888 non-asthmatic controls, from six population-based cohorts (**Table 1**). Genome-wide genotyping was conducted on available platforms and, after standard quality control (see Methods), subsequently imputed to ~2.5 million autosomal SNPs using the HapMap CEU reference panel (**Table S1**). Association studies with asthma used an additive genetic model and were adjusted for sex, ancestry-informative principal components and (in non-birth cohorts) for age; we also controlled for relatedness in family-based cohorts. We meta-analyzed the study-specific results using a fixed effect model and considered ~2.2 million SNPs with imputation quality >0.30 and minor allele frequency >5%. We applied genomic control at the individual study level and again after meta-analysis to correct for inflation of test statistics due to any systematic bias. The individual study genomic control inflation factors were modest ($\lambda_{GC} \leq 1.02$ for all studies; **Table S1**). We also explored whether stratifying the individual cohorts by smoking status or by the presence of allergic symptoms prior to meta-analysis substantially affected our results. We observed strong correspondences between the unstratified and stratified analyses (**Figure S1**), and therefore primarily report on the unstratified results as a simpler main analysis with slightly improved power.

Analysis of genetic variants previously associated with asthma

To test whether the population-based cohorts and the phenotype of self-reported doctor-diagnosed asthma could be used to detect associations, we analyzed thirteen SNPs in nine genomic loci that had shown genome-wide significant associations in previously published GWA studies of European ancestry

individuals. Encouragingly, the APCAT results (**Table 2**) are directionally consistently for all thirteen SNPs, and nominally replicate (one-tailed $P < 0.05$) ten of these thirteen SNPs. We were unable to assess the *PYHIN1* variant reported as associated with asthma in African-Americans [18] as this variant is monomorphic in European populations. For variants discovered in Japanese individuals [19,27], it is harder to interpret replication in our samples because of differences in LD. Nevertheless, we examined the associations in APCAT for the reported lead SNPs and additional SNPs in LD in the HapMap JPT sample. For two out of the four loci, the associations showed directional consistency, with one association nominally replicated (rs1701704 in *IKZF4*, $p = 0.00132$, see **Table S2**).

We also examined three SNPs where previous evidence was strongly suggestive ($P < 5 \times 10^{-7}$ for association with asthma) but not genome-wide significant (**Table 2**), and one of them, rs11071559 in *RORA*, is strongly supported in our data (Reported $P = 1.1 \times 10^{-7}$; $P_{APCAT} = 0.0031$). The combined evidence of association at *RORA* ($P = 2.4 \times 10^{-9}$) now surpasses the genome-wide significance threshold, and therefore rs11071559 represents a new genome-wide significant association with asthma. We note that the *RORA* locus is 6.4 Mb away from a known asthma variant (rs744910 near *SMAD3*), but these two loci are independent, because there is low linkage disequilibrium between these two variants (pairwise linkage disequilibrium, $r^2 = 0.01$ in HapMap CEU panel) and because the association estimates for these two variants are virtually unchanged when conditioned on each other (**Table S3**).

Search for novel asthma risk loci in APCAT

Having established the validity and utility of our population-based studies, we next examined the results of a genome-wide meta-analysis of the studies within APCAT (**Figure S2**), with the goal of identifying novel associations with asthma. An inspection of the quantile-quantile plots (**Figure S2**) and the low genomic inflation factor ($\lambda = 1.01$) of the meta-analyses indicate that there is little evidence of confounding by population stratification or other technical artifacts.

To test whether some of the top results from the APCAT analysis could represent valid associations with asthma, we considered the most strongly associated 14 SNPs from independent loci within the APCAT results (*Stage 1* in **Table 3**). We obtained *in silico* replication data for 14 top SNPs from several additional studies of asthma, including two population based cohorts. These replication studies consisted of the 1958 British Birth Cohort (B58C), Australian Asthma Genetics Consortium (AAGC), the second survey of the European Community Respiratory Health Survey (ECRHS2) and the European Prospective Investigation of Cancer in Norfolk (EPIC-Norfolk) (**Table S1**). The results from the replication studies (*Stage 2*) and a meta-analysis of these results with the APCAT data (*Stage 1* + *Stage 2*) are summarized in **Table 3**.

The most strongly replicated SNP is rs13408661 near the *IL1RL1* and *IL18R1* genes, with the combined P value reaching genome-wide significance ($P_{\text{Stage1}} = 3.9 \times 10^{-6}$; $P_{\text{Stage2}} = 3.2 \times 10^{-5}$; $P_{\text{Stage1+Stage2}} = 1.1 \times 10^{-9}$; **Figure 1**). This SNP lies approximately 31.1 kb from rs3771166 (pairwise $r^2 = 0.157$) identified by GABRIEL [17] and 2.6 kb from rs1420101 ($r^2 = 0.053$) identified by an earlier study [16] as genetic risk factors for asthma. Conditioning rs13408661 on either rs3771166 or rs1420101 did not substantially reduce the signal of association in APCAT (**Table S3**), implying that rs13408661 represents a different signal of association at this locus. Interestingly, a previous study focusing on previously reported risk loci for asthma in a

Table 1. Characteristics of Stage1 and Stage2 studies.

Study	# Asthmatics cases	# Non-asthmatic Controls	Mean age at questionnaire	% male	% non-smokers	% individuals without hayfever	Genotyping platform ^a
Stage1: APCAT studies for discovery							
FINRISK	160	1,705	52	57.7%	47.8%	58.8%	Illumina Quad 610K (CorroGene), Affymetrix 6.0 (MIGen)
Framingham Heart Study (FHS)	797	6,463	44	43.9%	43.0%	37.7%	Affymetrix 5.0
Health 2000 (H2000)	153	1,841	49	51.1%	47.1%	67.6%	Illumina Quad 610K (GeneMets), Illumina 370K (HDL)
Helsinki Birth Cohort (HBC)	123	1,533	62	43.3%	42.3%	70.4%	Illumina 670K
Northern Finland Birth Cohort 1966 (NFBC1966)	364	3,502	31	48.2%	37.0%	64.4%	Illumina CNV370- Duo
Young Finns Study (YFS)	119	1,844	37	45.7%	49.8%	77.1%	Illumina 670K
Total for Stage1	1,716	16,888			43.3%	55.6%	
Stage2: in-silico replication							
1958 British Birth Cohort (B58C)	986	3,211	42	49.8%	28.9%	78.4%	Affymetrix 500K (WTCCC), Illumina 550K (T1DGC), Illumina Quad 610 (GABRIEL)
Australian Asthma Genetics Consortium (AAGC) ^b	2,110	3,857	34	45.2%	NA	NA	Illumina CNV 370-Duo (28%), Illumina 610K (72%)
European Community Respiratory Health Survey follow-up (ECRHS 2)	600	1,268	42	46.4%	45.7%	63.6%	Illumina Quad 610K
EPIC-Norfolk ^c population based	216	2,005	59	46.8%	45.4%	87.9%	Affymetrix 500K
EPIC-Norfolk ^c obese cases	123	910	60	43.0%	44.3%	87.2%	Affymetrix 500K
Total for Stage2	4,035	11,251					

^aSee **Table S1** for more information on genotyping, imputation and software used. ^b The characteristics of the studies in the AAGC are presented in Ferreira et al., 2011 [8]. ^c EPIC = European Prospective Investigation into Cancer and Nutrition.
doi:10.1371/journal.pone.0044008.t001

Table 2. Results in APCAT for SNPs at loci with strong previously published evidence of association with asthma.

Loci with previous genome-wide significant associations to asthma ($P < 5 \times 10^{-8}$)									
Gene in region ^a	SNP	Study ^b	Effect Allele ^c	Reported		APCAT		APCAT+Reported	
				OR (95% CI)	P value	OR (95% CI)	P value ^d	OR (95% CI)	P value ^e
<i>GSDMA, GSDMB, ORMDL3</i>	rs3894194	1	A	1.17 (1.11,1.23)	4.6E-09	1.11 (1.04,1.18)	2.3E-03	1.15 (1.10,1.19)	1.4E-10
	rs2305480	1	A	0.85 (0.81,0.90)	9.6E-08	0.94 (0.87,1.01)	4.1E-02	0.89 (0.84,0.93)	1.6E-07
	rs7216389	2	T	1.18 (1.11,1.25) ^f	8.5E-08 ^f	1.11 (1.04,1.19)	2.1E-03	1.15 (1.11,1.20)	2.6E-09
<i>IL33</i>	rs3939286	3	T	1.12 (1.07, 1.17)	5.3E-06	1.18 (1.10,1.26)	4.8E-05	1.13 (1.09,1.18)	3.6E-09
	rs1342326	1	C	1.20 (1.13,1.28)	9.2E-10	1.18 (1.09,1.28)	2.5E-04	1.2 (1.15,1.25)	2.0E-12
<i>HLA-DQ</i>	rs9273349	1	C	1.18 (1.13,1.24)	7.0E-14	1.22 (1.07,1.38) ^g	5.6E-03 ^g	1.19 (1.14,1.23)	2.8E-15
<i>IL18R1, IL1RL1</i>	rs3771166	1	A	0.87 (0.83,0.91)	3.4E-09	0.89 (0.82,0.97)	2.4E-03	0.88 (0.84,0.92)	7.0E-11
	rs1420101	3	T	1.16 (1.11, 1.21)	5.5E-12	1.16 (1.08,1.23)	9.9E-05	1.16 (1.12,1.20)	4.9E-15
<i>SMAD3</i>	rs744910	1	A	0.89 (0.86,0.92)	3.9E-09	0.92 (0.85,1.00)	1.7E-02	0.90 (0.86,0.93)	5.7E-10
<i>IL2RB</i>	rs2284033	1	A	0.89 (0.86,0.93)	1.2E-08	0.98 (0.91,1.06)	3.1E-01	0.91 (0.88,0.95)	1.4E-07
<i>IL13</i>	rs1295686 ⁱ	1	C	0.87 (0.83,0.92)	1.4E-07	0.90 (0.82,0.98)	5.4E-03	0.88 (0.84,0.92)	5.9E-09
<i>DENND1B</i>	rs2786098	4	T	0.70 (0.63,0.78)	3.9E-11	0.93 (0.89,1.02)	5.8E-02	0.83 (0.76,0.90)	4.7E-08
<i>PDE4D</i>	rs1588265	5	G	0.85 (0.87,0.93)	2.5E-08	0.96 (0.85,1.06)	2.0E-01 ^h	0.87 (0.82,0.92)	1.1E-07
Loci with previous suggestive associations to asthma ($P < 5 \times 10^{-7}$)									
<i>RORA</i>	rs11071559	1	T	0.85 (0.80,0.90)	1.1E-07	0.86 (0.75,0.97)	3.1E-03	0.85 (0.80,0.91)	2.4E-09
<i>RAD50</i>	rs2244012	6	G	1.64 (1.36,1.97)	3.0E-07	1.05 (0.96,1.14)	1.3E-01	1.14 (1.06,1.22)	1.5E-03
<i>SLC22A5</i>	rs2073643	1	C	0.90 (0.87,0.94)	2.2E-07	0.96 (0.88,1.04)	1.6E-01	0.91 (0.88,0.95)	3.9E-07

^aGene shown is nearest gene to associated SNP. SNPs from the same locus are grouped together. ^bReferences: 1 = Moffatt et al. (2010) [17]; 2 = Moffatt et al. (2007) [7]; 3 = Gudbjartsson et al. (2009) [16]; 4 = Sleiman et al. (2010) [13]; 5 = Himes et al. (2009) [12]; 6 = Li et al. (2010) [22]. ^cAlleles are indexed to the forward strand of NCBI build36. ^dAPCAT *P* values are one-tailed with respect to the direction of the original association. ^e*P* values are from fixed-effect inverse-variance model of meta analysis. ^fResults shown are from Moffatt et al (2010), which is the larger and more recent study. ^g SNP rs9273349 is present in NFBC1966 data set only. ^hResults exclude the Framingham Heart Study, which contributed to the original report in Himes et al (2009) ⁱShown here are the random effects *P* value in Gabriel data, the *P* value for fixed effects model had a genome wide significance *P* value of 1.4E-08 with no evidence of heterogeneity. doi:10.1371/journal.pone.0044008.t002

subsample of the Australian samples in this report [8] had also suggested that an association with rs10197862, which is tightly correlated with rs13408661 ($r^2 = 0.932$), was independent of other associated variants in the region. Quite recently, a study of ethnically diverse sample of Americans [18] identified a genome-wide significant association at rs3771180/rs10173081, which are also tightly correlated with rs13408661 ($r^2 = 0.907$ and 1). Therefore, the combined data, including our conditional analyses, indicate that rs13408661 truly represents an additional genome-wide significant signal of association with asthma at this locus.

Our second strongest signal is at rs9268516 ($P_{\text{Stage1}} = 1.2 \times 10^{-7}$; $P_{\text{Stage2}} = 1.0 \times 10^{-3}$; $P_{\text{Stage1+Stage2}} = 1.1 \times 10^{-8}$; **Figure 2**) in the HLA region on Chromosome 6, approximately 246 kb away from the rs9273349 variant ($r^2 = 0.324$) identified by GABRIEL [17]. Conditional analysis (**Table S3**) indicates that the association at rs9273349 cannot completely explain the association at rs9268516, suggesting either the presence of multiple signals at this locus or the presence of a third variant (partially correlated with both rs9273349 and rs9268516) that explains the association at both of these variants.

None of the other variants had compelling evidence of replication in our Stage2 data (**Table 3**), although rs7861480 near *IFNE* showed a trend in the same direction ($P = 0.064$). We also examined the evidence for association of these 14 SNPs or their proxies in the GABRIEL data (estimates downloaded from <http://www.cng.fr/gabriel/results.html>), excluding B58C and ECRHS (which participated in Stage2 of the APCAT study) and the occupational asthma cohorts. Consistent with our Stage2 results, we

saw directionally consistent evidence of association at rs13408661 ($P = 7.0 \times 10^{-6}$) in the *IL1RL1/IL18R1* locus and rs9268516 in the *HLA* region ($P = 0.069$); rs7861480 near *IFNE* showed directional consistency but was not significantly associated with asthma in GABRIEL ($P = 0.19$).

We also estimated the variance explained in the APCAT by the most strongly associated SNP at each of the previously reported loci. Under a liability threshold [28], and assuming a prevalence of 9% (the prevalence in APCAT), these SNPs together explain ~1.6% of the population variance in asthma risk (**Table S4**). Of course, additional variance is explained by multiple signals at each locus (such as at *IL1RL1/IL18R1* and *HLA* regions), variants at additional loci (such as at *RORA*), and variants yet to be discovered by additional genetic studies.

Search for interactions with smoking and allergic status in APCAT

To our knowledge, there has been no genome-wide search for gene-environment interactions with smoking status or allergic status, two important modifiers for risk of developing asthma. We scanned ~2.2 million SNPs for gene-environment effects for smoking exposure in APCAT by comparing, for each SNP, the estimated association statistics with asthma in current smokers to the estimates in never smokers. The strongest evidence for interaction with smoking status did not reach genome-wide significance [estimated pooled odds ratio for interaction between rs1007026 (nearest genes are *MOCS1* and *DAAM2*) and smoking status = 1.89; 95% confidence interval 1.43 to 2.49,

Table 3. Replication results for top signals from APCAT (Stage1 N= 18,604) in additional studies (Stage 2 N = 15,576) and in GABRIEL.

Nearest Gene(s)	SNP	Chr (position ^a)	Effect/ alternate allele ^a	Effect allele frequency	Stage1 (APCAT) OR (95% CI) and p value	Stage2 (replication) OR (95% CI) and p value	Stage1 + Stage2 (pooled) OR (95% CI) and p value	GABRIEL ^b fixed-effects OR (95% CI) and p value	GABRIEL SNP (^r ² in HAPMAP CEU with APCAT SNP) ^c
IL1RL1/IL18R1	rs13408661	2 (102321514)	G/A	0.84	1.29 (1.16,1.44); P = 3.9E-06	1.19 (1.10,1.29); P = 3.2E-05	1.23 (1.15,1.31); P = 1.1E-09	1.15 (1.09,1.22); P = 7.0E-06	rs3213733 (^r ² = 0.83)
BTNL2/HLA-DRA	rs9268516	6 (32487467)	T/C	0.24	1.26 (1.17, 1.37); P = 1.2E-07	1.11 (1.04,1.17); P = 1.0E-03	1.15 (1.10,1.21); P = 1.1E-08	1.05 (0.99,1.12); P = 6.9E-02	rs8180664 (^r ² = 0.99)
IFNE	rs7861480	9 (21460997)	T/C	0.11	1.28 (1.17,1.39); P = 9.8E-06	1.09 (1.00,1.18); P = 5.7E-02	1.17 (1.10,1.24); P = 1.7E-05	1.04 (0.96,1.12); P = 1.9E-01	rs10811568 (^r ² = 0.87)
CSMD1	rs2977724	8 (4911545)	T/C	0.88	1.34 (1.21,1.46); P = 5.6E-06	1.07 (0.99,1.15); P = 1.0E-01	1.14 (1.08,1.21); P = 1.2E-04	0.98 (0.90,1.05); P = 7.4E-01	Same SNP
SLC6A15/TMT2	rs4882411	12 (83141816)	A/C	0.76	1.22 (1.14,1.31); P = 7.3E-06	1.05 (0.99,1.15); P = 1.0E-01	1.11 (1.06,1.16); P = 8.3E-05	1.03 (0.98,1.08); P = 1.5E-01	rs1564606 (^r ² = 0.91)
ATP2B2	rs26310	3 (10347840)	T/C	0.67	1.22 (1.13,1.30); P = 7.1E-06	1.05 (0.99,1.11); P = 9.2E-02	1.10 (1.05,1.15); P = 7.7E-05	1.01 (0.96,1.06); P = 3.2E-01	Same SNP
DACH1/C13orf37	rs1460456	13 (71402712)	G/A	0.21	1.22 (1.13,1.31); P = 1.5E-05	1.02 (0.93,1.04); P = 7.0E-01	1.10 (1.04,1.16); P = 1.7E-03	1.03 (0.96,1.09); P = 2.2E-01	Same SNP
ZNF479/MIR3147	rs10227804	7 (57341784)	G/A	0.48	1.19 (1.12,1.26); P = 2.5E-06	0.98 (0.93,1.04); P = 5.8E-01	1.06 (1.02,1.11); P = 1.1E-02	0.99 (0.93,1.05); P = 5.9E-01	rs1403937 (^r ² = 0.95)
TRERF1	rs4714586	6 (42352608)	A/G	0.41	1.18 (1.11,1.25); P = 9.3E-06	0.98 (0.92,1.03); P = 4.2E-01	1.05 (1.00,1.09); P = 3.9E-02	0.99 (0.94,1.04); P = 6.7E-01	Same SNP
SPRY1/ANKRD50	rs2553377	4 (170622584)	A/C	0.20	1.23 (1.13,1.32); P = 2.4E-05	0.97 (0.89,1.04); P = 3.8E-01	1.06 (1.00,1.12); P = 5.2E-02	1.01 (0.95,1.08); P = 3.4E-01	Same SNP
MECOM	rs1918969	3 (170622584)	T/C	0.55	1.18 (1.11,1.26); P = 9.0E-06	0.99 (0.93,1.04); P = 6.4E-01	1.05 (1.01,1.10); P = 2.4E-02	1.00 (0.95,1.05); P = 5.4E-01	rs4245909 (^r ² = 0.78)
NEDD4L	rs292448	18 (54047200)	A/C	0.69	1.21 (1.12,1.29); P = 1.2E-05	0.98 (0.92,1.04); P = 5.4E-01	1.05 (1.00,1.10); P = 3.7E-02	1.02 (0.97,1.07); P = 2.7E-01	rs292451 (^r ² = 0.96)
SLC7A11/PCDH18	rs6825001	4 (139122321)	G/A	0.10	1.27 (1.15, 1.39); P = 8.1E-05	1.06 (0.97,1.15); P = 2.3E-01	1.05 (0.98,1.12); P = 1.6E-02	0.90 (0.81,0.98); P = 9.9E-01	Same SNP
GPD1L/CMTM8	rs7620066	3 (32194428)	T/A	0.35	1.14 (1.06, 1.22); P = 8.7E-04	0.95 (0.89,1.01); P = 8.3E-02	1.02 (0.97,1.06); P = 5.3E-01	1.04 (0.99,1.09); P = 7.9E-02	rs7644491 (^r ² = 0.84)

^aPositions/alleles are relative to the forward strand of NCBI build36. ^bResults from GABRIEL are from a re-analysis using fixed-effects meta-analysis, excluding the B58C and ECRHS2 cohorts which are included in Stage2 or with occupational asthma (see Methods), and are for the APCAT SNP or the best available proxy. All p values are two-tailed. doi:10.1371/journal.pone.0044008.t003

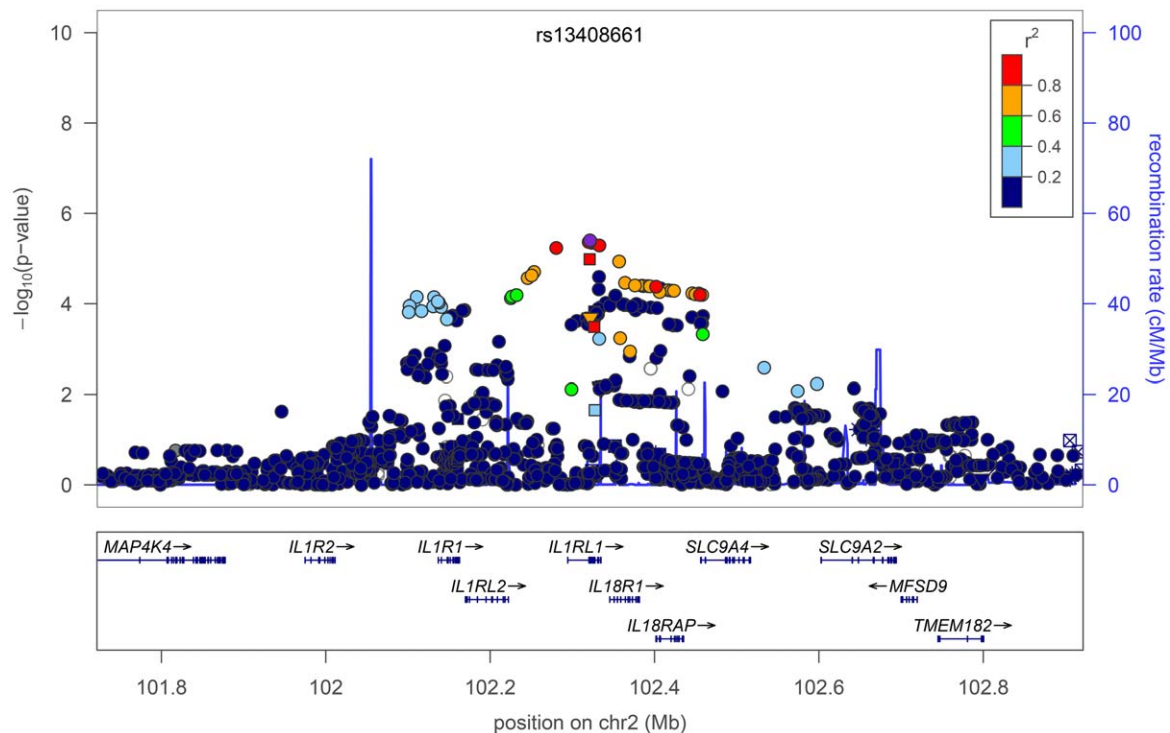


Figure 1. Regional association and forest plots for rs13408661 in the *IL1RL1/IL18R1* locus. For the regional plot, the lead SNP is indicated by a purple diamond, and the degree of linkage disequilibrium (r^2) of other SNPs in the region to the lead SNP is indicated by the color scale. Genes are shown below, and estimated recombination rate is indicated by the blue lines. Note that the regional plot is based on Stage1 (pooled) estimates only. For the forest plot, the estimated odds ratio and 95% confidence interval for each individual study is shown by the boxes (scaled to sample size) and lines; pooled estimates and 95% confidence intervals are indicated by diamonds.
doi:10.1371/journal.pone.0044008.g001

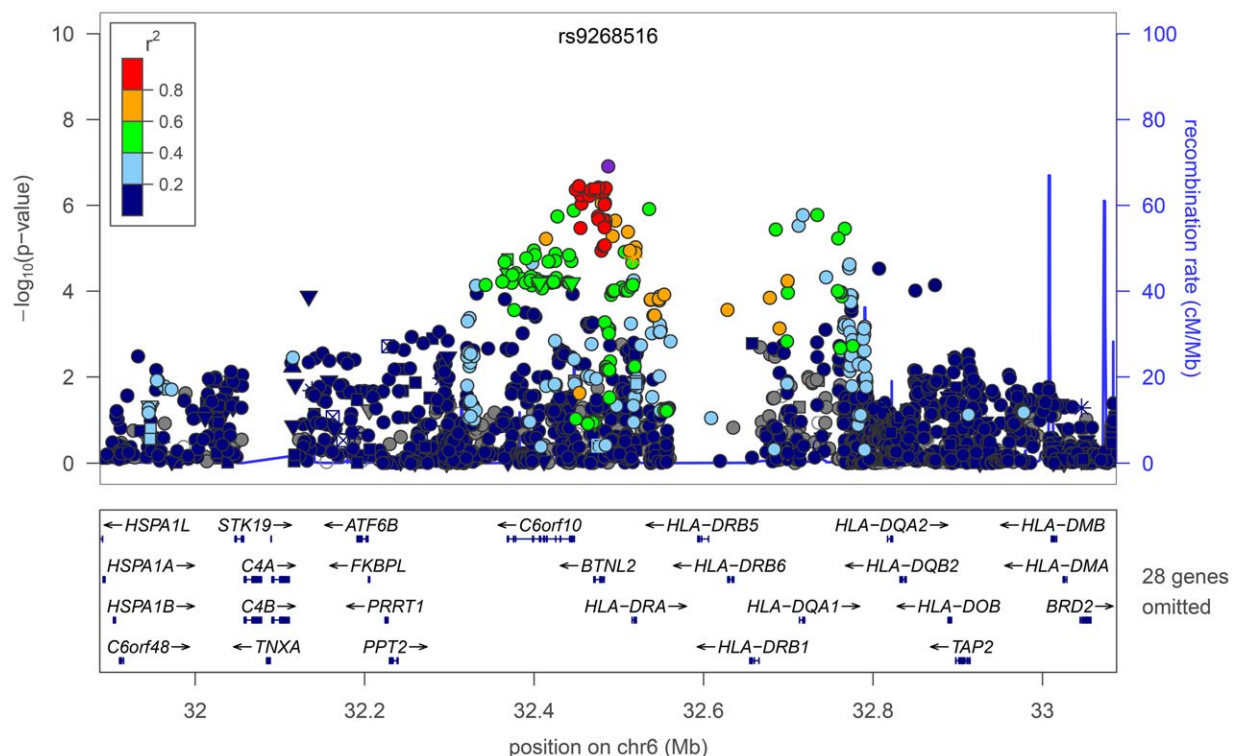


Figure 2. Regional association and forest plots for rs9268516 in the HLA region, with symbols as in Figure 1.
doi:10.1371/journal.pone.0044008.g002

$P = 8.6 \times 10^{-6}$; **Figure S3**]. Similarly, we scanned for interactions with allergic status in APCAT by comparing, for each SNP, the estimated association statistics for asthma in individuals with hay fever (the most commonly available measure of allergic risk factors in the APCAT studies) to the estimates in individuals without hay fever. The locus with the strongest evidence of interaction with allergic status did not reach genome-wide significance either [estimated pooled odds ratio for interaction between rs17136561 (located in SLC22A23 which overlaps with PSMG4 and TUBB2B) and hay fever status = 1.64; 95% confidence interval 1.33 to 2.02, $P = 2.3 \times 10^{-6}$; **Figure S3**]. We did not pursue either of these loci further.

We also investigated whether the associations of the previously known or suggestive loci, and the signals emerging from this paper (rs13408661 near the *IL1RL1/IL18R1* genes and rs9268516 in the HLA region), differed in association by smoking or allergic status. The direction of association for asthma in the never smokers and also in non-allergic individuals (*i.e.* the healthy subgroups) were generally consistent with the unstratified analysis, but with weaker signals as expected with reduced sample sizes (**Table S5**). A formal test for heterogeneity between smoking strata indicated no significant differences. Similarly, there was no significant heterogeneity for the allergic strata except possibly for rs2284033 in *IL2RB* ($P_{\text{heterogeneity}} = 0.038$), where opposite directions of association with asthma were observed in the two allergic strata, and for rs11071559 in *RORA* ($P_{\text{heterogeneity}} = 0.059$) where the signal for asthma association appears to be seen predominantly in the allergic individuals.

Materials and Methods

Participants and Studies

Cases and controls for the discovery study were drawn from six population-based studies of individuals of European ancestry: FINRISK [29], Framingham Heart Study [30], Health 2000 [31], Helsinki Birth Cohort [32], Northern Finland Birth Cohort of 1966 [33] and Young Finns Study [34]. All cohorts were genotyped using commercially available genotyping arrays and SNPs which passed QC filters were used to impute up to 2.5 million SNPs using HapMap CEU as the reference. Participants for the *in silico* replication were drawn from the 1958 British Birth Cohort (B58C) [7], the Australian Asthma Genetics Consortium (AAGC) [35], European Community Respiratory Health Survey followup (ECRHS) [36] and EPIC-Norfolk [37]. Study characteristics are given in **Table 1** and **Supplementary Methods**; genotyping and imputation details are given in **Table S1**. The most strongly associated SNPs from APCAT were checked for validity after re-analysis of data from the GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) study [17] (excluding B58C, ECRHS2 and cohorts with occupational asthma) made available at www.cng.fr/gabriel (Table 3).

Phenotype definition and stratification

Cases were defined as individuals who had given an affirmative questionnaire response to the question “Have you ever been diagnosed with asthma?” (exact wording varied among questionnaires – see **Supplementary Methods**). The remaining subjects served as healthy controls if they did not affirmatively respond to any of the following: self-reported asthma without a physician diagnosis, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, chronic cough associated with wheeze, other lung disease, or $FEV_1 < 70\%$ of predicted. Individuals with reports of chronic obstructive pulmonary disease, emphysema, chronic

bronchitis, or other lung diseases were also excluded from the cases.

We also conducted two stratified analyses of asthma: smoking-stratified and allergy-stratified. Allergic status was defined using an affirmative response to the question “Have you ever had hay fever or other allergic nasal symptoms?” (exact wording varied among questionnaires), as this was the most uniformly available information on allergy in APCAT. Participants were divided into three smoking categories: never smokers, ex-smokers if smoked regularly more than a year ago or current smokers if currently smoking or smoked regularly in past year.

Statistical analyses

The association statistic for each SNP oriented towards the forward strand and was calculated assuming an additive genetic model adjusting for sex, ancestry-informative principal components, and (in non-birth cohorts) for age. In the family-based Framingham Heart Study, the association analysis was done controlling for family structure using the GWA package in R [38]. The data for FINRISK, Health 2000, the Helsinki Birth Cohort, and the Young Finns Study were analyzed together with an adjustment term for cohort. The data from this combined dataset along with data from the Framingham Heart Study and the Northern Finland Birth Cohort of 1966 were meta-analyzed and verified using the fixed effect inverse-variance method implemented in METAL [39] and R. Genomic control was applied at the individual study level as well as in meta-analysis stage. Approximately 2.2 million SNPs with imputation quality (info) > 0.30 and minor allele frequency (MAF) $> 5\%$ were analyzed.

SNP selection for *in silico* replication

We selected a locus (defined as a region 500 kb wide) for further follow-up if it either contained a single SNP with $P < 10^{-6}$ or multiple SNPs with $P < 10^{-5}$. The “sentinel SNP” was defined as the SNP with the most significant P value and was included in the replication list.

Discussion

We completed a genome-wide association study of a simple asthma phenotype – self-report of ever having been diagnosed by a physician with asthma – from 18,604 participants from six population-based studies comprising the APCAT consortium. We performed follow-up analyses of the top signals from APCAT in 15,286 additional individuals. These results provided strong evidence for an additional associated variant in the *HLA* region, a known asthma locus, and confirmed recent reports of multiple associations at the *IL1RL1/IL18R1* locus. We also examined the evidence for association in APCAT of SNPs with previous genome-wide or suggestive evidence of asthma, and show that the results from our population-based studies validate and in one case newly establish genome-wide significant associations with asthma. Finally, we found no evidence of genes modifying the relation between smoking and asthma or the relationship between hay fever and asthma.

The present study has several strengths. First, it demonstrates the usefulness of a large untapped resource to complement genetic studies of asthma: population-based studies with genome-wide genotype data and a simple asthma phenotype: self-reported information on doctor-diagnosed asthma. We also present the first comprehensive search for genetic interactions with smoking status and with hay fever, two important modifiers of asthma development. Finally, we provide evidence for new genome-wide

significant associations with asthma: one novel signal where there was prior suggestive evidence of association (*RORA*), one independent novel signal at a previously associated locus (the *HLA* region), and one previously associated locus where we demonstrate multiple independent signals (*IL1RL1/IL18R1*).

It is important to recognize some limitations of this present study. First, the constituent studies in APCAT studies have limited information on asthma, such as age of asthma onset or severity of symptoms, which prevents a more detailed investigation of associated loci. Second, although controls were carefully selected to exclude individuals with other respiratory diseases or abnormalities on spirometry (see methods) that may share pleiotropic risk alleles with asthma, the choice of controls in population-based cohorts is potentially subject to misclassification bias due to the inclusion of undiagnosed cases among the controls. We note that this problem is common to many study designs for diseases with variable age at onset such as asthma, but does not prevent the identification of new disease markers. Third, the power to detect novel associations in population based studies is restricted by the low prevalence of disease (our prevalence was ~9%) compared to a case-control study of an equivalent total sample size and an equal number of cases to controls. However, we note that the large number of additional available population-based studies similar to the ones in APCAT still represent a large untapped pool of genotyped cases. Fourth, we only examined SNPs with frequencies above 5%, so have not tested rarer variants for association to asthma. Finally, the use of self-reported diagnosis of asthma in the APCAT cohorts, even though it was doctor-diagnosed, may also have led to misclassification within cases.

Despite these limitations, we were able to independently validate 10 of 13 SNPs previously reported as being associated for asthma through GWA studies, which indicates that population-based studies with simple asthma phenotypes can indeed complement ongoing genetic studies in asthma. For most (but not all) of these variants, the estimated effect sizes in APCAT are smaller than reported, which could be due to the “winner’s curse” phenomenon [40], slightly greater misclassification in our cohorts, or other differences between this study and previous reports. However, the fact that most of the APCAT odds ratios fall within the 95% confidence intervals of the original reports and the *p*-values typically becomes more significant when combined with the APCAT data suggests that the power gained by adding population-based studies still outweigh the effects of any misclassification there may be. When the reported estimates are combined with our data, the variant rs11071559 in *RORA* reaches genomewide significance. In the subsequent analysis, the association in *RORA* was perhaps more strongly associated with asthma in individuals with hay fever. This gene belongs to a subfamily of nuclear orphan receptors suggested to negatively regulate inflammatory response [41]. Interestingly, *RORA* deficient mice have diminished capacity to mediate allergic inflammatory response [41,42,43]. A recent paper found that *RORA* is critical for the development of nuocytes, which are part of the innate immune response and contribute to asthma response, in mice [44]. Another recent study that looked at asthma candidate genes found that *RORA* is differentially expressed during lung development in both mouse and human [45].

By analyzing the results of meta-analysis of APCAT studies, we identified and successfully replicated associations of rs13408661 in the *IL1RL1/IL18R1* region and of rs9268516 in the *HLA* region; the association with rs13408661 is in agreement with recent findings [8,18]. These two loci had originally been reported to contain other strongly associated variants [16,17], which we now show are distinct from the ones we identify. Observing multiple

signals at a locus could either be due to multiple true causal variants, as seen in genetic studies for height and other polygenic traits [22,46], or due to a single causal variant that is not well tagged by either signal. While fine mapping of data in these regions would be required to resolve this problem conclusively, we note that the variants we identified were only modestly correlated with published variants and remain significant even after conditioning on known risk variants. Thus, our data provide two genome-wide significant signals of association at known loci that are distinct from the first variants originally reported to be associated with asthma at genome-wide significance, and in one case (*HLA*) is not accounted for by the known associations.

Finally, we present the first comprehensive search for genetic interactions with smoking status and with hay fever, two potentially important modifiers of asthma development. In our studies, we found no convincing evidence for SNPs that interacted with either smoking or hay fever. A larger meta-analysis would be required to confirm the absence or existence of such gene-environment effects for asthma.

In conclusion, these results strongly suggest that GWA studies of population-based cohorts with simple asthma phenotypes are an effective approach to find novel asthma-associated variants, replicate signals identified in other studies, and provide estimates that are representative of the demography and disease spectrum. Such cohorts are an untapped resource that can be utilized to complement genetic studies of asthma. We anticipate that many more population-based studies could be leveraged to assist with discovery of asthma susceptibility loci, which could potentially lead to more effective or targeted therapies and preventions.

Supporting Information

Figure S1 Correlation between unstratified analysis and smoking or allergic-status stratified analyses. (TIF)

Figure S2 Quantile-quantile plot and Manhattan plot for meta-analyses of asthma (basic unstratified analysis) in APCAT. The shaded area in the quantile-quantile plot shows the 95% confidence intervals. (TIF)

Figure S3 Manhattan plot for interaction effect for smoking exposure and for allergic status and a graphical depiction of the odds ratios for the best signals. (TIF)

Supplementary Methods S1 Supplementary Methods and Materials. (DOCX)

Table S1 Genotyping platform, calling algorithm, imputation details and software used in Stage 1 and Stage 2 studies. (DOCX)

Table S2 Loci with previous genome-wide significant associations to asthma ($P < 5 \times 10^{-8}$) in Japanese populations. (DOCX)

Table S3 Conditional analysis on novel signals in APCAT. (DOCX)

Table S4 Variance explained by the GWAS hits under a liability threshold model. (DOCX)

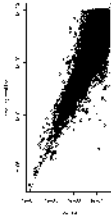
Table S5 Stratified association analysis of previously published SNPs and two novel loci emerging from this paper.
(DOCX)

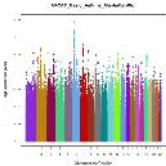
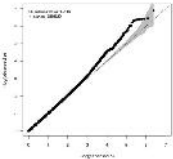
Author Contributions

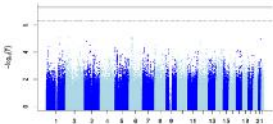
Conceived and designed the experiments: AR MK SV ZKG HNL ACA GTO VS M-RJ JNH. Performed the experiments: DPS JHZ MJA MAB

References

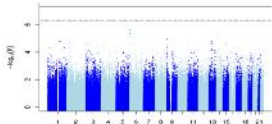
- Masoli M, Fabian D, Holt S, Beasley R (2004) The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* 59: 469–478.
- Bousquet J, Bousquet PJ, Godard P, Daures JP (2005) The public health implications of asthma. *Bull World Health Organ* 83: 548–554.
- Barnett SB, Nurmagambetov TA (2011) Costs of asthma in the United States: 2002–2007. *J Allergy Clin Immunol* 127: 145–152.
- Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD (1990) Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 142: 1351–1358.
- Nieminen MM, Kaprio J, Koskenvuo M (1991) A population-based study of bronchial asthma in adult twin pairs. *Chest* 100: 70–75.
- van Beijsterveldt CE, Boomsma DI (2007) Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* 29: 516–521.
- Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating ORM DL3 expression contribute to the risk of childhood asthma. *Nature* 448: 470–473.
- Ferreira MA, McRae AF, Medland SE, Nyholt DR, Gordon SD, et al. (2011) Association between ORM DL3, IL1RL1 and a deletion on chromosome 17q21 with asthma risk in Australia. *Eur J Hum Genet* 19: 458–464.
- Rogers AJ, Raby BA, Lasky-Su JA, Murphy A, Lazarus R, et al. (2009) Assessing the reproducibility of asthma candidate gene associations, using genome-wide data. *Am J Respir Crit Care Med* 179: 1084–1090.
- Galanter J, Choudhry S, Eng C, Nazario S, Rodriguez-Santana JR, et al. (2008) ORM DL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 177: 1194–1200.
- Barnes KC (2010) Genomewide association studies in allergy and the influence of ethnicity. *Curr Opin Allergy Clin Immunol* 10: 427–433.
- Himes BE, Hunninghake GM, Baurley JW, Rafals NM, Sleiman P, et al. (2009) Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* 84: 581–593.
- Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, et al. (2010) Variants of DENND1B associated with asthma in children. *N Engl J Med* 362: 36–44.
- Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, et al. (2010) Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 125: 328–335 e311.
- Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, et al. (2008) Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. *PLoS Genet* 4: e1000166.
- Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, et al. (2009) Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 41: 342–347.
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363: 1211–1221.
- Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, et al. (2011) Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet*.
- Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, et al. (2011) Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet*.
- Wan YI, Shrine NR, Soler Artigas M, Wain LV, Blakey JD, et al. (2012) Genome-wide association study to identify genetic determinants of severe asthma. *Thorax*.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9: 356–369.
- Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467: 832–838.
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948.
- Meyers DA (2010) Genetics of asthma and allergy: what have we learned? *J Allergy Clin Immunol* 126: 439–446; quiz 447–438.
- Wenzel SE (2006) Asthma: defining of the persistent adult phenotypes. *Lancet* 368: 804–813.
- Borish L, Culp JA (2008) Asthma: a syndrome composed of heterogeneous diseases. *Ann Allergy Asthma Immunol* 101: 1–8; quiz 8–11, 50.
- Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, et al. (2011) Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* 7: e1002170.
- So HC, Gui AH, Cherny SS, Sham PC (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* 35: 310–317.
- Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, et al. (2010) Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol* 39: 504–518.
- Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, et al. (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 165: 1328–1335.
- Aromaa A, Koskinen S, editors (2004) Health and functional capacity in Finland. Baseline Results of the Health 2000 Health Examination Survey. Helsinki: National Public Health Institute KTL.
- Eriksson JG, Forsen T, Tuomilehto J, Osmond C, Barker DJ (2001) Early growth and coronary heart disease in later life: longitudinal study. *BMJ* 322: 949–953.
- Rantakallio P (1969) Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr Scand* 193: Suppl 193:191+.
- Raitakari OT, Juonala M, Ronnema T, Keltikangas-Jarvinen L, Rasanen L, et al. (2008) Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 37: 1220–1226.
- Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, et al. (2011) Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* in press.
- (1997) Genes for asthma? An analysis of the European Community Respiratory Health Survey. *Am J Respir Crit Care Med* 156: 1773–1780.
- Myint PK, Luben RN, Surtees PG, Wainwright NW, Welch AA, et al. (2005) Respiratory function and self-reported functional health: EPIC-Norfolk population study. *Eur Respir J* 26: 494–502.
- Chen MH, Yang Q (2010) GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 26: 580–581.
- Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191.
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33: 177–182.
- Delerive P, Monte D, Dubois G, Trottein F, Fruchart-Najib J, et al. (2001) The orphan nuclear receptor ROR alpha is a negative regulator of the inflammatory response. *EMBO Rep* 2: 42–48.
- Jaradat M, Stapleton C, Tilley SL, Dixon D, Erikson CJ, et al. (2006) Modulatory role for retinoid-related orphan receptor alpha in allergen-induced lung inflammation. *Am J Respir Crit Care Med* 174: 1299–1309.
- Jetten AM (2009) Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nucl Recept Signal* 7: e003.
- Wong SH, Walker JA, Jolin HE, Drynan LF, Hams E, et al. (2012) Transcription factor RORalpha is critical for nuocyte development. *Nat Immunol* 13: 229–236.
- Melen E, Kho AT, Sharma S, Gaedigk R, Leeder JS, et al. (2011) Expression analysis of asthma candidate genes during human and murine lung development. *Respir Res* 12: 86.
- Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, et al. (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39: 638–644.



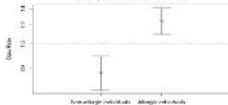




Stage 1 pooled odds ratio (and 95% CI) for asthma by smoking category per T allele of rs1007028



Stage 1 pooled odds ratio (and 95% CI) for asthma by allele per T allele of rs1710455



SUPPLEMENTARY METHODS

Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association in the HLA region

Stage 1 – Brief study description, funding acknowledgement and phenotype definition

FINRISK STUDY

Description: A population survey of risk factors for chronic diseases in Finland. The survey has been executed every five years from 1972 using independent, random and representative population samples from five geographical areas of the country. Participants have filled in a health-related questionnaire and undergone a physical examination including measurement of anthropometric traits and blood draw (E1).

Funding: Mainly funded by the Finnish National Institute for Health and Welfare with additional financial support from the Academy of Finland (grant number 129494, 139635), the Finnish Foundation for Cardiovascular Research, and the Sigrid Juselius Foundation. Participants were genotyped as part of the COROGENE (Genetic Predisposition of Coronary Heart Disease in Patients Verified with Coronary Angiogram) (E2) and MGen (Myocardial Infarction Genetics Consortium) (E3) efforts.

Asthma status: For the FINRISK 1992 cohort, individuals who have answered “YES” to the question “*Have you had any of the following diseases diagnosed or treated by a doctor during the past year (last 12 months)?*: Asthma of the lungs”. were considered to have asthma. For the FINRISK 1997, 2002 and 2007 cohorts, information about asthma was based on the following

question: “*Have you ever been diagnosed with asthma?*” Participants responding “YES” were considered to have asthma. In all FINRISK surveys the rest of the participants served as controls if their age was less than or equal to 70 years and they did not report pulmonary emphysema, or chronic bronchitis during the last 12 months.

Smoking status: Participants were considered as never smokers if they answered “NO” to the question “*Have you ever smoked?*” and as ex smokers if they had quit smoking at least a month ago. Current smokers were participants who smoked regularly at the time of the survey or had been smoking regularly less than 1 month ago.

Allergy status: Information about allergy was based on the question “*Have you ever had hay fever or other allergic nasal symptoms?*” Participants responding “YES, during the last 12 months”, and those responding “YES, over a year ago” were considered to have allergy.

Framingham Heart Study (FHS)

Description: Began in 1948 by recruiting an Original Cohort of 5,209 men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, who had not yet developed overt symptoms of cardiovascular disease or suffered a heart attack or stroke (E4). Since that time the study has added an Offspring Cohort of 5,124 individuals in 1971 and a Third Generation Cohort of 4,095 individuals in 2002 (E5, 6). The Offspring Cohort is made up of the offspring of the members of the Original Cohort and their spouses, and the Third Generation Cohort is made up of the children of the Offspring cohort. Individuals in the FHS had exams every two years consisting of a detailed medical history, physical examination, and laboratory tests. We used data from the exam questions described below for the purposes of determining asthma cases and controls and to determine smoking and allergy statuses.

Funding: Supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center.

Asthma status: For the Original Cohort, data from Exams 1 and 5 were used. Individuals who answered “yes” to any asthma question were coded as asthma cases (“yes” to “bronchial asthma, alone” or “allergy and asthma, together” at Exam 1 or “asthma, onset before age 16,” “asthma, onset age 16 or after,” or “asthma, age of onset unknown” at Exam 5). Individuals with a response of “yes” to the question “lifetime history of chronic pulmonary disease” at Exam 2 were excluded from both cases and controls. All remaining individuals were coded as asthma controls. For the Offspring Cohort, data from Exams 2 and 7 were used. Individuals who answered “yes” to any asthma question were coded as asthma cases (“yes” to “asthma” at Exam 2 or “yes, new” or “yes, old” to “have you had asthma in the interim?” at Exam 7). Individuals with a response of “yes” to “chronic obstructive lung disease” or “chronic bronchitis” at Exam 2 were excluded from both cases and controls. Individuals with a response of “yes, now” or “yes, not now” to “non-cardiovascular medications: bronchodilators and aerosols” were excluded from the controls. All remaining individuals were coded as asthma controls. For the Third Generation Cohort, individuals who answered “yes” to “asthma: was it diagnosed by a doctor or other health professional?” were coded as asthma cases. Individuals who answered “no” to both “asthma: was it diagnosed by a doctor or other health professional?” and “have you ever had asthma?” were coded as asthma controls. Individuals who answered “yes” to any of the following questions were excluded from both the cases and controls: “have you ever had COPD?,” “have you ever

69 had pulmonary fibrosis?,” “have you ever had chronic bronchitis?,” “have you ever had
70 emphysema?.”

71 *Smoking status:* For the Original Cohort, data from Exam 5 were used. Current smokers
72 answered “smokes now” to the question “history of smoking, present habit.” Former smokers
73 answered “does not smoke now, but smoked formerly,” and never smokers answered “never
74 smoked.” For the Offspring Cohort, data from Exams 1, 2, and 7 were used. Current smokers
75 answered “yes” to “smoked cigarettes regularly in last year” (Exam 7) or “smoking now” (Exam
76 2). Former smokers answered “no” to “smoked cigarettes regularly in last year” (Exam 7) and
77 “yes” to “ever smoked regularly” (Exam 2) or “smoked at least 1 year” (Exam 1). Never smokers
78 answered “no” to “ever smoked” or “ever smoked regularly” (Exam 2). For the Third Generation
79 Cohort, current smokers answered “yes” to “if ever smoked cigarettes regularly: do you now
80 smoke?.” Former smokers answered “yes” to “have you ever smoked cigarettes regularly?” and
81 “no” to “if ever smoked cigarettes regularly: do you now smoke?.” Never smokers answered
82 “no” to “have you ever smoked cigarettes regularly?.”

83 *Allergy status:* For the Original Cohort, data from Exam 1 were used. Individuals who answered
84 “allergy, alone” or “allergy and asthma, together” to “history of allergy or asthma” were
85 classified as allergy cases. For the Offspring Cohort, no allergy data were available. For the
86 Third Generation Cohort, individuals who answered “yes” to “have you ever had hay fever
87 (allergy involving the nose and/or eyes)?” were classified as allergy cases.

88

89

90 **Health 2000 Study**

Description: The study was conducted in 2000 (E7) and included home interview, several questionnaires, laboratory and anthropometrical measurements, spirometry with bronchodilator test and clinical examination of a physician. The data were completed by record linkage with the National Hospital Discharge Register and the National Social Insurance Institutions register data on reimbursement of asthma medication.

Funding: The study was mainly funded by the Finnish National Institute for Health and Welfare. H2000 participants were genotyped as part of the HDL extremes study (E8) and the Sanger Institute/Wellcome Trust efforts for GENMETs (Genetic Background and Molecular Pathogenesis of Metabolic Syndrome and Its component Risk Factor Traits).

Asthma status: Information about asthma was based on the following question: “*Has a doctor ever diagnosed you with one of the following illnesses?*” One of the listed illnesses was asthma and those who responded “YES” were considered to have asthma. The rest of the participants were taken as controls if their age was less than or equal to 70 years and they never had had chronic bronchitis.

Smoking status: Information about smoking was taken from a question with five alternatives: (1) I smoke daily; (2) I smoke occasionally; (3) I have quit smoking 1-12 months ago; (4) I have quit smoking more than a year ago; (5) I have never smoked. Participants choosing alternatives 1 or 2 were considered as current smokers; those choosing 3 or 4 were considered as ex-smokers; and those choosing 5 were considered as never smokers.

Allergy status: Information about allergy was based on responses under the subheading: “*Other diseases diagnosed by a doctor*”. One of the alternatives given was “*allergy, rhinitis; for*

example hay fever or other allergic rhinitis” Those responding “YES” to this item were considered as having allergy.

Helsinki Birth Cohort Study (HBCS)

Description: The study includes 8760 subjects born in Helsinki between 1934 and 1944. Between 2001 and 2004, a representative subset of 928 males and 1075 females participated in a clinical study focusing upon cardiovascular and metabolic outcomes and cognitive functions (E9). Information on asthma, smoking and alcohol intake is available from questionnaires for 2003 individuals who participated in the clinical study. Information on hospitalization due to alcohol abuse is available from the National Hospital Discharge Register. Psychological questionnaires have been used to assess personality characteristics including data on impulsivity. GWAS has been done on these individuals participating in the clinical study (n~2000) at the mean age of ~62 years.

Funding: Financial support was received from the Academy of Finland, Samfundet Folkhälsan, Finnish Diabetes Research Foundation, Finska Läkaresällskapet, Finnish Foundation for Cardiovascular Research; Yrjö Jahnsson Foundation, Foundation Liv och Hälsa and Academy of Finland (grants number 129287 and 134839).

Asthma status: Information about asthma was based on the following question: “*Have you ever had any of the following illnesses diagnosed or treated by a doctor?*” One of the listed illnesses was “*Asthma of the lungs*” and those responding “YES” to this item were considered as having asthma. Those participants who responded to the same question as having emphysema or chronic

bronchitis were excluded and the others were taken as controls. All participants were less than 70 years of age.

Smoking status: Information about smoking was taken from a question with three alternatives: (1) I smoke daily; (2) I smoke occasionally; (3) I have never smoked. Participants choosing alternatives 1 were considered as current smokers; those choosing 2 were considered as ex-smokers; and those choosing 3 were considered as never smokers.

Allergy status: Information about allergy was based on the following question: “Do you have following allergic symptoms? Nasal allergy, hay fever, stuffy nose” Participants responding “YES” were considered as having allergy.

Northern Finland Birth Cohort of 1966 (NFBC1966)

Description: This is a prospective follow-up study of 12,058 live births from the two northernmost provinces of Finland, Oulu and Lapland, and cover 96% of the children born in that region between January 1 and December 31, 1966. In 1997, at age 30 years, 8463 survivors still living in Northern Finland or in the capital area received a postal questionnaire and invitation to clinical examinations including DNA sample (71% participated) (E10, 11). Please see <http://kelo.oulu.fi/NFBC/> for more details about the study. Informed consent for the use of the data including DNA was obtained from all subjects. The study was approved by the ethics committees in Oulu (Finland) and Oxford (UK) universities in accordance with the Declaration of Helsinki.

Funding: Financial support from the Academy of Finland (project grants 104781, 1114194, 120315 and Center of Excellence in Complex Disease Genetics), Oulu University Hospital,

Biocenter Oulu, University of Oulu, Finland, the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council (studentship grant G0500539). We thank Professor Paula Rantakallio (launch of NFBC1966 and 1986), Ms Outi Tornwall and Ms Minttu Jussila (DNA biobanking). DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Institute for Health and Welfare, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki.

Asthma status: We defined asthmatic cases as those who answered yes to BOTH conditions which are parts of Question 31 of the main questionnaire administered in 1997 when participants were aged 31 years: 1) “*have you ever asthma during the last 12 months or more than a year ago?*” and 2) “*has this been verified or treated by doctor?*” The remaining individuals served as healthy controls if they did not have any of the following conditions: self-reported asthma but not verified or treated by doctor diagnosis (n=196); individuals with FEV1 less than 70% of predicted for sex and height (n=19); individuals who reported they ever had emphysema, chronic bronchitis, pulmonary bronchial (n=211) or doctor-diagnosed chronic cough (n=402).

Smoking status: First we defined the never/ever smoking status by Question 79 “*Have you ever smoked in your life?*” of 31-year questionnaire. Among the Ever smokers, we distinguish the Current Smokers if they answered said that they smoke on at least one day per week to Question 82 “*Do you smoke nowadays?*”. We excluded 43 ever smokers whom we could not distinguish if they were current or ex-smokers.

Allergy status: We defined an individual as allergic if they answered yes for either having hay fever or allergic rhinitis during the last 12 months or more than a year ago in Question 31 of main questionnaire.

Young Finns Study (YFS)

Description: This is a longitudinal population study sample on the evolution of cardiovascular risk factors from childhood to adulthood (E12). The first cross-sectional survey was conducted in 1980 in five Finnish university cities and included 3,596 participants who were in the age groups of 3, 6, 9, 12, 15, and 18 years and were randomly chosen from the national population register; equal ratios of males and females were selected in each age group. In 2007, 2204 subjects now aged 30 to 45 years participated in the latest follow-up study.

Funding: Financially supported by the Academy of Finland (grant numbers 126925, 121584, and 124282); the Social Insurance Institution of Finland; Kuopio, Tampere, and Turku University Hospital Medical Funds; the Juho Vainio Foundation; the Paavo Nurmi Foundation; the Tampere Tuberculosis Foundation; the Finnish Foundation of Cardiovascular Research; and the Finnish Cultural Foundation.

Asthma status: Information about asthma was based on a following question: “*Do you have at the moment or have you had a long-term illness, handicap or injury diagnosed by a doctor?*” Those responding “YES” and specifying among the given alternatives “Asthma of the lungs” were considered as having asthma. The rest of the participants were taken as controls, except those who reported having chronic bronchitis. All YFS participants were less than 70 years of age.

199 *Smoking status:* Information about smoking was based on a question with 5 response alternatives
200 as follows: (1) I smoke once a day or more often than once a day; (2) I smoke once a week or
201 more often than once a week but not daily; (3) I smoke less often than once a week; (4) I have
202 stopped smoking; (5) I have never smoked. Participants choosing any of the alternatives 1-3
203 were taken as current smokers; those choosing the alternative 4 were taken as ex-smokers and
204 those choosing the alternative 5 were taken as never smokers.

205 *Allergy status:* Information about allergy was based on the same question as above, i.e., “Do
206 you have at the moment or have you had a long-term illness, handicap or injury diagnosed by a
207 doctor?” Those responding “YES” and specifying among the given alternatives “Allergic rhinitis,
208 for example hay fever” were taken as having allergy.

209

Stage 2 – Brief study description, funding acknowledgement and phenotype definition

1958 British Birth Cohort (B58C)

Description: This is a nationwide British birth cohort including participants born in a particular week of 1958. Details of the B58C biomedical follow-up at age 44.5 have been previously reported and a full technical report is available online

(<http://www.b58cgene.sgul.ac.uk/report.php>).

Funding: The phenotype and genotype data are from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z (E13). The B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium (E14), a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic Award (079895). Genotyping for the B58C-GABRIEL and for ECRHS was supported by a contract from the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1

Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control and grants from the French Ministry of Research (E15).

Asthma status: Asthma is defined as self-reported ever at either age 33 or age 42 via administered interviews. Healthy controls exclude any history of asthma, wheezing or wheezing bronchitis as reported in the interviews at age 7, 11, 16 and 23, plus exclusions for bronchitis at age 42 and FEV1 less than 70% predicted for sex and height at age 42.

Australian asthma genetics consortium (AAGC)

We carried out a GWAS in 2,110 physician-diagnosed asthmatics and 3,857 controls of European ancestry from Australia. Participants were drawn from two cohorts that are described in detail elsewhere (Ferreira et al., in press): the Australian Asthma Genetics Consortium (AAGC) cohort (n=1,810) and the Queensland Institute of Medical Research (QIMR) GWAS cohort (n=4,157). Amongst the 2,110 asthmatic cases, 759 (36%) were diagnosed through clinical examination and 1,351 (64%) reported a lifetime doctor diagnosis of asthma in epidemiological questionnaires. With respect to disease onset, 1,269 (60%) subjects were classified as having childhood asthma (defined by an age-of-onset at or before age 16), 515 (24%) subjects with later onset asthma (age-of-onset after the age of 16) and 326 (16%) with unknown age-of-onset. Fifty-eight percent of asthmatics were atopic, as defined by a positive skin prick test (SPT) response to at least one common allergen; 68% had at least one first-degree relative with asthma; and 36% reported lifetime smoking.

The 3,857 controls included 2,030 (52.6%) individuals who were classified as asthma-free based on clinical examination (3.0%) or epidemiological questionnaires (49.6%). The remaining 1,827

(47.4%) individuals provided no information about their asthma status and were included in the analysis as controls to improve power. SPT information and lifetime smoking status was unavailable for most controls. Overall, the mean age of participants was 34 years (sd=16.2, range 2 to 92) and 45.2% were males. This dataset does not include 1,230 samples from the Busselton cohort analysed in the GABRIEL study.

Australian Asthma Genetics Consortium collaborators: Svetlana Baltic¹, Mark Jenkins², Dale R. Nyholt³, Catherine Hayden⁴, Patrick Danoy⁵, Faang Cheah¹, John L. Hopper², Stephen R. Leeder⁶, Haydn Walters⁷, Graham Jones⁸, Désirée Mészáros⁹, Mary Roberts¹⁰, Melissa C. Southey¹¹, Euan R. Tovey¹², Loren Price¹, Margaret J. Wright³, Scott D. Gordon³, Li P. Chung¹, Anjali K. Henders³, Graham Giles¹³, Paul S. Thomas¹⁴, Suzanna Temple¹, John B. Whitfield³, Ian Feather¹⁵, Stephen Morrison¹⁶, Chalermchai Mitrpant¹, Warwick J. Britton¹⁷, David John², Andrew S. Kemp¹⁸

¹ Lung Institute of WA and Centre for Asthma, Allergy and Respiratory Research, University of WA, Perth, Australia.

² Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Australia.

³ The Queensland Institute of Medical Research, Brisbane, Australia.

⁴ School of Paediatrics and Child Health, Princess Margaret Hospital for Children, Perth, Australia.

⁵ University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Australia.

⁶ Australian Health Policy Institute, University of Sydney, Sydney, Australia

⁷ Menzies Research Institute, Hobart, Australia.

⁸ University of Western Sydney, Penrith, Australia

⁹ Menzies Research Institute, Hobart, Australia.

¹⁰ Department of Respiratory Medicine, Royal Children's Hospital, Parkville, Australia.

¹¹ Department of Pathology, The University of Melbourne, Melbourne, Australia.

¹² Woolcock Institute of Medical Research, University of Sydney, Sydney, Australia.

¹³ Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Australia.

¹⁴ Faculty of Medicine, University of New South Wales, Sydney, Australia.

¹⁵ Gold Coast Hospital, Southport, Australia.

¹⁶ University of Queensland, Brisbane, Australia.

¹⁷ Centenary Institute and University of Sydney, Camperdown, Australia.

¹⁸ The Children's Hospital, Westmead, Sydney, Australia.

European Community Respiratory Health Survey (ECRHS)

Description: This is a multicentre, mainly European, population based survey carried out in 1990s (E16) with a follow-up starting in 1998 (E17). Questionnaire information is based on information collected through interviewer administered questionnaire (forms available at <http://www.ecrhs.org/>). Participants were aged 20 – 48 (mean of 34) at the first survey and 28 – 56 (mean age 43) at the second survey. Samples were genotyped as part of the GABRIEL consortium (E15) and were collected at the follow-up and were mainly European participants and included an enrichment of asthma cases.

Funding: The co-ordination of follow-up of 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) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02), Hospital Universitario de Albacete, Consejería de Sanidad; **Barcelona:** SEPAR, Public Health Service (grant code: R01 HL62633-01), Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02) CIRIT (grant code: 1999SGR 00241) Red Respira ISCII; CIBER Epidemiologia y Salud Pública (CIBERESP), Spain **Basel:** Swiss National Science Foundation, Swiss Federal Office for Education & Science, Swiss National

306 Accident Insurance Fund (SUVA), USC NIEHS Center grant 5P30 ES07048; **Bergen:**
307 Norwegian Research Council, Norwegian Asthma & Allergy Association (NAAF), Glaxo
308 Wellcome AS, Norway Research Fund; **Erfurt:** GSF-National Research Centre for Environment
309 & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code FR 1526/1-1); **Galdakao:**
310 Basque Health Dept; **Grenoble:** Programme Hospitalier de Recherche Clinique-DRC de
311 Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de
312 Grenoble, Ministere de l'Emploi et de la Solidarite, Direction Generale de la Sante, Comite des
313 Maladies Respiratoires de l'Isere; **Hamburg:** GSF-National Research Centre for Environment
314 & Health, Deutsche Forschungsgemeinschaft (DFG) (grant code MA 711/4-1); **Ipswich and**
315 **Norwich:** Asthma UK (formerly known as National Asthma Campaign); **Huelva:** Fondo de
316 Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02); **Oviedo:**
317 Fondo de Investigaciones Santarias (FIS) (grant code: 97/0035-01, 99/0034-01 and 99/0034-02) ;
318 **Paris:** Ministere de l'Emploi et de la Solidarite, Direction Generale de la Sante, UCB-Pharma
319 (France), Aventis (France), Glaxo France, Programme Hospitalier de Recherche Clinique-DRC
320 de Grenoble 2000 no. 2610, Ministry of Health, Direction de la Recherche Clinique, CHU de
321 Grenoble; **Tartu:** Estonian Science Foundation; **Umeå:** Swedish Heart Lung Foundation,
322 Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy
323 Foundation, Swedish Cancer & Allergy Foundation; **Uppsala:** Swedish Heart Lung Foundation,
324 Swedish Foundation for Health Care Sciences & Allergy Research, Swedish Asthma & Allergy
325 Foundation, Swedish Cancer & Allergy Foundation; *Financial support for ECRHS I for centres*
326 *in ECRHS II was provided by:* Ministère de la Santé, Glaxo France, Institut Pneumologique
327 d'Aquitaine, Contrat de Plan Etat-Région Languedoc-Rousillon, CNMATS, CNMRT (90MR/10,
328 91AF/6), Ministre délégué de la santé, RNSP, France; GSF, and the Bundesminister für

Forschung und Technologie, Bonn, Germany; Norwegian Research Council project no. 101422/310; Ministerio Sanidad y Consumo FIS (grants #91/0016060/00E-05E and #93/0393), and grants from Hospital General de Albacete, Hospital General Juan Ramón Jiménez, Consejería 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.

Asthma status: We define asthmatic cases if they responded positively to BOTH of the following questions: 1) ever had asthma at first survey (Question 13) or second survey (Question 14) and 2) if it was verified by doctor diagnosis in second survey (Question 14 part 1). The remaining individuals served as healthy controls if they did not have any of the following conditions: self-reported asthma without doctor diagnosis (n=36); reported ever having wheeze in second survey (n=285); FEV1 less than 70% predicted for age, sex and height (n=10).

European Prospective Investigation of Cancer, Norfolk (EPIC-Norfolk)

Description: EPIC-Norfolk is part of the large multi-centre Europe-wide EPIC programme looking at the connection between diet and cancer (E18). There are over 30,000 participants aged 45 – 70 at recruitment who lived in Norwich and the surrounding towns and rural areas. They have been contributing information about their diet, lifestyle and health through questionnaires, and through health checks carried out by EPIC nurses. In 2006, a GWAS for obesity was carried out on 3,867 individuals (E19). A case-cohort design was used in which the subcohort (N=2,566) was a random sample of the cohort at baseline and cases were part of the remaining individuals

with a value of BMI being 30 kg/m² or greater (N=1,301). Unlike controls in the commonly used case-control design, the subcohort was an unselected population sample allowing for a variety of traits to be investigated.

Funding: Supported by research programme grant funding from the Cancer Research Campaign and the Medical Research Council, with additional support from Stroke Association, British Heart Foundation, Department of Health, Europe Against Cancer Programme Commission of the European Union, the Food Standards Agency and the Wellcome Trust.

Asthma status: Asthmatics were defined as a positive response to the question "*Has the doctor ever told you that you have asthma?*" which was asked at baseline survey. The remaining individuals served as healthy controls unless they had bronchitis or FEV1 < 70% predicted for age, height and sex.

REFERENCES

- E1. Vartiainen E, Laatikainen T, Peltonen M, Juolevi A, Mannisto S, Sundvall J, Jousilahti P, Salomaa V, Valsta L, Puska P. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol*. 2010;39(2):504-18.
- E2. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki ML, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*. 2010;376(9750):1393-400.
- E3. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Merlini PA, Berzuini C, Bernardinelli L, Peyvandi F, Tubaro M, Celli P, Ferrario M, Fetiveau R, Marziliano N, Casari G, Galli M, Ribichini F, Rossi M, Bernardi F, Zonzin P, Piazza A, Yee J, Friedlander Y, Marrugat J, Lucas G, Subirana I, Sala J, Ramos R, Meigs JB, Williams G, Nathan DM, MacRae CA, Havulinna AS, Berglund G, Hirschhorn JN, Asselta R, Duga S, Spreafico M, Daly MJ, Nemesh J, Korn JM, McCarroll SA, Surti A, Guiducci C, Gianniny L, Mirel D, Parkin M, Burt N, Gabriel SB, Thompson JR, Braund PS, Wright BJ, Balmforth AJ, Ball SG, Hall AS, Linsel-Nitschke P, Lieb W, Ziegler A, Konig I, Hengstenberg C, Fischer M, Stark K, Grosshennig A, Preuss M, Wichmann HE, Schreiber S, Ouwehand W, Deloukas P, Scholz M, Cambien F, Li M, Chen Z, Wilensky R, Matthai W, Qasim A, Hakonarson HH, Devaney J, Burnett MS, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Knouff CW, Waterworth DM, Walker MC,

- 387 Mooser V, Epstein SE, Scheffold T, Berger K, Hogue A, Martinelli N, Olivieri O,
388 Corrocher R, McKeown P, Erdmann E, König IR, Holm H, Thorleifsson G,
389 Thorsteinsdóttir U, Stefánsson K, Do R, Xie C, Siscovick D. Genome-wide association of
390 early-onset myocardial infarction with single nucleotide polymorphisms and copy
391 number variants. *Nat Genet.* 2009;41(3):334-41.
- 392 E4. Dawber TR, Meadors GF, Moore FE. Epidemiologic approaches to heart disease: the
393 Framingham study. *Am J Public Health.* 1951;41:179-286.
- 394 E5. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham
395 Offspring Study. Design and preliminary data. *Prev Med.* 1975;4(4):518-25.
- 396 E6. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino
397 RB, Sr., Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D.
398 The Third Generation Cohort of the National Heart, Lung, and Blood Institute's
399 Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol.*
400 2007;165(11):1328-35.
- 401 E7. Aromaa A, Koskinen S, eds. Health and functional capacity in Finland. Baseline Results
402 of the Health 2000 Health Examination Survey. Helsinki: National Public Health Institute
403 KTL; 2004.
- 404 E8. Perttinen J, Merikanto K, Naukkarinen J, Surakka I, Martin NW, Tanhuanpää K, Grimard
405 V, Taskinen MR, Thiele C, Salomaa V, Jula A, Perola M, Virtanen I, Peltonen L,
406 Olkkonen VM. OSBPL10, a novel candidate gene for high triglyceride trait in
407 dyslipidemic Finnish subjects, regulates cellular lipid metabolism. *J Mol Med (Berl).*
408 2009;87(8):825-35.

- 409 E9. Barker DJ, Osmond C, Forsen TJ, Kajantie E, Eriksson JG. Trajectories of growth among
410 children who have coronary events as adults. *N Engl J Med*. 2005;353(17):1802-9.
- 411 E10. Jarvelin MR, Sovio U, King V, Lauren L, Xu B, McCarthy MI, Hartikainen AL, Laitinen
412 J, Zitting P, Rantakallio P, Elliott P. Early life factors and blood pressure at age 31 years
413 in the 1966 northern Finland birth cohort. *Hypertension*. 2004;44(6):838-46.
- 414 E11. Canoy D, Pekkanen J, Elliott P, Pouta A, Laitinen J, Hartikainen AL, Zitting P, Patel S,
415 Little MP, Jarvelin MR. Early growth and adult respiratory function in men and women
416 followed from the fetal period to adulthood. *Thorax*. 2007;62(5):396-402.
- 417 E12. Raitakari OT, Juonala M, Ronnema T, Keltikangas-Jarvinen L, Rasanen L, Pietikainen
418 M, Hutri-Kahonen N, Taittonen L, Jokinen E, Marniemi J, Jula A, Telama R, Kahonen
419 M, Lehtimäki T, Akerblom HK, Viikari JS. Cohort profile: the cardiovascular risk in
420 Young Finns Study. *Int J Epidemiol*. 2008;37(6):1220-6.
- 421 E13. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D,
422 Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD,
423 Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball
424 SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K,
425 Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren
426 OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD,
427 Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D,
428 Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier
429 IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs
430 P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D,
431 Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines

- 432 S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt
 433 S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G,
 434 Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren
 435 CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL,
 436 McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C,
 437 Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR,
 438 Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ,
 439 Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick, Ring SM,
 440 Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott
 441 CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ,
 442 Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z,
 443 Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker
 444 M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN,
 445 Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling
 446 TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP,
 447 Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough
 448 SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR,
 449 Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand
 450 W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. Genome-wide association
 451 study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls.
 452 *Nature*. 2010;464(7289):713-20.
- 453 E14. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C,
 454 Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens

- 455 H, Todd JA, Walker NM, Rich SS. Genome-wide association study and meta-analysis
456 find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41(6):703-7.
- 457 E15. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E,
458 Farrall M, Lathrop M, Cookson WO. A large-scale, consortium-based genomewide
459 association study of asthma. *N Engl J Med.* 2010;363(13):1211-21.
- 460 E16. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory
461 Health Survey. *Eur Respir J.* 1994;7(5):954-60.
- 462 E17. The European Community Respiratory Health Survey II. *Eur Respir J.* 2002;20(5):1071-
463 9.
- 464 E18. Day N, Oakes S, Luben R, Khaw KT, Bingham S, Welch A, Wareham N. EPIC-Norfolk:
465 study design and characteristics of the cohort. European Prospective Investigation of
466 Cancer. *Br J Cancer.* 1999;80 Suppl 1:95-103.
- 467 E19. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy
468 RM, Attwood AP, Beckmann JS, Berndt SI, Jacobs KB, Chanock SJ, Hayes RB,
469 Bergmann S, Bennett AJ, Bingham SA, Bochud M, Brown M, Cauchi S, Connell JM,
470 Cooper C, Smith GD, Day I, Dina C, De S, Dermitzakis ET, Doney AS, Elliott KS,
471 Elliott P, Evans DM, Sadaf Farooqi I, Froguel P, Ghorri J, Groves CJ, Gwilliam R, Hadley
472 D, Hall AS, Hattersley AT, Hebebrand J, Heid IM, Lamina C, Gieger C, Illig T,
473 Meitinger T, Wichmann HE, Herrera B, Hinney A, Hunt SE, Jarvelin MR, Johnson T,
474 Jolley JD, Karpe F, Keniry A, Khaw KT, Luben RN, Mangino M, Marchini J, McArdle
475 WL, McGinnis R, Meyre D, Munroe PB, Morris AD, Ness AR, Neville MJ, Nica AC,
476 Ong KK, O'Rahilly S, Owen KR, Palmer CN, Papadakis K, Potter S, Pouta A, Qi L,
477 Randall JC, Rayner NW, Ring SM, Sandhu MS, Scherag A, Sims MA, Song K, Soranzo

- 478 N, Speliotes EK, Syddall HE, Teichmann SA, Timpson NJ, Tobias JH, Uda M, Vogel CI,
479 Wallace C, Waterworth DM, Weedon MN, Willer CJ, Wraight, Yuan X, Zeggini E,
480 Hirschhorn JN, Strachan DP, Ouwehand WH, Caulfield MJ, Samani NJ, Frayling TM,
481 Vollenweider P, Waeber G, Mooser V, Deloukas P, McCarthy MI, Wareham NJ, Barroso
482 I, Kraft P, Hankinson SE, Hunter DJ, Hu FB, Lyon HN, Voight BF, Ridderstrale M,
483 Groop L, Scheet P, Sanna S, Abecasis GR, Albai G, Nagaraja R, Schlessinger D, Jackson
484 AU, Tuomilehto J, Collins FS, Boehnke M, Mohlke KL. Common variants near MC4R
485 are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008;40(6):768-75.
- 486 E20. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A,
487 Enomoto T, Miyagawa T, Adachi M, Tanaka H, Niimi A, Matsumoto H, Ito I, Masuko
488 H, Sakamoto T, Hizawa N, Taniguchi M, Lima JJ, Irvin CG, Peters SP, Himes BE,
489 Litonjua AA, Tantisira KG, Weiss ST, Kamatani N, Nakamura Y, Tamari M. Genome-
490 wide association study identifies three new susceptibility loci for adult asthma in the
491 Japanese population. *Nat Genet.* 2011.
- 492 E21. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, Kurosaka F,
493 Akasawa A, Yoshihara S, Kanno N, Yamada Y, Shimojo N, Kohno Y, Suzuki Y, Kang
494 MJ, Kwon JW, Hong SJ, Inoue K, Goto Y, Yamashita F, Asada T, Hirose H, Saito I,
495 Fujieda S, Hizawa N, Sakamoto T, Masuko H, Nakamura Y, Nomura I, Tamari M,
496 Arinami T, Yoshida T, Saito H, Matsumoto K. Genome-wide association study identifies
497 HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet.*
498 2011;7(7):e1002170.
- 499 E22. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM,
500 Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, Williams C, Hui J,

- 501 Beilby J, Warrington NM, James A, Palmer LJ, Koppelman GH, Heinzmann A, Krueger
502 M, Boezen HM, Wheatley A, Altmuller J, Shin HD, Uh ST, Cheong HS, Jonsdottir B,
503 Gislason D, Park CS, Rasmussen LM, Porsbjerg C, Hansen JW, Backer V, Werge T,
504 Janson C, Jonsson UB, Ng MC, Chan J, So WY, Ma R, Shah SH, Granger CB, Quyyumi
505 AA, Levey AI, Vaccarino V, Reilly MP, Rader DJ, Williams MJ, van Rij AM, Jones GT,
506 Trabetti E, Malerba G, Pignatti PF, Boner A, Pescollderung L, Girelli D, Olivieri O,
507 Martinelli N, Ludviksson BR, Ludviksdottir D, Eyjolfsson GI, Arnar D, Thorgeirsson G,
508 Deichmann K, Thompson PJ, Wjst M, Hall IP, Postma DS, Gislason T, Gulcher J, Kong
509 A, Jonsdottir I, Thorsteinsdottir U, Stefansson K. Sequence variants affecting eosinophil
510 numbers associate with asthma and myocardial infarction. *Nat Genet.* 2009;41(3):342-7.
- 511 E23. Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk
512 JB, Willis-Owen SA, Klanderman B, Lasky-Su J, Lazarus R, Murphy AJ, Soto-Quiros
513 ME, Avila L, Beaty T, Mathias RA, Ruczinski I, Barnes KC, Celedon JC, Cookson WO,
514 Gauderman WJ, Gilliland FD, Hakonarson H, Lange C, Moffatt MF, O'Connor GT, Raby
515 BA, Silverman EK, Weiss ST. Genome-wide association analysis identifies PDE4D as an
516 asthma-susceptibility gene. *Am J Hum Genet.* 2009;84(5):581-93.

Table S1: Genotyping platform, calling algorithm, imputation details and software used in Stage 1 and Stage 2 studies.

			QC filters applied before imputation								
Study name	GWAS platform	Calling algorithm	Subject call rate	SNP call rate	SNP HWE	SNP MAF	No of SNPs after filtering	Imputation Software	HapMap; CEU ver.	Genotype-phenotype association software	λ for asthma analysis
<i>Stage 1: GWAs for discovery</i>											
FINRISK – CoroGene	Illumina HumanHap 610 Quad	Beadstudio	0.95	0.95	1e-6	0.01	554,988	MACH 1.0.16	36;22	PLINK v1.07	1.0178
FINRISK – MIGen	Affymetrix 6.0 GeneChip	Birdseed	0.95	0.95	1e-6	0.01	727,478	MACH 1.0.16	36;22		
Health 2000 – GeneMets	Illumina HumanHap 610 Quad	ILLUMINUS	0.95	0.95	1e-6	0.01	555,388	Mach 1.0.10	36;22		
Health 2000 – HDL	Illumina 370k	ILLUMINUS	0.95	0.95	1e-6	0.01	339,416	Mach 1.0.16	36;22		
Helsinki Birth Cohort	Illumina 670 Quad	ILLUMINUS	0.95	0.95	1e-6	0.01	546,814	Mach 1.0.16	36;22		
Young Finns Study	Custom-made Illumina 670k	ILLUMINUS	0.90	0.95	1e-6	0.01	546,677	MACH 1.0.16	36; 22	GEE model in R package GWAF	1.0146
Framingham Heart Study	Affymetrix 500K	BRLMM	0.97	0.95	1.e-6	0.01	413,905	MACH 1.0.16	36;22		
Northern Finland Birth Cohort 1966	Illumina 370K	Beadstudio	none	0.95	1.e-4	0.01	328,007	IMPUTE v1.0	35; 21	QUICKTEST v0.94	1.0122
<i>Stage 2: in-silico replication</i>											
Australian Asthma G.C.	Illumina 310K (22%) Illumina 610K (72%)	Beadstudio	0.98	0.95	1e-6	0.01	~610K	IMPUTE	1000G	PLINK v1.07	NA
B58C – WTCCC	Affymetrix 500K	CHIAMO	0.98	No	No	No	490,033	IMPUTE 0.2.0	35;21	ProbABEL v0.0-7	NA
B58C - T1DGC	Illumina 550K	ILLUMINUS	0.98	No	No	No	520,010	MACH 1.0.13	35;21		
B58C - GABRIEL	Illumina Quad 610	GenCall	No	0.97	1e-4	0.05	582,892	MACH 1.0	36; 22		
ECRHS	Illumina Quad 610	GenCall	No	0.97	1e-4	0.05	582,892	MACH 1.0	36; 22	ProbABEL v0.0-7	NA
EPIC obese cases	Affymetrix 500K	BRLMM	0.94	0.90	1e-6	0.01	397,438	IMPUTE 0.3.1	35;21	SNPTEST 1.1.5	NA
EPIC population based	Affymetrix 500K	BRLMM	0.94	0.90	1e-6	0.01	397,438	IMPUTE 0.3.1	35;21	SNPTEST 1.1.5	NA

Table S2: Loci with previous genome-wide significant associations to asthma ($P < 5 \times 10^{-8}$) in Japanese populations

Gene region ^a	SNP ^b	Reported				APCAT			
		Risk Allele	Freq	OR (95% CI)	P value	SNPs examined in APCAT ^c	Risk Allele	OR_95%_CI	P value ^d
<i>USP38-GAB1</i>	rs7686660	T	0.27	1.16(1.11–1.21)	1.87E-12	rs7686660	T	0.97 (0.89,1.05)	8.00E-01
						rs4485768	G	0.97 (0.89,1.04)	8.08E-01
						rs10015501	G	0.97 (0.89,1.05)	7.61E-01
						rs13126430	A	0.98 (0.90,1.05)	7.47E-01
						rs13151714	G	0.97 (0.89,1.05)	7.67E-01
<i>HLA region (NOTCH4)</i>	rs404860	A	0.5	1.21(1.16–1.25)	4.07E-23	rs404860	A	0.94 (0.83,1.05)	8.65E-01
<i>HLA region (HLA-DPB1)</i>	rs987870	C	0.17	1.33(1.20–1.47)	2.3E-10	rs987870	C	0.92 (0.79,1.05)	8.84E-01
<i>IKZF4</i>	rs1701704	G	0.18	1.19(1.14–1.25)	2.33E-13	rs1701704	G	1.13 (1.05,1.21)	1.32E-03
						rs773108	G	1.13 (1.05,1.21)	1.09E-03
						rs773114	T	1.12 (1.04,1.19)	2.03E-03
						rs2292239	G	1.12 (1.04,1.2)	3.43E-03
						rs11171739	C	1.10 (1.03,1.18)	5.83E-03
						rs2069408	G	1.08 (1.00,1.16)	2.91E-02
<i>GATA3/TAF3</i>	rs10508372	C	0.43	1.16(1.12–1.21)	1.79E-23	rs10508372	C	1.08 (0.94,1.23)	1.40E-01
						rs10905491	T	1.16 (0.93,1.40)	8.99E-01
						rs10905488	T	1.08 (0.94,1.23)	8.60E-01

^aGene shown is nearest gene to associated SNP. SNPs from the same locus are grouped together and shaded in the same color. ^bAll SNPs are from Hirota et al. (2011) (E20) except for rs987870, which is from Noguchi et al. (2011) (E21). ^cAll SNPs in LD with $r^2 > 0.8$ with the reported SNP in the Hapmap JPT reference panel were collected, and a “clumped” set of SNPs was generated (only one SNP from a pair within 100 kb and pairwise $r^2 > 0.8$ in HapMap CEU). The association results in APCAT are reported for the most strongly associated SNPs in each clump, for each locus. ^dAPCAT P values are one-tailed with respect to the direction of the risk-increasing allele in the original report.

Table S3: Conditional analysis on novel signals in APCAT

IL1RL1/IL18R1 SNP1 = rs13408661 (APCAT signal) & coded allele is G SNP2 = rs3771166 (GABRIEL signal) (E15) & coded allele is G <i>Distance=31,140 bases and $r^2=0.157$</i>				
	<i>FINRISK, H2000, HBC, YFS</i>	<i>FraminghamHS</i>	<i>NFBC66</i>	<i>pooled</i>
Frequency of coded allele of SNP1	83.59%	83.92%	83.69%	83.81%
Frequency of coded allele of SNP2	58.97%	62.67%	57.87%	60.02%
SNP1 + covs	1.18 (0.99, 1.40); p=6.3e-2	1.31 (1.11, 1.54); p=1.5e-3	1.49 (1.17, 1.88); p=1.0e-3	1.29(1.18,1.4);p=3.84E-06
SNP2 + covs	1.04 (0.91, 1.17); p=5.8e-1	1.10 (0.98, 1.24); p=1.1e-1	1.29 (1.10, 1.52); p=1.5e-3	1.12(1.04,1.19);p=4.7E-03
SNP1 conditional on SNP2 + covs	1.21 (0.99, 1.47); p=6.3e-2	1.31 (1.08, 1.59); p=5.7e-3	1.33 (1.01, 1.74); p=4.3e-2	1.27(1.15,1.4);p=1.23E-04
SNP2 conditional on SNP1 + covs	0.97 (0.83, 1.11); p=6.3e-1	0.99 (0.86, 1.14); p=8.8e-1	1.17 (0.97, 1.40); p=9.9e-2	1.02(0.93,1.11);p=6.9E-01

IL1RL1/IL18R1 SNP1 = rs13408661 (APCAT signal) & coded allele is G SNP2 = rs1420101(Gudbjartsson et al.) (E22) & coded allele is T <i>Distance=2,634 bases and $r^2=0.053$</i>				
	<i>FINRISK, H2000, HBC, YFS</i>	<i>FraminghamHS</i>	<i>NFBC66</i>	<i>pooled</i>
Frequency of coded allele of SNP1	83.59%	83.92%	83.69%	83.81%
Frequency of coded allele of SNP2	31.95%	39.03%	31.95%	34.45%
SNP1 + covs	1.18 (0.99, 1.40); p=6.3e-2	1.31 (1.11, 1.54); p=1.5e-3	1.49 (1.17, 1.88); p=1.0e-3	1.29(1.18,1.4);p=3.84E-06
SNP2 + covs	1.13 (1.00, 1.26); p=7.1e-1	1.14 (1.02, 1.25); p=3.2e-1	1.24 (1.09, 1.40); p=6.5e-3	1.16(1.08,1.23);p=2.0E-04
SNP1 conditional on SNP2 + covs	1.14 (0.96, 1.32); p=1.6e-1	1.27 (1.09, 1.44); p=7.7e-3	1.40 (1.15, 1.64); p=7.83e-3	1.24(1.13,1.35);p=1.5E-04
SNP2 conditional on SNP1 + covs	1.09 (0.96, 1.23); p=1.9e-1	1.08 (0.96, 1.20); p=2.2e-1	1.16 (0.99, 1.32); p=7.8e-2	1.10(1.02,1.18);p=1.5E-02

HLA region SNP1= rs9268516 (APCAT signal) & coded allele is T SNP2 = rs9272346 (proxy for GABRIEL rs9273349 with $r^2=1$) & coded allele is G Distance=224,883 bases and Rsquared=0.336				
	FINRISK, H2000, HBC, YFS	FraminghamHS	NFBC66	Pooled
Frequency of coded allele of SNP1	20.76%	27.25%	22.46%	23.92%
Frequency of coded allele of SNP2	48.26%	39.94%	51.28%	46.27%
SNP1 + covs	1.24 (1.07, 1.43); p=3.6e-3	1.30 (1.14, 1.48); p=5.4e-5	1.22 (1.01, 1.46); p=3.5e-2	1.26(1.17,1.35);p=1.256E-07
SNP2 + covs	0.83 (0.73, 0.94); p=2.6e-3	0.91 (0.78, 1.05); p=2.1e-1	0.82 (0.7, 0.96); p=1.2e-2	0.83(0.74,0.91);p=2.92E-06
SNP1 conditional on SNP2 + covs	1.16 (0.99, 1.35); p=6.6e-2	1.24 (1.08, 1.42); p=2.1e-3	1.12 (0.91, 1.37); p=2.8e-1	1.19(1.09,1.28);p=3E-04
SNP2 conditional on SNP1 + covs	0.87 (0.76, 0.99); p=4.0e-2	0.91 (0.78, 1.05); p=2.1e-1	0.85 (0.72, 1.02); p=7.8e-2	0.88(0.79,0.97);p=3.42E-03

RORA/SMAD3 region SNP1 = rs11071559 (in RORA) & coded allele is T SNP2 = rs744910 (in SMAD3) & coded allele is G Distance=6,376,797				
	FINRISK, H2000, HBC, YFS	FraminghamHS	NFBC66	pooled
Frequency of coded allele of SNP1	15.53%	13.53%	16.07%	14.86%
Frequency of coded allele of SNP2	50.81%	48.52%	52.21%	49.76%
SNP1 + covs	0.86 (0.72, 1.03); p=9.7e-2	0.89 (0.75, 1.06); p=1.8e-1	0.81 (0.65, 1.01); p=6.1e-2	0.86(0.75,0.97);p=6.29E-03
SNP2 + covs	1.12 (1.00, 1.27); p=5.8e-2	1.06 (0.95, 1.19); p=3.0e-1	1.06 (0.91, 1.24); p=4.5e-1	1.08(1.01,1.16);p=3.3E-02
SNP1 conditional on SNP2 + covs	0.86 (0.72, 1.03); p=9.8e-2	0.89 (0.75, 1.06); p=2.0e-1	0.81 (0.65, 1.01); p=5.6e-2	0.86(0.75,0.97);p=6.09E-03
SNP2 conditional on SNP1 + covs	1.12 (1.00, 1.27); p=5.8e-2	1.06 (0.95,1.17);p=3.0e-1	1.07 (0.92, 1.24); p=4.0e-1	1.1(1.01,1.2);p=4.5E-02

Table S4: Variance explained by the GWAs hits under a liability threshold model.

LOCUS	SNP	allele	Frequency in APCAT	Odds Ratio in APCAT	Variance Explained	Used for calculating the total
GSDMA	rs3894194	A	0.4234	1.110	0.137%	yes
GSDMB	rs2305480	A	0.4822	0.940	0.048%	no
ORMDL3	rs7216389	T	0.4729	1.110	0.139%	yes
IL33	rs3939286	T	0.2511	1.180	0.267%	yes
IL33	rs1342326	C	0.1662	1.180	0.197%	no
HLA-DQ	rs9273349	C	0.5005	1.220	0.511%	yes
IL1RL1	rs1420101	T	0.3493	1.160	0.259%	yes
IL18R1	rs3771166	A	0.3977	0.890	0.164%	no
SMAD3	rs744910	A	0.4976	0.920	0.088%	yes
IL2RB	rs2284033	A	0.4741	0.980	0.005%	yes
IL13	rs1295686I	C	0.6938	0.900	0.119%	yes
DENND1B	rs2786098	T	0.2054	0.930	0.043%	yes
PDE4D	rs1588265	G	0.3035	0.960	0.018%	yes
Total Variance Explained					1.586%	

The variance explained was calculated using a liability threshold model by So et al (Genetic Epidemiology, 2011) assuming an asthma prevalence of 9% (the prevalence in APCAT). Only the best SNP from each locus is used to calculate the total variance explained.

Table S5: Stratified association analysis of previously published SNPs and two novel loci emerging from this paper.

SNP (Gene)	chr (position)	Effect/ Other Allele	Effect allele freq.	Smoking stratified analysis					Allergic status stratified analysis		
				No Stratification	in Never Smokers	in Ex-smokers	in Current Smokers	Heteroge neity p- value	in Non-allergic Individuals	in Allergic Individuals	Hetero geneity p-value
Previously reported asthma risk variants											
rs3894194 (GSDMA)	17 (35375519)	A/G	41.7%	1.11 (1.03, 1.19); p=2.4e-03	1.10(0.99,1.22); p=4.10E-02	1.15 (1.03,1.27); p=1.3E-02	1.04 (0.89,1.2); p=2.99E-01	0.623	1.07 (0.93,1.21); p=1.84E-01	1.12 (1.02,1.22); p=1.16E-02	0.570
rs2305480 (GSDMB)	17 (35315722)	A/G	51.0%	0.94 (0.87,1.01); p=4.1E-02	1.00 (0.89,1.11) p=5.10E-01	0.87 (0.75,0.99); p=1.3E-02	0.92 (0.77,1.07); p=1.34-01	0.242	1.01 (0.87,1.15); p=5.72E-01	0.90 (0.8,1.0); p=1.99E-02	0.186
rs7216389 (ORMDL3)	17 (35323475)	T/C	46.6%	1.11 (1.04,1.19); p=2.1E-03	1.06 (0.94,1.18); p=1.60E-01	1.20 (1.08,1.33); p=2.1E-03	1.10 (0.94,1.26); p=1.19E-01	0.340	1.05 (0.9,1.2); p=2.73E-01	1.17 (1.06,1.27); p=1.53E-03	0.245
rs3939286 (IL33)	9 (6200099)	T/C	25.1%	1.18 (1.10,1.26); p=4.8E-05	1.18 (1.05,1.31); p=6.40E-03	1.10 (0.95,1.25); p=1.0E-01	1.33 (1.16,1.5); p=4.66E-04	0.242	1.27 (1.11,1.43); p=1.43E-03	1.20 (1.09,1.32); p=7.75E-04	0.587
rs1342326 (IL33)	9 (6180076)	C/A	16.4%	1.18 (1.09,1.28); p=2.5E-04	1.22 (1.07,1.37); p=5.40E-03	1.05 (0.88,1.22); p=2.9E-01	1.33 (1.14,1.53); p=1.99E-03	0.175	1.33 (1.15,1.52); p=1.20E-03	1.15 (1.02,1.29); p=2.00E-02	0.218
rs9273349 ^a (HLA-DQ)	6 (32733847)	C/T	50.1%	1.22 (1.07,1.38); p=5.6E-03	1.12 (0.87,1.37); p=1.90E-03	1.34 (1.06,1.62); p=2.0E-01	1.19 (0.91,1.48); p=1.14E-01	0.637	1.25 (0.92,1.58); p=9.11E-02	1.30 (1.11,1.49); p=3.31E-03	0.853
rs3771166 (IL18R1)	2 (102352654)	A/G	60.1%	0.89 (0.82,0.97); p=2.5E-03	0.96 (0.84,1.07); p=2.20E-01	0.87 (0.75,1); p=1.8E-02	0.82 (0.67,0.97); p=4.50E-03	0.262	0.82 (0.67,0.97); p=4.26E-03	0.95 (0.84,1.06); p=1.85E-01	0.109
rs1420101 (IL1RL1)	2 (102324148)	T/C	34.4%	1.16 (1.08,1.23); p=9.9E-04	1.09 (0.98,1.21); p=6.70E-02	1.14 (1.01,1.27); p=2.6E-02	1.32 (1.17,1.46); p=1.28E-04	0.136	1.18 (1.03,1.33); p=1.43E-02	1.13 (1.02,1.23); p=1.16E-02	0.625
rs744910 (SMAD3)	15 (65233839)	A/G	50.6%	0.92 (0.85,1.00); p=1.7E-02	0.97 (0.86,1.07); p=2.60E-01	0.89 (0.76,1.02); p=3.5E-02	0.93 (0.78,1.08); p=1.68E-01	0.619	0.98 (0.84,1.11); p=3.63E-01	0.96 (0.86,1.06); p=2.24E-01	0.877
rs11071559 (RORA)	15 (58857280)	T/C	85.0%	0.86 (0.75,0.97); p=3.1E-03	0.9 (0.74,1.07); p=1.1E-01	0.91 (0.73,1.09); p=1.4E-02	0.77 (0.54,0.99); p=1.13E-02	0.463	1.00 (0.81,1.2); p=5.17E-01	0.79 (0.64,0.94); p=1.24E-03	0.059
rs2284033 (IL2RB)	22 (35863980)	A/G	51.9%	0.98 (0.91,1.06); p=3.1E-01	0.96 (0.84,1.07); p=2.40E-01	0.98 (0.85,1.11); p=3.7E-01	1.06 (0.9,1.21); p=7.56E-01	0.608	0.91 (0.78,1.05); p=1.01E-01	1.10 (1.0,1.2); p=9.62E-01	0.038
rs1295686 (IL13)	5 (132023742)	C/T	30.5%	0.90 (0.82,0.98); p=5.4E-03	0.91 (0.78,1.03); p=6.30E-02	0.87 (0.72,1.02); p=3.0E-02	0.97 (0.79,1.14); p=3.46E-01	0.658	0.88 (0.72,1.03); p=5.20E-02	0.91 (0.80,1.02); p=5.44E-02	0.694
rs2244012 (RAD50)	5 (131929124)	G/A	23.3%	1.05 (0.96,1.14); p=1.3E-01	0.95 (0.81,1.08); p=2.06E-01	1.08 (0.93,1.22); p=1.7E-01	1.04 (0.86,1.22); p=3.40E-01	0.958	1.08 (0.91,1.25); p=1.86E-01	0.99 (0.87,1.11); p=5.40E-01	0.432
rs2073643 (SLC22A5)	5 (131751187)	C/T	52.8%	0.96 (0.88,1.04); p=1.6E-01	0.98 (0.86,1.1); p=3.50E-01	0.94 (0.81,1.07); p=1.7E-01	0.97 (0.82,1.13); p=3.70E-01	0.892	0.89 (0.74,1.05); p=7.47E-02	0.98 (0.87,1.08); p=3.23E-01	0.353
rs2786098 (DENND1B)	1 (195592531)	T/G	79.8%	0.93 (0.89,1.02); p=5.8E-02	1.01 (0.88,1.14); p=5.60E-01	0.9 (0.74,1.06); p=9.1E-02	0.88 (0.7,1.07); p=9.48E-02	0.385	0.95 (0.77,1.12); p=2.65E-01	0.94 (0.81,1.07); p=1.74E-01	0.962

rs1588265 ^b (PDE4D)	5 (59405551)	G/A	70.1%	0.96 (0.85,1.06); p=2.0E-01		0.84 (0.71,0.97); p=3.60E-03	0.98 (0.85,1.12); p=4.1E-01	0.92 (0.76,1.09); p=1.72E-01	0.248		0.82 (0.66,0.99); p=1.02E-02	0.94 (0.83,1.06); p=1.64E-01	0.181
<i>Risk variants emerging from this paper</i>													
rs13408661 (<i>IL1RL1</i> <i>/IL18R1</i>)	2 (102321514)	G/A	0.84%	1.29 (1.18,1.4); p=3.80E-06		1.28 (1.27,1.47); p=1.41E-03	1.23 (1.05,1.41); p=1.24E-02	1.42 (1.19,1.65); p=1.26E-03	0.554		1.39 (1.18,1.60); p=1.03E-03	1.20 (1.05,1.32); p=7.17E-03	0.738
rs9268516 (HLA Region)	6 (32487467)	T/C	0.24	1.26 (1.17,1.35); p=1.30E-07		1.22 (1.09,1.36); p=1.30E-03	1.49 (1.27,1.71); p=8.7E-05	1.19 (1.02,1.36); p=2.06E-02	0.425		1.27 (1.11,1.43); p=1.58E-03	1.23 (1.11,1.35); p=4.32E-04	0.102

All the p-values presented in this table are based on one-sided testing except in the no stratification column. ^a SNP rs9273349 is present in NFBC66 data set only. ^bResults exclude the Framingham Heart Study, which contributed to the original report in Himes et al (2009) (E23).