



# UNIVERSITY OF TAMPERE

This document has been downloaded from  
Tampub – The Institutional Repository of University of Tampere

The permanent address of the publication is <http://urn.fi/URN:NBN:fi:uta-201301041004>

Author(s): Ramasamy, Adaikalavan; Kuokkanen, Mikko; Vedantam, Sailaja; Kähönen, Mika; Lehtimäki, Terho **et al.**

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

Year: 2012

Journal Title: Plos ONE

Vol and number: 7 : 9

Pages: 1-10

ISSN: 1932-6203

Discipline: Biomedicine

School /Other Unit: School of Medicine

Item Type: Journal Article

Language: en

DOI: <http://dx.doi.org/doi:10.1371/journal.pone.0044008>

URN: URN:NBN:fi:uta-201301041004

URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0044008>

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.

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

Adaikalavan Ramasamy<sup>1,2,3\*</sup>, Mikko Kuokkanen<sup>4\*</sup>, Sailaja Vedantam<sup>5,6\*</sup>, Zofia K. Gajdos<sup>5,6\*</sup>, Alexsander Couto Alves<sup>2</sup>, Helen N. Lyon<sup>5,6</sup>, Manuel A. R. Ferreira<sup>7</sup>, David P. Strachan<sup>8</sup>, Jing Hua Zhao<sup>9</sup>, Michael J. Abramson<sup>10</sup>, Matthew A. Brown<sup>11</sup>, Lachlan Coin<sup>2</sup>, Shyamali C. Dharmage<sup>12</sup>, David L. Duffy<sup>7</sup>, Tari Haahtela<sup>13</sup>, Andrew C. Heath<sup>14</sup>, Christer Janson<sup>15</sup>, Mika Kähönen<sup>16</sup>, Kay-Tee Khaw<sup>17</sup>, Jaana Laitinen<sup>18</sup>, Peter Le Souef<sup>19</sup>, Terho Lehtimäki<sup>20</sup>, Australian Asthma Genetics Consortium collaborators<sup>†</sup>, Pamela A. F. Madden<sup>14</sup>, Guy B. Marks<sup>21</sup>, Nicholas G. Martin<sup>7</sup>, Melanie C. Matheson<sup>12</sup>, Cameron D. Palmer<sup>5,6</sup>, Aarno Palotie<sup>22,23,24,25</sup>, Anneli Pouta<sup>26,27</sup>, Colin F. Robertson<sup>28</sup>, Jorma Viikari<sup>29</sup>, Elisabeth Widen<sup>23</sup>, Matthias Wjst<sup>30</sup>, Deborah L. Jarvis<sup>1,31</sup>, Grant W. Montgomery<sup>7</sup>, Philip J. Thompson<sup>32</sup>, Nick Wareham<sup>9</sup>, Johan Eriksson<sup>33,34,35,36</sup>, Pekka Jousilahti<sup>4</sup>, Tarja Laitinen<sup>37,38</sup>, Juha Pekkanen<sup>39,40</sup>, Olli T. Raitakari<sup>41,42</sup>, George T. O'Connor<sup>43,44</sup>, Veikko Salomaa<sup>4\*</sup>, Marjo-Riitta Jarvelin<sup>2,31,33,45,46\*</sup>, Joel N. Hirschhorn<sup>5,6,47\*</sup>

**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 EW (#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 *PYHINI* 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 ( $P_{\text{Stage 1}} = 3.9 \times 10^{-6}$ ;  $P_{\text{Stage 2}} = 3.2 \times 10^{-5}$ ;  $P_{\text{Stage 1+Stage 2}} = 1.1 \times 10^{-9}$ ; **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 <sup>a</sup>
<b>Stage1: APCAT studies for discovery</b>							
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
<b>Total for Stage1</b>	<b>1,716</b>	<b>16,888</b>			<b>43.3%</b>	<b>55.6%</b>	
<b>Stage2: in-silico replication</b>							
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) <sup>b</sup>	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 <sup>c</sup> population based	216	2,005	59	46.8%	45.4%	87.9%	Affymetrix 500K
EPIC-Norfolk <sup>c</sup> obese cases	123	910	60	43.0%	44.3%	87.2%	Affymetrix 500K
<b>Total for Stage2</b>	<b>4,035</b>	<b>11,251</b>					

<sup>a</sup>See **Table S1** for more information on genotyping, imputation and software used. <sup>b</sup>The characteristics of the studies in the AAGC are presented in Ferreira et al., 2011 [8]. <sup>c</sup> 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 <sup>a</sup>	SNP	Study <sup>b</sup>	Effect Allele <sup>c</sup>	Reported		APCAT		APCAT+Reported	
				OR (95% CI)	P value	OR (95% CI)	P value <sup>d</sup>	OR (95% CI)	P value <sup>e</sup>
<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) <sup>f</sup>	8.5E-08 <sup>f</sup>	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) <sup>g</sup>	5.6E-03 <sup>g</sup>	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 <sup>i</sup>	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 <sup>h</sup>	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

<sup>a</sup>Gene shown is nearest gene to associated SNP. SNPs from the same locus are grouped together. <sup>b</sup>References: 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]. <sup>c</sup>Alleles are indexed to the forward strand of NCBI build36. <sup>d</sup>APCAT *P* values are one-tailed with respect to the direction of the original association. <sup>e</sup>*P* values are from fixed-effect inverse-variance model of meta analysis. <sup>f</sup>Results shown are from Moffatt et al (2010), which is the larger and more recent study. <sup>g</sup> SNP rs9273349 is present in NFBC1966 data set only. <sup>h</sup>Results exclude the Framingham Heart Study, which contributed to the original report in Himes et al (2009) <sup>i</sup>Shown 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

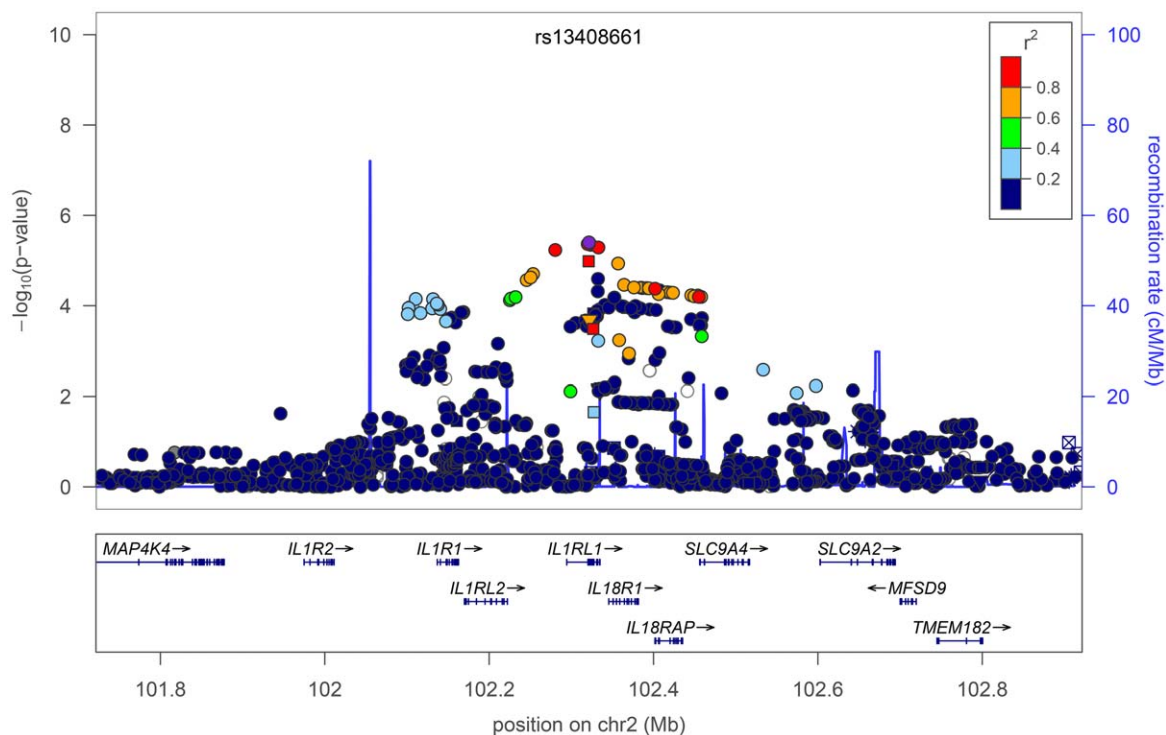
To our knowledge, there has been no genome-wide search for gene-environment interactions with smoking status or allergic status, two important modifiers for risk of developing asthma. We scanned ~2.2 million SNPs for gene-environment effects for smoking exposure in APCAT by comparing, for each SNP, the 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,



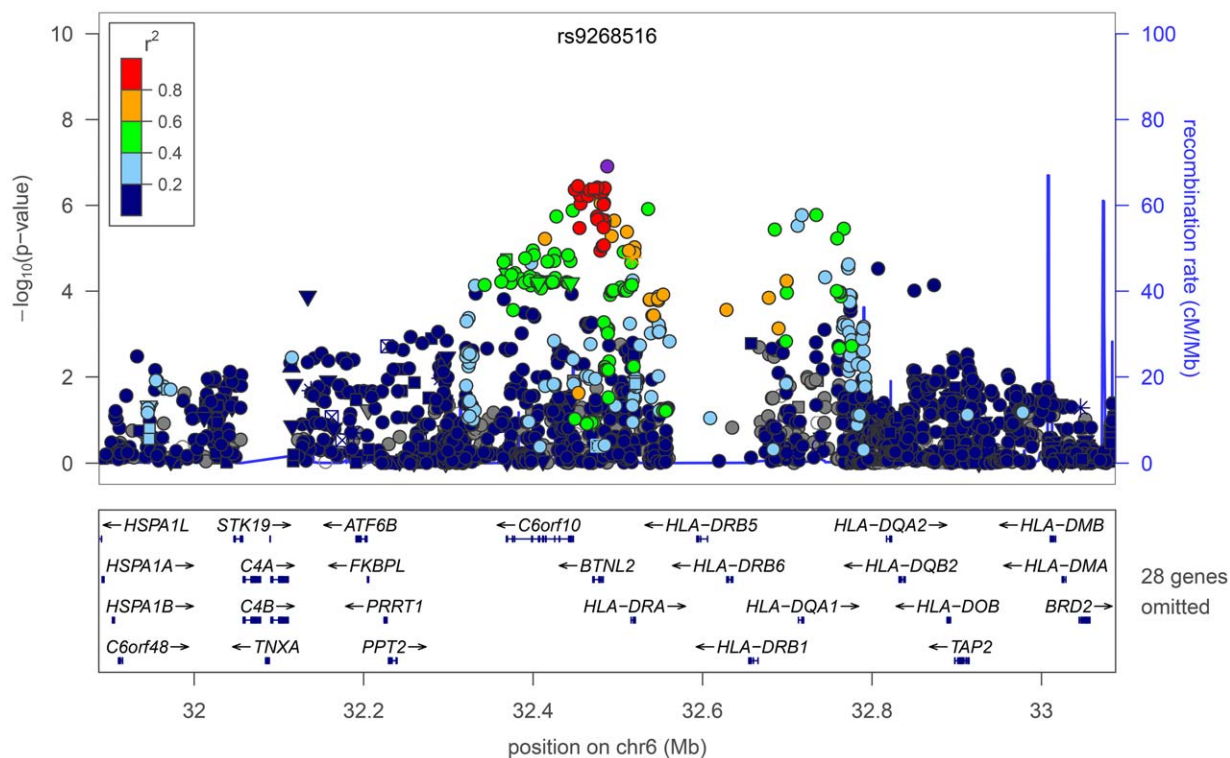
**Table 3.** Replication results for top signals from APCAT (Stage1 N= 18,604) in additional studies (Stage 2 N = 15,576) and in GABRIEL.

Nearest Gene(s)	SNP	Chr (position <sup>a</sup> )	Effect/alternate allele <sup>a</sup>	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 <sup>b</sup> fixed-effects OR (95% CI) and p value	GABRIEL SNP ( $r^2$ in HAPMAP CEU with APCAT SNP) <sup>c</sup>
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/TMTC2	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 ( $r^2 = 0.84$ )

<sup>a</sup>Positions/alleles are relative to the forward strand of NCBI build36. <sup>b</sup>Results 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](http://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 GWAF 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/IL18R* 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 neutrophils, 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 ORMDL3 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 ORMDL3, 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) ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 177: 1194–1200.
- Barnes KC (2010) Genome-wide association studies in allergy and the influence of ethnicity. *Curr Opin Allergy Clin Immunol* 10: 427–433.
- Himes BE, Hunninghake GM, Baurley JW, Rafaels 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, Helgadóttir 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 genome-wide 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 genome-wide 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.