Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks

DOI:
10.1038/s41588-017-0014-7

Document Version
Accepted author manuscript

Multi-ancestry association study identifies new asthma risk loci that co-localize with immune cell enhancer marks

Florence Demenais[1,2*], Patricia Margaritte-Jeannin[1,2*], Kathleen C Barnes[3*], William OC Cookson[4*], Janine Altmüller[5], Wei Ang[6], R Graham Barr[7], Terri H Beaty[8], Allan B Becker[9], John Beilby[10], Hans Bisgaard[11], Unnur Steina Bjornsdottir[12], Eugene Bleecker[13], Klaus Bønnelykke[11], Dorret I Boomsma[14], Emmanuelle Bouzigon[1,2], Christopher E Brightling[15], Myriam Brossard[1,2], Guy G Brusselle[16,17,18], Esteban Burchard[19], Kristin M Burkart[20], Andrew Bush[21,22], Moira Chan-Yeung[23], Kian Fan Chung[21,24], Alexessander Couto Alves[25], John A Curtin[26], Adnan Custovic[27], Denise Daley[23,28], Johan C de Jongste[29], Blanca E Del-Rio-Navarro[30], Kathleen M Donohue[7], Liesbeth Duijts[31], Celeste Eng[32], Johan G Eriksson[33], Martin Farrall[34,35], Yuliya Fedorova[36], Bjarke Feenstra[37], Manuel A Ferreira[38], Australian Asthma Genetics Consortium collaborators, Maxim B Freidin[39], Zofia Gajdos[40,41], Jim Gauderman[42], Ulrike Gehring[43], Frank Geller[37], Jon Genuneit[44], Sina A Gharib[45], Frank Gilliland[42], Raquel Granell[46,47], Penelope E Graves[48], Daniel F Gudbjartsson[49,50], Tari Haahtela[51], Susan R Heckbert[52], Dick Heederik[53], Joachim Heinrich[54,55], Markku Heliövaara[56], John Henderson[46,47], Blanca E Himes[57], Hiroshi Hirose[58], Joel N Hirschhorn[41,59,60], Albert Hofman[17,61], Patrick Holt[62], Jouke Jan Hottenga[14], Thomas J Hudson[63,64], Jennie Hui[10,65,66], Medea Imboden[67,68], Vladimir Ivanov[69], Vincent WV Jaddoe[70], Alan James[71,72], Christer Janson[73], Marjo-Riitta Jarvelin[74,75,76,77], Deborah Jarvis[21,78], Graham Jones[79], Ingileif Jonsdottir[49,80], Pekka Jousilahti[56], Michael Kabesch[81], Mika Kähönen[82], David B Kantor[83,84], Alexandra S Karunas[36,85], Elza Khusnutdinova[36,85], Gerard H Koppelman[86,87], Anita L Kozyrskyj[88], Eskil Kreiner[11], Michiaki Kubo[89], Rajesh Kumar[90,91], Ashish Kumar[67,68,92], Mikko Kuokkanen[56,93], Lies Lahousse[17,94], Tarja Laitinen[95], Catherine Laprise[96,97], Mark Lathrop[98], Susanne Lau[99], Young-Ae Lee[100,101], Terho Lehtimäki[102], Sébastien Letort[1,2], Albert M Levin[103], Guo Li[45], Liming Liang[61,104], Laura R Loehr[105], Stephanie J London[106], Daan W Loth[17], Ani Manichaikul[107], Ingo Marenholz[100,101], Fernando J Martinez[48], Melanie C Matheson[108], Rasika A Mathias[109], Kenji Matsumoto[110], Hamdi Mbarek[14], Wendy L McArdle[111], Mads Melbye[37,112,113], Erik Melén[92,114,115], Deborah Meyers[13], Sven Michel[81], Hamida Mohamdi[1,2], Arthur W Musk[116,117], Rachel A Myers[118], Maartje A E Nieuwenhuis[87,119], Emiko Noguchi[120], George T O'Connor[121,122], Ludmila M Ogorodova[123], Cameron D Palmer[41,59], Aarno Palotie[93,124,125], Julie E Park[23], Craig E Pennell[6], Göran Pershagen[92,114], Alexey Polonikov[69], Dirkje S Postma[87,119], Nicole Probst-Hensch[67,68], Valery P Puzyrev[39], Benjamin A Raby[126], Olli T Raitakari[127],

Adaikalavan Ramasamy[25,128], Stephen S Rich[107], Colin F Robertson[129], Isabelle Romieu[130,131], Muhammad T Salam[42,132], Veikko Salomaa[56], Vivi Schlünssen[133], Robert Scott[134], Polina A Selivanova[135], Torben Sigsgaard[133], Angela Simpson[26,136], Valérie Siroux[137,138], Lewis J Smith[139], Maria Solodilova[69], Marie Standl[55], Kari Stefansson[49,80], David P Strachan[140], Bruno H Stricker[17,141,142], Atsushi Takahashi[89], Philip J Thompson[143], Gudmar Thorleifsson[49], Unnur Thorsteinsdottir[49,80], Carla MT Tiesler[55,144], Dara G Torgerson[32], Tatsuhiko Tsunoda[89,145], André G Uitterlinden[142], Ralf JP van der Valk[146], Amaury Vaysse[1,2], Sailaja Vedantam[40,41], Andrea Von Berg[147], Erika Von Mutius[148,149], Judith M Vonk[87,150], Johannes Waage[11], Nick J Wareham[134], Scott T Weiss[126], Wendy B White[151], Magnus Wickman[92,152], Elisabeth Widén[93], Gonneke Willemsen[14], L Keoki Williams[153,154], Inge M Wouters[53], James J Yang[155], Jing Hua Zhao[134], Miriam F Moffatt[4*], Carole Ober[156*] & Dan L Nicolae[157*]

*These authors contributed equally to this work

[1]Genetic Variation and Human Diseases Unit, Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France.

[2]Institut Universitaire d'Hématologie, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France.

[3]Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado, Denver, Colorado, USA.

[4]Section of Genomic Medicine, National Heart and Lung Institute, London, UK.

[5]Cologne Center for Genomics and Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany.

[6]School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia.

[7]Departments of Medicine and Epidemiology, Columbia University, New York, New York, USA.

[8]Division of Genetic Epidemiology, Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.

[9]Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.

[10]Department of the Diagnostic Genomics Laboratory, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia.

[11]Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

[12]Department of Medicine, Landspitali, The National University Hospital of Iceland, Reykjavik, Iceland.

[13]Center for Genomics, Wake Forest University School of Medicine, Winston Salem, North Carolina, USA.
[14]Department of Biological Psychology, Amsterdam Public Health research institute, Vrjie Universiteit, Amsterdam, The Netherlands.
[15]University of Leicester, Institute for Lung Health, Glenfield Hospital, Leicester, UK.
[16]Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.
[17]Department of Epidemiology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[18]Department of Respiratory Medicine, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[19]Department of Bioengineering & Therapeutic Sciences and Medicine, University of California, San Francisco, California, USA.
[20]Division of Pulmonary, Allergy and Critical Care, College of Physicians and Surgeons, Columbia University, New York, New York, USA.
[21]National Heart and Lung Institute, Imperial College, London, UK.
[22]Royal Brompton Harefield National Health Service (NHS) Foundation Trust, London, UK.
[23]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.
[24]Biomedical Research Unit, Royal Brompton & Harefield National Health Service (NHS) Trust, London, UK.
[25]Department of Epidemiology and Biostatistics, Imperial College London, London, UK.
[26]Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.
[27]Department of Paediatrics, Imperial College London, London, UK.
[28]Centre for Heart and Lung Innovation, University of British Columbia, Vancouver, British Columbia, Canada.
[29]Department of Pediatrics - Division of Respiratory Medicine, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[30]Hospital Infantil de Mexico Federico Gomez, Mexico City, Mexico.
[31]Department of Pediatrics-Division of Respiratory Medicine and Department of Pediatrics-Division of Neonatology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[32]Department of Medicine, University of California, San Francisco, California, USA.
[33]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

[34]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.
[35]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[36]Institute of Biochemistry and Genetics, Ufa Scientific Center of Russian Academy of Sciences, Ufa, Russian Federation.
[37]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.
[38]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.
[39]Population Genetics Laboratory, Research Institute of Medical Genetics, Tomsk NRMC, Tomsk, Russian Federation.
[40]Divisions of Genetics and Endocrinology, Children's Hospital, Boston, Massachusetts, USA.
[41]Broad Institute, Cambridge, Massachusetts, USA.
[42]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.
[43]Division Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands.
[44]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany.
[45]Department of Medicine, University of Washington, Seattle, Washington, USA.
[46]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
[47]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
[48]Asthma and Airway Disease Research Center and BIO5 Institute, The University of Arizona, Tucson, Arizona, USA.
[49]deCODE genetics, Amgen Inc., Reykjavik, Iceland.
[50]School of Engineering and Natural Sciences, Univeristy of Iceland, Reykjavik, Iceland.
[51]Skin and Allergy Hospital, University of Helsinki, Helsinki, Finland.
[52]Department of Epidemiology, University of Washington, Seattle, Washington, USA.
[53]Division Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands.
[54]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, University Hospital in Munich, Munich, Germany.
[55]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany.
[56]National Institute for Health and Welfare (THL), Helsinki, Finland.

[57]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[58]Health Center, Department of Internal Medicine, Keio University, Tokyo, Japan.
[59]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, USA.
[60]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, Massachusetts, USA.
[61]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
[62]Cell Biology Telethon Kids Institute, The University of Western Australia, Subiaco, Western Australia, Australia.
[63]Ontario Institute for Cancer Research, Toronto, Ontario, Canada.
[64]AbbVie Inc., Redwood City, California, USA.
[65]Busselton Population Medical Research Institute, Perth, Western Australia, Australia.
[66]School of Population and Global Health, University of Western Australia, Nedlands, Western Australia, Australia.
[67]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland.
[68]University of Basel, Basel, Switzerland.
[69]Department of Biology, Medical Genetics and Ecology, Kursk State Medical University, Kursk, Russian Federation.
[70]The Generation R Study Group, Department of Pediatrics and Department of Epidemiology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[71]Department of Pulmonary Physiology and Sleep Medicine, Busselton Population Medical Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia.
[72]School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia.
[73]Department of Medical Sciences: Respiratory, Allergy & Sleep Research, Uppsala University, Uppsala, Sweden.
[74]Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK.
[75]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland.
[76]Biocenter Oulu, University of Oulu, Oulu, Finland.
[77]Unit of Primary Care, Oulu University Hospital, Oulu, Finland.
[78]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK.

[79]School of Science and Health, Western Sydney University, Sydney, New South Wales, Australia.
[80]Faculty of Medicine, Univeristy of Iceland, Reykjavik, Iceland.
[81]Department of Pediatric Pneumology and Allergy, University Children's Hospital Regensburg (KUNO), Regensburg, Germany.
[82]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland.
[83]Department of Anesthesiology, Perioperative and Pain Medicine, Division of Critical Care Medicine, Boston Children's Hospital, Boston, Massachusetts, USA.
[84]Department of Anaesthesia, Harvard Medical School, Boston, Massachusetts, USA.
[85]Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russian Federation.
[86]University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, the Netherlands.
[87]Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands.
[88]Department of Pediatrics, The University of Alberta, Edmonton, Alberta, Canada.
[89]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
[90]Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA.
[91]Divison of Allergy and Clinical Immunology, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
[92]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
[93]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.
[94]Pharmaceutical Care Unit, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium.
[95]Department of Pulmonary Medicine, University of Turku and Turku University Hospital, Turku, Finland.
[96]Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Chicoutimi, Quebec, Canada.
[97]Centre de santé et de services sociaux du Saguenay-Lac-Saint-Jean, Saguenay, Québec, Canada.
[98]McGill University and Genome Quebec Innovation Centre, Montréal, Quebec, Canada.
[99]Pediatric Pneumology and Immunology, Charité Universitätsmedizin, Berlin, Germany.
[100]Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany.
[101]Pediatric Allergology, Experimental and Clinical Research Center, Charité Universitätsmedizin, Berlin, Germany.
[102]Department of Clinical Chemistry, Fimlab Laboratories, University of Tampere Faculty of Medicine and Life Sciences, Tampere, Finland.
[103]Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan, USA.

[104]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
[105]Division of General Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
[106]National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA.
[107]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA.
[108]Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.
[109]Division of Allergy & Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA.
[110]Department of Allergy and Clinical Immunology, National Research Institute for Child Health and Development, Tokyo, Japan.
[111]Bristol Bioresource Laboratories, School of Social and Community Medicine, University of Bristol, Bristol, UK.
[112]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.
[113]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.
[114]Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden.
[115]Sachs Children's Hospital, Stockholm, Sweden.
[116]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia.
[117]Schools of Population Health and Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia.
[118]Center for Applied Genomics and Precision Medicine, Duke University School of Medicine, Durham, North Carolina, USA.
[119]University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands.
[120]Department of Medical Genetics, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.
[121]Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.
[122]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA.
[123]Department of Faculty Pediatrics, Siberian State Medical University, Tomsk, Russian Federation.
[124]Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA.

[125]The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.
[126]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
[127]Department of Clinical Physiology and Nuclear Medicine, University of Turku and Turku University Hospital, Turku, Finland.
[128]Department of Medical and Molecular Genetics, King's College London, London, UK.
[129]Respiratory Medicine, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.
[130]Hubert department of Global Health, Mory University, Atlanta, Georgia, USA.
[131]Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico.
[132]Department of Psychiatry, Kern Medical, Bakersfield, California, USA.
[133]Department of Public Health, Section for Environment, Occupation & Health, Aarhus University, Aarhus, Denmark.
[134]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK.
[135]Department of Faculty Therapy, Siberian State Medical University, Tomsk, Russian Federation.
[136]University Hospital of South Manchester, National Health Service (NHS) Foundation Trust, Manchester, UK.
[137]Institut National de la Santé et de la Recherche Médicale (INSERM) U1209, Institute for Advanced Biosciences, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Grenoble, France.
[138]Université de Grenoble Alpes / CNRS UMR5309, Institute for Advanced Biosciences, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Grenoble, France.
[139]Division of Pulmonary and Critical Care Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA.
[140]Population Health Research Institute, St George's University of London, London, UK.
[141]Netherlands Healthcare Inspectorate, The Hague, The Netherlands.
[142]Department of Internal Medicine, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[143]Institute for Respiratory Health and Harry Perkins Institute of Medical Research, University of Western Australia and The Lung Health Clinic, Nedlands, Western Australia, Australia.
[144]Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany.

[145]Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan.
[146]The Generation R Study Group, Department of Pediatrics - Division of Respiratory Medicine and Department of Epidemiology, Erasmus MC-University Medical Center Rotterdam, Rotterdam, The Netherlands.
[147]Department of Pediatrics, Marien-Hospital Wesel, Wesel, Germany.
[148]Dr. Von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany.
[149]Member of the German Center for Lung Research, Germany.
[150]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands.
[151]Undergraduate Training and Education Center (UTEC), Jackson Heart Study, Tougaloo College, Jackson, Mississippi, USA.
[152]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden.
[153]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA.
[154]Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan, USA.
[155]School of Nursing, University of Michigan, Ann Arbor, Michigan, USA.
[156]Department of Human Genetics, The University of Chicago, Chicago, Illinois, USA.
[157]Departments of Statistics, Human Genetics and Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA.

Correspondence should be sent to:
Florence Demenais
Genetic Variation and Human Diseases Unit
UMR-946, INSERM-Université Paris Diderot
27 rue Juliette Dodu
75010 Paris, France
Phone : +33 1 72 63 93 02 – Fax : +33 01 72 63 93 49 - e-mail: florence.demenais@inserm.fr
or
Dan Nicolae
Departments of Statistics
The University of Chicago
5734 S. University Avenue
Chicago, IL 60637, USA
Phone: +1 773 702 4837 - Fax: +1 773 702 9810 - e-mail: nicolae@galton.uchicago.edu

ABSTRACT

We examined common variation in asthma risk by conducting a meta-analysis of worldwide asthma genome-wide association studies (23,948 cases, 118,538 controls) from ethnically-diverse populations. We identified five new asthma loci, uncovered two additional novel associations at two known asthma loci, established asthma associations at two loci implicated previously in comorbidity of asthma plus hay fever, and confirmed nine known loci. Investigation of pleiotropy showed large overlaps in genetic variants with autoimmune and inflammatory diseases. Enrichment of asthma risk loci in enhancer marks, especially in immune cells, suggests a major role of these loci in the regulation of immune-related mechanisms.

312 Asthma is a complex disease affecting hundreds of millions of people worldwide. Asthma
313 prevalence varies between populations and ethnicities, ranging in the U.S. from 3.9% in Mexican
314 Americans to 12.5% in African Americans[1]. The contribution of genetic factors to asthma risk has
315 been demonstrated in family studies, where heritability estimates range from 25%-80%[2]. The large
316 variability in prevalence and heritability estimates reflects the role of environmental exposures on
317 disease risk and phenotypic heterogeneity that are hallmarks of asthma. These features may explain
318 why genome-wide association studies (GWAS) have uncovered a smaller number of asthma loci
319 than similarly sized studies of other multifactorial diseases[3]. Indeed, up to now, only 21 loci have
320 been associated with asthma *per se* in 20 studies, and these loci explain only part of the genetic
321 risk. Although an exome-chip study showed no evidence for low frequency or rare variants with
322 large effects on asthma risk[4], the role of noncoding rare variants in asthma remains unknown.
323 Future studies based on whole-genome sequencing may clarify the respective influence of
324 common and rare variants on asthma risk. To generate larger sample sizes for GWAS meta-analysis
325 of asthma enabling the discovery of common novel risk loci, we established the Trans-National
326 Asthma Genetic Consortium (TAGC) across worldwide groups of investigators with genome-wide
327 data available in >142,000 individuals of diverse ancestries. We constructed a comprehensive
328 catalog of asthma risk variants that are robust across populations and environmental exposures. By
329 combining TAGC meta-analysis results with existing databases, we assessed the genetic
330 architecture of asthma risk alleles with respect to functional effects and shared effects with other
331 diseases.

332

333 **RESULTS**

334 **Meta-analysis of asthma GWAS**

We combined data from asthma GWAS with high-density genotyped and imputed SNP data (2.83 million SNPs) in the following populations: European-ancestry (19,954 cases, 107,715 controls), African-ancestry (2,149 cases, 6,055 controls), Japanese (1,239 cases, 3,976 controls) and Latino (606 cases, 792 controls) (Supplementary Table 1). After extensive QC of summary data provided by each participating group (Online Methods, Supplementary Note and Supplementary Table 2), we first conducted ancestry-specific meta-analyses followed by a multi-ancestry meta-analysis of all populations (23,948 cases, 118,538 controls) to identify additional loci with pan-ancestry effects. Because childhood-onset asthma may be distinct from later-onset asthma[5] and may represent a more homogeneous subgroup, we also performed analyses on the pediatric subgroup (asthma onset ≤16 years; 8,976 cases, 18,399 controls). Meta-analyses of SNP effect sizes obtained from each asthma GWAS were performed using fixed-effects (significance of the combined SNP effect size summarized in P_{fixed}) and random-effects (P_{random}) models (Online Methods) and using a conventional P_{random} (or P_{fixed}) threshold of $5x10^{-8}$ to define genome-wide significance. Results were consistent between methods for detecting loci with at least one SNP significantly associated with asthma. We therefore present the results from the random-effects analysis for the European-ancestry and multi-ancestry meta-analyses, which include the largest number of studies and allow an accurate estimate of the between-study variance, and results from the fixed-effects analysis for the African-ancestry, Japanese and Latino meta-analyses. We observed little evidence of inflation in the test statistics in either the ancestry-specific (lambda: European-ancestry, 1.031; African-ancestry, 1.014; Japanese, 1.021; Latino, 1.044) or multi-ancestry (lambda=1.046) meta-analyses (Supplementary Fig.1).

We identified 673 genome-wide significant SNPs ($P_{random} \leq 5x10^{-8}$) at 16 loci in European-ancestry populations (Fig.1a, Table 1, Supplementary Tables 3 and 4; Online Methods for locus definition). No genome-wide significant risk loci were detected in African-ancestry, Japanese or

Latino populations (Supplementary Fig.2 and Supplementary Tables 5-7), possibly due to a lack of power. In the combined multi-ancestry meta-analysis, 205 additional SNPs were significant at $P_{random} \leq 5x10^{-8}$, including 12 SNPs at two loci not detected in the European-ancestry analysis (Fig.1b, Table 1, and Supplementary Tables 3 and 8). Altogether, 878 SNPs at 18 loci reached genome-wide significance, of which 69% were significant in both European-ancestry and multi-ancestry meta-analyses, 23% were significant in the multi-ancestry meta-analysis only, and 8% were significant in the European-ancestry meta-analysis only (Supplementary Tables 4 and 8 and Supplementary Fig.3 for the regional plots of the 18 loci). All 18 loci remained genome-wide significant after further genomic control correction of the test statistics, confirming the robustness of these results (Supplementary Table 9).

The 18 chromosomal regions included five new loci associated with asthma at 5q31.3, 6p22.1, 6q15, 12q13.3, 17q21.33; two new associations at 6p21.33 and 10p14 that were independent from previously reported signals at these loci in ancestry-specific populations (Latino[6] and Japanese[7], respectively); two associations at 8q21.13 and 16p13.13 that were previously reported for asthma plus hay fever but not for asthma alone in a study of European-ancestry populations[8]; and nine previously identified asthma loci.

None of the lead SNPs at the 18 loci showed evidence for heterogeneity in effect sizes across studies except for the lead variant at 9p24.1 (P_{het} for Cochran's Q test[9]=0.008 across European-ancestry studies and P_{het}=0.02 across multi-ancestry studies, Table 1, Supplementary Fig.4). There was also significant evidence for heterogeneity in ancestry-specific effect sizes (P_{ethnic} =0.003) for the 6p22.1 lead SNP rs1233578, which, consequently, did not reach significance in the multi-ancestry analysis (Table 1, Supplementary Table 3). The meta-analysis of the pediatric subgroup showed evidence for association ($P_{random} \leq 5x10^{-8}$) at five of the 18 loci (2q12, 5q31, 6p21.33 9p24.1

382 and 17q12-21) (Supplementary Figs.5 and 6 and Supplementary Table 10). No loci specific to that
383 group were identified.

384 The results provided genome-wide significant confirmation of nine previously reported loci
385 in both European-ancestry and multi-ancestry meta-analyses (Table 1; Supplementary Figs.3b and
386 4). Our results allow detailed analysis of the broad 17q12-21 locus. It is notable that the lead SNP
387 (rs2952156) at this locus is within *ERBB2* (P_{random}=2.2x10^{-30} in multi-ancestry meta-analysis), at
388 least 180 kb from the previously recognized asthma-associated signals at the *GSDMB/ORMDL3*
389 haplotype block[3] (Supplementary Fig.7). This is attributable to effect size heterogeneity across
390 studies ($0.001 \leq P_{het} \leq 0.05$) that extends over a 200 kb region that includes *ORMDL3* and *GSDMB*
391 (Supplementary Table 11). This heterogeneity is partly due to age of asthma onset, as previously
392 reported[5]. Indeed, in the pediatric group, the 17q12-21 SNPs did not show heterogeneity ($P_{het} \geq 0.09$)
393 and the lead SNP rs8069176 is 3.6 kb proximal to G*SDMB* ($P_{random}=P_{fixed}$=4.4 x10^{-26}), consistent
394 with previous studies[3,5]. The SNP effect sizes in the pediatric and non-pediatric studies show a
395 significant difference for rs8069176 at the *GSDMB* locus (P_{het}=7.4x10^{-4}) but no difference for
396 rs2952156 at the *ERBB2* locus (P_{het}=0.11). These two SNPs are only in moderate LD (r^2=0.30)
397 and, interestingly, each is in strong LD (r^2>0.9) with missense variants localized in *ERBB2* for
398 the proxy of rs2952156 and in *ZPBP2* and *GSDMB* for the proxies of rs8069176. Moreover, both
399 rs2952156 and rs8069176 are associated with *GSDMB* and *ORMDL3* expression in blood e-QTL
400 databases[10-13] and with expression of *GSDMA*, *CDK12*, *GSDMB*, *ORMDL3* in whole lung
401 tissue[12,14]. However, only rs2952156 is associated with *PGAP3* transcript in lung[12,14]
402 (Supplementary Table 12a). Further exploration of eQTL data from GTEx[12] indicated that
403 rs8069176 accounts for a large part of the association of the most significant SNP with *ORMDL3*
404 transcript in blood while rs2952156 accounts for a large part of the association of the most
405 significant SNP with *PGAP3* transcript in lung (Supplementary Table 12b), suggesting that the

asthma-associated signals near *PGAP3*/*ERBB2* and *ORMDL3*/*GSDMB* blocks may affect asthma risk through the expression of different genes in different tissues..

Finally, of the 21 published asthma loci, 12 did not reach genome-wide significance in TAGC (Supplementary Table 13). The most significant SNPs in the GWAS catalog[3] at seven of those loci had P-values>0.01 in TAGC analyses. Among these seven non-replicated loci, two (4q31.21[7], 8q24.11[15]) were reported in Japanese, three (4q12, 9p23, 10q24.2)[16] had SNPs with low minor allele frequencies (≤ 2%) and were reported in a childhood-onset asthma study, and two (1q31.3[17], 5q12.1[18]) were reported in European-ancestry children with asthma defined by current or persistent asthma symptoms with regular use of medication. The most significant SNPs at the remaining five loci had P-values $\leq 5 \times 10^{-4}$ in at least one TAGC meta-analysis, thus providing some replication. Amongst these five loci, the 1q23.1 locus is specific to African-ancestry populations[19]; the 12q13.2 SNP, reported by a Japanese study[7], showed heterogeneity in the TAGC Japanese meta-analysis as well as in European-ancestry and multi-ancestry meta-analyses (P_{het} ≤0.05); and the 7q22.3 SNP, reported in European-ancestry populations[20], was associated with a severe form of childhood asthma and also showed heterogeneity across studies in the original publication[20] (where the P_{random} value did not reach significance) as well as in our study (European-ancestry, multi-ancestry and pediatric meta-analyses, $0.006 \leq P_{het} \leq 0.03$). Finally, SNPs at 1q21.3 and 22q12.3 loci, previously reported in European-ancestry populations[21,22], did not show significant evidence for heterogeneity across TAGC studies in the European-ancestry and multi-ancestry meta-analyses ($0.11 \leq P_{het} \leq 0.19$). When we repeated these two meta-analyses, under a fixed-effects model and considering separately the set of TAGC datasets that were part of the original publication (set P) and the set of remaining TAGC datasets (set R), both 1q21.3 and 22q12.3 SNPs had higher effect sizes in set P than in set R. These differences in effect sizes did not reach significance for the 1q21.3 SNP (P_{het} for Cochran's Q test are 0.13 and 0.20 in European-

ancestry and multi-ancestry analyses) and were borderline significant for the 22q12.3 SNP ($0.04 \leq P_{het} \leq 0.06$) (Supplementary Table 14). Altogether, these results suggest that the lack of replication is mainly due to heterogeneity that is attributable to various factors, such as ethnicity, specificity of clinical phenotypes or other factors as further discussed.

To investigate whether the 18 asthma loci identified in this study contain multiple distinct signals, we performed approximate conditional regression analysis based on summary statistics for all loci (Online Methods), except for the 9p24.1 region which showed heterogeneity in SNP effect size across studies over the whole locus. For the 17q12-21 locus, this analysis was restricted to the pediatric sub-group in which there was no heterogeneity. .After conditioning on the lead SNP in each investigated region, four secondary signals (2q12, 5q22.1, 5q31, 6p21.32) remained significant ($P_{fixed} \leq 5x10^{-8}$) (Supplementary Table 15), yielding 22 distinct genome-wide significant signals.

To provide biological insight into our findings, we conducted a comprehensive bioinformatic assessment of the asthma association signals. To pinpoint the most likely candidate genes at the nine loci harboring novel associations with asthma *per se*, we interrogated results of six eQTL studies in tissues relevant to asthma, blood (including peripheral blood[11,12], lymphoblastoid cell lines (LCLs)[10,13], monocytes[23]) and whole lung tissue[12,14], and searched for missense variants potentially tagged by the association signals. To assess the level of overlap of asthma associations with susceptibility loci for other phenotypes, we interrogated the GWAS catalog[3] while varying the strength of association with asthma (thresholds from $5x10^{-8}$ to 10^{-3}). To get greater insight into how asthma associated variants may functionally influence disease, we interrogated the ROADMAP/ENCODE functional genomics data generated in a wide range of human cell types[24]. Finally, the degree of connectivity between the asthma-associated loci was assessed through text mining[25]. Results are described below.

Candidate genes at the nine loci showing novel associations

A summary of eQTL analysis for these nine loci is described in Table 2 and Supplementary Table 16; regional plots are shown in Supplementary Fig.3a.

New asthma susceptibility loci

Five new loci were identified in this study. The strongest new signal in both the European-ancestry (P_{random}=8.6x10^{-13}) and multi-ancestry (P_{random}=2.2x10^{-12}) meta-analyses was with SNP rs2325291 in an intron of *BACH2* at 6q15, which is strongly correlated with rs10455168 (r^2=0.91), a cis-eQTL altering expression of *BACH2* in blood[11]. *BACH2* encodes a zip transcription factor that regulates nucleic acid-triggered antiviral responses in human cells[26]. The second strongest signal in the European-ancestry and multi-ancestry analyses was with rs17637472 (P_{random}=3.3x10^{-9} and 6.6x10^{-9}), which lies between *ZNF652* and *PHB* at 17q21.33, and is a strong cis-eQTL for *GNGT2* (173 kb from rs17637472) in blood[10,11,13,23]. GNGT2 interacts with beta-arrestin 1 to promote G-protein-dependent Akt signaling for NF-kappaB activation[27].

Among the other new signals, the lead SNP rs1233578 at 6p22.1 (P_{random}=5.3x10^{-9} in European-ancestry populations) resides between *TRIM27* and *GPX5*. This SNP was not associated with gene expression in blood or lung but is in LD (r^2=0.6 in European-ancestry populations) with rs7766356 (312 kb from rs1233578), which is a cis-eQTL for *ZSCAN12* in blood[13] and *ZSCAN31* in lung[14]. These zinc finger protein encoding genes were associated with lung function[28]. The two SNPs, rs1233578 and rs7766356, represent the same association signal in European-ancestry populations (the association with rs7766356 becomes non-significant after conditioning on the lead SNP rs1233578). The 12q13.3 lead SNP (rs167769), which was only significant in the multi-ancestry analysis (P_{random}=3.9x10^{-9}), lies in an intron of *STAT6* and is strongly associated with *STAT6*

478 expression in blood[10,11,13] and lung[14]. STAT6 is a transcription factor, essential for the functional
479 responses of Th2 lymphocytes mediated by IL-4 and IL-13[29]. This result robustly establishes the
480 association of *STAT6* with asthma risk that has been disputed by candidate gene studies[30]. The
481 5q31.3 lead SNP rs7705042 (P_{random}=7.9x10^{-9} in multi-ancestry analysis) is within an intron of
482 *NDFIP1* and associated with *NDFIP1* expression in blood[11-13]. NDFIP1 is a potent inhibitor of
483 antiviral response[31] and inflammation processes[32].

484 ***New asthma signals at loci reported in specific populations***

485 Two associations in our study were with novel SNPs at loci previously reported to be associated
486 with asthma in Latinos[6] and Japanese[7]. The first one at 6p21.33 was reported in an admixture
487 mapping study in Latinos[6]. The lead TAGC SNP, rs2855812 (P_{random}=8.9x10^{-12} in the multi-
488 ancestry analysis, P_{random} =1.7x10^{-8} in the European-ancestry meta-analysis) lies in an intron of
489 *MICB*. This SNP was not correlated (r^2=0) with any of the SNPs reported in the Latino study[6]. The
490 6p21.33 region harbors many genes whose transcripts are associated with TAGC asthma signals,
491 including *TNF*, *LST1*, *HLA-C, LTA* in blood[10,11,13] and *MICB* in lung[12,14]. These genes are involved
492 in immune-related mechanisms. This 6p21.33 locus is about 600 kb apart from the previously
493 reported 6p21.32 locus that spans HLA-Class II genes. Intensive sequencing efforts will be needed
494 to further clarify the HLA region associations. The second association was at 10p14 locus where
495 a GWAS in Japanese[7] reported association (lead SNP rs10508372) with adult asthma. We detected
496 a new signal, rs2589561, in European-ancestry (P_{random}=1.4x10^{-8}) and multi-ancestry meta-
497 analyses (P_{random}=3.5×10^{-9}), that is not correlated with rs10508372 in either European-ancestry or
498 Japanese populations. The SNP rs2589561 is in a gene desert, 929 kb distal of *GATA3*. However,
499 recently published promoter capture Hi-C data in hematopoietic cells[33] revealed that two proxies
500 of rs2589561 (r^2>0.9) lie in a region that interacts with the *GATA3* promoter, especially in CD4+T
501 cells. This suggests that the SNP may be in a distal regulator of *GATA3* , which encodes a

transcription factor that is a master regulator of differentiation of Th2 cells and innate lymphoid cells type 2[34].

Asthma signals reported for asthma plus hay fever

Loci on chromosomes 8q21.13 and 16p13.13 were previously associated with asthma plus hay fever but not with asthma alone in one European-ancestry study[8]. In our results, the 8q21.13 lead SNP rs12543811 (P_{random}=3.4x10^{-8} and 1.1x10^{-10} in the European-ancestry and multi-ancestry analyses) lies between *TPD52* and *ZBTB10* and is in strong LD (r^2=0.79) with the previously reported asthma/hay fever SNP rs7009110. These two SNPs represent the same signal, as the association with rs12543811 becomes non-significant after conditioning on rs7009110. Thus, the 8q21.13 locus is likely implicated in allergic asthma. A functional analysis of the asthma/hay fever locus pinpointed *PAG1* as a promising candidate[35]. The chromosome 16p13.13 SNP rs17806299 is within an intron of *CLEC16A* (P_{random}=2.1x10^{-10} and 2.7x10^{-10} in European-ancestry and multi-ancestry meta-analyses). Although in moderate LD (r^2=0.66) with the previously reported asthma/hay fever signal (rs62026376)[8], the association of asthma with rs17806299 was removed after conditioning on rs12935657 (r^2=0.96 with rs62026376), indicating that these SNPs represent the same signal and 16p13.13 is probably also an allergic asthma locus. The SNP rs17806299 is strongly associated with the expression of a nearby gene, *DEXI* in blood[11,23]. Similar observations of associations of *CLEC16A* SNPs with autoimmune diseases and expression of *DEXI* together with chromosome conformation capture experiments implicated *DEXI* as the most likely candidate gene for autoimmune diseases[36]. The potential relevance of *DEXI* in allergic diseases has also been previously discussed[8].

It is notable that the lead SNPs at the nine new asthma-associated loci lie in non-coding regions and are not tagging missense variants.

Overlap of loci associated with asthma and other phenotypes

We next explored whether the nine loci that harbored new signals for asthma *per se* overlapped with GWAS loci reported for allergy-related phenotypes, lung function phenotypes, or other immune-related diseases using the GWAS catalog[3]. Six of these nine asthma loci showed overlapping associations with allergy-related phenotypes and eight of them with auto-immune diseases or infection-related phenotypes (Table 2). Moreover, three asthma loci overlapped with associations with lung function phenotypes.

We expanded our search of overlap between the asthma association signals having multi-ancestry $P_{random}<10^{-4}$ in this study and GWAS signals with all phenotypes and diseases in the GWAS catalog[3]. We examined 4,231 unique trait-loci combinations (Online Methods), and used the disease classification from Wang *et al*.[37] to group traits. We summarized the overlap with GWAS catalog signals as the proportion of catalog SNPs having asthma *P*-values$<10^{-4}$ in our analysis. This revealed significant overlap with autoimmune disease (10%, i.e. 49 out of 480 catalog SNPs show evidence for asthma association), consistent with the hypothesized shared susceptibility[38,39], moderate overlap with diseases having an inflammatory component (cardiovascular diseases, cancers, neuro-psychiatric diseases), and small to no overlap with other diseases (Table 3). When investigating specific diseases and traits (Supplementary Table 17), the most significant overlap is with allergic phenotypes. There is little to no overlap with other phenotypes that appear most frequent in the GWAS catalog (for example, no shared associations with type 2 diabetes).

When we broadened our analysis to a larger set of SNPs in the GWAS catalog to identify loci for diseases with potentially shared genetic architecture with asthma (i.e, SNPs associated with asthma at $P_{random}\leq10^{-3}$ in our multi-ancestry meta-analysis), additional pleiotropic signals emerged (Supplementary Table 18). This larger set of associations suggests a broader picture of asthma

risk, with a wide range of pleiotropic effects for traits ranging from lung cancer and multiple sclerosis (with rs3817963 in *BTNL2*) to coronary heart disease (with rs1333042 near *CDKN2B*). This analysis also generated an extended set of asthma candidate genes. Indeed, there are 210 SNPs in the GWAS catalog that are associated with asthma in TAGC at a threshold of 10^{-3}; the proportion of false positives among these is smaller than 1%.

Enrichment of asthma risk loci in epigenetic marks

Because nearly all lead SNPs at the 18 loci identified by this study lie in non-coding sequences, except for the *IL13* missense variant (rs20541), we investigated whether the asthma-associated variants and their proxies ($r^2 \geq 0.80$) were concentrated in cis-regulatory DNA elements. Only 16 of 18 identified asthma loci were explored because we excluded the two loci spanning the HLA region due to the large amount of variability and extensive LD in this region. We interrogated the 111 ROADMAP and 16 ENCODE reference epigenomes in a wide range of human cell types[24]. We focused on histone marks characterizing enhancers and promoters assayed in all 127 epigenomes and DNase I hypersensitivity sites available in 51 cell types. To assess enrichment of the asthma risk variants for co-localization with these regulatory elements, we used the Uncovering Enrichment through Simulation pipeline[40]. This approach generates random SNP sets that match the characteristics of the original asthma-associated SNPs (distance from the nearest transcription start site, number of LD partners, minor allele frequency). Empirical P-values for enrichment are calculated by comparing the observed frequency of co-localization of SNPs with a given type of regulatory element in the original asthma-associated SNP set to the co-localization frequency distribution obtained from the 10,000 random SNP sets generated. Benjamini-Hochberg false discovery rates (FDRs) are then computed to correct for multiple testing (Online Methods).

While the asthma-associated variants were strongly enriched for co-localization with enhancer marks, there was only weak enrichment in promoter marks (Table 4 and Supplementary Table 19). This enrichment was highest in leukocytes (27 leukocytes of which 19 (70%) are lymphocytes and monocytes). For example, a FDR≤5% for enrichment of asthma loci in active enhancers was observed in 100% of leukocytes compared to 50% of all cell types. The enrichment of asthma risk variants for co-localization with DNAse I hypersensitivity sites was intermediate between the enrichments in promoters and enhancers and was again increased in blood cells (FDR≤5% in 40% of leukocytes and 12% of all cell types) (Table 4 and Supplementary Table 20).

The strong enrichment of asthma loci in enhancer marks, especially in immune cells, indicates that the associated genetic variants are likely involved in regulation of immune-related functions. This also suggests that epigenetic mechanisms may be key to promoting asthma, as evidenced for IgE levels, an asthma-associated phenotype[41].

Connectivity between asthma-associated loci

To characterize the degree of connectivity between the 18 asthma-associated loci, we applied the Gene Relationships Across Implicated Loci (GRAIL) text-mining approach[25]. Genes at eleven of these loci showed connections with a GRAIL score, P_{GRAIL}, less than 5% (7 of them being highly connected with $P_{GRAIL}<10^{-3}$) (Fig.2 and Supplementary Table 21). These genes were connected by keywords such as 'asthma', 'allergy', 'atopic', 'interleukin', 'cytokines', 'airway', and 'inflammation', confirming the central role of immune-related mechanisms accounting for these connections.

DISCUSSION

In this meta-analysis of worldwide asthma GWAS in ethnically-diverse subjects, we discovered nine novel loci influencing asthma risk. This study confirms that immune-related mechanisms are prominent in asthma susceptibility and brings novel insights that open new routes for future asthma research. The asthma-associated loci identified by TAGC are enriched in enhancer marks and are likely to be involved in gene regulation. Although this was observed in immune cells, asthma genes (e.g., *IL1RL1, TSLP, IL33, ORMDL3/GSDMB*) are also expressed in the airway epithelium where they modulate airway inflammation. Investigation of epigenetic marks in airway epithelial cells may bring additional insight. The best candidates at many loci are involved in immune response to viruses or bacteria, which underlines the importance of infections in asthma risk. This study further provides evidence for overlap of asthma loci with loci underlying auto-immune diseases and other diseases that have an inflammatory component, which strengthens the growing importance of pleiotropy in multifactorial diseases.

Our meta-analysis doubles the number of cases from prior genome-wide studies[21,22]. We identified 878 SNPs that correspond to 22 distinct association signals at 18 loci meeting criteria for genome-wide significance in European-ancestry and/or multi-ancestry populations. Pooling data from ethnically-diverse populations can increase power to detect new loci (two loci reached the genome-wide threshold only in the multi-ancestry analysis) but may also increase heterogeneity. Besides differences in the genetic background, varying environmental exposures can modify genetic risks and result in heterogeneity in SNP effect size, and consequently reduce power of multi-ancestry analysis compared to ancestry-specific analysis. Assuming an asthma prevalence of 10%, the variance in liability to asthma explained by the 22 genome-wide significant variants of this study was estimated to be 3.5% (95% Confidence Interval : 2.0%-5.4%) of which 72% was accounted for by the known loci and 28% by the new loci. It is of note that the current study was

621 based on HapMap2-imputed data which was shared within the TAGC consortium and thus allowed
622 detection of associations with common genetic variants (MAF≥1%).

623 The overall relative paucity of asthma risk loci detected by large-scale GWAS compared to
624 other common diseases may be due to the clinical heterogeneity of asthma and the important role
625 of differing environmental exposures. It is recognized that asthma is not a single disease but that
626 the syndrome varies based on many characteristics[42], including age of asthma onset, the severity
627 of disease, the type of cellular inflammatory infiltrates, occupational exposures and the varying
628 response to treatment.. It is thus possible that additional asthma loci will be revealed by studies
629 targeting more specific asthma sub-phenotypes and/or taking into account environmental
630 exposures.

631 In conclusion, future discoveries might come by exploring more complex models of asthma
632 phenotypes and through the joint analysis of asthma and other immune-mediated and inflammatory
633 diseases. The central role of gene regulatory mechanisms highlighted by our study might prompt
634 genome-wide explorations of the epigenome in immune cells and the respiratory epithelium while
635 integrating information on genetic variation and environmental exposure histories.

636

637 **URLs**

638 National Human Genome Research Institute (NHGRI) and European Bioinformatics Institute
639 (EBI) catalog of published genome-wide association, https://www.ebi.ac.uk/gwas/ ; 1000
640 Genomes Project Consortium Phase 3, http://www.internationalgenome.org/; Genome-wide
641 Complex Trait Analysis (GCTA), http://cnsgenomics.com/software/gcta ; Blood-eQTL database,
642 https://omictools.com/blood-eqtl-browser-tool ; GTEx, http://www.gtexportal.org/ ; MuTHER
643 database, http://www.muther.ac.uk/ ; eQTL database in lymphoblastoïd cell lines from MRCA
644 and MRCE families, https://www.hsph.harvard.edu/liming-liang/software/eqtl/ ; GHS_Express,

http://genecanvas.ecgene.net/ ; HaploReg v4.1, http://archive.broadinstitute.org/mammals/haploreg/haploreg.php; ROADMAP and ENDCODE epigenomics data, http://egg2.wustl.edu/roadmap/web_portal/ ; Uncovering Enrichment through Simulation (UES) pipeline, https://github.com/JamesHayes/uesEnrichment ; GRAIL, https://software.broadinstitute.org/mpg/grail/ ; VIZGRAIL, http://software.broadinstitute.org/mpg/grail/vizgrail.html, LocusZoom, http://locuszoom.org/

ACKNOWLEDGMENTS

We thank all participants who provided data for each study and to our valued colleagues who contributed to data collection and phenotypic characterization of clinical samples, genotyping and analysis of individual datasets. Detailed acknowledgements and funding for individual studies can be found in the Supplementary Note.

AUTHOR CONTRIBUTIONS

TAGC study management: F.D., K.C.B., W.O.C.C., M.F.M., C.O., D.L.N.
F.D. and D.L.N. designed the study and wrote the manuscript. F.D., D.L.N., P.M-J designed and conducted the statistical analysis. K.C.B., W.O.C.C., M.F.M., C.O. designed the study and wrote the manuscript. M.B., A.V., S.L., H.M. carried out the quality control of the data and performed statistical analysis.

AAGC (Australia): Study PIs: M.A.F., M.C.M., C.F.R., P.J.T.; Data collection or analysis: M.A.F., M.C.M., C.F.R., G.J., P.J.T.

ALLERGEN CAPPS and SAGE: Study PIs: A.B.B., M.C-Y., D.D., A.L.K.; Data collection or analysis: D.D., J.E.P.; Study Phenotyping: A.B.B., M.C-Y.

SLSJ: Study PIs: C.L., T.J.H.; Study design and management: C.L.

APCAT Consortium: Study PIs: J.N.H., M.R.J., V.S.; **FHS**: Study PI: G.T.O.; Data collection or analysis: S.V., Z.G.; **EPIC-Norfolk**: Study PI: N.J.W.; Data collection or analysis: J.H.Z., R.S.; **NFBC1966**: Study PI: M.R.J.; Data collection or analysis: A.C.A., A.R.; **FINRISK**: Study PI: V.S.; Data collection or analysis: M.K., T.L.; **H2000**: Study PIs: M.H., P.J.; Data collection or analysis: M.K., T.H.; **HBCS**: Study PIs J.G.E.; Data collection or analysis: E.W., A.P.; **YFS**: Study PI: O.T.R.; Data collection or analysis: T.L., M.K.

CARe Consortium: Study PIs: J.N.H., S.S.R.; Data collection or analysis: C.D.P, D.B.K, L.J.S., R.K., K.M.B., W.B.W.

MESA: Study PIs: R.G.B, S.S.R; Data collection or analysis: K.M.D, A.M.

ARIC: Study PI: S.J.L.; Data collection or analysis: S.J.L., L.R.L.

CHS: Study PIs: S.A.G., S.R.H.; Data collection or analysis: G.L., S.A.G., S.R.H.

deCode Genetics: Study PIs: K.S., I.J., D.F.G., U.T., G.T.; Data collection or analysis: I.J., D.F.G., G.T.; Study Phenotyping: U.S.B.

EAGLE Consortium: PI: H.B. **COPSAC**: Study PIs: H.B, K.B.; Data analysis: E.K., J.W.; Study Phenotyping: K.B.; **DNBC:** Study PI: M.M.; Data collection or analysis: B.F and F.G.; **GENERATION R:** Study PI: J.C.dJ.; Data collection or analysis: R.J.P.vdV., L.D. and V.W.V.J. **GINIplus/LISAplus:** Study PI: J.H.; Genotyping, data collection or analysis: M.S., C.M.T.T.; Study Phenotyping: J.H.; **MAAS:** Study PIs: A.S., AC.; Data collection or analysis: J.A.C.; **RAINE:** Study PI: P.H.; Data collection or analysis: W.A.,C.E.P.

B58C: PI and Statistical Analysis: D.P.S.

EVE: Study-PIs: C.O., D.L.N., K.C.B., E.B., E.B., J.G., F.G., S.J.L., F.J.M., D.M., I.R., S.T.W., L.K.W., B.A.R.; Data collection or analysis: D.L.N., J.G., S.J.L., D.M., D.G.T., B.A.R., B.E.H., P.E.G., M.T.S., C.E., B.E.D., J.J.Y., A.M.L., R.A.M., R.A.M., T.H.B.

JAARC: Study PI: T.T.; Data collection or analysis: T.T., A.T., M.K.

JPAC: Study PI: E.N.; Data collection or analysis: H.H., K.M.

GABRIEL CONSORTIUM: Study PIs: W.O.C.C., E.V.M.; Genotyping: M.L.; Data Analysis: E.B., F.D., M.F., D.P.S. **EGEA:** Study PIs: V.S., F.D.; Genotyping, data collection or analysis: M.L., E.B.; **ALSPAC:** Study PI: J.H.; Genotyping, data collection or analysis: W.L.M., R.G.; Study phenotyping: J.H.; **ECRHS:** Study PI: D.J.; Data collection or analysis: C.J., J.H.; **BAMSE:** Study PIs: EM, M.W., G.P.; **BUSSELTON:** Study PIs: A.W.M., A.J., J.B.; Genotyping, data collection or analysis: A.W.M., A.J., J.H., J.B.; **GABRIEL Advanced Studies;** Study PI: E.V.M.; Data collection or analysis: M.K., J.G.; **KSMU:** Study PI: A.P.; Data collection or analysis: M.S., V.I.; **MRCA-UKC**: Study PIs: W.O.C.C., M.M.; Data collection or analysis: L.L.; **MAGICS:** Study PI: MK; Data collection or analysis: A.V.B., S.M.; **MAS:** Study PI: Y.L.; Data collection or analysis: S.L., I.M.; **PIAMA:** Study PI: G.H.K., D.S.P.; Data collection or analysis: G.H.K., D.S.P., U.G.; **SAPALDIA:** Study PI: N.P.H.; Data collection or analysis: M.I., A.K. **TOMSK:** Study PIs: L.M.O., V.P.P.; Data collection or analysis: M.B.F., P.A.S.; **UFA:** Study PI: E.K.; Data collection or analysis: A.S.K., Y.F. **INDUSTRIAL:** Study PIs: D.H., T.S.; Data collection or analysis: I.M.W., V.S.; **SEVERE**: Study PIs: A.B., K.F.C., C.E.B.

NTR: Study PI: D.I.B.; Genotyping, data collection or analysis: J.J.H., H.M., G.W.

Rotterdam: Study PIs: A.H., B.H.S., G.G.B.; Genotyping, data collection or analysis: G.G.B., B.H.S., D.W.L., L.L., A.G.U.

DAGC: Study PIs: G.H.K., D.S.P.; Genotyping, data collection or analysis: G.H.K., D.S.P., J.A., M.A.E.N., J.M.V.

All authors provided critical review of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors who are affiliated with deCODE (D.F.G., I.J., K.S., U.T., G.T.) are employees of deCODE genetics/AMGEN. All other co-authors did not declare any conflict of interest

REFERENCES

1. Akinbami, L.J. *et al.* Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief*, 1-8 (2012).
2. Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* **142**, 1351-8 (1990).
3. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-6 (2014).
4. Igartua, C. *et al.* Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nat Commun* **6**, 5965 (2015).
5. Bouzigon, E. *et al.* Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* **359**, 1985-94 (2008).
6. Galanter, J.M. *et al.* Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. *J Allergy Clin Immunol* **134**, 295-305 (2014).
7. Hirota, T. *et al.* Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* **43**, 893-6 (2011).
8. Ferreira, M.A. *et al.* Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J Allergy Clin Immunol* **133**, 1564-71 (2014).
9. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
10. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-9 (2012).
11. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
12. The GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
13. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res* **23**, 716-26 (2013).
14. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).

15. Noguchi, E. *et al.* Genome-wide association study identifies *HLA-DP* as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* **7**, e1002170 (2011).
16. Ding, L. *et al.* Rank-based genome-wide analysis reveals the association of ryanodine receptor-2 gene variants with childhood asthma among human populations. *Hum Genomics* **7**, 16 (2013).
17. Sleiman, P.M. *et al.* Variants of *DENND1B* associated with asthma in children. *N Engl J Med* **362**, 36-44 (2010).
18. Himes, B.E. *et al.* Genome-wide association analysis identifies *PDE4D* as an asthma-susceptibility gene. *Am J Hum Genet* **84**, 581-93 (2009).
19. Torgerson, D.G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-92 (2011).
20. Bonnelykke, K. *et al.* A genome-wide association study identifies *CDHR3* as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* **46**, 51-5 (2014).
21. Ferreira, M.A. *et al.* Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* **378**, 1006-14 (2011).
22. Moffatt, M.F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-21 (2010).
23. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
24. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
25. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* **5**, e1000534 (2009).
26. Hong, S.W., Kim, S. & Lee, D.K. The role of Bach2 in nucleic acid-triggered antiviral innate immune responses. *Biochem Biophys Res Commun* **365**, 426-32 (2008).
27. Yang, M., He, R.L., Benovic, J.L. & Ye, R.D. Beta-Arrestin1 interacts with the G-protein subunits beta1gamma2 and promotes beta1gamma2-dependent Akt signalling for NF-kappaB activation. *Biochem J* **417**, 287-96 (2009).
28. Artigas, M.S.L., D.W.; Wain, L.V.; Gharib, S.A.; Obeidat, M.; Tang, W.; et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* (2011).
29. Goenka, S. & Kaplan, M.H. Transcriptional regulation by *STAT6*. *Immunol Res* **50**, 87-96 (2011).
30. Qian, X., Gao, Y., Ye, X. & Lu, M. Association of *STAT6* variants with asthma risk: a systematic review and meta-analysis. *Hum Immunol* **75**, 847-53 (2014).
31. Wang, Y., Tong, X. & Ye, X. Ndfip1 negatively regulates RIG-I-dependent immune signaling by enhancing E3 ligase Smurf1-mediated MAVS degradation. *J Immunol* **189**, 5304-13 (2012).
32. Venuprasad, K., Zeng, M., Baughan, S.L. & Massoumi, R. Multifaceted role of the ubiquitin ligase Itch in immune regulation. *Immunol Cell Biol* **93**, 452-60 (2015).
33. Javierre, B.M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369-1384 e19 (2016).
34. Barnes, P.J. Pathophysiology of allergic inflammation. *Immunol Rev* **242**, 31-50 (2011).
35. Vicente, C.T. *et al.* Long-Range Modulation of *PAG1* Expression by 8q21 allergy risk variants. *Am J Hum Genet* **97**, 329-36 (2015).
36. Davison, L.J. *et al.* Long-range DNA looping and gene expression analyses identify *DEXI* as an autoimmune disease candidate gene. *Hum Mol Genet* **21**, 322-33 (2012).

37. Wang, L. *et al.* CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation. *Genome Biol* **16**, 190 (2015).
38. Rottem, M. & Shoenfeld, Y. Asthma as a paradigm for autoimmune disease. *Int Arch Allergy Immunol* **132**, 210-4 (2003).
39. Li, X. *et al.* Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. *J Allergy Clin Immunol* **130**, 861-8 e7 (2012).
40. Hayes, J.E. *et al.* Tissue-Specific enrichment of lymphoma risk loci in regulatory elements. *PLoS One* **10**, e0139360 (2015).
41. Liang, L. *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670-4 (2015).
42. Wenzel, S.E. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* **18**, 716-25 (2012).

Figure Legends

Figure 1. Manhattan plots of the results of European-ancestry and multi-ancestry random-effects meta-analyses of asthma risk. (**a**) The European-ancestry meta-analysis pertains to 19,954 cases and 107,715 controls. (**b**) The multi-ancestry meta-analysis pertains to 23,948 cases and 118,538 controls. Each locus is annotated by its cytogenetic band location. The x axis represents chromosomal location and the y axis represents $-\log_{10}$ (*P* value) for tests of association between SNPs and asthma. Black, previously known loci; red, new loci identified in the European-ancestry meta-analysis; blue, additional new loci identified in the multi-ancestry meta-analysis The dashed horizontal line denotes $P=5\times10^{-8}$.

Figure 2. GRAIL[25] circle plot of connectivity between genes at asthma risk loci. The 17 asthma risk loci are along the outer ring (the 10p14 locus was ignored because it corresponds to a gene desert); the internal ring represents the genes at these loci. The width of the lines drawn between genes corresponds to the strength of the literature-based connectivity, with thicker lines representing stronger connections.

.

Figure 1.

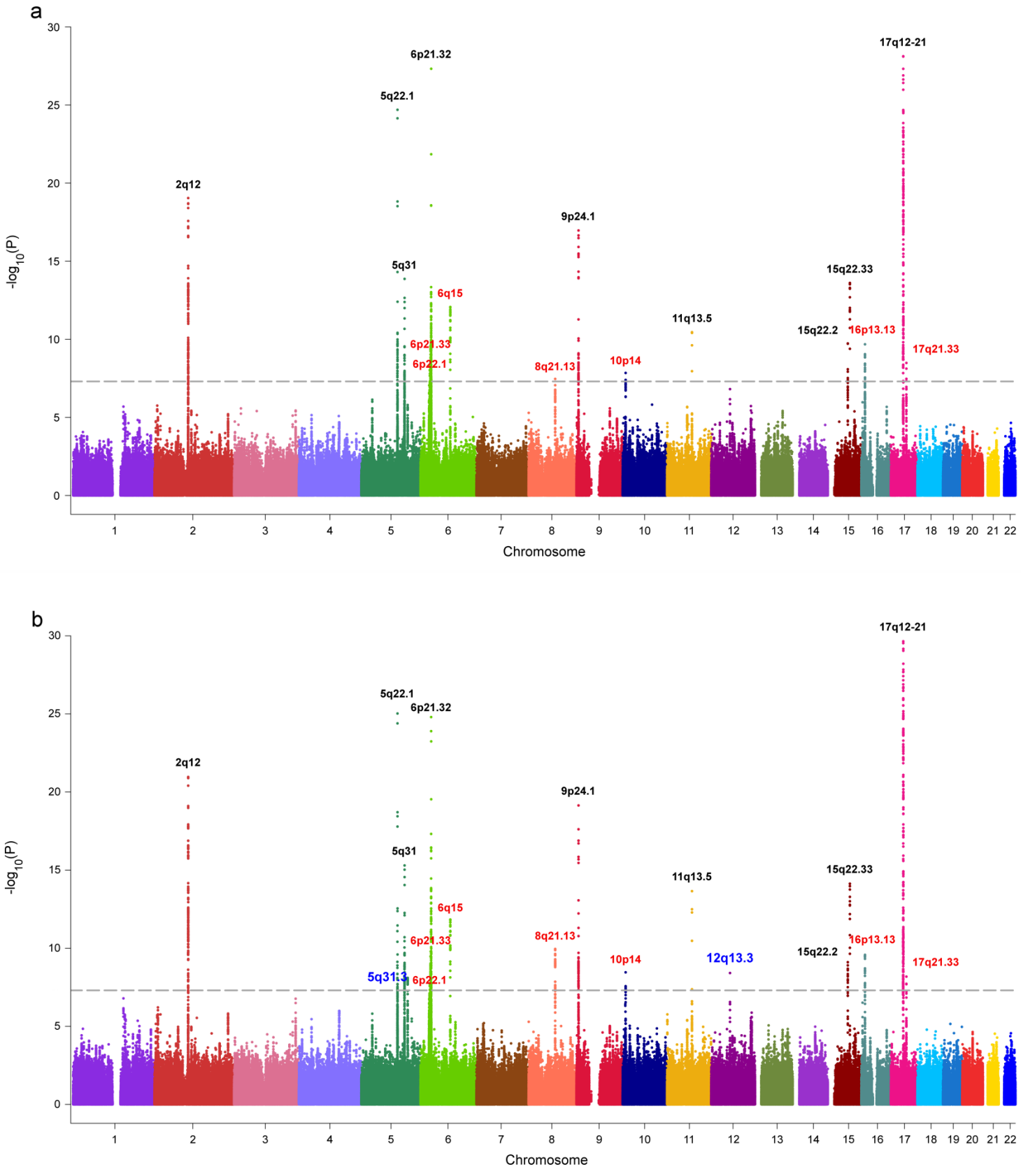
a
-log10(P)
Chromosome
2q12
6p21.32
5q22.1
9p24.1
5q31
6q15
6p21.33
6p22.1
8q21.13
10p14
11q13.5
15q22.33
15q22.2
16p13.13
17q12-21
17q21.33
b
-log10(P)
Chromosome
2q12
5q22.1
6p21.32
9p24.1
5q31
6q15
6p21.33
5q31.3
6p22.1
8q21.13
10p14
11q13.5
12q13.3
15q22.33
15q22.2
16p13.13
17q12-21
17q21.33

Figure 2.

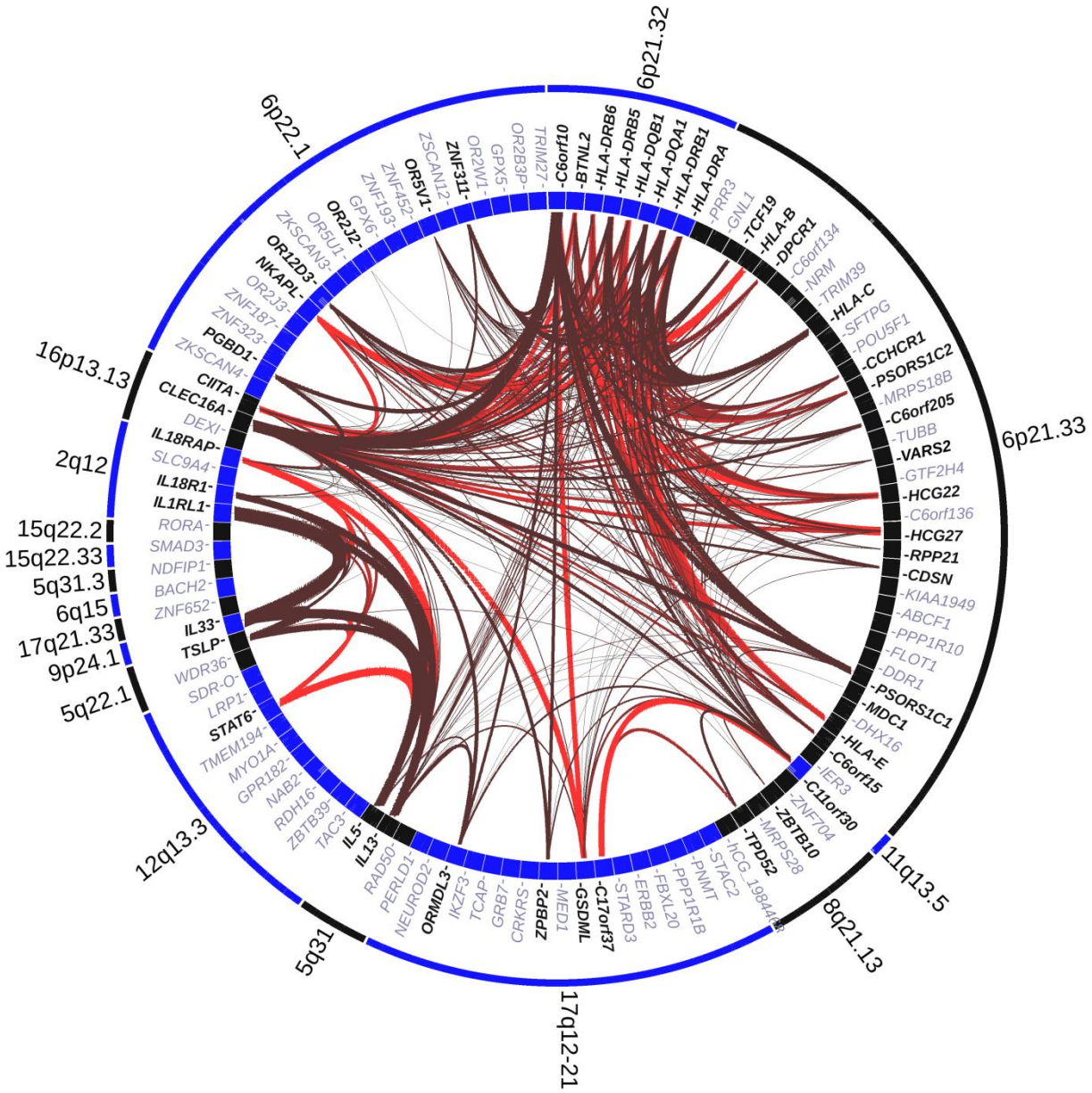
6p21.32
6p21.33
11q13.5
8q21.13
17q12-21
5q31
12q13.3
5q22.1
9p24.1
17q21.33
6q15
5q31.3
15q22.33
15q22.2
2q12
16p13.13
6p22.1
C6orf10
BTNL2
HLA-DRB6
HLA-DRB5
HLA-DQB1
HLA-DQA1
HLA-DRB1
HLA-DRA
PRR3
GNL1
TCF19
HLA-B
DPCR1
C6orf134
NRM
TRIM39
HLA-C
SFTPG
POU5F1
CCHCR1
PSORS1C2
MRPS18B
C6orf205
TUBB
VARS2
GTF2H4
HCG22
C6orf136
HCG27
RPP21
CDSN
KIAA1949
ABCF1
PPP1R10
FLOT1
DDR1
PSORS1C1
MDC1
DHX16
HLA-E
C6orf15
IER3
C11orf30
ZNF704
ZBTB10
MRPS28
TPD52
hCG_1984
STAC2
PNMT
PPP1R1B
FBXL20
ERBB2
STARD3
GSDML
C17orf37
MED1
ZPBP2
CRKRS
GRB7
TCAP
IKZF3
ORMDL3
NEUROD2
PERLD1
RAD50
IL13
IL5
TAC3
ZBTB39
RDH16
NAB2
GPR182
MYO1A
TMEM194
STAT6
LRP1
SDR-O
WDR36
TSLP
IL33
ZNF652
BACH2
NDFIP1
SMAD3
RORA
IL1RL1
IL18R1
SLC9A4
IL18RAP
DEXI
CLEC16A
CIITA
ZKSCAN4
PGBD1
ZNF323
ZNF187
OR2J3
NKAPL
OR12D3
ZKSCAN3
OR5U1
OR2J2
GPX6
ZNF193
ZNF452
ZSCAN12
OR5V1
ZNF311
OR2W1
GPX5
OR2B3P
TRIM27

Table 1 Genetic loci associated with asthma in European-ancestry and multi-ancestry meta-analyses

						European-ancestry meta-analysis					Multi-ancestry meta-analysis				
Locus[a]	No. Sig SNPs Eur-anc/ Multi-anc[b]	rsID	Position[c]	Nearby genes[d]	Allele[e]	EAF	OR	95% CI	P_{random}	P_{het}	OR	95% CI	P_{random}	P_{het}	P_{ethnic}
New asthma susceptibility loci															
5q31.3	0 / 11	rs7705042	141,492,419	*NDFIP1,GNDPA1,SPRY4*	C/A	0.63	1.08	1.05-1.11	1.6×10^{-6}	0.07	1.09	1.06-1.12	7.9×10^{-9}	0.11	0.55
6p22.1	8 / 5	rs1233578	28,712,247	*GPX5, TRIM27*	A/G	0.13	1.11	1.07-1.15	5.3×10^{-9}	0.82	1.09	1.05-1.12	5.9×10^{-7}	0.56	0.003
6q15	26 / 26	rs2325291	90,986,686	*BACH2,GJA10,MAP3K7*	G/A	0.33	0.91	0.89-0.93	8.6×10^{-13}	0.78	0.91	0.89-0.94	2.2×10^{-12}	0.80	0.39
12q13.3	0 / 1	rs167769	57,503,775	*STAT6,NAB2,LRP1*	C/T	0.40	1.08	1.05-1.11	1.6×10^{-7}	0.19	1.08	1.05-1.11	3.9×10^{-9}	0.31	0.85
17q21.33	4 / 3	rs17637472	47,461,433	*ZNF652,PHB*	G/A	0.39	1.08	1.05-1.11	3.3×10^{-9}	0.56	1.08	1.05-1.11	6.6×10^{-9}	0.35	0.12
New signals at loci previously associated with asthma in ancestry-specific populations															
6p21.33	66 / 53	rs2855812	31,472,720	*MICB, HCP5, MCCD1*	G/T	0.23	1.10	1.06-1.13	1.7×10^{-8}	0.23	1.10	1.07-1.13	8.9×10^{-12}	0.39	0.58
10p14	3 / 6	rs2589561	9,046,645	*GATA3,CELF2*	A/G	0.82	0.90	0.87-0.94	1.4×10^{-8}	0.78	0.91	0.88-0.94	3.5×10^{-9}	0.82	0.25
Asthma signals previously reported for asthma plus hay fever															
8q21.13	1 / 28	rs12543811	81,278,885	*TPD52,ZBTB10*	G/A	0.66	0.93	0.91-0.95	3.4×10^{-8}	0.47	0.92	0.90-0.95	1.1×10^{-10}	0.54	0.24
16p13.13	12 / 13	rs17806299	11,199,980	*CLEC16A,DEXI,SOCS1*	G/A	0.20	0.90	0.88-0.93	2.1×10^{-10}	0.51	0.91	0.88-0.94	2.7×10^{-10}	0.49	0.58
Known asthma loci															
2q12	133 / 144	rs1420101	102,957,716	*IL1RL1,IL1RL2,IL18R1*	C/T	0.37	1.12	1.10-1.15	9.1×10^{-20}	0.63	1.12	1.09-1.15	3.9×10^{-21}	0.61	0.64
5q22.1	35 / 32	rs10455025	110,404,999	*SLC25A46,TSLP*	A/C	0.34	1.15	1.12-1.18	2.0×10^{-25}	0.53	1.15	1.12-1.18	9.4×10^{-26}	0.57	0.27
5q31	33 / 62	rs20541	131,995,964	*IL13,RAD50,IL4*	A/G	0.79	0.89	0.86-0.91	1.4×10^{-14}	0.73	0.89	0.87-0.92	5.0×10^{-16}	0.77	0.62
6p21.32	101 / 124	rs9272346	32,604,372	*HLA-DRB1,HLA-DQA1*	G/A	0.56	1.16	1.13-1.19	4.8×10^{-28}	0.46	1.16	1.12-1.19	5.7×10^{-24}	0.14	0.43
9p24.1	65 / 71	rs992969	6,209,697	*RANBP6,IL33*	A/G	0.75	0.85	0.82-0.88	1.1×10^{-17}	0.008	0.86	0.83-0.88	7.2×10^{-20}	0.02	0.57
11q13.5	4 / 5	rs7927894	76,301,316	*C11orf30, LRRC32*	C/T	0.37	1.10	1.07-1.13	3.5×10^{-11}	0.38	1.10	1.08-1.13	2.2×10^{-14}	0.56	0.47
15q22.2	9 / 14	rs11071558	61,069,421	*RORA,NARG2,VPS13C*	A/G	0.14	0.89	0.85-0.92	1.9×10^{-10}	0.44	0.89	0.86-0.92	1.3×10^{-9}	0.19	0.06
15q22.33	13 / 13	rs2033784	67,449,660	*SMAD3,SMAD6,AAGAB*	A/G	0.30	1.11	1.08-1.14	2.5×10^{-14}	0.75	1.10	1.08-1.13	7.4×10^{-15}	0.76	0.48
17q12-21	160 / 198	rs2952156	37,876,835	*ERBB2,PGAP3,C17orf37*	A/G	0.70	0.86	0.84-0.88	7.6×10^{-29}	0.55	0.87	0.84-0.89	2.2×10^{-30}	0.52	0.35

SNP *P*-values for association with asthma are based on random-effects meta-analysis using Stata. A total of 878 SNPs, belonging to 18 loci, reached genome-wide significance ($P_{random} < 5 \times 10^{-8}$). Each locus, in this table, is represented by the SNP with the strongest evidence for association in the European-ancestry (127,669 subjects) or multi-ancestry meta-analysis (142,486 subjects from European-ancestry, African-ancestry, Japanese and Latino populations). The Cochran's Q test was used to test for heterogeneity in SNP effect sizes across studies (P_{het}) and to test for a difference between the four ancestry-specific summary effects (P_{ethnic}), EAF, effect allele frequency; OR, odds (log-additive) ratio; 95% CI, 95% confidence interval. [a]Cytogenetic band; [b]Number of genome-

wide significant SNPs (P_{random}<5x10^{-8}) at each locus in European-ancestry meta-analysis/multi-ancestry meta-analysis; [c]SNP position, build 37. [d]The gene where eventually the SNP lies is first indicated, followed by the previous gene and next gene; [e]Reference/Effect allele.

Table 2. Main characteristics of the nine loci harboring novel associations with asthma

Locus[a]	Location of lead SNP[b]	Cis-eQTLs in blood (B) and lung tissue (L)	Association with allergy-related and lung function phenotypes	Association with auto-immune diseases and other immune-related traits
New asthma susceptibility loci				
5q31.3	*NDFIP1* (intron)	B: *NDFIP1* ($2.7x10^{-9}$)		IBD
6p22.1	Intergenic	B: *ZSCAN12* ($3.0x10^{-8}$) L: *ZSCAN31* ($6.5x10^{-11}$)	Lung function	
6q15	*BACH2* (intron)	B: *BACH2* ($3.0x10^{-10}$)		MS, T1D, CD, IBD, V, IGG
12q13.3	*STAT6* (intron)	B: *STAT6* ($9.8x10^{-198}$) L: *STAT6* ($3.7x10^{-37}$)	IgE (total, specific) Lung function	Pso, ISP_IFN
17q21.33	Intergenic	B: *GNGT2* ($2.1x10^{-52}$)	Atopic dermatitis	ISP_IL2
New asthma signals at loci previously associated with asthma in ancestry-specific populations				
6p21.33	*MICB* (intron)	B: *TNF* ($4.8x10^{-14}$), *LST1* ($1.0x10^{-13}$), *HLA-C* ($3.2x10^{-13}$), *LTA* ($1.0x10^{-10}$) L: *MICB* ($4.6x10^{-13}$)	IgE (total, specific), Self-reported allergy, Atopic dermatitis, Lung Function	SLE, UC, RA, IBD, BS, GD, SS, AS, Pso, UC, V, WBC, MoC, DS, HIV-1, SJS, HB, HBV, IMN, CD4:CD8 ratio, HIV-1C
10p14	Intergenic	None	Self-reported allergy	RA, ISP_IL1B, ISPV
Asthma signals previously reported for asthma plus hay fever				
8q21.13	Intergenic	None	Atopic dermatitis, Asthma + hay fever Self-reported allergy	RA
16p13.13	*CLEC16A* (intron)	B: *DEXI* ($2.2x10^{-43}$)	Atopic dermatitis, Asthma + hay fever	T1D, PBC, MS, RA, IBD, CD, LEP

At each of the nine loci harboring novel associations with asthma, cis-genes whose expression (e-QTLs) is associated with the lead asthma-associated SNPs (shown in Table 1) or SNPs in LD ($r^2 \geq 0.5$) with the lead SNPs were searched using six eQTL databases from whole blood[11,12], lymphoblastoïd cell lines[10,13], monocytes[23] and lung[12,14]; only genes with the strongest associations (*P*-value $<5x10^{-8}$, as shown in parentheses) are presented here (Supplementary Table 16 for details). Overlap of these nine loci with associations with allergy-related and lung function phenotypes as well as with auto-immune diseases and other immune-related traits was annotated using the GWAS catalog[3] ; IBD=Inflammatory bowel diseases (Crohn's disease), MS=multiple sclerosis, T1D=type 1 diabetes, CD=celiac disease, V=vitiligo, IGG=IgG Glycosylation, Pso=psoriasis, ISP_IFN=Immune Response to Smallpox (secreted IFN-alpha), ISP_IL2 Immune Response to Smallpox (secreted IL2), SLE=Systemic Lupus Erythematosus, UC=Ulcerative colitis, RA=Rheumatoid arthritis, BS=Behçet syndrome, GD=Grave's disease, SS=Systemic sclerosis, AS=Ankylosing spondylitis, WBC=White Blood cell count, MoC=monocyte count, DS=Dengue shock, HIV-1=HIV-1-susceptibility, SJS=Stevens-Johnson syndrome, HB=Hepatitis B infection, HBV=Hepatitis B vaccine response, IMN=Idiopathic membranous nephropathy, CD4:CD8=CD4:CD8 lymphocyte ratio, HIV-1C= HIV-1 control, ISP_IL1B=Immune Response to Smallpox (secreted IL-1 beta), ISPV=Immune response to smallpox vaccine (IL-6), PBC=Primary biliary cirrhosis, LEP=Leprosy. [a]Cytogenetic band; [b]The protein coding genes flanking intergenic SNPs are shown in Table 1.

Table 3. Overlap of TAGC asthma-associated SNPs with GWAS catalog association signals by disease group

Disease Group	Number of GWAS catalog association signals	Number of SNPs associated with asthma at $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis	*P*-value for overlap
Cardiovascular	743	20	7.8×10^{-42}
Body size and morphology	346	2	5.0×10^{-4}
Immune/Autoimmune	480	49	3.0×10^{-129}
Nervous system	242	4	1.4×10^{-8}
Blood	594	10	1.3×10^{-19}
Neuropsychiatric	114	5	1.5×10^{-12}
Cancer	417	7	4.0×10^{-14}
Endocrine system	276	2	4.0×10^{-4}
Digestive system	347	16	1.4×10^{-37}
Eyes	177	2	2.0×10^{-4}
Respiratory system	85	2	3.6×10^{-5}
Infectious disease/Infection	104	2	5.3×10^{-5}
Urinary system	144	1	1.5×10^{-2}
Alcohol, smoking, and illicit substances	30	0	1
Musculoskeletal system	132	0	1

Overlap of TAGC asthma-associated SNPs with association signals of all diseases/traits in the GWAS catalog[3] was investigated for all TAGC SNPs having $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis; diseases from the GWAS catalog were grouped according to the disease classification proposed by Wang *et al*.[37] (note that the "Digestive system" group includes Crohn's Disease, a subtype of Inflammatory Bowel Disease). The significance of overlap was estimated by the binomial tail probability for observing the shown number of TAGC asthma SNPs among the number of SNPs reported in the GWAS catalog for a group of diseases (for example, the probability of observing 20 or more asthma SNPs with $P_{random} \leq 10^{-4}$ among the 743 cardiovascular SNPs is shown in the last column); a conservative Bonferroni adjusted significance threshold for enrichment in shared associations is 0.05/15=0.003 (for the 15 disease groups investigated).

Table 4. Enrichment of asthma risk loci in promoter and enhancer marks and DNase I hypersensitivity sites

Type of regulatory elements	Proportion of all cell types (blood cell types) showing enrichment with a given false discover rate (FDR)	
	FDR ≤ 10%	FDR ≤ 5%
All promoter states	6% (26%)	0
Active promoter states	13% (33%)	0
All enhancer states	57% (100%)	44% (89%)
Active enhancer states	66% (100%)	50% (100%)
DNase I hypersensitivity sites	16% (50%)	12% (40%)

The co-localization of SNPs at asthma risk loci with regulatory elements (promoters, enhancers, DNase I hypersensitivity sites) was assessed at 16 asthma-loci identified by this study (Table 1); the 6p21.33 and 6p21.32 loci that encompass the HLA region were excluded because of the high amount of variability and LD in this region. Enhancer and promoter states were defined using the ChromHMM 15-state model applied to functional data of 127 ROADMAP and ENCODE reference epigenomes in various cell types (including 27 leukocytes)[24]. DNase I hypersensitivity sites were identified in 51 cell types (including 10 leukocytes)[24]. Empirical-*P*-values for enrichment were obtained using 10,000 Monte-Carlo simulations of random sets of SNPs matching the original set of asthma-associated SNPs[40]; Benjamini-Hochberg's FDR was calculated to correct for multiple testing (Online Methods for details).

ONLINE METHODS

GWAS Studies and Data Shared

All 66 genome-wide association studies that form the TAGC consortium are described in the Supplementary Note and summarized in Supplementary Table 1. These studies included 56 studies of European-ancestry (19,954 cases, 107,715 controls), seven studies of African-ancestry (2,149 cases, 6,055 controls), two Japanese studies (1,239 cases, 3,976 controls) and one Latino study (606 cases, 792 controls), making a total of 23,948 cases and 118,538 controls. There were 27 studies including only childhood-onset asthma (defined as asthma diagnosed at or before 16 years of age) which allowed us analyzing separately a pediatric subgroup (8,976 cases, 18,399 controls). All subjects provided informed consent to participate in genetic studies and local ethics committees for each of the individual studies approved the study protocol. Definition of asthma was based on doctor's diagnosis and/or standardized questionnaires (see Supplementary Note for details). The samples were genotyped on a variety of commercial arrays, detailed in the Supplementary Note and Supplementary Table 2. GWAS were performed on imputed SNP data that were generated using HapMap2 as reference panel and one of the commonly used imputation software (Supplementary Note and Supplementary Table 2). In each dataset, the effect of each individual SNP on asthma, assuming an additive genetic model, was estimated through a logistic regression-based approach and expressed in terms of a regression coefficient with its standard error; the detailed methodology and software used for analysis by each study can be found in the Supplementary Note and Supplementary Table 2.

Imputation, quality control (including adjustments for population stratification) and analysis was done by each group independently and data on a predefined set of 3,952,683 autosomal SNPs was shared. These SNPs were those of the HapMap Phase 2, release 21 panel in subjects from European, Asian and African-ancestry that were filtered using SNP annotation from the build 37.3 of the reference sequence and dbSNP b135 (31,587 SNPs (0.8% of all

SNPs) from previous annotations that showed discrepancies with the chosen annotation were deleted). The variables that were shared contained the study name, general information on SNPs (rs number, chromosome, position, alleles (baseline and effect alleles as used in the analysis by each study), SNP status (imputed or genotyped SNP and whether the SNP genotype or imputed value was used in computation), quality control (QC) indicators (call rate and *P*-value for Hardy-Weinberg (HW) equilibrium test for genotyped SNPs, software used for imputation and imputation quality score for imputed SNPs), allele frequencies in cases and controls and information on association statistics (regression coefficient for SNP effect, standard error of regression coefficient, Z scores, *P*-values associated with Z score statistic).

Quality control of shared data

For each SNP, the alleles on the HapMap2 template (reference and alternate alleles on the positive strand) were compared to the alleles (baseline and effect alleles) used in the analysis by each group. When necessary, the association variables (allele frequencies, regression coefficient for SNP effect, Z score) were swapped to match the reference/alternate alleles of the template. Data for each SNP showing any ambiguity or error in assignment to the template were set to missing. In addition, a number of QC checks were done regarding the name, format, range of possible values for all shared variables mentioned in the previous paragraph as well as consistency checks across variables. Any problem or inconsistency was corrected, otherwise the data for that SNP were set to missing. After this first stage of QC procedure, association statistics for at least one SNP in at least one study were available for 2.83 million autosomal SNPs. Strict QC criteria were used for inclusion of a SNP in the analysis. When a SNP genotype was used in the study analysis, these criteria were: call rate ≥99%, *P*-value for HW test $\geq 10^{-6}$ and minor allele frequency (MAF) ≥0.01 in both controls and cases. When a SNP imputed value was used in the analysis, the criteria were: imputation quality score ≥0.5 and MAF≥0.01 in both

controls and cases. The distribution of the summary statistics (regression coefficient for SNP effect, standard error, Z score) of all SNPs passing QC was examined for each study; SNPs that still showed extreme Z scores (≥ 7 or ≤-7) after QC were excluded.

Meta-analysis of asthma GWAS

We conducted fixed-effects meta-analysis with inverse variance weigthing and random-effects meta-analysis using the Der Simonian and Laird[43] estimator of the between-study variance when the meta-analyses included a large number of studies (European-ancestry, multi-ancestry and pediatric sub-group meta-analyses), which allows an accurate estimate of the between-study variance. We used a fixed-effects model for the meta-analyses of the African-ancestry, Japanese and Latino populations..For all these meta-analyses, we used the SNP regression coefficient and standard error from each study for which the SNP passed QC. All meta-analyses were done with Stata version 14.1 (Stata Corp., College Station, Texas, USA). To minimize the false-positive findings and to obtain robust results, we examined the combined results for SNPs for which at least two-thirds of the studies contributed to a meta-analysis. Tests of significance of the combined effect sizes were performed using a standard normal distribution. We applied a threshold of P_{random} (or P_{fixed}) of $5x10^{-8}$ to declare a combined SNP effect as genome-wide significant. To verify the robustness of the results, we applied a genomic control correction to the association test statistics. The lead SNP at a locus was the variant with the strongest evidence for association in the European-ancestry or multi-ancestry meta-analysis. We defined a support interval around the lead SNP designated as "locus"; the bounds of this interval were the positions of the two most extreme SNPs among all SNPs lying within 500 kb on each side of the lead SNP and having P_{random} (or P_{fixed})$\leq 10^{-6}$. Heterogeneity of per-SNP effect sizes across all studies in a meta-analysis was assessed using the Cochran's Q test[9]. A difference between the four ethnic-specific summary effects was also tested with the Cochran's Q statistic.

Conditional analysis of asthma-associated loci

The Genome-wide Complex Trait Analysis (GCTA) software[44] (see URLs) was used to perform approximate conditional analysis for all loci with at least one SNP reaching the genome-wide significance level. This approximate conditional analysis is based on the summary meta-analysis statistics obtained under a fixed-effects model and takes into account the correlations among SNPs, that are estimated from a large reference population included in the meta-analysis. Approximate conditional analysis was only carried out in the European-ancestry ethnic group which can be assumed to share a similar LD pattern and represents the largest ancestry-specific dataset and the only one to show genome-wide significant results. As this analysis assumes no heterogeneity in SNP effect size across studies, the 9p24.1 and 17q12-21 loci, that show significant heterogeneity ($P_{het} \leq 0.05$ based on the Cochran's Q test) for a large portion of each locus, were not investigated. However, for the 17q12-21 locus, where there is no heterogeneity in the pediatric sub-group, GCTA was restricted to the European-ancestry pediatric sub-group. We used the large ECRHS (European Community Respiratory Health Survey) dataset as the reference sample to estimate LD. This dataset was genotyped using Illumina Human610Quad array and included 2,101 unrelated individuals after QC[22]. Imputation was done using the MACH software[45] and HapMap2, release 21 panel; only well-imputed SNPs (imputation quality score rsq>0.8) and with minor allele frequency (MAF) ≥1% were kept in this reference panel. For each asthma-associated locus, the region explored by conditional analysis extended by 500 kb on each side of the two extreme SNPs defining the support interval around the lead SNP (see preceding paragraph). However, we reduced that extension to 250 kb for the 6p21.33 and 6p21.32 loci to avoid overlap. The length of the regions explored by conditional analysis varied from 1.01 Mb to 1.63 Mb. Within each investigated region by conditional analysis, fixed-effects summary meta-analysis data for SNPs belonging to that region were adjusted for the

lead SNP using the --cojo-cond option; tests for the adjusted SNP effects were based on the Wald test If there was an additional SNP meeting the Bonferroni-corrected threshold for the total number of SNPs overall all regions investigated by GCTA (P=4.1x10^{-6}) after adjustment for the lead SNP, we performed an additional round including both SNPs. If the remaining SNPs had P-values greater than 4.1x10^{-6}, no further analysis was performed. The results of this analysis are reported in Supplementary Table 15.

Identification of cis-eQTLs at new asthma risk loci

To get greater insight into the potential genes driving the association signals at the novel asthma loci, we defined a list of SNPs to be interrogated that included the lead SNPs, the secondary signals identified by conditional analysis and all SNPs in LD with these SNPs (r^2 comprised between 0.5 and 1). To search for cis-expression quantitative trait loci (eQTLs) within at most 1 Mb of each investigated SNP, we interrogated six publically available eQTL databases by giving priority to cell types more likely to be involved in asthma biology (blood cell types and lung tissue): (i) a meta-analysis of the transcriptional profiles from peripheral blood cells of 5,311 European-ancestry subjects (the blood eQTL browser[11]); (ii) the gene expression data of 777 lymphoblastoïd cell lines (LCLs) from the MuTHER database[10]; (iii) the transcription profiles of 405 and 550 lymphoblastoïd cell lines from UK asthma (MRCA) and eczema (MRCE) family members, respectively[13]; (iv) the eQTL data from monocytes of 1,490, subjects included in the GH-express database[23]; (v) the GTEx eQTL Browser with data from multiple tissues including blood and lung[12]; (vi) the transcriptional profiles from lung tissues of 1,111 subjects[14] (see URLs).

Search for missense variants at new asthma risk loci

To complement the eQTL analysis, we searched whether the lead asthma-associated SNPs and secondary signals were in LD (r^2>0.5) with missense variants using the HaploReg v4.1 tool (see URLs).

Overlap of loci associated with asthma and other phenotypes

Overlap of novel asthma risk loci with associations with allergy-related phenotypes/diseases and immune-related diseases as well as lung function phenotypes was first annotated using the March 24, 2015 version of the NHGRI-EBI (National Human Genome Research Institute and European Bioinformatics Institute) GWAS catalog[3] (see URLs) We then used this catalog to systematically investigate the overlap of asthma signals having $P_{random} \leq 10^{-4}$ in the multi-ancestry meta-analysis with association signals of all diseases and traits in the catalog. That version of the catalog had 19,080 SNP entries, and 16,047 of those SNPs had a TAGC asthma association *P*-value. To investigate pleiotropy, we filtered out SNPs associated with asthma in the database, SNPs that have a reported GWAS *P*-value larger than 10^{-7} (with the intent of removing some of the potential false positives in the catalog) and SNPs that are duplicated (i.e., remove disease-SNP duplications). This reduced the number of entries to 5,927. Note that this process did not remove SNPs in perfect LD associated with the same disease, nor SNPs that were present multiple times in the database as associated with different phenotypes. For some diseases or quantitative traits, there were multiple SNPs in the same region reported in the catalog potentially yielding redundant information. Some of the SNPs could be in strong LD, but some could reflect independent signals. To avoid possible duplication of signals, we decided to keep only unique trait-loci combinations as reflected by the variables "Disease.Trait" and "Region" in the catalog. There were 4,231 unique entries left after this filtering step. Diseases/traits in the GWAS catalog were grouped using the classification from Wang *et al*.[37]

We summarized the overlap of GWAS catalog signals with asthma signals by the proportion of catalog SNPs with asthma *P*-values smaller than 10^{-4} in our analysis. The significance of overlap was estimated by the binomial tail probability for observing the number of TAGC SNPs with $P_{random} \leq 10^{-4}$ among the number of SNPs reported in the GWAS catalog for a group of diseases. The significance threshold for enrichment in shared associations between a disease group and asthma was set equal to 0.05 divided by the number of disease groups investigated using a Bonferroni correction. Finally, we examined a larger set of SNPs in the GWAS catalog that show an association with asthma at $P_{random} \leq 10^{-3}$ in TAGC multi-ancestry meta-analysis and estimated the proportion of false positives among those SNPs.

Enrichment of asthma risk loci in epigenetic marks

To get greater insight into the functional role of the genetic variants at the novel and known asthma loci identified by this study, we investigated whether the lead SNPs and their proxies ($r^2 \geq 0.80$) were concentrated in cis-regulatory DNA elements. We used the Uncovering Enrichment through Simulation pipeline[40] (se URLs) that was adapted to the current study. This approach tests if GWAS-identified SNPs are enriched in particular functional annotations through use of Monte Carlo simulations. The original set of asthma-associated SNPs included the lead SNPs at each asthma risk locus (ie one SNP per asthma-associated locus, as recommended by Hayes *et al*[40]). We excluded the two associated loci that span the HLA region (6p21.33 and 6p21.32) because of the high amount of variability and LD in this region. Each of the original lead SNPs is categorized by its distance from the nearest transcription start site (TSS) and number of LD partners ($r^2 \geq 0.8$). Quartiles for both the TSS distance and LD partner count are calculated and the initial SNPs are binned accordingly. Then, SNPs from the whole set of imputed SNPs used for analysis are binned according to the original SNP criteria (distance from the closest TSS, number of LD partners, and also MAF). Random SNP sets are chosen,

matching to the original bin frequencies. LD partners ($r^2 \geq 0.8$) for both the original lead SNPs and random SNPs are retrieved. The SNP data, including the original and random sets of SNPs and their corresponding LD partners ($r^2 \geq 0.8$), are intersected with the cell-specific epigenome tracks of regulatory elements using the BedTool's intersectBed[46], to determine which SNPs co-localize with a given type of regulatory elements (for example, enhancers or promoters). Those resultant SNPs are then collapsed into loci that co-localize with marks based on LD structure. We computed an empirical-*P* value for a specific track using 10,000 random SNP sets (this *P*-value is equal to r_{loci}/n where r_{loci} is the number of instances when the frequency of co-localization of the random SNP sets with the regulatory feature is greater than or equal to the frequency of co-localization with the feature for the original SNP set and n is the number of random SNP sets generated (here, 10,000). We used the Benjamini-Hochberg false discovery rates (FDR) to correct for multiple testing. We interrogated the functional data from 111 ROADMAP reference epigenomes and 16 additional epigenomes from ENCODE (Encyclopedia of DNA elements) that are available in a wide range of human cell and tissue types[24] (see URLs). We focused on enhancers and promoters that were defined using the ChromHMM 15-state model assayed in all 127 epigenomes. We also examined enrichment in DNase I hypersensitivity sites that are available in 51 cell types.

Connectivity between asthma-associated loci

We used GRAIL (Gene Relationships Across Implicated Loci)[25] to assess the relatedness between asthma associated loci. As described in detail previously[25], to define the genes near each SNP, GRAIL finds the furthest neighboring SNPs in the 3′ and 5′ direction that are in LD ($r^2 > 0.5$) and proceeds outward in each direction to the nearest recombination hotspot. All genes that overlap that interval are considered implicated by the SNP. If there are no genes in that region, the interval is extended by 250 kb in either direction. We took the genome-wide

significant signals identified by this study as a seed and queried loci to investigate biological connectivity among those loci. The connectivity between genes belonging to these loci was assessed through text-mining of PubMed abstracts. Each gene at each locus was scored for enrichment in GRAIL connectivity to genes located at the other loci by using statistical text-mining methods, as previously described[25]. The interconnectivity among genes at asthma risk loci was visualized using VIZGRAIL[47] (see URLs).

Variance explained by the asthma associated genetic variants

We estimated the variance in liability to asthma explained by the 22 distinct genome-wide significant SNPs (18 lead SNPs plus four secondary signals identified by approximate conditional analysis) at the 18 asthma-associated loci using a method based on the liability threshold model[48] and assuming a prevalence of asthma of 10%. The variance in liability to asthma explained by individual SNPs was summed over all 22 significant variants. For the loci that included two SNPs (lead SNP and secondary signal), we used the SNP effect sizes estimated by approximate joint analysis using GCTA[44]. We also estimated the variance in liability to asthma explained by the nine lead SNPs at the nine new asthma loci and by the 13 distinct genome-wide significant signals at the nine known loci.

Data availability statement

The summary statistics of the meta-analysis that support the findings of this study are available through a link from the GWAS Catalog entry for the TAGC study on the EMBL-EBI (European Bioinformatics Institute) web site (https://www.ebi.ac.uk/gwas/downloads/summary-statistics).

43. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177-88 (1986).
44. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).

45. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
46. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
47. Raychaudhuri, S. VIZ-GRAIL: visualizing functional connections across disease loci. *Bioinformatics* **27**, 1589-90 (2011).
48. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* **35**, 310-7 (2011).