Title page

Word count

Main text: 3250. Abstract: 249.

Title

Genetic associations and architecture of asthma-chronic obstructive pulmonary disease overlap

Running title

Genome-wide association study of asthma-COPD overlap

Authors

Catherine John[1*] MBBChir, Anna L. Guyatt[1*] MBChB PhD, Nick Shrine[1] PhD, Richard Packer[1] MBBS, Thorunn A. Olafsdottir[2] PhD, Jiangyuan Liu[3] PhD, Lystra P. Hayden[3] MD, Su H. Chu[3] PhD, Jukka T. Koskela[4] MD PhD, Jian'an Luan[5] PhD, Xingnan Li[6] PhD, Natalie Terzikhan[7] PhD, Hanfei Xu[8] PhD, Traci M. Bartz[9] MS, Hans Petersen[10] MS, Shuguang Leng[11] MD PhD, Steven A. Belinsky[10] PhD, Aivaras Cepelis[12] PhD, Ana I. Hernández Cordero[13] PhD, Ma'en Obeidat[13] BSPharm, PhD, Gudmar Thorleifsson[2,14] PhD, Deborah A. Meyers[6] PhD, Eugene R. Bleecker[6] MD, Lori C. Sakoda[15] PhD, Carlos Iribarren[15] MD PhD, Yohannes Tesfaigzi[16] PhD, Sina A. Gharib[17] MD, Josée Dupuis[9] PhD, Guy Brusselle[7,18] MD PhD, Lies Lahousse[7,19] PhD, Victor E. Ortega[20] MD PhD, Ingileif Jonsdottir[2,14] PhD, Don D. Sin[13] MD, Yohan Bossé[21] PhD, Maarten van den Berge[22] MD PhD, David Nickle[23,24] PhD, Jennifer K. Quint[25] MBBS PhD, Ian Sayers[26,27] PhD, Ian P. Hall[26] MBBCh PhD, Claudia Langenberg[5] MD PhD, Samuli Ripatti[4,28] PhD, Tarja Laitinen[29,30] MD PhD, Ann C. Wu[31] MD, Jessica Lasky-Su[3] ScD, Per Bakke[32] MD, Amund Gulsvik[32] MD PhD, Craig P. Hersh[3] MD, Caroline Hayward[33] PhD, Arnulf Langhammer[12] MD PhD, Ben Brumpton[34,35] MD PhD, Kari Stefansson[2,14] MD PhD, Michael H. Cho[3] MD, Louise V. Wain[1,36] PhD, Martin D. Tobin[1,36] MBChB PhD

1 Department of Health Sciences, University of Leicester, Leicester, UK.

2 deCODE genetics/Amgen, Reykjavik, Iceland.

3 Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA.

4 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland.

5 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK.
6 Division of Genetics, Genomics and Precision Medicine, Department of Medicine, University of Arizona, Tucson, USA.
7 Department of Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands.
8 Department of Biostatistics, Boston University School of Public Health, Boston, USA.
9 Cardiovascular Health Research Unit, Department of Medicine & Department of Biostatistics, University of Washington, Seattle, USA.
10 Lovelace Respiratory Research Institute, Albuquerque, USA.
11 Division of Epidemiology, Biostatistics, and Preventive Medicine, Department of Internal Medicine, University of New Mexico, Albuquerque, USA
12 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Levanger, Norway.
13 Centre for Heart Lung Innovation, University of British Columbia, Vancouver, British Columbia, Canada
14 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.
15 Division of Research, Kaiser Permanente of Northern California, Oakland, USA.
16 Brigham and Women's Hospital and Harvard Medical School, Boston, USA.
17 Computational Medicine Core, Center for Lung Biology & UW Medicine Sleep Center, Medicine, University of Washington, Seattle, USA.
18 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.
19 Department of Bioanalysis, Ghent University, Ghent, Belgium.
20 Department of Medicine, Wake Forest School of Medicine, Winston-Salem, USA
21 Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec, Canada
22 University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC Research Institute, Groningen, The Netherlands
23 Global Health, University of Washington, Seattle, USA
24 Gossamer Bio, San Diego, USA
25 National Heart and Lung Institute, Imperial College London, London, UK
26 Division of Respiratory Medicine and NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK

27 Biodiscovery Institute, University of Nottingham, Nottingham, UK
28 Broad Institute of MIT and Harvard, Cambridge, USA
29 Division of Medicine, Department of Pulmonary Diseases, Turku University Hospital, Turku, Finland.
30 Department of Pulmonary Diseases and Clinical Allergology, University of Turku, Turku, Finland.
31 Center for Healthcare Research in Pediatrics (CHeRP) and PRecisiOn Medicine Translational Research (PROMoTeR) Center, Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, USA.
32 Department of Clinical Science, University of Bergen, Bergen, Norway
33 MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
34 K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
35 Clinic of Thoracic and Occupational Medicine, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway
36 Leicester NIHR Biomedical Research Centre, Leicester, UK

* These authors contributed equally.

Corresponding authors: Catherine John, cj153@leicester.ac.uk; Anna L Guyatt, anna.guyatt@leicester.ac.uk

Financial / non-financial disclosures

This work is supported by BREATHE, the Health Data Research Hub for Respiratory Health [MC_PC_19004] in partnership with SAIL Databank. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

C.J. held a Medical Research Council Clinical Research Training Fellowship (MR/P00167X/1). A.L.G. was funded by internal fellowships at the University of Leicester from the Wellcome Trust Institutional Strategic Support Fund (204801/Z/16/Z) and the BHF Accelerator Award (AA/18/3/34220). L.P.H. is funded by NIH/NHLBI (5K23HL136851). S.H.C. is funded by NIH/NHLBI (1K01HL153941-01). A.I.H.C. is supported by MITACS accelerate. M.H.C. is

supported by R01 HL137927, R01 HL089856, and R01 HL147148. L.V.W. holds a GSK/British Lung Foundation Chair in Respiratory Research. M.D.T. is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). M.D.T. and L.V.W. have been supported by the MRC (MR/N011317/1). M.D.T. and I.P.H. hold NIHR Senior Investigator Awards. The research was partially supported by the NIHR Leicester Biomedical Research Centre and the NIHR Nottingham Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. C.J. and M.D.T. were involved in all stages of study development and delivery, and M.D.T. had full access to all data in the study and final responsibility for the decision to submit for publication. This research was funded in whole, or in part, by the Wellcome Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Please see e-Appendix for additional study acknowledgements and funding statements.

T.A.O., G.T., I.J. and K.S. are employees of deCODE genetics/Amgen Inc. E.R.B. has undertaken clinical trials through his employer, Wake Forest School of Medicine and University of Arizona, for AstraZeneca, MedImmune, Boehringer Ingelheim, Genentech, Novartis, Regeneron, and Sanofi Genzyme. E.R.B. has also served as a paid consultant for ALK-Abello, AztraZeneca, Glaxo Smith Kline, MedImmune, Novartis, Regeneron, Sanofi Genzyme, and TEVA. M.v.d.B. reports grants paid to the University from Astra Zeneca, TEVA, GSK and Chiesi outside the submitted work. D.C.N. has been a Merck & Co. employee during this study and is now an employee at Biogen Inc. I.S. has received support from GSK and Boehringer Ingelheim. I.P.H. has funded research collaborations with GSK, Boehringer Ingelheim and Orion. M.H.C. has received grant support from GSK and Bayer, and consulting and speaking fees from Genentech, Illumina and Astrazeneca. L.V.W. receives funding from GSK for a collaborative research project outside of the submitted work. M.D.T. receives funding from GSK and Orion for collaborative research projects outside of the submitted work. Other authors declare no conflicts of interest.

Prior abstract presentations

European Society of Human Genetics 2020 (virtual meeting)

European Respiratory Society 2020 (virtual meeting)

Abstract

Background Some individuals have characteristics of both asthma and chronic obstructive pulmonary disease (asthma-COPD overlap), and evidence suggests they experience worse outcomes than those with either condition alone.

Research Question What is the genetic architecture of asthma-COPD overlap, and do the determinants of risk for asthma-COPD overlap differ from those for COPD or asthma?

Study Design and Methods We conducted a genome-wide association study in 8,068 asthma-COPD overlap cases and 40,360 controls without asthma or COPD of European ancestry in UK Biobank (Stage 1). We followed up promising signals which had $p<5\times10^{-6}$, and that remained associated in analyses comparing: i) asthma-COPD overlap vs asthma-only controls, and ii) asthma-COPD overlap versus COPD-only controls). These variants were analysed in 12 independent cohorts (Stage 2).

Results We selected 31 independent variants for further investigation in stage 2, and discovered eight novel signals ($P<5x10^{-8}$) for asthma-COPD overlap (meta-analysis of Stage 1 and 2 studies). These signals suggest a spectrum of shared genetic influences, some predominantly influencing asthma (*FAM105A, GLB1, PHB, TSLP)*, others predominantly influencing fixed airflow obstruction (*IL17RD*, *C5orf56, HLA-DQB1*). One intergenic signal on chromosome 5 had not been previously associated with asthma, COPD or lung function. Subgroup analyses suggested that associations at these eight signals were not driven by smoking or age at asthma diagnosis, and in phenome-wide scans, eosinophil counts, atopy and asthma traits were prominent.

Interpretation We identified eight signals for asthma-COPD overlap, which may represent loci that predispose to type 2 inflammation, and serious long-term consequences of asthma.

Key words

epidemiology; genome-wide association study; asthma; chronic obstructive pulmonary disease; spirometry

Abbreviations

95% CI 95% confidence interval; ACO asthma-COPD overlap; COPD chronic obstructive pulmonary disease; eQTL expression quantitative trait locus; FEV_1 forced expiratory volume

in 1 second; FVC forced vital capacity; FDR false discovery rate; GWAS genome-wide association study; HLA Human leukocyte antigen; LDSC Linkage disequilibrium score regression; MHC Major histocompatibility complex; MAF Minor allele frequency; OR odds ratio; SNP single-nucleotide polymorphism

Asthma and COPD have substantial global impacts.[1] They are heterogeneous conditions[2-4] that share some common features, including airflow obstruction with differing degrees of reversibility. Inflammatory processes are important in both conditions, and cytokine profiles identify both distinct and overlapping groups of patients.[5] Individuals with characteristics of both conditions have until recently been referred to as having "asthma-COPD overlap" (ACO),[4] and a number of studies have suggested that such patients have significantly worse outcomes than those with either condition alone.[6-13] Recent guidelines emphasize that asthma and COPD are different conditions, but may coexist in the same patient.[14] Individuals with features of both diseases risk being excluded from research that might provide evidence about the most effective treatment strategies.[3]

Environmental risk factors – notably smoking in COPD – are central, but genetics also plays an important role in both asthma and COPD,[15-17] and it has long been hypothesized that there may be a shared, underlying genetic predisposition to both diseases.[2,18] Genome-wide association studies (GWAS) examine variants across the genome in an unbiased manner, to identify variant-trait associations that inform understanding of disease biology and potential treatment strategies. GWAS have identified many loci associated with asthma or COPD in European populations (**e-Appendix**). The genetic correlation (r_g) between asthma and COPD is 0.38 ($p=6.2\times10^{-5}$), suggesting shared genetic aetiology.[19] A GWAS of ACO compared to COPD alone (n=3570) did not identify any variants associated at the conventional threshold,[8] and a meta-analysis of an asthma and COPD GWAS found one association, driven by COPD.[20] Eighteen loci outside the HLA (human leucocyte antigen) region have been identified as associated with both asthma and lung function/COPD at $p<5x10^{-8}$, but have not been specifically described as ACO loci.

Notwithstanding the controversies of changing terminology for individuals with both asthma and COPD, we refer to this case status as "ACO". Improved knowledge of genetic variants associated with co-existing asthma and COPD would contribute to understanding of underlying molecular pathways, and potentially inform diagnostic terminology and specific management strategies for those with co-existing asthma and COPD.

Accordingly, using spirometry, self-report and electronic healthcare record (EHR) data to define cases with both asthma and COPD (ACO) and suitable controls, we undertook the

largest GWAS of coexisting asthma and COPD to date, including up to 12,369 cases and 88,969 controls, in a two-stage design incorporating 13 studies.

Study Design and Methods

Stage 1

The data source for this study was UK Biobank (http://www.ukbiobank.ac.uk).[21] Eligibility criteria, genotyping and quality control are described in the e-Appendix. 321,057 individuals and 37 million single-nucleotide polymorphisms (SNPs) were included.

We defined cases of ACO if they had self-reported asthma (see **e-Appendix**) AND FEV_1/FVC <0.7 with GOLD 2+ airflow limitation (FEV_1 <80% predicted). Cases who reported alpha-1-antitrypsin deficiency were excluded. Controls reported no asthma or COPD (**e-Appendix**), and had FEV_1 ≥80% predicted and FEV_1/FVC >0.7. Five controls were randomly selected for each case. Cases and controls were unrelated (second degree or closer). Two additional control sets were defined for signal prioritisation: individuals with asthma but without COPD, and individuals with COPD but without asthma. Asthma and COPD were defined as above.

Association testing was undertaken in SNPTEST ('score' option) (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html, version 2.5.2), under an additive model. Age, sex, smoking status (ever/never), genotyping array and 10 principal components were included as covariates. Variants were filtered based on minor allele frequency (MAF) >0.01 and imputation quality (INFO) >0.5. P-values and standard errors were adjusted for the LD score regression intercept (LDSC, https://github.com/bulik/ldsc) (**e-Figure 1**).

In stage 1, we defined distinct signals passing a P-value threshold of $P<5x10^{-6}$. We defined regions of association around the most strongly associated variant (sentinel variant) ±1Mb. To identify distinct signals, and additional signals within the regions described above, conditional analyses were undertaken using GCTA-COJO (http://cnsgenomics.com/software/gcta/#COJO) (**e-Appendix**, **e-Figure 2**).

Two further "signal prioritisation" analyses were undertaken to ascertain the extent to which signals were driven by association with COPD and/or asthma alone. These included the same cases as the primary analysis, plus the two additional control sets described above. Variants were selected for follow-up in Stage 2 if they were associated at $P<5x10^{-6}$ in the main Stage 1 analysis and at P<0.01 in both signal prioritisation analyses.

Stage 2 and joint analysis

SNPs identified in Stage 1 signal prioritisation analyses were tested for association in twelve independent studies of European ancestry (up to 4,301 cases and 48,609 controls, in CHS, COPDGene, deCODE, ECLIPSE, EPIC-Norfolk, FHS, Generation Scotland, GenKOLS, the Trondelag Health Study [HUNT], LOVELACE, Rotterdam Study, SPIROMICS) and one African-American ancestry cohort (COPDGene; 297 cases, 1335 controls) (**e-Appendix**, **e-Table 1**, **e-Table 2**).

Cases had both asthma and COPD. Asthma was defined as any lifetime self-report of asthma, or asthma diagnosis in the healthcare record, including billing codes (**e-Appendix** for further details and validation).[22] All cases had spirometry indicating $FEV_1/FVC<0.7$, and $FEV_1<80\%$ predicted. All controls had $FEV_1/FVC>0.7$, $FEV_1\geq80\%$ predicted and no asthma diagnosis. Where possible, studies excluded individuals with alpha-1-antitrypsin deficiency.

Details of statistical analysis in Stage 2 studies are in the **e-Appendix** (and **e-Table 3**). Results were combined across Stage 2 studies using fixed-effect meta-analysis. Heterogeneity was assessed using the I^2 statistic. We combined these results with those from UK Biobank (Stage 1).

We performed a sensitivity analysis to assess whether the way COPD was defined changed our Stage 2 results (**e-Appendix**).

To assess whether associations with our Stage 1 signals changed according to age of asthma diagnosis, we divided cases into those who self-reported their age at asthma diagnosis as <12 years, and >25 years.[23] We then repeated the association tests in UK Biobank. In addition, we repeated association testing after stratifying our sample into ever-/never-smokers.

Definition of top signals for bioinformatic analyses

We undertook bioinformatic analyses on ACO signals reaching $p<5x10^{-8}$ in the joint analysis of Stages 1 and 2, and which also had a lower p-value in the joint analysis than in UK Biobank (Stage 1) alone or had $p<0.05$ in Stage 2. For each of these, we identified the set of SNPs that was 99% likely to contain the causal variant ('99% credible set'), assuming that the causal variant was included in the dataset **(e-Appendix)**.[24] For bioinformatic analysis methods, see **e-Appendix**.

Using LD score regression,[25] we computed genetic correlations between ACO (Stage 1 results), asthma,[26] moderate-severe asthma,[27] COPD,[28] eosinophil counts,[29] and FEV_1/FVC.[30] We also computed genetic correlations between ACO and atopic, auto-immune, and smoking behaviour traits (http://ldsc.broadinstitute.org/).[31]

Approvals

The research was conducted using UK Biobank (http://www.ukbiobank.ac.uk), under application 648. UK Biobank has ethical approval from the UK National Health Service (NHS) National Research Ethics Service (11/NW/0382). All included studies were approved by the appropriate research ethics committee or institutional review board (**e-Appendix**). All participants gave informed consent.

Results

In Stage 1, 8,068 ACO cases were selected from UK Biobank, and 40,360 as healthy controls free of asthma and COPD. For signal prioritisation analyses, another 16,762 individuals were selected as controls with COPD alone (without asthma), and 26,815 as controls with asthma alone (without COPD). Descriptive statistics for cases and controls are in **Table 1**. ACO cases were slightly older than healthy controls, and included more males and ever-smokers.

After filtering on MAF and INFO, 7,693,381 variants were analysed. The LDSC regression intercept was 1.018, suggesting that results were not strongly inflated due to population structure (**e-Figure 1)**.[25]

ACO association signals

In stage 1, there were 83 distinct signals at $P<5x10^{-6}$ (**Figure 1**,[29] **e-Appendix** and **e-Figure 2** for the signal selection, **e-Table 4** for results). Of these, 31 retained significance (P<0.01) in signal prioritisation analyses comparing ACO cases separately with either COPD cases or asthma cases, to determine whether signals were driven by asthma or COPD alone (**e-Table 4**). In Stage 2, comprising 12 independent cohorts (4301 cases, 48609 controls) (**e-Table 1** and **e-Table 2**), 26/31 signals had a direction of effect concordant with Stage 1 (**e-Table 5**), and the median value for heterogeneity (I^2) across these signals was 15%. Whilst the sample size of individuals of African-American ancestry was small (297 cases, 1335 controls) and confidence intervals were broad, 22/31 signals had a direction of effect consistent with the European ancestry studies (**e-Table 5)**.

Results for the Stage 2 sensitivity analysis (9,638 cases and 128,273 controls from 15 studies), where COPD was defined either by available spirometry or, alternatively, by EHR diagnoses (**e-Appendix**), are in **e-Table 6**.

Subgroup analyses

Effect sizes for the 31 signals amongst cases with childhood-onset asthma were highly correlated with those amongst individuals with adult onset (R=0.883) (**e-Table 7, e-Figure 3**). Effect sizes in ever- and never-smokers were also closely correlated (R=0.911) (**e-Table 7** and **e-Figure 4**).

Eight top signals for ACO defined from joint analysis

After meta-analysis combining Stage 1 and Stage 2, 13 signals were genome-wide significant ($p<5x10^{-8}$) (**e-Table 4**; **e-Figure 2** for flow diagram). Of these, eight either had a lower p-value in the joint analysis than in Stage 1 alone, or p<0.05 in Stage 2 studies alone (**Table 2, e-Figure 5, e-Figure 6**). None of these eight signals are previously reported as associated specifically with ACO.[8]

For the novel intergenic ACO signal on chromosome 5 (rs80101740, effect allele frequency (EAF)=0.015, OR=1.42, $P=3.72x10^{-8}$, **e-Table 5**), which has not been previously associated with asthma, lung function or COPD, the sentinel SNP had the largest posterior probability (0.77) of being the true causal variant, assuming the causal variant was genotyped/imputed **(e-Table 8)**. There was no evidence of colocalisation with eQTL signals at this locus (**e-Tables 9 and 10**), and no chromatin interactions were identified.

Four of our novel signals for ACO were previously reported for asthma but not COPD/lung function.[32-34] For rs35570272 in *GLB1* (OR=1.10, EAF 0.398, $P=2.44x10^{-9}$), there were 11 SNPs in the credible set, and the intronic sentinel SNP had the highest posterior probability (0.655). There were significant chromatin interactions nearby in *GLB1* in fetal lung fibroblasts. *GLB1* encodes the beta-galactosidase enzyme involved in lysosomal function, and an elastin-binding protein involved in extracellular elastic fibre formation. Two SNPs (both with a posterior probability ~0.13) in the 99% credible set, rs7646283 and rs34064757, were eQTLs for cartilage-associated protein (*CRTAP*) in lung (**e-Table 9**), implicated in bone development and osteogenesis imperfecta.

Another signal (previously reported for asthma) was rs16903574 (EAF=0.077, OR=1.20, $P=3.8x10^{-10}$), a missense variant in *FAM105A*, deleterious according to its CADD score (22.6).[35] *FAM105A* encodes a pseudoenzyme, possibly involved in protein-protein interactions.[36] This sentinel had a posterior probability of 0.99. A previous study in asthma predicted *FAM105A* as the target based on chromatin interactions and correlation between enhancer epigenetic marks and gene expression, although we did not identify any eQTL evidence in lung or whole blood.[32] We also identified a highly significant chromatin interaction in fetal lung fibroblasts overlapping *FAM105A* and another nearby gene (*TRIO*), but not in adult lung.

An intergenic signal between *PHB* and *ZNF652* (rs2584662; EAF=0.42, OR=0.92, P=2.21x10^{-8}) was previously associated with asthma and reported as a blood eQTL for *GNGT2* (implicated in NF-κB activation),[26,32] although we did not identify this in our eQTL analysis. In our analysis, eight SNPs were in the credible set (posterior probabilities all ≤0.2). Hi-C data suggested a significant chromatin interaction in *ZNF652*, with another less significant peak close to *GNGT2*. Nearby loci in *ZNF652* have previously been associated with asthma/allergic disease and moderate-to-severe asthma.[32]

We also identified rs1837253, an intergenic signal near *TSLP* (EAF 0.739, OR=1.16, P=1.53x10^{-18}), with a posterior probability of 1, i.e. the only variant in the credible set. No eQTL evidence was identified. There were highly significant chromatin interactions with *SLC25A46* in fetal lung fibroblasts and in adult lung tissue, and with a region between *TSLP* and *SLC25A46* in fetal lung fibroblasts only. The cytokine TSLP was implicated in asthma and allergic disease prior to the GWAS era,[37] and an anti-TSLP antibody has been trialled in allergic asthma.[38]

Another signal, rs6787279 in *IL17RD* (EAF=0.169, OR=0.89, P=7.87x10^{-9}), has been previously reported for lung function and COPD.[28,39] There were 55 variants in the credible set, all with posterior probability ≤0.12, meaning it is not yet possible to fine-map this signal confidently. One SNP in the credible set was exonic and possibly damaging (rs17057718), but the posterior probability was only 0.012. Multiple SNPs at this locus were eQTLs for *IL17RD* in lung, with the ACO risk allele corresponding to decreased *IL17RD* expression. IL17RD is in the IL17 signalling pathway, which is implicated in asthma,[40] and in COPD pathogenesis,[41,42] potentially by mediating effects of cigarette smoke.

Two ACO signals have previously been reported separately for both asthma and lung function or COPD: rs9273410 in *HLA-DQB1* (EAF=0.445, OR=1.20, P=9.19x10^{-28}) and rs3749833 in *C5orf56* (EAF=0.263, OR=1.12, P=9.37x10^{-12}). *HLA-DQB1* encodes a major histocompatibility complex (MHC) type II molecule involved in antigen presentation. *HLA-DQB1* alleles are associated with numerous inflammatory and autoimmune diseases. In our analysis, the sentinel was the only SNP in the credible set. For lung function, an amino acid change in the gene product HLA-DQβ1 has been identified as the main driver of signals in the MHC region.[30] Analyses in asthma have identified *HLA-DQA1* as the likely driver gene.[32]

C5orf56 is located on a cytokine gene cluster on chromosome 5, including *IL3*, *IL4* and *IL5*. Several interleukins in this region have been considered as therapeutic targets in asthma. In severe eosinophilic asthma, the anti-IL5 monoclonal antibodies mepolizumab and reslizumab reduce exacerbations and improve quality of life.[43-45] SNPs in the credible set were eQTLs in lung and/or blood for *SLC22A5*, *AC116366*.6, *RAD50* and a non-coding Y RNA. *SLC22A4* has been identified as the most likely candidate gene for the lung function association.[30] The gene products of *SLC22A4* and *SLC22A5* are involved in bronchial uptake of bronchodilators and anti-cholinergic drugs.[46] An analysis in asthma predicted *C5orf56* (which encodes the interferon regulatory factor 1 antisense RNA, *IRF1-AS1*) as the causal gene.[32]

In our phenome-wide scan, all ACO loci previously associated with asthma showed association with blood cell counts, particularly eosinophils and neutrophils, and atopic traits (**e-Table 11**). The HLA locus was associated with a wide range of autoimmune/inflammatory traits. Another locus (rs2584662, near *PHB* and *ZNF652*), was associated with anthropometric traits, cardiovascular phenotypes and chronic diseases/multimorbidity, whilst rs3749833 (near *C5orf56*), was associated with anthropometric traits and inflammatory bowel disease. SNPs in the credible set for the intergenic chromosome 5 signal (rs80101740) were associated with cardiovascular and a range of other traits.

ACO shares genetic architecture with other traits

We observed genetic correlations (r_g) of broadly similar magnitude between ACO and COPD (r_g=0.828, p=3.19×10^{-299}), ACO and asthma (r_g =0.743, p=6.18×10^{-44}), and ACO and FEV_1/FVC (r_g=-0.692, p=7.48×10^{-33}) (**Figure 2, e-Table 12**). The genetic correlation (r_g) between asthma and FEV_1/FVC was -0.333 (p=8.71×10^{-7}), (i.e. increased asthma risk was correlated with lower FEV_1/FVC). Blood eosinophil counts were moderately correlated with ACO (r_g=0.292, p=4.87×10^{-11}), similar in magnitude to the correlation of eosinophils with asthma (r_g=0.371, p=3.15×10^{-7}), whereas the correlation of eosinophils with FEV_1/FVC (r_g =-0.070, p=0.002) and COPD (r_g=0.130, p=4.83×10^{-6}) were smaller. We additionally computed genetic correlations between ACO and 16 autoimmune traits, and ACO and smoking behaviour (r_g=0.046, p=0.417) (**e-Table 12**). After asthma, the next strongest correlation was with eczema (r_g=0.255, p=0.004), then multiple sclerosis (r_g=0.323, p=0.011).

Discussion

We conducted the largest GWAS of ACO to date, and identified 83 independent signals associated at $P<5x10^{-6}$ in Stage 1. After excluding variants associated with asthma only or COPD only, we studied 31 variants in stage 2, with eight distinct signals for ACO showing genome-wide significance ($P<5x10^{-8}$) in a Stage 1 and Stage 2 meta-analysis.

Our study contributes to understanding of the genetic architecture of ACO. We showed strong genetic correlation between ACO and COPD/lung function, and ACO and asthma, especially moderate-severe asthma. Furthermore, we showed that the genetic correlation of blood eosinophil counts with ACO was more similar in magnitude to the genetic correlation of eosinophils with asthma than of eosinophils with FEV_1/FVC and COPD. Increased eosinophils are associated with asthma and COPD exacerbations,[47-49] and with lung function decline in subjects with and without asthma.[50] Eosinophil counts, atopy and asthma traits were prominent in phenome-wide scans of our top eight signals, consistent with an important role for type 2 inflammation in ACO.[51,52]

One intergenic signal on chromosome 5 (rs80101740) is not reported as associated with asthma, COPD or lung function. Whilst near to a putative signal for lung function without replication support (rs377731, r^2=0.02 with rs80101740),[30] the ACO sentinel was independent of this lung function signal in conditional analyses. Evidence from eQTL studies suggests that the nearby lung function signal is associated with *RGMB* and *LINC02062* expression.

Four of the eight signals identified as novel (*GLB1, FAM105A, PHB, TSLP*) are known signals for asthma or allergic disease, but not COPD. Our results suggest that these loci also have a role in COPD. All four have been associated with child- and adult-onset asthma, and could represent an opportunity to intervene in early life to prevent serious long-term sequelae.[23] One ACO signal (*IL17RD*) is a known lung function and COPD locus; our findings demonstrate its relevance in reversible airflow obstruction. Together, these loci could represent targets for intervention, potentially to prevent development of fixed airflow obstruction.

Two signals are known to be associated with asthma and COPD/lung function, including the *HLA-DQB1* locus (the first signal identified as associated with both asthma and COPD), and a signal at *C5orf56*, encoding *IRF1-AS1*, on chromosome 5, near a cytokine gene cluster.

In subgroup analyses, there was a strong positive correlation between Stage 1 effect sizes for ACO in ever- and never-smokers, suggesting that ACO is not due solely to smoking in people with asthma, although childhood asthma in smokers increases COPD risk compared with non-asthmatics, possibly via early lung development.[53] Similarly, when stratifying by child- versus adult-onset asthma, there was a strong correlation between effect sizes in both groups. Nevertheless, for some of the eight top signals, we found evidence of chromatin interactions in fetal but not adult lung. Although this may implicate developmental processes in ACO, inference is difficult, due to differences in experimental conditions, sample sizes and reporting practices. Clearer conclusions may become possible as functional genomic assays advance.

Our study has some potential limitations. The stage 2 sample size (4,301 cases) was substantial, although relatively underpowered compared to stage 1 (8,068 cases). All signals reported met commonly-adopted criteria for genome-wide significance, but stricter criteria are starting to be used for genome sequencing studies;[54] future work using sequence data would provide an opportunity to re-evaluate the genomic regions we highlight. Misclassification of asthma and COPD diagnoses is possible: asthma in older patients may mimic COPD, and clinicians may be less likely to suspect COPD in non-smokers. To mitigate this, we utilised GOLD 2+ spirometric criteria to define COPD wherever possible, and note that self-reported asthma has been shown to accurately identify subjects with clinical and genetic characteristics of asthma.[53] We hypothesise that any remaining misclassification would attenuate effect estimates towards the null, i.e. reduce power to detect true genetic associations with ACO. Our main analysis was undertaken in European ancestry populations only; although for many loci there was good concordance in a small sample of African-American ancestry, it is essential to study this trait further in diverse populations.

Interpretation

In the largest genome-wide association study to date, we identified eight signals associated with ACO. Our findings suggest a spectrum of shared genetic influences, from variants predominantly influencing asthma, to those predominantly influencing fixed airflow obstruction. We focus on variants that tend towards an intermediate phenotype with features of both asthma and fixed airflow obstruction, with pathways implicating innate and adaptive immunity and potentially bone development, and signals for which the biology

remains unclear. Further biological understanding is likely to be important for therapeutics to prevent the development of fixed airflow obstruction among people with asthma.

References

1. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir. Med.* 2020;8(6):585-596.
2. Bateman ED, Reddel HK, van Zyl-Smit RN, Agusti A. The asthma-COPD overlap syndrome: towards a revised taxonomy of chronic airways diseases? *Lancet Respir. Med.* 2015;3(9):719-728.
3. Postma DS, Rabe KF. The Asthma-COPD Overlap Syndrome. *N. Engl. J. Med.* 2015;373(13):1241-1249.
4. Global Initiative for Asthma, Global Initiative for Chronic Obstructive Lung Disease. *Diagnosis of Diseases of Chronic Airflow Limitation: Asthma, COPD and Asthma-COPD Overlap Syndrome (ACOS).* GINA;2014.
5. Ghebre MA, Bafadhel M, Desai D, et al. Biological clustering supports both "Dutch" and "British" hypotheses of asthma and chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* 2015;135(1):63-72.e10.
6. de Marco R, Marcon A, Rossi A, et al. Asthma, COPD and overlap syndrome: a longitudinal study in young European adults. *Eur. Respir. J.* 2015;46(3):671-679.
7. Miravitlles M, Soriano JB, Ancochea J, et al. Characterisation of the overlap COPD-asthma phenotype. Focus on physical activity and health status. *Respir. Med.* 2013;107(7):1053-1060.
8. Hardin M, Cho M, McDonald ML, et al. The clinical and genetic features of COPD-asthma overlap syndrome. *Eur. Respir. J.* 2014;44(2):341-350.
9. Menezes AMB, Montes de Oca M, Perez-Padilla R, et al. Increased risk of exacerbation and hospitalization in subjects with an overlap phenotype: COPD-asthma. *Chest.* 2014;145(2):297-304.
10. Lange P, Colak Y, Ingebrigtsen TS, Vestbo J, Marott JL. Long-term prognosis of asthma, chronic obstructive pulmonary disease, and asthma-chronic obstructive pulmonary disease overlap in the Copenhagen City Heart study: a prospective population-based analysis. *Lancet Respir. Med.* 2016;4(6):454-462.
11. Kauppi P, Kupiainen H, Lindqvist A, et al. Overlap syndrome of asthma and COPD predicts low quality of life. *J. Asthma.* 2011;48(3):279-285.
12. Diaz-Guzman E, Khosravi M, Mannino DM. Asthma, chronic obstructive pulmonary disease, and mortality in the U.S. population. *COPD.* 2011;8(6):400-407.
13. Henriksen AH, Langhammer A, Steinshamn S, Mai XM, Brumpton BM. The Prevalence and Symptom Profile of Asthma-COPD Overlap: The HUNT Study. *COPD.* 2018;15(1):27-35.
14. Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease.* 2020.
15. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *Am. J. Respir. Crit. Care Med.* 2013;188(8):941-947.
16. Los H, Koppelman GH, Postma DS. The importance of genetic influences in asthma. *Eur. Respir. J.* 1999;14(5):1210-1227.
17. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of Asthma and Hay Fever in Australian Twins. *Am. Rev. Respir. Dis.* 1990;142(6_pt_1):1351-1358.
18. Sluiter HJ, Koëter GH, de Monchy JG, Postma DS, de Vries K, Orie NG. The Dutch hypothesis (chronic non-specific lung disease) revisited. *Eur. Respir. J.* 1991;4(4):479-489.
19. Hobbs BD, de Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat. Genet.* 2017;49(3):426-432.
20. Smolonska J, Koppelman GH, Wijmenga C, et al. Common genes underlying asthma and COPD? Genome-wide analysis on the Dutch hypothesis. *Eur. Respir. J.* 2014;44(4):860-872.

21. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.
22. John C, Reeve NF, Free RC, et al. Cohort Profile: Extended Cohort for E-health, Environment and DNA (EXCEED). *Int. J. Epidemiol.* 2019;48(3):678-679j.
23. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir. Med.* 2019;7(6):509-522.
24. Wakefield J. Reporting and interpretation in genome-wide association studies. *Int. J. Epidemiol.* 2008;37(3):641-653.
25. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 2015;47(11):1236-1241.
26. Demenais F, Margaritte-Jeannin P, Barnes KC, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* 2018;50(1):42-53.
27. Shrine N, Portelli MA, John C, et al. Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *Lancet Respir. Med.* 2019;7(1):20-34.
28. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat. Genet.* 2019;51(3):494-505.
29. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167(5):1415-1429.e1419.
30. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* 2019;51:481-493.
31. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2016;33(2):272-279.
32. Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* 2017;49(12):1752-1757.
33. Hirota T, Takahashi A, Kubo M, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* 2011;43(9):893-896.
34. Ferreira MAR, Matheson MC, Tang CS, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J. Allergy Clin. Immunol.* 2014;133(6):1564-1571.
35. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat. Methods.* 2010;7(4):248-249.
36. Ceccarelli DF, Ivantsiv S, Mullin AA, et al. FAM105A/OTULINL Is a Pseudodeubiquitinase of the OTU-Class that Localizes to the ER Membrane. *Structure.* 2019;27(6):1000-1012.e1006.
37. Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin Transl Immunology.* 2017;6(12):e165.
38. Gauvreau GM, O'Byrne PM, Boulet LP, et al. Effects of an anti-TSLP antibody on allergen-induced asthmatic responses. *N. Engl. J. Med.* 2014;370(22):2102-2110.
39. Wyss AB, Sofer T, Lee MK, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. *Nat. Commun.* 2018;9(1):2976.
40. Hynes GM, Hinks TSC. The role of interleukin-17 in asthma: a protective response? *ERJ Open Res.* 2020;6(2).
41. Miossec P, Kolls JK. Targeting IL-17 and TH17 cells in chronic inflammation. *Nat. Rev. Drug Discov.* 2012;11(10):763-776.

42. Mellett M, Atzei P, Horgan A, et al. Orphan receptor IL-17RD tunes IL-17A signalling and is required for neutrophilia. *Nat. Commun.* 2012;3:1119.
43. Ortega HG, Liu MC, Pavord ID, et al. Mepolizumab treatment in patients with severe eosinophilic asthma. *N. Engl. J. Med.* 2014;371(13):1198-1207.
44. Chupp GL, Bradford ES, Albers FC, et al. Efficacy of mepolizumab add-on therapy on health-related quality of life and markers of asthma control in severe eosinophilic asthma (MUSCA): a randomised, double-blind, placebo-controlled, parallel-group, multicentre, phase 3b trial. *Lancet Respir. Med.* 2017;5(5):390-400.
45. Castro M, Zangrilli J, Wechsler ME, et al. Reslizumab for inadequately controlled asthma with elevated blood eosinophil counts: results from two multicentre, parallel, double-blind, randomised, placebo-controlled, phase 3 trials. *Lancet Respir. Med.* 2015;3(5):355-366.
46. Mukherjee M, Cingolani E, Pritchard DI, Bosquillon C. Enhanced expression of Organic Cation Transporters in bronchial epithelial cell layers following insults associated with asthma - Impact on salbutamol transport. *Eur. J. Pharm. Sci.* 2017;106:62-70.
47. Zeiger RS, Schatz M, Li Q, et al. High blood eosinophil count is a risk factor for future asthma exacerbations in adult persistent asthma. *J Allergy Clin Immunol Pract.* 2014;2(6):741-750.
48. Price D, Wilson AM, Chisholm A, et al. Predicting frequent asthma exacerbations using blood eosinophil count and other patient data routinely available in clinical practice. *J Asthma Allergy.* 2016;9:1-12.
49. Siva R, Green RH, Brightling CE, et al. Eosinophilic airway inflammation and exacerbations of COPD: a randomised controlled trial. *Eur. Respir. J.* 2007;29(5):906-913.
50. Hancox RJ, Pavord ID, Sears MR. Associations between blood eosinophils and decline in lung function among adults with and without asthma. *Eur. Respir. J.* 2018;51(4):1702536.
51. Miravitlles M. Diagnosis of asthma–COPD overlap: the five commandments. *Eur. Respir. J.* 2017;49(5):1700506.
52. Christenson SA, Steiling K, van den Berge M, et al. Asthma-COPD overlap. Clinical relevance of genomic signatures of type 2 inflammation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2015;191(7):758-766.
53. Hayden LP, Cho MH, Raby BA, Beaty TH, Silverman EK, Hersh CP. Childhood asthma is associated with COPD and known asthma variants in COPDGene: a genome-wide association study. *Respir. Res.* 2018;19(1):209.
54. Pulit SL, de With SA, de Bakker PI. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet. Epidemiol.* 2017;41(2):145-151.

Take-Home Points

Study question

What are the genetic determinants of risk for asthma-COPD overlap, and how do these differ from those for COPD or asthma?

Results

We discovered eight novel signals for asthma-COPD overlap in a meta-analysis of 12,369 cases and 88,969 controls; most signals suggested a spectrum of shared genetic influences on asthma, COPD or lung function, and in phenome-wide scans of these signals, eosinophil counts, atopy and asthma traits were prominent.

Interpretation

We identified eight signals for asthma-COPD overlap, not driven by smoking or age at asthma diagnosis, which may represent loci that predispose to type 2 inflammation, and serious long-term consequences of asthma.

Acknowledgements

Guarantor statement

C.J., A.L.G. and M.D.T. will act as guarantors for the content of the manuscript.

Author Contributions

C.J., A.L.G., L.V.W., M.D.T. contributed to the conception and design of the study. C.J., A.L.G., N.S., T.A.O., J. Liu, L.P.H., S.H.C., J.K., J. Luan, X.L., N.T., H.X., T.M.B., H.P., S.L., A.C., A.I.H.C., M.O. undertook data analysis. C.J., A.L.G., N.S., R.P., S.A.B. M.O., G.T., D.A.M., E.R.B., L.C.S., C.I., Y.T., S.A.G., J.D., L.L., V.E.O., I.J., J.K.Q., D.D.S., Y.B., M.v.d.B, D.C.N., I.S., I.P.H., C.L., S.R., T.L., A.C.W., J. Lasky-Su, P.B., A.G., C.P.H., C.H., A.L., B.B., K.S., M.H.C., L.V.W., M.D.T. contributed to data acquisition and/or interpretation. C.J., A.L.G., L.V.W., M.D.T. drafted the manuscript. All authors critically reviewed the manuscript before submission.

Financial / non-financial disclosures

This work is supported by BREATHE, the Health Data Research Hub for Respiratory Health [MC_PC_19004] in partnership with SAIL Databank. BREATHE is funded through the UK

Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

C.J. held a Medical Research Council Clinical Research Training Fellowship (MR/P00167X/1). A.L.G. was funded by internal fellowships at the University of Leicester from the Wellcome Trust Institutional Strategic Support Fund (204801/Z/16/Z) and the BHF Accelerator Award (AA/18/3/34220). L.P.H. is funded by NIH/NHLBI (5K23HL136851). S.H.C. is funded by NIH/NHLBI (1K01HL153941-01). A.I.H.C. is supported by MITACS accelerate. M.H.C. is supported by R01 HL137927, R01 HL089856, and R01 HL147148. L.V.W. holds a GSK/British Lung Foundation Chair in Respiratory Research. M.D.T. is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). M.D.T. and L.V.W. have been supported by the MRC (MR/N011317/1). M.D.T. and I.P.H. hold NIHR Senior Investigator Awards. The research was partially supported by the NIHR Leicester Biomedical Research Centre and the NIHR Nottingham Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. C.J. and M.D.T. were involved in all stages of study development and delivery, and M.D.T. had full access to all data in the study and final responsibility for the decision to submit for publication. This research was funded in whole, or in part, by the Wellcome Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. Please see e-Appendix for additional study acknowledgements and funding statements.

T.A.O., G.T., I.J. and K.S. are employees of deCODE genetics/Amgen Inc. E.R.B. has undertaken clinical trials through his employer, Wake Forest School of Medicine and University of Arizona, for AstraZeneca, MedImmune, Boehringer Ingelheim, Genentech, Novartis, Regeneron, and Sanofi Genzyme. E.R.B. has also served as a paid consultant for ALK-Abello, AztraZeneca, Glaxo Smith Kline, MedImmune, Novartis, Regeneron, Sanofi Genzyme, and TEVA. M.v.d.B. reports grants paid to the University from Astra Zeneca, TEVA, GSK and Chiesi outside the submitted work. D.C.N. has been a Merck & Co. employee during this study and is now an employee at Biogen Inc. I.S. has received support from GSK and Boehringer Ingelheim. I.P.H. has funded research collaborations with GSK, Boehringer Ingelheim and Orion. M.H.C. has received grant support from GSK and Bayer, and consulting and speaking fees from Genentech, Illumina and Astrazeneca. L.V.W. receives funding from

GSK for a collaborative research project outside of the submitted work. M.D.T. receives funding from GSK and Orion for collaborative research projects outside of the submitted work.

Role of the funding source

The funders had no role in the design of the analyses or conduct of the study.

Data sharing statement

Individual-level participant data are from UK Biobank (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). Summary-level genome-wide association statistics will be made publicly available via the EBI GWAS catalog on publication (https://www.ebi.ac.uk/gwas/).