**A genome-wide association study of moderate-severe asthma in individuals of European ancestry.**

Nick Shrine PhD1†, Michael A. Portelli PhD2†, Catherine John MD1†, María Soler Artigas PhD1, Neil Bennett1, Robert Hall MRes2, Jon Lewis MRes2, Amanda P. Henry PhD 2, Charlotte K. Billington PhD 2, Azaz Ahmad PhD2, Richard J. Packer MD1, Prof. Dominick Shaw PhD 2, Zara E.K. Pogson PhD 3, Prof. Andrew Fogarty PhD 3, Prof. Tricia M. McKeever PhD3, Amisha Singapuri4, Prof. Liam G. Heaney MD5, Adel H. Mansur MD6, Rekha Chaudhuri MD7, Prof. Neil C. Thomson MD7, Prof. John W. Holloway PhD 8, Gabrielle A. Lockett PhD 8, Prof. Peter H. Howarth PhD8, Prof. Ratko Djukanovic PhD8, Jenny Hankinson PhD 9, Robert Niven MD9, Prof. Angela Simpson PhD9, Prof. Kian Fan Chung MD DSc10, Prof. Peter J. Sterk PhD11, John D. Blakey PhD12, Prof. Ian M. Adcock PhD10, Sile Hu PhD13, , Prof. Yike Guo PhD13, Maen Obeidat PhD14,Prof. Don D. Sin MD14, 15, Prof. Maarten van den Berge PhD16, Prof. David C. Nickle PhD17, Prof. Yohan Bossé PhD18, Prof. Martin D. Tobin PhD1,19,Prof. Ian P. Hall DM2, Prof. Christopher E. Brightling PhD4,19, Prof. Louise V. Wain PhD 1,19\*†& Prof. Ian Sayers PhD 2\*†.

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

*2Division of Respiratory Medicine, National Institute for Health Research, Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK.*

*3Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK.*

*4Institute for Lung Health, Department of Infection, Immunity and Inflammation, University of Leicester, Glenfield Hospital, Leicester, UK*

*5Centre for Infection and Immunity, Queen’s University of Belfast, Belfast, UK*

*6Respiratory Medicine, Birmingham Heartlands Hospital and University of Birmingham, Birmingham, UK*

*7Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, UK*

*8Human Development & Health & Clinical and Experimental Sciences, Faculty of Medicine and National Institute of Health Biomedical Research Centre, Southampton, University of Southampton, Southampton, UK.*

*9Division of Infection Immunity and Respiratory Medicine, , The University of Manchester, Manchester Academic Health Science Centre, and Manchester University NHS Foundation Trust, Manchester, UK.*

*10The National Heart and Lung Institute, Imperial College, London, United Kingdom.*

*11Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands.*

*12Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, WA.*

*13Data Science Institute, Imperial College London, London.*

*14The University of British Columbia Center for Heart Lung Innovation, St Paul's Hospital Vancouver, Vancouver, BC, Canada.*

*15Division of Respiratory Medicine, University of British Columbia, Vancouver, BC, Canada.*

*16University of Groningen, University Medical Center Groningen, Dept of Pulmonology, GRIAC Research Institute, Groningen, The Netherlands.*

*17Merck & Co Inc, Rahway, NJ, USA*

*18Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Quebec City, Quebec, Canada*

*19National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, University of Leicester, Leicester, UK.*

\* Corresponding author: Ian Sayers: ian.sayers@nottingham.ac.uk

† These authors contributed equally to this study

**Abstract**

**Background** Genetic studies focussing on moderate-severe asthma are limited. We aimed to identity genetic variants associated with moderate-severe asthma and provide initial mechanistic insight using expression analyses in asthma patients.

**Methods** We used a two stage design including; 5,135 moderate-severe asthma cases and 25,675 controls (stage 1) and 5,414 cases and 21,471 controls (stage 2). The effect of signals on all asthma (mild, moderate, severe) and gene expression in asthma patients and controls was investigated.

**Findings** We identified 24 genome-wide significant signals including a large number of signals in innate/adaptive immune response genes. Novel signals included; rs10905284 in *GATA3* (coded allele A, OR 0.90 [95% CI 0.88-0.93] P=1.76x10-10), rs11603634 in the *MUC5AC* region (coded allele G, OR 1.09 [95% CI 1.06-1.12] P=2.32x10-8) and rs560026225 near *KIAA1109* (coded allele GATT, OR 1.12 [95% CI 1.08-1.16] P=3.06x10-9). These signals were reported previously for blood eosinophil counts, pulmonary fibrosis and allergic sensitisation, respectively. The *MUC5AC* signal was not associated when analyses included mild asthma. The rs11603634G allele was associated with increased expression of *MUC5AC* mRNA in bronchial epithelial brush samples and *MUC5AC* mRNA was elevated in bronchial epithelial samples from severe asthma patients.

**Interpretation** To our knowledge, this is the largest study of moderate-severe asthma which suggests substantial shared genetic architecture between mild and moderate-severe disease. We also report for the first time genetic variants associated with risk of developing moderate-severe asthma that regulate mucin production. Finally, we identify potential candidate/causal genes in these loci and provide greater insight into this difficult to treat population.

**Research in Context**

**Evidence before this study:**

We searched the NHGRI-EBI catalog of published GWAS to identify studies that tested the association between genetic variants and asthma (January 2018). We used the search term “asthma” and manually curated the findings to identify studies that focussed to asthma diagnosis to define cases. We examined the original publication and included studies with more than 500 cases and 500 control subjects and considered signals of relevance those that met genome-wide significance (P<5 x 10-8). These previous studies have reported 38 regions associated at genome-wide significance for susceptibility to develop asthma, providing novel insight into disease biology. To date, only two GWAS have specifically investigated moderate-severe asthma, however these studies were limited by numbers of cases (<1,000) due to the challenges of recruitment of these patients.

**Added value of this study:**

To our knowledge, this study is the largest genetic study of moderate-severe asthma to date, with 10,549 cases and 47,146 controls. We identify three novel genome-wide significant genetic associations that implicate; mucin 5, subtypes A and C, tracheobronchial (*MUC5AC*), GATA-Binding Factor 3 (*GATA3*) and *KIAA1109,* an uncharacterised protein. Altered expression of the pathogenic mucin *MUC5AC* potentially contributes to mucus plugging and airway obstruction. GATA3, is a transcription factor linked to the T cell response in asthma and eosinophilia. The *KIAA1109* locus has previously been associated with allergic sensitisation. We also describe and further characterise the contribution of 21 previously described asthma signals to this phenotype including identifying potential candidate/causal genes.

**Implications of all the available evidence:**

The identification of genetic association with variants in multiple genes of the innate/adaptive immune (T2 inflammation) pathways suggest targeting this pathway may represent a therapeutic opportunity in moderate-severe asthma. The association between variants in the *MUC5AC* locus advance accumulating evidence of alterations in the airway epithelium and mucin dysregulation in more severe forms of asthma. These data potentially support the specific targeting of MUC5AC expression and induction in severe asthma. The identification of the *GATA3* and *KIA1109* signals further extend the data that genetic variants in these regions are associated with asthma potentially via eosinophilia and allergic sensitisation, two drivers of asthma that are also important in moderate-severe disease.

**Introduction**

Asthma is a common disease, and was recently identified as the most prevalent chronic respiratory disease 1,2. Severe asthma accounts for 10-15% of individuals with asthma and there is a significant unmet clinical need, with symptoms including debilitating breathlessness, associated frequent exacerbations and increased hospital admissions despite the high use of medicines3. Both genetic and environmental factors contribute to disease risk with genetic factors thought to account for 40–80% of the susceptibility to develop asthma4. Previous genome-wide association studies (GWAS) in asthma have identified 38 regions of association with asthma including signals in or near; *PEX14, IL6R, PYHIN1* (*African American only),* *ADAMTS4, CD247, TNFSF18, DENND1B, ADORA1, ID2, IL1RL1/IL18R1, D2HGDH, LPP, TLR1, USP38 (Japanese only), PDE4D, TSLP/WDR36, RAD50/IL13, NDFIP1, GPX5, HLAC/NOTCH4/HLADRB1/HLADQA1, GRM4, BACH2, CDHR3, SLC30A8 (Japanese only), ZBTB10, IL33, EQTN, GATA3, LRRC32, IKZF4, STAT6, RAD51B, RORA, SMAD3, CLEC16A, ERBB2/GSDMB/ORMDL3, ZNF652* and *IL2RB*5-20*.* There is significant overlap of asthma signals with signals reported in GWAS of self-reported allergy and allergic sensitisation, e.g. *TLR1, WDR36, IL1RL1, SMAD3,* STAT6, *C11orf30, IL1RL1, TLR1*21,22. More recently in a GWAS of allergic disease and asthma, the authors showed a genetic correlation between asthma and allergic disease with evidence of specific loci being unique to asthma e.g. *ORMDL323.*

Interestingly, the concept of shared genetic origins has recently been tested by a large GWAS involving 180,129 cases (asthma, allergic rhinitis or atopic dermatitis) and 180,709 controls24. This study identified 99 genetic susceptibility loci including 136 independent signals spanning genes involved predominantly in immune function with only six signals showing some disease specificity, e.g. *ORMDL3* for asthma.

In the first GWAS in severe, difficult to treat asthma utilising The Epidemiology and Natural History of Asthma: Outcomes and Treatment Regimes (TENOR) cohort, associations with known asthma loci were found, for example with single nucleotide polymorphisms (SNPs) in the *RAD50-IL13* and *HLA-DR/DQ* regions, although no signal reached genome-wide statistical significance criteria25. Similarly, the Asthma UK Genetics of Severe Asthma (AUGOSA) study of moderate-severe asthma replicated the known 17q21 association at *ORMDL3/GSDMB/ZPB2*19. Neither study identified any new signals meeting genome wide significance potentially because the number of cases was small. Therefore, we aimed to perform a large GWAS of moderate-severe asthma to address three specific objectives. Firstly, we aimed to identify novel variants predicting disease risk for moderate-severe asthma (as opposed to mild asthma). Secondly, we wanted to see if the previously described asthma signals were specifically associated with moderate-severe asthma. Thirdly, we undertook initial functional studies using the U-BIOPRED integrated asthma patient genomics resource in order to translate genetic findings into disease mechanisms, which might in turn identify new targets for therapeutic intervention.

**Methods**

***Study design and participants***

A two stage design was used to identify novel genome-wide significant associations with susceptibility to moderate-severe asthma (figure 1, table 1). To maximise power, in stage 1, a GWAS using cases and controls was performed with variants showing association (P<10-6) further analysed in an independent cohort of cases and controls (stage 2). When meta-analysing stages 1 and 2 we defined statistical significance as variants that met a genome-wide significant threshold of P<5x10-8.

Stage 1 included moderate-severe asthma subjects with European ancestry recruited from primary and secondary care settings across the United Kingdom as part of the Genetics of Asthma Severity & Phenotypes (GASP) initiative with additional cases included from the U-BIOPRED asthma cohort and UK Biobank May 2015 genetic data release (appendix p1). Moderate-severe asthma patients were identified in GASP and U-BIOPRED from clinical records indicating patients were taking medication required in patients defined as moderate, severe asthma as per the British Thoracic Society (BTS) guidelines (2015) (i.e. step 3 and above). In UK Biobank cases were defined by (i) having doctor diagnosed asthma, (ii) taking medication for asthma, (iii) exclusion of doctor diagnosed emphysema/chronic bronchitis and (iv) meeting the definition of moderate-severe asthma using British Thoracic Society (BTS) criteria. Cases were therefore selected from individuals for whom medication information was available and who met BTS stage 3-5 criteria; i.e. for Stage 3 taking - long acting 2 agonist, plus inhaled corticosteroid (ICS), Stage 4 – higher dose ICS and addition of a fourth drug e.g. leukotriene receptor antagonist, SR theophylline, and Stage 5 – oral corticosteroid (OCS) and/or Omalizumab. A complete list of medications to identify moderate-severe asthma is included in the supplementary methods. .Stage 1 and 2 controls were identified from UK Biobank by excluding individuals with doctor diagnosed asthma, rhinitis, eczema, allergy and emphysema/chronic bronchitis. Additional controls were included from U-BIOPRED. A case-control ratio of 1:5 was chosen to balance power and computational time. Cases and controls were matched across age and sex strata and in stage 1 across genotyping arrays. Stage 2 cases and controls were selected from UK Biobank full release as described. All cohorts included individuals with self-reported European ancestry, individuals of non-European ancestry were excluded using PC analysis.

By excluding allergic disease from the control subjects in stage 1 and 2 we could potentially have inflated shared genetic signals related to allergic comorbidities in the asthma population. Hence, we also completed a sensitivity analyses (stage 2a) which included moderate-severe asthma cases from stage 2 and controls from UK Biobank with rhinitis, eczema and allergy not excluded.

All studies had appropriate ethics approval.

***Procedures and statistical analyses***

Moderate-severe asthma cases and controls in stage 1 were genotyped using the Affymetrix Axiom® UK BiLEVE array26 and the later generation Affymetrix Axiom® UK Biobank array, which are 95% identical in content; stage 2 used only the latter array (Affymetrx, Santa Clara, CA, USA). Genotyping and imputation procedures are described in the appendix (p1). After imputation (UK10K + 1000 genomes Phase 3), 33,771,858 single nucleotide polymorphisms (SNPs) were available for association testing in stage 1 (appendix p2).

For stage 1, a GWAS was performed for moderate-severe asthma susceptibility using a logistic model of association and assuming an additive genetic model of asthma status with imputed genotype dose fitted using SNPTEST v2.527 adjusted for 10 ancestry principal components. Independent variants reaching a threshold of P<10-6 in stage 1 association testing were followed up in stage 2. Conditional analyses using GCTAv1.26.028 were used to identify additional independent signals in the same loci and a sensitivity analysis was performed to check for array effects.

Variants with the same direction of effect in stage 1 and 2, that showed a genome wide significant association (P<5 x 10-8) in the meta-analyses and P<0.05 in stage 2 were included in downstream analyses. As we had excluded atopy from the controls we also completed a sensitivity analyses using stage 2 cases and controls with rhinitis, eczema, allergy not excluded (stage 2a and stage 1 + 2a).

 To investigate if novel variants associated with moderate-severe asthma showed an association with susceptibility to all asthma (mild, moderate, severe) we interrogated a large GWAS of asthma, including; 28,399 cases (self- reported asthma) and 128,843 controls5. We also investigated SNPs previously associated with i) asthma5-20 and ii) allergic diseases23,24 at genome-wide significance (P < 5×10‑8) in our stage 1 dataset.

For all signals meeting genome-wide significant (P<5x10-8) association with moderate-severe asthma we investigated the lead SNP (and SNPs in Linkage Disequilibrium (LD) r2>0.4) for association with mRNA levels in cells and tissues using expression quantitative loci (eQTL) datasets: lung tissue29 (n=1,110), blood cells30 (n=5,311) using 10% False Discovery Rate (FDR) and an additional five U-BIOPRED eQTL datasets31,32; blood (n=345), sputum (n=91), biopsy (n=84), bronchial brushing (n=117) and nasal brushing (n=75) using 5%FDR (appendix p3). An association signal meeting FDR in any dataset was considered of interest.

For novel signals where an eQTL effect was observed we investigated the expression of proteins in human lung using Protein Atlas to identify relevant airway cell types for further study33. We also investigated mRNA levels in severe asthma patients using available Gene Expression Omnibus (GEO) datasets (appendix p3) GSE43696, which includes data for bronchial epithelial cells from 20 control, 50 mild and 38 severe asthma subjects34, and GSE89809 which includes data for 18 control, 14 mild, 13 moderate and 11 severe asthma subjects35. There were some differences in the definition of asthma severity in these studies, in GSE43696 mild/moderate-severe asthma patients had a FEV1 of >60% predicted, with/without low–moderate dose inhaled ICS and severe asthma subjects had continuous use of high-dose ICS and/or frequent use of oral CS with continuing symptoms and/or chronic airflow limitation. In GSE89809 mild patients with asthma were taking 2-agonists alone, those with moderate asthma were taking ICS and subjects with severe asthma had persistent symptoms despite high-dose ICS (n =14) and oral corticosteroids (n=4). Both of these studies included medication use to define moderate-severe asthma providing confidence that comparable definitions were used between the GWAS and these downstream analyses.

To provide initial mechanistic insight we investigated lead SNPs (and SNPs in LD r2>0.4) at the novel loci using the HaploReg v4 resource36 and using the deep-learning functional prediction resource DeepSEA (deep learning–based sequence analyzer)37. We used GRASP38 and GWAS catalog39 to identify if any variants in LD (r2 > 0.4) with the 3 novel signals had previously been reported in GWAS of other traits.

***Role of funding source***

The funders had no role in the collection, analysis, or interpretation of the data, and in the writing of the report. IS and LVW were involved in all stages of the study design and delivery, had access to all data and had final responsibility for the decision to submit for publication.

**Results**

In stage 1, 32 independent signals were associated with susceptibility to moderate-severe asthma (P<10-6) with 21 of these meeting genome-wide significance (P<5x10-8) in stage 1 (figure 2, table 2, appendix suppl. figure 1). Array sensitivity analyses did not identify any array effects (appendix suppl. figure 2). The 32 signals included independent secondary signals at the *TSLP*, *IL13/RAD50* and *HLA* loci (appendix suppl. figure 3).

All 32 signals showing association (P<10-6) in stage 1 (appendix suppl. figure 3), including 11 potentially novel signals, were further analysed in stage 2 using an independent cohort of 5,414 moderate-severe asthma cases and 21,471 controls (table 1, appendix suppl. table 1). In stage 2 analyses, 26 signals showed consistent direction of effect and P<0.05 for association (appendix suppl. table 1).

Following meta-analyses of stage 1 and stage 2 we identified 25 signals that showed consistent direction of effect and overall genome wide significance (table 2). The rs61816761 (Filaggrin, *FLG*) signal was excluded following sensitivity analyses as the stage 2 signal was at least in part driven by atopy (suppl. table 2).

Of the 24 signals, three are novel for asthma (figure 3) and include sentinel SNP rs10905284 in *GATA3* (coded allele A, OR 0.90 [95% CI 0.88-0.93] P=1.76x10-10) and rs11603634 in the *MUC5AC* region (coded allele G, OR 1.09 [95% CI 1.06-1.12] P=2.32x10-8). The third signal included an insertion/deletion (rs560026225, proxy in stage 2 rs72687036) in a locus covering *KIAA1109* (coded allele GATT, OR 1.12 [95% CI 1.08-1.16] P=3.06x10-9). rs10905284 (*GATA3*) is a novel signal for asthma independent from previously described signals in the *GATA3* region for asthma, rs1050837215, rs258956111 and rs124135785. We identified a second signal in the *GATA3* region (confirmed by conditional analyses), rs61840192 (labelled *LOC101918272* as distal from *GATA3*, table 2) which is in LD with the previously described signals at rs12413578 (r2=0.166) and rs2589561 (r2=0.161). Both the *KIAA1109* and *GATA3* signals met the threshold for independent replication, however *MUC5AC* signal did not meet a Bonferroni corrected threshold (P<0.005, based on 11 potentially novel signals) but was nominally significant (P=0.018).

We systematically assessed all previously described signals associated with asthma published to date5-20 (appendix suppl table 3). Investigation of 80 previously reported SNPs identified 60 SNPs showing association with moderate-severe asthma (P<6 x 10-4 Bonferroni correction). A further 10 SNPs showed nominally significant association at P<0.05 in our stage 1 dataset (appendix suppl table 3). Similarly, we investigated previously reported SNPs associated with allergic diseases in two recent large studies (appendix suppl table 4 and 5). For the 136 signals identified in the study of Ferreira et. al.24, 87 signals had a P<0.05 for moderate-severe asthma with 40 signals meeting correction (Bonferroni). Similarly, for the 38 signals identified in the study of Zhu et. al.23, 35 signals had a P<0.05 for moderate-severe asthma with 28 signals meeting correction (Bonferroni).

In order to identify if the novel signals for moderate-severe asthma were also associated with mild asthma we undertook a lookup of these signals in a published GWAS of all asthma from the personal genetics company 23andMe, Inc. that was independent of our dataset5 (appendix suppl. table 6). Both the *KIAA1109* and *GATA3* signals were significant (P<0.017, Bonferroni correction): *KIAA1109 (*proxy rs72687036 coded allele G, OR 1.06 [95% CI 1.04-1.08] P=3.96 x10-7) and *GATA3* (rs10905284, risk allele C, OR 1.04 [95% CI 1.02-1.06] P=2.75x10-5). An association between the *MUC5AC* signal and all asthma was not identified (rs11603634, coded allele G, OR 1.00 [96% CI 0.97-1.02] p=0.809).

GRASP and GWAS catalogue analyses of the three novel signals, and variants in LD (r2>0.4) identified genome-wide significant (p<5 x 10-8) associations; rs560026225 (*KIAA1109*) asthma risk allele (GATT) (allele associated with increased risk of moderate-severe asthma) associated with risk of allergic sensitisation (proxy rs17454584, with the rs560026225 GATT allele correlated with rs17454584 G allele)22, the rs10905284 (*GATA3*) asthma risk allele (C) associated with increased number of eosinophils in the blood40 and rs11603634 (*MUC5AC*) asthma risk allele (G) (proxy rs4077759, rs11603634 (G) allele correlated with rs4077759 T allele) was associated with risk of pulmonary fibrosis41 (appendix suppl table 7).

Next, a series of eQTL analyses were completed (appendix suppl table 8), of the three novel signals, rs11603634 (*MUC5AC*) was an eQTL for *MUC5AC* in bronchial epithelial brush cells (via proxy SNP rs11602802, r2=0.46). The rs11603634 asthma risk allele (G) was correlated with rs11602802 (A) allele, which was associated with elevated *MUC5AC* mRNA (figure 4A). While not meeting 5%FDR, *MUC5B* mRNA levels showed an opposite relationship with the rs11602802 SNP (appendix suppl figure 4). Similarly, proxy SNP rs17454584 (r2=0.57) for rs560026225 (*KIAA1109*) was an eQTL for *KIAA1109* in lung tissue with the asthma risk allele (GAATT) associated with reduced expression of *KIAA1109* (rs560026225 GAATT allele correlated with rs17454584 (G) allele) (figure 5A). No significant eQTL was observed for the rs10905284 (*GATA3*) signal. We identified significant eQTL associations for 16 of the 21 previously reported asthma signals in the lungs and/or blood providing additional insight (appendix suppl table 8). More specifically we identified a large number of potential candidate/causal genes in the T2 inflammatory pathway including adaptive and innate immune response genes; *CD247, IL1RL1*, *IL18R1, TSLP, HLA* genes, *BACH2, IL33* and *STAT6* and also genes that may be important in airway structural cell homeostasis/integrity/function*; MUC5AC, D2HGDH, ING5, WDR36, RAD50, SLC22A5, SMAD3, ORMDL3, GSDMA* and *GSDMB.*

MUC5AC and KIAA1109 expression is present in airway epithelium and was localised to the cytoplasm/membrane (appendix suppl figures 5 & 6), therefore we hypothesised that levels of *MUC5AC* and *KIAA1109* may be altered in airway epithelium in severe asthma patients. Elevated levels of *MUC5AC* mRNA were apparent in severe asthma patients (fig 4B & C). We did not identify differential expression of *KIAA1109* in the same bronchial epithelial datasets (fig 5B & C).

We used HaploReg and DeepSEA to investigate the *MUC5AC*, *KIAA1109* and *GATA3* sentinel SNPs and SNPs in LD (r2>0.4). Using HaploReg we identified a large number of potentially functional consequences of these genetic variants (appendix suppl tables 9-11). In the *MUC5AC* locus multiple SNPs including the lead SNP (rs11603634) altered Fox family transcription factors (appendix suppl table 9). For the KIA1109 and GATA3 signals, many potentially functional changes were apparent (appendix suppl table 10 & 11). Using DeepSEA, the moderate-severe asthma risk allele [G] at the sentinel SNP rs11603634 near MUC5AC was predicted to result in a more than 1.5 log2fold change in function at FOXA1 and FOXA2 transcription factor binding sites in airway epithelium (appendix suppl. table 12). For the rs10905284 (*GATA3*) asthma risk allele (C) (proxy rs3802597, r2=0.93) had an effect on USF1 and USF2 transcription factor binding in various cell types including airway epithelium (appendix suppl. table 13). The rs560026225 (*KIAA1109*) (proxy rs17389644, r2=0.57) had a functional effect on a DNase hypersensitivity site in HMVEC and HUVEC (appendix suppl. table 14).

**Discussion**

To our knowledge, we present the largest genetic association study of moderate-severe asthma to date with 10,549 cases and 47,146 controls and identify 24 signals reaching genome-wide significant association. 21 of the 24 signals have previously been reported in studies predominantly using mild asthma patients suggesting there is substantial shared genetic architecture between mild and moderate-severe asthma. Our findings potentially suggest additional factor(s) drive development of more severe forms of asthma e.g. environmental exposures/epigenetics and the presence of comorbidities. We also provide greater insight identifying potential casual/candidate genes in many of these loci including several genes related to T2 inflammation. Importantly, three signals have not previously been reported for asthma in GWAS: rs11603634 in the *MUC5AC* region (coded allele (G), frequency 50.4%, risk), rs1090584 in *GATA3* (coded allele (A), frequency 57.06%, protective) and rs560026225 (coded allele (GATT), 23.6%, risk) in a locus covering *KIAA1109.* The rs11603634 signal is specific to moderate-severe asthma and we identify that *MUC5AC* is elevated in bronchial epithelial cells of risk allele carriers, with *MUC5B* showing a decrease in risk allele carriers, albeit not reaching our significance threshold. This may be via alterations in FOXA transcription factor activity. For the *GATA3* and *KIAA1109* signals we see association with all asthma and there are previously reported associations for blood eosinophil levels and self-reported allergy respectively. Therefore, we provide a further insight into the genetic architecture of moderate-severe asthma and we report the first evidence that genetic variants associated with risk of developing moderate-severe asthma regulate mucin production.

In the current study we identified 21 previously reported signals including those associated with severe or difficult-to-treat asthma previously; *RAD50* and *HLA-DR/DQ*25 and *17q21* (*ORMDL3/GSDMB/ZPB2*)19.Using eQTL analyseswe identified candidate/causal genes; however it is important to note that the level of evidence for each potential candidate gene based on LD with sentinel SNP, relevant tissue/cell type and statistical significance varied from highly supportive to suggestive. Overall these data highlight the role of innate/adaptive immunity and T2 inflammation in moderate-severe asthma including; *CD247* which encodes T-cell receptor zeta, *GATA3,* an important transcription factor in T cells, *IL18R1* and *IL1RL1,* receptors for key cytokines IL18 and IL33 respectively, *TSLP,* that drives T2 inflammation, HLA genes encoding the major histocompatibility complex (MHC), *BACH2*, a transcriptional regulator in T2 inflammation, *IL33*, an innate cytokine, *STAT6*, a signalling molecule downstream of IL4/13 which are drivers of T2 inflammation. The other genes identified highlight roles in airway cell homeostasis; *D2HGDH* regulates alpha-ketoglutarate levels influencing histone and DNA methylation42, *SLC9A2*, a sodium-hydrogen exchanger involved in the regulation of cell pH and volume43, *ING5* a transcription factor involved in epithelial to mesenchymal transition44, *RAD50*, involved in DNA double-strand break repair, *SLC22A5*, an organic cation transporter with a role in epithelial cells, *CLEC16A*, a regulator of autophagy45 and *GSDMB* and *ORMDL3* linked to airway smooth muscle contraction and airway remodelling respectively46,47.

 Importantly, this is the first report of the *MUC5AC* locus being specifically associated with increased susceptibility to development of moderate-severe asthma in GWAS. We show association in stage 1 and stage 2 and in the meta-analyses, however the more severe nature of the asthma patients (and associated power) in stage 1 vs. stage 2, may explain the different significance levels across stages, i.e. stage 1 subjects had lower lung function and reported higher use of oral corticosteroids compared to stage 2 case subjects (6 vs. 3%, Table 1). Stage 1 and 2 cases did not differ significantly by % of allergic comorbidities (allergic rhinitis and/or eczema) (P=0.51) or smoking history (P=0.21). The association between the asthma risk allele (G) of rs11603634 and elevated *MUC5AC* mRNA levels in bronchial epithelial brush samples and the predicted effect on FOXA transcription factors provides a putative mechanisms as FOXA2 regulates MUC5AC production48. This eQTL association is the most significant for *MUC5AC* in the bronchial epithelial cell dataset. This moderate-severe asthma signal has also been identified as associated with pulmonary fibrosis (proxy rs4077759, r2=0.42)41, however rs11603634 reported in the current study and rs35705950 (the main IPF signal) are not in LD (r2=0.01), suggesting distinct signals. Similarly, the signal reported in the current study is independent from that reported (rs1132440) in a candidate gene study for asthma49. We also identified an effect of rs11603634 on *MUC5B* mRNA levels in bronchial epithelial brush samples with risk allele carriers having lower levels, however this did not meet the significance threshold. MUC5AC protein levels are elevated in the sputum of asthma patients during exacerbations compared to stable asthma and controls, while MUC5B are lower, i.e. alterations in the ratio of MUC5AC/MUC5B are a feature of asthma50. This may be partially explained by the opposite effect of our novel asthma risk allele on MUC5AC and MUC5B production. MUC5AC has pathogenic roles and has been linked to airway hyper-responsiveness and mucus plugging, during exacerbation51. *MUC5AC* deficient mice develop allergic airway disease, however the severity and abundance of mucus plugging is attenuated52. Loss of MUC5B leads to airway inflammation suggesting a role in homeostasis53. Overall, these data suggest targeting of specific mucins may be a therapeutic opportunity for moderate-severe asthma.

The *KIAA1109 (*rs72687036) novelsignal has previously been associated with self-reported allergy21 and allergic sensitization22 type 1 diabetes54, ulcerative colitis55, mean platelet volume40 and recently with allergic disease (asthma/hay fever/allergic rhinitis or eczema)23. The region is rich in candidate genes, e.g. *IL2* and *IL21,* however our eQTL data suggests that the potential causal gene is *KIAA1109*. Little is known about *KIAA1109*, mice deficient in *KIAA1109* have pre-weaning lethality (International Mouse Phenotyping Consortium) and a suggested role in synaptic vesicle recycling in *Drosophila* has been identified56.

The third novel signal, rs1090584 in *GATA3*, asthma risk allele (C) is also associated with Rheumatoid Arthritis (proxy rs3824660, r2 0.86)57 and elevated levels of blood eosinophils40, a known effector cell in asthma. More recently, this signal was associated with allergic disease (asthma/hay fever/allergic rhinitis or eczema)23. The rs1090584 signal in *GATA3* identified in the current studyis independent to those previously described for asthma including; rs10508372 (Japanese)15, rs2589561 (European/Multiancestry)11 and rs12413578 (European)5. The second signal in *GATA3* (rs61840192) we report is in LD with rs12413578 and rs2589561 (r2=0.16). These data suggest that multiple genetic signals within the *GATA3* locus may contribute to asthma. We identified potential effects of rs1090584 on upstream stimulatory factors (USF) 1 and 2 in airway cells, however did not identify an eQTL. USF1 is important in regulating GATA family genes including GATA558. *GATA3* is a transcriptional regulator associated with differentiation *e.g*. in Type-2 innate lymphoid cells (ILC2) differentiation59 an effector cell in T2 inflammation.

We also investigated all previously reported asthma signals, with general replication of previous signals. Those signals that did not replicate were associated with a specific asthma phenotype, e.g. *PDE4D* and mild-moderate childhood asthma with bronchial hyper-responsiveness16 or were reported in non-European ancestry e.g. *NOTCH4*15. The effect sizes for previously described asthma signals in this moderate-severe asthma study (OR range 1.09-1.24) are comparable with those reported in large studies of asthma5,11. Similarly, we investigated all recently described allergic disease signals identified in two large GWAS23,24, These data showed that a large proportion of these signals were associated with moderate-severe asthma, however it is important to note that our case-control design using moderate-severe asthma cases and controls (excluding individuals with doctor diagnosed asthma, rhinitis, eczema, allergy) will enhance to the ability to identify genetic signals associated with allergic comorbidities in the asthma population.

While the current study represents the largest study of moderate-severe asthma to date, several limitations require comment. Regarding the design of the study, we considered alternative approaches when planning the analysis, e.g. using mild asthma patients as the control group. However, we opted to compare moderate-severe asthma with healthy controls in the initial discovery analysis as we felt that this would minimise the risk of misclassification between mild and moderate asthma and so be a more powerful strategy for a genetic study. We included the 23andMe analysis to specifically address the issue of specificity of the signals to moderate-severe asthma. Stage 1 cases likely represent a more severe phenotype than those in stage 2 as GASP and U-BIOPRED cases in stage 1 were recruited from secondary care. This may have contributed to the attenuated association with *e.g.* the MUC5AC signal in stage 2 however potentially provided the power for stage 1. We also acknowledge that we defined asthma severity based on medication use alone and additional measures including symptoms, exacerbation frequency and other markers would have enhanced the definition although these data were not available. Importantly all cases in our analyses required a doctor diagnosis of asthma for inclusion prior to stratification based on medication. Severe asthma is defined by the requirement for high dose ICS or maintenance OCS with persistent poor control or at high risk of developing poor control if these therapies are stepped down. Similarly, the GINA treatment steps are used as surrogates of severity. Thus, using the treatment step as a measure of severity is in keeping with current guidelines and has the advantage that it provides maximum sample sizes for this study. We acknowledge the limitations that in the cases that were prescribed OCS we did not undertake individual case review to exclude the possibility that some subjects might have been receiving OCS for other co-morbid conditions and that we did not record additional information on current asthma control nor exacerbation frequency which would have been informative. Advances in imputation mean that the widely used threshold for genome-wide significance of P<5x10-8, historically defined on the basis of 1 million independent tests, could be considered too lenient for an analysis of 33 million SNPs (representing more than 1 million independent tests). Had we applied a threshold of P<5x10-9, recommended as an appropriate threshold for studies of whole genome sequence data from European populations60, the signals at *GATA3* and *KIAA1109* would still have been significant. Overall, the *MUC5AC* signal was weaker in terms of statistical significance compared to other signals identified. Finally, asthma is a complex disease involving both genetic and environmental influences and we have not formally assessed the role of the environment, which may be critical for the development of more severe asthma. Accumulating data suggesting a role for several factors in the development of severe disease; e.g. comorbidities such as atopic dermatitis in severe asthma61 and environmental/epigenetic mechanisms62 however more work is needed here.

In summary, to our knowledge, we present the largest GWAS of moderate-severe asthma published to date and i) identify that the genetic architecture is similar to mild disease, ii) report three novel genome-wide significant associations and iii) identify potential causal/candidate genes underlying these signals. These findings provide new understanding for this difficult to treat asthma population and adds to the accumulating evidence that strategies to target mucins may have therapeutic value. Similarly, our findings add to evidence that targeting T2 inflammation in asthma may be particularly useful for moderate-severe asthma potentially in carriers of genetic variants in genes of relevance to innate/adaptive immunity.

**Figure Legends**

**Figure 1: Flow chart showing quality control and sample selection**

GASP = Genetics of Asthma Severity & Phenotypes, U-BIOPRED = Unbiased BIOmarkers in PREDiction of respiratory disease outcomes. See appendix (p1-3) for more details of sample selection.

**Figure 2: Manhattan Plot for the Stage 1 analyses of moderate-severe asthma risk**

Data pertains to 5,135 moderate-severe asthma cases and 25,675 controls assessed for 33.6 million well-imputed variants, P value presented have genomic control applied. Shown in red are signals meeting criteria for stage 2 (P<10-6) and dotted line is genome-wide significant (P<5x10-8). Loci are labelled with the nearest gene. Quantile-quantile plot for this analysis is shown in the appendix (suppl. Figure 1).

**Figure 3: Regional association plots of novel signals *KIAA1109*, *GATA3* and *MUC5AC* associated with moderate-severe asthma**

Regional association plots from Stage 1 analyses for the three novel signals that show replicated association in Stage 1 and 2 and met genome-wide significance in the meta-analyses. Statistical significance of each SNP on the −log10 scale as a function of chromosome position (NCBI build 37). The sentinel SNP at each locus is shown in blue; the correlations (r2) of each of the surrounding SNPs to the sentinel SNP are shown in the indicated colours.

**Figure 4: rs11603634 is an eQTL for *MUC5AC* in bronchial epithelial brush samples and *MUC5AC* mRNA expression is elevated in bronchial epithelial cells from severe asthma patients**

1. Bronchial epithelial brush samples (n=117) were collected as part of the U-BIOPRED study, DNA was extracted and genotypes generated using the Affymetrix Axiom® UK Biobank array, RNA was extracted and transcriptomic analysis was performed using the Affymetrix HT HG-U133 1 PM GeneChip. rs11603634 was not directly genotyped, therefore proxy rs11602802 was used. The rs11603634 asthma risk allele (G) is correlated with rs11602802 (A) allele. The graph shows the data for *MUC5AC* mRNA expression stratified by rs11602802 genotype, the box and whiskers show the mean and IQR for each genotype.
2. Bronchial epithelial brush samples were collected from control (n=20), mild (n=50) and severe (n=38) asthma subjects and Agilent Human GE 4×44K V2 Gene Expression data generated and deposited on GEO (GSE43696). Robust Multi-array Average (RMA) data for MUC5AC (Probe A\_24\_P548274) was extracted and the graph represents median, IQR and min/max data stratified by subject group. MUC5AC levels were significantly higher in severe asthma vs. control (Kruskal-Wallace test, \*p<0.05).
3. Bronchial epithelial brush samples were collected from control (n=18), mild (n=14), moderate (n=13) and severe (n=11) asthma subjects and Affymetrix HT HG-U133+ PM GeneChip data generated and deposited on GEO (GSE89809). RMA data for MUC5AC (Probe 217182\_PM\_at) was extracted and the graph represents median, IQR and min/max data stratified by subject group. MUC5AC levels were significantly higher in severe asthma vs. control (Kruskal-Wallace test, \*p<0.05).

See appendix p3 for more details of datasets and analyses.

**Figure 5: rs560026225 is an eQTL for *KIAA1109* in lung tissue and *KIAA1109* expression levels in bronchial epithelial cells from asthma patients**

1. Non-tumour lung tissue was isolated from 1,110 individuals who had undergone lung resection across three centres to generate an eQTL dataset. rs560026225 was not directly genotyped, therefore rs17454584 was used as a proxy. rs560026225 asthma risk allele (GAATT) correlated with rs17454584 (G) allele. The graph shows the data for *KIAA1109* expression for each recruitment centre stratified by rs17454584 genotype, the box and whiskers show the mean and IQR for each genotype.
2. Bronchial epithelial brush samples were collected from control (n=20), mild (n=50) and severe (n=38) asthma subjects and Agilent Human GE 4×44K V2 Gene Expression data generated and deposited on GEO (GSE43696). Robust Multi-array Average (RMA) data for *KIAA1109* (probe A\_23\_P147869) was extracted and the graph represents median, IQR and min/max data stratified by subject group. No significant differences in *KIAA1109* expression levels between groups were observed (Kruskal-Wallace test).
3. Bronchial epithelial brush samples were collected from control (n=18), mild (n=14), moderate (n=13) and severe (n=11) asthma subjects and Affymetrix HT HG-U133+ PM GeneChip data generated and deposited on GEO (GSE89809). RMA data for KIA1109 (probe 1553792\_PM\_at) was extracted and the graph represents median, IQR and min/max data stratified by subject group. No significant differences in *KIAA1109* expression levels between groups were observed (Kruskal-Wallace test).

See appendix p3 for more details of datasets and analyses.

**Tables**

**Table 1: Baseline characteristics of stage 1 and stage 2 samples.**

Stage 1 subjects were recruited from GASP, U-BIOPRED and UK Biobank. Stage 2 subjects were recruited from UK Biobank alone. \*U-BIOPRED was not screened for Rhinitis/Eczema at sample selection. In Stage 1, 6.0% of asthma cases were taking oral corticosteroids (prednisolone) based on available data (n=3,710) and in Stage 2, 3.0% of asthma cases were taking oral corticosteroids based on available data (n=5,414). Oral corticosteroids (prednisolone) use corresponds to severe asthma (BTS Stage 5).

**Table 2: Gene variants with genome-wide significance for moderate-severe asthma.**

Results from case-control analyses for the variants that were significant in stage 1 and stage 2 showing the same direction of effect and reached genome-wide significance in the meta-analysis of stages 1 and 2. MAF corresponds to that from the stage 1 study. The OR per copy of the coded allele is presented. rs1438673 was conditioned on rs1837253, rs1986009 was conditioned on rs3749833, rs776111176 was conditioned on rs9273410. Stage 1 P value has genomic control applied. rsis.ukb - rs number of variant used in Stage 2 using UK Biobank imputed data. rs61816761 (*FLG*) was excluded following sensitivity analyses.

**Acknowledgements**

This research has been conducted using the UK Biobank Resource under application 8389 (lead I.S.). This report is independent research supported by the North West Lung Centre Charity and the National Institute for Health Research Manchester Clinical Research Facility at Manchester University NHS Foundation Trust (Wythenshawe). We acknowledge the Manchester Allergy, Respiratory and Thoracic Surgery Biobank for supporting this project and thank the study participants for their contribution. This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This research used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester.

**Funding**

Genotyping of the stage 1 samples was funded by; Asthma UK Grant to I.S., I.P.H., D.E.S., C.E.B. (AUK-PG-2013-188), Rosetrees Tust (Grant to I.S.), AirPROM, U-BIOPRED (EU-IMI 115010) and an MRC Strategic Award to I.P.H., M.D.T., L.V.W. and Professor David Strachan (MC\_PC\_12010). The U-BIOPRED project is supported through an Innovative Medicines Initiative joint undertaking under grant agreement 115010, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007-2013) and European Federation of Pharmaceutical Industries and Associations companies' in-kind contributions (www.imi.europa.eu). L.V.W holds a GSK/ British Lung Foundation Chair in Respiratory Research. Asthma UK funded the GASP initiative (AUK-PG-2013-188). Recruitment of asthma patients in Manchester was funded by the NIHR and the North West Lung Centre Charity. This work was part funded by the NIHR Nottingham Biomedical Research Centre and NIHR Leicester Biomedical Centre. A.S. is supported by the Manchester Biomedical Research Centre. C.J. was supported by a Medical Research Council Clinical Research Training Fellowship.

**Author contributions**

I.S. and L.V.W. designed and supervised the study; A.P.H., C.K.B., D.S., Z.E.K.P., A.F., T.M.M., A.S., L.H., A.H.M., R.C., N.C.T, J.W.H, G.A.L, P.H.H, R.D., J.H., R.N., A.S., K.F.C., P.J.S., J.D.B, I.M.A., D.D.S., M.V.D.B, D.C.N, I.P.H., and C.E.B recruited and genotyped cases; N.S., M.A.P, C.J., M.S.A, N.B., R.H., J.L A.A. R.J.P., M.O. M.T., I.M.A., Y.K.G. and S.H. performed analyses and experiments; I.S., L.V.W., N.S., M.A.P and C.J. wrote the manuscript with input from all authors.

Access to raw data: N.S., M.A.P, C.J., M.S.A, N.B., R.H., J.L., A.A. R.J.P., M.O. M.T., Y.K.G, I.M.A. and S.H.

**Conflict of Interest**

Dr. Blakey reports personal fees and non-financial support from Boehringer Ingelheim, personal fees and non-financial support from Astra Zeneca, personal fees and non-financial support from Napp, non-financial support from GSK, personal fees from Novartis, personal fees from Teva, non-financial support from Virginia Commonwealth University, outside the submitted work; Prof. Brightling reports grants from AirPROM FP7, grants from Asthma UK, grants from NIHR Biomedical Research Centre, during the conduct of the study; Dr. Chaudhuri reports personal fees and non-financial support from AstraZeneca, personal fees from GlaxoSmithkline, personal fees from Teva, personal fees and non-financial support from Novartis, outside the submitted work; Prof. Chung has received honoraria for participating in Advisory Board meetings of the pharmaceutical industry regarding treatments for asthma and chronic obstructive pulmonary disease and has also been renumerated for speaking engagements; Prof. Djukanovic reports receiving fees for lectures at symposia organised by Novartis, AstraZeneca and TEVA, consultation for TEVA and Novartis as member of advisory boards, and participation in a scientific discussion about asthma organised by GlaxoSmithKline. He is a co-founder and current consultant, and has shares in Synairgen, a University of Southampton spin out company; Prof. Hall reports grants from GSK, grants from BI, outside the submitted work; Prof. Heaney reports grants and personal fees from Astra Zeneca, Circassia, Amgen, Boehringer Ingelheim, Novartis, Glaxo Smith Kline and Teva, outside the submitted work; Prof. Howarth reports personal fees from GSK, grants from Boehringer-Ingelheim, outside the submitted work; Dr. John reports grants from Medical Research Council, during the conduct of the study; Dr. Mansur reports personal fees, non-financial support and other from GSK-AZ-Novartis, NAPP-Sanofi,-Teva- BI- others, outside the submitted work; Prof. Nickle is an employee of Merck & Co; Prof. Sayers reports grants from GlaxoSmithKline plc., grants from AnaptysBio Inc., grants from Boehringer Ingelheim, outside the submitted work; Prof. Shaw reports speaker fees from Novartis, Astra Zeneca and TEVA, and advisory board fees from GSK and Astra Zeneca; Prof. Simpson reports grants from National institute of health Research, grants from North West Lung Centre Charity, during the conduct of the study; Prof. Sin reports grants from Merck, personal fees from Sanofi-Aventis, personal fees from Regeneron, grants and personal fees from Boehringer Ingelheim, grants and personal fees from AstraZeneca, personal fees from Novartis, outside the submitted work; Dr. Singapuri reports grants from AirPROM, grants from Asthma UK, grants from NIHR BRC, during the conduct of the study; Prof. Sterk reports grants from Innovative Medicines Initiative (IMI), during the conduct of the study; Prof. Tobin reports grants from GSK, grants from Pfizer, outside the submitted work; Prof. Wain reports grants from GlaxoSmithKline, grants from Pfizer, outside the submitted work; The remaining authors have nothing further to disclose.

**References**

1. Collaborators GBDCRD. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Respir Med* 2017; **5**(9): 691-706.

2. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma P. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* 2004; **59**(5): 469-78.

3. Jarjour NN, Erzurum SC, Bleecker ER, et al. Severe asthma: lessons learned from the National Heart, Lung, and Blood Institute Severe Asthma Research Program. *Am J Respir Crit Care Med* 2012; **185**(4): 356-62.

4. Ullemar V, Magnusson PK, Lundholm C, et al. Heritability and confirmation of genetic association studies for childhood asthma in twins. *Allergy* 2016; **71**(2): 230-8.

5. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016; **48**(7): 709-17.

6. Ferreira MA, Matheson MC, Duffy DL, et al. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 2011; **378**(9795): 1006-14.

7. Torgerson DG, Ampleford EJ, Chiu GY, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 2011; **43**(9): 887-92.

8. Sleiman PM, Flory J, Imielinski M, et al. Variants of DENND1B associated with asthma in children. *N Engl J Med* 2010; **362**(1): 36-44.

9. Bonnelykke K, Sleiman P, Nielsen K, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014; **46**(1): 51-5.

10. Ramasamy A, Kuokkanen M, Vedantam S, et al. Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. *PLoS One* 2012; **7**(9): e44008.

11. 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.

12. Gudbjartsson DF, Bjornsdottir US, Halapi E, et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009; **41**(3): 342-7.

13. Ferreira MA, 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): 1564-71.

14. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; **363**(13): 1211-21.

15. 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): 893-6.

16. Himes BE, Hunninghake GM, Baurley JW, et al. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am J Hum Genet* 2009; **84**(5): 581-93.

17. Noguchi E, Sakamoto H, Hirota T, et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet* 2011; **7**(7): e1002170.

18. Almoguera B, Vazquez L, Mentch F, et al. Identification of Four Novel Loci in Asthma in European American and African American Populations. *Am J Respir Crit Care Med* 2017; **195**(4): 456-63.

19. Wan YI, Shrine NR, Soler Artigas M, et al. Genome-wide association study to identify genetic determinants of severe asthma. *Thorax* 2012; **67**(9): 762-8.

20. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**(7152): 470-3.

21. Hinds DA, McMahon G, Kiefer AK, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* 2013; **45**(8): 907-11.

22. Bonnelykke K, Matheson MC, Pers TH, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* 2013; **45**(8): 902-6.

23. Zhu Z, Lee PH, Chaffin MD, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat Genet* 2018; **50**(6): 857-64.

24. 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): 1752-7.

25. Li X, Howard TD, Zheng SL, et al. Genome-wide association study of asthma identifies RAD50-IL13 and HLA-DR/DQ regions. *J Allergy Clin Immunol* 2010; **125**(2): 328-35 e11.

26. Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med* 2015; **3**(10): 769-81.

27. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**(7): 906-13.

28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**(1): 76-82.

29. Lamontagne M, Couture C, Postma DS, et al. Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* 2013; **8**(7): e70220.

30. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**(10): 1238-43.

31. Hekking PP, Loza MJ, Pavlidis S, et al. Pathway discovery using transcriptomic profiles in adult-onset severe asthma. *J Allergy Clin Immunol* 2017.

32. Lefaudeux D, De Meulder B, Loza MJ, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. *J Allergy Clin Immunol* 2017; **139**(6): 1797-807.

33. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science* 2015; **347**(6220): 1260419.

34. Voraphani N, Gladwin MT, Contreras AU, et al. An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. *Mucosal Immunol* 2014; **7**(5): 1175-85.

35. Singhania A, Wallington JC, Smith CG, et al. Multi-tissue Transcriptomics Delineates the Diversity of Airway T Cell Functions in Asthma. *Am J Respir Cell Mol Biol* 2017.

36. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016; **44**(D1): D877-81.

37. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 2015; **12**(10): 931-4.

38. Leslie R, O'Donnell CJ, Johnson AD. GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics* 2014; **30**(12): i185-94.

39. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 2017; **45**(D1): D896-D901.

40. 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): 1415-29 e19.

41. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; **45**(6): 613-20.

42. Lin AP, Abbas S, Kim SW, et al. D2HGDH regulates alpha-ketoglutarate levels and dioxygenase function by modulating IDH2. *Nat Commun* 2015; **6**: 7768.

43. Moeser AJ, Nighot PK, Ryan KA, Simpson JE, Clarke LL, Blikslager AT. Mice lacking the Na+/H+ exchanger 2 have impaired recovery of intestinal barrier function. *Am J Physiol Gastrointest Liver Physiol* 2008; **295**(4): G791-7.

44. Zhao QY, Ju F, Wang ZH, Ma XZ, Zhao H. ING5 inhibits epithelial-mesenchymal transition in breast cancer by suppressing PI3K/Akt pathway. *Int J Clin Exp Med* 2015; **8**(9): 15498-505.

45. Tam RC, Li MW, Gao YP, et al. Human CLEC16A regulates autophagy through modulating mTOR activity. *Exp Cell Res* 2017; **352**(2): 304-12.

46. Das S, Miller M, Beppu AK, et al. GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proc Natl Acad Sci U S A* 2016; **113**(46): 13132-7.

47. Chen J, Miller M, Unno H, Rosenthal P, Sanderson MJ, Broide DH. Orosomucoid-like 3 (ORMDL3) upregulates airway smooth muscle proliferation, contraction, and Ca(2+) oscillations in asthma. *J Allergy Clin Immunol* 2017.

48. Lachowicz-Scroggins ME, Finkbeiner WE, Gordon ED, et al. Corticosteroid and long-acting ss-agonist therapy reduces epithelial goblet cell metaplasia. *Clin Exp Allergy* 2017; **47**(12): 1534-45.

49. Johnson L, Shah I, Loh AX, et al. MUC5AC and inflammatory mediators associated with respiratory outcomes in the British 1946 birth cohort. *Respirology* 2013; **18**(6): 1003-10.

50. Lachowicz-Scroggins ME, Yuan S, Kerr SC, et al. Abnormalities in MUC5AC and MUC5B Protein in Airway Mucus in Asthma. *Am J Respir Crit Care Med* 2016; **194**(10): 1296-9.

51. Bonser LR, Erle DJ. Airway Mucus and Asthma: The Role of MUC5AC and MUC5B. *J Clin Med* 2017; **6**(12).

52. Evans CM, Raclawska DS, Ttofali F, et al. The polymeric mucin Muc5ac is required for allergic airway hyperreactivity. *Nat Commun* 2015; **6**: 6281.

53. Livraghi-Butrico A, Grubb BR, Wilkinson KJ, et al. Contribution of mucus concentration and secreted mucins Muc5ac and Muc5b to the pathogenesis of muco-obstructive lung disease. *Mucosal Immunol* 2017; **10**(3): 829.

54. Plagnol V, Howson JM, Smyth DJ, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet* 2011; **7**(8): e1002216.

55. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 2011; **43**(3): 246-52.

56. Verstreken P, Ohyama T, Haueter C, et al. Tweek, an evolutionarily conserved protein, is required for synaptic vesicle recycling. *Neuron* 2009; **63**(2): 203-15.

57. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014; **506**(7488): 376-81.

58. Chen B, Hsu R, Li Z, et al. Upstream stimulatory factor 1 activates GATA5 expression through an E-box motif. *Biochem J* 2012; **446**(1): 89-98.

59. McKenzie AN. Type-2 innate lymphoid cells in asthma and allergy. *Ann Am Thorac Soc* 2014; **11 Suppl 5**: S263-70.

60. 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): 145-51.

61. Lee JK, Han D. Atopic dermatitis is an important comorbidity in severe asthma. *Ann Allergy Asthma Immunol* 2018; **120**(6): 661-2.

62. Perry MM, Lavender P, Kuo CS, et al. DNA methylation modules in airway smooth muscle are associated with asthma severity. *Eur Respir J* 2018; **51**(4).