Probation Review Report:

Discovery and fine-mapping of difficult-to-treat asthma susceptibility loci to identify putative targets for drug development

WTDTP PhD Student:

Noemi Nicole Piga

Supervisor 1:

Dr. Katherine Fawcett, University of Leicester

Supervisor 2:

Dr. Ian Sayers, University of Nottingham

Supervisor 3:

Dr. Glenda Lassi, AstraZeneca

Collaborator:

Dr. Michael Portelli, University of Nottingham

Abstract

Asthma is a heterogeneous disease with symptoms including cough, chest tightness, wheeze and breathlessness. Most patients' symptoms are controlled by existing treatment, however 3-10% of asthmatic individuals struggle to manage their condition despite high levels of therapy; they experience a higher burden of symptoms that affect their physical and mental health. It is imperative to understand the genetic and biological components of severe asthma to improve diagnosis and treatments.

A previous study led by our group discovered 24 genetic loci, including three novel loci, associated with moderate-to-severe asthma using European ancestry participants of UK Biobank[1]. In this study, moderate-to-severe asthma was defined based on self-reported medication data. However, since the publication of this study, UK Biobank has released electronic health records (EHRs), including prescriptions from primary care that provide a more comprehensive record of participants' medication history. Furthermore, whole-exome and whole-genome sequencing data has been released enabling analysis of previously untested rare variants and structural variants. Using this new data, this project aims to identify a subgroup of adult asthma participants with the greatest clinical unmet need through the creation of an ad-hoc phenotype and the use of genome-wide association studies to confirm or discover variants relevant to this more severe asthma type. Bayesian fine-mapping approaches will be applied to identify potential causal variants, and to map these to candidate genes and biological pathways.

To refine the signals of association and identify any population-specific risk factors, I will also extend my analysis to individuals of non-European ancestry through collaborations with other asthma cohorts and population-based biobanks.

Through my second supervisor (Prof Ian Sayers, University of Nottingham), I will have access to a large biobank of samples from asthma patients. Therefore, potential causal variants, genes and biological pathways from the first stage of my PhD can be tested using human airway models derived from patient cells. Finally, placements at AstraZeneca (AZ) (organised by my third supervisor, Glenda Lassi) will enable me to explore the therapeutic potential of my genetic findings.

During this first year, I wrote a literature review on severe asthma, created a phenotype algorithm to extract participants with asthma from UK Biobank using EHRs and identify the subgroup of those taking treatment within stage 4-5 of the British Thoracic Society (BTS) 2019 guidelines[2]. These individuals were used as cases for a genetic association analysis. This probation report will (i) give an inclusive summary of our current knowledge of severe asthma; (ii) outline the steps in my phenotype definition and provide a descriptive analysis of my extracted cases; (iii) show some preliminary genetic results; (iv) describe my plans for the rest of the PhD.

[1] Shrine, N. et al. (2019). Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. Lancet Respiratory Medicine, 7(1), 20-34. https://doi.org/10.1016/S2213-2600(18)30389-8

[2] SIGN Guideline development group and Alexander Mathioudakis (2019). SIGN158: British guideline on the management of asthma. Scottish Intercollegiate Guidelines Network. ISBN 978 1 909103 70 2

Table of Content

Introduction

What is asthma and severity in asthma

Asthma is a non-communicable disease presenting symptom and treatment heterogeneity and researchers struggle to create a well-defined classification for asthma due to its complexity. Important gaps in asthma knowledge includes etiology, biological pathways, genetic risk factors and their interaction with environmental factors.

Worldwide, asthma prevalence accounts for more than 300 million people[1]. In the UK alone, 5.4 million people use asthma treatments (~20% children, ~80% adults) costing 1.1 billion pounds to the National Health Service per year[2].

Severe asthma affects only 3-10% of asthma patients, but has a large impact on life quality and is a disproportionate burden on the health care system[2]. People experiencing this type of asthma show more frequent exacerbations and hospitalisation; it requires high intensity treatment to control the symptoms which can be still present despite therapy. Recurrent asthma events and treatments hit patients both physically and emotionally[3]. New biological treatments (monoclonal antibodies) have increased the number of patients able to control their asthma, but they are not effective for all of them. Improving the understanding of the pathology, endotypes and phenotypes in severe asthma is a priority which can lead to earlier diagnosis and new therapies aiming to change each patient's life for the better.

How genetic study can help

Genome-wide association study, an overview

A genome-wide association study, hereafter named GWAS, is a type of analysis that uses statistical methods to find correlation between human genetic variation (usually single nucleotide polymorphisms, SNPs) across the whole genome and a particular outcome of interest. A GWAS is a type of hypothesis-free testing, as it makes no/few assumptions about which genomic regions are involved in disease. Over the last two decades, GWASs have been used to identify many genetic risk loci associated with non-communicable diseases, such as asthma. These have helped to highlight biological pathways underlying asthma and identify potential new targets for drug development. Many of these results have been collated in the GWAS Catalog, an online and free access resource[4].

Although GWASs are helping to find genetic associations, a huge limitation is the lack of genetic diversity in the datasets used for these studies. The vast majority of genomic research is conducted in individuals of European ancestry and the results are not necessarily transferable into other genetic ancestries[5]. It is possible that the genetic risk factors may be different in other world populations, as a result of genetic background or population-specific environmental factors. Even where the underlying risk factors are the same, results from GWAS may vary due to differences in allele frequencies and patterns of linkage disequilibrium. Efforts like the Global Biobank Meta-analysis Initiative[6] are trying to address this disparity to ensure more equity in human genomic studies.

Another limitation in traditional GWASs is the use of genotyped and imputed genomic data. Historically, participants have been genotyped with micro-array chip technology, which capture primarily common SNPs. The remaining genomic variants are inferred using imputation techniques and reference panels such as Haplotype Reference Consortium panel[7], the UK10K[8] and 1000 Genomes Project (1000GP) Phase

3[9]; they are used to deduce the missing SNPs with probabilistic calculations. However, rare variants are often missed, therefore can't be investigated for association with diseases. Thanks to the advancement of sequencing technology and its cost accessibility, researchers have started to use whole-genome and whole-exome sequencing data which allow the study of rarer variants as well as other type of variants (e.g. structural).

Genetics of asthma

GWASs have helped to unveil loci and genes linked to asthma including more than 200 genetic signals[10]. The first identified locus was 17q12-21.1 including two genes, *ORMDL3* and *GSDMB*[11]. *ORMDL3* is involved in calcium homeostasis and has been suggested to influence airway smooth muscles during hyperresponsiveness events[12]; *GSDMB* encodes gasdermin B that acts in remodeling of the airway[13]. Another associated locus is 7q22.3 containing the *CDHR3* gene, which plays a role in cadherin-mediated cell adhesion[14]. Subsequent studies replicated these two loci and found other associated genes encoding signaling molecules, receptors or transcription factors involved in immune activities[15]. A preprint study led by AstraZeneca reported a whole exome-sequencing association analysis in UK Biobank and confirmed the role of other two genes in asthma, namely *IL33* and *FLG*[16].

GWAS on severe asthma

Genetic studies focused on a severe asthma phenotype highlighted same genomic regions as per all asthma, the 17q12-21.1 and the *IL33* gene locus on chromosome 9[17,18]. The latter is involved in the same biological pathway of *IL1RL1/IL18R1* genes (locus 2q21) which has been found associated with severe asthma as well[18,19] and at least three monoclonal antibodies directed against these interleukins have undertaken Phase II in clinical trials[20] (GSK3772847[21], REGN3500[22], Etokimab[23]). Other genes include *Tumor growth factor-β (TGF- β)* on chromosome 19 especially among studies on adults[18] and *CHI3L1*/*YKL-40* on chromosome 1 positively correlated with airway remodeling[24].

A study led by our group performed a GWAS on European individuals with 'moderate or severe' asthma in UK Biobank[19]. Twenty-four loci were identified, three of which were novel and located in the region of the genes *MUC5AC*, *GATA3* and *KIAA1109*. *MUC5AC* may be involved in mucus plugs, which have a causal role in airway obstruction and exacerbations. *GATA3* is a transcription factor implicated in T-cell response in asthma and eosinophilia. *KIAA1109* was previously found associated with allergic sensitization[25]. The other 21 signals had already been found associated with asthma in other studies and the likely causal genes were involved in adaptive immunity and type 2 inflammation[26].

Gaps in our knowledge and how to address them

Although GWASs have found several associated genetic variants, many of them have a modest effect size on asthma risk[27] with an odds ratio (OR) ~1.2[28]. Genetic factors are predicted to explain 35-70% of the total variance in asthma risk, with childhood onset asthma showing a higher genetic component compared to adulthood onset asthma[29]. Overall, the genetic risk factors identified so far explain a small fraction of the heritable risk of asthma. Possible sources of this missing heritability include interaction of genetic components with environmental factors, such as diet, lifestyle, and air pollution, as well as intrinsic factors like the microbiota[15], rare variants or other type of variants that have not been fully investigated yet, and population-specific risk factors.

Defining severe asthma for GWASs is quite complex due to the lack of a standard definition, information on each prescription's daily dosages and adherence. Each study defines it with similar but yet different terms, looking at specific conditions within the disease (e.g. exacerbation) and thus highlighting distinct genetic involvements.

As already described, a previous study led by our group discovered 24 genetic loci, three of which were novel, associated with moderate-to-severe asthma using European ancestry participants of UK Biobank[19]. Cases were defined using self-reported medications. After the publication of this study, UK Biobank released electronic health records (EHRs) including primary and secondary care records, which could potentially improve the definition of a severe asthma. Moreover, whole-exome and whole-genome sequencing data are now available[30] allowing us to pinpoint rarer variants and structural variants associated with severity in asthma.

Scope of PhD

Building on the previous knowledge on severity in asthma, this PhD project aims to add new information about the genetic risk factors for a severe asthma condition.

I will address five main questions:

- **'What is the state-of-the art in difficult-to-treat/severe asthma and how can I define it in UK Biobank?'**
  Write of a literature review of asthma and severe asthma; investigate the use of available phenotype information such as EHRs in UK Biobank to define cases with higher clinical unmet need.

- **'Which genomic variants and loci are associated with our definition of difficult-to-treat/severe asthma?'**
  Perform an updated GWAS of severe asthma in European ancestry cases in UK Biobank using firstly imputed data and then whole-exome/whole-genome sequencing data.

- **'Among these associated variants, which are the causal ones and what are their likely mechanisms of action?'**
  Post-GWAS analyses (e.g. fine-mapping of associated genomic regions); *in silico* variant-to-gene mapping analyses.

- **'What are the genes and biological pathways impacted by causal variants and how do they affect risk of severe asthma?'**
  As well as the *in silico* approaches described above, I will design *in vitro* experiments to test hypotheses regarding the genes and biological processes involved.

- **'Can we translate these findings into a genetically driven drug discovery?'**
  Explore the potential of new genetic findings to inform drug development using *in silico* resources and during placements in industry.

Progress to date

During this year, I focused on the first and second questions.
In order to have a comprehensive understanding of asthma and severe asthma, I firstly wrote a literature review and scoped out information, such as EHRs, that could be used to identify individuals with asthma and the greatest clinical unmet needs. I wrote an algorithm to extract individuals taking medication indicative of difficult-to-treat asthma and a more severe condition. I was also able to run a case-control GWAS in individuals of European ancestry in UK Biobank, obtaining preliminary genetic results.

Literature Review of asthma and severe asthma

Definition of asthma

The definition and interpretation of the word 'asthma' has changed over time. The name derives from the Greek 'ασθμα' ('asma') which refers to difficulty breathing[31]. In 1894, Sir William Osler identified the inflammatory component present in asthma, and he noticed the interaction with the environment[32]. Classically, asthma has been divided into two groups: allergic/atopic due to external irritant factors and intrinsic/non-atopic due to probable infectious triggers[33].

In more recent years, the Global Initiatives for Asthma (GINA) has collated information about definition, treatment and prevention of asthma. The GINA 2022 guidelines state that asthma is: *"a heterogeneous disease, usually characterized by chronic airway inflammation. It is defined by the history of respiratory symptoms, such as wheeze, shortness of breath, chest tightness and cough, that vary over time and in intensity, together with variable expiratory airflow limitation"*[26].

Specific to the United Kingdom, the British Thoracic Society (BTS) and the Scottish Intercollegiate Guidelines Network (SIGN) have teamed up to create clinical asthma guidelines. In the newest report in 2019 they defined asthma in agreement with GINA's definition adding the notion of airway hyper-responsiveness[34].

Asthma subtypes

Asthma as an umbrella term for multiple conditions rather than for one single disease[27,31]. Studies have tried to define asthma sub-phenotypes based on clinical, demographic and/or pathophysiological features, although these types of classification do not necessarily produce corresponding asthma phenotypes[2].

From a clinical point of view, the Severe Asthma Research Program (SARP) considered three broadly used clinical factors to cluster asthma phenotypes: baseline Forced Expiratory Volume for the first second (FEV_1), FEV_1 post-bronchodilator and age at onset[35]. They clustered asthma into 5 groups differing by age at onset and severity, all presenting atopic asthma except for the third category which is characterized by adult onset and increased prevalence of women with high body mass index (BMI)[31,35].

However, SARP clustering analysis was based on clinical features only, which ignores the heterogeneity in the biological processes impacting development and progression of asthma: it is possible to encounter people presenting the same clinical phenotype whilst having distinct underlying biological aetiologies -called endotypes. For this reason, many researchers promote studies that look also at the biological mechanism and potential causes for asthma. As an example, the Unbiased Biomarkers for the Prediction of Respiratory Disease Outcomes (U-BIOPRED) study has used not only clinical features, but also "omics" data –such as transcriptomics, proteomics, and metabolomics- in order to try to characterize different endotypes of asthma[36]. Up to now, the level of airway inflammation from type 2 immune response is the main discriminator to classify asthma endotypes. Type 2 high refers to consistent presence of eosinophils in the bloodstream; it is typical for either childhood atopic or late onset asthma and the majority of patients improve their health with traditional treatment targeted at components of type 2 inflammation [31] (i.e. corticosteroids). However, some individuals do not achieve control of their symptoms despite these treatments suggesting the involvement of type 2 independent processes[37]. Type 2 high presents adaptive immune system cytokines -such as interleukin 4 (IL4), IL5 and IL13- which inhabit the airways during asthmatic events[31,38]. These biomarkers have become good targets for monoclonal antibodies[31]. Type 2 low asthma is characterized by normal levels of eosinophils and different inflammatory profiles such as neutrophilic or pauci-granulocytic asthma[39]. The neutrophilic endotype presents an immune response with T helper cell type1 (Th1) producing interferon gamma (IFNγ), or T helper cell 17 (Th17) with IL17A[38]. Patients are typically female, smokers with non-atopic and adult-onset asthma[31].

Severe asthma

In 1981, for the first time, Carmichael et al. described severe asthma patients as those resistant to systematic corticosteroids[40]. In 1991, the National Asthma Education and Prevention Program (NAEPP) collated the first guidelines for diagnosis and management of asthma dividing it into intermittent and persistent, which was further divided into moderate and severe[41]. Severe asthma was characterised as the need for high-dose inhaled corticosteroids (ICSs) in combination with a second controller, and ongoing symptoms including reduced lung function and higher propensity to exacerbation despite treatments[42].

In 1993, GINA was launched, and the first report initially classified asthma severity into four steps based on treatment and management of the disease[43]; severe asthma was associated with step 4 and above, also called severe persistent asthma[43]. As the definition has changed over the years[44-46], the 2021 GINA report decided to abandon this step classification method and to refer to severe asthma as a subtype of difficult-to-treat asthma. In the latest release (GINA 2022)[26] difficult-to-treat asthma is defined as *'uncontrolled despite prescribing of medium or high dose ICS with a second controller (usually LABA) [long acting beta agonist] or with maintenance OCS, or that requires high dose treatment to maintain good symptom control and reduce the risk of exacerbations '*[26].

GINA approves and uses the definition of severe asthma by the European Respiratory Society (ERS)/American Thoracic Society (ATS) Task Force[47]. Established in 1997, the ERS/ATS Task Force first defined severe asthma based on clinical history and physiological evidence of intermittent airflow obstruction[47]. In 2014, it published the definition shared by GINA: after confirmation of an asthma diagnosis, severe asthma is present when the trait is *'uncontrolled despite adherence with maximal optimized high dose ICS-LABAs treatment and management of contributory factors, or that worsen when high treatment is decreased'*[26,48].

Finally, GINA interprets as 'uncontrolled' patients showing either poor symptoms control or frequent exacerbations happening two or more times per year, or serious exacerbation requiring more than one hospitalisation per year[26]. The ERS/ATS Task Force defines it in a similar way, adding airflow limitation[48].

Due to their in-depth knowledge and accuracy, GINA and ERS/ATS Task Force represent two milestones for severe asthma definition and management, used by clinicians and researchers. Based on these definitions, GINA estimates that difficult-to-treat asthma patients represent 17% of the asthma population, while severe asthma patients represent 3.7% according to a study of the Netherlands population[49].

Diagnosis process

There is no one single clinical measure to assess the presence of asthma. Typically, patients showing more than one symptom and having airflow obstruction or airway inflammation are more likely to be diagnosed with asthma. Spirometry is a widely used test for airflow obstruction in primary care and clinicians usually look at three measures: forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and FEV_1/FVC ratio. These measures are correlated with other diseases, especially with chronic obstructive pulmonary diseases (COPD). Since lung function can vary in asthma patients, FEV_1 variability is interpreted as the percentage difference between the highest and the lowest forced expiratory flows in an ideal period of two weeks. To test the airway responsiveness, clinicians rely on the change of FEV_1 after inhaling increasing doses of histamine or methacoline. To measure eosinophilic inflammation, clinicians use fractional exhaled nitric oxide (FeNO) –although being aware of several cofounders [34]- or sputum/blood eosinophil count.

The thresholds applied in these tests are based on epidemiological studies and are prone to false positive and false negative findings [34]. For example, some asthma treatments reduce T2 inflammation and so measures of eosinophilic inflammation should be interpreted differently depending on treatments[26].

Severe asthma is characterized by uncontrolled symptoms despite good adherence and high-dose treatment. In case of uncontrolled symptoms, the diagnosis is straightforward; if symptoms are controlled but require high doses of medication, the heaviest drug is removed and control is assessed again. If asthma becomes uncontrolled then the assessment suggests a diagnosis of severe asthma (Figure1).

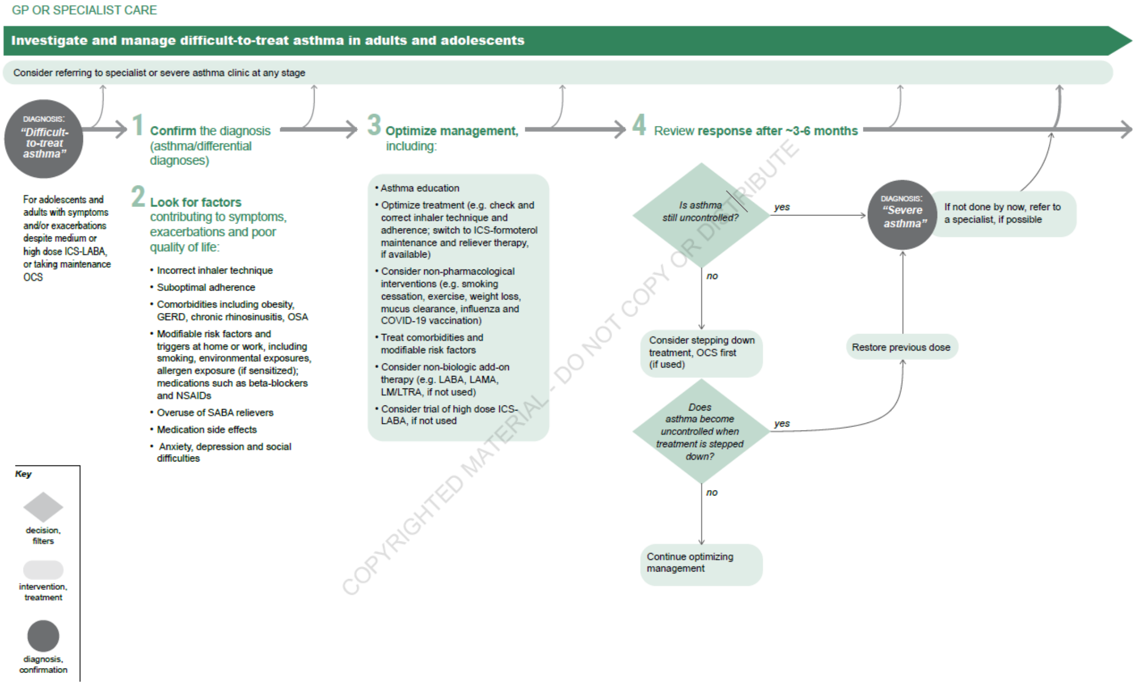

GP OR SPECIALIST CARE
Investigate and manage difficult-to-treat asthma in adults and adolescents
Consider referring to specialist or severe asthma clinic at any stage
DIAGNOSIS: "Difficult-to-treat asthma"
For adolescents and adults with symptoms and/or exacerbations despite medium or high dose ICS-LABA, or taking maintenance OCS
1 Confirm the diagnosis (asthma/differential diagnoses)
2 Look for factors contributing to symptoms, exacerbations and poor quality of life:
• Incorrect inhaler technique
• Suboptimal adherence
• Comorbidities including obesity, GERD, chronic rhinosinusitis, OSA
• Modifiable risk factors and triggers at home or work, including smoking, environmental exposures, allergen exposure (if sensitized); medications such as beta-blockers and NSAIDs
• Overuse of SABA relievers
• Medication side effects
• Anxiety, depression and social difficulties
3 Optimize management, including:
• Asthma education
• Optimize treatment (e.g. check and correct inhaler technique and adherence; switch to ICS-formoterol maintenance and reliever therapy, if available)
• Consider non-pharmacological interventions (e.g. smoking cessation, exercise, weight loss, mucus clearance, influenza and COVID-19 vaccination)
• Treat comorbidities and modifiable risk factors
• Consider non-biologic add-on therapy (e.g. LABA, LAMA, LM/LTRA, if not used)
• Consider trial of high dose ICS-LABA, if not used
4 Review response after ~3-6 months
Is asthma still uncontrolled?
yes
no
DIAGNOSIS: "Severe asthma"
If not done by now, refer to a specialist, if possible
Consider stepping down treatment, OCS first (if used)
Restore previous dose
Does asthma become uncontrolled when treatment is stepped down?
yes
no
Continue optimizing management
Key
decision, filters
intervention, treatment
diagnosis, confirmation

Figure 1.Decision Tree for difficult-to-treat and severe asthma in adult and adolescent in GINA 2022 guidelines[26]

Treatments and the unmet clinical need of severe asthma

Overall, asthma medications can be divided into three categories based on their function:

- *Controller:* they contain ICS and their aim is to reduce airway inflammation.
- *Reliever:* they are divided into ICS-formoterol and SABA, and are prescribed as-needed for quick relief of asthma symptoms.
- *Add-on therapies:* encompass different types, given when high dose of controller medication is unsuccessful

According to both BTS 2019 and GINA 2022 guidelines, treatment should achieve asthma control and take into account the patient's personal preferences[26,34]. GINA 2022 describes asthma control as the absence of mostly all the common symptoms combined with the reduction of disease-related future risks which include mortality, exacerbation, constant airflow limitation and side-effects of treatments[26].

Asthma therapies aim to achieve control by increasing drug doses or actuation per day, or adding medication for a combined treatment. In order to decide the right treatment, the patient needs to attend an assessment visit to understand the current disease condition and severity. Asthma can worsen over time for several reasons including poor adherence, lack of therapy technique, environmental factors, and smoking, therefore new treatments and/or higher doses may be required.

BTS 2019[34]

According to BTS 2019, when in presence of sporadic symptoms, the best option is the prescription of intermittent reliever short-acting bronchodilators, mainly represented by short-acting β_2 agonists (SABA) (Figure2).

In case of more frequent asthma symptoms while under SABA therapy, a regular reliever therapy is prescribed. This regular reliever therapy is an inhaled corticosteroid (ICS), which is usually taken as two puffs twice a day on either a low, medium or high dose. The practitioner should administer the lowest possible dose that ensures symptom control in order to avoid an excessive and superfluous amount of ICS that can increase the risk of side effects. Alternatives to this treatment are available, such as leukotriene receptor antagonists (LTRS), sodium cromoglicate or nedocromil sodium, and theophyllines.

In some cases, ICS alone can be insufficient to control symptoms. In this situation, practitioners will explore the administration of an add-on therapy that is the combination of a low-dose ICS with a LABA. Instead of using two inhalers (ICS and LABA), one single combination inhaler for maintenance and reliever (MART) can be used as well; the important point is that LABA must be taken with ICS and not as a single treatment.

If combined low-dose ICS and LABA is still not enough to control symptoms, ICS dose can be increased to medium or high. Also, other medications can be added on top of ICS-LABA therapy: LTRA, tiotropium (a long-acting muscarinic antagonist, LAMA) or theophylline.

GINA 2022[26]

GINA 2022 guidelines are organized in a slightly different way (Figure3). They no longer recommend SABA only as an intermittent controller therapy, but instead advocate the use of as-needed-only ICS-formoterol. GINA organises the treatment into five steps of medication and two tracks based on the baseline reliever: ICS-formoterol (track 1) or SABA (track 2). Overall, Track 1 is preferred because of the reduced risk of exacerbation compared to SABA; track 2 is used as an alternative when track 1 is not possible, or in the presence of evidence of good adherence, or no exacerbation in the last year of therapy.

In track 1, steps 1 and 2 are combined together for those participants showing symptoms less than 5 days a week. The recommended controller is an as-needed-only ICS-formoterol. Track 2 distinguishes step 1 for symptoms less than twice a month and suggests ICS as controller when SABA is taken; step 2 is for those whose symptoms are more than twice a month but less than 5 days per week and suggests low dose maintenance ICS with SABA reliever.

Step 3 is characterized by symptoms almost every day or waking with asthma once a week. Track 1 suggests a MART therapy with low-dose ICS-formoterol both as reliever and maintenance; track 2 recommends ICS-LABA as maintenance and as-needed SABA reliever.

If daily symptoms, or waking with asthma once or more per week, then step 4 is chosen: for track 1, continuation of the MART therapy with medium dose ICS-formoterol maintenance, and for track 2, ICS-LABA as maintenance at medium/high dose. Short courses of OCS can be needed at step 4 in both tracks if asthma remains uncontrolled.

BTS 2019[34]**and GINA 2022**[26]

At this stage, if patients show persistent uncontrolled symptoms despite optimal treatment and good adherence to triple treatment as per BTS 2019 or step 4 of GINA 2022, they need to be referred to a specialist for an investigation and assessment for severe asthma. They are therefore likely to be treated with add-on long course -more than three months- oral corticosteroids (OCS) as tablets, usually prednisolone/prednisone, per both BTS 2019 and GINA 2022 step 5. However, GINA 2022 defines OCS as the 'last resort': patients need to be monitored for risk of adrenal suppression and corticosteroid-induced osteoporosis. As per GINA 2022 step 5, other add-on therapies include high-dose ICS-containing controller formoterol or LABA based on the track 2, and add-on LAMA.

If the severe asthma phenotype includes type 2 inflammation, both BTS 2019 and GINA 2022 guidelines suggest the use of biologics as another add-on therapy, delivered as intravenous or subcutaneous monoclonal antibodies (mABs). Six different drugs are already in clinical use for eosinophilic and allergic severe asthma, and they target four different pathways[50,51]:

- Anti-IgE: Xolair (Omazilumab);
- IL4-Receptor blockade: Dupixent (Dupilumab);
- Anti-IL5/5R: Nucala (Mepolizumab), Cinqaero (Reslizumab) and Fasenra (Benralizumab);
- Anti-thymic stromal lymphopoietin: Tezspire (Tezepelumab)

Unfortunately, not all severe asthma patients are eligible or respond well to mABs therapy. For instance, monoclonal antibodies targeting type 2 low phenotypes are not in clinical use yet. As an alternative treatment, both guidelines suggest add-on bronchial thermoplasty for adult severe asthma. Such invasive treatment reduces the smooth muscle mass, which is typically increased in asthmatic airway epithelium. Although thermoplasty has shown some beneficial results in the long-term[52], it is still an invasive procedure and prone to adverse respiratory events in the short term.

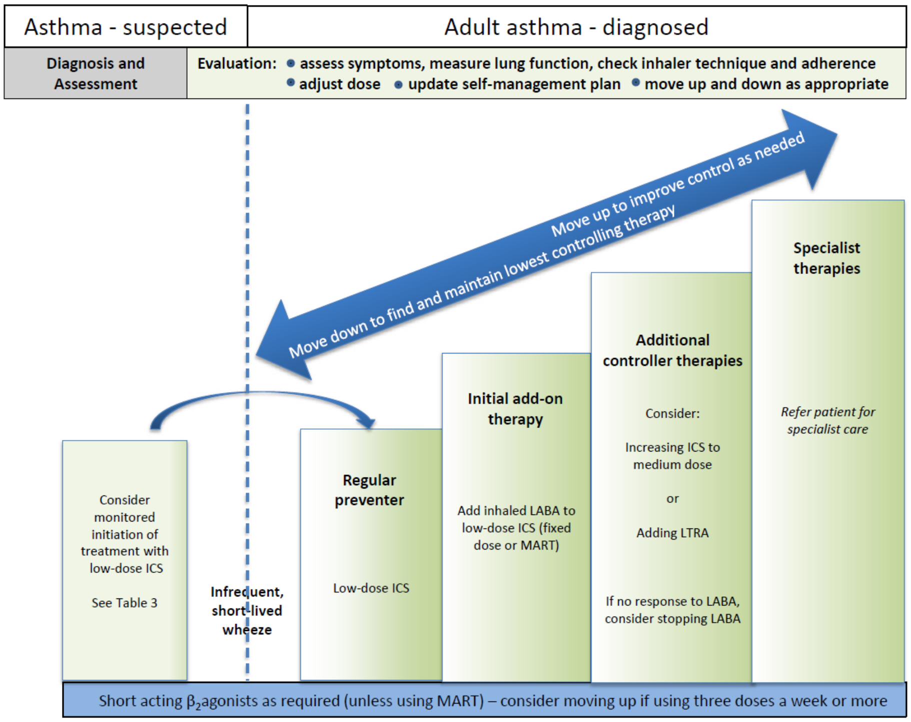
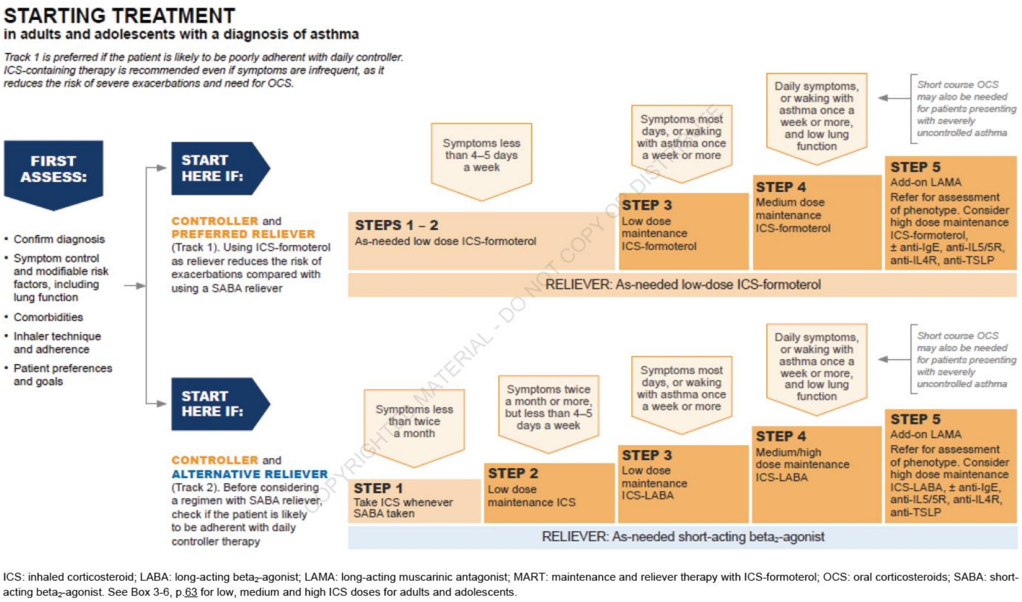

Asthma - suspected
Adult asthma - diagnosed
Diagnosis and Assessment
Evaluation: ● assess symptoms, measure lung function, check inhaler technique and adherence ● adjust dose ● update self-management plan ● move up and down as appropriate
Move up to improve control as needed
Move down to find and maintain lowest controlling therapy
Consider monitored initiation of treatment with low-dose ICS
See Table 3
Infrequent, short-lived wheeze
Regular preventer
Low-dose ICS
Initial add-on therapy
Add inhaled LABA to low-dose ICS (fixed dose or MART)
Additional controller therapies
Consider:
Increasing ICS to medium dose
or
Adding LTRA
If no response to LABA, consider stopping LABA
Specialist therapies
Refer patient for specialist care
Short acting $\beta_2$agonists as required (unless using MART) – consider moving up if using three doses a week or more

Figure 2.Treatment stages for asthma as in BTS2019[34]


STARTING TREATMENT
in adults and adolescents with a diagnosis of asthma
Track 1 is preferred if the patient is likely to be poorly adherent with daily controller. ICS-containing therapy is recommended even if symptoms are infrequent, as it reduces the risk of severe exacerbations and need for OCS.
FIRST ASSESS:
• Confirm diagnosis
• Symptom control and modifiable risk factors, including lung function
• Comorbidities
• Inhaler technique and adherence
• Patient preferences and goals
START HERE IF:
CONTROLLER and PREFERRED RELIEVER (Track 1). Using ICS-formoterol as reliever reduces the risk of exacerbations compared with using a SABA reliever
Symptoms less than 4–5 days a week
STEPS 1 – 2
As-needed low dose ICS-formoterol
Symptoms most days, or waking with asthma once a week or more
STEP 3
Low dose maintenance ICS-formoterol
Daily symptoms, or waking with asthma once a week or more, and low lung function
STEP 4
Medium dose maintenance ICS-formoterol
Short course OCS may also be needed for patients presenting with severely uncontrolled asthma
STEP 5
Add-on LAMA
Refer for assessment of phenotype. Consider high dose maintenance ICS-formoterol, ± anti-IgE, anti-IL5/5R, anti-IL4R, anti-TSLP
RELIEVER: As-needed low-dose ICS-formoterol
START HERE IF:
CONTROLLER and ALTERNATIVE RELIEVER (Track 2). Before considering a regimen with SABA reliever, check if the patient is likely to be adherent with daily controller therapy
Symptoms less than twice a month
STEP 1
Take ICS whenever SABA taken
Symptoms twice a month or more, but less than 4–5 days a week
STEP 2
Low dose maintenance ICS
Symptoms most days, or waking with asthma once a week or more
STEP 3
Low dose maintenance ICS-LABA
Daily symptoms, or waking with asthma once a week or more, and low lung function
STEP 4
Medium/high dose maintenance ICS-LABA
Short course OCS may also be needed for patients presenting with severely uncontrolled asthma
STEP 5
Add-on LAMA
Refer for assessment of phenotype. Consider high dose maintenance ICS-LABA, ± anti-IgE, anti-IL5/5R, anti-IL4R, anti-TSLP
RELIEVER: As-needed short-acting $beta_2$-agonist

ICS: inhaled corticosteroid; LABA: long-acting $beta_2$-agonist; LAMA: long-acting muscarinic antagonist; MART: maintenance and reliever therapy with ICS-formoterol; OCS: oral corticosteroids; SABA: short-acting $beta_2$-agonist. See Box 3-6, p.63 for low, medium and high ICS doses for adults and adolescents.

Figure 3.Treatment stages for asthma as in GINA 2022[26]

Available data

UK Biobank

UK Biobank is one of the biggest resources integrating information on genetics, habits, life-style, and health records[53]. Globally, more than 2,000 projects have been approved since 2012[54]. In UK Biobank, information is organized into Data-Field each with a different identifier; values can be numeric or categorical using specific Data-Coding; there are also Resource –report or table sheets- with explanations and additional data.

Electronic Health Records (EHRs)

Starting from 2013, UK Biobank have added EHRs and in 2019 they released general practice primary care records including coded clinical data and prescriptions[55]. Other biobanks have linked or are willing to link EHRs, such as All of us[56], Biobank Japan[57], China Kadoorie Biobank[58], Million Veteran Program[59]; or networks such as the Electronic Medical Records and Genomics (eMERGE)[60].

The primary aim of EHRs is to store medical information about a patient, not to be used for research. However, since the rise of population-scale studies, researchers understood the potential of applying it to identify a specific cohort manifesting the diseases of interest. Thus, the first EHRs-based genetic study was conducted in 2008 (Figure 4). Nowadays, EHRs are used in GWAS to create valid phenotype definition for the trait of interest, especially in the context of complex diseases, pharmacogenomics and phenome-wide association studies (PheWAS)[61]. The dialogue between domain experts (e.g. specialized doctors) and clinical informaticians is an important step to assess the good quality of a EHRs-derived phenotype[61]: the expert knows the symptoms, manifestation, comorbidities for the studied disease whilst the clinical informaticians know how EHRs are stored and where to look for the coded version of the diseases. The aim is to design a phenotype algorithm taking into account diagnosis, prescriptions, and death records in order to identify patients with the trait of interest while minimizing the risk of false positives.

EHRs present both structured (diagnosis codes, laboratory tests), semi-structured (prescriptions, adverse reactions), and unstructured data (clinical notes) or digital images (X-arrays, computed tomography). Individuals presenting evidence for the studied trait in more than one category are more likely to be true positives[61]. Especially with unstructured data, more sophisticated methods based on Natural Language Processing (NLP) or key term searches need to be applied. The level of accuracy for EHRs-derived phenotypes is limited by several factors: the structure of the data, biases in the recording of data, biases due to local practice in a specific facility, one-off (rather than longitudinal) measures such as blood cell counts and spirometry, missing data, typos and abbreviations[61].

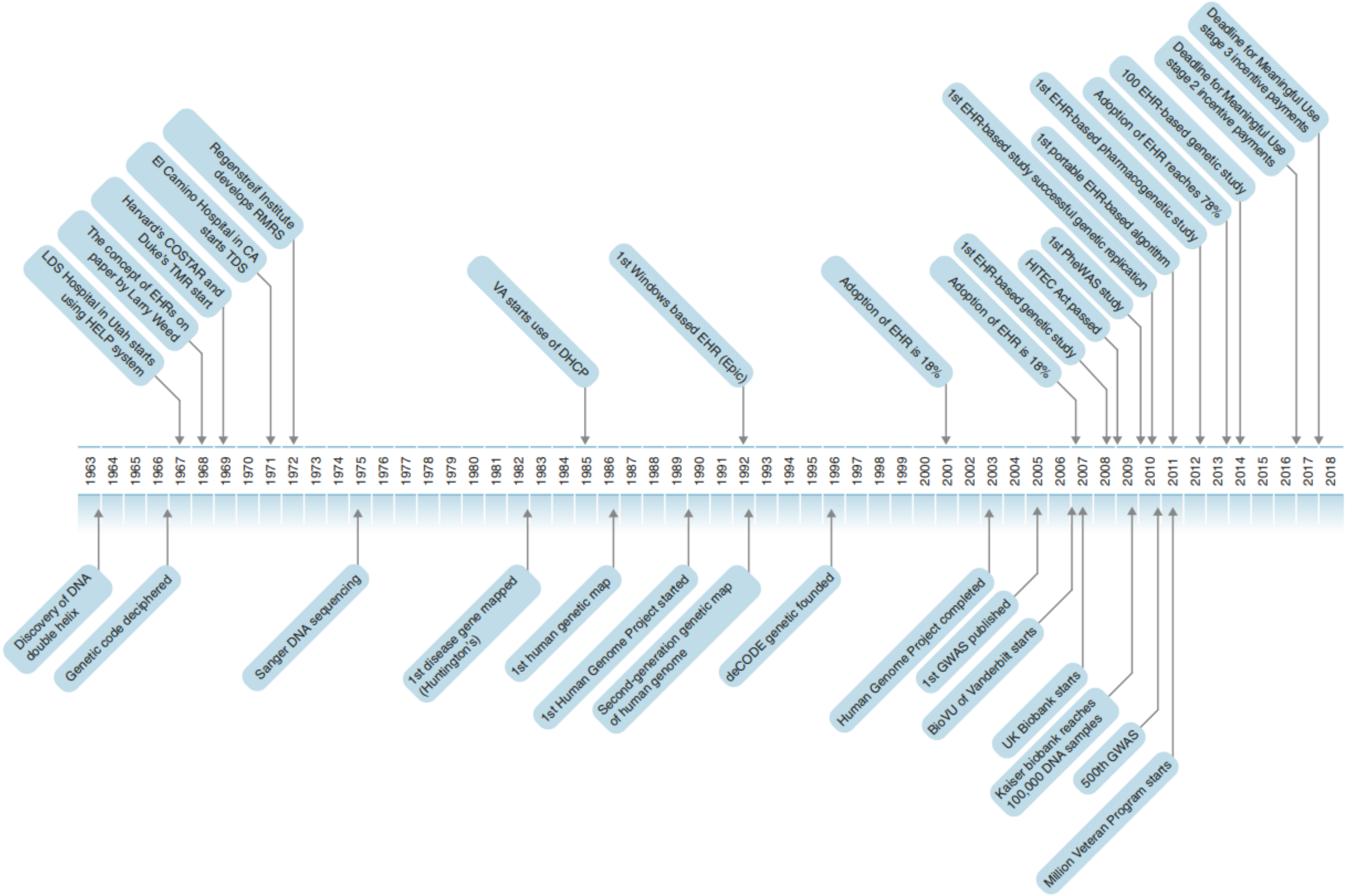

LDS Hospital in Utah starts using HELP system
The concept of EHRs on paper by Larry Weed
Harvard's COSTAR and Duke's TMR start
El Camino Hospital in CA starts TDS
Regenstreif Institute develops RMRS
VA starts use of DHCP
1st Windows based EHR (Epic)
Adoption of EHR is 18%
Adoption of EHR is 18%
1st EHR-based genetic study
HITEC Act passed
1st PheWAS study
1st EHR-based study successful genetic replication
1st portable EHR-based algorithm
1st EHR-based pharmacogenetic study
Adoption of EHR reaches 78%
100 EHR-based genetic study
Deadline for Meaningful Use stage 2 incentive payments
Deadline for Meaningful Use stage 3 incentive payments
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Discovery of DNA double helix
Genetic code deciphered
Sanger DNA sequencing
1st disease gene mapped (Huntington's)
1st human genetic map
1st Human Genome Project started
Second-generation genetic map of human genome
deCODE genetic founded
Human Genome Project completed
1st GWAS published
BioVU of Vanderbilt starts
UK Biobank starts
Kaiser biobank reaches 100,000 DNA samples
500th GWAS
Million Veteran Program starts

Figure 4.Timeline for EHRs and genomic advancements-Additional file 1 of Wei et al. 2015[61]

Methods

Phenotype definition

In any GWAS study, the phenotype definition is an important and preliminary step to select the population of interest to use for the association analysis. In this project, I want to look at asthmatic individuals with treatments related to a more severe disease condition. I therefore implemented a phenotype definition to identify this case population using information from UK Biobank including self-reported data and different types of EHRs. I used data from UK Biobank application 56607.

All-comer asthma

For the case definition in UK Biobank, I started by finding participants with asthma (Figure 5A). An individual was identified as having asthma if they had any of: self-reported doctor-diagnosed asthma (Data-Fields 6152 or 22127; value '8' and '1' respectively); readv2 or readCTV3 codes for asthma in general practice (GP) clinical records (Data-Fields 42040); an ICD-9 or ICD-10 code for asthma in hospital inpatient records (Data-Field 41234; ICD-9 and ICD-10); an ICD-10 code for asthma in death records (Data-Fields 40001 or 40002).

For diagnosis code (ICD-9, ICD-10) and clinical code readv2 I used the table 'Asthma - P2' in the Resource 594 from UK Biobank (Table S1). To retrieve clinical code readCTV3, I mapped them onto the clinical code readv2 specific for asthma using the table 'read_v2_read_ctv3' in the Resource 592 (Table S2).

As exclusion criteria, I filtered out withdrawals (up to February 2022, Figure 5A). I also excluded participants with a diagnosis for COPD/emphysema/chronic bronchitis (Figure 5B). These conditions can be comorbid with asthma and some medications may be prescribed for both conditions. An individual was identified as having emphysema/chronic bronchitis if they had any of: self-reported doctor-diagnosed emphysema or chronic bronchitis (Data-Fields 6152, 22128, or 22129; value '6', '1' and '1' respectively); readv2 or readCTV3 codes for emphysema/chronic bronchitis in GP clinical records (Data-Fields 42040); an ICD-9 or ICD-10 code for emphysema/chronic bronchitis in hospital inpatient records (Data-Field 41234; ICD-9 and ICD-10); an ICD-10 code for emphysema/chronic bronchitis in death records (Data-Fields 40001 or 40002). I used ICD-10 codes J40/J41/ J42/ J43 to retrieve clinical code readv2 and readCTV3 within UK Biobank Data-Coding 1834 and Data-Coding 1835 respectively (Table S3). To find participants with COPD, I used a list of diagnosis and clinical codes as obtained by an automated analysis made by my colleague Richard Packer (Table S4) for Data-Field 40002/40003/42040/41234, with information from self-reported diagnosis (Data-Field 20002). The overall workflow is illustrated in Figure S1.

Difficult-to-treat asthma

To identify patients with more severe or difficult-to-treat asthma, I used GP prescription records (available for ~44% of the whole cohort, Data-Field 42039). Prescriptions could be registered with at least one drug coding system among readv2/readCTV3, British National Formulary (BNF), and dictionary of medicines and devices (dm+d). The drug name was also specified, especially for records that used the dm+d drug system. I firstly extracted all prescriptions with a drug name. For records with no drug name, I mapped the read v2 codes to drug names using the 'read_v2_drugs_lkp' table from the UK Biobank Resource 592 (Table S5). I thus obtained a list of prescriptions for asthma participants (Figure 5B) from which my collaborator, Dr Mike Portelli (University of Nottingham), pulled out the asthma-specific prescriptions. I refined the list by removing medications with names including either 'nasal spray/nasal/spra/cream' terms or 'MOMETASONE FUROATE' prescriptions (Figure 5B) as these are not asthma-specific prescriptions. The final list of asthma prescriptions is reported in Table S6.

Firstly, we decided to follow GINA 2022 guidelines[26], dividing the medication into categories: 'mild' (step 1-2), 'moderate' (step 3-4), 'moderate-severe' (step 5 medications other than prednisolone and biologics), and 'severe' (prednisolone and biologics). Severe asthma patients were defined as those taking a severe asthma prescription and at least one moderate or moderate-severe medication. However, following advice from five clinicians with expertise in asthma (Dominick Shaw, Pete Bradding, John Busby, Ian Hall, Liam Heany), I decided to change the reference guidelines to BTS 2019, as UK Biobank primary care practitioners are more likely to follow this system[34], and to amend my case definition to individuals with any of the following in their records (corresponding to BTS stage 4-5 treatment):

- high dose ICS* + any other medication (except SABA or sodium cromoglicate);
- (not high dose ICS + LABA) + LTRA/LAMA/theophylline/prednisolone;
- ICS + LAMA +/- any other medication

*I used a list for high dose ICS primary care prescriptions from the Respiratory Dashboard Appendix 2 ('High Dose ICS Items' table) of the NHS Business Services Authority (BSA)[62].

To do this, I divided all the prescriptions for asthma medications into drug classes (e.g. SABA, LABA, LAMA) (Table S6). I then ascertained the combination of the classes of drugs for each participant and defined as cases those participants with combinations corresponding to BTS stage 4-5 (Table S7). Finally, I kept individuals of European ancestry.

I compared the cases with the moderate-to-severe UK Biobank cases in stage 1 and stage 2 of the Shrine et al 2019 study[19] in which they used self-reported medications (Data-Field 20003). For UK Biobank application 56607, I was able to retrieve 2,984 out of 2,996 and 5,306 out of 5,414 cases. Thus, the comparison was done for a total of 8,290 individuals.

Controls

Controls were defined as individuals without the following (Figure 6):

- Either asthma, emphysema, chronic bronchitis or COPD diagnostic codes (Table S8);
- Self-reported asthma or any other major respiratory condition (Data-field 20002, value '1111' for asthma and '1072/1466/1130/1131/1132/1133/1134/1135/1660/1139/1140/1141/1142/ 1143/1144/1467/1164/1168/1534/1559/1560/1561/1562/1563/1465' for other respiratory conditions);
- FEV_1/FVC ratio <= 70% (Data-Field 20150, 20151);
- Predicted percentage FEV_1 < 60% (Data-Field 20154);
- General practice prescriptions or self-reported medications for asthma (Data-Field 42039, 20003);
- Absence of GP prescription or self-reported medication records (Data-Field 42039, 20003)
- Non-European ancestry

I selected five age- and sex-matched controls for every case.

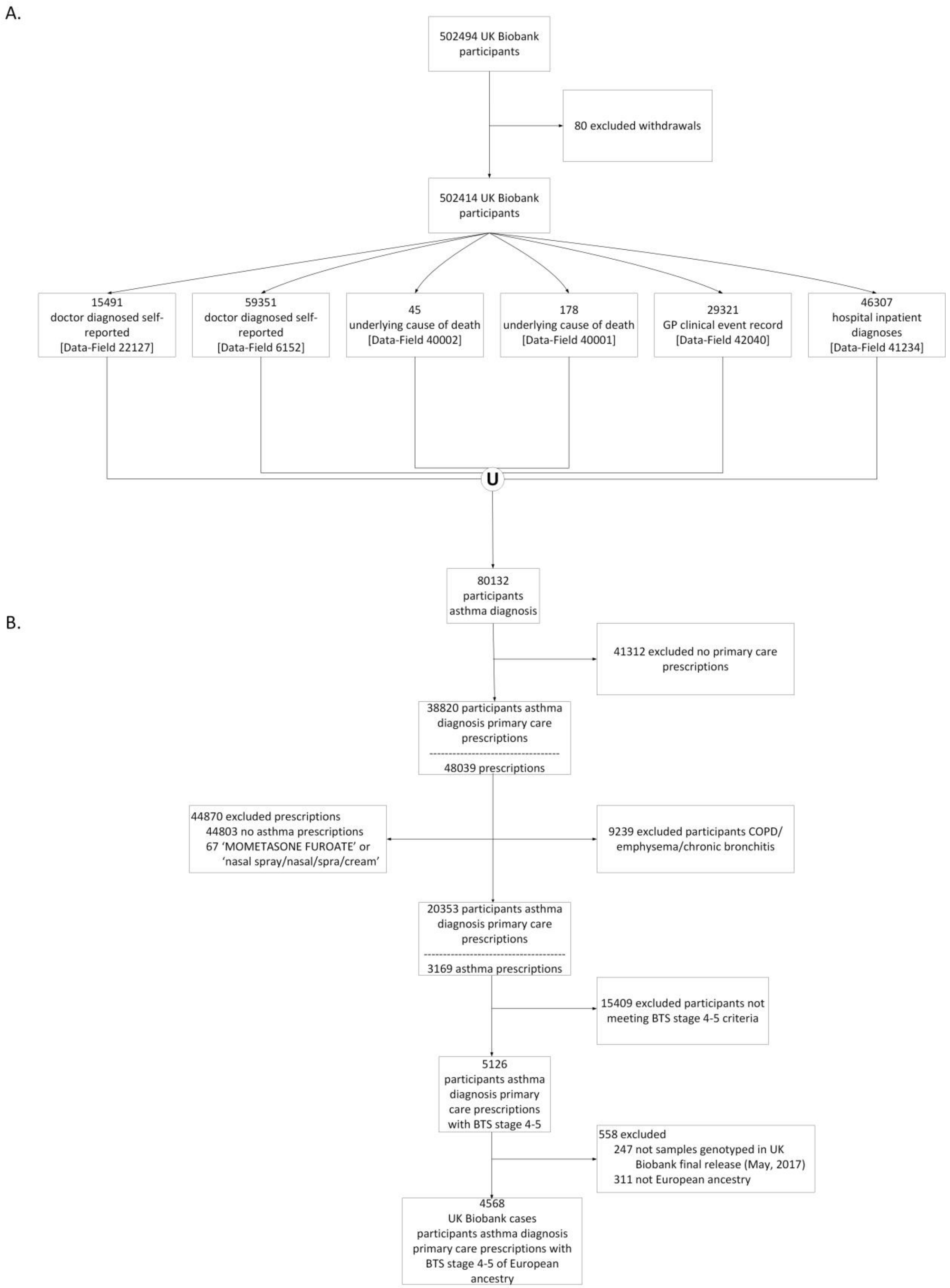

A.
502494 UK Biobank participants
80 excluded withdrawals
502414 UK Biobank participants
15491 doctor diagnosed self-reported [Data-Field 22127]
59351 doctor diagnosed self-reported [Data-Field 6152]
45 underlying cause of death [Data-Field 40002]
178 underlying cause of death [Data-Field 40001]
29321 GP clinical event record [Data-Field 42040]
46307 hospital inpatient diagnoses [Data-Field 41234]
U
80132 participants asthma diagnosis
B.
41312 excluded no primary care prescriptions
38820 participants asthma diagnosis primary care prescriptions
48039 prescriptions
44870 excluded prescriptions
44803 no asthma prescriptions
67 'MOMETASONE FUROATE' or 'nasal spray/nasal/spra/cream'
9239 excluded participants COPD/ emphysema/chronic bronchitis
20353 participants asthma diagnosis primary care prescriptions
3169 asthma prescriptions
15409 excluded participants not meeting BTS stage 4-5 criteria
5126 participants asthma diagnosis primary care prescriptions with BTS stage 4-5
558 excluded
247 not samples genotyped in UK Biobank final release (May, 2017)
311 not European ancestry
4568 UK Biobank cases participants asthma diagnosis primary care prescriptions with BTS stage 4-5 of European ancestry

Figure 5.Case selection workflow in UK Biobank. A) Asthma diagnosis definition according to several Data-Fields; B) Asthma with primary care and following BTS 2019 stage 4-5 criteria. GP=General Practitioner; BTS=British Thoracic Society.

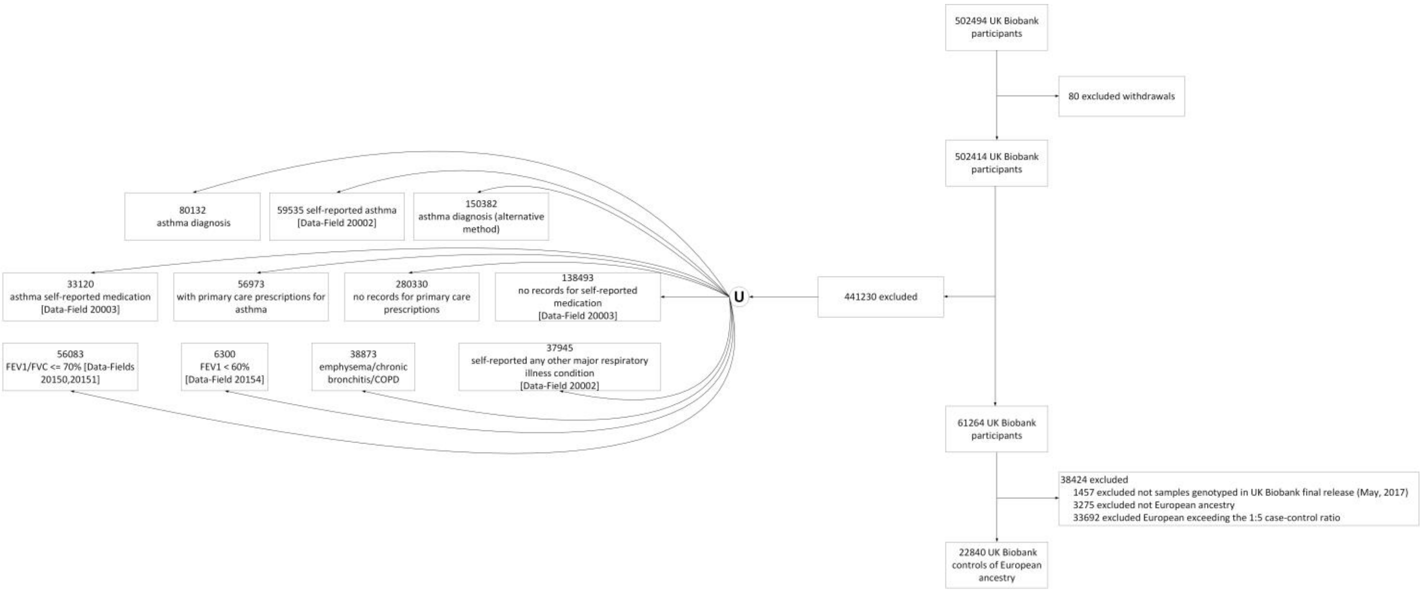

502494 UK Biobank participants
80 excluded withdrawals
502414 UK Biobank participants
80132 asthma diagnosis
59535 self-reported asthma [Data-Field 20002]
150382 asthma diagnosis (alternative method)
33120 asthma self-reported medication [Data-Field 20003]
56973 with primary care prescriptions for asthma
280330 no records for primary care prescriptions
138493 no records for self-reported medication [Data-Field 20003]
U
441230 excluded
56083 FEV1/FVC <= 70% [Data-Fields 20150,20151]
6300 FEV1 < 60% [Data-Field 20154]
38873 emphysema/chronic bronchitis/COPD
37945 self-reported any other major respiratory illness condition [Data-Field 20002]
61264 UK Biobank participants
38424 excluded
1457 excluded not samples genotyped in UK Biobank final release (May, 2017)
3275 excluded not European ancestry
33692 excluded European exceeding the 1:5 case-control ratio
22840 UK Biobank controls of European ancestry

Figure 6.Control selection workflow in UK Biobank. Several exclusion criteria were applied looking at comorbidities and lung function measures. FEV1=Forced Expiratory Volume for the first second; FVC= forced vital capacity; COPD=chronic obstructive pulmonary disease.

Descriptive analysis

In order to investigate my case-control cohort, I compared demographic information including genetic sex (Data-Field 22001), age at recruitment (Data-Field 21022), and smoking status (Data-Field 20116) as defined in U-BIOPRED[63] using pack years of smoking (Data-Field 20616). I also compared clinical traits: age at onset of asthma (merged Data-Field 3786 and 22147), BMI (Data-Field 23104), hospitalisation due to asthma (Data-Field 41234, ICD-9 and ICD-10, 'level 1'), eosinophil count (Data-Field 30150), neutrophil count (Data-Field 30140), percentage predicted FEV_1 (in house pre-processed from Data-Field 20154), FEV_1/FVC ratio (in-house pre-processed from Data-Field 20150, 20151), prednisolone use (from prescription records, Data-field 42039).

GWAS analysis

Genotype and Imputation data

UK Biobank individuals were genotyped using the UK Biobank Axiom Array including 825,927 variants[53]. Imputed data for ~93M autosomal variants[53] were obtained using a combined reference panel inclusive of the Haplotype Reference Consortium panel[7], the UK10K[8] and 1000GP Phase 3[9].

Association analysis

I performed a GWAS using the software REGENIE v2.2.4[64] with covariates age at recruitment, age at recruitment^2, genetic sex, and the first 10 principal components.

REGENIE is a machine-learning approach based on a two-step analysis and has shown better performance in terms of time and memory with respect to other similar methods[64]. In step 1, a set of genotyped and good quality SNPs is used to create a prediction model to be used as a covariate in step 2. Specifically, step 1 is further divided into two sub-processes, called levels in the REGENIE paper: at level 0, the set of SNPs (M) are segmented into blocks of fixed size (B) and they are used to create ridge linear regression predictions (J) using linkage disequilibrium (LD) information and different parameters aiming at capturing the true number and size of associated variants in each block. At level 1, these predictions are combined and a reduced set of SNPs, calculated as M*J/B, is used as predictors for another ridge regression, linear or logistic according to the studied trait, and a cross-validation method is used to choose the parameters of the model. The result is one predictor that is then decomposed into chromosomal predictions for a leave-one-chromosome-out (LOCO) approach. LOCO is a conditional testing in which imputed SNPs from a single chromosome are tested against all other variants except ones on the same chromosome; the reason is to minimize the risk of regional contamination that can lower the statistical power[65,66]. Finally, step 2 applies a linear or logistic regression model using the LOCO predictors as a fixed covariate value. For binary traits, REGENIE implements a score test as hypothesis testing, but it has also the possibility to use a time efficient approximate Firth logistic regression test which reduces biases derived for the maximum-likelihood estimates, and it gives good quality Type 1 errors, SNP estimates and standard errors[64]. As the score test, the approximate Firth test follows a X^2 distribution with one degree of freedom; user can set the p-value threshold to use the Firth (default 0.05).

Since REGENIE step 1 requires good quality genotype data in order to create the LOCO predictors, the authors recommend a preliminary step of filtering for good quality data using the software plink v2[67]: excluding variants with minor allele frequency (MAF) < 1%, minor allele count (MAC) < 100, missingness above 10%, and Hardy-Weinberg equilibrium p-value > 10e-15. Individuals with >10% missing genotypes are also excluded. The resulting variants and samples were kept as input for REGENIE step 1.

For REGENIE step 2, I used the option '--bt' (to specify a binary outcome) and the Firth test was applied for p-values below 0.01. Variants with MAF of at least 1%, MAC of at least 10 and INFO imputation score of at least 0.3 were used for this analysis.

Power calculation

I used the GAS Power Calculator, a free accessible online tool to compute statistical power for one-stage genetic association analysis[68]. The parameters used were (Figure S3A): sample size (4568 cases;22840 control); significance level (5e-8); multiplicative disease model (equivalent to additive model for this software); prevalence of disease as the proportion of asthmatic individuals taking high intensity treatments as a proxy for our case definition, which is estimated to be 24% of the asthma population[26] (24% of 12% as prevalence of all asthma in UK population, thus 0.028); disease allele frequency as 0.203 based on Hakonarson et al. 2019[69]; genotype relative risk of 1.2 as for Kuruvilla et al. 2019[70].

Post association analysis

Q-Q plot and Manhattan plots were obtained using the R package 'qqman'[71] and R v4.1.0. In order to identify any residual genomic factor that could have biased the analysis, I calculated the LDscore (LDSC) regression intercept using ldsc v1.0.1 [72] and including the LD patterns of the European super-population of 1000GP[9].

Variants with a p-value equal to or below the genome-wide significant threshold of 5e-8 were extracted and the sentinel variant for each genomic loci was chosen as the one with the lowest p-value in its surrounding region of 1Mb (+/- 500Kb).

Locus zoom plots were created using the free online tool LocusZoom v0.12[73] to investigate the sentinel variants, their associated genomic loci, and surrounding genes if any. I annotated the sentinel variants using the online tool variant effect predictor (VEP) release 107[74].

Finally, I compared the sentinel variants with the moderate-to-severe asthma study[19] using LDpair Tool[75] to calculate LD between variants. In order to investigate association with any trait in literature, I used the plugins 'Phenotypes' in VEP and 16 databases including GWAS Catalog[76].

Code implementation

I used bash, python (v.3.5.6) and R (v4.1.0) languages to run my analyses and I used the High Performance Computing (HPC) SPECTRE2 at the University of Leicester.

All the code is documented and available by request in the github repositories (currently private):

https://github.com/legenepi/UKBiobank_asthma_medication.git

https://github.com/legenepi/UKBiobank_asthmaMeds_stratification.git

https://github.com/legenepi/REGENIE_assoc_severe_asthma.git

Results

Phenotype definition

I obtained 80,132 asthma participants among which 48.45% had primary care prescriptions (Figure 5A and B) and a total of 3,169 records for asthma medications. After exclusion of asthma participants with a diagnosis for emphysema/chronic bronchitis/COPD and/or without prescription for asthma drugs, I obtained 20,353 participants with asthma-only and good quality asthma prescriptions (Figure 5B). Among these, 25.19% (5,126) had prescriptions for medications indicative of BTS stage 4-5 criteria; this percentage is in line with the proportion of high intensity treatment patients reported by GINA 2022 guidelines[26]. This resulted in 4,568 European ancestry cases (Figure 5B).

My case set included 1,370 out of 8,290 moderate-to-severe cases found in UK Biobank by Shrine et al.[19], which represented 30% of the cases in this study (Figure S4). My phenotype definition did not capture 6,920 cases from Shrine et al. among which 758 cases had primary care prescription records. Of these, 405 participants had prescriptions indicative of BTS stage 3 criteria, and therefore not consistent with my criteria.

Case-control cohort: descriptive analysis

I selected 4,568 cases and 22,840 age- and sex-matched controls for this analysis (Table 1). Overall, the mean age is 57 years old. The dataset has a higher proportion of females compared to males (63.95% vs 36.05%). Cases exhibit higher BMI, lower lung function, and higher blood cell counts of eosinophils and neutrophils compared to controls. There are a similar proportion of smokers and non-smokers between cases and controls. Amongst cases, 61.82% self-report adult-onset asthma, 8.37% have been hospitalised with asthma as primary cause at least once, and 66.70% have been prescribed for prednisolone at least once.

		Cases (4568)	Controls (22840)
Age, years		57(8)	57(8)
Sex	Female	2921 (63.95%)	14605 (63.95%)
	Male	1647 (36.05%)	8235 (36.05%)
BMI		28.69 (5.57)	27.37 (4.68)
Lung function measure	FEV1 % predicted	85.41 (15.62)	95.80 (12.23)
	FEV1/FVC	0.73 (0.07)	0.78 (0.04)
Blood cell count	Eosinophils	0.21 (0.13)	0.14 (0.08)
	Neutrophils	4.46 (1.31)	4.10 (1.16)
Smoking status	Smokers	1159 (25.37%)	5782 (25.32%)
	Non-smokers	2834 (62.04%)	13853 (60.65%)
	Unknown	575 (12.59%)	3205 (14.03%)
Hospitalisation	L1	411 (8.37%)	NA
	L2	3257 (66.29%)	NA
	Unknown	1245 (25.34%)	NA
Category onset	Adult	2824 (61.82%)	NA
	Childhood	876 (19.18%)	NA
	Unknown	868 (19.00%)	NA
Prednisolone use	Yes	3047 (66.70%)	NA
	No	1521 (33.30%)	NA

Table 1.Descriptive analysis in case-control cohort of European ancestry. For categorical traits: count (%). For quantitative traits: mean (standard deviation). BMI=Body Mass Index; FEV1=Forced Expiratory Volume for the first second; FVC=forced vital capacity. L1=level 1; L2=level 2.

Association Analysis

I tested 9,805,323 variants for association with BTS 4-5 stage asthma in 4,568 cases and 22,840 controls. Of these, 2,506 variants were associated with case status at a genome-wide significant threshold (P = 5e-08) (Table S9) (Figure 7A). Power calculations for this analysis showed an overall statistical power above 88.4% (Figure S3B) to detect associations with variants showing disease allele frequency of 20.3% and mean genetic effect size of 1.20; but it is underpowered to detect effects in variants with minor allele frequency less than 18% (Figure S3C).

LDSC regression intercept for the GWAS was 1.03 suggesting little systematic bias due to population stratification (Figure 7B). Since LDSC < 1.05, no genomic correction was applied.

After signal selection, I obtained ten sentinel variants representing genomic loci on chromosomes 2, 5, 6, 8, 9 and 10 (Table 2); OR ranged between 0.84-1.22 and allele frequency was common for all variants (24.29-74.50%). Regional association plots for these loci are shown in Figure S5: overall, sentinel variants are in high LD with variants in the surrounding region ($R^2 > 0.6$) with the exception of rs6462, which shows lower correlation with surrounding variants. LD values could not be plotted for rs113880645 and rs201499805 due to the lack of LD values in the data used by the software.

Comparing the results with the previous moderate-to-severe study, I observed that eight genomic loci out of ten had been previously associated with asthma: the locus on chromosome 8 showed the same sentinel variant (rs113880645/rs71266076) and the other seven loci showed correlated sentinels, calculated via LDpair Tool with European populations[75] (result not shown). The two loci not found in the moderate-to-severe were two genomic regions on chromosome 6 with sentinel variants rs2428494 and rs6462. None of the novel loci reported in Shrine et al.(mapped to GATA3, MUC5AC and KIAA1109 respectively)[19] exceeded genome-wide significance in my study (rs72687036 p-value 3.1e-2, rs10905284 p-value 1.9e-3, rs11603634 p-value 2.3e-3); however the OR for these loci in my study (rs72687036 OR 1.03, rs10905284 OR 0.93, rs11603634 OR 1.07) were similar to the stage 2 UK Biobank analysis in the previous study[19].

Using the online tool VEP, I was able to query several databases for association results in the literature, annotate the sentinel variants and find the nearest gene(s). Firstly, all the sentinel variants were found already associated for at least one trait (Table S10): five variants showed evidence of association with asthma and other, related traits including allergic or inflammatory disease and eosinophil count (rs12470864, rs10455025, rs2428494, rs9271365, rs113880645). VEP did not report a direct asthma association for the remaining five variants, but they have been reported associated with either chronic rhinosinusitis, eosinophil counts, nasal polyp, D-2-hydroxyglutaric aciduria, congenital adrenal hyperplasia or hip circumference adjusted for BMI (Table S11). Second, variants were mostly annotated as non-coding, mapping either upstream or downstream of genes, within introns, or in intergenic regions (Table 2). Finally, a total of 12 genes were mapped to these sentinels; rs9271365, rs113880645 and rs201499805 did not have associated mapped genes.

Ten genes had been previously associated with asthma traits. In particular, *ILR1RL1/IL18R1* genes were previously reported by two studies: Wan Yi et al. (p-value 5.59e-8)[17] and Shrine et al.[19]. The latter reported another three genes found associated in my analysis: *D2HGDH*, *TSLP*, *IRF1-AS1*. They reported the *MIR5708* gene for rs113880645 on chromosome 8, which was not found by VEP, possibly due to different rsid number or gene annotation tool. *HLA-B* and *MIR6891* were found associated to allergic disease and age at onset of asthma in two studies[77,78]. *TNXB*, is in the genomic region of the major histocompatibility complex on chromosome 6 and overlaps the *CYP21A2* gene; it was found associated in a study for atopic dermatitis in childhood with asthma[79]. *CB4*, *CB-AS1* and *CYP21A2* showed association with blood protein measures.

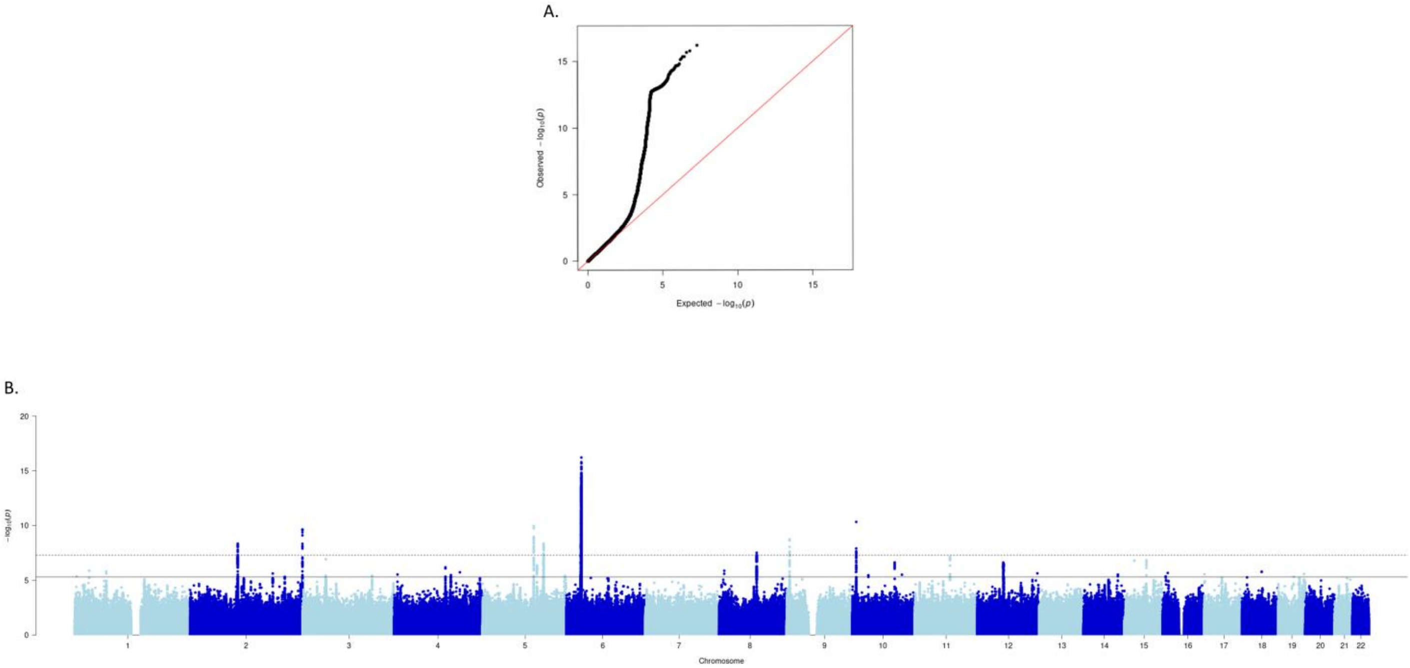


Figure 7.GWAS results visualisation. A) Q-Q plot; B) Manhattan Plot, dashed line is the genome-wide significant threshold (5e-8), continuous line is the suggestive genome-wide threshold (5e-6). p=p-value.

SNP	Chr	Position	Gene	Consequence	Effect allele	Non effect allele	Minor allele	Effect allele frequency	OR [95%CI]	P-value
rs12470864	2	102926362	IL1RL1	upstream_gene_variant	A	G	A	38.99%	1.15 [1.10-1.21]	4.51E-09
rs12470864	2	102926362	IL18R1	upstream_gene_variant	A	G	A	38.99%	1.15 [1.10-1.21]	4.51E-09
rs6761047	2	242692858	D2HGDH	intron_variant	T	C	T	24.29%	0.84 [0.79-0.89]	2.31E-10
rs6761047	2	242692858	D2HGDH	downstream_gene_variant	T	C	T	24.29%	0.84 [0.79-0.89]	2.31E-10
rs6761047	2	242692858	D2HGDH	upstream_gene_variant	T	C	T	24.29%	0.84 [0.79-0.89]	2.31E-10
rs10455025	5	110404999	TSLP	upstream_gene_variant	C	A	C	35.59%	1.17 [1.12-1.23]	1.11E-10
rs2188962	5	131770805	IRF1-AS1	intron_variant	T	C	T	42.71%	0.87 [0.83-0.91]	4.51E-09
rs2428494	6	31322197	HLA-B	intron_variant	A	T	A	47.16%	1.17 [1.11-1.22]	6.44E-11
rs2428494	6	31322197	HLA-B	downstream_gene_variant	A	T	A	47.16%	1.17 [1.11-1.22]	6.44E-11
rs2428494	6	31322197	MIR6891	downstream_gene_variant	A	T	A	47.16%	1.17 [1.11-1.22]	6.44E-11
rs6462	6	32006597	TNXB	downstream_gene_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs6462	6	32006597	C4B-AS1	upstream_gene_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs6462	6	32006597	C4B	downstream_gene_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs6462	6	32006597	CYP21A2	intron_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs6462	6	32006597	CYP21A2	upstream_gene_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs6462	6	32006597	CYP21A2	non_coding_transcript_exon_variant	T	C	C	72.20%	1.16 [1.10-1.22]	1.88E-08
rs9271365	6	32586794	-	intergenic_variant	G	T	G	42.89%	1.22 [1.16-1.28]	6.17E-17
rs113880645	8	81266924	-	downstream_gene_variant	CT	C	C	63.50%	0.87 [0.83-0.92]	3.06E-08
rs113880645	8	81266924	-	intron_variant	CT	C	C	63.50%	0.87 [0.83-0.92]	3.06E-08
rs1888909	9	6197392	GTF3AP1	downstream_gene_variant	C	T	T	74.50%	0.85 [0.81-0.90]	1.78E-09
rs201499805	10	9042744	-	intergenic_variant	C	CT	C	30.70%	0.84 [0.79-0.88]	4.67E-11

Table 2.Sentinel variants summary statistics and annotation in Variant Effect Predictor (VEP). Variants present different annotation terms, mainly within non-coding type. Chr: Chromosome. OR: Odds Ratio. CI: Confidence Interval

Discussion

During this first year, most of the time has been spent in reflecting on which criteria are the best suited to select a pool of cases most likely to be true difficult-to-treat and severe asthma patients. In doing so, we decided to select UK Biobank individuals with primary care prescriptions for treatments within the BTS 2019 guidelines stage 4-5, interpreting these treatments as a proxy for a difficult-to-treat and more severe type of asthma.

The longitudinal nature of primary care prescription records allows more comprehensive information for each individual than self-reported medications. Together with prescriptions, I included different sources of information for asthma diagnosis and excluded participants showing many comorbidities that could have biased the definition based on treatments, resulting in a more conservative phenotype. As a result, I obtained 4,568 cases of European ancestry in UK Biobank and conducted a GWAS with 22,840 respiratory disease-free controls. Ten sentinel variants were found associated below the genome-wide significant threshold, among which eight were already described or were proxies of associated variants in the moderate-to-severe study[19]. Twelve genes were mapped to the sentinel variants showing previous associations with asthma traits or allergic/inflammation traits type. Based on the results, the phenotype definition seems to capture true positive asthma participants.

However, this study had several limitations. Firstly, as suggested by the fact that BTS and GINA guidelines are continuously changed and updated, there is no definitive definition for severe asthma due to its complexity, heterogeneity and missing knowledge. In fact, UK Biobank primary care prescriptions included in this study span 52 years of records (1967-2019), a period in which guidelines have changed. Currently, the most accepted definition for severe asthma is based on a retrospective analysis looking at intensity of patients' treatment and symptoms control. Unfortunately, there is no systematic data on symptoms control in UK Biobank, thus I wasn't able to take this criteria into account in my phenotype algorithm. In addition, prescription records only confirm the evidence of a prescription, but they do not guarantee that the treatment is actually followed by the patient meaning it is not possible to measure adherence.

Moreover, primary care records present a hefty amount of free speech terms, which demands computational knowledge and time to process. I faced difficulties in extracting daily dosages ('puff per day' or 'mg/day') from each prescription when available. In addition, I found it challenging to devise code to take duration and date of prescriptions into account. As an example, I regarded prednisolone as an add-on treatment indicative of difficult-to-treat asthma, but did not distinguish between maintenance prednisolone and short-term prednisolone. Discerning individuals under long-term prednisolone should be a good improvement to this phenotype algorithm. Furthermore, prednisolone might be prescribed for conditions other than asthma. It was also challenging to distinguish high, medium and low dose ICS prescriptions as number of puffs per day are reported in an ambiguous way in the prescription records. I used the NHS-BSA resource for high dose ICS prescriptions as a proxy, but it is possible that some high ICS drugs are not present in this resource, leading to misclassification.

Finally, my definition collapsed together all the prescriptions for an individual assuming that they were taken at the same time, and also that they were taken continuously rather than as a one-off/short-term treatment. Once again, understanding the long-term course of the treatment would validate the presence of a severe condition avoiding misclassification due to episodic worsening of asthma that could also happen in individuals with mild/moderate conditions.

These limitations could partially explain why I did not identify the three novel loci associated to moderate-to-severe as reported by Shrine et al.[19]. As the discovery cohort in the moderate-to-severe study comprised cases from the Genetics of Asthma Severity and Phenotypes (GASP) initiative[19] and U-BIOPRED[63], it is possible that these associations were driven by the more severe cases in these cohorts. In my study, we have attempted to select more severe cases (BTS stage 4-5, rather than BTS stage 3-5) but the results from these three loci suggest that our cases may still be milder than the GASP and U-BIOPRED cases. This is supported by the fact that the rate of hospitalisations for asthma amongst our cases is lower than the rate amongst stage 1 and stage 2 cases in the previous study (Figure S6). UK Biobank alone might be not sufficient to highlight this specific subgroup of patients and defining high-intensity treatment might be the best possible approximation of difficult-to-treat or severe asthma. It is also an option that mild and more severe asthma share common genetic risk factors and that other non-genetic factors are involved in severity including symptoms control, adherence, inhaler technique, environment, and life style.

Project proposal for future PhD years

‘Which genomic variants and loci are associated with difficult-to-treat/severe asthma?’

I will:

1) Refine my definition of difficult-to-treat/severe asthma in UK Biobank
    a. By taking into account date/duration of treatments in prescription records, including distinguishing maintenance from short-term prednisolone
    b. Consider using self-reported medication data on the whole-cohort as increased power may counterbalance the limitations of this data
2) Perform seq-GWAS of difficult-to-treat asthma using UK Biobank whole-genome sequencing data in European individuals under application 88144
    a. Glenda Lassi is arranging a placement in the Centre for Genomics Research at AZ in the autumn so that I can train to use this data and perform seq-GWAS analysis
3) Replication of findings in additional cohorts
    a. As well as GASP and U-BIOPRED cohorts, we have an ongoing collaboration with the GERA and Lifelines cohorts. These cohorts have EHRs and genotyping data. We are scoping out other opportunities for collaboration with cohorts of non-European ancestry
    b. If enough cohorts of non-European ancestry, discuss the possibility of a multi-ancestry meta-analysis
    c. Sensitivity analyses (eg. excluding individuals with allergic conditions)

‘Among these associated variants, which are the causal ones and what are their likely mechanisms of action?’

I will:

4) Perform conditional analysis to find the final number of independent associated genomic signals
5) Run a fine mapping analysis using state-of-the-art Bayesian methods (eg. Susie)
6) Perform functional annotation and gene mapping/prioritization analyses (e.g. co-localisation)

‘What are the genes and biological pathways impacted by causal variants and how do they affect risk of severe asthma?’

I will

7) Validate the discovered variant-gene biological pathways and networks with *in silico* methods.
8) Have the possibility to follow-up variants/genes of interest with *in vitro* experiments: human-derived cell lines of severe asthma patients available at Nottingham University. Here I will have the opportunity to implement some state-of-the art techniques, such as CRISPR/Cas9 system, guided by my second supervisor Ian Sayers.

‘Can we translate these findings into a genetically driven drug discovery?’

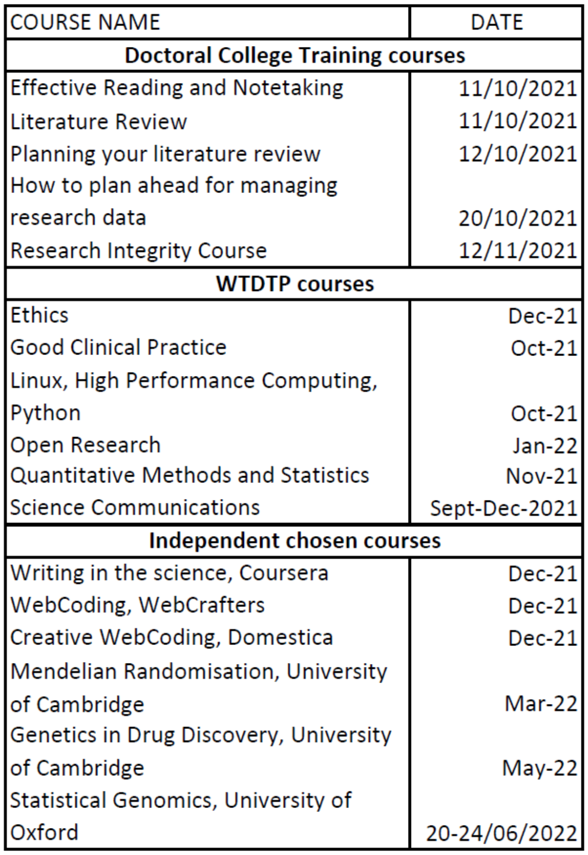
Towards the end of my PhD there will be opportunities for a second placement at AZ. Possible projects would include:

9) Investigating the potential of variants/genes of interest to guide development of a new drug or repurposing of an existing drug
10) Using genetic knowledge gained from the earlier stages of my PhD, identify subgroups of patients that might have differential presentation, prognoses, and/or response to medication

Appendix

Training plan

Year 1



COURSE NAME	DATE
Doctoral College Training courses	
Effective Reading and Notetaking	11/10/2021
Literature Review	11/10/2021
Planning your literature review	12/10/2021
How to plan ahead for managing research data	20/10/2021
Research Integrity Course	12/11/2021
WTDTP courses	
Ethics	Dec-21
Good Clinical Practice	Oct-21
Linux, High Performance Computing, Python	Oct-21
Open Research	Jan-22
Quantitative Methods and Statistics	Nov-21
Science Communications	Sept-Dec-2021
Independent chosen courses	
Writing in the science, Coursera	Dec-21
WebCoding, WebCrafters	Dec-21
Creative WebCoding, Domestica	Dec-21
Mendelian Randomisation, University of Cambridge	Mar-22
Genetics in Drug Discovery, University of Cambridge	May-22
Statistical Genomics, University of Oxford	20-24/06/2022

Future years

- Learn how to use the Research Analysis Platform (RAP) in order to run whole-genome and whole-exome sequencing GWASs in UK Biobank.
- Learn different methods to perform whole-genome or whole-exome sequencing analyses such as single variant or variant set (/collapsing) approaches as well as ad-hoc analyses for non-coding variants (STAARPipeline [80-82]). In addition, the AZ Centre for Genomics Research has recently released a deep learning method to prioritise non-coding genomic regions (gwRVIS and JARVIS[83]) and my third supervisor, Glenda Lassi, is organizing a placement in AZ in order to be able to learn about this and other tools for GWAS.
- Learn about post-GWAS analyses including:
  - fine-mapping
  - co-localisation
  - functional analyses (variants annotation; network and pathway analyses; omics-integration)
  - *in silico* validation of results and *in vitro* experiments
  - drug discovery
- How to write efficient bioinformatics pipelines

- Project management
- Leadership skills
- Budget management
- Community engagement and interdisciplinary co-production
- Experience in genomic counselling

Other achievements

RSG-ISCB Committee members

This academic year I enjoyed the Italian Regional Student Group of the International Society of Computational Biology, ISCB-RSG Italy for short[1]. It has been re-established in 2020 by two colleagues from my MSc and during this year we organized two online webinars ('How to choose a PhD', 'Life outside academia') and we ran an afternoon session within the Bioinformatics Italian Society (BITS) annual meeting. Here is a photo of the team with the speakers we invited. It was an afternoon focused on students and a conversation about what it means to be a principal investigator in the bioinformatics field. From this, I learnt that open communication, self-awareness and being pro-active and positive are important in research as well as professional values.

1.http://rsg-italy.iscbsc.org/

Creative Engagement Fellowship Phase 2

The Attenborough Arts Centre together with the University's Wellcome Trust ISSF Public Engagement Scheme have launched the second call for the Creative Engagement fellowship Programme Phase 2[1]. It consists of a co-production between academics and an artist. Academics were called to come up with an idea to address the problem of racism, classism, or ableism and the aim of the co-production is to produce a digital piece of art.

I came to Leicester two years ago driven by my will to do scientific research on African ancestry populations, which are currently underrepresented in human genomic studies. This motivation of mine to do human research in such a way that it is more respectful of the real genetic ancestry composition of our world is a core value of my professional life. In Leicester, I am not the first one who wants to speak about this. I am delighted to know of local studies such as EXCEED[2] and the different activities of the Ethnic Centre for Health Research[3]. Still, the road towards equity and fair representation of genomic ancestry and populations is still long. One problem is the lack of public awareness of this under-representation in genomic research, and potentially the lack of trust in academics that result in a reduction in enrollment in this type of study. This creates a vicious circle in which, at the end of the day, genomic studies will be mostly represented by middle aged European and white individuals. For this reason, when this fellowship was announced, I saw it as an opportunity to finally put into practice my thoughts and try to invest in local activities with a high intrinsic value. I am co-leading this fellowship with Dr. Chiara Batini, and with collaborators Dr. Winifred Ekezie, Barbara Czyznikowska, Dr. Katherine Fawcett, and Dr. Laura Venn.

I expect to work a total of one day per week on this project from September 2022 to the end of February 2023 with a possibility to extend the project until mid April 2023.

1.https://attenborougharts.com/research/creative-engagement-fellowship/

2.https://exceed.org.uk/

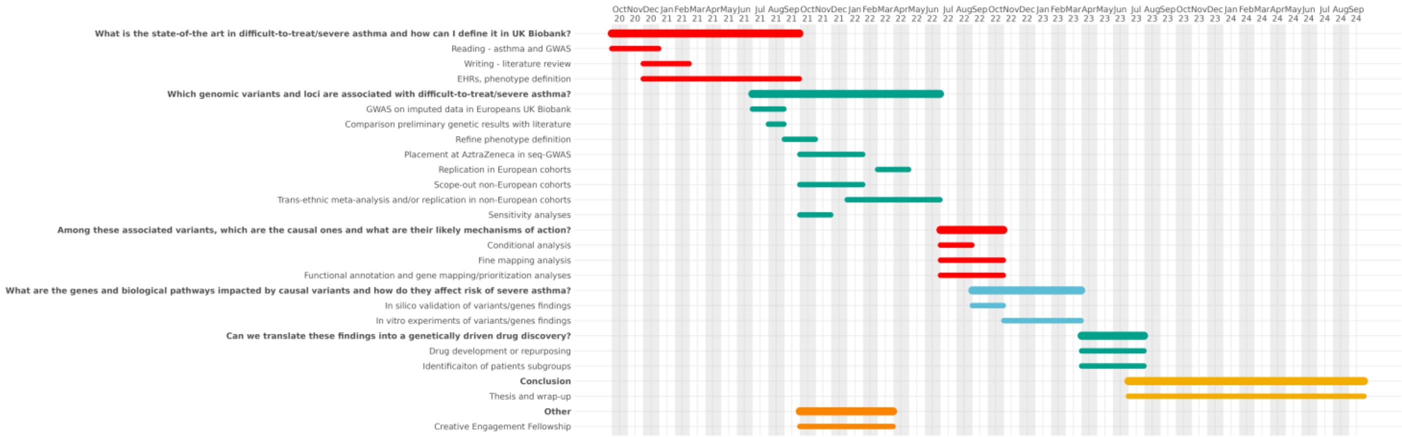
3.https://ethnichealthresearch.org.uk/

<u>Contribution to a manuscript as a first-author paper (manuscript still in revision)</u>

During my Master degree research project I have worked on a GWAS of smoking behavior traits in male individuals of African ancestry supervised by Dr. Chiara Batini and Dr. Ananyo Chadauri[1]. We are now answering reviewer's comments from Scientific Reports and will submit in September 2022.

[1] Piga et al., Genetic insights into smoking behaviours in 10,558 men of African ancestry from continental Africa and the UK, medRxiv 2021.12.09.21267119;
doi: https://doi.org/10.1101/2021.12.09.21267119

Gantt Chart



Gantt Chart.Legenda: colors for each main task (bold text); colors can be repeated.

References

1. NICE. What is the prevalence of asthma? (2021).
2. Mukherjee, M. *et al.* The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Medicine* **14**, 113 (2016).
3. McDonald, V.M., Kennington, E. & Hyland, M.E. *Understanding the experience of people living with severe asthma*, (2019).
4. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).
5. Martin, A.R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* **51**, 584-591 (2019).
6. Global Biobank Meta-analysis Initiative.
7. McCarthy, S.A.-O. *et al.* A reference panel of 64,976 haplotypes for genotype imputation.
8. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
9. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
10. Valette, K. *et al.* Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. *Communications Biology* **4**, 700 (2021).
11. Moffatt, M.F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470-3 (2007).
12. Cheng, Q. & Shang, Y. ORMDL3 may participate in the pathogenesis of bronchial epithelialmesenchymal transition in asthmatic mice with airway remodeling. *Mol Med Rep* **17**, 995-1005 (2018).
13. Das, S. *et al.* GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. *Proc Natl Acad Sci U S A* **113**, 13132-13137 (2016).
14. Bonnelykke, K. *et al.* Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* **45**, 902-906 (2013).
15. Tang, H.H.F., Teo, S.M., Sly, P.D., Holt, P.G. & Inouye, M. The intersect of genetics, environment, and microbiota in asthma-perspectives and challenges. *J Allergy Clin Immunol* **147**, 781-793 (2021).
16. Cameron-Christie, S. *et al.* A broad exome study of the genetic architecture of asthma reveals novel patient subgroups. *bioRxiv*, 2020.12.10.419663 (2020).
17. Wan, Y.I. *et al.* Genome-wide association study to identify genetic determinants of severe asthma. *Thorax* **67**, 762-8 (2012).
18. Jones, B.L. & Rosenwasser, L.J. Linkage and Genetic Association in Severe Asthma. *Immunol Allergy Clin North Am* **36**, 439-47 (2016).
19. Shrine, N. *et al.* Moderate-to-severe asthma in individuals of European ancestry: a genome-wide association study. *The Lancet. Respiratory medicine* **7**, 20-34 (2019).
20. Saikumar Jayalatha, A.K., Hesse, L., Ketelaar, M.E., Koppelman, G.H. & Nawijn, M.C. The central role of IL-33/IL-1RL1 pathway in asthma: From pathogenesis to intervention. *Pharmacol Ther* **225**, 107847 (2021).
21. GlaxoSmithKline. Efficacy and safety study of GSK3772847 in subjects with moderately severe asthma. (Clinicaltrials.Gov (2017), 2017).
22. Sanofi. Sanofi and Regeneron announce positive topline Phase 2 results for IL-33 antibody in asthma. (2019).

23. AnaptysBio, I. Proof of concept study to investigate ANB020 activity in adult patients with severe eosinophilic asthma. *NIH Website https://clinicaltrials. gov/ct2/show/NCT03469934* (2019).
24. Chupp, G.L. *et al.* A chitinase-like protein in the lung and circulation of patients with severe asthma. *N Engl J Med* **357**, 2016-27 (2007).
25. Hinds, D.A. *et al.* A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* **45**, 907-11 (2013).
26. Asthma, G.I.f. Global Strategy for Asthma Management and Prevention. (2022).
27. Cooper, R., Bingham, K., Portelli, M. & Sayers, I. Genetics of Asthma: Insights From Genome Wide Association Studies. (2021).
28. Kuruvilla, M.E., Vanijcharoenkarn, K., Shih, J.A. & Lee, F.E.-H. Epidemiology and risk factors for asthma. *Respiratory Medicine* **149**, 16-22 (2019).
29. Pividori, M., Schoettler, N., Nicolae, D.L., Ober, C. & Im, H.K. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med* **7**, 509-522 (2019).
30. UK Biobank Whole Genome Sequencing project.
31. Pavord, I.D. *et al.* After asthma: redefining airways diseases. *Lancet* **391**, 350-400 (2018).
32. Cohen, S.G. Sir John Floyer (1649-1734) British physician and pioneer clinical investigator. *Allergy Proc* **16**, 328-9 (1995).
33. Gerday, S. *et al.* Revisiting differences between atopic and non-atopic asthmatics: When age is shaping airway inflammatory profile. *World Allergy Organization Journal* **15**, 100655 (2022).
34. Mathioudakis, S.G.d.g.a.A. *SIGN158: British guideline on the management of asthma*, (Scottish Intercollegiate Guidelines Network, United Kingdom, 2019).
35. Moore, W.C. *et al.* Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* **181**, 315-23 (2010).
36. De Meulder, B. *et al.* U-BIOPRED accessible handprint: combining omics platforms to identify stable asthma subphenotypes. *European Respiratory Journal* **52**, OA3578 (2018).
37. Jackson, D.J. *et al.* Characterisation of patients with severe asthma in the UK Severe Asthma Registry in the biologic era. *Thorax* **76**, 220-227 (2021).
38. Lambrecht, B.N., Hammad, H. & Fahy, J.V. The Cytokines of Asthma. *Immunity* **50**, 975-991 (2019).
39. Hinks, T.S.C., Levine, S.J. & Brusselle, G.G. Treatment options in type-2 low asthma. *The European respiratory journal* **57**, 2000528 (2021).
40. Carmichael, J. *et al.* Corticosteroid resistance in chronic asthma. *Br Med J (Clin Res Ed)* **282**, 1419-22 (1981).
41. Guidelines for the diagnosis and management of asthma. National Heart, Lung, and Blood Institute. National Asthma Education Program. Expert Panel Report. *J Allergy Clin Immunol* **88**, 425-534 (1991).
42. Busse, W.W. *Definition and impact of severe asthma*, (2019).
43. Asthma, G.I.f. Global Strategy for Asthma Management and Prevention. (1995).
44. Asthma, G.I.f. GLOBAL STRATEGY FOR

ASTHMA MANAGEMENT AND PREVENTION. (2002).

45. Asthma, G.I.f. Global Strategy for Asthma Management and Prevention. (2017).
46. Asthma, G.I.f. Global Strategy for Asthma Management and Prevention. (2021).
47. National Asthma Education and Prevention Program. Guidelines for the Diagnosis and Management of Asthma: Expert Panel Report 2. Bethesda, National Heart, Lung and Blood Institute. National Institutes of Health

J Allergy Clin Immunol (1997).

48. Chung, K.F. *et al.* International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* **43**, 343-73 (2014).
49. Hekking, P.W. *et al.* The prevalence of severe refractory asthma. *J Allergy Clin Immunol* **135**, 896-902 (2015).
50. UK, A.L. Biologic therapies for severe asthma. (2021).
51. Wenzel, S.E. Severe Adult Asthmas: Integrating Clinical Features, Biology, and Therapeutics to Improve Outcomes. *Am J Respir Crit Care Med* **203**, 809-821 (2021).
52. Zhou, J.P. *et al.* Long-term efficacy and safety of bronchial thermoplasty in patients with moderate-to-severe persistent asthma: a systemic review and meta-analysis. *J Asthma* **53**, 94-100 (2016).
53. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
54. Biobank, U. Enable your research.
55. Biobank, U. Retrospective timeline of the data currently available.
56. Us, A.o. All of Us.
57. RIKEN. Biobank Japan.
58. Biobank, C.K. China Kadoorie Biobank.
59. Million Veteran Program.
60. eMerge.
61. Wei, W.Q. & Denny, J.C. Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Med* **7**, 41 (2015).
62. NHS-BSA. Respiratory dashboard. (2022).
63. Shaw, D.E. *et al.* Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *European Respiratory Journal* **46**, 1308-1321 (2015).
64. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).
65. Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819-1829 (2001).
66. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* **46**, 100-106 (2014).
67. Plinkv2.
68. Johnson, J.L. & Abecasis, G.R. GAS Power Calculator: web-based power calculator for genetic association studies. *bioRxiv*, 164343 (2017).
69. Hakonarson, H. *et al.* Allelic frequencies and patterns of single-nucleotide polymorphisms in candidate genes for asthma and atopy in Iceland.
70. Kuruvilla, M.E., Vanijcharoenkarn, K., Shih, J.A. & Lee, F.E. Epidemiology and risk factors for asthma.
71. Turner. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. . *Journal of Open Source Software* **3**, 731.
72. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015).
73. Boughton, A.P. *et al.* LocusZoom.js: Interactive and embeddable visualization of genetic association study results.
74. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
75. Machiela, M.J. & Chanock, S.J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.
76. EMBL-EBI. Ensembl Variation - Phenotype sources. (2022).
77. Ferreira, M.A.-O. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology.

78. Ferreira, M.A.-O. *et al.* Age-of-onset information helps identify 76 genetic variants associated with allergic disease.
79. Weidinger, S. *et al.* A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis.
80. xihaoli. STAARpipeline.
81. xihaoli. STAARpipelineSummary.
82. xihaoli. MetaSTAAR.
83. Vitsios, D., Dhindsa, R.S., Middleton, L., Gussow, A.B. & Petrovski, S. Prioritizing non-coding regions based on human genomic constraint and sequence context with deep learning. *Nature Communications* **12**, 1504 (2021).