INVITED REVIEW SERIES:
MOLECULAR TECHNIQUES FOR RESPIRATORY DISEASES
SERIES EDITORS: IAN A. YANG AND YUBEN MOODLEY

Genetic risk factors for the development of pulmonary disease identified by genome-wide association

ROBERT HALL, IAN P. HALL AND IAN SAYERS ⓘ

Division of Respiratory Medicine, NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK

ABSTRACT

Chronic respiratory diseases are a major cause of morbidity and mortality. Asthma and chronic obstructive pulmonary disease (COPD) combined affect over 500 million people worldwide. While environmental factors are important in disease progression, asthma and COPD have long been known to be heritable with genetic components playing an important role in the risk of developing disease. Identification of genetic variation contributing to disease progression is important for a number of reasons including identification of risk alleles, understanding underlying disease mechanisms and development of novel therapies. Genome-wide association studies (GWAS) have been successful in identifying many loci associated with lung function, COPD and asthma. In recent years, meta-analyses and improved imputation have facilitated the growth of GWAS in terms of numbers of subjects and the number of single nucleotide polymorphisms (SNP) that can be interrogated. As a consequence, there has been a significant increase in the number of signals associated with asthma, COPD and lung function. SNP that have shown association with lung function reassuringly show a significant overlap with SNP associated with COPD giving a glimpse at pathways that may be involved in COPD mechanisms including genes in, for example, developmental pathways. In asthma, association signals are often in or near genes involved in both adaptive and innate immune response pathways, epithelial cell homeostasis and airway structural changes. The challenges now are translating these genetic signals into a new understanding of lung biology, understanding how variants impact health and disease and how they may provide opportunities for therapeutic intervention.

INTRODUCTION

Chronic obstructive respiratory diseases are common and are leading causes of morbidity and mortality. The most common are chronic obstructive pulmonary disease (COPD) and asthma. Worldwide in 2015, COPD and asthma resulted in the death of approximately 3.2 million and 400 000 people, respectively, while affecting over 500 million people combined.[1] Asthma is characterized by chronic inflammation, reversible airway obstruction and airway hyper-responsiveness, whereas COPD is defined by irreversible obstruction of the airways with emphysema and/or chronic bronchitis.[2,3]

Currently, there is no cure for asthma or COPD and treatment focuses on the management of symptoms which can successfully improve the quality of life of patients. Recently, greater understanding of disease mechanisms and the specific subtypes of disease have led to the introduction of more tailored treatments such as anti-IgE (immunoglobulin E) and/or anti IL-5 therapy in asthma.[4] However, further insight is needed into disease mechanisms in order to develop more effective therapies.

COPD and asthma are heterogeneous diseases with both environmental and genetic factors contributing to disease development and progression. Smoking tobacco is well known as the major risk factor for developing COPD; however, only 10–20% of heavy smokers develop COPD.

Identification of genetic variants and disease genes is important for a number of reasons including (i) to identify risk alleles which might in time help with early diagnosis and possible disease prevention; (ii) the development of novel therapeutic interventions; and (iii) the identification of subgroups of patients who are most likely to gain benefit from particular therapies, an approach usually referred to as personalized or precision medicine. In addition, these studies can help to unravel the complex gene–environment interactions that lead to disease.

In the current review, we outline the significant progress that has been made in identifying genetic variants of relevance to asthma and COPD with a particular

focus to those identified in genome-wide association studies (GWAS).

Heritability of lung function, COPD and asthma

It is well established that asthma and COPD run in families. A twin study in 34 782 Danish twins suggested that genetic factors accounted for 34% of the variation of asthma age of onset with the risk of asthma for a co-twin of an affected twin being higher in monozygotic than in dizygotic twins.[5]

Lung function measurements are used in the diagnosis of obstructive pulmonary diseases making them useful quantitative traits for studying the genetics of COPD with the additional statistical power beyond case/control approaches. The most commonly used measures of lung function are forced expiratory volume in 1 s (FEV_1), the volume of air that can forcibly be blown out in 1 s; forced vital capacity (FVC), the volume of air that can forcibly be blown out after full inspiration; and the FEV_1/FVC ratio. Population-based studies on families with respiratory disease have provided evidence for familial aggregation of spirometric measurements and the heritability of lung function measures have been estimated to range anywhere from 10% to 80%.[6-10]

The first gene identified to be associated with COPD, specifically emphysema, was *SERPINA1* which encodes alpha-1 antitrypsin (A1AT). A1AT deficiency, caused by mutations in A1AT, affects up to 1 in 5000 individuals in Europe.[11] This deficiency was first identified in 1963 by Laurell and Eriksson who described individuals who have approximately 10% of normal A1AT levels in their serum. While some individuals with A1AT deficiency have emphysema, only about 1–3% of COPD sufferers have the deficiency.[12] Therefore, despite the well-understood genetic and molecular causes of COPD in this small subset of patients, it is still largely unknown what the underlying disease mechanisms in the majority of sufferers are.

METHODS FOR STUDYING GENETICS OF RESPIRATORY DISEASE

Until recently, identifying genes and variants involved in respiratory disease has been challenging. This is in part due to the heterogeneity in patient phenotypes. It is also due to the likelihood that complex, common diseases are driven by a combination of common genetic variants each contributing small phenotypic effects combined with a significant environmental component. This means that variation in phenotype is likely due to a combination of different genotypes making unravelling genetic causes of disease difficult.

Early genetics: Linkage and candidate gene studies

Linkage analysis is a method of genetic mapping which allows the identification of genetic variants that cosegregate with disease through generations based on their chromosomal location. Linkage analysis has been effective in identifying genetic variants causing monogenic disorders (e.g. cystic fibrosis). However, this method has proved less useful for asthma and COPD due to the lack of sufficient power and the complex genetic basis of these diseases.

Despite these limitations, there were a number of studies which found regions of linkage with lung function and COPD phenotypes on chromosomes 2, 4, 6, 8, 12 and 19.[13-17] Similarly, in asthma, several loci were identified including but not limited to *IL13*, *TNF*, *ADAM33*, *IL4RA*, *DPP10*, *PHF11*, *NPSR1*, *HLA-G*, *CYFIP2, IRAK3, COL6A5, OPN3/CHML* and *TBXA2R*.[18,19]

Candidate gene studies are a hypothesis-driven approach in which genes postulated to impact disease, due to known role in disease biology and/or chromosomal location, are investigated. In asthma and COPD, over 100 genes have been identified as potentially relevant; however, these studies have been limited by power and lack of replication of findings. For more details of early genetic studies, see excellent reviews for lung function and COPD[20] and asthma.[18]

Genome-wide association studies

In 2006, Dewan *et al.* performed a GWAS identifying a variant in the promoter region of *HTRA1* to be significantly associated with macular degeneration.[21] Since this first proof-of-concept study, GWAS have become highly effective for the discovery of disease genes across a large number of complex traits and diseases due to the rapid advances in genotyping technology and projects such as the Human Genome Project and the HapMap project.[22-25]

The basic concept of GWAS is that hundreds of thousands of common single nucleotide polymorphisms (SNP) (generally >5% allele frequency) are tested for association with a specific trait; for example, a lung function phenotype such as FEV_1 or a disease category such as asthma. A typical GWAS utilizes a case–control design whereby the allele frequency of SNP in the cases (individuals with the trait or disease) is compared to the allele frequency of SNP in the controls (Fig. 1).

Individuals are genotyped on an SNP array, which is a chip-based microarray technique that is used to assay hundreds of thousands to millions of SNP. Currently, there are two platforms commonly used for GWAS, which differ slightly in their methods to detect SNP variation in genomic DNA. First, Affymetrix (Central Expressway, Santa Clara, CA, USA) uses an approach by which short DNA sequences are printed on a spot on the chip which recognizes a specific allele by hybridization. Alternatively, the Illumina (Illumina Way, San Diego, CA, USA) platform uses a bead-based approach with slightly longer DNA sequences to detect specific alleles.[26]

A standard GWAS involves examining association of >500 000 SNP spanning the entire genome as well as the platform, the individual SNP to be assayed is an important consideration as it is not currently feasible to assay all known SNP in the human genome. To capture the maximum amount of information possible, panels are used that contain SNP that tag haplotype blocks. A haplotype block is a group of SNP that are statistically likely to be inherited together. Therefore, by choosing to include SNP on the panel that tag haplotype blocks, it is possible to infer the genotype for many more SNP
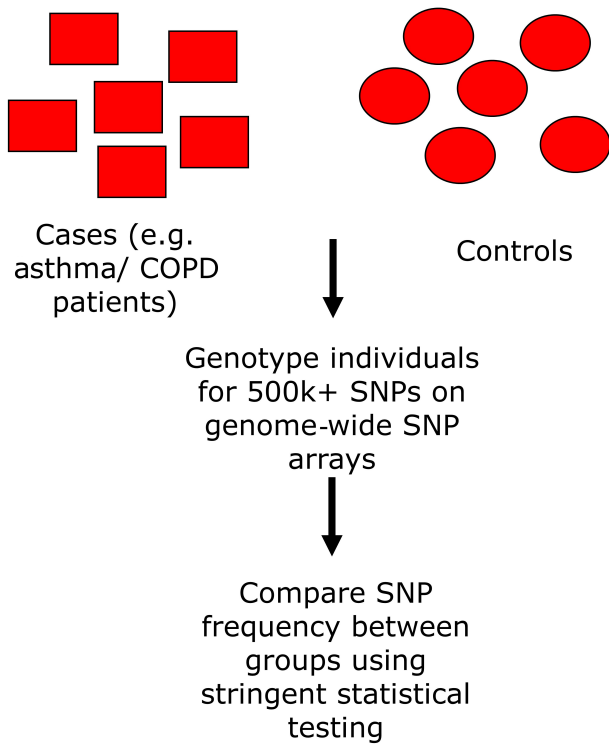
Figure 1 Genome-wide association studies test the frequency of common SNP between cases and controls. SNP, single nucleotide polymorphism.

than those that are actually on the panel. Other considerations for SNP selection generally are specific to the type of study. For example, the UK BiLEVE custom array was designed to (i) measure rare coding variants (variants which alter amino acid sequences); (ii) provide a framework which would allow optimum imputation of non-genotyped common or low frequency variants (therefore increasing coverage) and (iii) optimize coverage of genes and loci which have already been linked to respiratory health and/or disease (allows fine mapping of regions of interest).[27]

The allele frequency of each SNP is tested for association with the trait in question, requiring a stringent level of statistical significance due to the large number of tests. Typically $P < 5 \times 10^{-8}$ is considered significant to show genome-wide association. This stringent threshold is important to avoid the detection of false positives, which is more likely when there are a large number of variants (and therefore hypotheses) being tested. A genome-wide scan of common variants is equivalent to approximately 1 million independent tests, hence this value represents Bonferroni correction.[28,29] This level of stringency therefore requires large sample sizes for a signal to reach genome-wide significance.

Although, because of the improved ability to impute genotypes which are not directly genotyped, most recent studies where minor allele frequencies of less than 5% are being assessed, the required significance level would be set at $P < 5 \times 10^{-9}$.

Replication of findings in an independent cohort is also standard practice to validate findings. Once SNP are validated for association, further in silico, in vitro

and in vivo studies are required to explore the biological mechanisms underlying the association.

The advantage of GWAS is that they are mostly hypothesis free in the sense that associations are searched for over the entire genome. However, they are limited in that association signals are not directly informative with respect to the biology underlying the association signal and therefore further analysis and functional genetics are required to elucidate the causal variant(s) and mechanisms underlying the signal.[30]

Early GWAS

For asthma, the first GWAS was published in 2007 using 994 asthmatic patients and 1243 controls which identified that variants which regulate the expression of *ORMDL3* are associated with an increased risk of childhood asthma.[31] Similarly, early GWAS of lung function showed some success in identifying a number of loci associated with lung function and COPD. These loci include *HHIP* at the 4q31 locus, *CHRNA3/5* at 15q25 and *FAM13A* at 4q22.[32–34]

Meta-analyses and imputation

Early GWAS showed great success in identifying novel loci associated with lung function, COPD and asthma; however, these studies also identified that greater statistical power would be required to identify genetic variation associated with modest effect sizes. As a result of these studies, meta-analyses were used to analyse the results of independent GWAS together thereby increasing the statistical power for novel loci discovery.

Imputation describes a method whereby genetic variants which were not directly genotyped can be predicted with a degree of confidence.[35] This is done using reference genomes which are now widely available, thanks to studies such as the 1000 genomes, 10 000 genomes and the HapMap projects.[36] Imputation gives increased power because the reference panel is more likely to contain the causal variant than the original array while also facilitating fine-mapping of the region by giving a high-resolution overview of an association signal over a locus. In addition, imputation allows GWAS typed with different arrays to be combined to the reference panel, which is essential to combine independent GWAS for meta-analyses.

GWAS in lung function and COPD

In 2010, the SpiroMeta and Charge consortium published back-to-back papers describing meta-analyses of GWAS on FEV_1 and FEV_1/FVC.[37,38] These two studies had significantly greater numbers of subjects than previously used with the SpiroMeta study using 20 288 individuals for discovery analysis. This was followed up by a meta-analysis of 32 184 genotyped individuals with in silico association data from the Charge consortium's 21 092 individuals and an additional 883 individuals from the Health 2000 study.[38] The SpiroMeta study replicated the association locus at 4q31 (*HHIP*) and identified five additional novel loci associated with either FEV_1 or FEV_1/FVC in or near *GSTCD, HTR4,*

AGER (Advanced Glycation End Product Receptor) and THSD4.[38] The Charge meta-analyses included 20 890 participants of European ancestry in the discovery phase and 20 228 individuals from the SpiroMeta consortium in the replication cohort.[37] This study identified another nine loci associated with lung function. HHIP, GPR126, ADAM19, AGER-PPT2, FAM13A, PTCH1, PID1 and HTR4 were associated with FEV_1/FVC while one locus was associated with FEV_1 which was INTS12-GSTCD-NPNT, replicating results from the SpiroMeta study.[37,38]

Since these first meta-analyses of GWAS of lung function, the scale of the studies has increased significantly with respect to both the number of subjects and the number of markers and this has facilitated the growth in the number of signals found. For example, following up these studies in 2011, meta-analyses of 23 lung function GWAS from the SpiroMeta and CHARGE consortiums were published which used a total of 48 201 individuals. The result was the identification of 16 additional novel loci for lung function including 12 novel loci for FEV_1/FVC, 3 novel loci for FEV_1 and 1 locus for both traits.[39] In addition to the novel loci, the study provided further evidence for 10 previously identified loci.[39]

In more recent years, meta-analysis using SpiroMeta and CHARGE GWAS was performed by Loth et al. in 2014, this time focused to FVC. This analysis used 52 253 individuals from 26 studies, following up the top associations in 32 917 individuals. They identified six novel loci associated with FVC at EFEMP1, BMP6,

MIR129-2-HSD17B12, PRDM11, WWOX and KCNJ2 as well as association with known signals at GSTCD and PTCH1.[40] While most studies have focused on cross-sectional lung function, a study of longitudinal change in adult lung function provided suggestive evidence for association at novel loci IL16/STARD5/TMC3 on chromosome 15 and ME3 on chromosome 11.[41]

A further 16 novel signals were identified in a 2015 SpiroMeta paper in or near ENSA, RNU5F1, KCNS3, AK097794, ASTN2, LHX3, CCDC91, TBX3, TRIP11, RIN3, TEKT5, LTBP4, MN1 and AP1S2 including novel signals in known loci GPR126 and NPNT.[42] A study around the same time in the UK Biobank cohort identified six novel genome-wide significant signals of association with extremes of FEV_1. This study selected individuals from the middle and extremes of the FEV_1 distribution among heavy smokers and never-smokers. Signals of association were found at four novel loci KANSL1, TSEN54, TET2 and RBM19/TBX5 and two novel signals from previously reported loci NPNT and HLA-DQB1/HLA-DQA2.[43]

The most recent published large-scale lung function GWAS almost doubled the number of signals that have been associated with lung function by identifying 43 new signals, increasing the number of independent signals to 97 (Fig. 2).[43] There was enhancement in genes for epigenetic regulation pathways, elastic fibres and development.[43]

As one would expect, many of these signals have also shown an association with COPD (Fig. 3).[27,32,34,37–40,42–47]
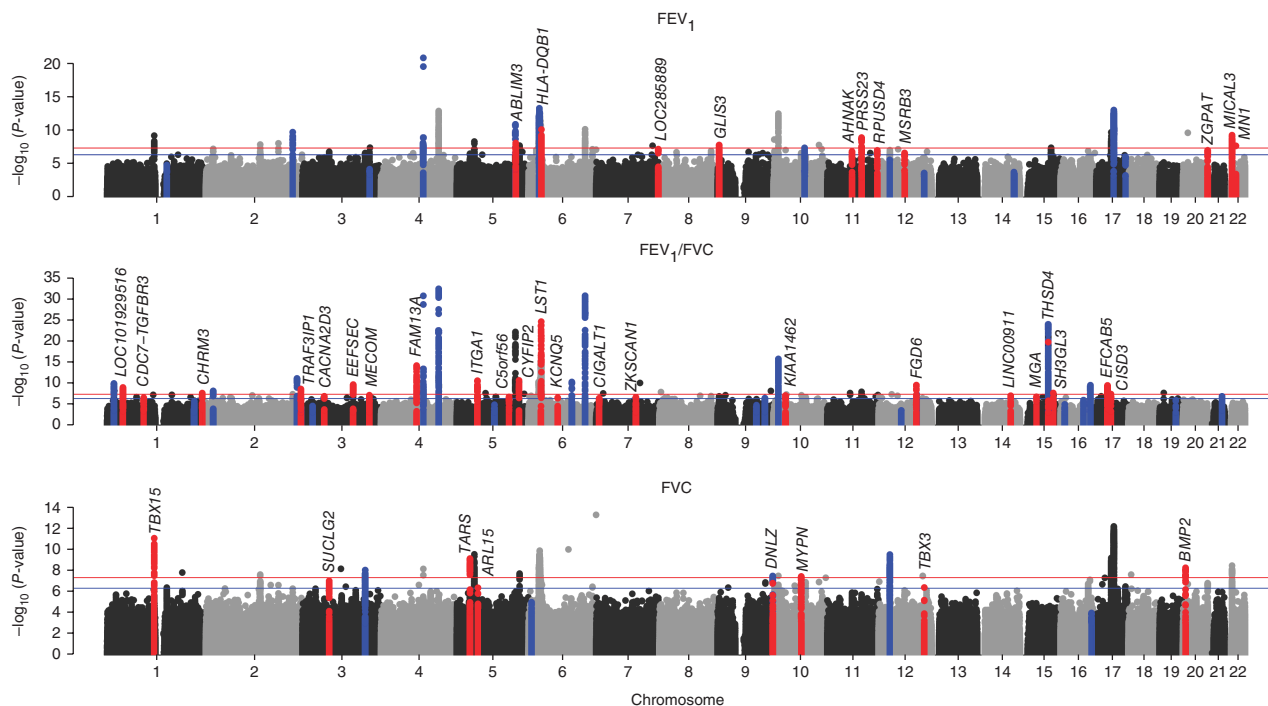

Figure 2 Genome-wide association results for FEV_1 (top), FEV_1/FVC (middle) and FVC (bottom). Previously reported signals are highlighted in dark blue (except signals with $P > 5 \times 10^{-4}$ in this study); and novel signals are coloured in red. Signals are highlighted for the trait with which they showed strongest association only. The red and blue lines correspond to the genome-wide significance level ($P = 5 \times 10^{-8}$, $-\log_{10}P = 7.3$) and the threshold used to select signals for follow-up in stage 2 ($P = 5 \times 10^{-7}$, $-\log_{10}P = 6.3$), respectively. Labels show the nearest gene to the novel sentinel variants. FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity. (Reproduced from Wain et al.,[43] with permission). Black and grey lines are used to show individual chromosomes.
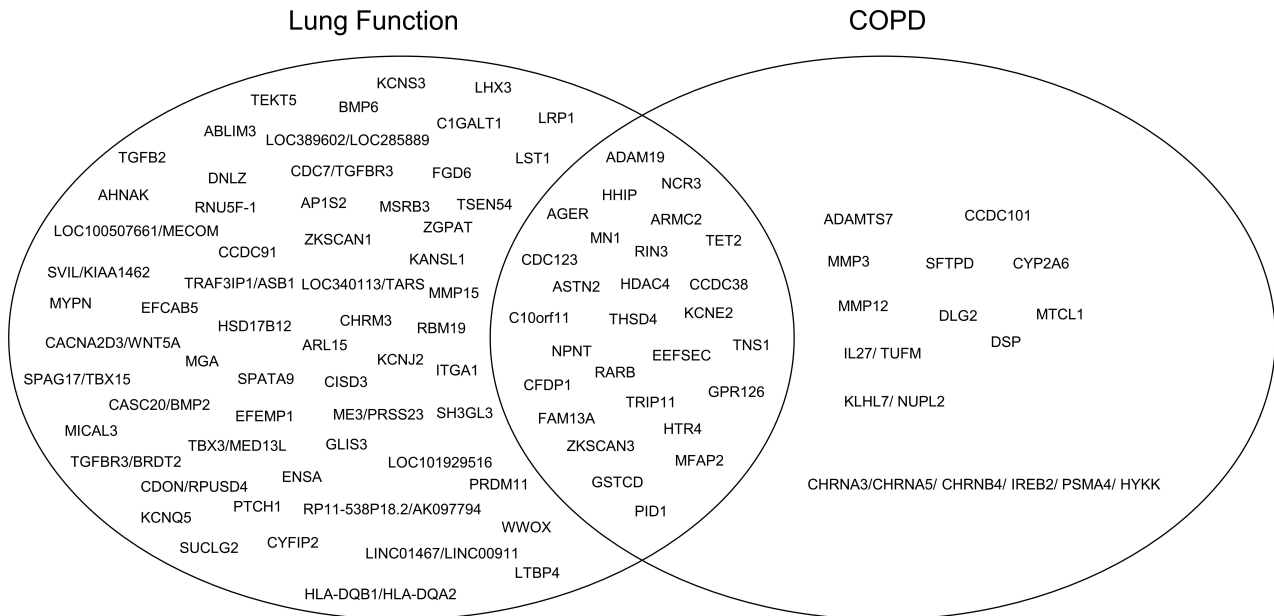
Figure 3 Reported genes associated with lung function and COPD. The diagram shows the genes which have association signals for lung function and COPD and the genes which have signals for both. The signals displayed are those which showed the highest levels of association, while additional signals which showed less significant association are not shown. In addition to this overlap, the genetic risk score based on 95 lung function variants is highly predictive of COPD risk.[43] COPD, chronic obstructive pulmonary disease.

A number of studies have succeeded in finding disease-specific SNP associated with COPD phenotypes including emphysema, COPD exacerbations, chronic bronchitis, airway responsiveness and severity of disease, for example AGER SNP rs2070600 and emphysema and gas trapping.[48] Significant association has been found for a number of loci already associated with lung function with COPD diagnosis including *FAM13A, TNS1, INTS12/GSTCD/NPNT, HTR4, HHIP, AGER, TSHD4, CHRNA5/CHRNA3* and *ADAM19*.[34,43,45–47,49–53]

A 2017 GWAS using 15 256 COPD cases and 47 936 controls identified 22 loci at genome-wide significance which included 13 new COPD associations, 9 of which have previously been associated with lung function.[44] The remaining four loci (*EEFSEC, DSP, MTCL1* and *SFTPD*) were novel.[44] Also in 2017, Wain *et al.* analysed 95 of 97 known independent signals associated with lung function for replication with association with COPD. Of the 95, 51 showed nominal association ($P < 0.05$) and 30 showed association with COPD susceptibility.[43] This finding that there is an overlap between genetic variants associated with lung function and COPD provides confidence that (i) the locus is a true association and (ii) it is important for COPD. By knowing that a variant contributes to lung function and a specific COPD phenotype, we can begin to discover the biological mechanisms that underlie the association with disease.

GWAS in asthma

Asthma is a heterogeneous condition involving a number of sub-phenotypes, which may have different clinical presentations, disease mechanisms, underlying causes and genetic associations. As of May 2018, there have been 28 GWAS on asthma according to the National Human Genome Research Institute (NHGRI)-European Bioinformatics Institute (EBI) GWAS catalogue (Table 1).[69] Multiple studies have begun to try to unravel the complex underlying causes of these subphenotypes. For example, the *ORMDL3/GSDMB/ZPB2* locus on chromosome 17 is associated with childhood asthma, severe asthma and asthma with allergic rhinitis.[31,56,63,68] *IL33* is involved in the recruitment and activation of inflammatory cells, and SNP in this gene region are associated with childhood severe asthma with exacerbation and asthma with an allergic rhinitis diagnosis.[56,58,62,63] Interestingly, SNP spanning *IL1RL1*, the IL-33 receptor, are associated with childhood asthma, severe asthma and with a diagnosis of allergic rhinitis.[56,63,68] For an excellent review of GWAS in asthma up until 2016, see the study by Vicente *et al.*[70]

A 2017 study found a further four novel loci after a meta-analysis on 21 644 European American and African American individuals.[67] 6p21.31, 9p21.2 and 10q21.3 were the loci identified in the European American population with *TEK* at the 9p21.2 locus of specific interest due to its known involvement in remodelling in the lung.[71,72] In the African American population, *PTGES* on 9q34.11 was identified as a novel locus.[67]

To date, the largest GWAS of asthma was completed in 2016 and included 28 399 cases of self-reported asthma and 128 843 non-asthmatic controls and identified 27 loci contributing to the risk to develop asthma.[54] This study identified many of the previous associations and identified new associations at *ADAMTS4, D2HGDH, CLEC16A, LRP1, ADORA1, BACH2, PEX14* and *STAT6*.[54] Several of these loci have recently been replicated using 23 948 self-reported asthma cases and

Table 1 Genetic loci associated with asthma and asthma phenotypes in GWAS

Chromosome	Phenotype (s)	Gene(s) reported	Reference
1	Asthma	PEX14	Pickrell (2016)[54]
1	Asthma	IL6R	Ferreira (2011)[55]
1	Asthma	PYHIN1	Torgerson (2011)[56]
1	Asthma	ADAMTS4	Pickrell (2016)[54]
1	Asthma	CD247	Pickrell (2016)[54]
1	Asthma	TNFSF18	Pickrell (2016)[54]
1	Moderate–severe childhood asthma	DENND1B	Sleiman (2010)[57]
1	Asthma	ADORA1	Pickrell (2016)[54]
2	Asthma	ID2	Pickrell (2016)[54]
2	Asthma, childhood asthma + exacerbations, asthma + hay fever	IL1R1/IL18R1	Bonnelykke (2014),[58] Pickrell (2016),[54] Ramasamy (2012),[59] Torgerson (2011),[56] Demenais (2018),[60] Gudbjartsson (2009),[61] Ferreira (2014),[62] Moffatt (2010)[63]
2	Asthma	D2HGDH	Pickrell (2016)[54]
3	Asthma	LPP	Pickrell (2016)[54]
4	Asthma, asthma + hay fever	TLR1	Pickrell (2016),[54] Ferreira (2014)[62]
4	Asthma	USP38	Hirota (2011)[64]
5	Mild–moderate childhood asthma with BHR	PDE4D	Himes (2009)[65]
5	Asthma, asthma + hay fever	TSLP/SLC225A46/WDR36	Torgerson (2011),[56] Hirota (2011),[64] Ferreira (2014),[62] Pickrell (2016),[54] Demenais (2018),[60] Gudbjartsson (2009)[61]
5	Asthma, childhood asthma + exacerbations	IL13/RAD50/IL4	Pickrell (2016),[54] Bonnelykke (2014),[58] Demenais (2018)[60]
5	Asthma	NDFIP1/GNDPA1/SPRY4	Demenais (2018),[60] Pickrell (2016)[54]
6	Asthma	GPX5/TRIM27	Demenais (2018)[60]
6	Asthma, asthma + hay fever, childhood onset asthma	HLAC/NOTCH4/MICB/HCP5/ MCCD1/HLADRB5/HLADRB1/ HLADQA1/HLA-DQB1/HLA-DPA1/ GRM4	Pickrell (2016),[54] Hirota (2011),[64] Demenais (2018),[60] Ferreira (2014),[62] Noguchi (2011),[66] Almoguera (2017)[67]
6	Asthma	BACH2, GJA10, MAP3K7	Pickrell (2016),[54] Demenais (2018)[60]
7	Asthma, childhood asthma + exacerbations	CDHR3	Bonnelykke (2014),[58] Pickrell (2016)[54]
8	Childhood onset asthma	SLC30A8	Noguchi (2011)[66]
8	Asthma, asthma + hay fever	TPD52, ZBTB10	Demenais (2018),[60] Pickrell (2016),[54] Ferreira (2014)[62]
9	Asthma, childhood asthma + exacerbations, asthma + hay fever	IL33/RANBP6	Ferreira (2014),[62] Moffatt (2010),[63] Torgerson (2011),[56] Pickrell (2016),[54] Demenais (2018),[60] Bonnelykke (2014)[58]
9	Asthma	EQTN	Almoguera (2017)[67]
10	Asthma	GATA3/CELF2	Hirota (2011),[64] Demenais (2018),[60] Pickrell (2016)[54]
11	Asthma + hay fever	LRRC32/C11orf30/EMSY	Ferreira (2011),[55] Ferreira (2014),[62] Pickrell (2016),[54] Demenais (2018)[60]
12	Asthma	IKZF4/STAT6/NAB2/LRP1	Hirota (2011),[64] Demenais (2018),[60] Pickrell (2016)[54]
14	Asthma	RAD51B	Pickrell (2016)[54]
15	Asthma	RORA/NARG2/VPS13C	Pickrell (2016),[54] Demenais (2018),[60] Moffatt (2010)[63]
15	Asthma, asthma + hay fever	SMAD3/SMAD6	Moffatt (2010),[63] Pickrell (2016),[54] Demenais (2018),[60] Ferreira (2014)[62]
16	Asthma, asthma + hay fever	CLEC16A, DEXI, SOCS1	Ferreira (2014),[62] Pickrell (2016),[54] Demenais (2018)[60]

Table 1 Continued

Chromosome	Phenotype (s)	Gene(s) reported	Reference
17	Asthma, childhood asthma + exacerbations, asthma + hay fever, childhood asthma, moderate–severe asthma	*ERBB2/PGAP3/MIEN1/ZPBP2/ GSDMB/ORMDL3/GSDMA*	Demenais (2018),[60] Pickrell (2016),[54] Moffatt (2010),[63] Bonnelykke (2014),[58] Torgerson (2011),[56] Moffatt (2007),[31] Wan (2012),[68] Ferreira (2014)[62]
17	Asthma	*ZNF652, PHB*	Demenais (2018)[60]
22	Asthma	*IL2RB*	Moffatt (2010)[63]

BHR, Bronchial hyperresponsiveness; GWAS, genome-wide association studies.

118 538 non-asthmatic controls from multiple ethnic groups which reported 18 loci (22 independent signals) and novel loci at *NDFIP1*, *GPX5* and *ZNF652*.[60]

TRANSLATIONAL APPROACHES: GWAS TO FUNCTION

GWAS have identified many signals relating to respiratory disease, the challenge now is translating these findings into biological insight. There are a number of considerations to take into account when selecting an association locus for follow-up analysis. Depending on the variant in question, the types of analyses will vary significantly. In brief, a few considerations are whether the SNP is coding or non-coding and whether there are other SNP that are in linkage disequilibrium with the sentinel SNP. Similarly, it must be considered if the SNP (or SNPs) are expression quantitative trait loci (eQTL), that is there is a relationship between the genotype and mRNA levels, which would suggest regulatory function. This analysis been done on a large scale with SNP identified as regulating mRNA levels in lung tissue overlapping with SNP identified associated with lung function demonstrating a clear enrichment of cis and trans eQTL in the associated SNP.[73] The various translational approaches and considerations have been extensively reviewed by Kheirallah *et al.*[20] For the purposes of this review, we will use *AGER* SNP rs2070600 as a case study of a coding region variant and variants upstream of *HHIP* as a case study of a non-coding variant.

Coding region variants

The *AGER*, encoding the RAGE protein, contains coding region variant rs2070600 which has been associated with FEV_1/FVC (Fig. 4).[37,38] The T allele of this variant results in a substitution in the amino acid sequence of RAGE from a glycine at position 82 to a serine.[37,38] Further analyses also showed the variant to be associated with COPD, emphysema and gas trapping and lung cancer, highlighting the need for further analysis of the role of the gene and variant Ser82 in COPD.[49,50,74,75]

Miller *et al.* conducted a study showing that the Ser82 variant had effects on lung function as well as acting as an eQTL for soluble RAGE in both the serum of UK smokers and RAGE production in a recombinant cell system.[76] RAGE is a member of the Ig superfamily which when activated, by a diverse range of ligands, has many downstream effects including inflammatory

regulation and cell migration, proliferation and adhesion. RAGE expression was identified by immunohistochemistry in the alveoli and bronchi of control and COPD lungs. *AGER* mRNA is also expressed in the foetal lung, with increasing levels across the pseudoglandular and canalicular stages of human lung development. This finding correlated with protein expression during lung development.[76] The study also showed that there was an association between rs2070600T and higher FEV_1 and FEV_1/FVC as well as providing evidence that individuals with the Ser82 variant had significantly lower levels of serum sRAGE compared to Gly82 genotype individuals.[76]

A reductionist RAGE recombinant cell model was used to study the function effects of the different alleles on the same genetic background and it was shown that the Ser82 variant of the receptor resulted in lower levels of secretion of soluble RAGE, demonstrating that the specific variant has functional consequences.[76] Rather counterintuitively, serum or plasma levels of
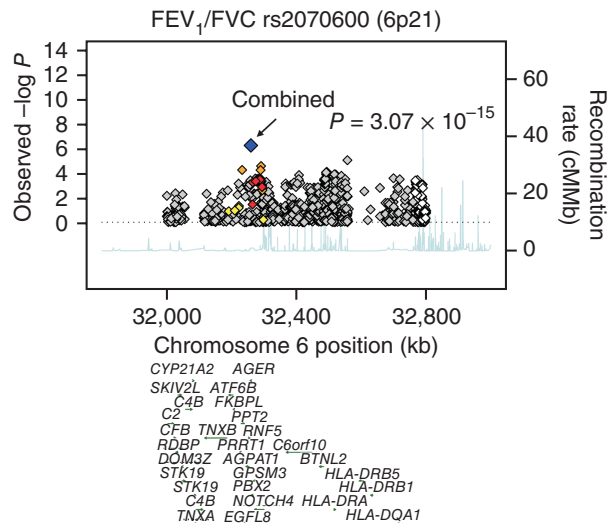

Figure 4 Region plot showing SNP associated with FEV_1/FVC in the *AGER* locus from the original SpiroMeta study. The x-axis represents the chromosomal location and is annotated with gene names and the y-axis shows the *P*-value, with more significant SNP higher up the axis. The key SNP (rs2070600) in this region is highlighted (♦). AGER, Advanced Glycation End Product Receptor; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; SNP, single nucleotide polymorphism. ♦, $r^2 > 0.8$; ♦, $r^2 > 0.5$; ♦, $r^2 > 0.2$; ♦, $r^2 < 0.2$; ◊, r^2 unknown (Reproduced from Repapi *et al.*,[38] with permission).

sRAGE in COPD patients are lower compared to controls. This led to the hypothesis that lower levels of sRAGE, which itself acts as a decoy to limit inflammation, actually may lead to greater RAGE-driven inflammation in COPD.[76] While the relationship between genetics, lung function, sRAGE levels and COPD is not fully defined, these findings provide evidence that the Ser82 variant, discovered through GWAS, is important for RAGE function in COPD and began to explain the underlying biological processes behind this association.

Non-coding region variants

The majority of SNP identified through GWAS localize outside of protein coding regions, indicating that regulation of gene expression is likely to be important in disease processes.[77] As an example, a number of SNP near the *HHIP* gene on chromosome 4q31 have been strongly associated with pulmonary function and COPD.[33,37,38] *HHIP* codes for a protein which is likely to be an important regulator in a number of pathways including those important for lung development.

HHIP mRNA and protein expression are reduced in COPD lung tissues. Two SNP, from the haplotype associated with COPD risk, rs6537296 and rs1542725, lie within an *HHIP* enhancer region.[78] These SNP have been shown to modulate expression of *HHIP*, suggesting that reduced expression of *HHIP* may be involved in the pathogenesis of COPD.[78] To determine the biological mechanism underlying the influence of *HHIP* on COPD susceptibility, a knockdown cell model was used.[79] Silencing of the gene resulted in differential expression of nearly 300 genes in the bronchial epithelial cell line Beas-2B.[79] Interestingly, there was enrichment for variants associated with COPD and seven of these genes were found to be differentially expressed in human COPD and control lung tissues.[79] Additional analyses found enrichment for extracellular matrix genes and genes involved in cell growth.[79] The functionality of the specific SNP in the long range enhancer was also demonstrated using promoter–reporter approaches.[79] The authors showed that *HHIP* regulates extracellular matrix remodelling and cell growth pathways and hypothesized that by doing so, *HHIP* might influence smoke-induced COPD susceptibility.[79] The *HHIP* case demonstrates how SNP in a non-coding region can begin to reveal mechanisms underlying disease process and highlight key pathways and genes for further study.

CONCLUSION AND FUTURE DIRECTIONS

To date, a large number of loci have been associated with lung function, COPD and asthma using GWAS particularly through recent consortia approaches maximizing sample sizes and improved imputation. While the last decade or so has been promising in terms of number of genes and variants associated with respiratory disease, there is still considerable work to do to translate these findings into clinical practice. In the future, translation of GWAS signals will be further facilitated by improved annotation of the human genome and improved eQTL data sets as well as advances in molecular and cell biology techniques such as clusters of regularly interspaced short palindromic repeats (CRISPR)/Cas9 gene editing to ultimately identify causal variants and mechanisms leading to new understanding and opportunities for therapies.

GWAS in asthma support genes involved in innate and adaptive immune responses being particularly important in the disease process (e.g. *MICB*, *CD247*, *IL13*, *IL33*, *LRCC32*, *HLA*, *CLEC16A* and *GATA3*), epithelial barrier/function (e.g. *IL33* and *TSLP*) or airway structural changes (e.g. *ORMDL3* and *GSDMB*). In the lung function/COPD signals, there is an overrepresentation of genes involved in developmental pathways (e.g. *CDON*, *PTCH1*, *PTHLH*, *TGFB2* and *HHIP*), elastic fibres (e.g. *EFEMP1*, *TGFB2*, *LTBP4*, *MFAP2* and *FBLN5*) and signals overlapping epigenetic marks.

Variants identified through GWAS only account for a small percentage of the heritability with the known lung function signals accounting for 9.6%, 6.4% and 14.3% of the heritability for FEV_1, FVC and FEV_1/FVC, respectively.[43] In asthma, 2.5% of heritability is estimated to be due to SNP discovered through GWAS.[70] This missing heritability is often attributed to the limitations of GWAS and it is a fact that GWAS only detect common variants and ignore other types of genomic variation such as rare variants and copy number variations. GWAS also cannot inform on gene–environment interactions, epigenetics or epistasis.[80] Of course, it may also just be likely that many variants which do not reach genome-wide significance may actually be true associations contributing to the overall heritability.

The availability of affordable whole-genome and/or whole-exome sequencing (WGS/WES) approaches promises to enhance our ability to identify genetic associations.[26,30] One limitation of the arrays is that they do not cover all genetic variants in the genome and although imputation and array designs have addressed this limitation to an extent, WGS gives the potential to identify all the variants in the genome and therefore improve our ability to fine-map regions of interest.[30]

A recent paper used WGS to sequence 821 individuals with severe COPD and 973 controls.[81] They performed single-variant and grouped-variant analyses and, importantly, assessed the overlap of variants found for COPD between array-based GWAS and sequencing. The only statistically significant variants were in the well-known and previously mentioned *HHIP* locus; however, other variants approaching significance were located in a previously reported loci, *CHRNA5*, *TNS1* and *SERPINA6*/*SERPINA1*.[81] Importantly, WGS identified >20 million novel variants which included >10 000 variants in regions previously identified in GWAS for COPD.[81] While the study provided evidence that novel and rare variants can be detected by sequencing, it also highlighted the need for considerably larger sample sizes to find variants of significance like those used in meta-analyses of GWAS.

With advances in technology and decrease in the price of sequencing in the coming years, it is anticipated that SNP array GWAS will be replaced by WGS eventually. However, GWAS by SNP array with large sample sizes have proven and continues to be a powerful tool for gene discovery. Using combinations of high-

throughput methods including GWAS, sequencing, transcriptomics and proteomics is starting to provide significant advances in our understanding. These technological advances as well as greater consideration of environmental factors and improved phenotype data are keys to providing insight into these diseases at the molecular level.

Acknowledgements

The Authors

R.H. is a final year PhD student with a background in molecular and cell biology. His current project is focused on the genetics of lung disease, specifically the mechanisms by which the adhesion G protein coupled receptor GPR126 is associated with lung function. Professor I.S. has a long-term interest in molecular and cellular mechanisms underlying asthma and COPD, particularly identifying susceptibility genes and translating findings to new understanding, drug targets and stratified medicine. His laboratory receives grant support from Asthma UK, British Lung Foundation, NIHR, MRC, GSK and Boehringer Ingelheim. Professor I.P.H. has been undertaking research on mechanisms underlying lung disease for over 30 years. His particular interest is in how genetic variation contributes to risk of lung disease. He is currently Director of the NIHR Nottingham Biomedical Research Centre. His research has been supported by a range of funders including an NIHR Senior Investigator Award, the MRC and Asthma UK.

Abbreviations: A1AT, alpha-1 antitrypsin; AGER, Advanced Glycation End Product Receptor; eQTL, expression quantitative trait loci; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; GWAS, genome-wide association studies; IL, interleukin; mRNA, messenger RNA; RAGE, receptor for advanced glycation endproducts; SNP, single nucleotide polymorphism; sRAGE, soluble RAGE; WGS, whole-genome sequencing.

REFERENCES

1 Soriano JBAA, Abate KH, Abera SF, Agrawal A, Ahmed MB, Aichour AN, Aichour I, Aichour MTE, Alam K, Alam N et al. 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**: 691–706.

2 Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* 2004; **364**: 709–21.

3 Abramson MJ, Perret JL, Dharmage SC, McDonald VM, McDonald CF. Distinguishing adult-onset asthma from COPD: a review and a new approach. *Int. J. Chron. Obstruct. Pulmon. Dis.* 2014; **9**: 945–62.

4 Hilvering B, Pavord ID. What goes up must come down: biomarkers and novel biologicals in severe asthma. *Clin. Exp. Allergy* 2015; **45**: 1162–9.

5 Thomsen SF, Duffy DL, Kyvik KO, Backer V. Genetic influence on the age at onset of asthma: a twin study. *J. Allergy Clin. Immunol.* 2010; **126**: 626–30.

6 McClearn GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. *J. Gerontol.* 1994; **49**: 264–8.

7 Palmer LJ, Burton PR, James AL, Musk AW, Cookson WO. Familial aggregation and heritability of asthma-associated quantitative traits in a population-based sample of nuclear families. *Eur. J. Hum. Genet.* 2000; **8**: 853–60.

8 Palmer LJ, Knuiman MW, Divitini ML, Burton PR, James AL, Bartholomew HC, Ryan G, Musk AW. Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur. Respir. J.* 2001; **17**: 696–702.

9 Hallberg J, Iliadou A, Anderson M, de Verdier MG, Nihlen U, Dahlback M, Pedersen NL, Higenbottam T, Svartengren M. Genetic and environmental influence on lung function impairment in Swedish twins. *Respir. Res.* 2010; **11**: 92.

10 Hukkinen M, Kaprio J, Broms U, Viljanen A, Kotz D, Rantanen T, Korhonen T. Heritability of lung function: a twin study among never-smoking elderly women. *Twin Res. Hum. Genet.* 2011; **14**: 401–7.

11 Blanco I, de Serres FJ, Fernandez-Bustillo E, Lara B, Miravitlles M. Estimated numbers and prevalence of PI*S and PI*Z alleles of alpha1-antitrypsin deficiency in European countries. *Eur. Respir. J.* 2006; **27**: 77–84.

12 Janciauskiene SM, Bals R, Koczulla R, Vogelmeier C, Kohnlein T, Welte T. The discovery of alpha1-antitrypsin and its role in health and disease. *Respir. Med.* 2011; **105**: 1129–39.

13 DeMeo DL, Celedon JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2004; **170**: 1294–301.

14 Wilk JB, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, Leppert MF, Province MA, Cupples LA, Gottlieb DJ et al. A genome-wide scan of pulmonary function measures in the National Heart, Lung, and Blood Institute Family Heart Study. *Am. J. Respir. Crit. Care Med.* 2003; **167**: 1528–33.

15 Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum. Mol. Genet.* 2002; **11**: 623–32.

16 Celedon JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N et al. The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Hum. Mol. Genet.* 2004; **13**: 1649–56.

17 Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum. Mol. Genet.* 2003; **12**: 1199–210.

18 Ober C, Yao T-C. The genetics of asthma and allergic disease: a 21(st) century perspective. *Immunol. Rev.* 2011; **242**: 10–30.

19 Undarmaa S, Mashimo Y, Hattori S, Shimojo N, Fujita K, Miyatake A, Doi S, Kohno Y, Okamoto Y, Hirota T et al. Replication of genetic association studies in asthma and related phenotypes. *J. Hum. Genet.* 2010; **55**: 342–9.

20 Kheirallah AK, Miller S, Hall IP, Sayers I. Translating lung function genome-wide association study (GWAS) findings: new insights for lung biology. *Adv. Genet.* 2016; **93**: 57–145.

21 Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, Zhao C, Tam PO, Chan WM, Lam DS, Snyder M et al. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 2006; **314**: 989–92.

22 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al. The sequence of the human genome. *Science* 2001; **291**: 1304–51.

23 International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005; **437**: 1299–320.

24 Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**: 928–33.

25 Syvanen AC. Toward genome-wide SNP genotyping. *Nat. Genet.* 2005; **37**(Suppl.): S5–10.

26 Bush WS, Moore JH. Chapter 11: genome-wide association studies. *PLoS Comput. Biol.* 2012; **8**: e1002822.

27 Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, Billington CK, Kheirallah AK, Allen R, Cook JP *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**: 769–81.

28 Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 2008; **32**: 381–5.

29 Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; **322**: 881–8.

30 Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 2017; **101**: 5–22.

31 Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E *et al*. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**: 470–3.

32 Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST *et al*. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* 2009; **5**: e1000429.

33 Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A *et al*. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009; **5**: e1000421.

34 Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D *et al*. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat. Genet.* 2010; **42**: 200–2.

35 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**: 906–13.

36 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 2010; **11**: 499–511.

37 Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.* 2010; **42**: 45–52.

38 Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V *et al*. Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.* 2010; **42**: 36–44.

39 Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE *et al*. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.* 2011; **43**: 1082–90.

40 Loth DW, Soler Artigas M, Gharib SA, Wain LV, Franceschini N, Koch B, Pottinger TD, Smith AV, Duan Q, Oldmeadow C *et al*. Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat. Genet.* 2014; **46**: 669–77.

41 Tang W, Kowgier M, Loth DW, Soler Artigas M, Joubert BR, Hodge E, Gharib SA, Smith AV, Ruczinski I, Gudnason V *et al*. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One* 2014; **9**: e100776.

42 Soler Artigas M, Wain LV, Miller S, Kheirallah AK, Huffman JE, Ntalla I, Shrine N, Obeidat M, Trochet H, McArdle WL *et al*. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. *Nat. Commun.* 2015; **6**: 8658.

43 Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, Obeidat M, Henry AP, Portelli MA, Hall RJ *et al*. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* 2017; **49**: 416–25.

44 Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, Wain LV, Hall IP, Jackson VE, Wyss AB *et al*. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat. Genet.* 2017; **49**: 426–32.

45 Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, Demeo DL, Sylvia JS, Ziniti J, Laird NM *et al*. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir. Med.* 2014; **2**: 214–25.

46 Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, Burkart KM, Enright PL, Rotter JI, Petersen H *et al*. A genome-wide association study of chronic obstructive pulmonary disease in Hispanics. *Ann. Am. Thorac. Soc.* 2015; **12**: 340–8.

47 Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, Himes BE, Sylvia JS, Klanderman BJ, Ziniti JP *et al*. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum. Mol. Genet.* 2012; **21**: 947–57.

48 Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, Bakke P, Gulsvik A, San Jose Estepar R, Van Beek EJ *et al*. A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *Am. J. Respir. Crit. Care Med.* 2015; **192**: 559–69.

49 Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, Johnson T, Zhao JH, Albrecht E, Dominiczak AF *et al*. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *Am. J. Respir. Crit. Care Med.* 2011; **184**: 786–95.

50 Castaldi PJ, Cho MH, Litonjua AA, Bakke P, Gulsvik A, Lomas DA, Anderson W, Beaty TH, Hokanson JE, Crapo JD *et al*. The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. *Am. J. Respir. Cell Mol. Biol.* 2011; **45**: 1147–53.

51 Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, Smolonska J, Tang W *et al*. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am. J. Respir. Crit. Care Med.* 2012; **186**: 622–32.

52 Hansel NN, Paré PD, Rafaels N, Sin DD, Sandford A, Daley D, Vergara C, Huang L, Elliott WM, Pascoe CD *et al*.; On Behalf of the Lung Health Study. Genome-wide association study identification of novel loci associated with airway responsiveness in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* 2015; **53**: 226–34.

53 Lee JH, Cho MH, Hersh CP, McDonald ML, Crapo JD, Bakke PS, Gulsvik A, Comellas AP, Wendt CH, Lomas DA *et al*. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir. Res.* 2014; **15**: 113.

54 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**: 709–17.

55 Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souef P, Danoy P, Baltic S, Nyholt DR, Jenkins M *et al*. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. *Lancet* 2011; **378**: 1006–14.

56 Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, Himes BE, Levin AM, Mathias RA, Hancock DB *et al*. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* 2011; **43**: 887–92.

57 Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, Wang K, Rafaels NM, Michel S, Bonnelykke K *et al*. Variants of DENND1B associated with asthma in children. *N. Engl. J. Med.* 2010; **362**: 36–44.

58 Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, den Dekker HT, Husby A, Sevelsted A, Faura-Tellez G *et al*. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat. Genet.* 2014; **46**: 51–5.

59 Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, Lyon HN, Ferreira MA, Strachan DP, Zhao JH, Abramson MJ *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**: e44008.

60 Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, Barr RG, Beaty TH, Becker AB, Beilby J *et al*. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* 2018; **50**: 42–53.

61 Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H *et al*. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat. Genet.* 2009; **41**: 342–7.

62 Ferreira MA, Matheson MC, Tang CS, Granell R, Ang W, Hui J, Kiefer AK, Duffy DL, Baltic S, Danoy P *et al*. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. *J. Allergy Clin. Immunol.* 2014; **133**: 1564–71.

63 Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson W. A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* 2010; **363**: 1211–21.

64 Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, Fujita K, Miyatake A, Enomoto T, Miyagawa T *et al*. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* 2011; **43**: 893–6.

65 Himes BE, Hunninghake GM, Baurley JW, Rafaels NM, Sleiman P, Strachan DP, Wilk JB, Willis-Owen SA, Klanderman B, Lasky-Su J *et al*. Genome-wide association analysis identifies PDE4D as an asthma-susceptibility gene. *Am. J. Hum. Genet.* 2009; **84**: 581–93.

66 Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, Kurosaka F, Akasawa A, Yoshihara S, Kanno N *et al*. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. *PLoS Genet.* 2011; **7**: e1002170.

67 Almoguera B, Vazquez L, Mentch F, Connolly J, Pacheco JA, Sundaresan AS, Peissig PL, Linneman JG, McCarty CA, Crosslin D *et al*. Identification of four novel loci in asthma in European American and African American populations. *Am. J. Respir. Crit. Care Med.* 2017; **195**: 456–63.

68 Wan YI, Shrine NR, Soler Artigas M, Wain LV, Blakey JD, Moffatt MF, Bush A, Chung KF, Cookson WO, Strachan DP *et al*. Genome-wide association study to identify genetic determinants of severe asthma. *Thorax* 2012; **67**: 762–8.

69 Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014; **42**: D1001–6.

70 Vicente CT, Revez JA, Ferreira MAR. Lessons from ten years of genome-wide association studies of asthma. *Clin. Transl. Immunology* 2017; **6**: e165.

71 Makinde TO, Agrawal DK. Increased expression of angiopoietins and Tie2 in the lungs of chronic asthmatic mice. *Am. J. Respir. Cell Mol. Biol.* 2011; **44**: 384–93.

72 Kanazawa H, Nomura S, Asai K. Roles of angiopoietin-1 and angiopoietin-2 on airway microvascular permeability in asthmatic patients. *Chest* 2007; **131**: 1035–41.

73 Obeidat M, Hao K, Bosse Y, Nickle DC, Nie Y, Postma DS, Laviolette M, Sandford AJ, Daley DD, Hogg JC *et al*. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. *Lancet Respir. Med.* 2015; **3**: 782–95.

74 Cheng DT, Kim DK, Cockayne DA, Belousov A, Bitter H, Cho MH, Duvoix A, Edwards LD, Lomas DA, Miller BE *et al*. Systemic soluble receptor for advanced glycation endproducts is a biomarker of emphysema and associated with AGER genetic variants in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2013; **188**: 948–57.

75 Yin NC, Lang XP, Wang XD, Liu W. AGER genetic polymorphisms increase risks of breast and lung cancers. *Genet. Mol. Res.* 2015; **14**: 17776–87.

76 Miller S, Henry AP, Hodge E, Kheirallah AK, Billington CK, Rimington TL, Bhaker SK, Me O, Melén E, Merid SK *et al*. The Ser82 RAGE variant affects lung function and serum RAGE in smokers and sRAGE production in vitro. *PLoS One* 2016; **11**: e0164041.

77 Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J *et al*. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190–5.

78 Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, Sliwinski P, Hersh CP, Mancini JD, Lu K *et al*. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum. Mol. Genet.* 2012; **21**: 1325–35.

79 Zhou X, Qiu W, Sathirapongsasuti JF, Cho MH, Mancini JD, Lao T, Thibault DM, Litonjua G, Bakke PS, Gulsvik A *et al*. Gene expression analysis uncovers novel Hedgehog interacting protein (HHIP) effects in human bronchial epithelial cells. *Genomics* 2013; **101**: 263–72.

80 Gibson G. Hints of hidden heritability in GWAS. *Nat. Genet.* 2010; **42**: 558–60.

81 Prokopenko D, Sakornsakolpat P, Loehlein Fier H, Qiao D, Parker MM, McDonald MN, Manichaikul A, Rich SS, Barr RG, Williams CJ *et al*. Whole genome sequencing in severe chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* 2018; **59**: 614–622.