Lessons from ten years of genome-wide association studies of asthma

Cristina T Vicente, Joana A Revez and Manuel AR Ferreira

Twenty-five genome-wide association studies (GWAS) of asthma were published between 2007 and 2016, the largest with a sample size of 157242 individuals. Across these studies, 39 genetic variants in low linkage disequilibrium (LD) with each other were reported to associate with disease risk at a significance threshold of $P<5 \times 10^{-8}$, including 31 in populations of European ancestry. Results from analyses of the UK Biobank data ($n=380\,503$) indicate that at least 28 of the 31 associations reported in Europeans represent true-positive findings, collectively explaining 2.5% of the variation in disease liability (median of 0.06% per variant). We identified 49 transcripts as likely target genes of the published asthma risk variants, mostly based on LD with expression quantitative trait loci (eQTL). Of these genes, 16 were previously implicated in disease pathophysiology by functional studies, including TSLP, TNFSF4, ADORA1, CHIT1 and USF1. In contrast, at present, there is limited or no functional evidence directly implicating the remaining 33 likely target genes in asthma pathophysiology. Some of these genes have a known function that is relevant to allergic disease, including F11R, CD247, PGAP3, AAGAB, CAMK4 and PEX14, and so could be prioritized for functional follow-up. We conclude by highlighting three areas of research that are essential to help translate GWAS findings into clinical research or practice, namely validation of target gene predictions, understanding target gene function and their role in disease pathophysiology and genomics-guided prioritization of targets for drug development.

INTRODUCTION

Asthma is a common and chronic inflammatory disease of the airways, specifically affecting the bronchi and bronchioli. Bronchial inflammation, which results in airway narrowing and shortness of breath symptoms, is generally caused by innate and adaptive immune responses to inhaled viruses and/or allergens. These and other environmental exposures are strong risk factors for disease onset and exacerbations but, based on twin studies, account for less than half of the overall disease liability. The remaining disease risk is largely explained by inherited genetic factors, with gene-by-environment interaction effects also thought to play a role. Given this high heritability, there has been a long-standing interest in identifying specific genetic risk factors for asthma, initially through linkage analysis (1989 to 2006) and subsequently through candidate-gene association studies (since 1995). Linkage studies were largely (if not entirely) unsuccessful because this approach is only adequately powered with realistic sample sizes to identify very large genetic effects, which we now know do not exist for asthma. On the other hand, most published candidate-gene studies suffered from a number of methodological limitations (for example, small number of samples and genetic markers tested), and so the reported candidate gene associations have been largely discounted.

In 2004, it became feasible to genotype hundreds of thousands of genetic variants in a single experiment. The development of genotyping arrays enabled the design of genome-wide association studies (GWAS), the first published soon after in 2005. In the years that followed, not only the cost of genotyping arrays decreased substantially, but also methods for the analysis of GWAS data were developed and refined; notably, these included statistical approaches to account for population structure and to infer individual genotypes for genetic variants not present in the genotyping arrays. As a result, GWAS including data from thousands of individuals genotyped for millions of genetic variants became a reality, at last providing a powerful tool to identify genetic associations with disease risk. For asthma, the first GWAS was published in 2007 by Moffatt et al. Since then, and until the end of 2016, 24 additional GWAS of asthma were published. In this review, we summarize and interpret the key genetic findings from these studies, specifically addressing the following questions: are any published risk variants likely to be false-positive associations? How much variation in disease liability do they explain? What are the likely target genes of those risk variants? Do those genes point to potential new mechanisms underlying disease pathophysiology? Lastly, we conclude by highlighting areas of research that are essential to help translate genetic findings into clinical research or practice.

SUMMARY OF GENETIC ASSOCIATIONS REPORTED IN ASTHMA GWAS PERFORMED BETWEEN 2007 AND 2016

We searched the NHGRI-EBI catalog of published GWAS to identify studies that tested the association between genetic variants and asthma...
risk, between 2007 and 2016. We used the search term ‘Asthma’ and applied no filters. The search was performed on the 2nd of August 2017, and returned 73 unique studies, which were individually reviewed for inclusion in our analysis. Of these, 48 (66%) were excluded (Supplementary Table 1), most (34 studies) because the phenotype tested in the GWAS was not asthma, but instead an asthma-related trait (for example, lung function). Studies were also commonly excluded because they were based on DNA pooling (five studies) or reported genetic interactions (for example, gene-by-environment; four studies) rather than main effects. We extracted data from the GWAS Catalog for the remaining 25 studies, which were included for analysis (Supplementary Table 2).

The definition of asthma was not always the same across the 25 published GWAS. For example, six GWAS ascertained asthma cases with disease onset in childhood, whereas in other GWAS cases had more severe symptoms or co-morbid allergies (Supplementary Table 2). The smallest GWAS included 66 cases and 42 controls, and the largest 28,399 cases and 128,843 controls. For most studies (18 of 25), the primary GWAS included exclusively individuals of European descent; two studies were based on populations of Asian ancestry, two of Latino ancestry, and one included multiple ancestries. Across these 25 studies, 73 unique genetic variants were reported to associate with disease risk at a genome-wide significance threshold of $P < 5 \times 10^{-8}$ (listed per study in Supplementary Table 3). Some single nucleotide polymorphisms (SNPs) were located in close proximity, and so we used the clump procedure in PLINK to determine which were likely to represent independent associations and which were simply correlated with previously published variants. Specifically, we assigned each SNP into groups of correlated variants based on pairwise linkage disequilibrium (LD). LD was estimated using data from the 1000 Genomes Project (release 20130502_v5a), separately for individuals of European, Asian and African ancestry, as appropriate. Using a conservative LD threshold of $r^2 > 0.05$, there were 31 groups of correlated risk variants reported in GWAS of European ancestry (Table 1 and Supplementary Table 4), with seven, one and three variants reported in GWAS of Asian, African and Latino ancestry, respectively (Supplementary Table 5). One variant reported in Asians (rs1837253) and two (rs9272346 and rs907092) in Latinos were in strong LD ($r^2 > 0.8$) with risk variants reported in Europeans, and so do not represent independent associations. Therefore, in total, 39 genetic variants (31 in Europeans and 8 additional in other ancestries) in low LD with each other were reported to associate with asthma risk in GWAS published between 2007 and 2016. In the sections below, we focus on the 31 associations reported in Europeans.

ARE ANY OF THE PUBLISHED RISK VARIANTS FOR ASTHMA LIKELY TO REPRESENT FALSE-POSITIVE ASSOCIATIONS?

GWAS, even when conducted using strict quality control procedures and an appropriate genome-wide significance threshold, are not completely protected from false-positive associations. That is, there is always a small chance that a new genome-wide significant association with asthma risk might not be a true-positive association and instead arise by chance or because of unaccounted methodological biases. As such, whenever possible, it is important to perform an independent and adequately powered replication study to confirm novel associations. This is not always feasible, particularly as GWAS become larger, because there might not be sufficiently large studies available for replication that were not included in the discovery stage.

In the last few years, ~500,000 individuals from the UK have been deeply phenotyped and genotyped as part of the UK Biobank study. The genotype data for the full dataset has just been made publicly available, providing a unique and timely opportunity to test if the 31 associations with asthma risk reported in Europeans between 2007 and 2016 are reproducible. Briefly, we analyzed data for 380,503 unrelated individuals (kinship coefficient indicating <3rd degree relatedness) with (a) European ancestry, confirmed based on analysis of allele sharing with individuals from the 1000 Genomes Project; and (b) non-missing information for field 6152 of the touchscreen questionnaire: ‘Has a doctor ever told you that you have had any of the following conditions?’ A total of 44,003 individuals selected ‘Asthma’ when answering that question, and so were considered as cases. On the other hand, 336,500 individuals did not select ‘Asthma’ and so were considered as controls. Mean age was 56.7 (range 38–72), with 54% of participants being female. We tested the association between individual SNPs and case-control status using SNPTEST, including age, sex and SNP chip as covariates in the model. We adjusted the association results for an LD Score intercept of 1.073, estimated using 1.2 million HapMap3 SNPs. In this analysis, we were able to test all 31 reported SNPs (all with imputation information >0.98), either directly (30 SNPs) or through a proxy SNP (rs1660709 instead of rs200634877, $r^2 = 0.75$). Of these, 28 (90%) had a statistically significant ($P < 0.05/31$ SNPs = 0.0016) and directionally consistent (same predisposing allele as originally reported) association with disease risk (Table 2 and Figure 1), thereby confirming the original findings as true-positive associations.

For the remaining three SNPs, results from this UK Biobank analysis of self-reported doctor-diagnosed asthma did not support an association with disease risk, and so it is possible that they represent false-positive associations. These associations are located in/near DENND1B/CRB1, PDE4D37 and CDHR3.20 Another explanation for the lack of association with these three SNPs is that the case-control definition we used in the UK Biobank analysis is not a good proxy for that used in the original studies. For example, the original association with rs6967330 in the CDHR3 gene, which was subsequently supported by results from Pickrell et al., was found when studying asthma cases with childhood onset and severe exacerbations. If such an association is specific to that subgroup of asthmatics, and if these only represent a small fraction of the UK Biobank asthma cases, then the power to replicate the original association might have been low. In this respect, it is noteworthy that the direction of effect in the UK Biobank for rs6967330 was the same as originally reported. In conclusion, at least 28 of the 31 SNPs previously reported in GWAS of European ancestry have a significant and consistent association with self-reported doctor-diagnosed asthma in the UK Biobank study and so represent bona fide asthma risk variants. It was beyond the scope of this review to report associations found in the UK Biobank study that were not located in previously reported asthma risk loci. Studies that report the full results from the UK Biobank study will be reported elsewhere in the near future.

HOW MUCH VARIATION IN DISEASE LIABILITY IS EXPLAINED BY THE PUBLISHED ASTHMA RISK VARIANTS AND BY OTHERS YET TO BE DISCOVERED?

Twin studies have estimated the heritability of asthma to be between 55 and 74% in adults, with even larger estimates reported in young children.4,44 The aim of GWAS is to identify variants that contribute to this heritability. So it is important to understand to what extent the heritability of asthma is explained by the asthma risk variants discovered to date, and how much heritability remains to be discovered. To answer the first question, we estimated the total
Lessons from ten years of genome-wide association

CT Vicente et al

Clinical & Translational Immunology

### Table 1 Variants in low LD with each other ($r^2<0.05$) reported to associate with asthma risk in GWAS conducted between 2007 and 2016 in populations of European ancestry

| Index | Chr | Bp | Context | Top SNP | Effect allele | OR  | P-value | PMID | Year |
|-------|-----|----|---------|---------|-------------|-----|---------|------|------|
| 1 | 1 | 10557251 | [PEX14] | rs662064 | T | 0.94 | 3.2E-08 | 27182965 | 2016 |
| 2 | 1 | 154426264 | [IL6R] | rs4129267 | T | 1.09 | 2.4E-08 | 21907864 | 2011 |
| 3 | 1 | 161159147 | PPOX-[]-ADAMTS4 | rs423366 | T | 1.09 | 4.8E-15 | 27182965 | 2016 |
| 4 | 1 | 167434320 | [CD247] | rs1723018 | G | 0.95 | 1.4E-08 | 27182965 | 2016 |
| 5 | 1 | 173152036 | TNFSF18-[]-TNFSF4 | rs6691738 | T | 0.94 | 2.9E-08 | 27182965 | 2016 |
| 6 | 1 | 197325908 | [CRB1] | rs2786098 | A | 0.63 | 8.6E-09 | 20032319 | 2009 |
| 7 | 1 | 203100504 | [ADORA1] | rs6683383 | T | 1.06 | 1.1E-08 | 27182965 | 2016 |
| 8 | 2 | 8458800 | [LINC00299] | rs1341275 | G | 1.06 | 1.3E-08 | 27182965 | 2016 |
| 9 | 2 | 102986222 | [IL18R1] | rs3771166 | A | 0.87 | 3.4E-09 | 20860503 | 2010 |
| 10 | 2 | 242698640 | [D2HGDH] | rs34290285 | G | 1.11 | 1.8E-15 | 27182965 | 2016 |
| 11 | 3 | 188402471 | [LPP] | rs7319673 | T | 0.92 | 6.5E-09 | 27182965 | 2016 |
| 12 | 4 | 38799710 | [TLR1] | rs4833095 | T | 1.20 | 5.0E-12 | 24388013 | 2013 |
| 13 | 5 | 59369794 | [PDE4D] | rs1588265 | G | 0.60 | 2.5E-08 | 19426955 | 2009 |
| 14 | 5 | 38799710 | [TLR1] | rs4833095 | T | 1.20 | 5.0E-12 | 24388013 | 2013 |
| 15 | 5 | 131969874 | SLC25A46 | rs110401872 | S | 1.09 | 4.8E-15 | 27182965 | 2016 |
| 16 | 5 | 141529762 | [NDFIP1] | rs200634877 | I | 0.94 | 2.5E-08 | 27182965 | 2016 |
| 17 | 6 | 90985198 | [BACH2] | rs58521088 | T | 0.93 | 7.1E-11 | 27182965 | 2016 |
| 18 | 6 | 32728261 | HLA-DQA1-[]-HLA-DQB1 | rs1837253 | C | 1.14 | 2.4E-09 | 21804549 | 2011 |
| 19 | 6 | 199085198 | [BACH2] | rs58521088 | T | 0.93 | 7.1E-11 | 27182965 | 2016 |
| 20 | 7 | 10568451 | [CDHR3] | rs6967330 | A | 1.45 | 1.4E-08 | 24241537 | 2013 |
| 21 | 8 | 81291879 | MIR5708-[]-ZBTB10 | rs7009110 | T | 1.14 | 4.0E-09 | 24388013 | 2013 |
| 22 | 9 | 6190076 | RANBP6-[]-IL33 | rs1342326 | C | 1.20 | 9.2E-10 | 20860503 | 2010 |
| 23 | 10 | 9049253 | GATA3-[]-SFTA1P | rs12413578 | T | 0.89 | 8.1E-12 | 27182965 | 2016 |
| 24 | 11 | 76270683 | WNT11-[]-LRRC32 | rs7130588 | G | 1.09 | 1.8E-08 | 27182965 | 2016 |
| 25 | 12 | 57590505 | [STAF6] | rs3001426 | T | 0.94 | 1.4E-10 | 27182965 | 2016 |
| 26 | 14 | 68749927 | [RAD51B] | rs3784099 | G | 0.94 | 1.6E-08 | 27182965 | 2016 |
| 27 | 15 | 61099988 | [RORA] | rs11071559 | T | 0.85 | 3.8E-09 | 27182965 | 2016 |
| 28 | 15 | 67446785 | [SMAD3] | rs744910 | A | 0.89 | 3.9E-09 | 20860503 | 2010 |
| 29 | 16 | 11228712 | [CLEC16A] | rs62026376 | C | 1.17 | 1.0E-08 | 24388013 | 2013 |
| 30 | 17 | 38069949 | [GDMB] | rs7216389 | T | 1.45 | 9.0E-11 | 17611496 | 2007 |
| 31 | 21 | 37530430 | [IL12B] | rs2284033 | A | 0.89 | 1.2E-08 | 20860503 | 2010 |

Abbreviations: OR, odds ratio; PMID, PubMed identifier.

*If the top SNP is located within the boundaries of a gene, then the gene name is shown inside square brackets. Otherwise, the two nearest genes (upstream and downstream) are listed, with the distance to each represented by the number of ‘-’ between the square bracket and the gene name.

Correlated SNPs reported in other asthma GWAS

| Correlated SNP | Effect allele | PMID | Year |
|---------------|--------------|------|------|
| 1 | 2.5% | of disease liability is explained by the 31 asthma risk variants discovered to date.

The second question of interest is how much heritability is likely to be explained by risk variants that remain to be discovered. One approach to address this question might be to simply subtract the SNP heritability explained by the 31 published associations (2.5%) from the overall asthma heritability estimates reported in twin studies (e.g. 55%). Therefore, potentially, asthma risk variants yet to be discovered could account for at least ~52% (55%–2.5%) of disease liability. However, heritability estimates from twin studies can be inflated, for example, because of violations of study design assumptions. Thus, such estimates should be taken as the upper boundary of the total variation in disease liability that is explained by genetic variants.

A more conservative approach to address the same question involves first estimating the disease heritability that is explained collectively by all genetic variants studied in a GWAS, not just those with a strong association with disease risk. This is referred to as the SNP-based disease heritability, which can be estimated using, for example GCTA, BOLT-REML, or LD Score regression. In theory, the SNP-based heritability can be lower than the twin-based heritability if genetic variants not tested (or not well tagged) in GWAS contribute to disease risk, which could be the case for uncommon variants (e.g. with minor allele frequency [MAF] <1%). Differences in heritability estimates could also arise if the twin-based heritability estimate is inflated, as discussed above. When we applied the LD Score regression approach to genome-wide results from the UK Biobank GWAS analysis described above (n = 380 503), we found that the overall SNP-based heritability for asthma was 14%, with a standard error (SE) of 1%. This estimate was obtained based on results for 1.2
Lessons from ten years of genome-wide association
CT Vicente et al

Table 2 Association between 31 variants reported in GWAS of European ancestry and self-reported doctor-diagnosed asthma in the UK Biobank study (44 003 cases and 336500 controls)

| Index | SNP          | Chr | Bp     | Context      | Effect allele | MAF | OR   | P-value     | Asthma h² explained (%) |
|-------|--------------|-----|--------|--------------|---------------|-----|------|-------------|-------------------------|
| 1     | rs662064     | 1   | 10557251 | [PEX14]      | C             | 0.31| 1.04 | 9.7E-006    | 0.02                    |
| 2     | rs4129267    | 4   | 154426264 | [IL6R]       | T             | 0.41| 1.03 | 5.8E-006    | 0.02                    |
| 3     | rs4233366    | 1   | 161159149 | [PPOX-ADAMTS4] | T             | 0.27| 1.04 | 4.7E-007    | 0.02                    |
| 4     | rs1723018    | 1   | 167433420 | [CD247]      | G             | 0.41| 0.95 | 5.7E-012    | 0.04                    |
| 5     | rs6691738    | 1   | 173152036 | [TNFSF18-TNFSF4] | G     | 0.29| 1.04 | 5.8E-007    | 0.02                    |
| 6     | rs2786098    | 1   | 197325908 | [CRB1]       | G             | 0.22| 1.00 | 5.82E-007  | 0.02                    |
| 7     | rs6683383    | 1   | 203100504 | [ADORA1]     | A             | 0.33| 0.95 | 5.9E-011    | 0.04                    |
| 8     | rs13412757   | 2   | 8458080  | [LINC00299]  | A             | 0.34| 0.94 | 1.8E-013    | 0.05                    |
| 9     | rs3771166    | 2   | 109286222 | [IIL8R1]     | A             | 0.38| 0.90 | 7.4E-040    | 0.15                    |
| 10    | rs34290285   | 2   | 242698640 | [D2HGDI]     | A             | 0.26| 0.89 | 1.5E-039    | 0.15                    |
| 11    | rs73196739   | 3   | 188402471 | [LPP]        | T             | 0.17| 0.94 | 4.1E-010    | 0.03                    |
| 12    | rs4833095    | 4   | 38799710 | [TLR1]       | C             | 0.21| 1.02 | 2.85E-009   | 0.00                    |
| 13    | rs1837253    | 5   | 110401872 | [PEDE4D]     | G             | 0.31| 1.01 | 2.85E-009   | 0.00                    |
| 14    | rs166079     | 5   | 141528959 | [NDFIP1]     | T             | 0.38| 1.04 | 3.4E-041    | 0.16                    |
| 15    | rs6871536    | 5   | 131969874 | [RAD50]      | C             | 0.19| 1.10 | 1.0E-024    | 0.09                    |
| 16    | rs1588265    | 5   | 59369794 | [PDE4D]      | G             | 0.31| 1.01 | 2.85E-009   | 0.00                    |
| 17    | rs3428494    | 6   | 31322197 | [HLA-B]      | G             | 0.47| 1.12 | 6.1E-055    | 0.20                    |
| 18    | rs17844604   | 6   | 32620283 | [HLA-DQA1-HLA-DQB1] | T     | 0.42| 1.18 | 1.9E-105   | 0.41                    |
| 19    | rs58521088   | 6   | 90985198 | [BACH2]      | T             | 0.35| 1.03 | 2.5E-021    | 0.08                    |
| 20    | rs9697330    | 7   | 10564851 | [CDH3]       | A             | 0.17| 1.01 | 2.23E-003  | 0.00                    |
| 21    | rs7009110    | 8   | 81291879 | [MIR5708]    | C             | 0.38| 0.94 | 9.6E-018    | 0.06                    |
| 22    | rs1342326    | 9   | 6190076 | [RAI8P3-LIL3] | C             | 0.16| 1.15 | 6.4E-048    | 0.17                    |
| 23    | rs12413578   | 10  | 9049253 | [GATA3-SFTA1P] | T             | 0.11| 0.86 | 1.0E-033    | 0.13                    |
| 24    | rs7130588    | 11  | 76270683 | [WNT11-LRRC32] | G             | 0.36| 1.10 | 8.9E-033    | 0.12                    |
| 25    | rs3001426    | 12  | 57500905 | [STAT6]      | C             | 0.45| 1.07 | 1.4E-021    | 0.08                    |
| 26    | rs3784099    | 14  | 68749927 | [RAD51B]     | A             | 0.28| 1.06 | 1.4E-021    | 0.08                    |
| 27    | rs11071559   | 15  | 61096986 | [RORA]       | T             | 0.13| 0.91 | 7.3E-017    | 0.06                    |
| 28    | rs744910     | 15  | 67446785 | [SMAO3]      | A             | 0.48| 0.95 | 3.5E-014    | 0.05                    |
| 29    | rs62026376   | 16  | 11228712 | [CLEC16A]    | T             | 0.25| 0.92 | 7.7E-023    | 0.08                    |
| 30    | rs7216389    | 17  | 38069949 | [GSMB]       | T             | 0.48| 1.11 | 1.7E-044    | 0.16                    |
| 31    | rs2284033    | 22  | 37534034 | [IL2RB]      | A             | 0.43| 0.97 | 4.2E-004    | 0.01                    |

Abbreviations: MAF, minor allele frequency; OR, odds ratio; h², heritability.
Three SNPs with a P<0.05 are highlighted in gray.

 HAVE WE FOUND FEWER ASTHMA RISK VARIANTS THAN EXPECTED BASED ON GWAS SAMPLE SIZE?

The largest asthma GWAS published between 2007 and 2016 included 157 242 individuals and identified 27 independent associations with disease risk. This figure is smaller than reported by GWAS of some complex diseases using similar or smaller sample sizes (Table 3). For example, with a similar sample size (n = 150 064), Ripke et al.49 found 128 independent associations with schizophrenia (SCZ), which is 4.7-fold greater than those found for asthma. What underlies this difference in GWAS yield? First, the power to detect an association with a SNP depends on the proportion of variance in disease liability it explains. As discussed above, this can be calculated from the risk allele frequency and the odds ratio. For example, power is about the same to detect an association with two SNPs, one with a risk allele of frequency 0.5 and odds ratio of 2.5, and the other with a risk allele of frequency 0.01 and odds ratio of 2.5—both SNPs explain about 0.5% of variance in disease liability (see formula in section above). When comparing two diseases, one needs to consider that the power to detect an association with a SNP that explains the same proportion of variance in disease liability depends on the disease prevalence.50,51 Using the formula derived by Yang et al.,51 if we consider a SNP that explains 0.05% of the variation in disease liability, then the expected non-centrality parameter (NCP; which reflects power, but is linearly related to sample size) for the SCZ GWAS listed in Table 3 is 102 (assuming a prevalence of 1%). This is 2.6-fold greater than obtained for the similar-sized asthma GWAS (NCP = 40, assuming a prevalence of 15%). In other words, although the overall sample size is the same, the SCZ GWAS has substantially greater power to detect an association with such a SNP when compared to the asthma GWAS.

The genetic architecture (number of risk loci, their frequency and effect size), which is unknown and may differ across
Table 3 Number of associations reported in published GWAS of other polygenic diseases, which were based on a similar or smaller sample size than the largest asthma GWAS published

| Disease              | N cases | N controls | N Total | N of independent associations in discovery GWAS (P<5x10⁻⁸) | PMID     | Disease prevalence | NCP for a SNP with h² = 0.05% |
|----------------------|---------|------------|---------|----------------------------------------------------------|----------|--------------------|-----------------------------|
| Atopic dermatitis    | 18900   | 84166      | 103066  | 21                                                       | 26482879 | 20%                | 24                          |
| Asthma               | 28399   | 128843     | 157242  | 27                                                       | 27182965 | 15%                | 40                          |
| Type 2 diabetes      | 26676   | 132532     | 159208  | 42                                                       | 28566273 | 5%                 | 52                          |
| Schizophrenia        | 36989   | 113075     | 150064  | 128                                                      | 25056061 | 1%                 | 102                         |
| Rheumatoid arthritis | 29880   | 73758      | 103638  | 101                                                      | 24390342 | 1%                 | 77                          |

Abbreviations: PMID: PubMed identifier; NCP: non-centrality parameter (which reflects power); h2: heritability.
All studies are based on the analysis of 1000 Genome Project SNPs imputed from GWAS arrays (not including fine-mapping arrays).

Figure 1 Effect size (odds ratio) for 31 previously reported asthma risk SNPs, comparing results reported in the original GWAS describing each association with those obtained in the analysis of the UK Biobank study. Each SNP is represented by a circle, with its size and color indicating total GWAS sample size (as per Supplementary Table 2) and PubMed identifier (PMID), respectively. The red diagonal line represents equality of odds ratio between published GWAS and UK Biobank (that is, x=y). The genomic location of the four SNPs that deviate markedly from the diagonal are shown next to the respective circle. For three of these (SNPs in PDE4D, CRBL1 and CDHR3), the association in the UK Biobank study had a P-value >0.05 (see Table 2).

Different diseases.52 For example, asthma might have a smaller number of common risk variants than SCZ, or SNP effects might be smaller. So to adequately compare the number of associations reported for different diseases (and quantitative traits), one needs to consider not just the sample size but also disease lifetime risk and the likely genetic architecture, as highlighted previously.50,53

Another factor that influences power, and that could have a more severe effect for some diseases than others, is disease heterogeneity and misclassification. If the genetic architecture of asthma is not homogeneous, and instead has components that are specific to clinically distinct subtypes, then lumping all individuals who reported ever having asthma in the same case group could lead to underestimation of SNP heritabilities, which decreases power. As examples of this, the risk allele for susceptibility variants near ORMDL3 is significantly more common in cases with childhood onset asthma6,54 while the association near CDHR3 might be specific to children with early onset and severe exacerbations20 as discussed above. Similarly, including in the control group individuals who do not have asthma but suffer from other common allergic diseases (for example, hay fever) can significantly decrease power to detect associations with SNPs that affect allergies in general, not just asthma.55

Shared genetic effects are expected to be widespread amongst common allergic diseases, with pairwise genetic correlations estimated to be >50% in twin studies.2,4,56 One approach to capitalize on this high genetic correlation is to define cases as those suffering from two or more allergic diseases, and controls as those who do not suffer from any allergic disease.55 We have shown empirically that the estimated SNP effects are indeed larger with such approach, and so power is increased.19

Another approach that also increases power to detect shared genetic effects is to define cases as those who suffer from any allergic disease.55 We have recently performed a large GWAS using this approach (n = 360 838) and identified 136 independent associations for allergic disease.57 Of note, the expected NCP in this study design for a SNP with a 0.05% heritability and assuming a 30% disease prevalence was 123, which is comparable to the SCZ GWAS that identified 128 independent associations.

Lastly, if gene-by-environment interactions have a substantial contribution to asthma risk, then ignoring the relevant environmental exposures in GWAS could also result in underestimated SNP effects. For example, gene-by-environment interactions have been reported for variants in the ORMDL3 locus, with larger SNP effects found in children with rhinovirus wheezing illness in early life63 or exposed to tobacco smoke in early life.59 The latter interaction, however, was not replicated in a large independent study.50 A small number of genome-wide interaction analyses between environmental exposures and asthma risk have been published,61-64 but are not discussed in this review. For reviews of this topic, see for example.65,66 Thus, in summary, fewer genetic associations have been reported for asthma as compared to some complex diseases with GWAS of similar or smaller size. This likely reflects lower statistical power arising from the relatively high disease prevalence, but also study design limitations, such as not accounting for disease subtypes, information from genetically-correlated diseases and gene-by-environment interaction effects.

WHAT ARE THE LIKELY TARGET GENES OF THE PUBLISHED ASTHMA RISK VARIANTS?

The aim of asthma GWAS per se is to identify genetic variants associated with disease risk. A genetic variant is associated with asthma risk most likely because that same variant, or another in strong LD with it (for example, r²>0.8), affects the protein sequence or the
transcription patterns of a gene (that is, the ‘target gene’) that plays a role in disease pathophysiology. Therefore, knowing which variants are associated with disease risk might highlight specific genes and molecular pathways dysregulated in asthma, and ultimately help better understand why asthma develops in the first place. How do we identify the likely target genes of risk variants? First, we can determine if the associated variant, or another in strong LD with it, is a non-synonymous coding variant, using for example ANNOVAR. If we include the additional correlated risk variants reported in other asthma GWAS (Supplementary Table 4), then this approach identifies eight associations with asthma in the UK Biobank study (Table 2), and also focus on the 28 published risk variants that had a reproducible

### Table 4 Likely target genes of published asthma risk variants identified based on LD with sentinel eQTLs

| Index | GWAS SNP | Results for sentinel eQTL in strongest LD with GWAS SNP | eQTL reported in other studies? |
|-------|----------|----------------------------------------------------------|---------------------------------|
|       | eQTL     | LD (r²) Target gene | P-value | Study | Tissue |
| 1     | rs662064 | rs668805 1.00 | PEPI4 | 1.2E-019 | Zeller | Monocytes | Yes |
| 2     | rs662064 | rs12028449 0.81 | DFFA | 1.2E-012 | Westra | Whole blood | No |
| 3     | rs419267 | rs4537545 0.93 | IL6R | 2.0E-029 | Westra | Whole blood | Yes |
| 4     | rs423366 | rs423366 1.00 | FCER1G | 5.0E-215 | Jansen | Whole blood | Yes |
| 5     | rs423366 | rs423366 1.00 | BAGALT3 | 1.5E-026 | GTEx | Fibroblasts | No |
| 6     | rs423366 | rs423366 1.00 | ADAMTS4 | 1.8E-024 | GTEx | Fibroblasts | No |
| 7     | rs423366 | rs423366 1.00 | PPOX | 6.2E-012 | GTEx | Fibroblasts | No |
| 8     | rs423366 | rs423366 1.00 | FHL1 | 3.4E-011 | Fehrman | Whole blood | No |
| 9     | rs423366 | rs423366 1.00 | USF1 | 3.4E-011 | Fehrman | Whole blood | Yes |
| 10    | rs423366 | rs423366 0.96 | TOMM40L | 7.3E-010 | Naranbhai | Neutrophils | No |
| 11    | rs172018 | rs2988279 0.93 | CD247 | 3.1E-062 | Zhernakova | Whole blood | Yes |
| 12    | rs6691738 | rs7553711 0.99 | TNFSF4 | 1.9E-035 | Yao | Whole blood | No |
| 13    | rs6683383 | rs6683383 1.00 | ADORA1 | 6.0E-096 | Zeller | Monocytes | Yes |
| 14    | rs6683383 | rs3766568 1.00 | CHIT1 | 6.6E-030 | Zhernakova | Whole blood | Yes |
| 15    | rs6683383 | rs10920570 1.00 | MYBPH | 5.2E-018 | Zeller | Monocytes | Yes |
| 16    | rs6683383 | rs11655198 0.99 | RPI1-35013.7 | 8.1E-021 | Zhernakova | Whole blood | Yes |
| 17    | rs6683383 | rs7555556 0.98 | PPIA4 | 8.1E-162 | Zhernakova | Whole blood | Yes |
| 18    | rs10173081 | rs3771180 1.00 | MFS99 | 1.7E-015 | Yao | Whole blood | No |
| 19    | rs10173081 | rs11674302 0.80 | IL18RAP | 7.9E-091 | Yao | Whole blood | No |
| 20    | rs10173081 | rs10189629 0.80 | IL1R1 | 3.0E-012 | Zhernakova | Whole blood | Yes |
| 21    | rs3771166 | rs11688559 1.00 | ACO07278.3 | 4.3E-249 | Zhernakova | Whole blood | No |
| 22    | rs4833090 | rs12233670 0.98 | TLR1 | 2.8E-057 | Battle | Whole blood | Yes |
| 23    | rs1438673 | rs7273819 0.86 | WDR36 | 2.4E-031 | Yao | Whole blood | No |
| 24    | rs1438673 | rs10073416 0.85 | TSLP | 3.3E-012 | Zhernakova | Whole blood | Yes |
| 25    | rs1438673 | rs2289277 0.84 | CTG-551A13.2 | 7.0E-029 | Zhernakova | Whole blood | Yes |
| 26    | rs1438673 | rs1001830 0.82 | CAMK4 | 1.8E-016 | Yao | Whole blood | No |
| 27    | rs1438673 | rs2246176 0.99 | SL222A5 | 8.7E-014 | Westra | Whole blood | No |
| 28    | rs166079 | rs12655645 1.00 | NDFIP1 | 6.8E-115 | Zhernakova | Whole blood | Yes |
| 29    | rs2428494 | rs2428494 1.00 | MCB | 2.3E-011 | Walsh | Whole blood | No |
| 30    | rs9268516 | rs9268400 0.99 | HLA-DRB6 | 6.2E-011 | Dinarzo | Whole blood | No |
| 31    | rs9272346 | rs9272346 1.00 | HLA-DQB1 | <4.9E-324 | Zeller | Monocytes | Yes |
| 32    | rs9272346 | rs9272346 1.00 | HLA-DQB5 | 2.1E-121 | Westra | Whole blood | Yes |
| 33    | rs9272346 | rs9272346 1.00 | TAP2 | 4.1E-011 | Westra | Whole blood | No |
| 34    | rs9273373 | rs1063355 0.99 | HLA-DOA1 | 3.6E-154 | Raj | Monocytes | Yes |
| 35    | rs9273373 | rs3134993 0.95 | HLA-DQB1-AS1 | 5.6E-058 | GTEx | Lung | No |
| 36    | rs9273373 | rs1063349 0.92 | HLA-DOAB2 | 1.1E-089 | Geuvadis | LCLs | No |
| 37    | rs9273373 | rs9272545 0.87 | HLA-DOA2 | 1.1E-022 | Quach | Monocytes | Yes |
| 38    | rs10151068 | rs11633029 0.86 | RP11-554D20.1 | 3.7E-017 | Zhernakova | Whole blood | Yes |
| 39    | rs56375023 | rs17293632 0.98 | AGAB | 1.7E-013 | Zhernakova | Whole blood | Yes |
| 40    | rs11078927 | rs12946510 0.80 | GSDMA | <2.2E-016 | Hao | Lung | No |
| 41    | rs1165490 | rs9003250 1.00 | RP1L-941L15.2 | 5.2E-064 | Zhernakova | Whole blood | Yes |
| 42    | rs1165519 | rs1165519 1.00 | ORMELD3 | 7.4E-041 | Kasa | CDB T-cells | Yes |
| 43    | rs1165519 | rs2305479 0.97 | IKZF3 | 5.6E-021 | Zhernakova | Whole blood | Yes |
| 44    | rs1165519 | rs8067378 0.96 | ZPB2P2 | 2.4E-017 | Grundberg | LCLs | Yes |
| 45    | rs2271308 | rs1053651 0.99 | STARD3 | 1.4E-016 | Fairfax | Monocytes | Yes |
| 46    | rs2271308 | rs1053651 0.99 | PGAP3 | 9.9E-014 | Yao | Whole blood | No |
| 47    | rs2271308 | rs4795388 0.83 | PPP1R1B | 4.1E-010 | Andiappan | Neutrophils | No |
| 48    | rs4794820 | rs4794820 1.00 | GSDMB | 1.1E-296 | Zhernakova | Whole blood | Yes |
likely target genes: GSDMA, GSDMB, HLA-DQA1, HLA-DQB1, IL1RL1, IL6R, TL1R and ZPB2 (Supplementary Table 6).

Table 5 Number of publications co-mentioning gene/alias terms and allergy-related terms

| Gene | Prior to 2007 | Since 2007 |
|------|--------------|------------|
| TSLP | 21           | 814        |
| IL1RL1 | 16           | 260        |
| TNFSF4 | 9             | 50         |
| TL1R | 8            | 43         |
| HLA-DQB1 | 30          | 31         |
| HLA-DQB2 | 28          | 30         |
| HLA-DQA1 | 13          | 13         |
| ADORA1 | 12           | 5          |
| TAP2 | 9            | 2          |
| ORMDL3 | 0            | 115        |
| GSDMB | 0            | 57         |
| ZPB2 | 0            | 18         |
| IL6R | 2            | 15         |
| GSDMA | 0            | 15         |
| CHI1 | 2            | 14         |
| IKZF3 | 0            | 11         |
| FCER1G | 1            | 9          |
| SLC22A5 | 0            | 7          |
| WDR3 | 0            | 6          |
| IL18RAP | 1           | 5          |
| HLA-DQA2 | 0           | 5          |
| NDFIP1 | 0            | 5          |
| F11R | 2            | 3          |
| HLA-DRB5 | 1           | 2          |
| MCB | 1            | 2          |
| STARD3 | 1            | 2          |
| PGAP3 | 0            | 2          |
| PPP1R1B | 0           | 2          |
| AAGAB | 2            | 1          |
| USF1 | 2            | 1          |
| ADAMTS4 | 0           | 1          |
| B4GALT3 | 0           | 1          |
| DFFA | 0            | 1          |
| PAG1 | 0            | 1          |
| AC007278.3 | 0           | 0          |
| CAMK4 | 0            | 0          |
| CD247 | 0            | 0          |
| CTC-551A13.2 | 0     | 0          |
| HLA-DQB1-AS1 | 0     | 0          |
| HLA-DRB6 | 0            | 0          |
| MFSD9 | 0            | 0          |
| MYBPH | 0            | 0          |
| PEK14 | 0            | 0          |
| PPIA4 | 0            | 0          |
| PPOX | 0            | 0          |
| RP11-335013.7 | 0     | 0          |
| RP11-654020.1 | 0     | 0          |
| RP11-94L15.2 | 0     | 0          |
| TOMM40L | 0            | 0          |

A second approach that can be used to identify likely target genes is to determine if the associated SNPs are in strong LD with variants that have been reported to associate with variation in gene expression levels, known as expression quantitative trait loci (eQTL). A plethora of eQTL studies have been reported in recent years, including 39 conducted using tissues relevant to asthma pathophysiology (Supplementary Table 7), for example, whole-blood, lung, skin and individual immune cell types, such as CD4+ T cells. For each of these studies, we extracted eQTL results (that is, SNP, gene and P-value) from the original publication, keeping only associations in cis (that is, SNP within 1 Mb of gene) that were significant at a conservative threshold of $P<2.3 \times 10^{-9}$, which corresponds to a Bonferroni correction for testing each of 21472 genes for association with 1000 independent SNPs. In each study and for each gene, we then used the clump procedure in PLINK to reduce the published list of eQTLs (which typically includes many correlated variants) to the subset of strongest eQTLs that were in low LD with each other ($r^2<0.05$), which we refer to as sentinel eQTLs. Finally, we asked if any of the 28 variants with a reproducible association with asthma risk in the UK Biobank study (Table 2), or the additional correlated risk variants reported in other asthma GWAS (Supplementary Table 4), were in strong LD ($r^2>0.8$) with a sentinel eQTL. Using this approach, we found 48 likely target genes (Table 4 and Supplementary Table 8). Of note, these include all eight genes with non-synonymous variants listed above, indicating that for these variation in both protein sequence and transcript levels might be important determinants of cellular function and disease risk.

For many asthma risk variants (12 of 28, or 43%), the two approaches described above failed to identify any likely target gene. This could arise, for example, if the effect of those variants on gene expression is (a) specific to tissues, cell types and/or cellular conditions (e.g., hypoxia) for which eQTL information is not available at present; or (b) too weak to be detected with the sample sizes included in published eQTL studies and at the conservative significance level that we used. In this case, the likely target genes could potentially be identified using functional studies; these can include, for example, experiments to determine the effect of risk variants on promoter-enhancer chromatin interactions, promoter activity or transcription factor binding. We recently used such approaches to identify PAG1 as a likely target gene of the asthma risk variants located on chromosome 8q21, for which no eQTL support was available at the strict significance threshold used above. In summary, at least 49 genes (including PAG1) are likely targets of published asthma risk variants, most (90%) identified based on the LD between risk variants and eQTLs.

**DO ANY OF THE LIKELY TARGET GENES REPRESENT POTENTIAL NEW PLAYERS IN THE PATHOPHYSIOLOGY OF ASTHMA AND ALLERGIC DISEASE MORE GENERALLY?**

To address this question, we performed a PubMed query using the HGNC-approved gene symbols listed in Table 4, as well as all known aliases (Supplementary Table 9), and the allergy-related terms ‘asthma OR rhinitis OR eczema OR atopic OR dermatitis OR allergy OR allergi* OR hayfever OR ’hay fever’’. We downloaded results from the PubMed query in XML format and then counted the number of unique articles in that file citing both the allergy-related terms and the gene name or aliases in the title, abstract or keyword fields. On the basis of results from this query, we classified the 49 genes into three groups (Table 5). Group one consisted of nine genes co-mentioned frequently (5 or more studies) with those allergy-related terms prior to 2007, the year the first GWAS of any allergy-related trait was
published. The genes were TSLP, IL1RL1, TNFSF4, TLR1, HLA-DQB1, HLA-DQB2, HLA-DQA1, ADORA1 and TAP2. For these genes, GWAS findings did not provide the first clue for a key role in disease pathophysiology.

Group two consisted of 13 genes co-mentioned frequently with allergy-related terms since, but not before, 2007: ORMDL3, GSDMB, ZPPB2, IKZF3, GSDMA, IL6R, CHIT1, Fcer1g, SLC22A5, WDR36, IL18RAP, HLA-DQA2 and NDFIP1. The first five are located in the same asthma risk locus, and were first suspected to contribute to asthma pathophysiology because of GWAS findings. Of the remaining eight genes, for five it can be argued that functional studies provided the first suggestion of a key role in asthma/allergies, namely IL6R,21 CHIT1,72 Fcer1g,73 IL18RAP4 and NDFIP1. But that is unlikely to be the case for the other three genes, SLC22A5 (organic cation transporter involved in pulmonary absorption of asthma-related drugs67,77), WDR36 (nucleolar protein involved in processing of 18S rRNA75) and HLA-DQA2 (HLA class II molecule expressed in epidermal Langerhans cells79).

Lastly, group three was composed of 27 genes co-mentioned infrequently with allergy-related terms, both before and after 2007. To our knowledge, only two of these genes have been suggested to play a role in allergic disease through functional studies: USF180 and STARD83. However, many have a known function (Supplementary Table 10) that is directly relevant to allergic disease pathophysiology, such as F11r82,83, MICB84, CD24785,86, PGAP387, AAGAB,88 CAMK489 and PEX1490. In summary, of the 49 likely target genes of published asthma risk variants, only 16 were previously implicated by functional studies in disease pathophysiology.

ARE THERE EXAMPLES OF GENETIC FINDINGS SUBSEQUENTLY TRANSLATED INTO CLINICAL RESEARCH OR PRACTICE?

Based on the Thomson Reuters Cortellis™ Drug database, drugs against five of the 49 likely target genes of asthma risk variants are being considered for clinical development (Table 6). Six of these drugs are being (or have been) tested in clinical trials of asthma. To our knowledge, only one of these clinical studies was motivated directly by results from a GWAS, our clinical trial of tocilizumab (TCZ) in participants with mild to moderate asthma. In 2011, we reported the association between a variant in the IL6R gene (rs4129267) and asthma risk.26 A consistent association with this variant was later reported also for eczema91,92 and asthma severity.93 We and others noted that the disease protective allele (rs4129267:C) was strongly associated with decreased protein levels of the soluble form of the receptor (sIL-6R),94 but increased mRNA levels of the full length IL6R transcript,95–97 which encodes for the membrane bound form of the receptor (mIL-6R). That is, decreased asthma risk is associated with decreased sIL-6R but increased mIL-6R. By extension, decreased disease risk is likely associated with decreased IL-6 trans-signaling (which requires sIL-6R and is mainly pro-inflammatory) but increased IL-6 classic signaling (which requires mIL-6R and is thought to be mainly regenerative and protective).98 On the basis of these genetic findings, it was not immediately obvious what to expect from a drug such as TCZ (approved therapeutic for rheumatoid arthritis and other auto-immune diseases), which blocks both sIL-6R and mIL-6R. A small case study published in 2011 reported decreased clinical activity of atopic dermatitis in three patients treated with TCZ for up to 12 months.99 This was the first suggestion of a protective effect of TCZ in allergic conditions. To characterize the effect of TCZ in asthma, in 2013 we performed pre-clinical studies using mouse models of acute allergic asthma.100 In these studies, we found that TCZ had a protective effect on allergen-induced airway inflammation only when the experimental model used resulted in increased levels of sIL-6R in the airways, and so that was likely to involve activation of the IL-6 trans-signaling pathway. When that was not the case, dual receptor blockade resulted in worse airway inflammation when compared to control mice. On the basis of the genetic findings and the mouse studies, in 2014 we initiated a clinical trial of tocilizumab in participants with mild asthma, specifically those with CT or TT genotype for rs4129267, as these have markedly increased levels of sIL-6R levels.96 Results from this trial are expected to be published in 2018. On the other hand, if our prediction from the genetic studies and findings from the mouse studies are correct, then a drug that blocks sIL-6R but not mIL-6R might be more desirable. There is one such drug (sgpl130Fc, also known as FE301 or olamkicept), which was shown to be safe in phase 1 clinical trials101 and is now in phase 2 trials for Crohn's disease (Stefan Rose-John, personal communication). Future studies that test the safety and efficacy of this drug in asthma patients are warranted.

OPPORTUNITIES AND CHALLENGES

In this section, we highlight three broad research areas that should be addressed in the short-term to help translate findings from asthma GWAS into clinical research or practice.

Improving and confirming target gene predictions

As discussed above, eQTL information provides a valuable tool to identify a set of genes for which variation in SNP genotype for an asthma risk SNP is likely to directly cause variation in transcription levels. However, this approach has some caveats, of which we highlight two. First, SNP effects on gene expression can be tissue-specific102, context-dependent103 and/or relatively small. For these reasons, some genes might not be predicted to be target genes because appropriate (right tissue, context and sample size for that gene) eQTL studies have not been performed. For example, to our knowledge, the largest eQTL study conducted using airway epithelial cells was based on just 105 individuals104 (compared to 5311 for whole-blood105), and so our understanding of the genetic control of gene expression in this relevant cell type is still very poor. We thus consider essential to continue expanding eQTL studies to include more relevant tissues, diverse cellular stimuli and larger sample sizes. The second caveat of this approach is that a genetic overlap between risk variants and eQTLs can often arise by chance, because eQTLs are widespread.106 We therefore argue that eQTL information should be used to generate predictions that can be subsequently tested by functional studies. However, such studies are laborious and time consuming, often deployed one locus at a time (for example, refs. 76,77). As the number of known asthma risk variants increases in the near future, efficient validation of target gene predictions will require high-throughput approaches, such as capture Hi-C,109 multiplexed reporter assays110 and genome editing.111 Ideally, these studies should be performed in relevant primary cell types, which at present is still technically challenging.

Understanding target gene function and their role in disease pathophysiology

As highlighted above, for many genes it is unclear how variation in gene expression or protein sequence ultimately leads to variation in asthma risk. Therefore, for these, there is an opportunity to make new discoveries regarding the molecular mechanisms underlying the regulation of gene expression, and the impact that this might have on cellular function and disease pathophysiology. The best examples of this to date are ORMDL3112−118 and GSDMB,119 for which recent
| Gene   | Drug name            | Originator company          | Indications (approved or in trials)                                                                 | Target-based actions                          | Highest status      | Clinical trials in asthma |
|--------|----------------------|-----------------------------|---------------------------------------------------------------------------------------------------|----------------------------------------------|---------------------|---------------------------|
| ADORA1 | Adenoscan            | King Pharmaceuticals R&D Inc | Coronary artery disease                                                                         | Adenosine A1 (and A2) receptor agonist       | Launched            | NA                        |
| ADORA1 | Trabodenoson         | Inotek Pharmaceuticals Corp  | Ocular hypertension; Open angle glaucoma; Optic nerve disorder                                  | Adenosine A1 receptor agonist                | Phase 3 Clinical    | NA                        |
| ADORA1 | Neladenoson bialanate| Bayer AG                    | Cardiac failure                                                                                  | Adenosine A1 receptor partial agonist         | Phase 2 Clinical    | NA                        |
| ADORA1 | PBF-680              | Palobiofarma SL              | Asthma; COPD                                                                                     | Adenosine A1 receptor antagonist              | Phase 2 Clinical    | NCT02635945               |
| IKZF3  | Iberdomide           | Celgene Corp                | Multiple myeloma; Systemic lupus erythematosus                                                   | Aiolos inhibitor                             | Phase 2 Clinical    | NA                        |
| IL1RL1 | RG-6149              | Amgen Inc                   | Asthma                                                                                           | IL-33 receptor antagonist                    | Phase 2 Clinical    | NCT01928368               |
| IL1RL1 | Nerofe               | Immune System Key Ltd        | Autoimmune disease; Cancer; Diabetes mellitus; Myocardial infarction                             | IL-33 receptor agonist                       | Phase 1 Clinical    | NA                        |
| IL1RL1 | GSK-3772847          | Janssen Research & Development LLC | Asthma                                                                                           | IL-33 receptor antagonist                    | Phase 2 Clinical    | NCT03207243               |
| IL6R   | Tocilizumab          | Chugai Pharmaceutical Co Ltd | Asthma; Chronic lymphocytic leukemia; Motor neuron disease; Rheumatoid arthritis; Sclerodema; and others. | IL-6 receptor antagonist                      | Launched            | ACTRN1261400123640       |
| IL6R   | Sarilumab            | Regeneron Pharmaceuticals Inc | Arthritis; Juvenile rheumatoid arthritis; Rheumatoid arthritis; Uveitis                         | IL-6 receptor antagonist                      | Launched            | NA                        |
| IL6R   | Siltuximab           | Janssen Biotech Inc          | Castleman's disease; Multiple myeloma                                                             | IL-6 antagonist                               | Launched            | NA                        |
| IL6R   | Sirukumab            | Janssen Biotech Inc          | Asthma; Major depressive disorder; Polymyalgia rheumatica; Rheumatoid arthritis; Temporal arteritis | IL-6 antagonist                               | Launched            | NCT02794519               |
| IL6R   | SA-237               | Chugai Pharmaceutical Co Ltd | Neuromyelitis optica                                                                            | IL-6 antagonist                               | Phase 3 Clinical    | NA                        |
| IL6R   | Olamikcept           | Conaris Research Institute AG | Inflammatory bowel disease                                                                       | IL-6/sIL-6 inhibitor                         | Phase 2 Clinical    | NA                        |
| IL6R   | Vobamilizumab        | Abyln NV                    | Rheumatoid arthritis; Systemic lupus erythematosus                                               | IL-6 receptor antagonist                      | Phase 2 Clinical    | NA                        |
| TSLP   | Tezepelumab          | Amgen Inc                   | Asthma; Atopic dermatitis                                                                         | TSLP antagonist                               | Phase 2 Clinical    | NCT01405963, NCT02054130 |
functional and animal studies have provided the first insights into their contribution to disease pathophysiology. Similarly, we have recently been interested in understanding how variation in PAG1 expression might contribute to asthma pathophysiology. Our functional studies suggested that decreased PAG1 expression is associated with decreased disease risk, contrary to what was expected based on the widely accepted contribution of this transmembrane adaptor protein to the development of immune responses. We are currently testing this possibility using mouse models of experimental asthma. Various models have been described, including those that mimic mechanisms underlying acute allergic asthma; chronic asthma; and viral-induced asthma in humans. It is not always straightforward to decide which model is most appropriate to use, and different models can result in different findings, presumably because of the different underlying pathophysiology. In this respect, understanding the contribution of gene expression to cellular function might provide the best clue as to which models to begin with. An additional caveat of these models might be encountered when comparing germine knock-out mice against wild-type mice, because of possible genetic compensation mechanisms in the former, that is, the up-regulation of functionally related genes. Potential solutions include temporally controlled gene deletion or using therapeutic agents that block gene activity. Mouse models are also laborious and so are invariably applied on a gene-by-gene basis, which constitutes a significant bottleneck when dealing with the rapidly increasing number of asthma risk genes. Nonetheless, despite the potential limitations of these models, they will continue to represent essential tools to understand the contribution of target genes of risk variants to asthma pathophysiology.

**Genomics-guided prioritization of targets for drug development**

GWAS findings, in combination with eQTL information, can provide important clues into the predicted directional effect of gene expression on disease risk. This, in turn, can be used to inform drug repositioning or the development of novel drugs: specifically, if the disease protective allele of an asthma risk variant is associated with decreased (increased) gene expression, then the prediction is that a drug that also decreases (increases) gene expression should similarly have a protective effect on asthma symptoms. For example, the disease-protective allele rs1438673:T19 on chromosome 5q22.1 is strongly associated with increased expression of gene deletion or using therapeutic agents that block gene activity. Mouse models are also laborious and so are invariably applied on a gene-by-gene basis, which constitutes a significant bottleneck when dealing with the rapidly increasing number of asthma risk genes. Nonetheless, despite the potential limitations of these models, they will continue to represent essential tools to understand the contribution of target genes of risk variants to asthma pathophysiology.

GWAS findings, in combination with eQTL information, can provide important clues into the predicted directional effect of gene expression on disease risk. This, in turn, can be used to inform drug repositioning or the development of novel drugs: specifically, if the disease protective allele of an asthma risk variant is associated with decreased (increased) gene expression, then the prediction is that a drug that also decreases (increases) gene expression should similarly have a protective effect on asthma symptoms. For example, the disease-protective allele rs1438673:T19 on chromosome 5q22.1 is strongly associated with increased expression of gene deletion or using therapeutic agents that block gene activity. Mouse models are also laborious and so are invariably applied on a gene-by-gene basis, which constitutes a significant bottleneck when dealing with the rapidly increasing number of asthma risk genes. Nonetheless, despite the potential limitations of these models, they will continue to represent essential tools to understand the contribution of target genes of risk variants to asthma pathophysiology.

This is not always so straightforward, as the IL6R example discussed above illustrates. Other examples that are harder to interpret occur when the directional effect of a risk variant on gene expression is not the same, or is poorly characterized, in different relevant tissues. For example, the asthma protective allele rs6683383:A is strongly associated with increased expression of ADORA1 in whole-blood (for example, beta = 0.40, P = 2 × 10\(^{-15}\)) and whole-blood (beta = 0.76, P = 10\(^{-18}\)). The same direction of effect on gene expression is observed in at least eight different tissues studied by the GTEx project, including adipose, colon and esophagus. Thus, based on these genetic findings, we can postulate that TSLP antagonists (but not agonists) might improve asthma symptoms. This is consistent with the known contribution of TSLP to disease pathophysiology and with results from a recent clinical trial of an anti-TSLP antibody.

This could translate into many hundreds or thousands of risk variants, if we consider that the mean SNP heritability is likely to be low (for example, 0.01-0.05%). Larger GWAS will also have increased power to detect individual associations with uncommon variants (for example, MAF<1%), and to quantify their overall contribution to disease liability, which has not been addressed to date. Other areas of research that remain largely unexplored in the field of asthma genetics include understanding if different inflammatory subtypes (for example, neutrophilic and eosinophilic asthma) have distinct genetic components and whether disease remission is a heritable trait. The next decade of research might provide new insights into these important questions.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

This research has been conducted using the UK Biobank Resource. CTV was supported by a PhD scholarship (QPhD2015025) from Equity Trustees (Australia); IAR was supported by a PhD scholarship (SRFH/BD/92907/2013) from Fundacao para a Ciencia e Tecnologia (Portugal); MAF was supported by a Senior Research Fellowship (APP1124501) from the National Health and Medical Research Council (Australia). We also thank Naomi Wray for providing comments on the original manuscript.

1. Murray CS, Simpson A, Custovic A. Allergens, viruses, and asthma exacerbations. *Proc Am Thorac Soc* 2004; 1: 99–104.
2. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genes of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 1990; 142(6 Pt 1): 1351–1358.
3. Thorsen SF,Ulrik CS,Kyvik KO,Feerreira MA,Backer V. Multivariate genetic analysis of atopy phenotypes in a selected sample of twins. *Clin Exp Allergy* 2006; 36: 1382–1390.
4. van Beijsterveldt CE,Boomsma DI.Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* 2007; 29: 516–521.
5. Cookson WO,Sharp PAFaux JAHopkin JM.Linkage between immunoglobulin E responses underlying asthma and rhinitis and chromosome 11q. *Lancet* 1989; 1: 1292–1295.
6. Bouzigon E,Forabosco P,Koppelman GH,Cookson WD,Dizer MH,Duffy DL et al. Meta-analysis of 20 genome-wide linkage studies evidenced new regions linked to asthma and atopy. *Eur J Hum Genet* 2010; 18: 700–6.
7. Ohe M,Munakata M,Hizawa N,Itoh A,Doi I,Yamaguchi E et al. Beta 2 adrenergic receptor gene restriction fragment length polymorphism and bronchial asthma. *Thorax* 1999; 50: 353–359.
Lessons from ten years of genome-wide association

CT Vicente et al.

11

36 Hancock DB, Romieu I, Shi M, Siens-Monge JJW, Wu H, Chiu CY et al. Genome-wide association study implicates chromosome 9q21.31 as a susceptibility locus for asthma in Mexican children. PLoS Genet 2009; 5: e1000623.

37 Himes BE, Hunninghake GM, Baurett JW, Rafaels NM, Steliman P, Strachan DP et al. Genome-wide association study reveals PSMD4 as an asthma-susceptibility gene. Am J Hum Genet 2009; 84: 581–593.

38 Moffatt MF, Gul IG, Demenais F, Strachan DP, Bouzignon E, Heath S et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010; 363: 1211–1221.

39 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Eur J Hum Genet 2007; 15: 599–575.

40 Falconsole J, Manning AK, Florez JC. Genome-wide association studies by imputation of genotypes. Annu Rev Genomics Hum Genet 2015; 16: 1–35.

41 Allen NE, Sudlow C, Peakman T, Collins R, Biobank UK. The UK biobank data: coming soon. BioRxiv 2017.

42 Vinkhuyzen AA, Wray NR, Yang J, Ferreira MA. Improving the power to detect risk variants for allergic disease by deconfounding from polygenicity in genome-wide association studies. Nat Genet 2015; 47: 291–295.

43 Li X, Howard TD, Zheng SL, Haselkorn T, Peters SP, Meyers DA. HLA-DQA1 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol 2014; 133: 1564–1571.

44 Bonnelykke K, Steinmetz J, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet 2014; 46: 51–55.

45 Lasky-Su J, Himes BE, Raby BA, Klaendler B, Bjerregaard P, Lange C et al. HLA-DQ associations in asthma: further confirrmation and get it. Am J Hum Genet 2011; 88: 76–82.

46 Loh PR, Bhatia G, Gusev A, Finucane HK, Pollack SJ et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 2014; 47: 1385–1392.

47 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from the SAGE II Study. Sci Transl Med 2016; 8: 347ra1-48.

48 Loh PR, Bhatia G, Gusev A, Finucane HK, Pollack SJ et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat Genet 2014; 47: 1385–1392.

49 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from the SAGE II Study. Sci Transl Med 2016; 8: 347ra1-48.

50 Yang J, Lee SH, Goddard ME, Visscher PM. Estimation and partition of heritability in human populations using whole-genome analysis methods. Ann Rev Genet 2015; 49: 75–95.

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

52 Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet 2016; 99: 139–153.

53 Levinson DF, Mostafavi S, Milaneschi Y, Riveria M, Ripke S, Wray NR et al. Genetic studies of major depressive disorder: new results, meta-analysis, and lessons learned. Nat Genet 2016; 48: 36–48.

54 Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol 2010; 34: 254–257.

55 Thi H, Kichaev G, Pasaniuc B. Contrasting the genetic complexity of 30 traits from summary association data. Am J Hum Genet 2016; 99: 139–153.

56 Levinson DF, Mostafavi S, Milaneschi Y, Riveria M, Ripke S, Wray NR et al. Genetic studies of major depressive disorder: new results, meta-analysis, and lessons learned. Nat Genet 2016; 48: 36–48.

57 Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol 2010; 34: 254–257.

58 Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet 2016; 99: 139–153.

59 Levinson DF, Mostafavi S, Milaneschi Y, Riveria M, Ripke S, Wray NR et al. Genetic studies of major depressive disorder: new results, meta-analysis, and lessons learned. Nat Genet 2016; 48: 36–48.

60 Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol 2010; 34: 254–257.

61 Thi H, Kichaev G, Pasaniuc B. Contrasting the genetic complexity of 30 traits from summary association data. Am J Hum Genet 2016; 99: 139–153.

62 Levinson DF, Mostafavi S, Milaneschi Y, Riveria M, Ripke S, Wray NR et al. Genetic studies of major depressive disorder: new results, meta-analysis, and lessons learned. Nat Genet 2016; 48: 36–48.

63 Ege MJ, Strachan DP, Cookson WO, Moffatt MF, Gut I, Lathrop M. Lessons from ten years of genome-wide association studies. Clin Transl Immunol 2015; 7: e1172716.
Bonnylaieke K, Ober C. Leveraging gene-environment interactions and allergies to asthma gene discovery. J Allergy Clin Immunol 2016; 137: 667–679.

Kauffmann F, Demenais F. Gene-environment interactions in asthma and allergic diseases: challenges and perspectives. J Allergy Clin Immunol 2012; 130: 1229–1240, 411–412.

Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 2015; 10: 1556–1566.

Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. Trends Genet 1997; 13: 163–170.

Davis JR, Fresald L, Knowles DA, Pala M, Bustamante CD, Battle A et al. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. Am J Hum Genet 2016; 98: 216–224.

Vicario DT, Edwards SL, Ramlakhan MS, Sauvageau M, Mitchell H, Bain L et al. Long Range Modulation of PAG1 Expression by 8q21 Allergy Risk Variants. Genet Immunol J 2015; 11: 313–325.

Seibold MA, Donnelly S, Solon M, Innes A, Woodruff PG, Boot RG et al. Chitotriosidase is the primary active chitinase in the human lung and is modulated by genotype and smoking habit. J Allergy Clin Immunol 2012; 129: 944–950 e3.

Yu M, Tsai M, Tam SY, Jones C, Zehnder J, Galli SJ. Mast cells can promote the development of multiple features of chronic asthma in mice. J Clin Invest 2006; 113: 1633–1644.

Lewkowich IP, Rempel JD, HayGlass KT. Prevention of allergen-specific IgE-mediated reactions using a chimeric IgE receptor expressed by human mast cells. J Allergy Clin Immunol 2008; 121: 2474–2485.

Lemmens BA, Bongers J, van der Ploeg L, van Linde van Oosterhout SM, Pinkert L et al. A novel regulatory (yeast)-like protein isoform 3 (ORMDL3) expression dysregulates ceramide homeostasis in asthma gene discovery. PLoS Genet 2015; 11: e1005245.

Losier C, Youtube PN, Cibulskis K, Deaver AG, Oblinger AM et al. Lessons from ten years of genome-wide association studies. J Allergy Clin Immunol 2015; 136: 1389–1412.

Chu M, Song Y, Wu L, Song H, Zhang Y et al. Integrating eQTLs with clinical and functional annotation for better understanding of asthma susceptibility variants. J Allergy Clin Immunol 2017; 140: 925–933.

Luo W, Obeidat M, Di Narzo AF, Chen R, Sin DD, Pare PD et al. Airway epithelial expression quantitative trait loci reveal genes underlying asthma and other airway diseases. Am J Respir Cell Mol Biol 2016; 54: 177–187.

Westa HJ, Peters MJ, Ekelund J, Schmeider RE, Gieger C et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. J Allergy Clin Immunol 2015; 136: 1238–1243.

Chen X, Asa S, Otsu T, Kato T, Takeuchi M et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat Genet 2017; 49: 600–605.

Berliet S, Mousselet S, Quinnet M, Verlaan DJ, Koka V, Ali Taiwani A et al. Interaction between genetic and epigenetic variation defines gene expression patterns at the asthma-associated locus 17q12-q21 in lymphoblastoid cell lines. Hum Immunol 2012; 73: 1161–1171.

Verlaan DJ, Berliet S, Hunninghake GM, Madore AM, Lariviére M, Mousselet S et al. Allele-specific chromatin remodelling in the ZFP292/GMDS/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet 2009; 85: 377–393.

Javierre BM, Burren OS, Wilder SP, Kreuzhuber H, Hill SM, Sewitz S et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to human diseases. Nature 2015; 533: 379–384.

Novakova DV, Beeken P, Veraart M, van Ierson M, van Galen M, Arinantu W et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 2017; 49: 139–145.

Luo W, Obeidat M, Di Narzo AF, Sin DD, Pare PD et al. Airway epithelial expression quantitative trait loci reveal genes underlying asthma and other airway diseases. Am J Respir Cell Mol Biol 2016; 54: 177–187.

Airway epithelial expression quantitative trait loci reveal genes underlying asthma and other airway diseases. Am J Respir Cell Mol Biol 2016; 54: 177–187.
118 Miller M, Tarn AB, Cho JY, Doherty TA, Pham A, Khorsam N et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. Proc Natl Acad Sci USA 2012; 109: 16648–16653.

119 Das S, Miller M, Beppu AK, Mueller J, McGough MD, Vuong C et al. GSDMB induces an asthma phenotype characterized by increased airway responsiveness and remodeling without lung inflammation. Proc Natl Acad Sci USA 2016; 113: 13132–13137.

120 Lloyd CM. Building better mouse models of asthma. Curr Allergy Asthma Rep 2007; 7: 231–236.

121 Barry J, Loh Z, Collison A, Mazzone S, Lalwani A, Zhang V et al. Absence of Toll-IL-1 receptor B/single immunoglobulin IL-1 receptor-related molecule reduces house dust mite-induced allergic airway inflammation in mice. Am J Respir Cell Mol Biol 2013; 49: 481–490.

122 Dullaers M, Schuijs MJ, Willart M, Fierens K, Van Moorleghem J, Hammad H et al. House dust mite-driven asthma and allergen-specific T cells depend on B cells when the amount of inhaled allergen is limiting. J Allergy Clin Immunol 2017; 140: 76–88 e7.

123 McMillan SJ, Lloyd CM. Prolonged allergen challenge in mice leads to persistent airway remodelling. Clin Exp Allergy 2004; 34: 497–507.

124 Anikkatt J, Ullah MA, Short KR, Zhang V, Gan WJ, Loh Z et al. RAGE deficiency predisposes mice to virus-induced paucigranulocytic asthma. Elife 2017; 6: e21199.

125 El-Brolosy MA, Stainier DYR. Genetic compensation: a phenomenon in search of mechanisms. PLoS Genet 2017; 13: e1006780.

126 Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J et al. The genetic architecture of gene expression in peripheral blood. Am J Hum Genet 2017; 100: 228–237.

127 Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348: 10804–10820.

128 Watson B, Gauvreau GM. Thymic stromal lymphopoietin: a central regulator of allergic asthma. Expert Opin Ther Targets 2014; 18: 771–785.

129 Gauvreau GM, O'Byrne PM, Boulet LP, Wang Y, Cockcroft D, Bigler J et al. Effects of an anti-TSLP antibody on allergen-induced asthmatic responses. N Engl J Med 2014; 370: 2102–2110.

130 Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J et al. The genetic architecture of gene expression in peripheral blood. Am J Hum Genet 2017; 100: 228–237.

131 Wells A, Kopp N, Xu X, O'Brien DR, Yang W, Nehorai A et al. The anatomical distribution of genetic associations. Nucleic Acids Res 2015; 43: 10804–10820.

132 Finucane HK, Bulik-Sullivan B, Gusev A, Tyinka G, Reshef Y, Loh PR et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 2015; 47: 1228–1235.

133 Fakh A, Marson A, Zhu J, Kleineiewiedfeld M, Housley WJ, Beik S et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 2015; 518: 337–343.