BASIC SCIENCE FOR THE CHEST PHYSICIAN

Genome-wide association studies in lung disease

María Soler Artigas, Louise V Wain, Martin D Tobin

1. WHAT IS A GENOME-WIDE ASSOCIATION STUDY?

A genetic association study assesses the relation between genetic variation and a disease or a disease-related phenotype. Until about 2006, such studies usually included a handful of common sequence variants (single nucleotide polymorphisms, SNPs) in potentially relevant genes or regions. Although these candidate gene studies provided some robust associations, the choice of genes studied was limited by existing biological knowledge. Furthermore, the limited size and power of most candidate gene studies, coupled with liberal criteria for statistical significance, selective reporting and severe publication bias, led to a body of literature which is unreliable and difficult to interpret.

Improved cataloguing of SNPs in the human genome and an understanding of the inter-relationship between the SNPs through the HapMap project, together with improvements in genetic assays, made it possible to measure 300,000 or more SNPs spaced throughout the genome in each participant in case-control or cohort studies. These genome-wide association studies (GWAS) tend to focus on common SNPs (allele frequency >5%). As closely co-located common genetic variants tend to be inherited together over many generations, the genotype at one SNP can be strongly predictive of a nearby SNP (referred to as ‘linkage disequilibrium’, LD). Therefore, not all common variants need to be measured to have a fairly complete picture of common genetic variation across the genome. We can also exploit this concept to infer from GWAS data the genotypes at unmeasured SNPs—referred to as imputation—which is a valuable approach when we want to combine information across different studies, each of which has measured a different set of SNPs.

GWAS enable a hypothesis-free approach to studying association across the genome and are therefore much less prone to the reporting and publication biases which have plagued candidate gene studies. Crucially, they also have the potential to identify genes involved in biological pathways not previously connected to the disease under study.

2. INTERPRETATION OF A GENOME-WIDE ASSOCIATION STUDY

A GWAS typically tests association at about 2.5 million common genotyped and imputed SNPs in a first (discovery) stage and takes forward SNPs reaching a certain statistical threshold to a follow-up stage comprising independent samples. As an extremely large number of hypotheses are tested, genome-wide significance is typically defined as $p < 5 \times 10^{-8}$ after meta-analysis of findings across discovery and follow-up stages. At any SNP there are three possible genotypes representing 0, 1 or 2 copies of a particular allele. For example, a SNP which results in substitution of adenine (A) by cytosine (C) will result in three possible genotypes—AA, AC and CC corresponding to 0, 1 or 2 copies of the C allele. Association tests usually employ an additive genetic model which expresses effect size per copy of a particular allele. If dominant and recessive genetic models are also tested, then additional statistical corrections are needed to account for the greater number of hypotheses tested.

GWAS undertaken within a cohort, often to study quantitative traits, tend to be less susceptible to severe biases than case-control studies. Where cases and controls have been genotyped using different genotyping platforms, or have been genotyped in different laboratories or at different times, differential biases often result. Stricter quality control can actually make such biases worse, for example, by calling many homozygotes at a given SNP as missing in cases only. Approaches to filter rogue signals include comparing differences in missingness of genotypes between cases and controls and examining cluster plots of hybridisation intensities for each allele at a SNP to reveal possible genotyping errors. Evidence from further genotyping (validation) or from independent populations (follow-up and replication) can also help to distinguish between real and artefactual association signals.

A real association signal will tend to reflect the local characteristics of the genome, so if there are many nearby SNPs in LD, then the association signal may be manifest across all of these SNPs (figure 1A,B). While some association signals are localised, others include many genes and this may make it difficult to identify which gene is most likely to harbour one or more causal variants.

Confounding is a major concern in observational epidemiology studies, but the situation is somewhat different in genetic association studies. As
genotypes are assigned at gamete formation by an essentially random process, lifestyle factors do not tend to confound genetic associations. Genetic association studies therefore need not adjust for covariates in a primary analysis. How does this affect our interpretation of the role of smoking, for example, which has a very large effect on respiratory phenotypes? A SNP which is causally associated with tobacco addiction will in turn increase susceptibility to chronic obstructive pulmonary disease (COPD), but in this case smoking is an intermediate on the causal pathway rather than a confounder. Therefore, it is preferable to adjust for smoking in a secondary analysis in order to assess whether smoking is indeed an intermediate on the causal pathway. In contrast, differences in ancestry can confound genetic associations. These ancestral differences are checked and accounted for using principal components approaches and genomic control.

Two key approaches have been employed to improve the power of GWAS to discover modest genetic effects. The first has been to recruit severe cases and/or refine the phenotype definition. A second approach has been to attain very large sample sizes, sometimes with some relaxation of phenotype definition, in order to discover variants that can be followed up subsequently in deeply phenotyped studies. Most of the discoveries to date relate to the latter approach. Meta-analysis of genome-wide association studies of asthma showed or confirmed association with \textit{IL1RL1/IL18R1}, \textit{HLA-DQ}, \textit{IL33}, \textit{SMAD3} and \textit{IL2RB} and reported variants specific to childhood-onset asthma in \textit{ORMDL3/GSDMB}.

GWAS for COPD have reported associations near \textit{HHIP} and in the region of \textit{FAM13A} and the \textit{z}-nicotinic acetylcholine receptors \textit{CHRNA3} and \textit{CHRNA5}.

Studies of quantitative lung function may improve understanding of susceptibility to, and severity of, COPD. To date, GWAS in over 20 000 individuals have reported associations with forced expiratory volume in 1 s (FEV$_1$) or FEV$_1$/forced vital capacity (FVC) in or near \textit{TNS1}, \textit{FAM13A}, \textit{GSTCD/NPNT}, \textit{HHIP}, \textit{HTR4}, \textit{ADAM19}, \textit{AGER}, \textit{GPR126}, \textit{PTCH1}, \textit{TSHD4} and \textit{PIDD5}.

A subset of these has already been shown to be associated with COPD and further reports of associations are likely as case–control studies of COPD attain larger sample sizes.

**UTILITY OF GENOME-WIDE ASSOCIATION FINDINGS**

GWAS have yielded novel findings for a range of common diseases and related traits. The majority of associated variants have a modest phenotypic effect (e.g. RR of disease 1.1–1.2 or 1/10–1/20 SD for a quantitative trait). The major utility of such findings is therefore in what they tell us about the biological pathways involved in development and progression of disease. Indeed, GWAS have highlighted pathways that have not previously been implicated in respiratory disease. Such knowledge can inform the choice of molecular targets for the development of new drugs and, in some instances, will inform pharmacogenetics.

**LIMITATIONS OF GWAS**

Unless very large and accompanied by large-scale follow-up studies, GWAS have little prospect of discovering novel associations that reach the stringent statistical thresholds of significance used to avoid false positive findings. All GWAS also require careful design and analysis to avoid the biases described above.

**GWAS FINDINGS AND RESPIRATORY DISEASE**

The study of genome-wide association in over 10 000 cases of asthma showed or confirmed association with \textit{IL1RL1/IL18R1}, \textit{HLA-DQ}, \textit{IL33}, \textit{SMAD3} and \textit{IL2RB} and reported variants specific to childhood-onset asthma in \textit{ORMDL3/GSDMB}.

GWAS for COPD have reported associations near \textit{HHIP} and in the region of \textit{FAM13A} and the \textit{z}-nicotinic acetylcholine receptors \textit{CHRNA3} and \textit{CHRNA5}.

Studies of quantitative lung function may improve understanding of susceptibility to, and severity of, COPD. To date, GWAS in over 20 000 individuals have reported associations with forced expiratory volume in 1 s (FEV$_1$) or FEV$_1$/forced vital capacity (FVC) in or near \textit{TNS1}, \textit{FAM13A}, \textit{GSTCD/NPNT}, \textit{HHIP}, \textit{HTR4}, \textit{ADAM19}, \textit{AGER}, \textit{GPR126}, \textit{PTCH1}, \textit{TSHD4} and \textit{PIDD5}.

A subset of these has already been shown to be associated with COPD and further reports of associations are likely as case–control studies of COPD attain larger sample sizes.

**UTILITY OF GENOME-WIDE ASSOCIATION FINDINGS**

GWAS have yielded novel findings for a range of common diseases and related traits. The majority of associated variants have a modest phenotypic effect (e.g. RR of disease 1.1–1.2 or 1/10–1/20 SD for a quantitative trait). The major utility of such findings is therefore in what they tell us about the biological pathways involved in development and progression of disease. Indeed, GWAS have highlighted pathways that have not previously been implicated in respiratory disease. Such knowledge can inform the choice of molecular targets for the development of new drugs and, in some instances, will inform pharmacogenetics.

**LIMITATIONS OF GWAS**

Unless very large and accompanied by large-scale follow-up studies, GWAS have little prospect of discovering novel associations that reach the stringent statistical thresholds of significance used to avoid false positive findings. All GWAS also require careful design and analysis to avoid the biases described above.

Some robust findings from GWAS are referred to below. These explain a small proportion of the variance of the trait attributable to common sequence variation in the genome and leave unexplained variation that might be attributable to rare sequence variation or structural variation.

**GWAS FINDINGS AND RESPIRATORY DISEASE**

The study of genome-wide association in over 10 000 cases of asthma showed or confirmed association with \textit{IL1RL1/IL18R1}, \textit{HLA-DQ}, \textit{IL33}, \textit{SMAD3} and \textit{IL2RB} and reported variants specific to childhood-onset asthma in \textit{ORMDL3/GSDMB}.

GWAS for COPD have reported associations near \textit{HHIP} and in the region of \textit{FAM13A} and the \textit{z}-nicotinic acetylcholine receptors \textit{CHRNA3} and \textit{CHRNA5}.

Studies of quantitative lung function may improve understanding of susceptibility to, and severity of, COPD. To date, GWAS in over 20 000 individuals have reported associations with forced expiratory volume in 1 s (FEV$_1$) or FEV$_1$/forced vital capacity (FVC) in or near \textit{TNS1}, \textit{FAM13A}, \textit{GSTCD/NPNT}, \textit{HHIP}, \textit{HTR4}, \textit{ADAM19}, \textit{AGER}, \textit{GPR126}, \textit{PTCH1}, \textit{TSHD4} and \textit{PIDD5}.

A subset of these has already been shown to be associated with COPD and further reports of associations are likely as case–control studies of COPD attain larger sample sizes.

**UTILITY OF GENOME-WIDE ASSOCIATION FINDINGS**

GWAS have yielded novel findings for a range of common diseases and related traits. The majority of associated variants have a modest phenotypic effect (e.g. RR of disease 1.1–1.2 or 1/10–1/20 SD for a quantitative trait). The major utility of such findings is therefore in what they tell us about the biological pathways involved in development and progression of disease. Indeed, GWAS have highlighted pathways that have not previously been implicated in respiratory disease. Such knowledge can inform the choice of molecular targets for the development of new drugs and, in some instances, will inform pharmacogenetics.

**LIMITATIONS OF GWAS**

Unless very large and accompanied by large-scale follow-up studies, GWAS have little prospect of discovering novel associations that reach the stringent statistical thresholds of significance used to avoid false positive findings. All GWAS also require careful design and analysis to avoid the biases described above.

Some robust findings from GWAS are referred to below. These explain a small proportion of the variance of the trait attributable to common sequence variation in the genome and leave unexplained variation that might be attributable to rare sequence variation or structural variation.

**GWAS FINDINGS AND RESPIRATORY DISEASE**

The study of genome-wide association in over 10 000 cases of asthma showed or confirmed association with \textit{IL1RL1/IL18R1}, \textit{HLA-DQ}, \textit{IL33}, \textit{SMAD3} and \textit{IL2RB} and reported variants specific to childhood-onset asthma in \textit{ORMDL3/GSDMB}.

GWAS for COPD have reported associations near \textit{HHIP} and in the region of \textit{FAM13A} and the \textit{z}-nicotinic acetylcholine receptors \textit{CHRNA3} and \textit{CHRNA5}.

Studies of quantitative lung function may improve understanding of susceptibility to, and severity of, COPD. To date, GWAS in over 20 000 individuals have reported associations with forced expiratory volume in 1 s (FEV$_1$) or FEV$_1$/forced vital capacity (FVC) in or near \textit{TNS1}, \textit{FAM13A}, \textit{GSTCD/NPNT}, \textit{HHIP}, \textit{HTR4}, \textit{ADAM19}, \textit{AGER}, \textit{GPR126}, \textit{PTCH1}, \textit{TSHD4} and \textit{PIDD5}.

A subset of these has already been shown to be associated with COPD and further reports of associations are likely as case–control studies of COPD attain larger sample sizes.

**UTILITY OF GENOME-WIDE ASSOCIATION FINDINGS**

GWAS have yielded novel findings for a range of common diseases and related traits. The majority of associated variants have a modest phenotypic effect (e.g. RR of disease 1.1–1.2 or 1/10–1/20 SD for a quantitative trait). The major utility of such findings is therefore in what they tell us about the biological pathways involved in development and progression of disease. Indeed, GWAS have highlighted pathways that have not previously been implicated in respiratory disease. Such knowledge can inform the choice of molecular targets for the development of new drugs and, in some instances, will inform pharmacogenetics.

**LIMITATIONS OF GWAS**

Unless very large and accompanied by large-scale follow-up studies, GWAS have little prospect of discovering novel associations that reach the stringent statistical thresholds of significance used to avoid false positive findings. All GWAS also require careful design and analysis to avoid the biases described above.
A 67-year-old woman with fever, multiple lung opacities, visual impairment and acute respiratory failure

**CASE PRESENTATION**

A 67-year-old woman presented to the emergency room in acute respiratory failure. She had a 4-month history of dry cough, fever, intense fatigue, progressive dyspnoea and recent visual impairment. Her prior medical history included a mitral valve replacement in 2005 due to post-endocarditis mitral regurgitation. On examination, the patient was in hypoxic acute distress. She was intubated and mechanically ventilated. A chest x-ray (figure 1A) just prior to endotracheal intubation revealed bilateral air space disease and high-resolution CT of the chest (figure 1B) revealed irregular nodules, bilateral areas of conglomerate masses with air bronchograms of peribronchovascular orientation surrounded by ground glass opacities and no mediastinal lymphadenopathy. Laboratory findings indicated mild normocromic and normocytic anaemia, haemoglobin 10 g/dl, lymphopenia 168 cells/mm³, C-reactive protein 157 mg/litre (0.0–3.0 mg/litre). Tests for HIV, serum cytoplasmic antineutrophil cytoplasmic antibody, antinuclear antibody and rheumatoid factor were negative. A transesophageal echocardiogram displayed a bioprosthetic mitral valve with mild stenosis, no evidence of endocarditis, and minimum regurgitation with negative blood cultures. MRI of the brain displayed bilateral optic neuritis.

An open lung biopsy was performed: angiocentric nodules comprised a mixture of lymphocytes with variable numbers of associated plasma cells and histiocytes (figure 2A). The lymphocytic infiltrate was mainly composed of small and intermediate size cells that were CD3 positive T lymphocytes. Scattered large atypical cells (figure 2B) were CD20 positive B lymphocytes. The polymorphic lymphoid infiltrate showed a distinctive predilection to infiltrate vessel walls (figure 2C), resulting in thickening of the lumens of affected vessels (figure 2D).

**QUESTION**

What is the diagnosis?

See page 280 for the answer

---

**FUTURE STUDIES**

Evidence from other complex traits suggests that, with larger sample sizes, GWAS will reveal many more loci harbouring common sequence variants underlying respiratory diseases. Other studies will usefully focus on the role of rare sequence variants and structural genomic variants in respiratory disease. Understanding precisely how the molecular pathways highlighted by these genetic studies influence the development and progression of disease, and how they can be modified, provides a major challenge for the coming decade.

**Contributors**

The literature review and drafting of the manuscript was wholly undertaken by the named authors.

**Provenance and peer review**

Commissioned; internally peer reviewed.

**REFERENCES**

Genome-wide association studies in lung disease

María Soler Artigas, Louise V Wain and Martin D Tobin

Thorax 2012 67: 271-273 originally published online August 19, 2011
doi: 10.1136/thoraxjnli-2011-200724

Updated information and services can be found at:
http://thorax.bmj.com/content/67/3/271

These include:

References
This article cites 5 articles, 0 of which you can access for free at:
http://thorax.bmj.com/content/67/3/271#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/