

# The genetics of obstructive lung disease: big is beautiful

Ian P Hall,<sup>1</sup> David A Lomas<sup>2</sup>

Chronic obstructive pulmonary disease (COPD) is defined as airflow obstruction that is not fully reversible. It results from abnormal inflammation following exposure to noxious particles or gases.<sup>1</sup> This is typically exposure to cigarette smoke but may also include exposure to biomass fuels and some industrial dusts. COPD clusters within families, suggesting that heritable factors play a role in the pathogenesis of this disease.<sup>2,3</sup> The only genetic factor that is widely accepted to be associated with COPD is severe deficiency of  $\alpha_1$ -antitrypsin.<sup>4</sup> This depletes an important member of the antiprotease screen and causes excessive intrapulmonary inflammation.<sup>5,6</sup> Indeed, even asymptomatic non-smoking heterozygotes for the Z allele (PiMZ) without airflow obstruction have increased intrapulmonary inflammation.<sup>7</sup>

Severe  $\alpha_1$ -antitrypsin deficiency is found in only 1–2% of all individuals with COPD. There is growing interest in the genetic factors that predispose to COPD in those individuals who do not have  $\alpha_1$ -antitrypsin deficiency. Early papers focused on association studies that looked at genetic variation within individuals with COPD compared with those individuals who were matched for all factors that are associated with COPD (most importantly age, smoking history and ethnic background) but who did not have airflow obstruction. The studies were typically small (100–150 cases and controls) and were often confounded by failure to match cases and controls carefully. There was often the issue of correcting for multiple comparisons as well as the inherent complexity of the COPD phenotype.<sup>8</sup> Larger family-based studies have shown the independent clustering of the airway disease and emphysema component of COPD within families.<sup>9</sup> This suggests that different genetic factors

predispose to each of these components of the phenotype. The only way of overcoming the inherent variation in COPD is to focus on groups of well-characterised individuals with components of the COPD phenotype or to undertake studies with large sample sizes and then to replicate any positive findings in other cohorts. This is now the case with candidate gene studies and there is good evidence that heterozygotes for  $\alpha_1$ -antitrypsin deficiency (PiMZ) and polymorphisms in genes involved in oxidative stress (microsomal epoxide hydrolase, glutathione S-transferase, haeme oxygenase and superoxide dismutase 3) are associated with an increased risk of COPD.<sup>10</sup> Furthermore, polymorphisms in transforming growth factor-beta, a protein important in maintaining the extracellular matrix, are also likely to be important in the pathogenesis of this disease. More recently, a minor allele of a single nucleotide polymorphism (SNP) in matrix metalloproteinase-12 (MMP-12) has been shown to protect against COPD in adult smokers.<sup>11</sup>

The limitation of association studies is that genes are assessed in pathways that are already recognised to be associated with COPD—the proteinase–antiprotease and oxidative stress pathways—and genes that maintain the integrity of the extracellular matrix. In recent years the collection of large cohorts has allowed unbiased genome-wide association studies (GWAS) in individuals with COPD. The largest study was undertaken in a cohort from Bergen, Norway, and then replicated in the International COPD Genetics Network, the National Emphysema Treatment Trial with controls from the Normative Ageing Study and then finally in the Boston Early Onset COPD cohort.<sup>12</sup> Top hits from this analysis were SNPs in the alpha-nicotinic acetylcholine receptor and the hedgehog interacting protein (HHIP). The SNP in the alpha-nicotinic acetylcholine receptor (rs8034919) was also identified in three GWAS of lung cancer and is thought to be important in peripheral vascular disease and nicotine addiction.<sup>13–16</sup> It is likely that this SNP functions as a marker for an addiction gene. Individuals who carry this SNP may require more cigarettes to satisfy

nicotine addiction, may inhale more deeply and may find it more difficult to withdraw from cigarette smoking. Alternatively, the association may result from linkage disequilibrium with SNPs in the iron responsive element binding protein 2 (IREB2).<sup>17</sup> This was identified by expression analysis in lung tissue from individuals with COPD and then confirmed in three separate COPD cohorts. IREB2 is localised to the human epithelial cell surface and may play a role in protecting against epithelial damage from oxidative stress. Finally, the analysis of even larger numbers of individuals with COPD has identified a SNP in FAM13A as being associated with COPD. The role of this gene in the disease is unclear but expression has been associated with hypoxia.<sup>18</sup> FAM13A has also been associated with lung function in a second independent study.<sup>19</sup> A detailed analysis of these genes in well-characterised cohorts showed that SNPs in the  $\alpha$ -nicotinic acetylcholine receptor are associated with smoking intensity, airflow obstruction and emphysema,<sup>20,21</sup> SNPs in the hedgehog interacting protein are associated with systemic features of COPD (low body mass index) and exacerbations,<sup>21</sup> while SNPs in FAM13A are associated with airflow obstruction.<sup>21</sup>

Taken together, there is clear evidence that genes associated with the protease–antiprotease pathway, oxidative stress and the integrity of the extracellular matrix are involved in the susceptibility of smokers to COPD. GWAS have identified genes that are likely to have a role in addiction and other genes that may protect against oxidative stress. It is clear that the gene mining has only just begun, with good prospects for identifying further novel genes associated with COPD. This is important as they will allow us to stratify individuals with COPD and identify pathways that provide new insights into the mechanisms of disease. Clearly, the long-term aim is to intervene in these pathways and so prevent the relentless progression of the pulmonary and systemic manifestations of COPD. It is clear that further collaboration is required to bring together even larger numbers of well-phenotyped cohorts to identify more genes that are associated with COPD. The potential value of this approach has been demonstrated in two large meta-analyses examining lung function in general population samples. These studies, run by the Spirometa and CHARGE consortia, identified a number of novel genes which predict forced

<sup>1</sup>Division of Therapeutics and Molecular Medicine, Queen's Medical Centre, Nottingham, UK; <sup>2</sup>Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge, UK

**Correspondence to** Ian P Hall, Division of Therapeutics and Molecular Medicine, Queen's Medical Centre, Nottingham NG7 2UH, UK; [ian.hall@nottingham.ac.uk](mailto:ian.hall@nottingham.ac.uk)

expiratory volume in 1 s (FEV<sub>1</sub>) and/or the ratio of FEV<sub>1</sub> to forced vital capacity. Four genes identified were common to both studies- the previously observed HHIP, GSTCD (glutathione S transferase C-terminal domain), HTR4 (the serotonin 4 receptor subtype) and AGER (the receptor for advanced glycosylation end products).<sup>19–22</sup> While these studies concentrated on general population samples, it is highly likely that genetic factors predicting FEV<sub>1</sub> will be strong risk factors for COPD and other diseases where airflow obstruction is present; analyses addressing this issue are ongoing. The populations studied in each consortium included primary discovery samples of >20 000 in each. It seems to us therefore that the future for COPD genetics is very bright, but only if we think big!

There have been many more studies addressing the genetics of asthma than have been performed examining COPD, although the overall conclusion that very large studies are required to identify reproducible effects is the same when considering asthma. Initial attempts at hunting an asthma gene used predominantly family-based linkage approaches, which resulted in multiple loci on almost every chromosome being identified in different populations. Candidate gene studies have also been widely performed but, as with COPD, the ability to replicate findings has been very limited. The genes which have attracted the most interest from these studies include ADAM33, PHF11, GPRA and DPP10. While there have been a number of studies which have replicated initial associations, the effect size of individual SNPs for asthma risk has generally been small and functional mutations remain to be identified. For example, in a study which involved genotyping the complete 1958 birth cohort for risk SNPs in these genes prioritised on the basis of linkage data, the strongest signals seen were for alleles with odds ratios of <1.2 in ADAM33 and PHF11, with little evidence being observed for major effects at each locus.<sup>23</sup>

It is possible that the key SNPs in these genes remain to be defined or that the contribution of these genes is at best minimal. More recently, GWAS approaches have taken over as the method of choice to identify the genetic factors underlying asthma risk. The first two studies to examine this area resulted in the identification of the region containing the gene for ORM DL3 as an asthma locus, and

a number of regions which were associated with eosinophilia in patients with and without asthma, including the regions containing the genes such as IL5, IL1RL1 and GATA2.<sup>24–25</sup> There have been a number of recent publications using GWAS approaches in small studies but these have generally failed to identify signals which reach conventional genome-wide significance, probably because the studies have been underpowered. Greater clarity on the key genetic regions involved in determining asthma risk will undoubtedly come from larger studies which are shortly to be published, such as that organised by the GABRIEL consortium—again it seems ‘biggest is best’ when it comes to asthma genetics. Finally, the ability to interrogate the human genome is about to shift up another step with the use of deep sequencing approaches to perform, for example, whole exome sequencing studies which could allow us to move from identifying common variants associated with disease to the identification of rare or intermediate variants of interest.

**Funding** Work in both authors’ laboratories is supported by the MRC. In addition, IPH is supported by Asthma UK and the Nottingham Biomedical Research unit and DAL receives support from the Wellcome Trust, EPSRC, BBSRC, the Alpha-1 Foundation and Papworth Hospital.

**Competing interests** None.

**Provenance and peer review** Commissioned; not externally peer reviewed.

*Thorax* 2010;**65**:760–761.  
doi:10.1136/thx.2010.147173

## REFERENCES

1. **Rabe KF**, Hurd S, Anzueto A, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;**176**:532–55.
2. **Silverman EK**, Chapman HA, Drazen JM, *et al*. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;**157**:1770–8.
3. **McCloskey SC**, Patel BD, Hinchliffe SJ, *et al*. Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *Am J Respir Crit Care Med* 2001;**164**:1419–24.
4. **DeMeo DL**, Silverman EK. Genetic aspects of  $\alpha_1$ -antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. *Thorax* 2004;**59**:259–64.
5. **Morrison HM**, Kramps JA, Burnett D, *et al*. Lung lavage fluid from patients with  $\alpha_1$ -proteinase inhibitor deficiency or chronic obstructive bronchitis: anti-elastase function and cell profile. *Clin Sci* 1987;**72**:373–81.
6. **Goopu B**, Lomas DA. Polymers and inflammation: disease mechanisms of the serpinopathies. *J Exp Med* 2008;**205**:1529–34.
7. **Malerba M**, Ricciardolo F, Radaeli A, *et al*. Neutrophilic inflammation and IL-8 levels in induced

- sputum of alpha-1-antitrypsin PiMZ subjects. *Thorax* 2006;**61**:129–33.
8. **Lomas DA**, Silverman EK. Genetics of chronic obstructive pulmonary disease. *Respir Res* 2001;**2**:20–36.
9. **Patel BD**, Coxson HO, Pillai SG, *et al*. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008;**178**:500–5.
10. **Marciniak SJ**, Lomas DA. What can naturally occurring mutations tell us about the pathogenesis of COPD? *Thorax* 2009;**64**:359–64.
11. **Hunninghake GM**, Cho MH, Tesfaigzi Y, *et al*. MMP12, lung function, and COPD in high-risk populations. *N Engl J Med* 2009;**361**:2599–608.
12. **Pillai SG**, Ge D, Zhu G, *et al*. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;**3**:e1000421.
13. **Hung RJ**, McKay JD, Gaborieau V, *et al*. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;**452**:633–7.
14. **Amos CI**, Wu X, Broderick P, *et al*. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;**40**:616–22.
15. **Thorgerisson TE**, Geller F, Sulem P, *et al*. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;**452**:638–42.
16. **Saccone SF**, Hinrichs AL, Saccone NL, *et al*. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007;**16**:36–49.
17. **DeMeo DL**, Mariani T, Bhattacharya S, *et al*. Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *Am J Hum Genet* 2009;**85**:493–502.
18. **Cho MH**, Boutaoui N, Klanderman BJ, *et al*. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. In press.
19. **Hancock DB**, Eijgelsheim M, Wilk JB, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;**42**:45–52.
20. **Lambrechts D**, Buyschaert I, Zanen P, *et al*. The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *Am J Respir Crit Care Med* 2010;**181**:486–93.
21. **Pillai SG**, Kong X, Edwards LD, *et al*. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. In press.
22. **Repapi E**, Sayers I, Wain LV, *et al*. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010;**42**:36–44.
23. **Blakey JD**, Sayers I, Ring SM, *et al*. Positionally cloned asthma susceptibility gene polymorphisms and disease risk in the British 1958 birth cohort. *Thorax* 2009;**64**:381–7.
24. **Gudbjartsson DF**, Bjornsdottir US, Halapi E, *et al*. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. *Nat Genet* 2009;**41**:342–7.
25. **Moffatt MF**, Kabisch M, Liang L, *et al*. Genetic variants regulating ORM DL3 expression contribute to the risk of childhood asthma. *Nature* 2007;**448**:470–3.