

EDITORIAL

Genetics of COPD: present and future

J.R. Hoidal

Chronic obstructive pulmonary disease (COPD) is a devastating clinical problem. In the USA, it affects >16 million people, accounts for 13% of hospitalizations, and is the fourth leading cause of death [1]. Similarly, COPD is a major cause of morbidity and mortality worldwide. Over the next half-century, one and a half million deaths·yr⁻¹ due to COPD are predicted in China alone [2]. Although cigarette smoking is the major risk factor, only 10–20% of smokers develop symptomatic COPD. The extent of cigarette smoking (pack-yrs and duration of smoking) only account for ~15% of the variation in lung function [3], indicating that differences in susceptibility to COPD must exist. Numerous epidemiological studies provide compelling evidence that genetic factors influence the development of COPD. Case-control studies have demonstrated an increased prevalence of COPD in relatives of patients with COPD that cannot be accounted for by known risk factors such as smoking [4–6]. Twin studies also suggest a genetic basis to the development of COPD [7].

To date, α_1 -antitrypsin (α_1 -AT) deficiency remains the only proven genetic risk factor for COPD. The homozygous-deficient phenotype is rare (<0.1%), however, so only a small percentage (1–2%) of the genetic susceptibility to cigarette smoke is due to this phenotype. The best models of genetic components of COPD indicate that with the exception of α_1 -AT it is a genetically complex disease [8]. This is as expected given that the pathogenesis of COPD probably involves several cell types, enzymes and inflammatory mediators interacting in an intricate manner to influence the development of airways inflammation and parenchymal destruction.

For genetically complex diseases, recent emphasis has been placed on the common disease-common variant hypothesis, which states that common disease susceptibility or resistance variants are expected to be few at each locus, relatively common in the human population and enriched in the coding and regulatory sequence of genes [9]. Despite the small effects of such genes individually, the magnitude of their attributable risk (the proportion of people affected by them) may be large because they are quite frequent in the population, making them of public health significance. Thus, in COPD, it is likely that multiple genes are operating and the genetic susceptibility may depend

on the coincidence of several gene polymorphisms acting together. Moreover, expression of different gene combinations probably affects the heterogeneous histopathological and clinical profile of COPD seen among individuals.

To date, the genes implicated in the pathogenesis of COPD have been found to be involved in antiproteolysis, metabolism of toxic substances in cigarette smoke, the inflammatory response to cigarette smoke, and the efficiency of mucociliary clearance in the lung. A variety of studies have examined candidate gene loci with association studies, comparing the distribution of variants in genes hypothesized to be involved in the development of COPD in patients and control subjects. Polymorphisms in the genes for α_1 -AT, α_1 -antichymotrypsin, microsomal epoxide hydrolase, vitamin D-binding protein, glutathione-S transferase, cytochrome P450 1A1, immunoglobulin-A, tumour necrosis factor (TNF)- α and haemoxygenase have all been associated with the development of COPD in cigarette smokers [10–16]. Many of the associations are controversial because they have not been replicated [17–19].

One key to establishing that a gene modifies the risk for a disease is replication of the association in different populations. This can be a difficult task. Several factors could contribute to the inconsistent results of case-control genetic association studies in COPD. Genetic heterogeneity or different genetic mechanisms in different populations could play a part in making it difficult to replicate associations between studies. In addition, false-positive or false-negative results could contribute. To be useful, polymorphisms need to be real, not errors or artefacts. A further well-recognized potential problem is ascertainment bias. Case-control association studies are susceptible to supporting associations based purely on population stratification that may result from incomplete matching between cases and controls, including differences in ethnicity and geographical origin. To the authors' knowledge, there are as yet no reported association studies in COPD that use family-based controls, a study design that is immune to population stratification effects.

In this issue of the *European Respiratory Journal*, HIRANO *et al.* [20] suggest for the first time that variants in the tissue inhibitor of metalloproteinases (TIMP)-2 gene predispose to COPD. The phenotype used is pulmonary function, as measured by forced expiratory volume in one second, which is a strong correlate of COPD and, in population studies, is

predictive of all-cause mortality [21–23]. The association of TIMP-2 polymorphisms with COPD has good biological plausibility. As referenced by HIRANO *et al.* [20], there are substantial *in vitro* and *in vivo* data to suggest that metalloproteinases may be important in the pathobiochemistry of COPD. The authors also indicate that the results are not due to the confounder of population stratification, although this possibility cannot be completely excluded by the information provided. In future studies, it will be important to determine whether the observation can be confirmed in different populations and to ascertain how the observed polymorphisms affect the expression or activity of TIMP-2.

What lies ahead in efforts to uncover the genetic basis of COPD? RISCH and MERIKANGAS [24] suggested that the future of genetic studies for complex polygenic diseases such as COPD will require large-scale testing by association analysis, rather than the linkage analysis approach that has been used successfully to find major genes. They indicate that many genetic effects, too weak to identify by linkage, can be effectively detected by genomic association studies. The limitation of association studies is that the actual polymorphisms within the gene (or at least a polymorphism in strong disequilibrium) must be available. A large number of genes and polymorphisms must first be identified, and an extremely large number of such polymorphisms will need to be tested. Testing such a large number of polymorphisms on several hundred individuals has, until recently, seemed unrealistic in scope. However, with the publication of drafts of human genome sequences [25, 26], the human genome projects have provided the first complete sequence of the reference human genome. This landmark feat will accelerate the pace at which the spectrum of genetic variation in the human gene pool and the genetic basis of many complex human phenotypes can be elucidated. There are now numerous ongoing efforts to catalogue sequence variants. Emphasis is being placed on detecting single nucleotide polymorphisms (SNPs), the most common human sequence variation. Two approaches include the targeted resequencing of genomic deoxyribonucleic acid (DNA) from diverse individuals, and random, whole genome approaches that utilize either pre-existing datasets (*e.g.* expressed sequence tag databases) or collected random sequence samples of the genome. An example of the latter approach is the work spearheaded by the SNP Consortium Ltd, a group of 10 pharmaceutical companies, and the Wellcome Trust. The SNP Consortium is set to produce 300,000 publicly available SNPs within the next year. Improved technologies for scoring SNPs are reducing hands-on time and cost, although truly high-throughput methods are still lacking for genome-wide, population-based studies.

For sequence variants to be fully utilized, there must be up-to-date, complete, coherent and organized views of this complex dataset so that investigators are empowered and not overwhelmed by its use. Investigators using the genome sequence must have access to electronic resources that allow them to gather, manipulate and display sequence-based information

so that they can derive the intended benefits. Integrated, up-to-date sequence-based views of complementary DNA and genomic sequences are essential for the analysis of common variation in candidate genes, and the prioritization of candidate genes and regions for further analysis. Bioinformatics tools will be at the core of this approach. Powerful new types of bioinformatics will be required to assimilate and interpret the data that will result from various types of genomics research.

The rate of detection and reporting of polymorphisms is increasing logarithmically. In the last decade, ~1,000 polymorphisms were reported. This number has increased 1,000-fold in the past year alone. A map of publicly available SNPs, fully integrated with the sequence as well as the physical and genetic maps of the human genome, has recently been described [27]. This and other such maps that are certain to be generated in the near future, provide unprecedented tools for studying the character of human sequence variation and will be immediately applicable for candidate-gene studies for disease association. The benefit of SNP maps is that they cover the entire genome. Thus, by comparing patterns and frequencies of SNPs in patients and controls, SNPs associated with specific diseases can be identified. As the speed and efficiency of SNP genotyping increases, studies of SNPs and diseases will become more efficient. The hypothesis that common variants contribute significantly to the risk of common diseases will be comprehensively tested. To the extent that such studies are successful, they should profoundly affect the understanding of COPD and ultimately, the development of new and more effective therapies.

A word of caution. The choice of phenotype is critical to the success of gene discovery programmes. If the cases include a heterogeneous collection of aetiologies for a complex disease, the power to detect a significant association will be reduced. Similarly, if the control group includes subjects who are affected but undiagnosed, power will be reduced. This is particularly relevant to gene discovery in COPD because of its variable pathological features and late onset of disease.

In summary, establishing molecular variants that may constitute inherited predispositions to chronic obstructive pulmonary disease can be a daunting task. However, it will be far more challenging to establish causation for a chronic obstructive pulmonary disease phenotype; that is, to understand the mechanism by which a genetic factor may predispose to the disorder. Ultimately, the studies of disease gene identification will have to be turned back over to biologists and physiologists because the identification of the functional disease mutation depends upon the demonstration of a biological basis for one or more of the putative genetic variants in the development of disease. As with the tissue inhibitor of metalloproteinases-2, the product of the gene under examination may be highly pleiotropic, due to its involvement with multiple physiological processes in multiple tissues, and determination of the role of a specific variant may require extensive manipulation in intact organisms.

References

1. Ventura SJ, Peters KD, Martin JA, Maurer JD. Births and deaths: United States, 1996. *Mon Vital Stat Rep* 1997; 46: Suppl. 2, 1–40.
2. Peto R, Chen ZM, Boreham J. Tobacco - the growing epidemic. *Nat Med* 1999; 5: 15–17.
3. Bascom R. Differential susceptibility to tobacco smoke: possible mechanisms. *Pharmacogenetics* 1991; 1: 102–106.
4. Tager I, Tishler PV, Rosner B, Speizer FE, Litt M. Studies of the familial aggregation of chronic bronchitis and obstructive airways disease. *Int J Epidemiol* 1978; 7: 55–62.
5. Khoury MJ, Beaty TH, Newill CA, Bryant S, Cohen BH. Genetic-environmental interactions in chronic airways obstruction. *Int J Epidemiol* 1986; 15: 65–72.
6. Kueppers F, Miller RD, Gordon H, Hepper NG, Offord K. Familial prevalence of chronic obstructive pulmonary disease in a matched pair study. *Am J Med* 1977; 63: 336–342.
7. Sandford AJ, Weir TD, Pare PD. Genetic risk factors for chronic obstructive pulmonary disease. *Eur Respir J* 1997; 10: 1380–1391.
8. Givelber RJ, Couropmitree NN, Gottlieb DJ, *et al.* Segregation analysis of pulmonary function among families in the Framingham Study. *Am J Respir Crit Care Med* 1998; 157: 1445–1451.
9. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997; 278: 1580–1581.
10. Eriksson S, Lindmark B, Lilja H. Familial α_1 antichymotrypsin deficiency. *Acta Med Scand* 1986; 220: 447–453.
11. Kauffmann F, Kleisbauer J-P, Cambon-de-Mouzon A, *et al.* Genetic markers in chronic airflow limitation: a genetic epidemiologic study. *Am Rev Respir Dis* 1983; 127: 263–269.
12. Beatty TH, Menkes HA, Cohen BH, Newill CA. Risk factors associated with longitudinal change in pulmonary function. *Am Rev Respir Dis* 1984; 129: 660–667.
13. Cantlay AM, Lamb D, Gillooly M, *et al.* Association between the CYP1A1 gene polymorphism and susceptibility to emphysema and lung cancer. *Clin Mol Pathol* 1995; 48: M210–M214.
14. Huang S-L, Su C-H, Chang SC. Tumor necrosis factor-alpha gene polymorphism in chronic bronchitis. *Am J Respir Crit Care Med* 1997; 156: 1436–1439.
15. Horne SL, Cockcroft DW, Dosman JA. Possible protective effect against chronic obstructive airways disease by the GC2 allele. *Hum Hered* 1990; 40: 173–176.
16. Yamada N, Yamaya M, Okinaga S, *et al.* Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. *Am J Hum Genet* 2000; 66: 187–195.
17. Higham MA, Pride NB, Alikhan A, Morrell NW. Tumour necrosis factor- α gene promoter polymorphism in chronic obstructive pulmonary disease. *Eur Respir J* 2000; 15: 281–284.
18. Yim J-J, Park GY, Lee C-T, *et al.* Genetic susceptibility to chronic obstructive pulmonary disease in Koreans: combined analysis of polymorphic genotypes for microsomal epoxide hydrolase and glutathione S-transferase M1 and T1. *Thorax* 2000; 55: 121–125.
19. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Pare PD. Susceptibility genes for rapid decline of lung function in the lung health study. *Am J Respir Crit Care Med* 2001; 163: 469–473.
20. Hirano K, Sakamoto T, Uchida Y, *et al.* Tissue inhibitor of metalloproteinases-2 gene polymorphisms in chronic obstructive pulmonary disease. *Eur Respir J* 2001; 18: 748–752.
21. Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, Hawthorne VM. Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley Prospective Population Study. *BMJ* 1996; 313: 711–715.
22. Jedrychowski W, Maugeri U, Gomola K, Tobias-Adamczyk B. Ventilatory lung function level as a predictor of survival among the elderly. *Monaldi Arch Chest Dis* 1994; 49: 293–297.
23. Sorlie PD, Kannel WB, O'Connor GT. Mortality associated with respiratory function and symptoms: the Framingham Study. *Am Rev Respir Dis* 1989; 140: 379–384.
24. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273: 1516–1517.
25. Venter JC, Adams MD, Meyers EW, *et al.* The sequence of the human genome. *Science* 2001; 291: 1304–1351.
26. Lander ES, Linton LM, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
27. Sachidanandam R, Weissman D, Schmidt SC, *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409: 923–933.