

REVIEW

What can genetics tell us about the cause of fixed airflow obstruction?

L. V. Wain, M. Soler Artigas and M. D. Tobin

Departments of Health Sciences & Genetics, University of Leicester, Leicester, UK

Clinical & Experimental Allergy

Correspondence:

Martin D. Tobin, Departments of Health Sciences & Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

E-mail: mt47@le.ac.uk

Cite this as: L. V. Wain, M. Soler Artigas, and M. D. Tobin, *Clinical & Experimental Allergy*, 2012 (42) 1176–1182.

Summary

Chronic obstructive pulmonary disease (COPD) is a major cause of chronic morbidity and mortality worldwide with smoking being the most important risk factor of the disease. However, lung function and COPD are known to also have a genetic component and a deeper knowledge of the genetic architecture of the disease could lead to further understanding of predisposition to COPD and also to development of new therapeutic interventions. Genetic linkage studies and candidate gene association studies have not provided evidence to convincingly identify the genes underlying lung function or COPD. However, recent large genome-wide association studies (GWAS) including tens of thousands of individuals have identified 26 variants at different loci in the human genome that show robust association with quantitative lung function measures in the general population. A growing number of these variants are being shown to be associated with COPD. Following the identification of these new lung function loci, the challenge now lies in refining the signals to identify the causative variants underlying the association signals and relating these signals to the molecular pathways that underlie lung function.

Keywords COPD, genetics, genome-wide association studies, lung function

Submitted 01 November 2011; revised 16 December 2011; accepted 18 January 2012

Introduction

Fixed airflow obstruction is a characteristic and diagnostic feature of chronic obstructive pulmonary disease (COPD), one of the major causes of chronic morbidity and mortality worldwide [1]. Smoking is the single most important risk factor in predicting COPD risk, but only a modest proportion of the population variability in lung function is explained by smoking, suggesting that other factors are also important [2]. Childhood respiratory infections, air pollution and occupational exposures are among other risk factors for fixed airflow obstruction [3]. Several sources of evidence point to the importance of genetic contributions to fixed airflow obstruction. Identifying the genetic determinants of fixed airflow obstruction can elucidate the molecular mechanisms underpinning development and progression of COPD, and highlight new molecular targets for drug treatments. Furthermore, if genetic variation can explain the clinical heterogeneity in COPD, or the heterogeneous responses to treatment, it could lead to new approaches to diagnosis and stratified treatment for COPD [4].

The arguments for identifying the genetic determinants of respiratory diseases and related traits have

been well rehearsed but, until a few years ago, the genetic determinants proved to be more elusive than was originally hoped [5]. Over the last few years, improved molecular tools and more powerful study designs have provided a plethora of new findings. In this article, we will provide a summary of progress in understanding the genetic determinants of fixed airflow obstruction, summarize recent findings and discuss the potential utility of these findings.

Heritability, linkage studies and candidate gene association studies

Lung function measurements and COPD aggregate in families [6–8]. The proportion of variance in lung function measures attributable to the additive effects of genes, termed the narrow sense heritability, is probably around 40% [9–12]. Though estimates of heritability do vary, and caution needs to be exercised regarding their interpretation [13], these studies provide a justification for searching for the genetic determinants. When microsatellite markers became available to undertake genome-wide linkage studies, there was optimism that linkage studies would provide an approximate genomic

location for variants contributing to the complex disease causation. Regions of linkage to COPD or lung function were reported in a variety of different genomic locations, with some regions overlapping those reported in other studies – for example on chromosome 2q [14–16]. Linkage signals indicate broad genomic regions often containing hundreds of genes within which causal genetic variants may lie. Testing association with COPD across multiple single nucleotide polymorphisms ('SNPs') in selected ('candidate') genes within linkage regions has been used to fine map these signals. However, inconsistent results have been produced across different studies, making it difficult to interpret findings. For example, although *SERPINE2* lies under a reported linkage peak at chromosome 2q33, reports of association have been highly variable between studies [17, 18]. Alternatively, candidate genes have been selected due to their involvement in pathways already known to influence COPD, such as in proteins that maintain the extracellular matrix or which are involved in oxidative stress [19]. There has been a plethora of candidate gene association studies reported for COPD, and many for lung function. The most extensive evidence probably relates to the *GSTM1* null polymorphism, and to SNPs in *TNF*, *TGFbeta1* and *SOD3*. These have been reviewed elsewhere [20, 21]. However, the evidence across all published candidate gene association studies of COPD and lung function is hard to interpret [22, 23]. The majority of published candidate gene associations are expected to be false positive reports [24, 25] and it is difficult to disentangle the very many false positive associations from any causal associations. Inconsistent findings were sometimes attributed to incomplete adjustment for smoking behaviour or for differences in population ancestry. More important, however, were small sample sizes and inadequate adjustment for multiple statistical testing, coupled with reporting biases (such as non-reporting of non-significant findings or undue emphasis on subgroup findings) and severe publication bias. Many of these problems have been overcome in complex disease genetics with the introduction of well powered genome-wide association studies (GWAS) [26].

Genome-wide association studies (GWAS)

Introduction to Genome-wide association studies (GWAS)

The GWAS became possible with reducing costs and miniaturization of genotyping assays such that hundreds of thousands of common SNPs could be affordably assayed on a single chip in every individual in a study. An extremely large number of associations are tested in a genome-wide association study, typically

between 0.5 million and 2.5 million SNPs in 2011, providing a hypothesis-free approach for discovery. Rigorous corrections for multiple testing are required and a P value $< 5 \times 10^{-8}$ is commonly used as a threshold for 'genome-wide significance'. Together with appropriate quality control and analytical approaches, the rigorous statistical thresholds employed in GWAS have minimized the reporting of false positive associations [27]. The design is also much less susceptible to the selective publication or selective reporting that has proved so problematic for candidate gene association studies.

Common variants individually have generally modest effects on common diseases and complex traits. Thus, to have power to detect genome-wide significant associations, consortia have been established to facilitate GWAS in very large sample sizes [28, 29]. To enable data from multiple studies using different commercial genome-wide genotyping platforms to be combined, 'imputation' can be undertaken in each study to infer the genotypes at SNPs which have not been directly assayed in that particular study. SNPs that are located close together in the genome are unlikely to be separated by recombination events at meiosis and tend to be correlated as they are passed down together through many generations. Thus, we can predict an individual's genotype at a given SNP from the genotypes at surrounding SNPs, often with a very high degree of certainty, from a panel of reference genotypes such as HapMap [30]. Where a genotype is imputed with some uncertainty, the uncertainty can be taken account of in the association testing, and SNPs with low-quality imputation are excluded altogether.

As logistic issues, intellectual property or consent issues can limit the extent of sharing of individual-level data, consortium GWAS have employed meta-analysis approaches, where participating studies impute to the same reference panel, undertake analyses to a common analysis plan and combine the findings. This is very different from a meta-analysis based on published data (especially given the publication biases described above). A common study design for a GWAS involves two stages. An initial discovery stage combines genome-wide association data from a number of studies in meta-analysis, and identifies a much smaller set of independent genetic variants which reach genome-wide significance, or near genome-wide significance, to be tested for association in additional studies in a follow-up stage. The recommended threshold for genome-wide significance ($P < 5 \times 10^{-8}$) refers to association tests after meta-analysis across the discovery and follow-up stages [26].

For readers unfamiliar with GWAS studies, it can seem baffling that little adjustment, or sometimes no adjustment, is made for covariates that would typically confound inferences in observational studies. There are

two reasons why this is the case. The first is that genetic studies are not prone to confounding by life-style factors. This issue, referred to as Mendelian randomization, is considered further below (subsection entitled '*The role of smoking*'). The second is that very large sample sizes are critical in GWAS studies to achieve the power to detect the individually modest effects of common genetic variants on most complex traits and common diseases, and greater complexity of the study design leads to attrition of sample size.

Genome-wide association studies (GWAS) findings for lung function and Chronic obstructive pulmonary disease (COPD)

The study of quantitative lung function traits in the general population provides a powerful means to study the genetics underlying COPD, diagnosis of which is based on forced expiratory volume in 1 s (FEV₁) and FEV₁/FVC. To date, GWAS have identified 26 loci associated with (FEV₁) or the ratio of FEV₁ to forced vital capacity (FVC), corrected for age, gender and height, in sample sizes exceeding 20 000 individuals [28, 29, 31] (Fig. 1). Ten loci discovered by Wilk et al. [32] in 2009, and by Repapi et al. [29] and Hancock et al. [28] in 2010 were subsequently reconfirmed in a larger meta-analysis [31] with a further 16 loci identified as showing genome-wide significant association with lung function after meta-analysis of data from up to 96 612 individuals. Of the variants at 26 loci associated with lung function described above, so far seven have been shown to be associated with COPD [33–36] (Fig. 1) and this number seems set to increase as additional well-powered studies are in progress.

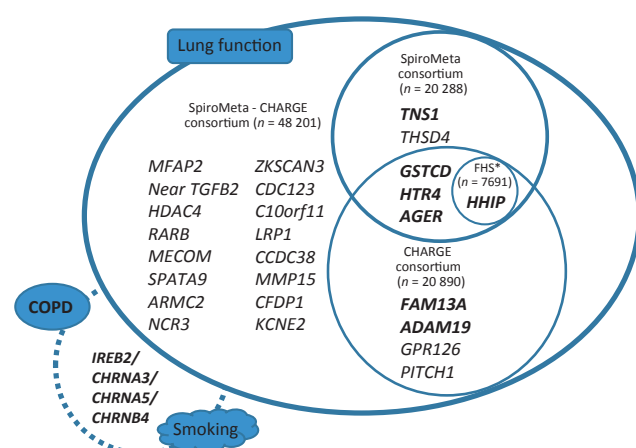


Fig. 1. Key genes representing the 26 loci showing association with lung function and the 15q25 locus showing significant association with COPD and smoking. Lung function genes in bold have also shown association with COPD. *Wilk et al. 2009 [32].

The SNPs upstream of *HHIP* show strongly significant association with lung function [29, 32] and COPD [35, 36]. SNPs in linkage disequilibrium with the SNP showing the strongest lung function associations also show association with height [37] suggesting a role in skeletal growth and development. However, the SNP, rs7689420, showing the strongest association with height lies within an intron of *HHIP* [38] and does not show association with FEV₁ or FEV₁/FVC [31]. SNPs within *HTR4* also show significant association with both FEV₁ and FEV₁/FVC. *HTR4* encodes 5-hydroxytryptamine receptor-4, a G-protein coupled receptor, activation of which protects spontaneous respiratory activity [39]. SNP rs3995090 in *HTR4* has also since been demonstrated to show association with COPD [36]. Other genes of note which are encompassed by the signal at this locus include members of the SPINK family of serine protease inhibitors. SNP rs7068966 in *CDC123* at 10p13 has also shown significant association with both FEV₁ and FEV₁/FVC.

The Major Histocompatibility Complex (MHC) region on chromosome 6 shows extended regions of linkage disequilibrium and has yielded many disease associations, particularly for autoimmune and inflammatory disorders [40]. Three independent signals for association with lung function (two showing strongest association with FEV₁/FVC and one showing strongest association with FEV₁) have been identified between 6p21.32 and 6p22.1. A non-synonymous SNP in *AGER* (also known as *RAGE*) shows association with FEV₁/FVC [29, 31] and also COPD [33]. *AGER* encodes a multi-ligand receptor of the immunoglobulin superfamily with a potential role in epithelium-extracellular matrix interactions and *Ager*^{-/-} mice develop age-related pulmonary fibrosis [41]. The two further regions of lung function association within the MHC identified by Soler Artigas et al. [31] encompass a number of genes with a variety of functions. The signal centred around SNP rs2857595 upstream of *NCR3* showing association with FEV₁/FVC also includes the plausible candidate genes lymphotoxin alpha (*LTA*) and tumour necrosis factor A (*TNFA*) which have previously been reported as showing association with lung function and COPD [42–46], and signals of association with psoriasis, lung cancer, neonatal lupus and systemic lupus erythematosus. The signal of association with FEV₁ (led by rs6903823 in an intron of *ZKSCAN3/ZNF323*) encompasses a number of zinc finger transcription factors and overlaps a signal of association with lung cancer [47].

Other genes with variants showing strongest association with FEV₁ include *GSTCD* at 4q24, *TNS1* at 2q35, *MECOM* at 3q26.1 and *C10orf11* at 10q22.3. The SNP within *GSTCD* appears to have the largest influence on FEV₁ with estimates of the effect size for each copy of the risk allele on FEV₁ being up to 52 mL [29] – equivalent to approximately 3 years of FEV₁ decline in non-smokers

[48]. *GSTCD* and *TNS1* have also shown significant association with COPD [33, 36]. The region of association which peaks within an intron of *C10orf11* also includes *KCNMA1* which encodes a potassium large conductance calcium-activated channel which is present on airway smooth muscle cells and pulmonary endothelium and its activation may lead to bronchial relaxation [49] and pulmonary arterial vasodilation [50].

Among the genes showing strongest association with FEV₁/FVC are several which are good candidates for involvement in lung function. *GPR126* encodes a G protein-coupled receptor involved in cell adhesion and signalling. Variants within this gene have also been associated with trunk height, although the association with lung function is robust to adjustments for trunk height [28]. Expression studies in mice also link this gene with lung development [51]. *FAM13A*, thought to play a role in signal transduction, has been previously associated with COPD [34]. *ADAM19*, encodes a member of a disintegrin and metalloprotease (ADAM) family of membrane-anchored glycoproteins involved in the control of cell-matrix interactions and regulation of growth and morphogenesis, and has also shown association with COPD [33]. Variants in another ADAM family protein, *ADAM33*, have been associated with bronchial hyperresponsiveness and accelerated lung function decline in both asthmatics and the general population [52–54]. *RARB*, encoding retinoic acid receptor B, has a role in controlling cell proliferation and differentiation and has been implicated in early lung development in mice [55, 56]. The SNP rs2284746, showing strongest association within *MFAP2* with FEV₁/FVC, is also strongly associated with height [38]. Within the same region as *MFAP2*, further plausible candidate genes include *CROCC*, which encodes a component of cilia [57], and *PADI2* which encodes a protein with increased levels in bronchial mucosa and bronchoalveolar lavage cells of smokers [58]. *HDAC4* encodes histone deacetylase 4; reduced activity of other histone deacetylases has been reported in lung tissue from COPD patients and correlated with disease severity and inflammatory cytokine expression [59, 60]. *Hdac4*-null mice die soon after birth due to skeletal defects and growth retardation that impair motility and breathing [61]. The intronic SNP rs12447804 showing strongest association with FEV₁/FVC within *MMP15* is highly correlated with a non-synonymous SNP, rs3743563 which results in a Glycine to Arginine change, in *MMP15*. Matrix metalloproteinases (MMPs) are a family of proteolytic enzymes which play an important role in tissue remodelling and repair associated with development and inflammation [62]. A recent study has shown that *Mmp15* expression is increased in the lungs of mice exposed to particulate matter [63]. There is some evidence supporting associations of *MMP1* and *MMP12*

with smoking-related injury [64], and associations of *MMP12* and COPD [65]. Conflicting results have been found for *MMP9* [21]. *MMP9* seemed to show an association with COPD in Chinese and Japanese populations [66, 67]; however, the association did not appear in Caucasian populations [64].

Other regions implicated by large-scale GWAS as having a role in lung function include *THSD4*, *PTCH1*, downstream of *TGFB2*, *SPATA9*, *ARMC2*, *LRP1*, *CCDC38*, *CFDP1* and *KCNE2*.

The role of smoking

One of the benefits of genetic studies is that they can avoid the confounding that has limited observational epidemiology studies. The essentially random allocation of genetic variants during meiosis means that lifestyle factors do not confound genetic associations – a principle known as Mendelian randomization [68, 69]. However, as smoking is a powerful predictor of lung function, it is important to consider whether smoking could be an intermediate on the causal pathway – that is, could a genetic association with lung function be mediated via an effect on tobacco smoking? In fact, all 26 variants showed genome-wide significant association after taking smoking status into account [31] and in a large consortium study including over 15 000 smokers, none of these variants were found to be associated with smoking behaviour [70]. A further consideration is whether or not the effects of a genetic variant might differ systematically between smokers and non-smokers. This is generally tested by a statistical test for interaction between the SNP and smoking status. In the case of the 26 variants associated with lung function [31], none showed an interaction with smoking. It is possible that GWAS that account for the presence of interactions could detect additional variants associated with lung function or COPD, and such studies are in progress.

A broad locus at 15q25, including the genes *IREB2*, *CHRNA3*, *CHRNA5* and *CHRNA4*, is strongly associated with smoking amount [70–72] and COPD [35, 73] and shows some evidence of association with lung function [31] although the latter association test does not reach genome-wide significance (Fig. 1).

Genetic architecture of lung function

Collectively, the findings from large-scale GWAS of lung function explain a modest proportion of the variance of lung function attributed to genetic factors (3.2% for FEV₁/FVC and 1.5% for FEV₁) [31]. The number of undiscovered genetic variants with similar effect sizes to the 26 described above is estimated to be 76 but collectively, these 102 variants still only explain approximately 7.5% of the additive polygenic variance

for FEV₁/FVC and 3.4% for FEV₁ [31]. Thus, a large proportion of the genetic variance in lung function remains unexplained. Furthermore, GWAS with larger sample sizes would be expected to yield further loci with similar or smaller effect sizes and fine-mapping of the existing signals could reveal rare variants with larger effect sizes. Experience from the study of other complex traits suggests that both common and rare sequence variants (SNPs) contribute to lung function and COPD risk. Copy number variation is another form of human genetic variation whereby sections of the genome can vary between individuals by deletion or duplication [74]. This area has not yet been well-studied in relation to lung function and could help to explain more of the genetic variance. Epigenetic inheritance, where variation in gene expression and regulation is not passed on directly via DNA changes, but is instead mediated by heritable factors such as DNA methylation and histone deacetylation, could also play a role.

Potential utility of GWAS findings

The studies described above provide us with novel insights into pathways involved in the development and, potentially, in the progression of airway obstruction. Many of the pathways highlighted by SNP associations have not previously been implicated in respiratory disease. It is possible that one or more of these pathways could prove to be amenable to modification by new drugs. However, developing a comprehensive understanding of how some of the genetic variants mediate their effect on lung function and COPD risk could take many years of painstaking functional work including, for example, expression studies in *ex-vivo* human tissues and experiments in animal models to further understand the functions of the genes. Several of the regions showing association with lung function described above encompass a number of genes and fine-mapping the signal to one particular gene or mutation would ensure that expensive functional studies are targeted appropriately. Such fine mapping can be helped by the identification of associated rare variants within these regions, for example using re-sequencing approaches.

In general, common genetic variants have not proven to have clinical utility to screen for disease risk as they individually have modest effects and collectively add little to established risk scores (for example in prediction of breast cancer risk [75] or type 2 diabetes risk [76]). However, the situation may be different for COPD risk prediction, because the absolute risk of developing COPD in smokers is very high [77]. Private companies are already marketing genetic tests for COPD and lung cancer prediction directly to the public. As with other screening tests, evidence of improvement of clinical outcomes attributable to the screening is warranted before widespread adoption of such testing.

Concluding remarks and future prospects

Major breakthroughs in understanding the common genomic variants influencing lung function have been made through the study of genome-wide association in very large samples sizes. At least seven of the 26 variants showing genome-wide association with lung function show association with COPD, providing a proof of concept that this can be a successful strategy to detect common genetic variants associated with COPD. Further studies to understand the effects of some of these variants in well-characterized cohorts of COPD patients are ongoing. Additional common variants remain to be discovered and may require even larger sample sizes. New molecular tools, including re-sequencing and new genotyping arrays, and genomic resources such as the 1000 Genomes project are improving the potential to assess the role of low frequency (minor allele frequency < 5%) genetic variants in determining lung function and COPD susceptibility, and the study of rare genetic variants is likely to be a major focus of research in 2012–2013. Rare variant discovery could assist with fine mapping and well-targeted functional follow-up to fully define the molecular mechanisms involved and to assess their potential for therapeutic intervention.

Acknowledgements

LVW and MDT have received funding for research from Pfizer.

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 Beck GJ, Doyle CA, Schachter EN. Smoking and lung function. *Am Rev Respir Dis* 1981; **123**:149–55.
- 3 Silverman EK, Speizer FE. Risk factors for the development of chronic obstructive pulmonary disease. *Med Clin North Am* 1996; **80**:501–22.
- 4 Hall IP. How will genetic approaches assist in the management of respiratory diseases?. *Curr Opin Pharmacol* 2009; **9**:256–61.
- 5 Marciniak SJ, Lomas DA. What can naturally occurring mutations tell us about the pathogenesis of COPD?. *Thorax* 2009; **64**:359–64.

- 6 Larson RK, Barman ML. The familial occurrence of chronic obstructive pulmonary disease. *Ann Intern Med* 1965; 63:1001–8.
- 7 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.
- 8 Lebowitz MD, Knudson RJ, Burrows B. Family aggregation of pulmonary function measurements. *Am Rev Respir Dis* 1984; 129:8–11.
- 9 Gottlieb DJ, Wilk JB, Harmon M *et al.* Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med* 2001; 164:1655–9.
- 10 Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. *Am J Epidemiol* 1984; 120:518–30.
- 11 Palmer LJ, Knuiman MW, Divitini ML *et al.* Familial aggregation and heritability of adult lung function: results from the Busselton Health Study. *Eur Respir J* 2001; 17:696–702.
- 12 Wilk JB, Djousse L, Arnett DK *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* 2000; 19:81–94.
- 13 Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet* 2005; 366:941–51.
- 14 Silverman EK, Palmer LJ, Mosley JD *et al.* Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet* 2002; 70:1229–39.
- 15 Palmer LJ, Celedon JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum Mol Genet* 2003; 12:1199–210.
- 16 DeMeo DL, Celedon JC, Lange C *et al.* Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004; 170:1294–301.
- 17 DeMeo D, Mariani T, Lange C *et al.* The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Proc* 2006; 3:502.
- 18 Zhu G, Warren L, Aponte J *et al.* The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am J Respir Crit Care Med* 2007; 176:167–73.
- 19 Hall IP, Lomas DA. The genetics of obstructive lung disease: big is beautiful. *Thorax* 2010; 65:760–1.
- 20 Castaldi PJ, Cho MH, Cohn M *et al.* The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. *Hum Mol Genet* 2010; 19:526–34.
- 21 Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. *Am J Respir Crit Care Med* 2009; 180:618–31.
- 22 Obeidat M, Wain LV, Shrine N *et al.* A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. *PLoS ONE* 2011; 6:e19382.
- 23 Silverman EK, Vestbo J, Agusti A *et al.* Opportunities and challenges in the genetics of COPD 2010: an International COPD Genetics Conference report. *COPD* 2011; 8:121–35.
- 24 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; 2:e124.
- 25 Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004; 96:434–42.
- 26 McCarthy MI, Abecasis GR, Cardon LR *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; 9:356–69.
- 27 Soler Artigas M, Wain LV, Tobin MD. Genome-wide association studies in lung disease. *Thorax* 2011; 2011:19.
- 28 Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marcianti KD *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010; 42:45–52.
- 29 Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010; 42:36–44.
- 30 Altshuler DM, Gibbs RA, Peltonen L *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; 467:52–8.
- 31 Soler Artigas M, Loth DW, Wain LV *et al.* Genome-wide association and large scale follow-up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43:1082–90.
- 32 Wilk JB, Chen TH, Gottlieb DJ *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009; 5:e1000429.
- 33 Castaldi PJ, Cho MH, Litonjua AA *et al.* The association of genome-wide significant spirometric loci with COPD susceptibility. *Am J Respir Cell Mol Biol* 2011; 45:1147–53.
- 34 Cho MH, Boutaoui N, Klanderman BJ *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010; 42:200–2.
- 35 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; 5:e1000421.
- 36 Soler Artigas M, Wain LV, Repapi E *et al.* Effect of 5 Genetic Variants Associated with Lung Function on the Risk of COPD, and their Joint Effects on Lung Function. *Am J Respir Crit Care Med* 2011; 184:786–95.
- 37 Loos RJ, Lindgren CM, Li S *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; 40:768–75.
- 38 Lango Allen H, Estrada K, Lettre G *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; 467:832–8.
- 39 Manzke T, Guenther U, Ponimaskin EG *et al.* 5-HT4(a) receptors avert opioid-induced breathing depression without loss of analgesia. *Science* 2003; 301:226–9.
- 40 Hindorff LA, Sethupathy P, Junkins HA *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; 106:9362–7.
- 41 Englert JM, Hanford LE, Kaminski N *et al.* A role for the receptor for advanced glycation end products in idiopathic pulmonary fibrosis. *Am J Pathol* 2008; 172:583–91.
- 42 Celedon JC, Lange C, Raby BA *et al.* The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease

- (COPD). *Hum Mol Genet* 2004; 13:1649–56.
- 43 Matheson MC, Ellis JA, Raven J, Walters EH, Abramson MJ. Association of IL8, CXCR2 and TNF-alpha polymorphisms and airway disease. *J Hum Genet* 2006; 51:196–203.
 - 44 Sakao S, Tatsumi K, Igari H, Shino Y, Shirasawa H, Kuriyama T. Association of tumor necrosis factor alpha gene promoter polymorphism with the presence of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2001; 163:420–2.
 - 45 van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Nolte IM, Boezen HM. Decorin and TGF-beta1 polymorphisms and development of COPD in a general population. *Respir Res* 2006; 7:89.
 - 46 Yucesoy B, Kurzius-Spencer M, Johnson VJ *et al.* Association of cytokine gene polymorphisms with rate of decline in lung function. *J Occup Environ Med* 2008; 50:642–8.
 - 47 Landi MT, Chatterjee N, Yu K *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009; 85:679–91.
 - 48 Kohansal R, Martinez-Camblor P, Agusti A, Buist AS, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* 2009; 180:3–10.
 - 49 Morin C, Sirois M, Echave V, Gomes MM, Rousseau E. Functional effects of 20-HETE on human bronchi: hyperpolarization and relaxation due to BKCa channel activation. *Am J Physiol Lung Cell Mol Physiol* 2007; 293:L1037–44.
 - 50 Vang A, Mazer J, Casserly B, Choudhary G. Activation of endothelial BKCa channels causes pulmonary vasodilation. *Vascul Pharmacol* 2010; 53:122–9.
 - 51 Moriguchi T, Haraguchi K, Ueda N, Okada M, Furuya T, Akiyama T. DREG, a developmentally regulated G protein-coupled receptor containing two conserved proteolytic cleavage sites. *Genes Cells* 2004; 9:549–60.
 - 52 Jongepier H, Boezen HM, Dijkstra A *et al.* Polymorphisms of the ADAM33 gene are associated with accelerated lung function decline in asthma. *Clin Exp Allergy* 2004; 34:757–60.
 - 53 van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Schouten JP, Boezen HM. A disintegrin and metalloprotease 33 polymorphisms and lung function decline in the general population. *Am J Respir Crit Care Med* 2005; 172:329–33.
 - 54 Van Eerdewegh P, Little RD, Dupuis J *et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 2002; 418:426–30.
 - 55 Chazaud C, Dolle P, Rossant J, Mollard R. Retinoic acid signaling regulates murine bronchial tubule formation. *Mech Dev* 2003; 120:691–700.
 - 56 Massaro GD, Massaro D, Chan WY *et al.* Retinoic acid receptor-beta: an endogenous inhibitor of the perinatal formation of pulmonary alveoli. *Physiol Genomics* 2000; 4:51–7.
 - 57 Yang J, Liu X, Yue G, Adamian M, Bulgakov O, Li T. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J Cell Biol* 2002; 159:431–40.
 - 58 Makrygiannakis D, Hermansson M, Ulfgren AK *et al.* Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis* 2008; 67:1488–92.
 - 59 Bourbeau J, Johnson M. New and controversial therapies for chronic obstructive pulmonary disease. *Proc* 2009; 6:553–4.
 - 60 Ito K, Ito M, Elliott WM *et al.* Decreased histone deacetylase activity in chronic obstructive pulmonary disease. *N Engl J Med* 2005; 352:1967–76.
 - 61 Vega RB, Matsuda K, Oh J *et al.* Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. *Cell* 2004; 119:555–66.
 - 62 Nagase H, Woessner JF Jr. Matrix metalloproteinases. *J Biol Chem* 1999; 274:21491–4.
 - 63 Park EJ, Roh J, Kim Y, Park K, Kim DS, Yu SD. PM 2.5 collected in a residential area induced Th1-type inflammatory responses with oxidative stress in mice. *Environ Res* 2011; 111:348–55.
 - 64 Joos L, He JQ, Shepherdson MB *et al.* The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet* 2002; 11:569–76.
 - 65 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.
 - 66 Minematsu N, Nakamura H, Tateno H, Nakajima T, Yamaguchi K. Genetic polymorphism in matrix metalloproteinase-9 and pulmonary emphysema. *Biochem Biophys Res Commun* 2001; 289:116–19.
 - 67 Zhou M, Huang SG, Wan HY, Li B, Deng WW, Li M. Genetic polymorphism in matrix metalloproteinase-9 and the susceptibility to chronic obstructive pulmonary disease in Han population of south China. *Chin Med J (Engl)* 2004; 117:1481–4.
 - 68 Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; 32:1–22.
 - 69 Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008; 5:e177.
 - 70 Liu JZ, Tozzi F, Waterworth DM *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010; 42:436–40.
 - 71 Thorgeirsson TE, Gudbjartsson DF, Surakka I *et al.* Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet* 2010; 42:448–53.
 - 72 Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010; 42:441–7.
 - 73 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.
 - 74 Wain LV, Armour JA, Tobin MD. Genomic copy number variation, human health, and disease. *Lancet* 2009; 374:340–50.
 - 75 Wacholder S, Hartge P, Prentice R *et al.* Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010; 362:986–93.
 - 76 Talmud PJ, Hingorani AD, Cooper JA *et al.* Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 2010; 340:b4838.
 - 77 Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. *Thorax* 2006; 61:935–9.