3

Genetic susceptibility

Introduction

Key points

- While some diseases, such as cystic fibrosis, are almost wholly genetic, genetic factors play a role in making us more or less vulnerable to the whole spectrum of respiratory disease.
- The search for genetic links to disease is two-pronged: linkage studies begin with a disease and look for genes associated with it, while association studies begin with polymorphisms or mutations, and look for associations with disease.
- Several genes have been identified that influence susceptibility to both asthma and COPD, reflecting similarities in disease presentation between the two conditions.
- Nondisease traits that influence respiratory health, such as tendency to nicotine addiction, and lung function, have also been related to genetic variations.

Respiratory diseases occur as a result of interactions between genotype and environment. Environmental influences include allergens, irritants, smoking, environmental tobacco smoke (ETS), diet, nutrients, drugs, infections and injuries. When a single gene has a very high impact on the development of a disease, this disease is called a "monogenic disease" (figure 1). Examples of such diseases are cystic fibrosis (CF) and α_1-antitrypsin deficiency, which are inherited in a classical "Mendelian" fashion, with recessive or dominant forms of the gene in question being passed from generation to generation. Other diseases are triggered mainly by major environmental exposures; examples include carbon monoxide poisoning, acute lung injury and acute respiratory distress syndrome (due to severe pneumonia or major trauma). However, in the most common lung diseases, such as asthma, chronic obstructive pulmonary disease (COPD), pulmonary fibrosis and sarcoidosis, both the genotype and the environment play major roles in disease susceptibility: these diseases are called "complex diseases".

Humans have 23 pairs of chromosomes (one pair of sex chromosomes and 22 pairs of autosomes). This provides every human with two versions of each gene: one on the maternally inherited chromosome and one on the paternally inherited chromosome. The human genome includes 3.3 billion base (nucleotide) pairs (the "building blocks" of DNA) and more than 25 000 genes, which code for proteins that build cells and tissues, and enzymes that catalyse biochemical

"The DNA sequence is more than 99% identical between different individuals; but this still leaves scope for more than 10 million potential differences or variations between the genomes of two humans"

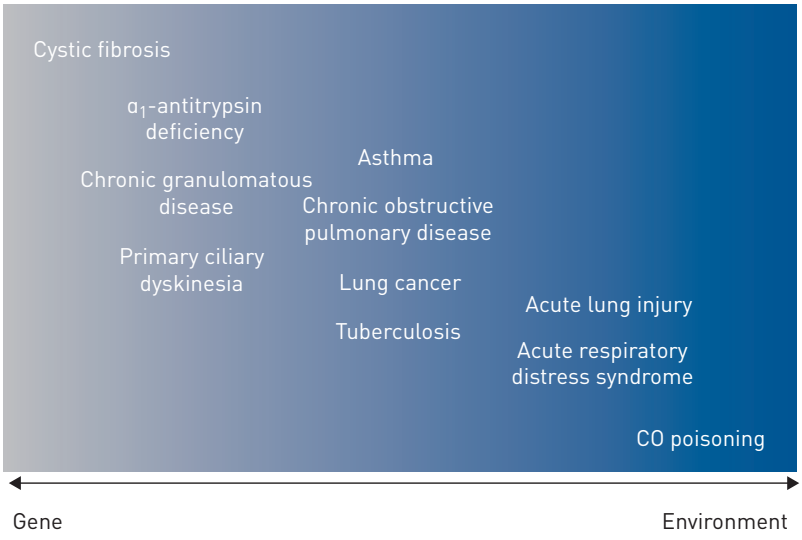

Figure 1 – Most common respiratory diseases are complex diseases. They arise as a consequence of interactions between an individual genotype and environmental exposures. CO: carbon monoxide.

reactions within cells. The DNA sequence is more than 99% identical between different individuals; but this still leaves scope for more than 10 million potential differences or variations between the genomes of two humans. Variations in gene structure that occur frequently in a population (1% or more of people) are called polymorphisms, whereas genetic variations that occur infrequently (less than 1%) are called mutations. There are several forms of polymorphism, of which single-nucleotide polymorphisms (SNPs) are by far the most common. SNPs constitute a single base-pair change in the DNA sequence at a particular point, relative to the common or "wild type" sequence. SNPs in parts of genes that code for proteins can lead to a change in the amino acid sequence of the protein, affecting its structure and/or function.

There are several methods to study the genetic factors that contribute to the development of an individual's specific characteristics (referred to as phenotypes; for example, height or lung function) or complex diseases, such as asthma and COPD. Linkage studies are performed in families: these are based on the tendency of genetic loci (the site on a chromosome at which one or several genes for a particular disease or trait are located) or alleles that are physically close to one another on a chromosome to be inherited together (this is known as genetic linkage). Once a genetic locus for the phenotype or disease of interest has been identified through linkage analysis, positional cloning is performed to further delineate the susceptibility gene(s). For many years, genetic linkage combined

Since there are some similarities between the disease phenotypes and patho-physiological pathways of asthma and COPD, several susceptibility genes are suspected to be common to both diseases

with positional cloning has offered a rational way of discovering gene mutations that cause monogenic diseases, such as CF. These searches have led to the discovery of rare mutations (present in less than 1% of the population) that alter the amino acid sequence of a protein and increase the risk of disease enormously (very high effect size) (figure 2).

In contrast, association studies begin with the polymorphism or mutation rather than with the disease. They are typically based on a case–control design (*i.e.* "cases" – people with the disease – are compared with healthy control subjects) in which SNPs are tested for association with a specific phenotype or disease. In single-candidate gene association studies, only one or a few SNPs near or in the gene under study are investigated for association with the disease of interest, based on an *a priori* hypothesis concerning the possible function and role of the particular gene. In genome-wide association studies, hundreds of thousands of SNPs across the entire human genome are genotyped and tested for association with the phenotype or disease of interest in hundreds or thousands of individuals. Without an *a priori* hypothesis, genome-wide association studies identify common genetic variants (which are present in more than 5% of the population) that confer a small risk of disease (small effect size, typically with odds ratios of 1.1 to 2.0).

This chapter on genetic susceptibility to respiratory diseases is neither exhaustive nor complete, but is intended as an introduction to the exponentially growing field of genetics and genomics in respiratory medicine and science.

Monogenic diseases

Monogenic diseases (table 1) are rare diseases attributable to genetic variants with large effects on disease status. Because

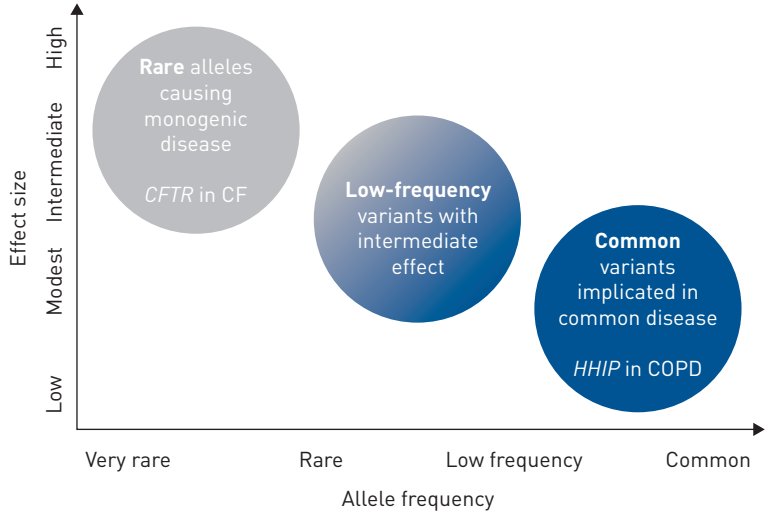

Figure 2 – Allele frequency *versus* disease risk. *CFTR*: cystic fibrosis conductance regulators; CF: cystic fibrosis; *HHIP*: hedgehog-interacting protein; COPD: chronic obstructive pulmonary disease.

of the high penetrance of such variants, the disease is typically inherited in a classical Mendelian fashion (*e.g*. dominant or recessive). The best-known monogenic respiratory diseases are CF and α_1-antitrypsin deficiency, but hundreds of rare monogenic diseases affecting the respiratory system have been described. We refer the interested reader to the Online Mendelian Inheritance in Man (OMIM) website, which is a comprehensive, authoritative and continuously updated compendium of human genes and genetic phenotypes (see Further reading).

Cystic fibrosis

CF is an autosomal recessive genetic disorder (*i.e*. both inherited copies of the gene need to be mutated in order for disease to result; or to put it another way, one healthy copy of the gene is enough to prevent CF), caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene on chromosome 7. The CFTR protein is an ion channel that regulates transport of chloride ions (Cl^-) in epithelial cells in the airways, as well as in the pancreas, liver, intestine and skin. More than 1000 mutations of the *CFTR* gene have been described. In Europe, the most common is the ΔF508 mutation (a deletion of three DNA bases). The resulting CFTR protein has a missing amino acid (phenylalanine) in position 508. One in 25 people of European descent carries one mutant allele of CFTR, and one person in 2000–3000 is affected by CF. The various *CFTR* mutations cause different CFTR protein defects, which impair transport of chloride and sodium across epithelial surfaces, leading to thick viscous secretions (*e.g*. mucus or phlegm).

CFTR modulators and potentiators are drugs that aim to correct the underlying defect that leads to CF by modifying the function of the CFTR protein. Since the therapeutic effects of CFTR modulators are based on individual protein defects, knowledge of the genotype of both alleles of the *CFTR* gene is necessary for appropriate patient selection. As an example, ivacaftor, which was approved for use by the US Food and Drug Administration in January 2012, targets the specific *CFTR* mutation G551D (in which glycine in position 551 is substituted with aspartic acid), improves lung function and reduces respiratory symptoms and pulmonary exacerbations in patients with CF who have at least one G551D *CFTR* mutation.

α_1-antitrypsin deficiency

α_1-antitrypsin is a protease inhibitor, produced mainly in the liver, which protects the lungs against proteolytic damage by the enzyme neutrophil elastase. α_1-antitrypsin is encoded by the *SERPINA1* gene (also known as *PI*). Like CF, α_1-antitrypsin deficiency is an autosomal recessive inherited disorder affecting 1 in 2000–5000 persons in Europe. It increases the risk of liver disease, COPD and emphysema. Those with two copies of the most severe "Z" mutation (*PI* ZZ genotype) have very low serum protein levels of α_1-antitrypsin. Cigarette smoking greatly increases the risk of COPD in α_1-antitrypsin-deficient patients, leading to severe, early-onset emphysema due to destruction of alveolar septa in the lung as a consequence of the protease–antiprotease imbalance.

Primary ciliary dyskinesia

Primary ciliary dyskinesia (PCD), or immotile cilia syndrome, is a genetically heterogeneous autosomal recessive disorder, caused by loss of function of different parts of the cilia, which line the epithelial cells of the airway mucosa and are responsible for clearing secretions and foreign material. Patients with PCD suffer from recurrent upper and lower respiratory tract infections, often leading to bronchiectasis,

Gene/locus	Gene name	Chromosomal location#	Gene product: protein function	Disease
CFTR	CF transmembrane conductance regulator	7q31.2	Ion channel: chloride transport	CF
SERPINEA1	α_1-antitrypsin	14q32.13	Serine protease inhibitor	α_1-antitrypsin deficiency (COPD, emphysema, liver disease)
DNAI1	Dynein, axonemal, intermediate chain 1	9p13.3	Dynein arm: ciliary function	CILD1, with or without situs inversus (Kartagener syndrome)
CYBB	p91-phox (phagocyte oxidase): beta subunit of cytochrome b, component of the phagocyte NADPH oxidase complex	Xp11.4	Killing of microbes in phagocytes by generation of reactive oxygen species	CGD, X-linked
CYBA	p22-phox (phagocyte oxidase): alpha subunit of cytochrome b, component of the phagocyte NADPH oxidase complex	16q24.3	Killing of microbes in phagocytes by generation of reactive oxygen species	CGD, autosomal recessive
SFTPC	Surfactant, pulmonary-associated protein C	8p21.3	Surfactant proteins are essential for lung function, preventing lung collapse by lowering surface tension	Respiratory distress syndrome of prematurity
SFTPB	Surfactant, pulmonary-associated protein B	2p11.2	Surfactant proteins are essential for lung function, preventing lung collapse by lowering surface tension	Respiratory distress syndrome of prematurity

Table 1 – Monogenic respiratory diseases (inherited in a Mendelian fashion). Only seven out of more than 100 known monogenic respiratory diseases are presented as illustration. #: p refers to the short arm of the chromosome. q refers to the long arm of the chromosome. The location numbers after p and q reflect the relative distance to the centromeres of the chromosomes (numbering by convention). CF: cystic fibrosis; COPD: chronic obstructive pulmonary disease; CILD1: ciliary dyskinesia, primary 1; NADPH: nicotinamide adenine dinucleotide phosphate; CGD: chronic granulomatous disease.

an abnormal widening of the airways. About half of people with PCD have Kartagener syndrome, in which PCD is combined with situs inversus (a condition in which the position of the major organs is a mirror image of the normal arrangement).

Other monogenic diseases encompass diseases caused by mutations in surfactant proteins, which are crucial in decreasing tension forces during breathing. Dysfunction of surfactant caused by mutations in surfactant protein genes leads to respiratory distress syndrome of prematurity. Mutations in cytochrome b, an enzyme involved in killing microbes in phagocytic cells, predispose individuals to recurrent respiratory infections (see the section on chronic granulomatous diseases later in this chapter).

Complex diseases

Asthma

Using the candidate gene approach (discussed earlier in this chapter), many genes have been associated with asthma or asthma-related traits such as allergy and high concentrations of immunoglobulin E (IgE) in serum (table 2). Not all of these suspected asthma susceptibility genes have been replicated in multiple independent studies. One group of (allergic) asthma susceptibility genes is involved in innate immunity responses, encompassing pattern-recognition receptors, immunoregulatory cytokines and molecules involved in antigen presentation. A second group of asthma susceptibility genes are key players in T-helper type 2 (Th2)-cell differentiation and Th2-cell effector function. Th2 cells are T-lymphocytes that drive the production of allergic immunoglobulins (IgE) and the chronic airway inflammation in (allergic) asthma.

Linkage studies in families have discovered several novel asthma susceptibility genes that are expressed in epithelial cells and/or smooth muscle cells in the airways (table 2). Although the functional role of these asthma susceptibility genes is not yet fully understood, they are thought to be involved in maintaining the integrity of the epithelial barrier, airway remodelling and bronchial hyperresponsiveness. These asthma susceptibility genes indicate the importance of altered communication between the epithelium and the underlying smooth muscle cells in the pathogenesis of asthma.

The first genome-wide association study of asthma showed that multiple markers at chromosomal location 17q21, encompassing genetic variants of *ORMDL3* and *GSDMB*, were strongly associated with childhood asthma. The association of the *ORMDL3/GSDMB* locus with early-onset asthma is further increased in children exposed to environmental tobacco smoke, implicating an interaction between gene and environment. In infancy, passive smoking does indeed significantly increase the risk of developing asthma. A large-scale genome-wide association study of asthma performed by the European GABRIEL consortium revealed that some genes involved in communication of epithelial damage to the adaptive immune system are susceptibility genes for asthma (table 2). This genome-wide association study of asthma confirmed the role of antigen presentation and of the Th2-cytokine gene *IL13* (interleukin 13) in the pathogenesis of asthma. Many of these asthma susceptibility genes have been confirmed by the American EVE consortium. Finally, several loci have been linked to increased serum total IgE levels in genome-wide association studies. These are *IL13*, *IL4R*, *STAT6* (signal transducer and activator of transcription 6), *FCER1A* (high-affinity Fc receptor for IgE) and *HLA-DRB1* (a human leukocyte antigen).

Gene/locus	Chromosomal location#	Gene name	Mechanism
Candidate gene association studies			
TLR2	4q31.3	Toll-like receptor 2	Pathogen recognition/ innate immunity
CD14	5q31.3	Cluster of differentiation 14: monocyte antigen	LPS signalling
IL10	1q32.1	Interleukin-10	Anti-inflammatory/ regulatory T-cells
TGFB	19q13.2	Transforming growth factor-ß	Anti-inflammatory/ airway remodelling
HLA-DR	6p21.32	Human leukocyte antigens	Antigen presentation
HLA-DQ	6p21.32	Human leukocyte antigens	Antigen presentation
HLA-DP	6p21.32	Human leukocyte antigens	Antigen presentation
IL4	5q31.1	Interleukin-4	Th2 responses/IgE production
IL13	5q31.1	Interleukin-13	Mucus production/IgE production
IL4R	16p12.1	Interleukin-4 receptor	Th2 responses/IgE production
STAT6	12q13.3	Signal transducer and activator of transcription 6	Transcription factor (Th2 responses)
Genome-wide association studies			
ORMDL3	17q21	Orosomucoid like 3	Unknown
IL2R	10p15.1	Interleukin-2 receptor	T-cell proliferation Th1 responses
IL18R1	2q12.1	Interleukin-18 receptor 1	T-cell proliferation Th1 responses
IL13	5q31.1	Interleukin-13	Mucus production/IgE production
IL33	9p24.1	Interleukin-33	Innate immunity/danger signal
Linkage studies and positional cloning			
ADAM33	20p13	A disintegrin and metalloproteinase	Airway remodelling/ bronchial hyperresponsiveness
DPP10	2q14.1	Dipeptidyl peptidase 10	
GPRA		G-protein coupled receptor for asthma susceptibility	Unknown

Table 2 – Genetic susceptibility to asthma. This is a partial list of selected genes intended as an illustrative example of genetic susceptibility to asthma. LPS: lipolysaccharide; Th: T-helper type 2; IgE: immunoglobulin E. #: p refers to the short arm of the chromosome. q refers to the long arm of the chromosome. The location numbers after p and q reflect the relative distance to the centromeres of the chromosomes (numbering by convention).

Chronic obstructive pulmonary disease and emphysema

Since only about 20% of smokers develop COPD, genetic risk factors are thought to be involved in the pathogenesis of the disease. The best known genetic risk factor

for emphysema is α_1-antitrypsin deficiency, implicating an imbalance of protease (neutrophil elastase) and antiprotease (α_1-antitrypsin) in the pathogenesis of the disease. Two meta-analyses of candidate gene studies in COPD concluded that only a few other COPD susceptibility genes have been firmly identified. These include *TNFA* (tumour necrosis factor-α), *TGFB1* (transforming growth factor-β_1), *GSTP1* and *GSTM1* (glutathione S-transferases P1 and M1), and *SOD3* (superoxide dismutase 3).

Genome-wide association studies in COPD have identified three major susceptibility loci: the *FAM13A* locus on chromosome 4q22, the locus near *HHIP* on chromosome 4 and the *CHRNA3/CHRNA5* locus on chromosome 15 (see the nicotine addiction and smoking section later in this chapter). Recently, several of the genetic determinants of lung function, encompassing genes involved in lung development and growth, such as *HHIP* (hedgehog-interacting protein), have been confirmed as genetic risk factors for COPD (see the lung function section later in this chapter).

Since there are some similarities between the disease phenotypes and pathophysiological pathways of asthma and COPD, several susceptibility genes are suspected to be common to both diseases, whereas other susceptibility genes will be specific to asthma or COPD. Both asthma and COPD are very heterogeneous diseases with multiple distinct phenotypes, suggesting that the degree of overlap between the genetic susceptibilities will depend on the asthma or COPD phenotypes examined. Using the candidate gene approach, several genes, such as *TNFA*, *TGFB1*, *MMP12* (matrix metalloproteinase 12) and *ADAM33*, have been implicated as susceptibility genes for both asthma and COPD. Common pathogenetic pathways in airway inflammation and remodelling might explain this common genetic susceptibility. Some genes, such as *IL13*, have been specifically associated with allergy and allergic asthma, but not with COPD. In contrast, *SERPINE* (serine protease inhibitors) genes such as that for α_1-antitrypsin have been specifically implicated in the pathogenesis of emphysema, an important phenotype of COPD.

Pulmonary fibrosis

Although the cause of pulmonary fibrosis is unknown (*i.e.* it is idiopathic), it is estimated that 0.5–2.0% of cases of idiopathic pulmonary fibrosis (IPF) are familial. Several mutations and polymorphisms in different genes have been shown to increase susceptibility to IPF: mutations in *TERT* (the telomerase reverse transcriptase gene), the catalytic subunit of the telomerase enzyme; mutations in *TERC* (the telomerase RNA component gene); and a promoter mutation in the *MUC5B* gene, which codes for the mucin B protein. A polymorphism in the *SFTPA1* gene encoding pulmonary surfactant protein A1 influences susceptibility to IPF in nonsmokers, and a mutation in the *SFTPA2* gene encoding pulmonary surfactant protein A2 can cause IPF.

Sarcoidosis

Sarcoidosis is suspected to be caused by a combination of environmental exposure to a still-unknown agent (*e.g.* a microorganism or inorganic material) and genetic susceptibility. Class II molecules of the major histocompatibility complex (MHC), also called human leukocyte antigens (HLA), are cell surface proteins that present processed foreign antigens to T-lymphocytes. These T-lymphocytes are then stimulated to become effector cells of adaptive immune responses. There is a high degree of polymorphism in class II MHC genetic loci.

Variation in the *HLA-DRB1* gene on chromosome 6p21.3, affecting antigen presentation to T-lymphocytes, is a major contributor to susceptibility to sarcoidosis (the locus is called susceptibility locus for sarcoidosis 1 (SS1)). The second susceptibility locus for

sarcoidosis, SS2, is on chromosome 6p21.32 and is associated with variation in the *BTNL2* (Butyrophilin-like protein 2) gene, which may regulate T-cell activation. The strongest association signal from a genome-wide association study for sarcoidosis mapped to the *ANXA11* gene, belonging to the annexin family, on chromosome 10q22.3.

Genetic variation of another MHC class II molecule, *HLA-DPB1*, has been shown to confer susceptibility to sarcoidosis and chronic beryllium disease, a hypersensitivity disorder of the lung caused by exposure to beryllium (used in diverse industries such as aerospace). Both sarcoidosis and chronic beryllium disease are characterised by chronic adaptive immune responses, leading to the formation of granulomas in the lung and lymph nodes.

Respiratory infections and pneumonia

Genetic factors can increase the risk of respiratory infections, including acute bronchitis and pneumonia. Most often, genetic polymorphisms underlie vulnerability to recurrent infections, but in rare cases monogenic defects are responsible (table 1). Repeated respiratory infections can be precipitated by structural defects of the lungs (*e.g.* bronchiectasis due to CF or PCD) or by genetic defects in the immune system. This defence system can be divided into the innate immune system, which recognises broadly conserved, generic structures of microbes *via* cell surface receptors (called pattern-recognition receptors), and the adaptive immune system, which recognises specific parts of microbial structures *via* very specific receptors on T-cells (which produce cytokines) and B-cells (which produce immunoglobulins (Ig)). These immunoglobulins, also called antibodies, are present in serum (*e.g.* IgM and IgG) and in sputum (IgA).

A disorder characterised by impaired immune responses towards infectious agents is called an 'immunodeficiency'. This can be either inherited or acquired (*e.g.* acquired immune deficiency syndrome (AIDS) caused by the human immunodeficiency virus (HIV)). Numerous genetic defects can impair the host's immune response to infection, leading to inherited immunodeficiencies. Genetic defects in innate immunity lead to several groups of immunodeficiencies. Firstly, chronic granulomatous diseases (table 1) are caused by immunodeficiencies due to impaired intracellular killing of microbes within phagocyte cells (neutrophils and macrophages). Secondly, defective recognition of microbes caused by genetic polymorphisms or mutations in pattern-recognition receptors can increase the risk of infection by particular micro-organisms. Deficiency of Toll-like receptor 3 (TLR3), which recognises double-stranded RNA, confers susceptibility to viral infections (*e.g.* herpesvirus), whereas deficiency of TLR5, which recognises

flagellin, increases the risk of *Legionella* infections (*e.g.* pneumonia due to *Legionella* (Legionnaires' disease)). Lastly, the common deficiency of mannose-binding lectin, which activates complement, increases the risk of infections with bacteria and fungi.

Genetic defects in adaptive immunity can affect the development and function of B-cells, leading to decreased levels of immunoglobulins (*e.g.* IgA deficiency), or of T-cells, impairing cellular immunity and predisposing to opportunistic infections. The most severe cases of inherited immunodeficiency are already apparent in infancy and are caused by impairment of both B- and T-cell immunity (*e.g.* X-linked severe combined immunodeficiency syndrome (SCID)).

Tuberculosis

One-third of the global population is latently infected with *Mycobacterium tuberculosis*. Exposure to *M. tuberculosis* can lead to asymptomatic 'latent' infection or to overt clinical tuberculosis. Why only 10% of individuals infected with *M. tuberculosis* develop active disease is not known, but variation in many genes has been associated with susceptibility to, or resistance against, *M. tuberculosis* (table 3). These genetic variants encompass a spectrum from causal susceptibility in rare cases, to very mild predisposition in the general population.

Lung cancer

Smoking is a major risk factor for lung cancer, and several studies have shown that a first-degree family history of lung cancer confers an approximately two-fold increased risk

Gene/locus	Location#	Name of gene or locus	Mechanism
Susceptibility to tuberculosis			
CISH	3p21.2	Cytokine-inducible SH2-containing protein	Adaptive immunity
CD209	19p13.2	DC-SIGN: membrane lectin receptor of dendritic cells	Pathogen recognition/ cell adhesion
MCP1	17q12	Monocyte chemotactic protein 1 or CCL2	Chemo-attractant
VDR	12q13.11	Vitamin D receptor	Innate and adaptive immunity
MTBS1	2q35	*M. tuberculosis* susceptibility locus 1	Unknown
MTBS2	8q12-q13	*M. tuberculosis* susceptibility locus 2	Unknown
MTBS3	20q13.31-q33	*M. tuberculosis* susceptibility locus 3	Unknown
Protection against tuberculosis			
TIRAP	11q24.2	TIR domain-containing adaptive protein	TLR4 signalling
IFNG	12q15	Interferon-γ	Th1 adaptive immunity
IFNGR1	6q23.3	Interferon-γ receptor 1	Th1 adaptive immunity

Table 3 – Genetic susceptibility to, or protection against, *Mycobacterium tuberculosis*. This is a partial list of selected genes and loci intended as an illustrative example of genetic susceptibility to tuberculosis. #: p refers to the short arm of the chromosome. q refers to the long arm of the chromosome. DC-SIGN: dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin; CCL2: chemokine ligand 2; TIR: Toll/IL1R; TLR4: Toll-like receptor 4; Th1: T-helper type 1.

of lung cancer, implicating a familial aggregation of lung cancer. Genome-wide association studies have identified a region on chromosome 15 (15q25.1), containing the nicotinic acetylcholine receptor subunit genes *CHRNA3* and *CHRNA5*, that is associated with nicotine addiction (*i.e.* number of cigarettes smoked per day) and lung cancer. Whether genetic variation in the nicotinic acetylcholine receptor increases the risk of lung cancer only indirectly *via* nicotine addiction or whether it also influences the lung epithelium directly in pulmonary carcinogenesis, is currently the subject of intense investigation (figure 3).

> *"The locus encompassing the* CHRNA3 *and* CHRNA5 *(nicotinic acetylcholine receptor) gene cluster on chromosome 15q24–25 has been associated with nicotine dependence"*

Both in small cell lung cancer and in nonsmall cell lung cancer, numerous somatic mutations and chromosomal aberrations have been described within the tumour cells. However, a detailed description of these is beyond the scope of this chapter. We refer the interested reader to one of the excellent reviews that are available on the genomics of lung cancer (see Further reading).

Pulmonary embolism

Most pulmonary embolisms arise from blood clots in the deep veins (*i.e.* deep vein thrombosis) of the legs. Risk factors for deep vein thrombosis and acute pulmonary embolism include immobilisation, surgery, stroke, malignancy, obesity and pregnancy, but also genetic susceptibility. If the former risk factors are absent (*i.e.* unprovoked venous thromboembolism), or if there is a positive family history of deep vein thrombosis or pulmonary embolism, then an inherited thrombophilia, or hypercoagulable state, should be suspected.

The most common inherited hypercoagulable state is due to a mutation in the coagulation factor V gene (called the factor V Leiden mutation), which causes resistance to the anticoagulation factor, activated protein C. Heterozygosity (one copy of the mutated gene) for the factor V Leiden mutation is present in approximately 5% of a Caucasian population, and

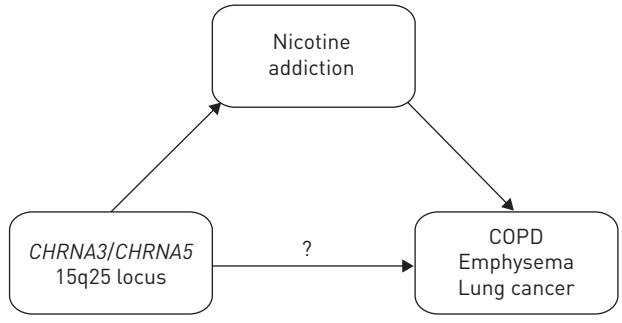

Figure 3 – Direct *versus* indirect effects of genetic variation in the nicotinic acetylcholine receptor cluster (including the *CHRNA3* and *CHRNA5* genes) on susceptibility to COPD, emphysema and lung cancer.

homozygosity (two copies of the mutated gene) in 1%. Homozygotes for the factor V Leiden mutation have a more than two-fold increased lifetime risk of developing deep vein thrombosis, with or without pulmonary embolism. Other inherited thrombophilias include a mutation in the prothrombin gene (coagulation factor II), antithrombin (ATIII) deficiency, protein C deficiency and protein S deficiency. Deficiencies of these anticoagulation factors increase the lifetime risk of venous thromboembolism seven- to eight-fold. Use of oral contraceptives (mainly third-generation oral contraceptives) is associated with an increased risk of venous thromboembolism, especially in heterozygote and homozygote carriers of the factor V Leiden mutation, implicating a gene–environment interaction.

Complex traits

Nicotine addiction and smoking

The locus encompassing the *CHRNA3* and *CHRNA5* (nicotinic acetylcholine receptor) gene cluster on chromosome 15q24–25 has been associated with nicotine dependence, as measured by the number of cigarettes smoked per day. This gene cluster has also been associated with smoking-related diseases such as peripheral arterial disease, lung cancer, COPD and emphysema. Whether the nicotinic acetylcholine receptor gene cluster confers an increased risk of developing smoking-related diseases such as COPD, emphysema and lung cancer, in addition to its major impact on smoking behaviour, is a matter of debate and intense investigation (figure 3).

Lung function

Asthma and COPD are classed as obstructive airway diseases. The ratio of a person's forced expiratory volume in 1 second (FEV_1) to forced vital capacity (FVC) (*i.e.* the FEV_1/FVC ratio) is an indicator of airflow obstruction; a reduced FEV_1/FVC ratio is the primary criterion for defining airway obstruction. The first genome-wide association study of pulmonary function, performed in the Framingham Heart Study in the USA, identified SNPs near the *HHIP* gene on chromosome 4q31 that were associated with the FEV_1/FVC ratio. Two large genome-wide association studies confirmed the *HHIP* locus and identified multiple novel loci associated with the FEV_1/FVC ratio (table 4). Thanks to collaboration between two large scientific research consortia, CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) and SpiroMeta, 16 additional genetic loci have been shown to be associated with lung function, including *RARB* (retinoic acid receptor B). Since several identified genes (*e.g.* *RARB* and *HHIP*) play crucial roles in lung development by regulating branching morphogenesis (the development of the bronchial 'tree') during fetal life, the results of these genome-wide association studies suggest that genetic variations associated with lung development and growth might be important genetic determinants of lung function in childhood and adulthood, both in healthy subjects and in patients with airway disease (asthma and COPD).

Conclusions and future prospects

The monogenic diseases CF and α_1-antrypsin deficiency are inherited in a recessive Mendelian fashion (*i.e.* mutations in both alleles are required for the disease to be present). However, the term 'monogenic' is an oversimplification, since the causal gene interacts both with other genes and with environmental exposures in the course of the

disease. Indeed, several modifier genes influence the severity of the disease in CF, implicating gene–gene interactions in its development. Active and passive smoking have deleterious effects in subjects with α_1-antitrypsin deficiency, implicating important gene–environment interactions in the pathogenesis of panlobular emphysema.

The most common chronic respiratory diseases – asthma and COPD – are complex airway diseases that result from interaction between multiple environmental exposures and many genetic risk factors. Thanks to the development of novel, powerful tools for genetic studies, many genetic loci have been discovered that are associated with asthma, allergy, smoking behaviour, lung function and COPD. Despite the impressive advances in the genetics of asthma and COPD in the past decade, major challenges remain. Firstly, a large proportion of the genetic variance in disease risk remains unexplained. Most genetic variants identified so far by genome-wide association studies confer relatively small increments in risk, and explain only a small proportion of familial clustering. The remaining, 'missing' heritability can be attributed to additional genetic variation as yet unidentified, including structural variation (*e.g.* copy number variation of genes) and rare sequence variation. Secondly, the biological pathways and molecular mechanisms involved in the pathogenesis of chronic airway disease need to be elucidated in order to translate these new genetic insights into better strategies for prevention and treatment.

Gene	Gene name	Gene function
HHIP	Hedgehog-interacting protein	Lung development
GPR126	G-protein-coupled receptor 126	Unknown
ADAM19	A disintegrin and metalloproteinase 19	Cell migration and adhesion, cell-matrix interactions
AGER	Advanced glycation end products receptor	Receptor for danger signals, pro-inflammatory gene activation
FAM13A	Family with sequence similarity 13, member A	Signal transduction
GSTCD	Glutathione S-transferase, C-terminal domain containing	Detoxification
HTR4	5-hydroxytryptamine receptor-4	Receptor for serotonin, modulates release of neurotransmitters
PTCH1	Patched 1	Receptor for HHIP, lung development
MMP15	Matrix metalloproteinase 15	Breakdown of extracellular matrix
TGFB2	Transforming growth factor-ß2	Embryonic development
HDAC4	Histone deacetylase 4	Transcriptional regulation, cell cycle progression and development
RARB	Retinoic acid receptor, beta	Transcriptional regulation, limits cell growth

Table 4 – Genes associated with lung function.

Current and future applications of genetic testing in respiratory medicine encompass screening (*e.g.* newborn screening for CF), antenatal diagnosis, early diagnosis and prediction of disease risk (*e.g.* risk of recurrent venous thromboembolism according to underlying inherited thrombophilia). Pharmacogenetic and pharmacogenomic applications will improve our ability to use drugs more effectively and with less risk (*e.g.* optimising the dosing of the anticoagulant warfarin according to the genetic constitution of the patient). Finally, this genetic revolution will lead to the discovery of novel causal pathways, guiding mechanistic research in respiratory diseases and revealing new therapeutic targets.

Further reading

General

- Attia J, Ioannidis JP, Thakkinstian A, *et al*. How to use an article about genetic association. A: Background concepts. *JAMA* 2009; 301: 74–81.
- Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010; 363: 166–176.
- Manolio TA, Collins FS, Cox NJ, *et al*. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747–752.
- McCarthy M. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010; 363: 2339–2350.
- McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Online Mendelian Inheritance in Man. www.ncbi.nlm.nih.gov/omim
- National Human Genome Research Institute. A catalog of published genome-wide association studies. www.genome.gov/GWAstudies
- Nature ENCODE. The encyclopedia of DNA elements. www.nature.com/encode
- Ober C, Butte AJ, Elias JA. Getting from genes to function in lung disease. A national Heart, Lung, and Blood Institue workshop report. *Am J Respir Crit Care Med* 2010; 182: 732–737.
- Soler Artigas M, Wain LV, Tobin MD. Genome-wide association studies in lung disease. *Thorax* 2012; 67: 271–273.
- Todd JL, Goldstein DB, Ge D, *et al*. The state of genome-wide association studies in pulmonary disease. A new perspective. *Am J Respir Crit Care Med* 2011; 184: 873–880.

Specific

- Borczuk AC, Toonkel RL, Powell CA. Genomics of lung cancer. *Proc Am Thorac Soc* 2009; 6: 152–158.
- 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.
- Heist RS, Sequist LV, Engelman JA. Genetic changes in squamous lung cancer: a review. *J Thorac Oncol* 2012; 7: 924–933.
- Kabesch M. Novel asthma-associated genes from genome-wide association studies. What is their significance? *Chest* 2010; 137: 909–915.
- Moffatt MF, Gut IG, Demenais F, *et al*. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211–1221.
- Netea MG, van der Meer JW. Immunodeficiency and genetic defects of pattern-recognition receptors. *N Engl J Med* 2011; 364: 60–70.
- Silverman EK, Sandhaus RA. Alpha1-antitrypsin deficiency. *N Engl J Med* 2009; 360: 2749–2757.
- 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: 1082–1090.