Genetic and environmental interactions in lung diseases

Petros Bakakos,
Apostolos Travlos,
Nikos Koufos

1st Department of Respiratory Medicine,
Medical School, National and Kapodistrian University of Athens, “SOTIRIA” Hospital for Chest Diseases, Athens, Greece

Key words:
- genotype
- phenotype
- environment
- lung diseases

The genotype is the specific composition of genes of an individual and it influences the phenotype. In contrast to the genotype, which is simply inherited, a phenotype is shaped also by epigenetic phenomena and environmental and other external factors.

Some respiratory diseases are caused by the alteration of a single gene, such as cystic fibrosis (CF) and α1-antitrypsin deficiency. For most diseases, however, such a clear-cut relationship between a gene and a disease is not evident and other factors, such as geographical distribution or familial clusters indicate a genetic background to the disease.

Emerging information indicates that even genetic-based disorders are influenced by the environment and that environmental-based disorders are modified by individual genetic factors in specific physiological responses. Accordingly, genetic aspects need to be considered in all areas of pulmonary medicine. Knowledge of the genotype causing a respective phenotype may be a useful tool for the prediction of outcome or selection of therapeutic options, and in particular the search for individual genotype-phenotype-based forms of treatment.

The increasing prevalence of asthma and other environmentally induced chronic pulmonary diseases is a serious public health concern. Although cigarette smoking is a risk factor for chronic obstructive pulmonary disease (COPD), only 15-20% of cigarette smokers develop COPD, suggesting that genetic susceptibility plays an important role in the risk of the disease.

Asthma is considered to result from the effects of environmental stimuli in genetically susceptible individuals.

It is now clear that the pathogenesis of these complex diseases involves both gene-gene and gene-environment interactions. These interactions are not fully understood and present a great challenge for future studies. Genome-wide association analysis has led to the identification of an increased number of possible susceptibility genes for chronic lung diseases. Most of the relevant studies, however, still lack information regarding the mechanisms by which genetic polymorphisms affect the susceptibility of individuals to a pulmonary disease. Careful characterization of clinical and pathological phenotypes is essential in order to validate any observed association.

Two broad research strategies have been used to identify the genes (quantitative trait loci) that determine susceptibility. The first is a genome scan or positional cloning, which attempts to associate the expression of
genes or markers with phenotypes in segregate populations. This approach was designed to identify genes that are polymorphic between two strains of mice and may account for the different response phenotypes under study. The second strategy is the candidate gene approach, in which genes are chosen a priori as likely determinants of the phenotype under study. Linkage is then assessed between the phenotype of interest and markers flanking the candidate genes, or the candidate genes themselves, but without a genome scan, the candidate gene approach may exclude other important loci that determine a quantitative trait, and the interaction between them. These two strategies have been used to identify genes and genetic mechanisms underlying diseases such as Huntington’s disease, Duchenne muscular dystrophy, insulin-dependent diabetes mellitus (DM), Alzheimer’s disease, chronic granulomatous disease and others. Linkage relationship has been identified of homologous loci in humans and mice, supporting the design of genetic studies in mice, since identification of the chromosomal location of a susceptibility gene in the mouse provides the basis for potential localization of a homologous gene in the human. Genetic epidemiology has progressed from candidate gene studies to fine mapping of linked regions and to whole genome association studies. Many of the original findings, however, have not been confirmed in subsequent studies. This could be due to the fact that gene-environment interactions may vary significantly across ethnic groups. Replication of positive findings continues to be an important issue in genetic association studies. In addition, epigenetic changes may exert profound effects on disease susceptibility, increasing the complexity of these interactions. Epigenetics refers to the study of processes that alter gene activity without changing the DNA sequence. Thus, epigenetics study investigates the changes that occur in gene transcription which are dependent on molecules that bind to DNA. Some of these changes in gene expression are heritable while others are not. The term epigenetics describes a wide range of DNA and histone modifications that contribute to the regulation of gene transcription. There are three primary mechanisms that govern gene expression, namely DNA methylation, non-coding RNAs and histone modification. DNA methylation is controlled by cytosine-methyltransferase, which transfers the methyl group from S-adenosylmethionine to the cytosine C-5 position. Cytosine is methylated only in CG islands, and single cytosines are not methylated. Accordingly, hypermethylation of CpG motifs inhibits binding of RNA polymerases to the gene and results in gene transcription silencing, while hypomethylation enhances gene transcription. Histones are protein spheres that bind DNA. Together with the bound DNA they make up a nucleosome, the core of the chromosome. Histones can be modified mainly by acetylation or methylation and by other mechanisms. Histone methylation leads to the accumulation of other proteins leading to a compacted nucleosome, which inhibits gene transcription. mRNAs are short highly conserved noncoding RNAs binding to 5' untranslated regions (5'UTR) of messenger RNAs. Incomplete binding leads to silencing, while complete binding leads to degradation of the RNA. Activation of mRNAs may suppress certain mediator genes, giving rise to a specific pattern of mediator activation. Epigenetic mechanisms have profound effects on phenotyping and have been shown to be involved in carcinogenesis, cancer progress (and thus, prognosis) and stem cell differentiation. Epigenetic mechanisms that lead to preferential expression of the maternal or paternal allele have been shown to be the cause of rare genetic anomalies such as the Prader-Willi, Angelman and Silver-Russell syndromes. Epigenetic marks can be modified by environmental influences during the lifetime, accounting for phenotypic differences between twins who are genetically identical at birth. Epigenetic mechanisms have so far been associated with a number of tumours, and links with non-malignant diseases are starting to be identified. For example, increased histone deacetylase (HDAC) activity has been observed in patients with asthma treated with steroids, resulting in lowered airway inflammation. In a murine model, in utero supplementation with methyl donors could modify the heritable risk of allergic airway disease by directing the differentiation of T cells toward a Th2 phenotype. Such observations are extremely important as they provide opportunities for intervention. Tobacco smoke is another environmental factor that can act in utero to modify gene expression through DNA methylation, and it has been associated with childhood asthma. Drugs can also modify the epigenome. Two DNA methylation inhibitors, 5-aza-deoxycytidine and 5-azacytidine have been approved for the treatment of myelodysplastic syndrome. One HDAC inhibitor, suberoylanilide hydroxamic acid, has been approved for the treatment of T cell cutaneous lymphoma, and another, trichostatin A, was shown to decrease ovalbumin-induced allergic airway disease in a murine model of asthma.
notypes, but the response is determined by a complex interaction of genes. The consideration of gene-gene interactions, or epistasis, is also important, considering that a complex gene interaction is likely to explain phenotype variation. This phenomenon adds more complexity to the model and dictates the need for larger sample sizes in studies. The majority of studies published in recent years have been genetic association studies in various clinical settings, which have demonstrated a positive association of single nucleotide polymorphisms with lung diseases such as asthma.

The main issue is how this knowledge can ultimately be put to use in clinical medicine, and many believe that the most immediate impact on clinical practice will be in the area of prediction. This is not an easy matter, as, for example, assessment of the SNPs of genes identified by genome-wide association studies (GWAS) was able to explain less that 4% of the variability in the phenotype of height and less than 1% in the variability in FEV1. For this reason, some geneticists criticise GWAS and judge them to be a failure because of their lack of predictive powers, and they support, instead, complete sequencing of the genome. It must be borne in mind that genes operate to be a failure because of their lack of predictive powers, and they support, instead, complete sequencing of the genome. It must be borne in mind that genes operate as asthmas. Beryllium is one of the most toxic elements on the periodic table, eliciting in susceptible humans three disease in persons carrying a mutation. Mutations in BMPR2 confer a 15-20% chance of developing PAH in a carrier’s lifetime, so there must be gene-gene or gene-environment interactions that either enhance or prevent the development of the vascular disease in persons carrying a mutation.

Finding provides evidence that the effects of air pollution can be modified by genetic polymorphisms. Epigenetic changes that do not alter the DNA sequence have been demonstrated in subjects exposed to particulate matter. Oxidative stress has been identified as a significant pathway underlying the toxic effects of air pollutants by triggering a number of redox-sensitive signalling pathways, such as those of inflammatory response and cytokine production. Accordingly, the genes involved in the pathways controlling oxidative stress and inflammation are reasonable candidates for the study of interaction with air pollutants. Polymorphisms in the oxidative stress genes GSTM1, GSTP1 and NQO1 have been studied the most extensively, but the inconsistency of the findings of the studies is a limitation to drawing firm conclusions.

Inhaled particles comprise a complex mixture and several of their components have been linked with toxic effects. Diesel exhaust particles are particularly toxic. In order to explain the dose dependent response to air pollutant exposure, a hierarchical model of oxidative stress has been proposed. Low exposure would lead to the formation of a reactive oxygen species (ROS) activating antioxidant response, followed by the transcription of genes that are significant in detoxification, cytoprotective and antioxidant responses. High exposure activates the nuclear factor (NF)-κB transcription factor and activator protein 1 (AP-1) leading to an increased expression of proinflammatory cytokines such as tumour necrosis factor (TNF)-α, interleukin (IL)-6 and IL-8 and adhesion genes.

Pulmonary arterial hypertension (PAH) is a rare disorder that may be hereditable (HPAH), idiopathic (IPAH), or associated with exposure to certain drugs or toxic substances or other medical conditions. Familial cases have long been recognized and are usually due to mutations in the bone morphogenetic protein receptor type 2 gene (BMPR2), or, much less commonly, two other members of the transforming growth factor (TGF)-β superfamily, activin-like kinase-type 1 (ALK1) and endoglin (ENG), which are associated with hereditary haemorrhagic telangiec-tasia. Mutations in BMPR2 confer a 15-20% chance of developing PAH in a carrier’s lifetime, so there must be gene–gene or gene–environment interactions that either enhance or prevent the development of the vascular disease in persons carrying a mutation.

A number of gene variations, regarded as neutral variations, are not harmful per se but together with specific external stimuli foster the development of certain diseases. Beryllium is one of the most toxic elements on the periodic table, eliciting in susceptible humans three conditions: a) an allergic immune response known as...
beryllium sensitization (BeS), b) acute beryllium disease, an acutely toxic, pneumonitis-like lung condition resulting from exposure to high beryllium concentrations that are, however, rarely seen in modern industry, and c) chronic beryllium disease (CBD), following high or even very low levels of exposure. CBD is a systemic granulomatous lung disorder caused by a specific delayed immune response to beryllium a few months to several decades after exposure, and it has been called the “unrecognized epidemic”. Although not a disease in itself, BeS is a population-based predictor of CBD. Genetic susceptibility to CBD is associated with alleles of the major histocompatibility gene, human leukocyte antigen DP (HLA-DP) containing glutamic acid at the 69th position of the beta chain (HLA-DPβ-E69)37. Glutamine at position 69 in the human leukocyte antigen (HLA)-DPB1 gene is not considered an illness in itself, but when there is contact with beryllium dust, carriers of Glu69+ HLA-DPB1 are at increased risk of developing CBD. Up to 97% of patients with CBD are Glu69+ HLA-DPB1 positive. The frequency of Clu69+ HLA-DPB1 is also increased in beryllium-sensitized healthy individuals and is considered to be a risk factor for the progression from sensitization to CBD.

Mutations may occur outside the germline, called somatic mutations, most of which are silent and either do not cause any defect or are corrected by a respective counterpart, but certain somatic mutations cause tumour genesis. Such an example is a mutation in the MYC gene, which leads to overexpression of c-myc. C-myc is a regulatory protein inducing histone acetylation and its overexpression leads to histone hyperacetylation and subsequently to transcription of a variety of genes. It has been shown that overexpression of c-myc is an important factor in the pathogenesis of small cell lung cancer (SCLC)38, although a single event, such as the mutation of c-myc is not solely responsible for tumour genesis, and in general many genetic alterations are involved in lung cancer oncogenesis, such as DNA methylation, histone modifications and alternative splicing.

Although most lung cancers are the consequence of smoking, a substantial number of molecular-epidemiological studies point to high-prevalence, low-penetrance genetic polymorphisms as modifiers of environmental lung cancer risk39. During tumour development the cell accumulates multiple genetic changes, which generate the transformed phenotype, i.e., a cell with increased genetic instability. Lung cancer is a useful model for the study of the interplay between genetic factors and environmental exposure, since the primary aetiology is well established.

Over the last two decades, a considerable mass of data has been generated, mostly addressing the interactions between smoking and xenobiotic-metabolizing enzymes in smoking-related cancers40. Many of the metabolic enzymes have recently been shown to express genetic polymorphisms in the population, and an association has been found between cigarette smoke-induced lung cancer and CYP1A1, CYP2D6, and GSTM1 genes. In addition, GSTM1 and NAT2 polymorphisms have been associated with susceptibility to bladder cancer41. A family history of lung cancer in any family member has been found to be associated with increased lung cancer risk of an individual; the odds associated with a history of lung cancer in the father, mother and sibling were 1.41, 2.14 and 1.53, respectively. These associations were generally stronger in never smokers, younger subjects and for the adenocarcinoma and squamous cell carcinoma subtypes42.

There are indications that interstitial lung diseases such as sarcoidosis and idiopathic pulmonary fibrosis (IPF) have a specific genetic background, and familial clusters are observed for both these conditions. An increased risk of sarcoidosis in close relatives of patients has been documented in populations from Spain and in African-Americans, suggesting that multiple small or moderate genetic effects cause a predisposition to sarcoidosis. As more detailed clinical classification of patients is made, stronger genetic associations between distinct clinical phenotypes and specific gene variants have been revealed. Associations have been found between genes of HLA class II antigens, and especially HLA-DRB1*03, and mild disease with spontaneous resolution in Swedish, Polish, Croatian and Czech populations43. Genome-wide association studies have identified two additional candidate genes, the butyrophilin-like 2 gene (BTNL2) and the Annexin A11 (ANXA11) gene. BTNL2 is necessary for the downregulation of T cell activation, so a dysfunctional gene product contributes to the exaggerated T cell activation in sarcoidosis. Annexin A11 is involved in apoptosis and proliferation. Löfgren's syndrome has been associated with a variant of the TNFa2 gene related to higher cytokine production44.

Emerging concepts suggest that IPF is the result of epithelial-mesenchymal interaction. The initiation of this fibrotic response may depend upon both genetic factors and environmental triggers45. Genome-wide linkage scans in familial interstitial pneumonia demonstrate linkage to chromosomes 4, 5 and 1146. Familial IPF is linked with two mutations in the surfactant protein C (SP-C) gene, resulting in protein misfolding and causing type-II epithelial cell injury. In the first mutation, there is a change from guanine to adenine at the starting point of exon 4,
resulting in a final protein lacking 37 amino acids. This shortened protein is dysfunctional and it becomes misfolded and accumulates in a perinuclear pattern in the cells. It aggregates with normal SP-C, which culminates in a lack of mature SP-C in the alveolar lumen. The second mutation involves a change from thymidine to adenine in exon 5 that results in the substitution of glutamine by leucine in position +188. Eventually, an accumulation of a pro-SP-C occurs in the cell. Different histological patterns were observed in the affected subjects, suggesting the influence of modifier genes and/or environmental factors. In both mutations, the pathological pattern is nonspecific interstitial pneumonitis in younger patients and usual interstitial pneumonia in the elderly.

In contrast to IPF, immunological inflammation appears to be more prominent in the pathogenesis of scleroderma lung fibrosis, which is an autoimmune disease with specific autoantibodies; antitopoisoenzyme antibodies in patients with diffuse lung disease, and anticitromere antibodies in patients with pulmonary vascular disease. Antitopoisoenzyme antibody positivity is associated with the carriage of human leukocyte antigen DRB1*11 and DPB1*1301 alleles, suggesting the recognition of a specific amino-acid motif. Extended haplotype analysis also supports the conclusion that TNF may be the primary association with anticitromere positivity. Associations with TGFβ1 and genes involved in extracellular matrix homeostasis have also been reported in this disease.

In a recent study, eighteen microRNAs, including let-7d, were found to be significantly decreased in IPF. The down-regulation of let-7d in IPF and the profibrotic effects of this down-regulation in vitro and in vivo suggest a key regulatory role for this microRNA in preventing lung fibrosis.

Using an experimental mouse model of infection with virulent Mycobacterium tuberculosis (MTB) for the genetic analysis of host resistance to this pathogen, several tuberculosis susceptibility loci have been identified in otherwise immunocompetent mice. The sst1 locus has been mapped in mouse chromosome 1 and shown to be especially important for control of pulmonary tuberculosis. Rampant progression of tuberculosis infection in the lungs of the sst1-susceptible mouse was associated with the development of necrotic lung lesions, which was suppressed by the sst1-resistant allele. Using a positional cloning approach, a novel host resistance gene, Ipr1, has been identified, which is encoded within the sst1 locus and mediates innate immunity to the intracellular bacterial pathogens MTB and Listeria monocytogenes. The sst1 locus and the Ipr1 gene participate in the control of intracellular multiplication of virulent MTB and have an effect on the mechanism of cell death of infected macrophages.

Data support the concept of Wegener’s granulomatosis (WG) as a multifactorial disease in which both genetic and environmental determinants are involved. Genetic investigations have identified various candidate genes, with α1-antitrypsin deficiency being the most consistently reported genetic susceptibility factor to date.

It is definite that careful characterization of clinical and pathological phenotypes is essential. Further study is required in order to provide high-level evidence on the interaction between genes, environment and disease expression in many lung diseases, which would facilitate the development of novel preventive and/or therapeutic strategies.

REFERENCES

12. Feinberg AP. Epigenetics at the epicentre of modern medicine.