

## Review

# Genetic polymorphisms in lung disease: bandwagon or breakthrough?

Michael C Iannuzzi\*, Mary Maliarik† and Benjamin Rybicki‡

\*Division of Pulmonary, Critical Care, Henry Ford Health System, Detroit, Michigan, USA

†Division of Allergy and Immunology, Henry Ford Health System, Detroit, Michigan, USA

‡Department of Biostatistics and Epidemiology, Henry Ford Health System, Detroit, Michigan, USA

**Correspondence:** Michael C Iannuzzi, Division of Pulmonary, Critical Care, Allergy and Immunology, Henry Ford Health System, 2799 West Grand Boulevard, Detroit, MI 48202, USA. Tel: +1 313 916 3757; fax: +1 313 916 9102; e-mail: [Miannuz1@HFHS.org](mailto:Miannuz1@HFHS.org)

Received: 30 August 2001

Revisions requested: 17 October 2001

Revisions received: 18 January 2002

Accepted: 22 January 2002

Published: 19 February 2002

*Respir Res* 2002, **3**:15

© 2002 BioMed Central Ltd

(Print ISSN 1465-9921; Online ISSN 1465-993X)

## Abstract

The study of genetic polymorphisms has touched every aspect of pulmonary and critical care medicine. We review recent progress made using genetic polymorphisms to define pathophysiology, to identify persons at risk for pulmonary disease and to predict treatment response. Several pitfalls are commonly encountered in studying genetic polymorphisms, and this article points out criteria that should be applied to design high-quality genetic polymorphism studies.

**Keywords:** genetic predisposition to disease, genetics, human, polymorphism (genetics), pulmonary

## Introduction

Genetic polymorphisms are defined as variations in DNA that are observed in 1% or more of the population. Genetic polymorphisms may alter protein structure and function through a single nucleotide base substitution in a gene's coding region, and may increase or decrease gene expression either by affecting mRNA stability when occurring in a gene's 3' untranslated region or by altering transcription factor binding when occurring in the 5' promoter region. Alternatively, a polymorphism may have no discernable effect on the protein product and may lie within DNA regions that are not involved in gene transcription or translation. Polymorphisms that exist in these regions as variations in repeat sequences throughout the genome have served the basis for genetic linkage analysis [1].

The study of genetic polymorphisms promises to help define pathophysiologic mechanisms, to identify individuals at risk for disease and to suggest novel targets for

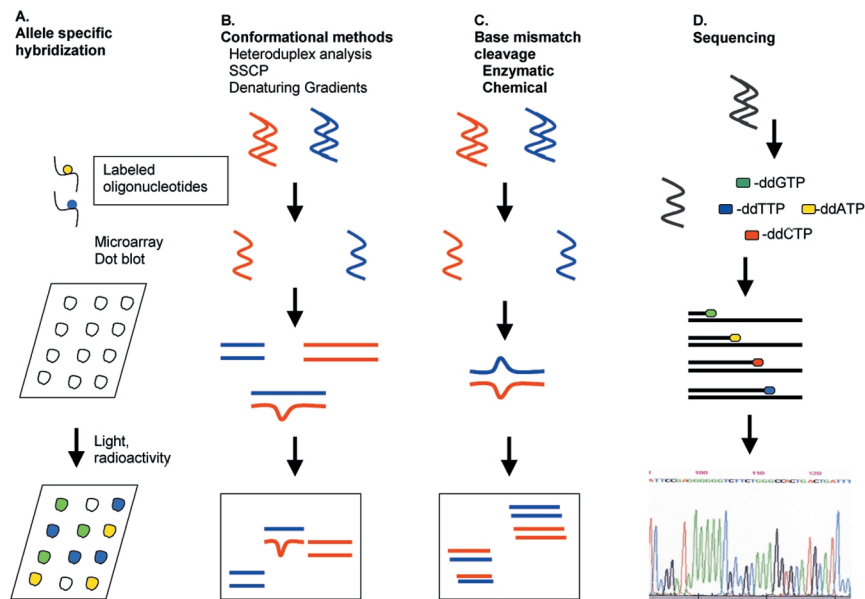
drug treatment. The methodology to study polymorphisms is simple, requiring only access to a polymerase chain reaction machine, funding for reagents, and DNA samples from cases and controls (Fig. 1 illustrates the methods used to detect polymorphisms). The seemingly unlimited potential of genetics to help predict who will get lung disease or who, once diagnosed with disease, will have an unfavorable prognosis has inspired many investigators to jump on the bandwagon of studying genetic polymorphisms. While progress in understanding and treating pulmonary diseases has occurred through investigating genetic polymorphisms, the limitations and potential pitfalls of this approach may be under-appreciated.

## Approaches to genetic polymorphism analysis

Polymorphisms are most often assessed as contributing to disease susceptibility or progression using association studies [2]. We will thus focus on factors affecting the

COPD = chronic obstructive pulmonary disease; HLA = human leukocyte antigen; HRR = haplotype relative risk; IL = interleukin; IPF = idiopathic pulmonary fibrosis; LD = linkage disequilibrium; LHS = Lung Health Study; mEH = microsomal epoxide hydrolase; TDT = transmission disequilibrium test; TNF = tumor necrosis factor.

Figure 1



(a) Several methods to detect specific nucleotide changes (polymorphisms) exist. One method relies on hybridization of oligonucleotides of known sequences to target DNA. The target DNA is generally obtained using the polymerase chain reaction and specific primers. Allele-specific oligonucleotides are then used to detect single base changes in the DNA samples. Typically, target DNA is immobilized on a solid support and denatured. Labeled (radioactive or fluorescent) oligonucleotides are then allowed to anneal. Complementary sequences bind while noncomplementary sequences do not. Sequences that match the oligonucleotide are detected by fluorescence or when the oligonucleotide is radiolabeled by exposure to X-ray film. (b) Another means of rapid screening for DNA variations relies on detecting conformational changes in secondary structure caused by the nucleotide sequence alteration. The change in structure can be detected in a number of ways including denaturing gradient electrophoresis and denaturing gradient high-performance liquid chromatography. SSCP, single-stranded conformational polymorphism. (c) Base mismatch methods begin with creating heteroduplexes between wild-type or normal DNA and target DNA. Heteroduplexes with mismatches are detected by enzymatic or chemical cleavage, with the cleavage products resolved by electrophoresis. (d) DNA sequencing can also be used to detect polymorphisms but is the most labor intensive. The method involves synthesis of DNA using DNA polymerase. Dideoxynucleotides are included in the synthesis mix to randomly terminate synthesis at each nucleotide in the sequence. Generally, each dideoxy nucleotide is labeled with a fluorescent tag. Terminated strands are separated by denaturing gel or capillary electrophoresis and are detected using fluorescence.

quality of association study design. To identify susceptibility loci, association studies involve typing a genetic polymorphism in unrelated affected individuals and in a group of healthy, ethnically matched controls. A given polymorphism is associated with the disease if that allele occurs at a significantly higher frequency among cases compared with controls. When evaluating polymorphisms as disease progression factors, rather than comparing affected individuals with healthy controls, investigators compare individuals with extreme phenotypes. If a significant association emerges, three possibilities exist: the polymorphism itself is the locus of interest, the polymorphism is in linkage disequilibrium (LD) with the locus, or confounding factors are present.

LD exists when alleles at two separate genetic loci are found more often together in a population than would be expected based on their individual allelic frequencies. Possible causes of LD include recent mutation, founder effects, or selection. Another potential cause of LD is a population admixture, where two populations that have been apart for a significant amount of time (and subse-

quently may have very different allelic frequencies at numerous genetic loci) combine to form a hybrid population. Depending on the nature of the admixture, the resulting LD can be expanded beyond distances generally observed in more stable populations [3]. The best estimates of LD within an outbred population suggest that LD is unlikely to extend over distances more than 1–2 cM or about 1–2 million base pairs, although in an inbred population, such as the Hutterites, LD may exist over 10 times that distance [4].

Confounding factors must be considered particularly when polymorphisms identified in one study cannot be duplicated in a similar ethnic group. One confounding factor is population stratification. This may occur with an unbalanced ethnic admixture, such as a Caucasian admixture in the African-American gene pool [5]. Fortunately, this problem can be overcome by careful planning in either the analysis or study design phase [6]. Family-based association studies, which require the genotyping of affected individuals and their parents (and/or unaffected sibs), are specifically designed to control for the genetic background that can cause con-

founding. Two common test statistics for these types of studies are the transmission disequilibrium test (TDT) and the haplotype relative risk (HRR). The TDT compares the frequency with which each allele is transmitted from a heterozygous parent to an affected offspring [7,8]. The HRR is similar to the TDT, but compares transmission of genotypes (haplotype) rather than alleles [9,10].

A critical factor to consider in genetic polymorphism studies is the choice of what phenotype to investigate. In fact, many studies that have evaluated genetic polymorphisms have been hindered by the case sample containing multiple phenotypes. For example, some studies of asthma have included samples consisting of patients with intrinsic, extrinsic, adult onset, mild and severe asthma. Studies that seek to determine whether an association between a polymorphism and disease exists can be greatly improved by studying a more narrowly defined intermediate phenotype. In the case of asthma, intermediate phenotypes include total IgE and bronchial hyperresponsiveness. Another consideration is that genes responsible for disease susceptibility may not be the same genes involved with disease progression. It may be beneficial to limit the sample to those with a particular stage or severity of disease. For example, to evaluate chronic obstructive pulmonary disease (COPD) candidate genes, investigators have focused on patients with severe early onset COPD [11,12].

Association studies are limited to evaluating DNA polymorphisms near or within candidate genes. To perform a genome screen to search for candidate genes, linkage analysis using families or affected siblings is required. Linkage analysis is comprehensive and locates genes that exert a major effect on disease susceptibility, but linkage analysis has relatively low power and will fail to detect genes conferring only mild to moderate disease risk. For instance, if a disease susceptibility allele exerts a twofold disease risk compared with the wild-type allele, several hundred to several thousand families need to be typed, a sample size that may not be achievable. Association studies have greater power, but associations are detected over much smaller genetic regions (thousands of base pairs) compared with that detected by linkage analysis (millions of base pairs). To perform a genome scan with association studies, tens of thousands of markers would be needed, which is not possible with current technology; although it is anticipated that this may soon be possible [13].

Most lung diseases require some type of environmental inciting agent to be manifest. For most complex diseases where genetic susceptibility alone accounts for only a fraction of disease variation, not considering the environment can severely underpower gene-finding studies. Furthermore, genome screens or association studies performed on populations not stratified nor selected on the basis of environmental exposure may only identify

genes for which the relevant environmental exposure is ubiquitous in that population. For example, studying a random population of asthmatics selected from the Midwest region of the US would stand a reasonable chance of identifying genes important in determining house dust mite response, but would not be likely to identify genes important in isocyanate-induced asthma.

Gene–environment interaction may manifest in various ways, including differential exposure risk effects based on an individual's genotype, or differential gene risk effects based on an individual's exposure. Methods to study gene–environment interaction have been reviewed by Yang and Khoury [14]. Two main interactions exist: statistical and biologic. A statistical interaction of risk factors (gene and environment) involves the coefficient of the product term of the genetic and environmental risk factors with the interaction measured in terms of departure from a multiplicative model. This method is arbitrary, model dependent and can ignore interaction or synergy on the biologic level. In the biologic interaction model, interaction between two factors is defined as their co-participation in the same causal mechanism to disease development, and in some instances may only be detectable in terms of a departure from an additive model.

### **Disease gene associations**

Having reviewed the general approach for evaluating polymorphisms, we turn to some recent examples. Table 1 displays examples of recently published reports of polymorphisms that were evaluated in a variety of lung diseases.

#### **Chronic beryllium disease**

Genetic polymorphisms in human leukocyte antigen (HLA) genes associated with resistance or susceptibility to disease exist. HLA polymorphisms determine immune response variation to individual antigens, including autoantigens, and thus make for excellent candidate genes for a variety of immune-mediated disorders [15]. One striking example is in the genetic analysis of patients with chronic beryllium disease. These studies revealed an association with alleles of HLA-DPB1 encoding a DP beta chain with glutamic acid at residue 69 [16]. Individuals with glutamic acid at position 69 have nearly a 10-fold increased disease risk [17]. Furthermore, functional studies have demonstrated that the presence of glutamic acid at residue 69 is essential for reactivity in T-cell clones generated from three patients with disease [18]. Few examples in pulmonary diseases exist where the HLA association with disease risk is as strong.

Given the clinical and pathohistologic similarities of chronic beryllium disease and sarcoidosis, we have evaluated this same polymorphism in sarcoidosis patients and controls, and found no association [19]. The lack of association of glutamic acid at residue 69 in sarcoidosis illus-

Table 1

## Examples of recently published studies of polymorphisms in lung disease

Disease	Reference	Summary
<b>ARDS</b>		
Polymorphisms of human SP-A, SP-B and SP-D genes: association of SPB Thr131ILE with ARDS	[40]	Data presented suggest that SP-B or a linked gene contributes to susceptibility to ARDS
<b>Asthma</b>		
Effect of polymorphism of the $\beta$ -2-adrenergic receptor on response to regular use of albuterol in asthma	[41]	Arg/Arg subjects who used albuterol regularly had AM PEF lower than Arg/Arg patients who had used albuterol as needed only. Subjects homozygous for glycine at $\beta$ -2-adrenergic receptor-16 showed no such decline
Association of a promoter polymorphism of the CD14 gene and atopy	[42]	-159 C to T promoter polymorphism in the CD14 gene was found associated with expression of a more severe allergic phenotype
The role of the C-C chemokines receptor-5 Delta32 polymorphism in asthma and in the production of regulated on activation, normal T cells expressed and secreted	[43]	Data indicate that the CCR5* $\Delta$ 32 allele is not a genetic risk factor for the development of asthma and does not influence disease severity nor influence RANTES production
<b>COPD</b>		
TNF- $\alpha$ gene promoter polymorphism in COPD	[44]	TNF gene promoter allele was not found to influence the risk of developing COPD in a Caucasian population of smokers and there was no association with severity of airflow obstruction
A polymorphism in the TNF- $\alpha$ gene promoter region may predispose to a poor prognosis in COPD	[45]	Homozygosity for adenine substitution polymorphism at position -308 was found associated with more severe airflow obstruction and a worse prognosis
Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema	[46]	Findings suggest that the large size of a GT( <i>n</i> ) repeat in the heme oxygenase-1 gene promoter may reduce the gene's inducibility by reactive oxygen species in cigarette smoke, thus resulting in emphysema
<b>Cystic fibrosis</b>		
HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype	[47]	DR7 allele was significantly associated with an increase in total IgE and <i>Pseudomonas aeruginosa</i> colonization in cystic fibrosis patients
An $\alpha$ 1-antitrypsin enhancer polymorphism is a genetic modifier of pulmonary outcome in cystic fibrosis	[48]	An enhancer polymorphism in the AAT gene was found associated with better pulmonary prognosis in cystic fibrosis patients
<b>Hypersensitivity pneumonitis</b>		
Major histocompatibility complex and TNF- $\alpha$ polymorphisms in pigeon breeder's disease	[49]	Results suggest that genetic factors located with the major histocompatibility complex region contribute to the development of pigeon breeder's disease
TNF- $\alpha$ -308 promoter gene polymorphism and increased TNF serum bioactivity in farmer's lung patients	[50]	The frequency for the TNFA2 allele, a genotype associated with high TNF- $\alpha$ production <i>in vitro</i> , was significantly higher in farmer's lung patients
<b>Idiopathic pulmonary fibrosis</b>		
Analysis of TNF- $\alpha$ , lymphotoxin alpha, TNF receptor II, and IL-6 polymorphisms in patients with idiopathic pulmonary fibrosis	[28]	This is the first paper to suggest that disease progression in idiopathic pulmonary fibrosis may be linked to a particular genetic marker or to functional polymorphisms
<b>Sarcoidosis</b>		
HLA-Gm/ $\kappa$ interaction in sarcoidosis. Suggestions for a complex genetic structure	[51]	This study addresses the interplay between IgG heavy chain/ $\kappa$ light chain markers and major histocompatibility complex genes
Lack of association with IL-1 receptor antagonist and IL-1 $\beta$ gene polymorphisms in sarcoidosis patients	[52]	No bias in the IL-1 receptor antagonist and IL-1 $\beta$ genotype was found in Japanese sarcoidosis patients
CC chemokine receptor gene polymorphisms in Czech patients with pulmonary sarcoidosis	[53]	CCR5 $\Delta$ 32 and CCR2-64I were found associated with sarcoidosis
<b>Silicosis</b>		
Polymorphisms of the IL-1 gene complex in coal miners with silicosis	[54]	This is the first report showing an association between the IL-1 receptor antagonist polymorphism and silicosis

PEF, Peak expiratory flow; ARDS, acute respiratory distress syndrome; COPD, chronic obstructive pulmonary disease; HLA, human leukocyte antigen; IL, interleukin; TNF, tumor necrosis factor.

trates the pitfall of studying polymorphisms in candidate genes, however attractive, chosen based on a limited understanding of the pathophysiology of disease.

### Chronic obstructive pulmonary disease

That only 10–20% of cigarette smokers develop symptomatic COPD suggests that genetic factors are likely to be important. In addition, several studies have shown an increased prevalence of COPD within families. COPD thus appears to be ripe to investigate genetic polymorphisms in disease susceptibility [20].

Studies have implicated oxidant–anti-oxidant interaction in the pathogenesis of COPD, which led Smith and Harrison to investigate genetic polymorphisms of the xenobiotic metabolizing enzyme, microsomal epoxide hydrolase (mEH) [21]. A mEH slow allele and a mEH fast allele exist. The homozygous state for the slow alleles results in very slow microsomal epoxide hydrolase activity. The presence of epoxides in the lung for longer periods following cigarette smoke exposure could lead to greater tissue damage and inflammation.

The study design and statistical analysis in the Smith and Harrison study is commonly encountered in reports of disease-associated genetic polymorphisms. The investigators studied blood donor controls ( $n = 203$ ), patients with asthma ( $n = 57$ ), patients with lung cancer ( $n = 50$ ), patients with COPD ( $n = 68$ ) and patients with emphysema ( $n = 94$ ). The proportion of individuals with innate slow activity was significantly higher in both the COPD group and the emphysema group than in the control group: COPD 19% versus control 6%, and emphysema 22% versus control 6%. The odds ratios for homozygous slow activity versus all other phenotypes were 4.1 for COPD and 5.0 for emphysema. Koyama and Geddes have noted that caution is needed over the interpretation of these findings [22]. The groups are small and the emphysema group was unusual, being defined from the morbid anatomy of lung samples resected for cancer.

Sakao *et al.* have argued that tumor necrosis factor (TNF)- $\alpha$ , a potent pro-inflammatory cytokine, may be involved in the development of COPD [23]. TNF- $\alpha$  has been reported to be elevated in bronchoalveolar lavage, bronchial biopsies and induced sputum of COPD patients. A polymorphism at position -308 of the TNF- $\alpha$  gene promoter is associated with alteration of TNF- $\alpha$  secretion *in vitro* [24]. The polymorphism consists of a guanine to adenine substitution. The guanine allele was denoted as 1 and the adenine allele as 2. TNF- $\alpha$ -308\*2, the rarer allele, has been associated with higher baseline and induced expression of TNF- $\alpha$ .

Sakao *et al.* compared TNF- $\alpha$ -308 1/2 allele frequencies in 106 Japanese patients with 110 asymptomatic smoker/ex-smoker control subjects matched for sex and age, and 129 population control blood donors. The authors

reported that TNF- $\alpha$ -308 1/2 alleles were significantly associated with the presence of smoking-related COPD. Allele frequencies were significantly different among the groups: in patients with COPD, the 1/2 allele frequencies were 0.835/0.165; in smoker/ex-smoker control subjects, 0.918/0.082; and in the population control subjects, 0.922/0.078. However, the TNF- $\alpha$ -308\*2 allele was not found to be associated with COPD in a white population. Furthermore, studies have been inconsistent in demonstrating an association with the presence of the TNF- $\alpha$ -308\*2 allele in a number of inflammatory diseases such as sarcoidosis and asthma [25–27].

Sandford *et al.* noted that widely divergent rates of decline in lung function in smokers would be a robust phenotype for detecting genes that contribute to COPD severity [11]. This association study has enhanced features, including reducing phenotypic heterogeneity by focusing on the decline of lung function rather than COPD and by comparing extreme phenotypes. From 5887 male and female smokers recruited to the Lung Health Study (LHS) conducted by the National Heart, Lung and Blood Institute, Sandford *et al.* selected 283 smokers with the fastest rate of decline of forced expiratory volume in 1 s, and 308 smokers who had no decline. All subjects were white and continued to smoke during the 5-year period of the LHS. These subjects were genotyped for polymorphisms in  $\alpha$ 1-antitrypsin, mEH, vitamin D binding protein, and TNF- $\alpha$  and TNF- $\beta$  genes. TNF- $\beta$  is also known as lymphotoxin alpha. The authors found [11] that the  $\alpha$ 1-antitrypsin MZ genotype and the mEH His113/His139 slow haplotype were associated with increased rate of decline of lung function. Both of these associations were strong when the subject had a family history of COPD, suggesting an interaction with other familial risk factors. No association of the TNF haplotypes with rate of decline of lung function was found.

It would be helpful to re-evaluate these polymorphisms in a family-based association study using TDT or HRR. Unlike asthma or sarcoidosis, however, COPD generally occurs later in life. Those recruited to the LHS range in age from 35 to 60 years, making it more difficult to recruit their parents.

### Idiopathic pulmonary fibrosis

A strong link between overexpression of pro-inflammatory mediators and idiopathic pulmonary fibrosis (IPF) exist. Recently, Pantelidis *et al.* reported an evaluation of TNF, lymphotoxin alpha, TNF receptor II and IL-6 polymorphisms in patients with IPF [28]. These investigators, through their thorough analysis, raised several issues regarding genetic polymorphism studies. While allele frequencies did not differ between a normal, white, British control population and a sample with IPF, they did observe a significant increase in the frequency of a particular TNF haplotype in females with IPF compared with males. Similar gender association has been observed in the distri-



bution of TNF- $\alpha$  haplotypes in ulcerative colitis [29]. This raises the issue that polymorphisms may act differently in women, and separate analysis of associations with polymorphisms in women may need to be considered. Panteleidis *et al.* [28] also noted an increased frequency of co-carriage of the IL-6 (intron 4G) allele located on chromosome 7p21-p14 and the TNF receptor II (1690C) allele located on chromosome 1p36.2. Since complex disorders, such as IPF, can be expected to involve several genes, combination of alleles on different chromosomes may need to be considered.

### Tuberculosis

Convincing evidence exists that genetic factors are important in tuberculosis susceptibility. Stead *et al.* found, among over 25,000 tuberculin-negative nursing home residents, that black subjects were twice as likely to become infected with tuberculosis as white subjects living in the same environment [30]. Twin studies in tuberculosis have consistently found much higher disease concordance among monozygotic than dizygotic twins [31].

Studies on murine models of susceptibility to mycobacterial infection led to the discovery of the natural resistance-associated macrophage protein gene and its human homolog NRAMP1 [32,33]. Bellamy *et al.* found a significant association of NRAMP polymorphisms and tuberculosis in a Gambian population ( $n=800$ ) [34]. Functional studies indicate that NRAMP1 is involved in the early stages of macrophage priming and activation, making NRAMP1 an attractive candidate gene for both tuberculosis and sarcoidosis [35].

We analyzed the same NRAMP gene polymorphisms found associated with tuberculosis in an association case-control study of 157 African-American sarcoidosis patients and 111 control subjects [36]. These polymorphisms included a microsatellite repeat in the 5' region (5'-CA), a non-conservative single base substitution at codon 543 changing aspartic acid to asparagine (D543N) position 543, a TGTG deletion in the 3' untranslated region and a single nucleotide change in intron 4. Our results, in contrast to those reported in tuberculosis patients, showed that the genotypes found associated with tuberculosis were under-represented in the sarcoidosis patients, suggesting a potential protective effect in sarcoidosis. One could speculate a mechanism whereby altered NRAMP expression could lead to susceptibility to tuberculosis and a protective effect in sarcoidosis, but it is more important to confirm these findings. We are therefore presently analyzing NRAMP polymorphisms in 240 sarcoidosis families using TDT.

### Pharmacogenetics

An exciting development has been the evaluation of genetic polymorphisms in determining treatment response. The best data concerns beta-2-adenoreceptor

polymorphisms. It appears that individuals who are homozygous for the glycine 16 variant of the adenoreceptor, which downregulates to a greater extent than other forms of the receptor, show a reduced response following chronic beta-agonist use [37,38]. Treatment response to 5-lipoxygenase inhibitors also appears determined by polymorphisms in the promoter regions of this gene [39].

### Bandwagon or breakthrough?

Several reasons exist why the genetic evaluation of lung diseases is not proceeding as rapidly as one might expect. There is a need to apply several criteria (Table 2) in designing high-quality association studies and in the interpretation of their results.

### Accurate narrow definition of the phenotype

Disease heterogeneity may obscure association between a disease subtype and a polymorphism. By narrowing the phenotype, the investigator improves chances of uncovering an association. An alternative strategy is to analyze an intermediate biologic or clinical phenotype (e.g. elevated IgE in patients with asthma).

### Large sample size

Given the number of conflicting reports of disease-associated genetic polymorphisms, it is important to point out that most negative studies lack power. A strong inverse relationship between the allele frequency in a population and the sample size required to test the allele contribution to the phenotype exists. Furthermore, complex diseases are polygenic in nature and we generally evaluate one genetic variant or a small fraction of genetic variants of the many genes likely to contribute, thus further requiring a larger sample size. Networks of investigators are probably needed when addressing complex diseases. Given the cost and time required to recruit such populations, investigators should be encouraged to enter into large-scale collaborations. For example, in the US, in the multicenter A Case Control Etiologic Sarcoidosis Study (ACCESS), 10 clinical centers recruited over 700 cases and 700 controls; and in the Sarcoidosis Genetic Analysis (SAGA) study, 11 centers are recruiting 350 affected sibling pairs and their family members for linkage analysis.

### Well-matched controls

Besides lack of power, most genetic polymorphism studies are confounded by lack of well-matched controls. In fact, rather than the recruitment of patients, the real challenge often lies in the recruitment of controls. Control groups are generally confounded by a selection bias that may influence the genetic makeup of that population.

### Biological plausibility and functional significance of candidate genes

Most association studies evaluating candidate genes choose candidates based on their potential role in the

**Table 2****Criteria for designing high-quality association studies and interpreting the results**


---

Accurate narrow definition of the phenotype
<i>A priori</i> estimate of study power
Large sample size
Well-matched controls
<i>P</i> values adjusted for the number of polymorphisms tested
Biological plausibility and functional significance of candidate genes
Independent replication in other populations
Confirmation in family-based studies (transmission disequilibrium test, haplotype relative risk)

---

pathophysiology of the disease. When evaluating polymorphisms, it is best that they have some plausible biologic role. However, when a candidate gene is chosen based on previous linkage with the disease or an intermediate phenotype, it is more likely that the candidate gene is in some way involved.

**Independent replication in other populations**

Polymorphism–disease associations from different populations are extremely difficult to interpret. Different genetic variants contributing to the phenotype may be different in different populations. We therefore cannot expect to generalize the findings in one ethnic group to another ethnic group, but we can insist that when an association is found using a case–control design, it is repeated using a family-based association study to minimize population stratification effects.

**Conclusions**

We have defined genetic polymorphisms, have pointed out why their study has become common, and have reviewed several pitfalls that need to be considered in designing genetic polymorphism studies and their interpretation. We have discussed association and linkage studies. Linkage studies can ensure that genes exerting large effects on disease susceptibility have not been missed, but will fail to identify genes exerting mild to moderate effects on disease risk. Association studies can detect genes that exert smaller effects, but cannot be used to screen the genome. Association studies look at known genes. TDTs greatly reduce the likelihood that any allele frequency differences between cases and controls might be due to unsuspected genetic differences among subgroups within the population.

A search for disease genes is best accomplished when both association and linkage strategies are used. We briefly cited some of the genetic polymorphism studies in

COPD, sarcoidosis, IPF and tuberculosis. Criteria for a high-quality association study were listed and discussed. While many have jumped on the genetic polymorphism bandwagon, those investigators who address the potential pitfalls will use genetic polymorphisms to provide breakthroughs in understanding disease.

**References**

- Weber JL, Broman KW: **Genotyping for human whole-genome scans: past, present, and future.** *Adv Genet* 2001, **42**:77-96.
- Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273**:1516-1517.
- Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD: **Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium.** *Am J Hum Genet* 2001, **68**:198-207.
- Ober C, Cox NJ: **The genetics of asthma. Mapping genes for complex traits in founder populations.** *Clin Exp Allergy* 1998, **28**(suppl 1):101-105 (discussion 108-110).
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: **Estimating African American admixture proportions by use of population-specific alleles.** *Am J Hum Genet* 1998, **63**:1839-1851.
- Satten GA, Flanders WD, Yang Q: **Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model.** *Am J Hum Genet* 2001, **68**:466-477.
- Spielman RS, McGinnis RE, Ewens WJ: **Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).** *Am J Hum Genet* 1993, **52**:506-516.
- Hodge SE: **Linkage analysis versus association analysis: distinguishing between two models that explain disease-marker associations.** *Am J Hum Genet* 1993, **53**:367-384.
- Li T, Arranz M, Aitchison KJ, Bryant C, Liu X, Kerwin RW, Murray R, Sham P, Collier DA: **Case-control, haplotype relative risk and transmission disequilibrium analysis of a dopamine D2 receptor functional promoter polymorphism in schizophrenia.** *Schizophr Res* 1998, **32**:87-92.
- Martin RB, Alda M, MacLean CJ: **Parental genotype reconstruction: applications of haplotype relative risk to incomplete parental data.** *Genet Epidemiol* 1998, **15**:471-490.
- Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, Pare PD: **Susceptibility genes for rapid decline of lung function in the lung health study.** *Am J Respir Crit Care Med* 2001, **163**:469-473.
- Silverman EK, Speizer FE, Weiss ST, Chapman HA Jr, Schuette A, Campbell EJ, Reilly JJ Jr, Ginns LC, Drazen JM: **Familial aggregation of severe, early-onset COPD: candidate gene approaches.** *Chest* 2000, **117**(5 suppl 1):273S-274S.
- Bader JS: **The relative power of SNPs and haplotype as genetic markers for association tests.** *Pharmacogenomics* 2001, **2**:11-24.
- Yang Q, Khoury MJ: **Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research.** *Epidemiol Rev* 1997, **19**:33-43.
- Theofilopoulos AN: **The basis of autoimmunity: Part II. Genetic predisposition.** *Immunol Today* 1995, **16**:150-159.
- Richeldi L, Sorrentino R, Saltini C: **HLA-DPB1 glutamate 69: a genetic marker of beryllium disease.** *Science* 1993, **262**:242-244.
- Richeldi L, Kreiss K, Mroz MM, Zhen B, Tartoni P, Saltini C: **Interaction of genetic and exposure factors in the prevalence of berylliosis.** *Am J Ind Med* 1997, **32**:337-340.
- Lombardi G, Germain C, Uren J, Fiorillo MT, du Bois RM, Jones-Williams W, Saltini C, Sorrentino R, Lechler R: **HLA-DP allele-specific T cell responses to beryllium account for DP-associated susceptibility to chronic beryllium disease.** *J Immunol* 2001, **166**:3549-3555.
- Maliarik MJ, Chen KM, Major ML, Sheffer RG, Popovich J Jr, Rybicki BA, Iannuzzi MC: **Analysis of HLA-DPB1 polymorphisms in African-Americans with sarcoidosis.** *Am J Respir Crit Care Med* 1998, **158**:111-114.

20. Barnes PJ: **Genetics and pulmonary medicine. 9. Molecular genetics of chronic obstructive pulmonary disease.** *Thorax* 1999, **54**:245-252.
21. Smith CA, Harrison DJ: **Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema.** *Lancet* 1997, **350**:630-633.
22. Koyama H, Geddes DM: **Genes, oxidative stress, and the risk of chronic obstructive pulmonary disease.** *Thorax* 1998, **53** (suppl 2):S10-S14.
23. 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**:420-422.
24. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW: **Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation.** *Proc Natl Acad Sci USA* 1997, **94**:3195-3199.
25. Albuquerque RV, Hayden CM, Palmer LJ, Laing IA, Rye PJ, Gibson NA, Burton PR, Goldblatt J, Lesouef PN: **Association of polymorphisms within the tumour necrosis factor (TNF) genes and childhood asthma.** *Clin Exp Allergy* 1998, **28**:578-584.
26. Chagani T, Pare PD, Zhu S, Weir TD, Bai TR, Behbehani NA, Fitzgerald JM, Sandford AJ: **Prevalence of tumor necrosis factor-alpha and angiotensin converting enzyme polymorphisms in mild/moderate and fatal/near-fatal asthma.** *Am J Respir Crit Care Med* 1999, **160**:278-282.
27. Yamaguchi E, Itoh A, Hizawa N, Kawakami Y: **The gene polymorphism of tumor necrosis factor-beta, but not that of tumor necrosis factor-alpha, is associated with the prognosis of sarcoidosis.** *Chest* 2001, **119**:753-761.
28. Pantelidis P, Fanning GC, Wells AU, Welsh KI, Du Bois RM: **Analysis of tumor necrosis factor-alpha, lymphotoxin-alpha, tumor necrosis factor receptor II, and interleukin-6 polymorphisms in patients with idiopathic pulmonary fibrosis.** *Am J Respir Crit Care Med* 2001, **163**:1432-1436.
29. Koss K, Satsangi J, Fanning GC, Welsh KI, Jewell DP: **Cytokine (TNF alpha, LT alpha and IL-10) polymorphisms in inflammatory bowel diseases and normal controls: differential effects on production and allele frequencies.** *Genes Immun* 2000, **1**:185-190.
30. Stead WW, Senner JW, Reddick WT, Lofgren JP: **Racial differences in susceptibility to infection by Mycobacterium tuberculosis.** *N Engl J Med* 1990, **322**:422-427.
31. Comstock GW: **Tuberculosis in twins: a re-analysis of the Proffit survey.** *Am Rev Respir Dis* 1978, **117**:621-624.
32. Vidal SM, Malo D, Vogan K, Skamene E, Gros P: **Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg.** *Cell* 1993, **73**:469-485.
33. Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, Sanchez F, Skamene E, Schurr E, Gros P: **Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression.** *J Exp Med* 1994, **180**:1741-1752.
34. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV: **Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans.** *N Engl J Med* 1998, **338**:640-644.
35. Blackwell JM, Searle S, Goswami T, Miller EN: **Understanding the multiple functions of Nramp1.** *Microbes Infect* 2000, **2**:317-321.
36. Maliarik MJ, Chen KM, Sheffer RG, Rybicki BA, Major ML, Popovich J Jr, Iannuzzi MC: **The natural resistance-associated macrophage protein gene in African Americans with sarcoidosis.** *Am J Respir Cell Mol Biol* 2000, **22**:672-675.
37. Hall IP, Wheatley AP, Dewar JC: **Genetic polymorphisms of adrenergic receptors.** *Methods Mol Biol* 2000, **126**:117-126.
38. Israel E, Drazen JM, Liggett SB, Boushey HA, Cherniack RM, Chinchilli VM, Cooper DM, Fahy JV, Fish JE, Ford JG, Kraft M, Kunselman S, Lazarus SC, Lemanske RF, Martin RJ, McLean DE, Peters SP, Silverman EK, Sorkness CA, Szeffler SJ, Weiss ST, Yandava CN: **The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma.** *Am J Respir Crit Care Med* 2000, **162**:75-80.
39. In KH, Asano K, Beier D, Grobholz J, Finn PW, Silverman EK, Silverman ES, Collins T, Fischer AR, Keith TP, Serino K, Kim SW, De Sanctis GT, Yandava C, Pillari A, Rubin P, Kemp J, Israel E, Busse W, Ledford D, Murray JJ, Segal A, Tinkleman D, Drazen JM: **Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription.** *J Clin Invest* 1997, **99**:1130-1137.
40. Lin Z, Pearson C, Chinchilli V, Pietschmann SM, Luo J, Pison U, Floros J: **Polymorphisms of human SP-A, SP-B, and SP-D genes: association of SP-B Thr131Ile with ARDS.** *Clin Genet* 2000, **58**:181-191.
41. Israel E, Drazen JM, Liggett SB, Boushey HA, Cherniack RM, Chinchilli VM, Cooper DM, Fahy JV, Fish JE, Ford JG, Kraft M, Kunselman S, Lazarus SC, Lemanske RF Jr, Martin RJ, McLean DE, Peters SP, Silverman EK, Sorkness CA, Szeffler SJ, Weiss ST, Yandava CN; National Heart, Lung, and Blood Institute's Asthma Clinical Research Network: **Effect of polymorphism of the beta(2)-adrenergic receptor on response to regular use of albuterol in asthma.** *Int Arch Allergy Immunol* 2001, **124**:183-186.
42. Koppelman GH, Reijmerink NE, Stine OC, Howard TD, Whittaker PA, Meyers DA, Postma DS, Bleecker ER: **Association of a promoter polymorphism of the CD14 gene and atopy.** *Am J Respir Crit Care Med* 2001, **163**:965-969.
43. Sandford AJ, Zhu S, Bai TR, Fitzgerald JM, Pare PD: **The role of the C-C chemokine receptor-5 Delta32 polymorphism in asthma and in the production of regulated on activation, normal T cells expressed and secreted.** *J Allergy Clin Immunol* 2008, **121**:69-73.
44. Higham MA, Pride NB, Alikhan A, Morrell NW: **Tumour necrosis factor-alpha gene promoter polymorphism in chronic obstructive pulmonary disease.** *Eur Respir J* 2000, **15**:281-284.
45. Keatings VM, Cave SJ, Henry MJ, Morgan K, O'Connor CM, FitzGerald MX, Kalsheker N: **A polymorphism in the tumor necrosis factor-alpha gene promoter region may predispose to a poor prognosis in COPD.** *Chest* 2000, **118**:971-975.
46. Yamada N, Yamaya M, Okinaga S, Nakayama K, Sekizawa K, Shibahara S, Sasaki H: **Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema.** *Am J Hum Genet*, **66**:187-195.
47. Aron Y, Polla BS, Bienvenu T, Dall'ava J, Dusser D, Hubert D: **HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype.** *Am J Respir Crit Care Med* 1999, **159**:1464-1468.
48. Henry MT, Cave S, Rendall J, O'Connor CM, Morgan K, FitzGerald MX, Kalsheker N: **An alpha1-antitrypsin enhancer polymorphism is a genetic modifier of pulmonary outcome in cystic fibrosis.** *Eur J Hum Genet* 2001, **9**:273-278.
49. Camarena A, Juarez A, Mejia M, Estrada A, Carrillo G, Falfan R, Zuniga J, Navarro C, Granados J, Selman M: **Major histocompatibility complex and tumor necrosis factor-alpha polymorphisms in pigeon breeder's disease.** *Am J Respir Crit Care Med* 2001, **163**:1528-1533.
50. Schaaf BM, Seitzer U, Pravica V, Aries SP, Zabel P: **Tumor necrosis factor-alpha-308 promoter gene polymorphism and increased tumor necrosis factor serum bioactivity in farmer's lung patients.** *Am J Respir Crit Care Med*, **163**:379-382.
51. Martinetti M, Dugoujon JM, Tinelli C, Cipriani A, Cortelazzo A, Salvaneschi L, Casali L, Semenzato G, Cuccia M, Luisetti M: **HLA-Gm/kappam interaction in sarcoidosis. Suggestions for a complex genetic structure.** *Eur Respir J* 2000, **16**:74-80.
52. Niimi T, Sato S, Tomita H, Yamada Y, Akita K, Maeda H, Kawaguchi H, Sugiura Y, Ueda R: **Lack of association with interleukin 1 receptor antagonist and interleukin-1beta gene polymorphisms in sarcoidosis patients.** *Respir Med* 2000, **94**:1038-1042.
53. Petrek M, Drabek J, Kolek V, Zlamal J, Welsh KI, Bunce M, Weigl E, Du Bois R: **CC chemokine receptor gene polymorphisms in Czech patients with pulmonary sarcoidosis.** *Am J Respir Crit Care Med* 2000, **162**:1000-1003.
54. Yucesoy B, Vallyathan V, Landsittel DP, Sharp DS, Matheson J, Burleson F, Luster MI: **Polymorphisms of the IL-1 gene complex in coal miners with silicosis.** *Am J Ind Med* 2001, **39**:286-291.