

## REVIEW

# Post-genome respiratory epidemiology: a multidisciplinary challenge

F. Kauffmann and the Post Genome Respiratory Epidemiology group

The Post Genome Respiratory Epidemiology group: J. Anto, M.P. Baur, H. Bickeboller, D. Clayton, W.O.C. Cookson, F. Demenais, P.J. Helms, I. Humphrey-Smith, S. Imbeaud, F. Kauffmann, B.M. Knoppers, M. Lathrop, J. Little, N. Pearce, D. Schaid, E. Silverman, S. Weiss, M. Wjst

*Post-genome respiratory epidemiology: a multidisciplinary challenge. F. Kauffmann and the Post Genome Respiratory Epidemiology group. ©ERS Journals Ltd 2004.*

**ABSTRACT:** The introduction of genetic approaches in respiratory epidemiology is novel for most epidemiologists, and the post-genome phase poses new challenges. After describing specific questions pertinent to the field of asthma and chronic obstructive pulmonary disease, two main methodological aspects regarding technological and scientific advances are presented in this review. The first one concerns biological aspects in the genome and post-genome phases, *i.e.* how to study the genome, the transcriptome and the proteome. The second area concerns genetic epidemiology, considering design (case control and family based) and statistical analytical issues. Key aspects are large sample size, good phenotyping and the consideration of environment-by-gene interaction according to windows of opportunity.

Needs that have been identified include the following. 1) Networking for setting standards in the field and access to sufficiently large samples. 2) Multidisciplinarity; the collaboration of epidemiologists, clinicians, geneticists and specialists in bioinformatics, in addition to specialists in disciplines less familiar to epidemiologists, to be prepared for new phenotypic characterisations based on transcriptome and proteome. 3) Training in genetic analytical techniques for some respiratory epidemiologists, as well as in respiratory epidemiology for some genetic epidemiologists.

Implications for research, considering ethical aspects, public health aspects and organisational aspects in the field of genetic and environmental respiratory epidemiology also need to be addressed.

*Eur Respir J 2004; 24: 471–480.*

Correspondence: F. Kauffmann  
INSERM U 472-IFR 69 Epidemiology and  
Biostatistics  
16, Avenue P.V. Couturier  
F-94807 Villejuif Cedex  
France  
Fax: 33 145595169  
E-mail: kauffmann@vjf.inserm.fr

Keywords: Epidemiology  
ethics  
genetic epidemiology  
genome  
lung diseases

Received: July 4 2003  
Accepted after revision: April 5 2004

This work was initiated in January, 2002, at a European Respiratory Society research seminar on this topic in Cernay, France.

Will genetics overwhelm the field of epidemiology? Should epidemiology maintain its focus on environmental and public health issues in this changing world? Is there a need for very large population studies? Whereas the shared interests between geneticists and clinicians may be pathophysiology, *i.e.* interest in pathways, the shared interests of geneticists and epidemiologists may well be design and analytical issues, as developed by genetic epidemiologists. One major problem is the belief that one discipline is too complicated to be understood and used by those outside the field, and the subsequent lack of confidence in those from other disciplines. A major aim of the present review is to list relevant questions pertinent to the setting up of multidisciplinary research involving respiratory epidemiologists, geneticists, environmental epidemiologists and clinicians, a topic of increasing relevance with the current development of national and international platforms. After describing specific questions pertinent to the field of asthma and chronic obstructive pulmonary disease (COPD), two main methodological aspects regarding relevant technological and scientific advances are presented. The first one concerns biological aspects in the genome and post-genome phases, *i.e.* how to study the genome, the transcriptome and the proteome. The second area concerns genetic epidemiology, with aspects related to design (case control and family based) and statistical analytical issues. In addition, implications for research, ethical considerations, public health and organisational aspects in the field of genetic and environmental respiratory

epidemiology will be presented. In conclusion, key points are summarised, and challenges and needs for the next steps in respiratory epidemiology for the genome and post-genome phases are discussed.

## Current challenges in the genetics of asthma and COPD, and in respiratory epidemiology

A rigorous approach to the study of the genetics of asthma and COPD is needed in which all aspects of the aetiology of such complex multifactorial diseases are identified. Appropriate phenotypic characterisation of asthma and COPD is crucial. The heterogeneity of the effects of the environment (short-term effects, long-term effects, cohort effects, cumulative exposure, protective and deleterious effects, *etc.*) needs to be considered. Questions arise regarding ethical issues in relation to the potential identification of individuals particularly susceptible to disease, mainly resistant to a specific hazard. Detailed reviews on the genetics of asthma and COPD may be found elsewhere [1–9].

Genetics are studied to understand underlying disease mechanisms, to define subgroups of disease, to identify individuals at risk, to prevent disease and to define new targets for treatment. Genetic epidemiology is studied because the current state of knowledge is inadequate, there are many

polymorphisms and many inconclusive results from small selected samples. Large samples in population-based studies are required in order to study disease subtypes, gene-gene interactions and gene-environment interactions [1, 2]. The nonrandom combination of alleles at different loci on a chromosome (linkage disequilibrium), which depends on the genetic distance of the loci, the age of the mutation at either locus and the history of the population, complicates the study of relevant genes in population samples. In general, this means that several single nucleotide polymorphisms (SNPs) need to be studied simultaneously for each gene under investigation.

It is likely that the genes that predispose to asthma have in the past been selected to be common in the population. The situation of Australian Aborigines serves as a good illustration of selection pressure. In this population, which is highly infected with parasites, the total serum immunoglobulin E concentration is negatively related to skin-prick test positivity. The nature of genetic polymorphisms differs from those of Caucasian Australians and has stronger associations to atopic phenotypes. In Europe as well, the population geography may contribute much of the overall observed variation in disease phenotype. Europe contains many environments and has a nonuniform genetic structure. Genetic polymorphisms in Europeans show signs of the spread of farming from the Middle East between 10,000 and 6,000 yrs ago, followed by the probable spread of Uralic people to the North East of Europe and the spread of pastoral nomads (and their successors) who domesticated the horse in the steppes and appeared towards the end of the farming expansion [10]. The genetic gradients remaining from these changing populations will be relevant for the allele frequency and presence of asthma genes, and will add to the complexity of gene identification. Examples of haplotypes that may vary markedly between populations include the  $\beta_2$  adrenergic receptor and the genes of the major histocompatibility complex.

Most known genetic influences on asthma have been identified by the study of candidate genes, *i.e.* genes of known function. Several groups are now attempting to identify new genes by positional cloning. This laborious process involves three steps: 1) the study of genetic linkage (*i.e.* the co-inheritance of disease and a chromosomal region); 2) fine mapping of a region of interest; and 3) gene identification by sequencing and the study of gene expression. Linkage studies correspond to a group of methods that analyse the co-inheritance of DNA markers within families, to determine if a particular region of the genome contains a gene related to the phenotype of interest. The lod score which is often used in such studies is the result of a statistical test used to determine if genetic loci are linked, expressed as  $\log_{10}$  of the odds of the observed data under two hypotheses, linked *versus* nonlinked. Linkage studies by genome-wide screens have revealed that linkages to some chromosomal regions coincide for asthma and other immune disease, and that there are also organ-specific coincidences, *e.g.* for atopic dermatitis and psoriasis. Such observations suggest that genes affecting disease expression may be concentrated within a few chromosomal regions, that there are clusters of genes with general effects, *e.g.* influencing dermal inflammation and immunity, and that the atopic component of atopic dermatitis may have a secondary rather than primary influence. In asthma, positional cloning has been successful, in particular, on chromosome 13 [11], and has allowed the identification of a disintegrin and metalloprotease 33 (ADAM33) on chromosome 20 [12] and dipeptidyl peptidase 10 (DPP10) on chromosome 2 [13].

Although cigarette smoking is the major environmental risk factor for the development of COPD, there is marked variability in the development of airflow obstruction in response to smoking. The frequent development of COPD in individuals

with severe  $\alpha$ -1-antitrypsin (AAT) deficiency (*e.g.* protease inhibitor (PI) phenotype Z), a proven genetic risk factor for COPD, has provided a foundation for the protease-antiprotease hypothesis for the pathogenesis of emphysema. Although only a small percentage of COPD patients (estimated at 1–2%) inherit severe AAT deficiency, AAT deficiency can serve as a model of the manner in which genetic and environmental factors may interact to lead to COPD [14]. Additional genetic determinants, which have not yet been identified, probably influence the variable development of airflow obstruction in PI Z individuals. Besides PI, a variety of association studies have compared the distribution of variants in candidate genes hypothesised to be involved in the development of progressive and irreversible airway obstruction in COPD patients and control subjects [15]. For each of the candidate genes, at least one study refutes the association. Several factors could contribute to the inconsistent results of case-control genetic association studies in COPD, including genetic heterogeneity and population stratification. In case-control studies, observed associations of gene polymorphisms with disease may be due to: 1) a direct effect of a candidate gene; 2) linkage disequilibrium, that is an association of susceptibility gene and genetic marker (between families, at population level); and 3) population stratification, resulting from incomplete matching between cases and controls, including confounding by ethnic background. In the Boston early-onset COPD study [16], a variety of phenotypes that demonstrate smoking-related susceptibility in first-degree relatives of early-onset COPD probands were identified including forced expiratory volume in one second (FEV<sub>1</sub>), FEV<sub>1</sub>/forced vital capacity, bronchodilator responsiveness and chronic bronchitis. Linkage analysis results on both qualitative and quantitative spirometric phenotypes show that regions suggestive (*i.e.*  $p < 7.4 \times 10^{-4}$ ) [17] for linkage for COPD or spirometric values were located on chromosomes 1, 2 and 12 [18, 19].

The role that genetics could play in the understanding of the aetiology of asthma is a matter of debate. The allergen hypothesis does not appear to explain the global increase in asthma prevalence. This hypothesises that the primary cause of asthma (and, hence, of the global increases) is that allergen exposure causes atopic sensitisation, which, in turn, induces bronchial hyperresponsiveness (BHR) and asthma. In fact, BHR is not a good surrogate measure of asthma prevalence, and is more relevant to specific mechanisms of asthma [20, 21]. The proportion of cases attributable to atopy is <40% in both children and adults [22]. There is not enough evidence to conclude that allergen exposure early in life is a major primary risk factor for developing asthma, as insufficient cohort studies have been conducted to demonstrate that link and, in fact, several studies have found no association. An alternative hypothesis is that changes in susceptibility could be occurring through better hygiene, producing reduced infant infections, and thus a reduced T-helper (Th)-1 cytokine response and an increased Th2 immune response. Key features of westernisation, which have been suggested to play a role, include small family size, reduction of infections in early life (particularly gastro-intestinal), large size at birth and changes in diet (with more fatty acids and fewer vegetables). However, replacing the allergen hypothesis with the hygiene hypothesis may be replacing one dogma with another. The hygiene hypothesis is conceptualised within the allergen Th1/Th2 paradigm, which appears to account for at most one out of two asthma cases [22–24]. Even this may be an overestimate because the association may not be entirely causal and sensitisation may be a consequence of an asthmatic predisposition. If the package of changes involved in increasing affluence and westernisation can account for the

increase in asthma prevalence, it does not necessarily invoke classic atopic (Th2) mechanisms [25].

Genetic epidemiological studies may help in understanding the origin of the disease. The greatest potential for population-based genetic interventions occurs when the disease is known to occur through a single well-defined mechanism, when only a small number of genes are relevant to this mechanism and when there is little temporal or geographical variation. Asthma satisfies none of these conditions. It occurs through multiple mechanisms that are not well understood; many genes appear to be related to different aspects of the main (allergic) mechanism that has been studied to date, and there is major temporal or geographical variation that cannot be explained by genetic factors. To date, asthma genetics has had limited success in explaining a significant proportion of asthma cases and even less success in explaining the population patterns. The lack of replication does not mean that the observed associations are invalid, but it does limit their usefulness in both scientific and public health terms. The expression of asthma as the increase of asthma and allergy in the last 30 yrs could be primarily ascribed to environmental changes. Genetics cannot account for the global increase in asthma prevalence. Some considered that the importance of genetic factors for health has been overestimated, and that the ethical and practical problems in applying genetic knowledge to interventions have been underestimated. Searching for environmental causes of asthma is likely to yield useful and useable results, whilst the study of gene-environment interactions could play an important secondary role. A simplistic approach would be to ignore genetics, although this would not serve the epidemiological community well. To get to the roots of the disease will be so complex that all of the skills of the molecular geneticist and the epidemiologist will be required, together with work on animal models. It is important that new theories of asthma causation, including nonallergic mechanisms, are developed and tested. The overall goal should be to incorporate more complex paradigms in order to disentangle the aetiology of such a complex disease or group of diseases. Subtypes should be considered; genetics have shown that diabetes is no longer one disease. Intermediate phenotypes are now popular, but difficult to define, because the definition of intermediate pathways is debatable. From the point of view of finding genes, the most promising intermediate phenotype would be one that is the closest to gene, in order to find linkage.

### Perspectives of the genome and post-genome phases

Post genome refers to the period after the establishment of the complete human genome sequence. The true post-genome phase is yet to come. The knowledge of the genome sequence is only a first step, the search for SNPs and haplotypes within a given gene now requires a phase of resequencing, as a result of errors in the currently available maps. Genes need to be characterised by their sequence, their function, their regulation and polymorphisms. In the post-genome phase, both genetic epidemiology and technological considerations will play a key role. Phenotyping may be improved in the future through the study of the transcriptome and proteome.

The sequencing of the human genome has shown that there are fewer genes than anticipated. The drafts already available, whilst rendering genetics more accessible are imposing other aspects of complexity, as a result of the smaller than anticipated functional genes. A current challenge is to use this knowledge for epidemiological applications. The requirements for genetic epidemiology include the availability of DNA from large populations of patients, family members

and controls, high quality phenotypes, and clinical data and methods for identifying gene or whole genome variation by fast, low-cost and flexible genotyping. Technically, numerous possibilities already exist. SNPs show variations in the genome in every few hundred bases [26–28]. A large proportion of putative SNPs in public databases have not been verified and much more information is required on allelic frequencies in different populations. Assessing only two polymorphisms in the coding region of each gene will not provide sufficiently dense information. The frequency of the variants is heavily dependent on the population and, in this regard, information on the population distribution is critical. Haplotype pattern information may be defined with a small number of SNPs [29–32]. The sites within a gene that are common between different haplotypes and may be related to a quantitative phenotype (such as serum angiotensin converting enzyme (ACE) for the ACE gene) may provide critical information on the causal variant.

Epidemiological applications in large populations raise technical challenges after SNP discovery. Mass spectrometry is based on primer extension assay products of different molecular weights run on a matrix and separated by their size. Its strengths include speed (2 s per analysis), the possibility of analysis of large numbers of individuals, the high quality due to absolute result, the miniaturisation of sample preparation (0.5–2 ng DNA per reaction) and the flexibility for the choice of SNPs. Future technological trends include electric field arrays, which are good for small number of individuals and SNPs, oligonucleotide ligation, which is good for large numbers of SNPs, and period-based technology and mass spectrometry, which are the methods of choice for a large number of individuals. The most appropriate and valid methods for both large numbers of SNPs and individuals are not yet established, and it is an intensive area of research [33, 34].

The study of the transcriptome, through cDNA arrays, is difficult but technically feasible and provides a new area of quantitative phenotypic characterisation. Genome-wide gene expression profiling using DNA arrays is the equivalent of performing tens of thousands of Northern blots [35, 36]. Recent developments in this technology make it possible to investigate interactions of a large number of molecules with a large gene library, thus providing fundamental insights into biological processes ranging from gene function to development, cancer, ageing and pharmacology. These techniques offer the possibility of a shift from small-scale, low-throughput molecular biology to a functional genomics approach. It will help further understanding of biological processes, genetic networks and living systems by using large-scale systematic, high-throughput technology, either in a purely discovery mode or in combination with traditional hypothesis-driven approaches. Systems biology is an emerging field that aims at system-level understanding of biological systems, which extends beyond the static description of their isolated parts and basic mechanisms based on a reductionist approach. It enables a systematic description of the entire sets of DNA, RNA and functional molecules of complete biological systems (genome, transcriptome and proteome) and their metabolic interactions (the interactome and metabolome).

Proteomics is the study of whole patterns of changes in protein expression and their modifications. The situation regarding proteomics is promising but even more difficult from a technological viewpoint than genomics. Whereas there are ~30,000 genes, there may be 400,000–1,000,000 proteins [37]. Generating hypotheses approaches and candidate (*a priori* hypothesis) protein approaches are both of interest. Technological aspects are still major considerations and have not yet been solved. The human proteome project will not be

deliverable without access to high-affinity ligands, produced at genomically relevant cost and time. There is a need for an enormous increase in throughput and reduction in the cost of antibody development and ensuring that high-throughput array avoids cross reactivity. It is anticipated that in the next 5 yrs, the cost will decrease by a factor of 100, thus making the technology more accessible. Specialists aim to develop a matrix capable of following the protein output of each and every gene within the human genome. An aspect familiar to epidemiologists is the dynamic nature of the human proteome, which varies according to environmental characteristics and the time of the day. Potential future applications include so-called patient "cohorting", which aims to characterise patients in order to develop markers of drug rejection, prognosis, early diagnosis and disease management, *etc.* Understanding biological networks is a challenge. Tools are currently only available for <5% of proteins but practical applications may be expected in the next few years in this area as a direct application of knowledge gained from sequencing the human genome [38].

Unsolved questions relate to the validity and quantitative accuracy of the observed changes including the precision and reproducibility of massive amounts of data, and the inferences on levels of gene expression that can be drawn from the hybridisation signal intensities of cDNA probes to gene expression levels. A central issue is the choice of genes, which should be the priority for further study. Studies of single genes and clusters of genes, analysed in terms of common functionality, interaction, coregulation and, ultimately, the understanding of protein networks through their underlying genes, may be informative. Setting standards will be essential in order to facilitate the comparability of gene expression data from different sources, and ensure the compatibility of different gene expression databases and data analysis software.

The limiting factor in future studies is likely to remain the phenotype rather than the genetic technology. Other aspects to be considered are quality control issues, including background noise *versus* signal, the need of replicates in relation to inter- and intra-subject variability, the number of samples studied, and protection of results and the sharing of data (and methods) in the context of public or private partnerships. The amount of variability concerned at the proteomic level is enormous, hence the challenge in applying these techniques. Issues of validation at the population level are critical.

### Genetic epidemiology: designs and analyses

The classical case-control design has entered into a new phase in genetic epidemiology as a consequence of the development of large-scale facilities for genotyping (table 1). Furthermore, for diseases such as COPD, for which age of onset is generally late, family-based designs have limitations. New types of case-control designs in the context of the search for genetic factors have appeared. Although well known among genetic epidemiologists, such designs could be used by respiratory epidemiologists without extensive training. Contrary to classical linkage analyses, association studies are intuitive to epidemiologists familiar with associations, selection, confounding and interactions in the context of environmental factors.

#### Case-control design

Case-control studies in the context of post-genome epidemiology are generally similar to classical ones, with some specificities for the post-genome context. In the selection of

Table 1. – Key points

Design issues	Case control, simple statistics but population stratification
	Case-only design for studying interaction
	Trios (TDT) based on siblings according to Mendelian inheritance: simple statistics (Mac Nemar test), linkage and linkage disequilibrium, avoid population stratification but overmatching
	Need for replication
Phenotype	Essential to be of good quality, problem of measurement error
	Intermediate phenotype (strength: quantitative, close to gene; limit: cost)
	Subphenotypes for investigating pathways
Environment	Essential to understand the aetiology and address public health issues
	Protective and deleterious
	Interaction with gene needs hypothesis before testing (pathways)
	Distinguish statistical and biological interaction
	Consider windows of opportunity
Large sample size	Subjects $n > 100$ per group for adequate power in case control studies, otherwise spurious results, debate about the usefulness of very large numbers
	200 families for TDT for one gene
	Minimum allele frequency, <i>i.e.</i> consideration of allele of epidemiological magnitude, such as 2%
Time	Related to phenotype (onset, severity)
	Related to environmental exposure (cohort effect, age (lifespan))
	Timing of environmental exposure prior to effect
	Window of expression for genes
	High throughput for studying genome, transcriptome, proteome
Ethics and public health	Consider to incorporate genetic aspects at the group level (issues of mutuality, consent, banking, communication, commercialisation)
	Primary, secondary prevention
	Relevance for public health

TDT: transmission disequilibrium test.

cases and controls population stratification has attracted a lot of attention. An unmeasured confounder needs to be extremely strong in order to change a relationship and, therefore, it is likely that the importance given to population stratification is somewhat exaggerated [39], an aspect recently highlighted [40]. As population stratification may affect a range of polymorphisms, any effects can be minimised by using a very large number of unselected markers (genome-wide control method [41]) or by selecting markers involved in the population substructure and employing a latent class analysis [42]. Close attention to study design may also correct for population stratification. Matching on ethnic background is not always obvious, as it can be difficult to define ethnicity when it is complicated by intermarriage in multi-ethnic populations. Group matching is the usual technique used, the extreme form using sibling controls with analysis by the sibling transmission disequilibrium test (S-TDT). However, because siblings are more similar genetically than unrelated subjects (overmatching), it has been shown that twice the sample size is required when using sibling controls rather than unrelated controls.

Measurement errors have major effects, even when they are nondifferential, *i.e.* unrelated to disease or exposure. The need to properly measure disease cannot be overemphasised and a particular type of measurement error is when the locus associated with disease is not itself a functional variant and, as a consequence, a flawed measure of true genetic risk will appear. Assessment of interactions is also hampered by

nondifferential measurement errors. Statistical approaches correcting for these distorting effects have been proposed [43], and involve integrated validation, test-retest reliability and substudies, although they still have to assume that repeated measurements of risk factors are conditionally independent given their true values.

There are similarities between the analyses of genetic and environmental factors. New approaches, such as DNA pooling are similar to ecological studies. Analyses regarding genetic factors are easier than for environmental ones from a certain point of view. Genes are time-independent for a given individual and nature delivers genes in a balanced, factorial design.

### *Family-based studies*

Design issues of family-based studies [9, 44–46] and the concepts behind the methods of analysis of such designs [47] should first be considered. For the identification and description of susceptibility genes for complex diseases using genetic markers two basic strategies are employed: genome scans and the investigation of candidate genes [48]. Candidate genes should be identified on the basis of their function. Clear ("narrow sense") candidates are genes with known function or structure related to the disease and to identified abnormalities, or to relevant animal models. Less clear ("broad sense") candidates are genes that are part of a biological system with more speculative influences on the disease. They can be of interest, especially when they are in a region that has suggested linkage on previously reported genome scans or when based on data from expression profiling in relevant tissues. Susceptibility genes for complex diseases can be roughly subdivided into three classes: 1) major genes (generally rare, highly penetrant and the main research focus in the past); 2) oligogenes (that might be quite frequent in the population, which contribute a moderate risk that might become higher in combination with other risk factors); and 3) polygenes (contributing only a small effect and with many required to influence disease expression). Current interest in public health research is now focusing on oligogenes. Genetic epidemiological studies using genetic marker data for the investigation of susceptibility genes have two principal goals [48]. The first goal is the identification of the risk factor susceptibility gene, which can be subdivided into localisation and provision of evidence for the influence. The second goal is modelling the role of this risk factor.

The goal of the genome-wide search is the localisation of a susceptibility gene by linkage analyses, using genetic markers. Genetic markers are parts of the DNA with known and unique locations, and can be used to identify the different alleles (subtypes) and thus an individual's genotype. Linkage is the cosegregation of marker alleles and disease in relatives [49]. A very common design is the investigation of ~350 markers evenly distributed on the total genome (average marker distance 10 cM) in families with two affected siblings and their parents (affected sib-pair design). More generally, designs for linkage include large families with many affected, 2–3 generations with multiple affected, families with affected sibs, families with one affected child and the study of isolated populations. Scores derived from qualitative and quantitative traits may be used as phenotypes, and, in the case of large cohorts, consideration of extreme phenotypes may be of interest. The major advantage of a genome-wide scan is that no biological mechanism must be known or assumed. However, because of the multiple testing of hundreds of markers, a balance between false-positives and false-negatives may be found [50], and statistical significance is, therefore, difficult to

attain. Regions suggestive for linkage according to a genome scan can then be used as a starting point for fine-mapping, including the investigation of possible candidate genes in this region.

In the traditional case-control design, it can be difficult to ensure the quality of the control sample since controls must be from the same genetic background as the cases. To avoid spurious findings as a result of population stratification, family-based controls are sometimes used. The principle of TDT [47, 51, 52] is based on parental control alleles. The TDT, equivalent to a McNemar's statistic familiar to epidemiologists, simultaneously tests for linkage and gametic disequilibrium. Independent of the issue of population stratification, family-based study designs may be preferred as they can sometimes be easier to implement than traditional case-control designs. Two examples are when cases are young and parents are easy to recruit, and when the cases represent an ill-defined target population, so that an appropriate population control group is difficult to define, therefore, making parental or sib-controls an appealing alternative. It also allows use of existing collections of families. The case-parent trio design is based on sampling affected cases and their parents, measuring phenotype information on only the cases, but measuring a genetic marker for the cases and all parents. Each parent is classified according to which allele is transmitted and which is nontransmitted to the affected child. The alleles not transmitted to the affected child are used as a family-based control ("pseudo-sib"). Only heterozygous parents can yield information. Another powerful genetic epidemiological perspective is that the case-parent trio can be viewed as a matching of a case with three "pseudo-sib controls". Although traditional conditional logistical regression can measure the main effects of genes and gene-by-environment interactions (*i.e.* whether allele transmission depends on the child's exposure to environment factors of interest), it cannot measure the main effects of environmental exposures, as it is a case-only design and environmental factors are measured only in the affected children. The relative risk can be biased towards one, because it is assumed that the pseudo sibs are never affected, which is rarely true for a common disease. The true relative risk may also be greater than the observed odds ratio if there is recombination or incomplete linkage disequilibrium. Another potential source of bias occurs when a marker is in linkage disequilibrium of a lethal foetal gene. This bias can go undetected because allele transmission is compared to expectation of Mendelian transmission (*i.e.* there are no true controls). Therefore, caution is necessary in the interpretation of case-parent trios designs. The power of the TDT approach depends on the number of heterozygous parents and on the relative transmission probabilities of the marker alleles. Two hundred families might be appropriate for a wide range of genetic models to detect a locus with genotype relative risk of  $\geq 2$  with a significance level of 5% and a power of 80%, but ~600 families may be required when a wide range of genetic models are to be investigated.

Sib controls are a reasonable alternative to parental controls, particularly when parents are not available, such as when cases only express the disease of interest in late adult life. The most efficient way to analyse a case with sib controls is with a matched design, where cases and their sibling controls define a matched set, analysed using traditional conditional logistical regression models. Sibling controls are not biased by segregation distortion, but they may be difficult to match for age and sex. They are informative if their genotypes differ, although the design still suffers from the problems associated with overmatching. Larger pools of cousin controls matched on age and sex could be used, but potential biases can appear. Therefore, pseudo-sibs remain

the most efficient controls for rare recessive major genes [53]. Regarding efficiency, population controls have similar efficiency as pseudo-sib controls, and both types of controls are more efficient than cousin controls and much more efficient than sib controls. Siblings are less efficient than population controls for gene main effects due to overmatching, and generally require twice the sample size as for population controls. However for gene environment interaction, siblings are more efficient than population controls as overmatching on genotype helps to find significant gene-environment interactions. Although families may be collected for a study, there can still be very good reasons to sample controls from the population, in order to compare the familial multiplex cases *versus* the population controls.

In conclusion, weaknesses of family-based studies include "overmatching" genetic background (a big problem for sib controls and major gene effects), loss of cases when there are no eligible controls, nonpaternity and more difficulties in recruitment than traditional case-control studies. Strengths of family-based studies include the ability to control population stratification, efficiency when parents are available (young cases) and efficiency for gene-by-environment interactions. Family-based studies also have genetic benefits, namely inference on haplotypes (more informative than single SNPs), checking genotype errors (Mendelian inconsistencies) and detection of parent-of-origin effects (imprinting). It is important to note that, in the context of a family-based study, the genetic relative risk has a different meaning than in association studies, which compare cases and controls. An attributable risk cannot be assessed using families; a case-control study is required.

### *Gene-environment interaction*

Before the current interest in gene-environment interaction, interactions between environmental factors had been the subject of numerous discussions. Three important forms of "interaction" appear in genetic epidemiology, including between two copies of an allele on an autosome ("dominance"), between two loci ("epistasis"), and between gene and environment. Since the original description by Fisher in 1918, statistical interaction (which is scale dependent and only corresponds to deviation from additivity of the effects of two factors) sheds little light on biological mechanisms, as recognised by many epidemiologists by the late 1980s. For the joint effects of two risk factors, there are numerous possible null hypotheses for independence of a causal action [54]. Targeted interventions are often aimed at risk subgroups, although whole population interventions may be more effective [55]. In the presence of a strong interaction with a known risk factor, a test which allows for such an interaction may have more power to detect a novel factor, although the gain in power is modest unless there is a reversal of the direction of the effect. New designs include case-only studies, which rely on the Mendelian randomisation between genotype and environmental cause or between two genotypes within the population. When the interest is only on interaction, it can be simply studied by departure from the multiplicative model.

Gene-environment interactions are a matter of debate at a conceptual level. Often it is viewed as an antidote to determinism, with the idea that some interactions may counter-balance genetic risk in subjects identified at high genetic risk. Arguments to study interactions include biological plausibility, improvement in specificity, increased strength of such interactions in susceptible groups, and reduced likelihood of bias and confounding. Arguments against the study of interactions are their scale dependency, the problems of

multiple comparisons and the large samples needed for adequate power. In the 21st century, there will be increased focus on effect modification, rather than association, as the basis for causal inference. Success can be anticipated in diseases where synergistic biological interactions are already suspected whereas success in common multifactorial diseases is less likely and will depend on aggregating susceptibility. Functional traits may be more informative than genotypes for this purpose. The impact of ignoring gene-by-environment interaction on the detection of complete linkage disequilibrium may hamper the identification of the potential function variant. It is important to use models that take into account both linkage disequilibrium and gene-by-environment interaction in order to disentangle the mechanisms underlying complex diseases.

Illustrations of potential gene-by-environment interaction in the context of asthma and COPD epidemiology may be found by looking at the main hypotheses for asthma and COPD in the epidemiological literature [9]. The British, the Dutch and the protease-antiprotease hypotheses for COPD, and the hygiene hypothesis for asthma provided interesting examples. The British hypothesis may be considered an environmental hypothesis and the Dutch hypothesis a gene-(host factor, represented by asthma, BHR and allergy) by-environment (*e.g.* irritant) interaction. The protease-antiprotease hypothesis starts as a genetic hypothesis (PI) and ends in the hypothesis that both genes (PI) and environment (smoking) act on the same physiological pathway. Finally, the hygiene hypothesis has moved from an environmental hypothesis (contact with infectious agents), to an interaction between native immune pathway genes and exposure to endotoxin. Interactions of PI and smoking [56], CD14 and toll-like receptors in the context of the hygiene hypothesis [57, 58] may be found. Interactions with oxidant exposure, such as smoking [59], air pollution [60] or occupational exposure [61], and with aspirin [62] have been evidenced as well as gene-by-gene interactions [63]. In conclusion, it has become essential to consider all the genes in a pathway, and due to the number of potential candidate genes already identified, there is a need for well-founded gene-by-environment hypotheses and their subsequent testing (table 2). Potential conflicts between disciplines due to their internal logic need to be addressed. The consideration of pathways and time effects is centrally important. The history of the individual's environmental exposures, the history of disease expression, which led to consideration of intermediate phenotypes, the phenotypes of the disease itself, as well as phenotypes describing severity and exacerbations. Respiratory epidemiologists are well aware of issues of selection and confounding that were just as relevant to genetic epidemiology, but less aware of the potential of family-based designs. Conversely, biologists and geneticists often raise methodological questions which had been previously addressed (and possibly solved) in the context of environmental factors. Clearly, more multidisciplinary work would be of mutual benefit for the full range of disciplines involved (table 2) in further studies of the origins of asthma and COPD, and the potential for disease prevention.

### **From genetic epidemiology to public health**

Studying the genetics of complex diseases is a multidisciplinary enterprise. Studying families can allow gene identification by positional cloning for monogenic diseases, with the major role devoted to human geneticists, along with an important contributing role for genetic epidemiologists, although a lesser role for clinicians. Positional cloning is still a

Table 2. – The challenge of respiratory epidemiology in the genome and post-genome phase

---

Building studies on professional approach for phenotypic characterisation, environmental characterisation, choice of design, choice of genes, technical aspects of genotyping and statistical analysis and standards setting
Large sample size (power issues for genetic aspects considering the whole genome and gene-environment interaction, technological advances, new issues regarding ethical aspects, public health aspects)
Networking required for large sample size, study replication, variability of environment and genetic background between countries
Combining exploratory approaches (finding new genes, new proteins) and hypothesis-driven research, especially focused on physiological pathways
Building capacity to react quickly to new genetic developments
Preparation for near and distant future for new phenotypic characterisation through expression profiling based on transcriptome and proteomics
Need of high-level informatic specialists
Training in genetic analytical techniques for some respiratory epidemiologists, a challenge but of moderate difficulty for those already with a classical epidemiological training
Training in respiratory epidemiology for some genetic epidemiologists, a challenge to motivate them on phenotypic characterisation issues, importance and complexity of environmental factors

---

paradigm for complex diseases, but it needs a different structure. Debates between scientists interested in function (gene-driven research) and those interested in disease aetiology (disease-driven research) are constructive and these two approaches need to be combined. Geneticists are no longer in the position to set up a project on their own, and the competence of genetic epidemiologists and clinicians is necessary. From the genetic epidemiologist's perspective, environmental contributions fall into the remit of molecular epidemiology. For example, a National Genomics Research Network has been established in Germany with disease-oriented networks and several platforms, amongst which there is one for high throughput genotyping, and one for design, data management and analyses called the Centers for Genetic Epidemiologic Methodology. It is only by fulfilling the need for competence and expertise in phenotypes, genotypes and in methodology that there is any chance of solving some of the genetic components of the aetiology of complex diseases (table 2).

The Human Genome Epidemiology (HuGE) initiative is a collaborative effort of individuals and organisations that are committed to the development and dissemination of population-based genome data worldwide [64] with a public health perspective. Potential applications include primary prevention (such as aiding at understanding the biology of diseases, potentially leading to the development of strategies based on mass intervention or target risk intervention approaches), secondary prevention (such as the development of primary tests for population-wide screening) and therapy (such as choosing among alternative interventions) [65–67]. Relevant information includes: 1) prevalence of gene variants; 2) assessment of the contribution of gene variants to the burden of disease and disability (penetrance and attributable fraction); 3) the magnitude of disease risk associated with gene-gene and gene-environment interactions in different populations and relevant to targeting intervention; 4) the clinical validity and utility of genetic tests; and 5) the determinants and impact of genetic tests and services. Epidemiological data on their own are clearly not enough, and it is important to develop policies to integrate genetic testing into medical or public health programmes for public health professionals and policy makers, researchers, healthcare providers (primary and specialists) and consumers. HuGE Net aims at developing information exchange, the knowledge base (with reviews on specific aspects), training and policy.

The term epidemiologist encompasses several dimensions, with the respiratory epidemiologist, the risk factor epidemiologist, the environmental epidemiologist and the genetic epidemiologist. From a geneticist's point of view, epidemiologists and clinicians are often viewed as collectors of good family cohorts (*i.e.* providing phenotypes and DNA) for groups with

genotyping facilities. One of the challenges of the post-genome period is to understand and get the appropriate training that will ensure a true integration of the various facets of the work of an epidemiologist interested in one given disease (table 2).

### Ethics

Genetic epidemiology, and DNA sampling and have banking raised new ethical questions. Presently, four different kinds of banks exist according to their mode of constitution: 1) those built through research activities in universities and hospitals; 2) those established following legislative decisions, such as in Iceland and Estonia, where whole countries have decided that DNA is a national resource to be exploited in their country; 3) commercial banks established by pharma or by start-up companies with the intent of commercialising results of research; and 4) virtual banks corresponding to entities which are soliciting DNA donors *via* the worldwide web and sending the results to individuals or through physicians.

Furthermore, it is important to point out the increased interest in the DNA of heterogeneous populations as opposed to high-risk populations, or of isolates (in population terms). The target has changed with the mapping of the genome and the study of SNPs in large populations [68]. This implies, in turn, that ethical aspects not only have to consider the issues raised by the typology of emerging banks, but also have to move away from the monogenic model of research in homogeneous populations. Most ethics codes have concentrated on individual rights (autonomy, consent, discrimination, confidentiality, *etc.*). The guidelines proposed no longer completely respond to the new issues raised by population-banking studies. The development of normative frameworks appears necessary. Some of the new relationships between research, populations and society have already been addressed by bodies such as the United Nations Educational Scientific and Cultural Organisation, World Health Organization, Human Genome Organisation, and Council for Internal Organizations of Medical Sciences [69–71].

At least six important issues need to be addressed as follows. 1) Recruitment. 2) The principle of mutuality *i.e.* involving healthy populations and communities with no direct immediate benefit and consent to participate. Finding a way for the community to express itself in such a context is needed [72]. Anonymised samples generally have limited use, thus it may be possible that in the next 10 yrs coded DNA could be used in an ongoing fashion provided there is ethical review. 3) Confidentiality as individual anonymisation or coding cannot mask the issue of genetic susceptibility profiles, which can

emerge on a group or a region. 4) Banking has to be considered prospectively due to its sociological, psychological and cultural implications. In particular, policy clarifying aspects such as the change in principal investigator, lack of funds, the structural aspects of the institution and responsibility issues should be transparent. 5) Communication of results needs to be clear upfront. 6) Commercialisation issues also need to be addressed [73].

Moreover, while people generally cannot own their DNA, they do exercise control, and the institution, custodianship. Transformation of the DNA adds intellectual know-how to the sample. Patenting can only be on what is novel, inventive and has economic utility. There is ongoing debate between countries on how to distinguish between discovery and invention. Intellectual property rights may have commercial value and that should be clear from the outset. The notion of shared benefits for the people as a group has also emerged in the last 5 yrs. Current discussions centre on whether populations get recognised for their contribution or receive something back as immediate investments in public health or for humanitarian purposes [72]. It is in that context that some countries have requested to co-patent the DNA, whilst others request that DNA should always remain in the country of origin or only leave the country after approval by a central representative body. The post-genome phase has raised various new aspects regarding confidentiality, traceability, commercialisation and patenting, feedback to the subjects, and differences between countries, which all need to be addressed.

### Requirements for the inclusion of genetic research in respiratory epidemiology

The relevance of incorporating genetics in respiratory epidemiology is not obvious to all, although it is generally agreed that new concepts and approaches need discussion before putting them into practice. More information, training (both of respiratory epidemiologists and geneticists) and discussion appear necessary (table 2).

From the epidemiological perspective the importance of valid case-control designs for association studies should be emphasised, which is within the area of expertise of the epidemiologist, and hence reassuring. Geneticists could be reminded of the central importance of good phenotypic characterisation, which is strongly grounded in respiratory epidemiology. The need for large sample size and replication, and the concept of population stratification (one type of confounding) are now well understood. Future studies should systematically collect information on the origins of individuals participating in such studies. Building nested case-control studies within cohorts and getting ready to react quickly to new genetic developments, as they became available would be efficient. Defining standards of study, thereby allowing the full realisation of the opportunities within existing cohorts, has been suggested. Improvements in organisational aspects that respect the existing framework of various studies are needed to encourage networking and possible aggregation of data banks, including recognition of the work requested to set up biological collections [74]. In the field of respiratory epidemiology, there is already an excellent foundation with regard to standardisation, as evidenced by initiatives such as the British Medical Research Council questionnaire for COPD, the European Community Respiratory Health Survey, and the International study of Asthma and Allergy in Childhood studies for asthma.

Whilst not forgetting the major importance of environmental factors, and in particular of smoking in the context of COPD, there is now a consensus among epidemiologists that genetic

research merits serious consideration. To understand the complexity of environmental influences, it is necessary to take into account the pertinent window of exposure for a disease with a variable expression over time. A real strength of the respiratory epidemiological community resides in methodological aspects, as many of the complex issues raised in genetic research have a great deal in common with issues and challenges in the design and analysis of traditional environmental epidemiology. New possibilities of collaboration could emerge between environmentalists and geneticists through the high level in statistical modelling of epidemiologists, with a current focus on environment. Design and statistical aspects are probably the bridge between the respiratory and genetic communities, at their common interface with genetic epidemiology. Training epidemiologists in these fields will provide the knowledge for deciding whether it is the time to incorporate the "new genetics" into epidemiological research and public health, and, conversely, training of basic scientists and clinicians in epidemiological methods is a necessary step to develop true multidisciplinary research. International networking, in particular at the European level, is now emerging with the support of various initiatives of European scientific societies, in particular for asthma and allergic diseases, and should provide the framework for the challenges of post-genome respiratory epidemiology [75, 76].

**Acknowledgements.** The authors would like to thank all those who made the Cernay seminar possible. It was organised by the European Respiratory Society and supported by the International Epidemiological Association (IEA), the International Genetic Epidemiology Society (IGES) and Institut National de la Santé et de la Recherche Médicale (INSERM). They also thank all those who contributed with insightful ideas during and after the seminar.

### References

1. Cookson W. The alliance of genes and environment in asthma and allergy. *Nature* 1999; 402: Suppl. 6760, B5–B11.
2. Ober C, Moffatt MF. Contributing factors to the pathobiology: the genetics of asthma. In: Wenzel SE, ed. The pathobiology of asthma: implications for therapy. Clinics in Chest Medicine. Philadelphia, USA, WB Saunders and Co., 2000; 21: pp. 245–261.
3. Cookson W. Genetics and genomics of asthma and allergic diseases. *Immunol Rev* 2002; 190: 195–206.
4. Hoffjan S, Ober C. Present status on the genetic studies of asthma. *Curr Opin Immunol* 2002; 14: 709–717.
5. Vercelli D. Genetic polymorphism in allergy and asthma. *Curr Opin Immunol* 2003; 15: 609–613.
6. Sandford AJ, Silverman EK. Chronic obstructive pulmonary disease. 1: Susceptibility factors for COPD the genotype-environment interaction. *Thorax* 2002; 57: 736–741.
7. Sandford AJ, Joos L, Pare PD. Genetic risk factors for chronic obstructive pulmonary disease. *Curr Opin Pulm Med* 2002; 8: 87–94.
8. Barnes PJ. Genetics and pulmonary medicine. 9. Molecular genetics of chronic obstructive pulmonary disease. *Thorax* 1999; 54: 245–252.
9. Kauffmann F, Demenais F. Chronic obstructive pulmonary disease. In: King RA, Rotter JI, Motulsky AG, eds. The genetic basis of common diseases. New York, Oxford University Press, 2002; pp. 155–179.
10. Cavalli-Sforza LL. The DNA revolution in population genetics. *Trends Genet* 1998; 14: 60–65.
11. Zhang Y, Leaves NI, Anderson GG, et al. Positional cloning of a quantitative trait locus on chromosome 13q14 that



- influences immunoglobulin E levels and asthma. *Nat Genet* 2003; 34: 181–186.
12. Van Eerdeghe P, Little RD, Dupuis J, *et al.* Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 2002; 418: 426–430.
  13. Allen M, Heinzmann A, Noguchi E, *et al.* Positional cloning of a novel gene influencing asthma from chromosome 2q14. *Nat Genet* 2003; 35: 258–263.
  14. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC. Family study of alpha 1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. *Gen Epidemiol* 1992; 9: 317–331.
  15. Lomas DA, Silverman EK. The genetics of chronic obstructive pulmonary disease. *Respir Res* 2001; 2: 20–26.
  16. Silverman EK, Chapman HA, Drazen JM, *et al.* Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 1998; 157: 1770–1778.
  17. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; 11: 241–247.
  18. Silverman EK, Mosley JD, Palmer LJ, *et al.* Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum Mol Genet* 2002; 11: 623–632.
  19. Silverman EK, Palmer LJ, Mosley JD, *et al.* Genome wide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet* 2002; 70: 1229–1239.
  20. Pekkanen J, Pearce N. Defining asthma in epidemiological studies. *Eur Respir J* 1999; 14: 951–957.
  21. Pearce N, Pekkanen J, Beasley R. Role of bronchial responsiveness testing in asthma prevalence surveys. *Thorax* 2000; 55: 352–354.
  22. Pearce N, Pekkanen J, Beasley R. How much asthma is really attributable to atopy? *Thorax* 1999; 54: 268–272.
  23. Pearce N, Douwes J, Beasley R. Is allergen exposure the major primary cause of asthma? *Thorax* 2000; 55: 424–431.
  24. Douwes J, Gibson P, Pekkanen J, Pearce N. Non-eosinophilic asthma: importance and possible mechanisms. *Thorax* 2002; 57: 643–648.
  25. Pearce N, Douwes J, Beasley R. The rise and rise of asthma: a new paradigm for the new millennium? *J Epidemiol Biostat* 2000; 5: 5–16.
  26. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature* 2001; 2: 930–942.
  27. Sachidanandam R, Weissman D, Schmidt SC, *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409: 928–933.
  28. The SNPs consortium Ltd. [snp.cshl.org/](http://snp.cshl.org/); The SNP consortium. Date last updated: October 21, 2003. Date last accessed: June 3, 2004.
  29. Johnson GCL, Esposito L, Barratt BJ, *et al.* Haplotype tagging for the identification of common disease genes. *Nature Genet* 2001; 29: 233–237.
  30. Daly MJ, Rioux JD, Schaffner ST, *et al.* High-resolution haplotype structure in the human genome. *Nature Genet* 2001; 29: 229–232.
  31. Stephens JC, Schneider JA, Tanguay DA, *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 2001; 293: 489–493.
  32. Patil N, Berno AJ, Hinds DA, *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294: 1719–1723.
  33. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* 2003; 3: 77–96.
  34. Sauer S, Gut IG. Extension of the GOOD assay for genotyping single nucleotide polymorphisms by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom* 2003; 17: 1265–1272.
  35. Cheung VG, Morley M, Aguilar F, Massimi A, Kuchelapati R, Childs G. Making and reading microarrays. *Nature Genet* 1999; 21: Suppl. 1, 15–19.
  36. Khan J, Bittner ML, Chen Y, Meltzer PS, Trent JM. DNA microarray technology: the anticipated impact on the study of human disease. *Biochim et Biophys Acta* 1999; 1423: M17–28.
  37. Lefkowitz. Functional and structural proteomics: a critical appraisal. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003; 787: 1–10.
  38. Tyers M, Mann M. From genomics to proteomics. *Nature* 2003; 422: 193–197.
  39. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiological studies of common genetic variants and cancer quantification of bias. *J Natl Cancer Inst* 2000; 92: 1151–1158.
  40. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; 361: 598–604.
  41. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; 55: 997–1004.
  42. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; 65: 220–228.
  43. Clayton DG. Models for analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer JH, Feinleib M, Lippert P, Hoffmeister H, eds. *Statistical models for longitudinal studies of health*. Oxford, Oxford University Press, 1992; pp. 301–331.
  44. Bickeböller H. Investigation of linkage and association: issues on study design. *Int Stat Rev* 2000; 68: 75–81.
  45. Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. *J Natl Cancer Inst* 1999; 26: 31–37.
  46. Risch N. Searching for genetic determinants in the new millennium. *Nature* 2000; 405: 847–856.
  47. Schaid DJ. *Transmission Disequilibrium Methods for Family-Based Studies*. Draft of a book chapter. [www.ersnet.org/Working\\_docs/2/SchaidChapter.pdf](http://www.ersnet.org/Working_docs/2/SchaidChapter.pdf). Date last updated: 23 January, 2002. Date last accessed: 2 April, 2004.
  48. Clerget-Darpoux F, Bonaiti-Pellie C. Strategies based on marker information for the study of human diseases. *Ann Hum Genet* 1992; 56: 145–153.
  49. Ott J. *Analysis of human genetic linkage*. 3rd Edn. Baltimore, Johns Hopkins University Press, 1999.
  50. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; 11: 241–247.
  51. 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.
  52. Schaid DJ. Transmission disequilibrium, family controls, and great expectations. *Am J Hum Genet* 1998; 63: 935–941.
  53. Witte JS, Gauderman WJ, Thomas DC. Asymptotic bias and efficiency in case-control studies of candidate genes and gene-environment interactions: basic family designs. *Am J Epidemiol* 1999; 149: 693–705.
  54. Thompson WD. Effect modification and the limits of biological inference from epidemiologic data. *J Clin Epidemiol* 1991; 44: 221–232.
  55. Rose G. *The strategy of preventive medicine*. Oxford, Oxford University Press, 1992.
  56. Corbo GM, Forastiere F, Agabiti N, *et al.* Passive smoking and lung function in alpha(1)-antitrypsin heterozygote schoolchildren. *Thorax* 2003; 58: 237–241.
  57. Baldini M, Vercelli D, Martinez FD. CD14: an example of gene by environment interaction in allergic disease. *Allergy* 2002; 57: 188–192.
  58. Eder W, Klimecki W, Yu L, *et al.* Toll-like receptor 2 as a major gene for asthma in children of European Farmers. *J Allergy Clin Immunol* 2004; 113: 482–488.
  59. Gilliland FD, Li YF, Dubeau L, *et al.* Effects of glutathione

- S-transferase M1, maternal smoking during pregnancy, and environmental tobacco smoke on asthma and wheezing in children. *Am J Respir Crit Care Med* 2002; 166: 457–463.
60. Romieu I, Sienra-Monge JJ, Ramirez-Aguilar M, *et al.* Genetic polymorphism of GSTM1 and antioxidant supplementation influence lung function in relation to ozone exposure in asthmatic children in Mexico City. *Thorax* 2004; 59: 8–10.
  61. Nadif R, Jedlicka A, Mintz M, Bertrand JP, Kleeberger S, Kauffmann F. Effect of TNF and LTA polymorphisms on biological markers of response to oxidative stimuli in coal miners: a model of gene-environment interaction. *J Med Genet* 2003; 40: 96–103.
  62. Sanak M, Pierzchalska M, Bazan-Socha S, Szczeklik A. Enhanced expression of the leukotriene C(4) synthase due to overactive transcription of an allelic variant associated with aspirin-intolerant asthma. *Am J Respir Cell Mol Biol* 2000; 23: 290–296.
  63. Howard TD, Koppelman GH, Xu J, *et al.* Gene-gene interaction in asthma: IL4RA and IL13 in a Dutch population with asthma. *Am J Hum Genet* 2002; 70: 230–236.
  64. The Human Genome Epidemiology Network. [www.cdc.gov/genomics/hugenet](http://www.cdc.gov/genomics/hugenet). Date last updated: 3 June, 2004. Date last accessed: 6 June, 2004.
  65. Little J, Bradley L, Bray MS, *et al.* Reporting, appraising and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156: 300–310.
  66. Little J, Houry MJ, Bradley L, *et al.* The Human Genome Project is complete. How do we develop a handle for the pump? *Am J Epidemiol* 2003; 157: 667–673.
  67. Houry MJ, Little J, Burke W, eds. Human Genome Epidemiology: a scientific foundation for using genetic information to improve health and prevent disease (Monograph in epidemiology and biostatistics). Oxford, Oxford University Press, 2004.
  68. Knoppers BM, Laberge C. Ethical guideposts for allelic variation databases. *Hum Mut* 2000; 15: 30–35.
  69. Council for international organizations of medical sciences (CIOMS). [www.cioms.ch/](http://www.cioms.ch/). Date last updated: 21 April, 2004. Date last accessed: 6 June, 2004.
  70. Canada Research Chair in Law and Medicine, Université de Montréal. HumGen. International database on the legal, social and ethical aspects of human genetics. [www.humgen.umontreal.ca/en/](http://www.humgen.umontreal.ca/en/). Date last accessed: 3 March, 2004.
  71. HUGO (Human Genome Organization), Ethics Committee, Statement on Benefit-Sharing, 2000. [www.hugo-international.org/hugo/benefit.html](http://www.hugo-international.org/hugo/benefit.html). Date last updated: 9 April, 2004. Date last accessed: 6 June, 2004.
  72. Deschênes M, Cardinal G, Knoppers BM, Glass KC. Human genetic research, DNA banking and consent: a question of 'form'? *Clin Genet* 2001; 59: 221–239.
  73. Knoppers BM, Hirtle M, Glass KC. Commercialization of genetic research and public policy. *Science* 1999; 286: 2277–2278.
  74. Cambon-Thomsen A. Assessing the impact of biobanks. *Nat Genet* 2003; 34: 25–26.
  75. European Respiratory Society research seminar. Post Genome respiratory epidemiology [www.ersnet.org/cernay](http://www.ersnet.org/cernay). Date last updated: February, 2002. Last accessed: 3 March, 2004.
  76. Bousquet J, Ansotegui IJ, van Ree R, Burney PG, Zuberbier T, van Cauwenberge. European Union meets the challenge of the growing importance of allergy and asthma in Europe. *Allergy* 2004; 59: 1–4.