REVIEW ARTICLE

GENOMIC MEDICINE

Alan E. Guttmacher, M.D., and Francis S. Collins, M.D., Ph.D., *Editors*

Genomics as a Probe for Disease Biology

Wylie Burke, M.D., Ph.D.

Although our understanding of pathology has grown rapidly in recent decades, the underlying mechanisms of many diseases remain obscure. Genomic research offers a new opportunity for determining how diseases occur, by taking advantage of experiments of nature and a growing array of sophisticated research tools to identify the molecular abnormalities underlying disease processes.[1] In this review I examine examples in which genomic research has improved our understanding of molecular pathobiology and consider its potential for contributing to the study of common complex diseases.

From the Department of Medical History and Ethics, University of Washington, Seattle. Address reprint requests to Dr. Burke at the Department of Medical History and Ethics, Box 357120, University of Washington, 1959 NE Pacific, Rm. A204, Seattle, WA 98195, or at wburke@u.washington.edu.

N Engl J Med 2003;349:969-74.

Effect of Mutations on the Severity of Hemophilia A

Before the advent of therapy for hemophilia A, some affected patients had only moderate bleeding problems, lived to adulthood, and led relatively normal lives in the absence of trauma or surgery.[2] Others had severe, spontaneous bleeding beginning in early childhood and rarely survived to adulthood; still others had disease of intermediate clinical severity. The main cause of this variation is the different mutations of the hemophilia A gene that cause hemophilia. When a mutation results in the complete loss of factor VIII protein — usually because a large gene deletion or genetic inversion results in the failure of gene transcription — severe hemophilia occurs.[2] The absence of endogenous factor VIII also increases the likelihood of an immune response when factor VIII–replacement therapy is used.[3] This immune response is a severe complication because it can prevent effective treatment. In contrast, other mutations, such as those in which there is a small change in the DNA sequence of the gene, lead to amino acid substitutions in the factor VIII protein. Depending on the nature of the substitution and its effect on the function of factor VIII, these mutations cause mild-to-moderate disease.

This example demonstrates that there is often a logical relation between a person's DNA sequence (genotype) and health outcome (phenotype). The functional effect of the genotype is the key factor, providing a molecular explanation for the severity of a given disease.

Reclassification of the Dystrophinopathies through Genomic Understanding

Duchenne's muscular dystrophy is caused by mutations in the dystrophin gene — usually deletions or gene inversions — that produce total or near-total loss of the dystrophin protein from skeletal muscle, resulting in early and progressive loss of muscle function (Fig. 1).[5] Mutations in the dystrophin gene that cause less severe deficits in the final protein product result in Becker's muscular dystrophy, a milder disease that was historically considered a separate clinical entity. This disorder differs from Duchenne's muscular dystrophy in its later onset and milder course.[5]

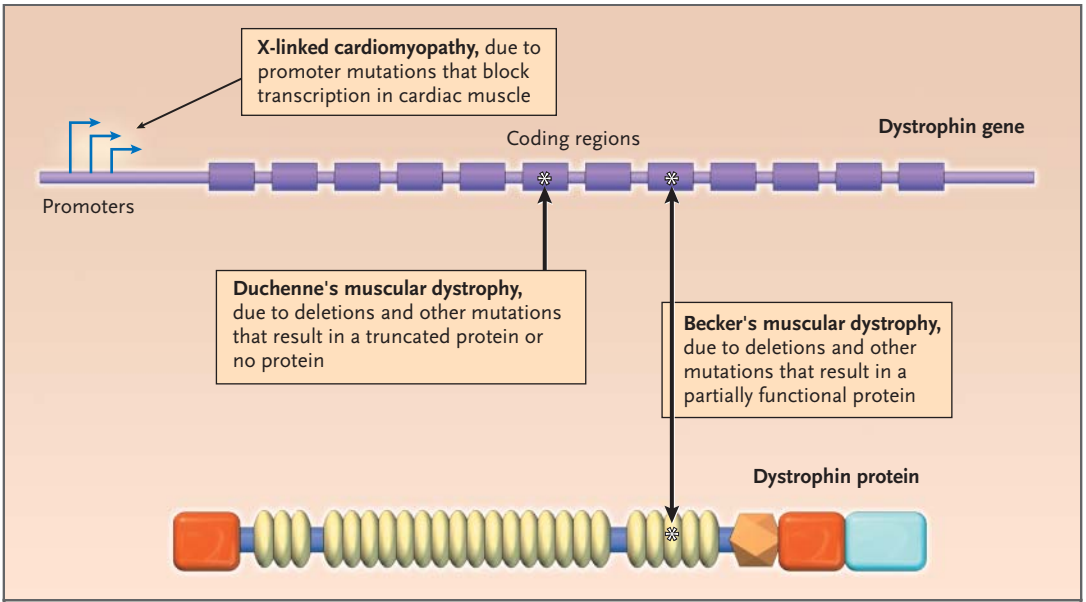

Figure 1. Clinical Effects of Various Mutations in the Dystrophin Gene.

The dystrophin gene consists of promoter regions, regions coding for the dystrophin protein, and noncoding regions; in addition, smaller protein products of uncertain function are produced through the action of additional internal promoters and alternative splicing of transcripts.[1,4] The dystrophin protein consists of binding regions at either end and a central portion consisting of units known as spectrin-like repeats that are thought to give the protein a flexible, rod-like structure.[4] Mutations that result in either no protein product or a truncated protein product — usually large gene deletions or inversions — cause Duchenne's muscular dystrophy. Mutations that change the protein structure so as to partially alter its function result in the milder Becker's muscular dystrophy. X-linked dilated cardiomyopathy is caused by promoter mutations that result in the selective loss of dystrophin from heart-muscle tissue owing to the loss of transcription. Adapted from Blake et al.[4]

When genomic research showed that these two clinically distinct disorders involved the same gene, a family of clinical disorders known as dystrophinopathies was identified.[4] A third dystrophinopathy, X-linked dilated cardiomyopathy, was subsequently discovered. This disorder is caused by specific mutations in the dystrophin gene that lead to the selective loss of dystrophin from cardiac muscle, while dystrophin levels in skeletal muscles remain normal or nearly normal.[4,6] These mutations appear to affect one of the promoter regions of the gene, resulting in the selective loss of gene transcription in cardiac tissue (Fig. 1).[4,6] Discovery of the gene coding for dystrophin thus provided the means to understand a molecular relation among three seemingly different clinical disorders. As with hemophilia A, the relation between the genotype and the clinical outcome is the result of the functional effect of different mutations on dystrophin.

Another genetic disorder, hereditary hemorrhagic telangiectasia,[7] provides an interesting contrast. This disorder (also known as Osler–Weber–Rendu disease) causes vascular dysplasia, resulting in epistaxis, hemoptysis, and gastrointestinal bleeding. It is inherited as an autosomal dominant disorder and was assumed to be due to mutations in a single gene, but molecular studies revealed the involvement of two genes, one encoding activin-receptor–like kinase 1 (*ALK1*)[8] and one encoding endoglin.[9] Thus, the discovery of the genetic causes in this case revealed an unexpected complexity. Molecular studies also provided an explanation for the similar clinical outcome of mutations in two different genes: both protein products appear to function in the same or related biologic pathways.[8,9]

VARIABLE EFFECT OF THE SAME GENOTYPE ON HEALTH OUTCOMES IN CYSTIC FIBROSIS

Although the clinical effect of genetic mutations can often be predicted on the basis of their functional

effect, the relation between the genotype and phenotype is sometimes less direct, as illustrated by cystic fibrosis. Cystic fibrosis results from mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene; different mutations in the gene have various effects on the function of CFTR protein.[10,11] More severe mutations are associated with a loss of function, through defective synthesis, defective maturation, or blocked activation of the CFTR protein. Other mutations result in only partial loss of CFTR function, and as expected, these mutations are often associated with less severe disease,[10,11] including a later onset of symptoms, pancreatic sufficiency, and sometimes milder pulmonary disease.[3,12,13]

However, the correlation between genotype and phenotype in cystic fibrosis is imprecise. In particular, the severity of pulmonary disease cannot be predicted for most *CFTR* genotypes, including the most common one — ΔF508/ΔF508[11] — since both the age at onset of pulmonary symptoms and the rate of decline in pulmonary function vary.[14-16] For example, in some patients with the ΔF508/ΔF508 genotype, chronic respiratory symptoms do not develop until adolescence or adulthood and pulmonary function in young adulthood can range from highly compromised to normal.[11,15,16] This discrepancy is important, because pulmonary complications are typically the most serious and life-threatening manifestations of cystic fibrosis.

Both genetic and nongenetic factors appear to modify the effect of the cystic fibrosis genotype.[11] A Danish study found that certain genetic variants of mannose-binding lectin, a protein that functions in innate immune responses, were associated with greater loss of lung function in patients with cystic fibrosis.[17] Interestingly, this adverse effect was limited to patients with chronic *Pseudomonas aeruginosa* infection, indicating an interaction between the gene variant and the environment. Other studies, some of which used animal models of cystic fibrosis, provide additional evidence of the existence of genetic modifiers of the clinical effect of *CFTR* genotypes.[11,18-20] Some modifiers appear to be organ-specific; for example, a gene locus has been identified that influences whether an infant with cystic fibrosis will have neonatal meconium ileus, but it does not appear to influence the severity of lung disease.[21] Nongenetic modifiers of the cystic fibrosis phenotype, such as exposure to environmental tobacco smoke as a factor in the poor outcome of the disease[22] or exposure to respiratory pathogens as a contributor to the progression of lung disease, have been described.[10,11,23] The existence of environmental modifiers is also suggested by studies demonstrating a correlation between life expectancy and health insurance status[24] or socioeconomic status.[25] These findings could reflect the modifying effect of factors such as nutritional status or access to antibiotic treatment or could represent the effect of other environmental modifiers that have yet to be described.

Further evidence of the complexity of the genetics of cystic fibrosis comes from data on other phenotypes associated with *CFTR* mutations. Some mutations cause male infertility through the congenital absence of the vas deferens, without accompanying lung disease,[26] and others cause idiopathic chronic pancreatitis and mild, late-onset pulmonary disease.[11,27] Limited data suggest that *CFTR* mutations may also contribute to some cases of chronic sinusitis and allergic bronchopulmonary aspergillosis.[27] As we learn more about the function of CFTR and the effect of other genetic and nongenetic factors on this protein, we will most likely increase our understanding of the biology of both cystic fibrosis and related disorders.

GENETIC VARIANTS AS A COMMON PHENOMENON

Variants in many different genes form the basis for the genetic contribution to disease across the spectrum, from rare disorders such as cystic fibrosis to common complex disorders such as cancer and heart disease. Genetic variants occur because new mutations arise at a low but continual rate in human tissues. Mutations that arise in germ-line tissues can subsequently be inherited, increasing the genetic variation in the population. The recently completed draft sequence of the human genome includes a catalog of 1.4 million single-nucleotide polymorphisms — sites where variations occur in the bases that form the building blocks of the DNA sequence.[28] Most DNA-sequence variations occur in noncoding regions of the genome — that is, in regions that do not code for protein products. Changes that occur in coding regions, however, can affect the function or efficiency of the protein that a gene encodes.

These differences can have physiological effects that are clinically important, such as causing differences in the response to drugs or environmental exposures or differences in susceptibility or the pre-

disposition to diseases. For example, many of the enzymes responsible for drug metabolism occur in variant forms, leading to differences between people in the efficacy of a drug and the risk of adverse effects.[29,30] Similarly, many genetic variants that contribute to the risk of common diseases such as cancer and heart disease have been identified.[31] Unlike the gene changes that cause genetic disease, which tend to be rare and result in clinically significant loss of function, most common genetic variants cause relatively small changes in function.

Genetic changes may be adaptive in certain circumstances and harmful in others. An often-cited example is the association of sickle cell trait — a mutation in the hemoglobin A gene — with resistance to malaria.[32] The selective advantage conferred by this resistance accounts for the high prevalence of sickle trait in populations originating from regions where malaria is endemic. However, the presence of two copies of the sickle cell trait results in sickle cell anemia, a disease associated with multiple complications and premature mortality. In this case, genetics offered insights into disease biology that would not otherwise have been attainable. The Human Genome Project is likely to document many instances in which the effects of genetic variation are dependent on the context. As an example, a variant in the interleukin-1–receptor agonist, a protein that inhibits the inflammatory response, is associated with an increased risk of certain autoimmune diseases, but it may reduce the morbidity of some infections.[33]

ASTHMA AS AN EXAMPLE OF THE GENETICS OF COMMON COMPLEX DISEASES

Both genetic and environmental risk factors have an important role in most common diseases.[31] The same strategies developed for the study of genetic diseases can be applied to the study of common diseases, but the task is more difficult because complex patterns of gene–gene and gene–environment interactions must be evaluated. Asthma provides an illustration of this challenge.

Epidemiologic studies provide evidence that multiple genetic and environmental factors contribute to the causation of asthma, a clinical condition that is best viewed as a cluster of related disorders.[34] Patients with asthma vary with respect to the age at onset, course, sensitivity to specific environmental precipitants, and response to medications, and the relative contribution of genetic and nongenetic factors may also vary considerably among patients. Furthermore, the prevalence of asthma has risen dramatically in the past two decades, indicating that environmental risk factors have a key role.[35] Control of environmental risk factors and improved treatment are the primary public health strategies for the prevention of asthma.

Nevertheless, genetic factors contribute substantially to the risk of asthma.[34,36-38] How, in this context, might the study of the genetics of asthma contribute to better health care outcomes? There are several possibilities. A classification of asthma based on genetics might provide a more accurate means of defining clinical subtypes that benefited from specific treatments. Genetic classification might also provide improved prognostic information, including the identification of patients who are at highest risk for severe or life-threatening episodes of asthma. A better understanding of the molecular processes involved in the different pathways of asthma is also likely to lead to a more detailed understanding of the pathophysiology of the disease. This effort could lead to a more precise definition of the environmental modifications most likely to reduce the risk of asthma. It could also lead to improved drug treatment through genetic testing to predict a patient's responses to a drug or through the development of new drug therapies.[36,38]

Studies to identify genes associated with asthma use mapping techniques to pinpoint gene loci linked to asthma[1] and physiological studies to identify genes that are likely to affect the disease process. Both approaches have been productive. Mapping techniques have identified several genes associated with asthma.[37,38] One of these has yielded information about a metalloproteinase, ADAM-33, that may have a role in inflammatory responses or smooth-muscle hypertrophy and hyperreactivity.[39,40] Physiological studies have led to the characterization of genetic variants associated with asthma or atopic airway inflammation in several biologic pathways potentially related to asthma — for example, the beta-adrenergic receptor,[41] cytokines associated with the secretion of immunoglobulin E and airway inflammation,[42] and a transferase presumed to be involved in the detoxification of inhaled irritants.[43] One study reported a gene–environment interaction in which the effect of smoking on the risk of asthma was increased by a specific beta-adrenergic–receptor genotype.[44]

This research is still in the early stages and faces

a number of technical problems that will also apply to the study of other common diseases. These include the need for standardized definitions of asthma phenotypes and intermediate biologic measures associated with the risk of asthma,[45-47] well-defined populations in unbiased studies with sufficient power to detect small effects,[48-50] and the concurrent measurement of both environmental and genetic risk factors.[51] Finally, any reported association between a genetic variant and disease risk cannot be considered established until the results of the study have been replicated.[52,53]

Determining associations between genes and diseases is just the first step in the translation of genomic research into clinical insights. This effort will require increasing attention to the study of the functions of proteins, or "proteomics,"[54] including the characterization of proteins identified as a result of genomic research. Of the estimated 30,000 to 35,000 genes in the human genome, approximately half code for unknown proteins.[28] With use of the genetic techniques that are now available, these proteins, their functions, and their interactions will be increasingly identifiable.[55] Analysis of the interaction between genetic and environmental effects in relevant biologic pathways, some (perhaps many) of which remain to be discovered, will form an important part of the study of common diseases such as asthma. As methods to achieve this goal are developed, they will provide a more complete and detailed picture of disease processes than has ever previously been possible.[56]

CONCLUSIONS

The study of gene mutations has provided a new model of pathophysiology in which the molecular causes of disease are illuminated by genetics. Evidence is now emerging of the complex interactions between genes and between genes and the environment in the causation of many diseases, and the study of these interactions represents the next important step in genomic research. Efforts to understand the molecular mechanisms that underlie common complex diseases will build on insights and strategies developed in the study of single-gene diseases. However, the scope of the analysis is far greater and will require continuing efforts to develop and improve molecular and informatic tools that allow the simultaneous analysis of many genetic variants and environmental risk factors. The success of genomic research to date[28] suggests that this ambitious research enterprise will also ultimately succeed.

REFERENCES

1. Guttmacher AE, Collins FS. Genomic medicine — a primer. N Engl J Med 2002;347:1512-20.
2. Hoyer LW. Hemophilia A. N Engl J Med 1994;330:38-47.
3. Goodeve AC, Williams I, Bray GL, Peake IR. Relationship between factor VIII mutation type and inhibitor development in a cohort of previously untreated patients treated with recombinant factor VIII (Recombinate). Thromb Haemost 2000;83:844-8.
4. Blake DJ, Weir A, Newey SE, Davies KE. Function and genetics of dystrophin and dystrophin-related proteins in muscle. Physiol Rev 2002;82:291-329.
5. Korf BR, Darras BT, Urion DK. Dystrophinopathies [includes: Duchenne muscular dystrophy (DMD), pseudohypertrophic muscular dystrophy, Becker muscular dystrophy (BMD), and X-linked dilated cardiomyopathy (XLDCM)]. Seattle: University of Washington, 2003. (Accessed August 8, 2003, at http://www.geneclinics.org/profiles/dbmd/details.html.)
6. Beggs AH. Dystrophinopathy, the expanding phenotype: dystrophin abnormalities in X-linked dilated cardiomyopathy. Circulation 1997;95:2344-7.
7. Guttmacher AE, Marchuk DA, White RI Jr. Hereditary hemorrhagic telangiectasia. N Engl J Med 1995;333:918-24.
8. Johnson DW, Berg JN, Baldwin MA, et al. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. Nat Genet 1996;13:189-95.
9. McAllister KA, Grogg KM, Johnson DW, et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. Nat Genet 1994;8:345-51.
10. Mickle JE, Cutting GR. Clinical implications of cystic fibrosis transmembrane conductance regulator mutations. Clin Chest Med 1998;19:443-58.
11. Zielenski J. Genotype and phenotype in cystic fibrosis. Respiration 2000;67:117-33.
12. Gan K-H, Veeze HJ, van den Ouweland AMW, et al. A cystic fibrosis mutation associated with mild lung disease. N Engl J Med 1995;333:95-9.
13. De Braekeleer M, Allard C, Leblanc JP, Simard F, Aubin G. Genotype-phenotype correlation in cystic fibrosis patients compound heterozygous for the A455E mutation. Hum Genet 1997;101:208-11.
14. The Cystic Fibrosis Genotype–Phenotype Consortium. Correlation between genotype and phenotype in patients with cystic fibrosis. N Engl J Med 1993;329:1308-13.
15. Santis G, Osborne L, Knight RA, Hodson ME. Independent genetic determinants of pancreatic and pulmonary status in cystic fibrosis. Lancet 1990;336:1081-4.
16. Burke W, Aitken ML, Chen SH, Scott CR. Variable severity of pulmonary disease in adults with identical cystic fibrosis mutations. Chest 1992;102:506-9.
17. Garred P, Pressler T, Madsen HO, et al. Association of mannose-binding lectin gene heterogeneity with severity of lung disease and survival in cystic fibrosis. J Clin Invest 1999;104:431-7.
18. Rozmahel R, Wilschanski M, Matin A, et al. Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. Nat Genet 1996;12:280-7. [Erratum, Nat Genet 1996;13:129.]
19. Salvatore F, Scudiero O, Castaldo G. Genotype-phenotype correlation in cystic fibrosis: the role of modifier genes. Am J Med Genet 2002;111:88-95.
20. Mekus F, Laabs U, Veeze H, Tummler B. Genes in the vicinity of CFTR modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pairs. Hum Genet 2003;112:1-11.
21. Zielenski J, Corey M, Rozmahel R, et al. Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13. Nat Genet 1999;22:128-9.
22. Campbell PW III, Parker RA, Roberts

BT, Krishnamani MR, Phillips JA III. Association of poor clinical status and heavy exposure to tobacco smoke in patients with cystic fibrosis who are homozygous for the F508 deletion. J Pediatr 1992;120:261-4.
23. Dipple KM, McCabe ER. Modifier genes convert "simple" Mendelian disorders to complex traits. Mol Genet Metab 2000;71:43-50.
24. Curtis JR, Burke W, Kassner AW, Aitken ML. Absence of health insurance is associated with decreased life expectancy in patients with cystic fibrosis. Am J Respir Crit Care Med 1997;155:1921-4.
25. Britton JR. Effects of social class, sex, and region of residence on age at death from cystic fibrosis. BMJ 1989;298:483-7.
26. Dork T, Dworniczak B, Aulehla-Scholz C, et al. Distinct spectrum of CFTR gene mutations in congenital absence of vas deferens. Hum Genet 1997;100:365-77.
27. Noone PG, Knowles MR. 'CFTR-opathies': disease phenotypes associated with cystic fibrosis transmembrane regulator gene mutations. Respir Res 2001;2:328-32.
28. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921. [Errata, Nature 2001;411:720, 412:565.]
29. Evans WE, McLeod HL. Pharmacogenomics — drug disposition, drug targets, and side effects. N Engl J Med 2003;348:538-49.
30. Weinshilboum R. Inheritance and drug response. N Engl J Med 2003;348:529-37.
31. Collins FS. Shattuck Lecture — medical and societal consequences of the Human Genome Project. N Engl J Med 1999;341:28-37.
32. Ashley-Koch A, Yang Q, Olney RS. Sickle hemoglobin (HbS) allele and sickle cell disease: a HuGE review. Am J Epidemiol 2000;151:839-45.
33. Witkin SS, Gerber S, Ledger WJ. Influence of interleukin-1 receptor antagonist gene polymorphism on disease. Clin Infect Dis 2002;34:204-9.
34. Patino CM, Martinez FD. Interactions between genes and environment in the development of asthma. Allergy 2001;56:279-86.
35. Mannino DM, Homa DM, Akinbami LJ, Moorman JE, Gwynn C, Redd SC. Surveillance for asthma — United States, 1980–1999. MMWR CDC Surveill Summ 2002;51:1-13.
36. Palmer LJ, Silverman ES, Weiss ST, Drazen JM. Pharmacogenetics of asthma. Am J Respir Crit Care Med 2002;165:861-6.
37. Cookson WO. Asthma genetics. Chest 2002;121:Suppl:7S-13S.
38. Hakonarson H, Wjst M. Current concepts of the genetics of asthma. Curr Opin Pediatr 2001;13:267-77.
39. Van Eerdewegh P, Little RD, Dupuis J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 2002;418:426-30.
40. Shapiro SD, Owen CA. *ADAM-33* surfaces as an asthma gene. N Engl J Med 2002;347:936-8.
41. Liggett SB. Polymorphisms of the beta2-adrenergic receptor and asthma. Am J Respir Crit Care Med 1997;156:S156-S162.
42. Renauld JC. New insights into the role of cytokines in asthma. J Clin Pathol 2001;54:577-89.
43. Spiteri MA, Bianco A, Strange RC, Fryer AA. Polymorphisms at the glutathione S-transferase, GSTP1 locus: a novel mechanism for susceptibility and development of atopic airway inflammation. Allergy 2000;55:Suppl 61:15-20.
44. Wang Z, Chen C, Niu T, et al. Association of asthma with beta(2)-adrenergic receptor gene polymorphism and cigarette smoking. Am J Respir Crit Care Med 2001;163:1404-9. [Erratum, Am J Respir Crit Care Med 2002;166:775.]
45. Postma DS, Meijer GG, Koppelman GH. Definition of asthma: possible approaches in genetic studies. Clin Exp Allergy 1998;28:Suppl 1:62-6.
46. Koppelman GH, Meijer GG, Postma DS. Defining asthma in genetic studies. Clin Exp Allergy 1999;29:Suppl 4:1-4.
47. Ayres J. Severe asthma phenotypes: the case for more specificity. J R Soc Med 2001;94:115-8.
48. Palmer LJ, Cookson WO. Using single nucleotide polymorphisms as a means to understanding the pathophysiology of asthma. Respir Res 2001;2:102-12.
49. Weiss ST. Association studies in asthma genetics. Am J Respir Crit Care Med 2001;165:2014-5.
50. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002;156:300-10.
51. Khoury MJ, Yang Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. Epidemiology 1998;9:350-4.
52. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306-9.
53. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45-61.
54. Childs B, Valle D. Genetics, biology and disease. Annu Rev Genomics Hum Genet 2000;1:1-19.
55. Abbott A. A post-genomic challenge: learning to read patterns of protein synthesis. Nature 1999;402:715-20.
56. Burley SK, Almo SC, Bonanno JB, et al. Structural genomics: beyond the Human Genome Project. Nat Genet 1999;23:151-7.