PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Genetics of common disease: implications for therapy, screening and redefinition of disease

John Bell

Phil. Trans. R. Soc. Lond. B 1997 **352**, 1051-1055
doi: 10.1098/rstb.1997.0084

Genetics of common disease: implications for therapy, screening and redefinition of disease

JOHN BELL

Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

SUMMARY

Susceptibility to most common human diseases is, at least in part, determined by genetic factors. Rapid progress is being made in defining these genetic determinants for a range of diseases including breast cancer, colon cancer, diabetes, arthritis and dementia. The ability to define susceptibility in genetic terms has already led to a reclassification of some of these diseases on genetic and mechanistic grounds. This information is likely to have a profound effect on our approach to human diseases as it will allow a better definition of these disorders, permitting more effective therapeutic intervention, and will lead to both a more precise understanding of the natural history of these diseases and the possibility of identifying populations at risk. An understanding of the mechanisms underlying disease susceptibilty will also improve our ability to develop rational therapeutic interventions for many of these diseases. The role of genetic screening in these common diseases will be discussed, particularly in regard to the application of health care in populations.

1. INTRODUCTION

Most of the major common diseases have a significant genetic component. Due to developments in genetic technology it has become possible to identify some of these factors and evaluate their contribution to disease. This approach has been applied systematically to (i) a wide range of cancers; (ii) autoimmune disorders such as juvenile diabetes, rheumatoid arthritis and inflammatory bowel disease; (iii) degenerative diseases and diseases of ageing such as osteoarthritis and Alzheimer's disease; (iv) metabolic diseases such as type II diabetes and obesity; and (v) cardiovascular diseases. This approach has led to the discovery of a substantial number of genetic localizations of disease genes throughout the human genome, and in some cases the identification of the disease genes themselves, and the DNA variants that contribute to disease susceptibility.

Progress in this area has been very rapid in the past four years. Initially, this area of genetic activity greatly benefited from an extensive genetic map of highly polymorphic markers throughout the human genome. These markers have permitted the detection of areas of linkage for most of the common disorders studied. In addition the field has benefited greatly from the introduction of automated technology for genotyping such markers, and by the development and application of sophisticated statistical techniques to identify the basis for linkage in families with multiply affected members.

In virtually all the diseases studied to date a common pattern of disease susceptibility has emerged. In a small percentage of individuals fully penetrant genes have been detected in a subset of patients. These genetic factors by themselves are sufficient to cause disease. This is true in breast cancer (BRCA1, BRCA2) (Miki *et al*. 1994; Wooster *et al*. 1995), in type II diabetes (MODY1, MODY2, MODY3) (Froguel *et al*. 1993; Yamagata *et al*. 1996), in colon cancer (APC) (Kinzler & Vogelstein 1996) and in Alzheimer's disease (Roses 1995) (pre-senilin 1, pre-senilin 2, and APP). In general, these fully penetrant subtypes of disease contribute to only 5% or less of individuals with these disorders, and from a genetic screening point of view can be handled like other single gene disorders of full penetrance. The remainder of these common disease genes appear to be mediated by a more complex interaction between the genes and the environment. These genetic effects are likely to define a variety of further subtypes of disease, each interacting with a particular environmental component. The exact contribution of these genetic factors is certain in only very few diseases, and they are best treated as 'risk factors' in the evaluation of susceptibility to disease as they are unlikely to provide complete penetrance.

The clinical genetics community is very familiar with handling the information generated by fully penetrant disease susceptibility genes and much less able to deal with that emanating from genetic 'risk factors'. This has been much more the domain of epidemiologists, and hence has been somewhat overlooked from a genetic point of view.

2. USING GENETICS TO DEFINE DISEASE MECHANISMS

Perhaps the most important contribution to be made from the discovery of novel genes is a better understanding of the mechanism that underlies these common diseases. Most common human disorders are

described in phenotypic terms; disorders like type II diabetes are simply defined as individuals with elevated blood sugar without any specific understanding of the mechanisms responsible. It is likely that most of these disorders are in fact the result of a range of different mechanisms and that diabetes is a cluster of several different independent diseases. This may in part explain the difficulty epidemiologists have had in identifying the coexisting environmental factors, as they have been dealing with a heterogeneous pool of phenotypically defined disease. One of the opportunities provided by modern genetic techniques is that it should be possible to clarify the pathogenic basis of many of these disorders, and thereby more clearly define most diseases by mechanism.

It is easy to underestimate the potential impact on genetics in this area. In the history of medical practice, only the developments in microbiology at the turn of the last century had the potential to clarify disease mechanisms as much as our new understanding of genetics. Diseases such as 'yellow jaundice' which were at the turn of the century simply believed to reflect toxins in the blood (Osler 1892), are now defined as being the result of a variety of pathogens including the range of hepatitis viruses, hepatitis A, B, C, D and E, as well as autoimmune and alcoholic hepatitis. The importance of this mechanistic classification of hepatitis is reflected in the varying prognosis of these different diseases and increasingly the different management therapeutically of these disorders. It is likely, therefore, that genetics will provide the same mechanistic insights into a range of other diseases and hence improve our therapeutic accuracy and our ability to understand and define the natural history and environmental factors involved in a range of these disorders.

There are now many examples where genetics has begun to uncover the mechanistic events involved in common disease. Type II diabetes is clinically an obviously heterogeneous disorder. Many patients are considerably overweight and have high insulin levels, while others are thin. There is, in some patients, evidence of islet cell failure. The first gene for type II diabetes to be cloned was *glucokinase*, responsible for phosphorylation of glucose in pancreatic islet cells (Froguel *et al*. 1993). This event is the pivotal point in the determination of the set point for insulin secretion within the islets, and hence in this disorder mutations around the binding site for glucose which alter the affinity of glucokinase for its substrate lead to an alteration in the set point for insulin secretion. In these patients, who account for less than 3% of the population of type II diabetics, the disease is clearly and entirely the result of islet cell function mediated by mutations in this single enzyme.

Even more exciting has been the discovery that mutations in hepatic transcription factors of the HNF (hepatic nuclear family) accounts for a significant subset of patients with the disease (Yamagata *et al*. 1996). Such a pathophysiology has not been previously predicted and highlights the power of a genome-wide linkage approach.

Similarly, a substantial understanding of Alzheimer's disease has come from our characterization of the genes responsible for that disease. Three genes appear to function as fully penetrant monogenic contributors to disease, pre-senilin 1, pre-senilin 2 and APP. An allele at a fourth locus, *Apo E*, functions much more as a multifactorial gene and contributes to the disease in a very large number of patients (Roses 1995). An understanding of these key components to disease pathogenesis has contributed substantially to our understanding of the mechanism of the disease, although clear explanation as to how they might mediate Alzheimer's disease is still not available.

In disorders such as ulcerative colitis and Crohn's disease, collectively known as inflammatory bowel disease, again little has been known about disease pathogenesis. Recent studies from Oxford have identified a number of loci responsible for this disease susceptibility and have demonstrated that genes exist not only producing susceptibility to inflammatory bowel disease generally, but also specifically for each of the subtypes, ulcerative colitis and Crohn's disease (Satsangi *et al*. 1996). In addition to allowing one to specifically identify individual subtypes of disease, it is also possible to use genetics to demonstrate similarities in disease not previously recognized. This is particularly true for a range of diseases shown to be associated with variants at the angiotensin-converting enzyme locus (Cambien *et al*. 1992). Myocardial infarction, various forms of cardiomyopathy and renal disease in diabetics have all been shown to be associated with these DNA variants, and it is interesting to speculate how the function of this enzyme might contribute to all three disorders not previously associated with each other.

3. GENETIC CONTRIBUTIONS TO THE DRUG DISCOVERY AND DRUG DEVELOPMENT PROCESS

It is now recognized throughout the pharmaceutical industry that efforts need to be applied to developing drugs in a rational way based on disease mechanisms. An obvious step in this process is to understand the genetic basis of these disorders in an attempt to identify targets for drug development. The genetic approach to drug discovery has been used in a variety of disorders including cancer and obesity. In cancer, the identification of the *BRCA1* gene and the subsequent recognition that it and the *BRCA2* locus, also associated with familial breast cancer, were members of the granin family (Jensen *et al*. 1996) has led to the suggestion that expression of these molecules could potentially modify tumour behaviour. Experiments have already suggested this is the case, and there is considerable interest in developing this interesting biological product for therapeutic purposes. Similarly, the definition of leptin as a major mediator of obesity in the mouse model of obesity, the *Ob/Ob* mouse, has rapidly led to the utilization of this product as a potential appetite suppressant (Zhang *et al*. 1995). The exact mechanism of this protein is still uncertain but its identification and the subsequent cloning of its receptor has led to a substantial drug discovery programme on the basis of these genetic observations. It is important

to note that in some cases this genetic information may rapidly lead to drug development strategies, while in other situations it may prove to be considerably less useful. The contribution of *HLA* to a range of autoimmune diseases is substantial but it has not proved possible yet to convert that information into a novel approach to therapy despite repeated attempts to do so over a 25 year time-frame.

Perhaps the most important application of genetic information to the development and application of new drugs will be the use of genetic variation to define populations with a specific disease in order to more accurately target and prove efficacy of individual new therapies. Because genetics helps to define disease, response to therapy is likely to vary between mechanistically distinct diseases. An ability to target drugs more precisely is likely to have a profound effect on how the industry develops and markets drugs. If the pharmaceutical industry avoids tackling this problem, health insurers/providers will inevitably apply the information themselves. It will no longer be possible to treat individuals with inappropriate therapy to which they are unlikely to respond if response can be determined by genetic definition of disease. In cancer therapy, for example, those responding well (or badly) to chemotherapy or radiotherapy may depend on the variation that exists in enzymes involved in nucleoside metabolism or in radiosensitivity genes (i.e. ataxia-telangiectasia). Screening individuals before therapy is very likely to benefit the patient and the health insurer and will inevitably become a standard practice.

4. GENETIC SCREENING

The third important application of our new understanding of genetics will be its use in screening populations. In principle, there is relatively little that is new about this approach. Many patients have been screened for single monogenic disorders in the past, and hence the application of screening in these common diseases is unlikely to be significantly different than for rare single gene disorders. There is, however, a substantial difference in the magnitude of the problem. Clinical genetics units currently screen only a thousand patients a year for the rare single gene disorders. A much larger number of patients are likely to be screened in the future for a large range of diseases. This will substantially increase the amount of genetic information available in the population. Information about multifactorial genetic contributions to disease may also be handled as with any other risk factor, such as hypertension or hypercholesterolaemia. This large-scale screening may be used to characterize individuals in a population at risk of disease, and who can be treated early to reduce the risk of serious outcomes.

Screening for disease will fall into two broad categories. The first is predictive screening in pre-symptomatic individuals. There are really two groups of diseases in which predictive screening will be applicable. The first are those with relatively high penetrance single gene contributions to disease. These include genes with single mutations that are easy to screen, and those with multiple mutations. The recognition of highly penetrant genes in common disease has led to a substantial crisis in the ability of medical science to provide appropriate genetic screening. Haemochromatosis is the most recent example of such a situation (Feder *et al*. 1995). This recessive disorder has an allele frequency, responsible for the disease, of roughly 10 % and a disease frequency of 2–5 per thousand and is the commonest known genetic disorder. The disease is associated with a syndrome of hepatic cirrhosis accompanied by arthritis, hypogonadism and, occasionally, cardiomyopathy. An important feature of this disease is that cost–benefit analyses have demonstrated that population screening might well be cost-effective in this disorder as there is a relatively simple intervention, that of venesection, which would allow individuals homozygous for the mutated gene to be treated pre-symptomatically and for risk of the disease to be essentially eliminated. The gene has been cloned recently and proves to be a novel form of the *HLA* gene with a mutation affecting the cystine involved in the intrachain disulphide bond in the first extracellular domain. This mutation is likely to eliminate the immunoglobin fold in this portion of the molecule and would considerably disable its function. It is possible that population screening may be necessary to identify those with this mutation early so that early venesection can be used to prevent disease.

These examples provide little difficulty either in terms of accuracy of screening or predictive value. They may, however, confront the insurance industry with a significant problem in determining whether it is valuable to screen for such genes in the pre-symptomatic population.

Although predictive testing confined to highly penetrant single gene disorders is of considerable current interest, it is clear that future interests will lie with detecting the contributions to common disease from multifactorial trait loci that do not contribute entirely to the risk of disease. These genetic factors are best thought of as 'risk factors' in the same way that hypertension or hypercholesterolaemia are risk factors contributing to myocardial infarction or stroke. The ability to deal with such genetic risk factors will require a substantial amount of epidemiological data as to their contribution to disease susceptibility. We know already that some genetic factors such as *Apo E4* have a significant effect on the time of onset of Alzheimer's disease, and the magnitude of that effect appears to be at least as great as the contribution of hypertension to the risk of myocardial infarction (Roses 1995). Epidemiologists have been used to dealing with less than certain risks (with relatively low positive and negative predictive values), but where at a population level, treatment of people at the extremes of a phenotype may have impressive beneficial effects. Given that the insurance industry might consider both blood pressure and cholesterol levels as factors in determining insurability, it is possible in the future that genetic factors may also contribute to this process.

The evidence on Alzheimer's disease is perhaps the best example of such a multifactorial susceptibility factor. Although the presence of the *Apo E* allele, an

allele with a 15% frequency in Caucasian populations, is a significant risk factor in the development of this disease, it is not in any sense fully predictive. The presence of Apo E4 homozygosity does substantially shift the age of onset curves for Alzheimer's disease almost 20 years from those not possessing an Apo E4 allele (Roses 1995). Such a substantial shift and increase in risk associated with this factor may be comparable with other risk factors that are currently ascertained by the insurance industry and hence, particularly for health insurance and for long-term care insurance, it seems likely that such genetic markers will relatively rapidly be used for the stratification of populations for insurance purposes. The suggestion that some drugs for dementia work preferentially in populations that are either Apo E4 positive or negative, indicates that health care providers may soon be considering the use of this marker in defining disease subtypes. One of the major limitations in evaluating this data, and other genetic susceptibility data available, is the lack of large-scale prospective studies evaluating risk. These will be essential in the future if genetic 'risk' data is to be effectively applied.

Genetic tests can also be used not in the predictive sense but in a diagnostic sense. Within the *HLA* region, the *HLA DR4 Dw4Dw14* genotype is associated with a very dramatic, increased relative risk of developing rheumatoid arthritis compared with non-associated *HLA* genotypes. This relative risk figure is as high as 49, and hence may dictate therapeutic decisions in the future. Similarly, the *Apo E* genotype can be used diagnostically, and is currently being used to refine the diagnosis in substantial numbers of individuals who are cognitively impaired. The positive predictive value of the test is 99% (A. Roses, personal communication). The application of these genetic tests diagnostically is likely to have little effect on life insurance premiums, but will increase the amount of genetic information that is available. This may have implications for normal relatives of individuals detected to have particular genotypes and as a means of aiding diagnosis.

5. WHO WILL SCREEN AND WHY

There are three groups of people who are likely to use information available from genetic screening. The first is the patient who may wish particularly to know genetic information to assist with assessing reproductive risk, either to aid a marital decision or in the process of having a family. They may wish to have genetic information because of a strong family history and a wish to know particular susceptibility.

Physicians are likely to utilize genetic screening to help them modify existing screening programmes such as for the breast cancer genes or for colon cancer. They may use the genetic information to choose optimal therapeutic interventions, particularly if genetic stratification is shown to correlate with therapeutic efficacy of individual drugs, or they may wish to know if genetic information may permit them to intervene early in the disease process. It is unlikely that this activity will all go on within clinical genetics departments and it is probable that for many common diseases, genetic susceptibility will be tested in centralized facilities, and the responsibility for the interpretation and application of this data will lie with individual physicians in sub-specialties of medicine.

The final group of individuals with an interest in knowing the outcome of genetic testing may prove to be third parties such as the insurance industry or employers. The discussion on this issue has been diverted by two side issues. First, there is a current problem of obtaining life insurance premiums for those carrying genes for monogenic disorders such as Huntington's disease or myotonic dystrophy. The second confounding factor is the restriction of such discussions to life insurance, rather than health insurance more generally.

The major single gene disorders diagnosed within clinical genetics departments produce a relatively small problem for the insurance industry. Because the numbers of individuals afflicted with these disorders are small, it is unlikely that the approach taken by the insurance industry to providing life insurance for these individuals is likely to substantially alter their profit margins or indeed lead to the use of genetic information by patients to their advantage. This is particularly true if life insurance policies are less than £100 000 and the number of diseases in which this is relevant remains small. This may become more difficult to manage as highly penetrant genes in common diseases become available for screening. Although such genes in breast and colon cancer account for only 5% of patients, the frequency of these diseases suggests that there are large numbers of people carrying these genes, an estimated 200 000 in the UK carrying the *HNPCC* genes for colon cancer (Kinzler & Volgelstein 1996).

More difficult to deal with will be the expansion in understanding of genetic susceptibility due to high frequency, low penetrance genes, where the number of individuals susceptible due to genetic factors will be extremely large. Also, given an appropriate amount of prospectively obtained epidemiological data, it may be possible to predict outcomes for a large segment of the population much more accurately. These genes are less useful for predicting individual risk, but they can identify populations at risk. These genetic effects behave like other risk factors (i.e. hypertension or hypercholesterolaemia), and the industry already has some experience dealing with these. Genetic effects discovered to date are not small in terms of the sort of risk factors already considered by the insurance industry. These genetic effects are therefore likely to provide substantial information about future risk, and society as well as the insurance industry is likely to grapple with this problem. This is particularly true for health insurers rather than life insurers, as one could rapidly stratify populations as a function of their likely health care needs in the future and vary premiums appropriately. The discussion about Huntington's disease and myotonic dystrophy does not address this larger and more serious problem. It can only be addressed by (i) careful evaluation of epidemiological data relating to genetic contribution to disease; and (ii)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the acceptance that these factors can and should be used as risk factors to permit early intervention and, where possible, refine insurance premiums appropriately. Because information about these genetic effects is likely to be necessary for the proper application of health care in populations, the information will be available. The only remaining issue is whether the insurance industry will wish to utilize the information. Their ability to do so will, however, depend strongly on the quality of available data and the ability to accurately estimate risk. The failure to consider this will lead to substantial and unfair bias in the underwriting process.

6. CONCLUSIONS

The advent of molecular genetics in the study of common human disease has led to a remarkably rapid proliferation of information about disease susceptibility factors in many common human disorders. This will predominantly lead to improved health care through its unveiling of mechanisms of a wide range of diseases. In addition, this information is likely to contribute substantially both to the design of new therapies and the development of such new agents. Genetic information available through screening is likely to become increasingly prevalent both because of a wish to know from patients and because of its use by physicians to improve health care. How it will be applied by others will depend on how they see their balance of commercial opportunity and societal responsibility.

REFERENCES

Cambien, F. *et al.* 1992 Delection polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature* **359**, 641–644.

Feder, J. N. *et al.* 1995 A novel MHC Class I-like gene is mutated in patients with hereditary haemochromatosis *Nature Genet.* **13**(4), 399–409.

Froguel, P. *et al.* 1993 Familial hyperglycemia due to mutations in glykinase: definition of a subtype of diabetes mellitus. *New Engl. J. Med.* **328**, 697–702.

Jensen, R. A. *et al.* 1996 BRAC1 is secreted and exhibits properties of a granin. *Nature Genet.* **12**(3), 303–309.

Kinzler, K. W. & Vogelstein, B. 1996 Lessons from heredity colorectal cancer. *Cell* **87**, 159–170.

Miki, Y. *et al.* 1994 A strong candidate for the breast and ovarian cancer susceptibility gene BRAC1. *Science* **266**, 66–71.

Osler, W. 1892 Principles and practice of medicine. New York: D. Appleton & Co.

Roses, A. 1995 Apolipoprotein E genotyping in the differential diagnosis, not prediction, of Alzheimer's disease. *Ann. Neurol.* **38**(1), 6–14.

Satsangi, J. *et al.* 1996 Two-stage genome-wide search in inflammatory bowel disease: evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nature Genet.* **14**, 199–202.

Wooster, R. *et al.* 1995 Identification of the breast cancer susceptibility gene BRAC2. *Nature* **378**, 789–792.

Wordsworth, B. P. *et al.* 1992 HLA heterozygosity contributes to susceptibility to rheumatoid arthritis. *Am. J. Hum. Genet.* **51**, 585–591.

Yamagata, K. *et al.* 1996 Mutations in the hepatocyte nuclear factor-1a gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455–458.

Zhang, Y. *et al.* 1995 Positional cloning of the mouse obese gene and its human homologue. *Nature* **374**, 479.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES