1

Introduction

Public health genetics is a new discipline. It brings together the insights of genetic and molecular science as a means of preventing disease and of protecting and improving the health of the population. Its scope is wide, and requires an understanding of genetics, epidemiology, public health, the principles of ethics, law and the social sciences and much else besides.

At the core of public health genetics is the notion that genes, like the classic environmental factors that have been shown over many decades to be causally implicated in disease, are themselves important determinants of health; and that they play as important a role as exposures to physical and biological agents or to social and structural factors such as poverty and unemployment. But, as with environmental determinants, genes act not on their own but in combination with other factors. Every gene interacts with others in the genome and with a host of external exposures to produce the full range of human characteristics. The complexities of these relationships mean that, while genetic factors are at work in all diseases, no single genetic variant (except in the case of relatively rare 'genetic diseases', discussed further below) will be predictive of when or whether disease will strike, or of its severity.

The health and social policy issues that form much of the practice of public health genetics are equally complex, including legal and regulatory frameworks in genetic testing; science funding and policy; consent, confidentiality and data protection; the pharmaceutical and biotechnology industries; the patenting of genes and genetic sequences; and the education and training of health professionals and of the public in the implications of genetic science.

The definition of public health genetics

Two widely used definitions of public health genetics come from the United States. The University of Washington in Seattle defines it as *the application of advances in human genetics and molecular biotechnology to improve public health and prevent*

disease. The University of Michigan tells us that it provides an opportunity for public health professionals to gain *an understanding of the effects of genes on health and disease and to apply genetic information to public health practice*.

In the United Kingdom we have built on the well respected Acheson definition of public health and defined it *as the impact of genetics on the art and science of promoting health and preventing disease through the organised efforts of society*.

The broad scope implied by this definition was endorsed recently, as discussed further later in this chapter, by an international expert group that defined public health genetics as *the effective translation of genome-based knowledge and technologies for the benefit of population health*.

All of these formulations emphasise that the subject matter of public health genetics is how the health of the population and individuals within it, and the way in which public health and clinical medicine are practised will be affected by genetic science and technology.

A detailed analysis of the modified Acheson definition that we have used in the UK brings out a number of points:

1. It is important to note the different meanings that can be attached to the word 'genetics': genetics as the study of inheritance and inherited diseases, and genetics as the study of DNA, molecular and cellular biology. The discipline of public health genetics uses 'genetics' in the second of these senses, moving beyond inherited and congenital diseases and dealing, in addition, with the role of genetic factors in the complex disorders, such as coronary heart disease, cancers, diabetes, asthma, stroke or dementia, that contribute most to mortality and morbidity.
2. The definition of public health genetics makes clear that its aim is to promote health and to prevent disease, both for individuals and for the population as a whole. The public health perspective is in essence one that asks of scientific advances, technological interventions, policy and legislation whether they add to or detract from improvements in human health.
3. The word 'prevention', as used in public health, includes not just measures designed to prevent the onset of disease but, in addition, clinical interventions that lead to the reduction of disability and the progression of disease.
4. The definition states that the practice of public health genetics is both an art and a science, implying that it requires not just technical competencies, but also a sensitivity to the plural views that exist in society about genetic science and its consequences, and an ability to work across a whole range of disciplines and cultures.
5. The sphere of influence of public health genetics goes far beyond health service boundaries. The phrase 'organised efforts of society' refers to a panoply of health determinants that includes not just the practices of health services

themselves but factors as diverse as fiscal policy, patent law, educational curricula, data protection, genetic test regulation, the funding of science and many others besides. Ethical, legal and social perspectives run across all these activities and contribute to a key aspect of public health genetics practice.

Genetic and environmental factors as determinants of health

Environmental exposures and social factors have provided much of the material for classical analytical epidemiology. Using case–control and cohort observational study designs, epidemiologists have sought to demonstrate associations between various exposures and the incidence or prevalence of disease, and to determine those that might be causal. The association of smoking and lung cancer; of the human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS); of unemployment and all-cause mortality; of fat consumption and coronary heart disease; and of social isolation and depression are classical examples of associations that have emerged from such studies. The implicit assumption in all these studies was that the population was homogeneous. The investigators did not, of course, actually believe that this was so, but they analysed the results as if genetic heterogeneity between individuals in the population under study did not exist. The public health community, in turn, saw little in genetics to interest them, and did not regard genetic variation as an important matter that contributed to public health practice.

Public health genetics seeks to remedy this incomplete characterisation of health determinants by showing how genes play a major role. Figure 1.1 is a conceptual diagram of health determinants. It is set out in that particular way to illustrate a number of points. First, genetic factors are included explicitly as a health determinant. Second, the arrows demonstrate that the determinants interact in a complex way. For example, the presence of radon in the natural environment will

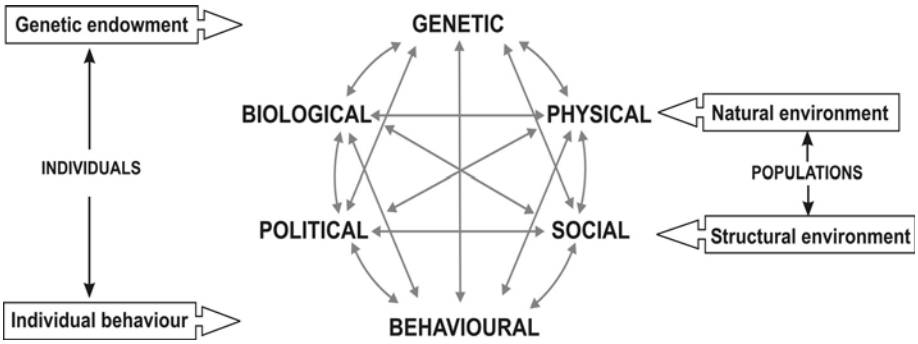

Figure 1.1 Determinants of human health

increase the mutation rate in exposed populations; genetic factors can affect certain aspects of human behaviour; the hepatitis virus and aflatoxins will affect the somatic genome and predispose exposed individuals to cancer.

Third, it brings out the point that natural environmental and structural determinants are external factors best influenced by interventions at the population level, while, by contrast, genetic factors and behaviour are essentially individual. In other words, how an individual chooses to behave, whether to smoke or indulge in other harmful behaviours, can only be changed or affected by the will of that particular individual, albeit that external or structural factors have an influence on such individual decisions.

The fourth point is that the multiple interactions in the figure can be reduced to a more simple representation, which simply shows a mutual relationship between genes and environment. Determinants of health and disease can only be genetic or environmental, since an 'environmental factor' is defined as anything that is not genetic. Genes and the environment interact with each other such that the risk of disease differs from individual to individual depending on their genetic constitution and environmental exposure. We normally call this 'gene–environment interaction', but it is probably more accurate to speak about the combined effects of genetic and environmental factors. The reason is that the term gene–environment interaction has a technical meaning for the statistician (see Chapter 3 for further discussion). The statistician takes the term interaction to refer to a risk estimation that does not conform to a pre-stated statistical model. No such assumption is made in our use of the term. All that we wish to imply is that genetic and environmental factors combine to influence risk.

Fifth and finally, the categorisation of determinants into those that are primarily population determined and those that are determined at an individual level reflects a distinction that is crucial to the understanding of preventive strategies and interventions. The epidemiologist Geoffrey Rose (1985) pointed out that there were two ways of preventing disease:

- by identifying those at greatest risk and directing preventive interventions at those individuals; and
- by trying to reduce the risk across the population as a whole through structural or environmental change that affects the whole population.

Blood pressure provides a good example. It is a risk factor for stroke. The high-risk approach identifies and treats individuals with blood pressure levels above a given threshold. The population approach seeks to reduce the mean blood pressure in the population through the reduction of salt consumption. Figure 1.2 is a conceptual diagram of the distribution of blood pressure levels across the population; it also shows the risk of stroke according to blood pressure levels. What Rose argued was that although treatment of high blood pressure in individuals at

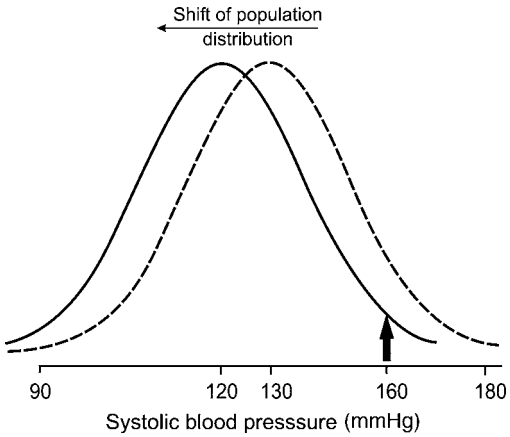

Figure 1.2 Population blood pressure levels and risk of stroke. Systolic blood pressure is shown in mmHg; 160 mmHg is used as a cut-off point (see arrow) above which individuals are considered to be at a significantly increased risk of stroke. The height of the curve at a given point represents the number or proportion of individuals within the population with that blood pressure level

the far end of the distribution brought about a much greater absolute risk reduction for each individual, a far greater number of strokes would be prevented in the population if the distribution curve could be shifted to the left by small reductions in mean blood pressure levels, by lowering the mean population salt consumption.

That genetic and environmental factors work together provides a key insight for public health genetics. The distribution of a risk factor such as blood pressure or cholesterol level across a population is determined by genetic, environmental and stochastic (random) factors. These factors between them are responsible for the shape and position of the frequency distribution of risk factors across the population. If we envisage a theoretical situation where all the individuals are exposed to exactly the same environmental factors, the variance of the resulting distribution will be reduced and will be a function of only the genetic variation in the population under study. The shape of the residual distribution will be due entirely to genetic factors, to the interaction between the genetic factors and the common environment, and to stochastic (random) variation. The position of the curve along the risk axis will in turn be determined by the nature of the environment – with some environments providing a greater, and others a lesser, health risk.

Rose's analysis has been used by some to argue that the targeting of preventive interventions at individuals who are at high genetic risk is not a defensible public health strategy. Rose himself did not draw this conclusion, however, stating that the whole-population and high-risk approaches were complementary and should both be pursued.

Genetic disease, complex disease and the combined effects of genetic and environmental factors

We have seen that genetic and environmental components both contribute to the development and risk of disease. However, we designate some diseases, cystic fibrosis or Duchenne muscular dystrophy for example, as 'genetic diseases'. What do we mean by this term and how do such diseases differ from those that we do not place in this category? All genetic factors are transmitted across generations in patterns that conform to the laws of inheritance first formulated by Mendel. The details of this mechanism are discussed in Chapter 2. The important point to appreciate here is that, in most cases, the physical manifestations that are influenced by these genetic factors – the diseases or physiological traits themselves – do not show such a pattern across generations and between family members, even though some traits and diseases have a tendency to cluster within families. A genetic disease is one where, in contrast to most, the manifestations of the disease show a pattern of transmission that is seen to conform to Mendel's laws.

The question then is why some diseases have these characteristics and not others. The diseases that pass across generations, genetic diseases, are usually single-gene disorders where the presence of the genetic abnormality is sufficient (on its own or with environmental factors) to give rise to the observable features of the disease. Most diseases, including those that are of greatest public health importance such as heart disease, diabetes, cancer or schizophrenia, are much more complex and come about as a result of the combined effects of multiple genes and their interaction with each other and with a range of environmental factors. For these reasons they are often referred to as complex or multifactorial diseases.

In multifactorial diseases, individual genes may make a contribution to the disease but each is insufficient in itself to cause the disease. The disease only manifests itself if other factors, such as other genetic variants and/or environmental factors, are also present such that together the factors comprise a sufficient cause of the disease. Suppose an individual possesses a set of genetic variants, and is exposed to a range of environmental factors, that together are sufficient to cause a specific disease. What are the implications for his or her offspring? Each gene only has a 50% chance of being passed on to the next generation (see Chapter 2 for further explanation). It is unlikely, therefore, that the full constellation of genetic components associated with the disease in that individual will be passed on; nor may it be assumed that the individual in the next generation is exposed to the same raft of environmental factors. Examples such as this illustrate why complex diseases do not conform to simple patterns of inheritance.

The manifestation or expression of genetic factors as disease or physiological traits is referred to by geneticists as the phenotype. The genetic factors themselves

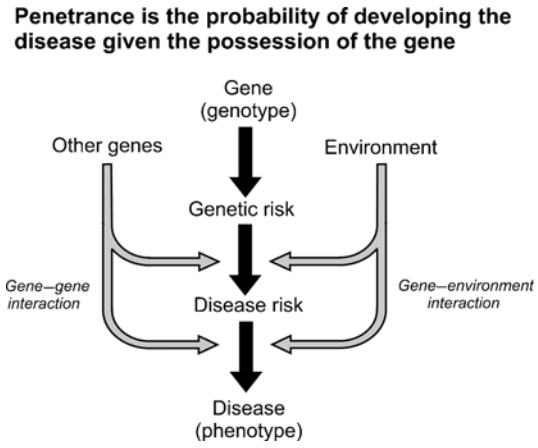

Figure 1.3 Penetrance. The penetrance of a genetic variant is the probability that traits or characteristics associated with that variant will manifest within a stated period of time

are called the genotype. Geneticists refer to the likelihood that a genetic variant will manifest its traits or characteristics within a stated time interval as penetrance (Figure 1.3). Penetrance refers to phenotype and is a property not just of the gene in question but of the genetic and external environment. Diseases such as cystic fibrosis or Duchenne muscular dystrophy are conditions whose penetrance is virtually 100% and may therefore be described as genetic diseases. If the genetic variant is inherited then it is certain that the individual will develop the disease (though, as discussed further in Chapter 2, even highly penetrant single-gene diseases may show variability in factors such as age of onset or range and severity of symptoms).

At the other end of the spectrum, is it possible to establish that there are diseases or traits that are entirely environmental, without any genetic influence at all? Accidental injury, for example as a result of a motor vehicle accident, may on first sight appear to be entirely environmental in origin, but a little reflection might lead one to question that initial impression. We know that there are certain conditions that predispose to sudden loss of consciousness, some of which are due to single-gene disorders. These include, for example, long QT syndrome, cardiomyopathies and certain forms of epilepsy. There is also reasonable evidence that certain genes are involved in traits such as impulsiveness, risk-taking behaviours or clumsiness, all of which could have an impact on the probability of being injured in a motor vehicle accident. The epidemiology of such accidents shows a huge sex difference; this is again evidence of a genetic factor at work.

The genetic disorder phenylketonuria (PKU; Box 1.1) provides another instructive example of the combined effect of genetic and enviromental factors: both a genetic mutation *and* a phenylalanine-containing diet are needed for the disease to

Box 1.1 Phenylketonuria: both a genetic and an environmental disease
Phenylketonuria (PKU) is a disease of the newborn and causes profound mental retardation. The disease is due to an abnormality in the gene that produces the enzyme phenylalanine hydroxylase. The enzyme converts the amino acid phenylalanine to tyrosine. Its absence allows the build up of phenylalanine which is toxic for the developing brain. The treatment is to restrict phenylalanine in the diet of the infant. The genetic abnormality *and* the presence of the environmental factor, phenylalanine, are both required for the disease PKU to develop. Neither one nor the other factor is sufficient on its own. The penetrance of the disease is 100% if phenylalanine is present but 0% if it is absent from the diet.

develop. We refer to PKU as a genetic disease because the dietary factor, phenylalanine, is ubiquitous while the genetic defect is rare, occurring in around 1 in 10 000 births. If, in an alternative world, a population all had the genetic abnormality that we associate with PKU but phenylalanine was not found in the diet of that population, the few cases of PKU observed in that world would be deemed to be toxic or nutritional in origin.

The phenylalanine-free world is obviously a hypothetical one. Real examples can also be found, however, to illustrate the point that whether we choose to label a disease as genetic or environmental in origin is dependent on the relative prevalence of those factors. If the genetic factor is rare against a common set of environmental factors we label the disease genetic, and vice versa if the environmental factor is rare. The relationship between smoking and lung cancer provides one example. If we study a population in which everyone smoked 60 cigarettes a day, the variation in lung cancer among individuals would by and large be determined by genetic factors, an insight due to Geoffrey Rose. When the tubercle bacillus was ubiquitous, medical students of the time were taught that individual constitution was a prime determinant of whether or not one developed tuberculosis and of its severity. They also learnt that rickets was a disease of vitamin D deficiency. But now, at least in the developed world, the few cases of rickets that we see are more often than not due to a plethora of individually rare genetic disorders of vitamin D metabolism, and very infrequently to dietary deficiency. These examples show that, as a general rule, to question whether a disease is genetic or environmental is meaningless; both contribute to the disorder and any attempt to segregate one set of factors from the other is conceptually unsound.

Public health genetics does not attempt to argue against the importance of environmental exposures as determinants of health – quite the reverse. Both genetic and environmental factors play a role in disease risk but in terms of prevention, a central goal of public health, it is generally only the environmental

part of this interaction that can be altered. The high incidence of diabetes mellitus in Pima Indians has been explained as the result of exposure of a particularly 'thrifty' genotype to a Western diet. Genetic factors are clearly at work but in practical terms the dietary exposure is much more important, as it is only by changing the diet that it will be possible to reduce the incidence of diabetes mellitus in this population.

The emergence and development of public health genetics

Public health genetics as a recognised and separate discipline of public health emerged only in the mid 1990s. It represented a coming together of insights and influences from several quarters: the increasing power and sophistication of epidemiology, the burgeoning discipline of genetic epidemiology, the excitement generated by the human genome project, and the recognition that there was a need to understand and resolve the many ethical, legal and social issues raised by advances in genetic science.

Advances in epidemiology and its application to public health

Epidemiology is now recognised as the core science that underpins the work of public health practitioners. In the last five decades it has developed beyond recognition. Basic concepts such as disease incidence and prevalence, relative and absolute risks and rates have been refined; the nature of causation and of how multiple factors interact in a web of causative factors are now better understood. Major methodological developments in study design have taken place, for cohort and case–control studies, and for randomised clinical trials.

Descriptive studies, documenting the relation of disease to person, place and time, have given way to analytical studies designed to elucidate association between exposures and disease, and causative pathways. Understanding the roles of bias, confounding and random factors in the interpretation of epidemiological data is now fundamental to epidemiological training, and advanced biostatistical techniques such as multiple logistic regression are now commonplace for the professional epidemiologist.

The approaches of epidemiology have also had an impact on mainstream clinical medicine, as clinical epidemiology and evidence-based medicine, both placing emphasis on numeracy, measurement and statistics in the evaluation of medical interventions, have been integrated into medical curricula.

The rise of genetic epidemiology

Genetic epidemiology began to be recognised as a separate subset of epidemiological research in the 1950s. In its early decades, most of its practitioners focused

on the genetic analysis of pedigree data. They carried out studies on families and siblings, on twins and adoptees, and their interests lay mainly in the single-gene disorders and chromosomal abnormalities. Their background was mathematical and the techniques that they used were highly complex and statistical, involving tools such as segregation analysis and linkage. This branch of epidemiology is still a robust and active one that has contributed much to the understanding of genetic diseases and the pinpointing of causative genes.

Another approach to genetic epidemiology, one that stresses its role in studying genetic–environmental interactions in disease aetiology, began to emerge during the 1970s and 1980s. The epidemiologists who took this approach regarded genetic factors in the same way as environmental exposures. They took the view that each played a role in the determination of disease risk and that the exposures, whether genetic or environmental, could be analysed using the same techniques derived from classical epidemiology such as case–control studies. The populations they studied were not family members or pedigrees but community-based population samples. Publications in the mid to late 1980s and early 1990s started to use the paradigms of genetic epidemiology to underpin public health work. Its techniques were applied to improve the estimation of disease risk for diseases such as familial breast and ovarian cancer; to evaluate screening programmes for genetic diseases; to study the role of folic acid in the pathogenesis of neural tube defects and much else besides.

The growth of genetic science

New approaches to genetics also began to emerge during the last two decades of the twentieth century. Since the re-discovery of Mendel's work in 1900, and the recognition by Boveri, Bateson and others of the chromosomal basis of inheritance, the emphasis had been largely on single-gene diseases. It was William Bateson who first coined the word 'genetics', defining it as the study of heredity and variation. The 1940s and 1950s saw the first demonstrations that defects in specific proteins were responsible for diseases including glycogen storage disease type 1 and sickle cell anaemia. By 1966, Victor McKusick was able to document a catalogue of 1500 single-gene disorders or traits; this formed the basis of the first edition of his book *Mendelian inheritance in man*.

The discovery of the structure of DNA by Watson and Crick in 1953, and the subsequent elucidation of the genetic code through which the DNA sequence specifies the sequence of proteins paved the way for a new molecular era in genetics. Recombinant DNA technology was developed during the 1970s and methods for sequencing DNA were reported independently in 1977 by Fred Sanger, and by Walter Gilbert and Alan Maxam. The polymerase chain reaction – a means of amplifying tiny quantities of DNA – was invented in 1983.