Duncan memorial lecture

Medical genetics, the human genome project and public health

M Bobrow, A H Grimbaldeston

I want to describe, in non-technical terms, what I believe the human genome project and modern genetics are. I will leave it to you to see where the ramifications of this presentation go. Because, although I will give you a few current examples, most of the returns from this new era in human genetics will be reaped in the future, not the present. Indeed, the challenge of connecting the genetics and public health agendas will only be appreciated during the 21st century.

The size of the issue

The body of a human being contains approximately 10^{14} cells. Each one, in order for the whole to function properly, needs an appropriate repertoire of biochemical functions. Every cell has a unique history as to where it came from and how it got to be what it is; the history of a liver cell is not the same as that of a hair root cell. However, they all start with exactly the same genetic information. Every specialised biochemical function a cell performs is encoded within the same set of genes in the nucleus of that cell. Hence, genetics is important because genes determine what cells can do, and what cells can do is what organs, tissues and bodies can do!

In each nucleus there are approximately 100 000 genes. Deoxyribonucleic acid (DNA), the substance of genes, is a code in very simple molecular language. The code consists of four "letters", "A" "T" "G" and "C". The words of this code are three letters (or bases) long, each set of three coding for one amino acid in the resulting protein product. However, there is rather a lot of DNA; each human nucleus contains about three thousand million "letters" occupying approximately two metres of DNA (physical length unmagnified). A visual reconstruction of the DNA can be developed if you imagine it to be around one centimetre thick; it would then form a strand, about 25 000 miles long that would go neatly around the equator.

Until 10 to 15 years ago the problems of isolating and decoding individual human genes seemed insuperable (box 1) and genetics was based almost entirely on inference—genes existed because you could see the effects of what happened when "mistakes" occurred. For example, we knew there was a gene that coded for a blood clotting factor (which we called Factor VIII), not because we knew anything

about that gene, but because we could see that a person suffered from haemophilia (and their blood would not clot) when the gene was defective. Technical advances now enable us to fragment the DNA, choose the parts we are interested in, make multiple copies (to purify them) and determine their base sequence. Putting the sequence back together in the right order again should give a complete picture of the genetic content of a human being. That, in essence, is the human genome project (where "genome" refers to the total genetic complement of an individual or species). However, the word "project" is a bit misleading. A project requires a "plan", a director giving specific instructions and people or organisations working together to complete the scheme. The human genome project is much more fluid, less tangible. It is undertaken largely by academics, doing what academics do best, combining competitiveness with collaboration to achieve results. More recently, industrial partners have shown increasing interest in this field.

The purpose of the human genome project is to find genes and describe what they do. This is an important distinction from some of the previous approaches to genetics because the aim is not to define function at this stage but to focus on mapping the genes (determining where they fit into the total chromosome set) and clarifying their sequence.

In all large major genomes, humans more so than most, genes form a minority of the total DNA, probably less than 10%. In between the genes is what has been called "junk DNA", whose function remains a mystery. That is, an individual gene is not a single recipe, a single coding sequence that makes sense, it is a patchwork of bits that make sense and bits in between, called introns, which do not. These latter pieces are thrown out by the cells during the process of "reading" and translating the DNA to produce proteins.

Box 1 The challenges involved when locating human genes

Many genes: 50 000–100 000
A lot of DNA: 3 000 000 000 base pairs
Most DNA is not "genes"

Department of
Medical Genetics,
Cambridge Institute
for Medical Research,
Addenbrooke's
Hospital, Cambridge
M Bobrow

NHS Executive North
West, 930–932
Birchwood Boulevard,
Millennium Park,
Birchwood,
Warrington WA3 7QN
A H Grimbaldeston

Correspondence to: Dr
Grimbaldeston
(hgrimalb@doh.gsi.gov.uk)

Presented at the University
of Liverpool, 9 December
1996 and accepted for
publication on 17 February
2000.

The current approach to finding genes

POSITIONAL CLONING

Positional cloning has become the standard way of finding disease related genes. This involves “mapping” the gene by correlating the presence of disease with DNA markers through family pedigrees. A search is undertaken to discover portions of DNA where all or most affected people in the pedigree have gene variant A while the unaffected people have a different variant, B. Knowing where these variants are situated in the chromosome complement fixes a map position. It probably does not get you to the responsible disease gene, but gets you within a few million bases. Once this close, you have a “positional fix” and can focus down onto a limited stretch of DNA to find the desired gene.

This technique has produced many of the cloned disease genes, such as cystic fibrosis, Huntington's chorea, fragile X... and many others. However, it is not easy work. Huntington's chorea is an adult onset, progressive, neurological disorder with a very characteristic phenotype (the appearance (physical, biochemical and physiological) of an individual that results from the interaction of their genotype and the environment). Because it is a fairly common illness investigators found a good positional fix quite early on in the process of studying the disease. As a high profile disorder, 8 to 10 laboratories around the world were funded and employed hundreds of people to find the gene. Despite all this effort it took six to seven years to complete the process!

CANDIDATE GENE APPROACH

The other technique, which is currently popular, is called “the candidate gene approach”. Instead of becoming concerned in the early stages of the investigation as to the gene's position (though it is helpful if you know roughly on which chromosome the gene is), the investigator postulates that the disease of interest has biological links with, for example, lipoprotein metabolism. Genes known to be related to lipoprotein metabolism are then reviewed, systematically. By virtue of what is known about the genes function, and what has been guessed about the disorder, these sites are candidates for being the right gene. These are then systematically examined for disease causing changes. A number of genes have been cloned using this technique. However, it is not yet generalisable as, for most diseases and developmental disorders, we have at best modest knowledge of what sort of gene we are looking for.

THE COMPLETE SEQUENCE APPROACH

The third, very broad approach is that advocated by the “gene sequencers”. This group of enthusiasts had a vision, predicated on the belief that if one person could sequence a 100 bases in a couple of hours, then “n” million people, given a few years, would complete the entire human DNA sequence picture. Sense could be made of the whole once it was finished. This approach has an increasing number of adherents as sequencing technology

improves, and progress is being made. However, this is quite expensive science. It costs at least 30 pence per base; 3000 million times 30 works out at to be a total of £900 million—not pocket money!

THE EXPRESSED SEQUENCE APPROACH

The last approach involves isolating and sequencing the DNA bits that are the “real genes”, that is the expressed sequences. These are expressed in the sense that they are the ones the cell uses to make proteins. Cells “use” genes by making temporary RNA copies of the gene's DNA as a template for protein manufacture. Capturing RNA fragments, and deriving the DNA sequence from which they are sourced, is technically feasible and should deliver all the real “gene” sequences, by only looking at about 10% of the total DNA (because none of the “junk” DNA is sequenced).

Progress to date

The number of human genes, which can be clearly defined, rose to just fewer than 2500 between 1985 and 1992. In a recent issue of *Science*, an international consortium of laboratories, which followed the expressed sequence strategy, updated that figure. As of a few months ago, the “expressed sequences” in the public domain were in excess of 450 000.

That is much more than the total number of human genes referred to earlier (50–100 000). The reason for this is that they are the portions of DNA that the cell is using as a gene template. As each gene is made up of many different fragments, different pieces of the same gene are being recognised many times over. So the 450 000 expressed sequences have been organised to see how many clusters on the actual chromosome they fit. It comes to about 49 000 clusters and that number ties in quite nicely with the 50–100 000 genes being searched for.

Does that mean that the task is nearly complete? Well, the easy way to answer that is to go back to the genes that are already known about from other routes, for example, those that code for known enzymes. Of all the genes recognised previously, how many cropped up again in this exercise? About 50–70% were represented in the 49 000 clusters, depending on the particular series looked at.

Therefore, over half of all human genes are probably among those entities that are physically known to exist as pieces of a DNA sequence; and about 16 000, nearly a third of them, have been mapped to individual chromosomes. This represents huge progress in four years, given that in 1992 only 2500 were recognised (box 2).

Box 2 Human gene map 1996¹

Public databases contain 450 000 gene fragments (expressed sequences)
Organisable into 49 625 clusters (genes)
Include about 50–70% “known” genes
About 16 000 mapped

That leaves us today, with an obvious conclusion, which carries with it an essential caveat.

The conclusion is that early in the 21st century we will have a catalogue of the sequence of, for practical purposes, every human gene. This was viewed as a remote possibility 5 to 10 years ago.

The caveat is, we will know the gene sequences, but in general we will not know what the sequences do. This gene finding activity is akin to that of the “explorers” who, some hundreds of years ago, set out to discover the New World. Every time land was found, in the large mass of water upon which they sailed, a flag was planted and the territory claimed. The “explorers” then moved on, knowing nothing about the island, its culture or people. Others came back later to exploit it. That is almost exactly what is happening now, things are being “discovered” but not “explored”.

Medical genetics, the genome project and public health

Having summarised the background, let me give you a medical geneticist's view of the purpose of all this activity and try to link it, and the specialty of medical genetics, to public health practice. A number of inevitable consequences will follow the identification of genes (box 3).

Box 3 Gene identification will lead to:

Better understanding of biology
Better understanding of disease
Better diagnosis
Opportunities for screening and prevention

BETTER UNDERSTANDING OF BIOLOGY AND DISEASE

Perhaps the most important consequence will be a better understanding of biology. This will result in clearer understanding of disease—two sides of the same coin. Gene products are the building blocks of all biological processes so decoding genes is a short cut to understanding the fundamental chemistry of living organisms.

For example, cystic fibrosis is a disorder of a protein that moves chloride ions through membranes. That is not a guess, we know the gene sequence, the protein it makes and what it does.

Similarly, a version of amyotrophic lateral sclerosis (a relatively uncommon, inherited disorder that presents as a form of motor neurone disease), is attributable to a change in a gene that codes for an enzyme that mops up superoxide radicals formed during oxidation. (I doubt if anyone would have looked at the processes of free radical management and late onset motor neurone degeneration, and connected the two).

The most common form of muscular dystrophy that affects young boys is called Duchenne muscular dystrophy. We now know, through genetics, that the gene, which is damaged in this disorder, produces a very large protein called dystrophin. The dystrophin molecule sits just below the surface of the mus-

cle cell with one end bound to actin filaments (which form the skeleton on the inside of the muscle cell). The other end of the molecule is attached to a group of proteins that bridge the cell membrane (whose existence was discovered by clarifying what dystrophin “sticks to”). It is now evident that a complete chain of proteins reach from the interior of the cell, through the muscle cell membrane and into the extracellular matrix. This chain of proteins must have an integral function in keeping muscle membranes intact under the stress of movement. If damaged, the stability of muscle membrane is destroyed and the muscle breaks down. There are several related but different forms of muscular dystrophy that are attributable to damage to various proteins in this chain.

BETTER DIAGNOSIS AND THE GENETICS OF COMMON DISEASES

Another early consequence of gene discovery is better diagnosis, and again I will start by using Duchenne muscular dystrophy as an example. By undertaking simple, reliable tests on a small sample of blood from a young boy it is possible to confirm whether he is or is not suffering from this disease. This is a precise diagnostic tool based on the presence or absence of a gene defect. Using a similar test, the sister of this boy will be able to find out, when she reaches childbearing age, if she could have an affected child. In the same way, Huntington's disease, fragile X, cystic fibrosis and many other illnesses can be diagnosed in affected people and in carriers who are clinically normal. This process allows individuals and families to receive better advice, which carries with it real practical benefits.

Genetics may one day also impact on the diagnosis of more common diseases. Indeed, much of the interest in genetics has now moved from the traditional genetics of simply inherited disorders to complex disorders or the genetics of common diseases.

It has long been known that for most common disorders, siblings of those with a disease are more frequently affected than a member of the general population, and identical twins often share conditions. There is little doubt that the communal genes of family members contribute to this correlation. However, for the majority of conditions it is also recognised that the genes themselves are not sufficient to determine that someone will suffer from a disease—just because a person has an illness does not mean that their identical twin will also succumb.

Modern genetics has not discovered that more diseases are genetic, but has confirmed previous studies showing that some diseases have genetic influences in their aetiology, and has offered techniques that might discover exactly which genes are concerned with which disease. This genetic predisposition, or possession of genes whose particular variants determine a greater or lesser disease susceptibility, works in concert with environmental factors to produce a clinical picture (box 4).

Box 4 Modern genetics has not:

Discovered that more diseases are “genetic”
Diminished the importance of environmental influence in medicine

Let me use the following examples to illustrate the above and show why separating diseases into “common” and “rare” is something of an illusion, serving at worst to discriminate against rare “orphan” diseases that have a considerable impact on people’s lives.

Diabetes is recognised to be a common disease. Indeed, diabetes is one of the few common diseases in which identifying contributory genes has already produced novel results. For example, three genes that are important in “maturity-onset diabetes of the young” or MODY (non-insulin dependent diabetes with onset at a young age), have been isolated, accounting for a small percentage of all diabetics. MODY is a simply inherited dominant disorder; hence a simple genetic disease that presents as diabetes (a common disorder) has been isolated. Should this be classified as a common or a rare disorder? Similarly, colon cancer is very common, yet familial adenomatous polyposis (FAP) is a well known dominantly inherited disease, which, though rare causes colon cancer. Is it possible that a disorder is called “common” because we do not understand enough about it to distinguish its relatively uncommon component causes?

Likewise, within the overall phenotype of breast cancer, there are two or three relatively uncommon, simply inherited dominant disorders. They were not known before, because breast cancer of genetic origin could not be differentiated from other breast cancers. If those same genes, called BRCA 1 and 2, happened by chance to have caused breast cancer and a blue nose, this would have been described a century ago as a dominantly inherited disorder. However, until the “genetic power” was available to find the genes their existence was obscured.

Box 5 Predictive power of a genetic test (the likelihood that a person with a “positive” genetic test will develop the associated disorder).

Duchenne muscular dystrophy 100%
Huntington’s disease (<70 years) 95%
Breast cancer (BRCA) 80%
Alzheimer’s disease (E4 homo) 50%
Ankylosing spondylitis (B27, back pain present) 30%
Diabetes (IDDM DR3/4) 0.5%

Box 5 shows the range of predictive power of a few selected genes in relation to disease. If a genetic test reveals a mutation, with what certainty can it be predicted that the person carrying the mutation will eventually succumb to the disease? If a sample from a boy reveals that some of his dystrophin gene is missing, that boy has, or will develop, muscular dystrophy.

The disease is entirely determined by defects in that one gene. For Huntington’s disease, it is not 100% certain but close; there are a few examples of people with Huntington’s mutations who have reached old age and not shown the disease. However, even if a person has two copies of the Apo E4 variant, a strong predictor of Alzheimer’s disease, and they live to an old age, only about half of them will eventually manifest Alzheimer’s disease. The insulin dependent diabetes mellitus predisposing genes (IDDM) DR 3 and 4, are currently the best genetic predictors of early onset diabetes. However, if you pick out random children with this high risk genotype, the chance that one of them will develop the disease is not much more than one in 200; the relative risk is very high but the absolute risk is small.

Not only does this illustrate some limitations of genetic tests in practice but it also enables the impact of other variables to be evaluated. If the predictive power of one factor is low, it must mean there are other factors at work. Geneticists use the term “penetrance” to describe the likelihood that a genetic change will lead to a particular phenotype. Complex diseases, such as diabetes and dementia, are in general associated with low penetrance genetic variation, because they require the interaction of several genetic and environmental factors.

PUBLIC HEALTH

There are two related areas in which public health may be touched by genetics. One is the potential for setting up screening programmes that identify people at high risk of a particular disorder by testing for genes. The other is occupational health, which presents similar issues.

The prime reason for establishing a screening programme is to target preventive action. For example, if a group of people can be identified who are at high risk of vascular disease because of some unusual property of their lipoprotein metabolism, it might be possible to persuade that group, more effectively than you can others, to change their eating habits. Those at high genetic risk of getting breast cancer may be candidates for increased surveillance. The national screening programme could be reoriented to provide very careful surveillance for a subset of the population who really are at risk of getting breast cancer. Other groups could be informed that the risks associated with screening outweigh the benefits. For defined subsets of the population the cost-benefit ratios of screening programmes will differ considerably.

To justify a population based genetic screening programme of this sort it must be possible to demonstrate real benefit. There are many potential genetic screening programmes that have not been implemented because of doubts concerning their deliverable benefit. An obvious example is Huntington’s disease; although technically feasible, population screening programmes have not been suggested because there is nothing that can currently be done to prevent or ameliorate disease progression after discovery of a positive result.

Should we be screening every adult in the country to detect carriers of cystic fibrosis?

Pilot studies indicate that the great majority of people offered the opportunity to be screened do not take it up—it does not seem to be important enough for them. No one has been sufficiently persuaded of the benefit of screening to mount a really aggressive recruitment drive—and so a social consensus has been reached, at least for now.

Researchers are currently investigating the possibility of genetic screening for breast cancer predisposing genes. Increased surveillance and early treatment may be beneficial, but as yet there is insufficient evidence to prove that the proposed preventative strategies would be effective.

These examples illustrate that a great deal more research is required before the leap can be made from finding a gene to instituting a public health programme. Managing the professional and public pressure to implement screening programmes will often be difficult as the case may not be judged by the efficacy of the proposed programme but the public profile of the disease. If this pressure is not resisted, large opportunity, financial and personal costs will be incurred without commensurate health gain.

The breast cancer genes again provide an excellent example. Let us imagine that treatment with tamoxifen, or some similar drug, was thought to prevent people at high genetic risk of getting breast cancer. Validation would require a randomised trial of those people. This would take a long time and cost a large amount of money. I fear that implementation may not wait for that sort of evaluation. Public health practitioners will have to reconcile the pressures for implementation with the requirement for scientific rigour.

Before concluding I would like to remind you that there are ethical issues that pervade most aspects of this presentation. The majority of these are not insurmountable (box 6).

Conclusion

Genetics is going to transform medicine over the next 10 years. Eric Lander made a perspicacious comparison between the human genome project and the discovery of the periodic table,

Box 6 Ethical issues in medical genetics²

Issues concerning the “value” of life
abortion
disability
eugenics

Issues concerning privacy and consent
rights and duties of family members
consent to testing children
counselling and the right to ignorance
rights of employers and insurers
Patenting of human genes and their derivatives

“The human genome project is the...20th century’s version of the ...periodic table. The building blocks (of biology) rendered finite”.³ Elements of the periodic table are the building blocks of matter, and the physical sciences were transformed once they and their interactions were understood. From that came quantum mechanics and all modern physics and chemistry.

Cataloguing the human genes and eventually their functions, makes biology similarly finite. Our understanding of all biology, including human biology, will grow explosively over the coming decades. That will teach us how to manage, diagnose and treat disease better. Not all diseases will be amenable to a gene based research approach, but many will. Which the important advances are and when they will come, we do not yet know.

It has been a great privilege to give the 1996 Duncan Lecture. I would like to thank Professor John Ashton and the Duncan Society for this opportunity.

Funding: none.

Conflicts of interest: none.

- 1 Schuler GD, Boguski MS, Stein LD, *et al.* A gene map of the human genome. *Science* 1996;274:540–6.
- 2 Bobrow M. *Ethical issues of the human genome project*. Edinburgh: Royal Society of Edinburgh, occasional papers number 5, 1996.
- 3 Lander ES. The new genomics: global views of biology. *Science* 1996;274:536–9.