Br J Clin Pharmacol 1996; **42**: 683–695

Genes, disease and medicine

COLIN W. DYKES
Genomics Research Unit, Glaxo Wellcome Research and Development, Stevenage, UK

1 Information and technologies derived from genomic research are beginning to revolutionize the study of disease. It is now being predicted that the human genome sequencing project will be more than 90% complete by the end of this decade and that most of the major genes involved in common diseases will have been identified by that time. Correlations between genetic mutations, disease susceptibility, and adverse reactions to drugs are already being established and it seems inevitable that this will lead to the development of novel therapies accurately targeted at subsets of patients most likely to show a favourable response.

2 Targeted therapies present a significant challenge to the pharmaceutical industry because the potential market for such drugs is likely to be smaller than for current 'phenotypic' treatments. However, application of molecular genetic technologies may allow parallel reductions in development costs since it should be more straightforward to demonstrate efficacy and safety in accurately selected patient groups, reducing the requirement for large clinical trials.

3 Gene-based diagnostics have the potential to radically improve medical practice, and progress in genetic testing technology has been impressive. Yet there is growing concern that commercial and other pressures may result in genetic testing being made widely available before the biological consequences of mutations in disease susceptibility genes are fully understood and before the legal, ethical and psychological consequences of testing have been fully debated.

Keywords pharmacogenetics genetic screening gene therapy ELSI drug development targeted therapy

Introduction

Human diseases result from complex interactions between genetically-determined biological mechanisms and the environment. Genetic effects are most apparent in diseases such as cystic fibrosis, muscular dystrophies or Huntington's disease that are inherited in a Mendelian fashion. Inherited susceptibility to more common human diseases including hypertension, diabetes, arthritis, and different types of cancer is less obvious, and is more likely to be obscured by environmental influences such as diet and smoking. Nevertheless, the importance of genetic determinants is now widely accepted [1–4]. Even infectious diseases, which clearly have a strong environmental component, occur in the context of host factors, most obviously the immune system [5], that are genetically specified and can show wide functional variation between different individuals.

Enormous effort is now being focused on identifying the genes and mutations associated with predisposition to common diseases in the expectation that this will provide key insights into the underlying pathophysiology as well as generating DNA-based diagnostic and prognostic tools to assist physicians in selection of appropriate therapies. Considering that the gene responsible for Huntington's disease, a Mendelian disorder, was not identified until 1993 [6], some 10 years after the trait had been mapped to chromosome 4, it may seem optimistic to consider attempting the far more complex task of mapping and cloning genes that merely

Correspondence: Dr C. W. Dykes, Genomics Research Unit, Glaxo Wellcome Research and Development, Gunnels Wood Road, Stevenage SG1 2NY, UK

contribute a predisposition to common diseases [7]. Yet there have already been striking successes, notably in Alzheimer's disease, and breast cancer [8–12]. This has been made possible by significant recent advances in key areas of genome research.

Recent progress in genomic research

Genetic maps

Mapping and identification of disease genes by linkage analysis usually begins with a systematic genome-wide search for chromosomal regions that are frequently co-inherited from affected progenitors along with the disease phenotype. This requires DNA markers, from known chromosomal locations spread evenly across the genome, that can be used to distinguish between DNA inherited from affected and unaffected parents (Figure 1). A high-density map of informative markers is also required for other types of genetic analysis such as sib-pair analysis and linkage disequilibrium mapping.

At the outset of the Human Genome Project in 1990 one of the immediate objectives, scheduled for completion during 1995, was the production of a human genetic map with an average marker density of 2–5 centimorgans [13]. In fact this was achieved a year ahead of schedule [14]. Nine hundred and seventy markers were positioned accurately to provide a framework map with an average density of 4 centimorgans. Almost 5000 additional markers were mapped to a lower confidence level to give an overall average density of 0.7 centimorgans.

Production of this map was a major milestone in genome research, providing a universal tool for mapping disease loci to relatively small genomic intervals, using a standard set of reagents. This resource will be crucial for the identification of genes causing predisposition to complex diseases.

Sequencing the human genome

The major goal of the Human Genome Project was to sequence all 3 billion base pairs of human genomic DNA by 2005 [13]. This appeared wildly optimistic to many observers, seemingly requiring dramatic improvements in contemporary technology. Yet it is now being suggested that 99% of the human genome could be sequenced to 99.9% accuracy, within only 5 years, simply by scaling up currently-available methods [15].

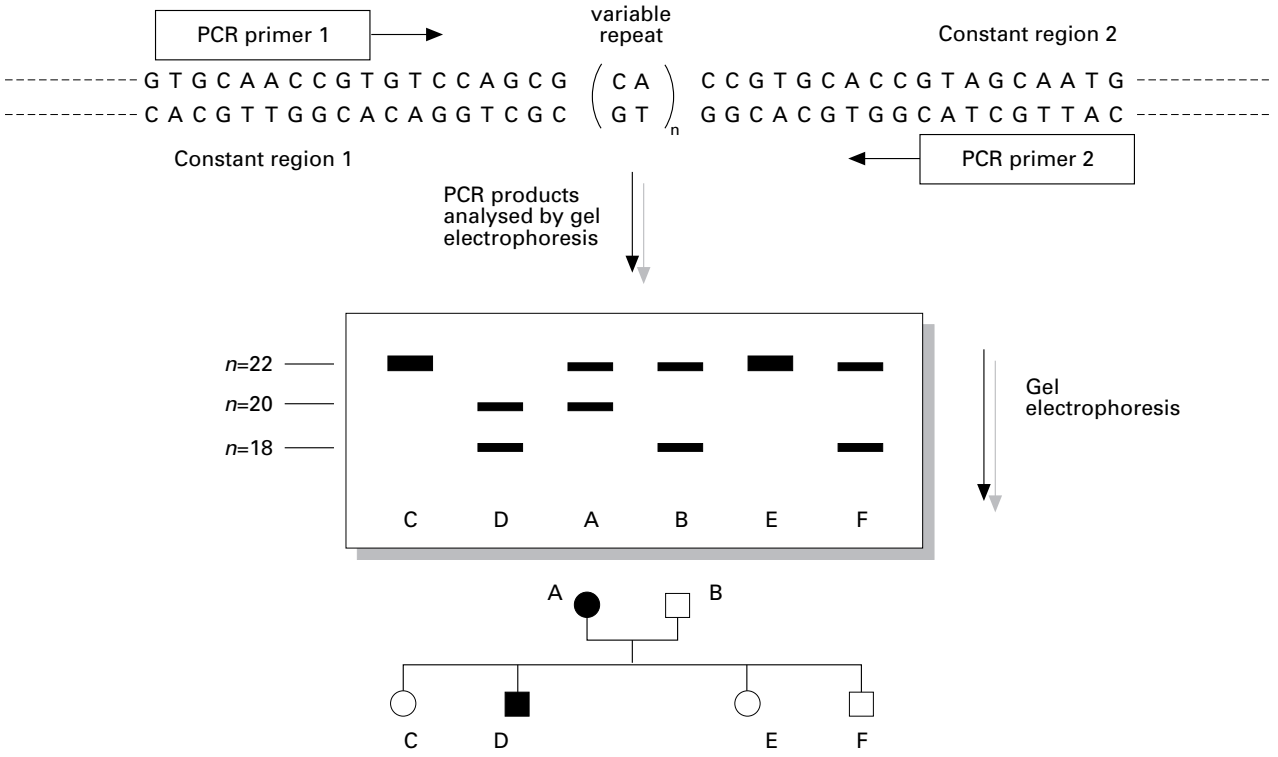

Figure 1 Schematic representation of genotyping using microsatellite markers

Fragments containing microsatellite repeats are amplified from genomic DNA by PCR, using primers that hybridise to the unique, constant regions flanking the variable sequence. The example shown is a di-nucleotide repeat but tri- and tetra-nucleotide repeats are also used. The unique regions identify the chromosomal location containing the marker. PCR products are sized by gel electrophoresis to reveal the size of the repeated region. This marker reveals that son D inherited the disease from his affected mother (A) along with the maternal chromosome carrying the 20-CA repeat unit. The three unaffected children inherited the maternal chromosome carrying a 22-CA repeat unit. (Filled symbols = affected; open symbols = unaffected; circles = female; squares = male). Typically, hundreds of such markers covering the entire genome are tested in large numbers of families to obtain evidence that certain chromosomal regions are co-inherited with the disease more often than would be expected by chance.

Determination of the entire genome sequence will identify the complete repertoire of human genes and their chromosomal locations. However, since most of the genome (around 97%) consists of non-transcribed DNA, genomic sequencing is considered to be a relatively inefficient means of identifying protein-coding genes. An alternative approach is to sequence cDNAs, which are derived only from the transcribed portion of the genome. cDNA libraries, prepared by *in vitro* reverse-transcription of mRNA extracted from particular organs or cell-types, not only provide a rich source of protein-coding sequences, but also information about patterns of gene expression in different tissues. A number of commercial organisations and academic consortia are pursuing this approach vigorously in the expectation that this will lead to the identification of novel genes with therapeutic potential.

DbEST, the largest public-domain sequence database, is expected to contain over 400 000 partial cDNA sequences (also known as ESTs–Expressed Sequence Tags) by the Summer of 1996. The vast majority of these have been contributed by a sequencing programme at the University of Washington in St Louis supported by Merck and Co. Ltd. It is not straightforward to calculate how many of the estimated 60 000 to 100 000 protein-coding genes in the human genome are represented in DbEST because it is possible to generate several non-overlapping EST fragments from a single, full-length cDNA. However, as of May 1996, DbEST contained around 45 000 distinct groups (clusters) of sequences. In the best possible case this could represent 45 000 individual genes, although the actual number is likely to be somewhat lower.

Commercial organisations such as Incyte Therapeutics Inc. (Palo Alto, Ca., USA) and Human Genome Sciences Inc. (Gaithersburg, Md, USA), have accumulated even larger private collections. The HGS database is reported to contain over 850 000 ESTs corresponding to perhaps 60 000 individual genes [16].

Human transcript map

EST databases contain information about the relative abundance of particular cDNA sequences in different libraries, which can provide clues about tissue-specific processes contributing to disease pathology. Similarly, changes in gene expression associated with pathological conditions may be identified by comparing sets of cDNAs isolated from normal and diseased tissues. These databases can also be used to find novel subtypes of receptors, growth factors and other proteins with known commercial importance. However databases *per se* are not particularly useful for establishing *causal* relationships between genes and disease.

The value of these cDNA collections would be further increased if they could be correlated with disease-susceptibility loci. A number of international collaborations are now involved in the production of a 'transcript map' by establishing the chromosomal locations of selected ESTs [17]. The number of mapped ESTs stood at around 1000 at the end of 1994, increasing to over 10 000 by the end of 1995 and is expected to exceed 20 000 by the end of 1996. As new susceptibility loci are identified, the transcript map will indicate which genes are already known to lie in the region of interest, providing an immediate list of candidates to be screened for disease-associated mutations.

The ultimate transcript map will be provided by the completion of the entire human genome sequence, possibly by the year 2001. This will identify any genes not previously mapped as ESTs, including structural RNA genes, and genes expressed at very low levels, that are less likely to be represented in cDNA libraries.

At the current rate of progress there is every likelihood that the large majority of human genes will have been sequenced and assigned to chromosomal loci within the next 2–3 years and that many of the genetic mutations associated with inherited susceptibility to common diseases will also have been identified. Information such as this is already proving valuable in the search for new therapeutic targets and treatments.

Genes and new medicines

Protein therapeutics

Despite pharmaceutical companies' historical preference for small molecules, by 1993 five of the top twenty pharmaceuticals were proteins, based on sales. The world-wide market for 'biopharmaceuticals' expanded from just 5% of the total pharmaceutical market in 1992, to 26% ($9 billion/year) in 1993 [18]. It has been predicted that the US market alone will increase from $5 billion in 1994 to over $15 billion/year by 2004 [19].

cDNA databases and disease genetics programmes will provide a rich new source of novel protein therapeutics, and commercial organisations are racing to exploit this untapped potential. This approach is already beginning to bear fruit. Human Genome Sciences reported recently that they had received three US patents for full-length human genes involved in osteoporosis, inflammation, and cell damage, and had another 150 gene patents pending. These genes were identified initially by analysis of the HGS cDNA sequence database.

There are several precedents that attest to the patentability and commercial potential of protein therapeutics. Successful products such as recombinant human erythropoietin have provided valuable insights into the issues involved in developing and marketing protein therapeutics, and there is no reason to believe that new proteins will face any insuperable obstacles.

Development of antimicrobial vaccines will also receive added impetus from the microbial genome sequencing programmes currently underway or already completed. Interest in vaccine development has been heightened by the emergence of drug-resistant strains of pathogenic bacteria. The genomes of *Haemophilus influenzae* [20] and *Helicobacter pylori* (by Genome Therapeutics Corp., Waltham, Massachusetts, USA)

have already been sequenced. A number of organisations have expressed their intention to sequence additional pathogen genomes during 1996. Analysis of the sequence data will be used to identify cell-surface proteins as candidate antigens for vaccine preparation, and potential targets for the development of novel antibiotics.

Gene therapy

After the excitement of encouraging early results [21], gene therapy now appears to be suffering the consequences of over-optimistic initial expectations. Results from several studies have failed to show clinical benefit to treated patients despite evidence that gene transfer and expression had been achieved to a certain extent in the target tissue [22]. It may be argued that some of these patients had proven resistant to conventional treatment, or were in advanced stages of disease. Furthermore in some of these trials, including treatment of patients suffering from familial hypercholesterolaemia and adenosine deaminase (ADA) deficiency, other therapies were being coadministered. Demonstrating clinical benefit attributable to the gene therapy is likely to be particularly demanding under these circumstances.

The fact that gene transfer and expression have been achieved in target cells does provide cause for optimism. Evidence for gene transfer has been obtained in 28 *ex vivo* studies (where patient cells are first removed from the body, treated with the gene preparation, then re-introduced) and 10 *in vivo* studies [23]. Success has been achieved with each of the three major types of vector preparation, using retrovirus, adenovirus, or plasmid-liposome complexes respectively, to mediate gene transfer (see Table 1 and [24–28]). Furthermore, appropriate biological responses have been observed in some of these trials [23]. Increases in circulating T-cell numbers and T-cell ADA levels were observed following retroviral-mediated *ex vivo* transfer of ADA cDNA into T-lymphocytes isolated from two ADA-deficient children. CFTR (cystic fibrosis transmembrane conductance regulator) cDNA has been delivered to the nasal epithelium (as an accessible surrogate for airway epithelium) in human cystic fibrosis patients, using adenoviral vectors and plasmid-liposome complexes. Partial correction of the abnormal ion-transport phenotype was obtained for 1–2 weeks using both viral and non-viral delivery [23, 29].

Unfortunately, it has not proved possible to achieve such positive results consistently, and many problems remain. There have been reports of inflammation and immune responses associated with the use of adenovirus vectors. The possibility of producing replication competent viruses (by recombination) in therapeutic preparations of supposedly 'safe' retrovirus and adenovirus derivatives continues to be a concern. Plasmid-liposome mediated delivery has the advantage that no infectious agents are used, but the gene transfer efficiencies achieved so far have been disappointing.

Gene therapy has attracted enormous investment of effort and capital, notably from some of the major pharmaceutical companies, and there seems little doubt that, in the longer term, this approach will have a significant impact on the treatment and prevention of human diseases.

Nevertheless, it is clear that significant improvements in targeted gene delivery and expression methodologies are required for clinical applications of gene therapy. This was reflected in a recent report from a committee instituted by the U.S. National Institutes of Health recommending greater emphasis on research applications of gene 'therapy' for identifying disease mechanisms and development of improved animal models, with less emphasis on clinical applications.

Table 1 Main features of common gene therapy vectors

	Retrovirus	*Adenovirus*	*Adeno-associated virus*	*Plasmid-liposome complexes*
Production of therapeutic preparation	Problematical. Low titres. Vector rearrangements.	Good. High titres achievable.	Good. Stable high-titre preparations.	Good, but requires large-scale plasmid DNA purification
Insert size	up to 9kb	up to 7.5 kb	up to 4.5 kb	unlimited
Gene transfer efficiency	Excellent, but replicating cells only.	Excellent, replicating AND quiescent cells.	Excellent, replicating and quiescent cells.	Relatively inefficient, but all cell-types.
Expression	Integrated, stable.	Episomal, transient.	Integration possible, on chromosome 19.	Mainly episomal, transient.
Drawbacks	Possibility of insertional mutagenesis. RCVs possible.	Current vectors produce immune response and inflammation. RCVs possible.	Possibly restricted host cell range? small insert size. RCVs possible.	Inefficient gene transfer. Low level transient expression.
Applications tested	*ex vivo* trials.	*in vivo* trials.	one *in vivo* trial.	*ex vivo* and *in vivo*.

RCV = Recombination Competent Virus

Insights into pathogenesis

Despite decades of biochemical and pharmacological investigation, the mechanisms underlying many common diseases, particularly those affecting the central nervous system, have remained elusive. Increasing emphasis is now being placed on establishing the molecular basis of inherited disease susceptibility as a means of gaining insight into the pathophysiological processes involved. Alzheimer's disease (AD) provides an example of this approach.

Mutations leading to development of AD were first identified in the amyloid precursor protein (APP) gene on chromosome 21 [8]. This is a comparatively rare cause of AD, with only 11 families in the world known to be affected by mutations in APP [30]. Subsequently, it was shown that more than 60% of all cases of AD are associated with alleles of the ApoE gene, which is the major susceptibility gene for late-onset disease [31]. The role of ApoE in the pathology of AD is not understood, although it is known that the ApoE4 allele is associated with increased risk, whereas the E2 allele appears to be protective.

Early-onset autosomal dominant AD accounts for less than 5% of all Alzheimer's patients. However, closer correlation of phenotype and genotype in rare Mendelian variants of complex diseases makes them more amenable to genetic analysis compared with the more common forms. The advantages of this approach were demonstrated in June of 1995, when it was reported that mutations in a gene designated 'S182' (subsequently renamed presenilin-1), on chromosome 14, co-segregated with the disease in certain families [9].

This sparked furious activity in Alzheimer's research. The value of EST databases, when correlated with genetic information, was rapidly demonstrated by the discovery that S182 had sequence homology to a number of other human partial cDNA sequences. One of these was mapped to a region of chromosome 1 known to be associated with early-onset AD disease in another small group of patients [10], and it was shown that mutations in this gene, 'STM2', cosegregated with the disease in the affected families. By the following September it had been reported that a novel gene involved in intercellular signalling in the nematode *Caenorhabditis elegans*, also had sequence homology to S182 [32]. Since there are human homologues of other components of this pathway, it should be possible to use *C. elegans* as an experimentally amenable model to establish the function of S182 and its role in cellular signalling.

Thus, the report describing a susceptibility gene affecting a relatively small number of Alzheimer's patients led, within a period of 4 months, to the identification of an additional AD susceptibility gene, recognition of a family of related human genes, and identification of an experimentally amenable system that may provide crucial insights into a fundamental pathway governing cellular interactions.

Ultimately, information such as this will define the mechanisms involved in Alzheimer's disease at the molecular level. This is already happening in other common diseases, including hypertension. Two important consequences are that it will become possible (a) to redefine disease by mechanism rather than merely by description of the phenotype and (b) to develop novel, mechanism-based therapies targeted at subsets of patients most likely to show a favourable response.

Genetic subdivision of disease and consequences for therapy

It is clear that in many diseases the observed 'phenotype' may result from dysfunction in any one of a number of different biological systems. For example, severe hypertension may result from excess fluid retention, dysfunction in the renin-angiotensin system (RAS), and other factors that affect smooth muscle tone. Although diuretics, angiotensin-converting-enzyme inhibitors, or β-adrenoceptor blockers may be used to treat subsets of hypertensive patients, none of these treatments is effective in all hypertensive individuals.

Heterogeneity such as this is evident in other common diseases but, to date, it has been difficult to identify the different mechanisms giving rise to the common phenotype, and to select subsets of patients likely to respond to particular treatments. This is beginning to change as a result of information coming from molecular genetics programmes. Multiple susceptibility loci have been mapped for hypertension, Type I and Type II diabetes, schizophrenia, breast cancer, colon cancer, rheumatoid arthritis, and as noted above, Alzheimer's disease.

In contrast to Alzheimer's disease, major mechanisms involved in control of blood pressure have been characterised biochemically and pharmacologically. Nevertheless the pathogenesis of hypertension remains unknown in the overwhelming majority of patients [1], because of complex interactions between different physiological pathways.

Recent studies have identified genetic variants in a small proportion of patients with essential hypertension (which accounts for 95% of all hypertensive patients) [33], and other mutations responsible for rare Mendelian disorders of blood pressure regulation. This information can be used to begin grouping subsets of hypertensive patients according to the nature of the genetic lesion, as described below.

Certain alleles of the angiotensinogen gene are associated with elevated plasma angiotensinogen levels [33]. Overexpression of angiotensinogen in transgenic animals has been shown to correlate with increased blood pressure [34] presumably because of the consequent overproduction of the active vasoconstrictor angiotensin II (A-II) (Figure 2). A-II binds to vascular receptors causing vasoconstriction, and to receptors on the adrenal glomerulosa, activating production of the mineralocorticoid, aldosterone. Aldosterone activates the kidney amiloride-sensitive sodium channel (indirectly, acting through mineralocorticoid receptors) which stimulates sodium reabsorption, causing increased plasma volume and, as a consequence, elevated blood pressure. A-II also has a direct effect on sodium reabsorption, mediated by A-II receptors on kidney

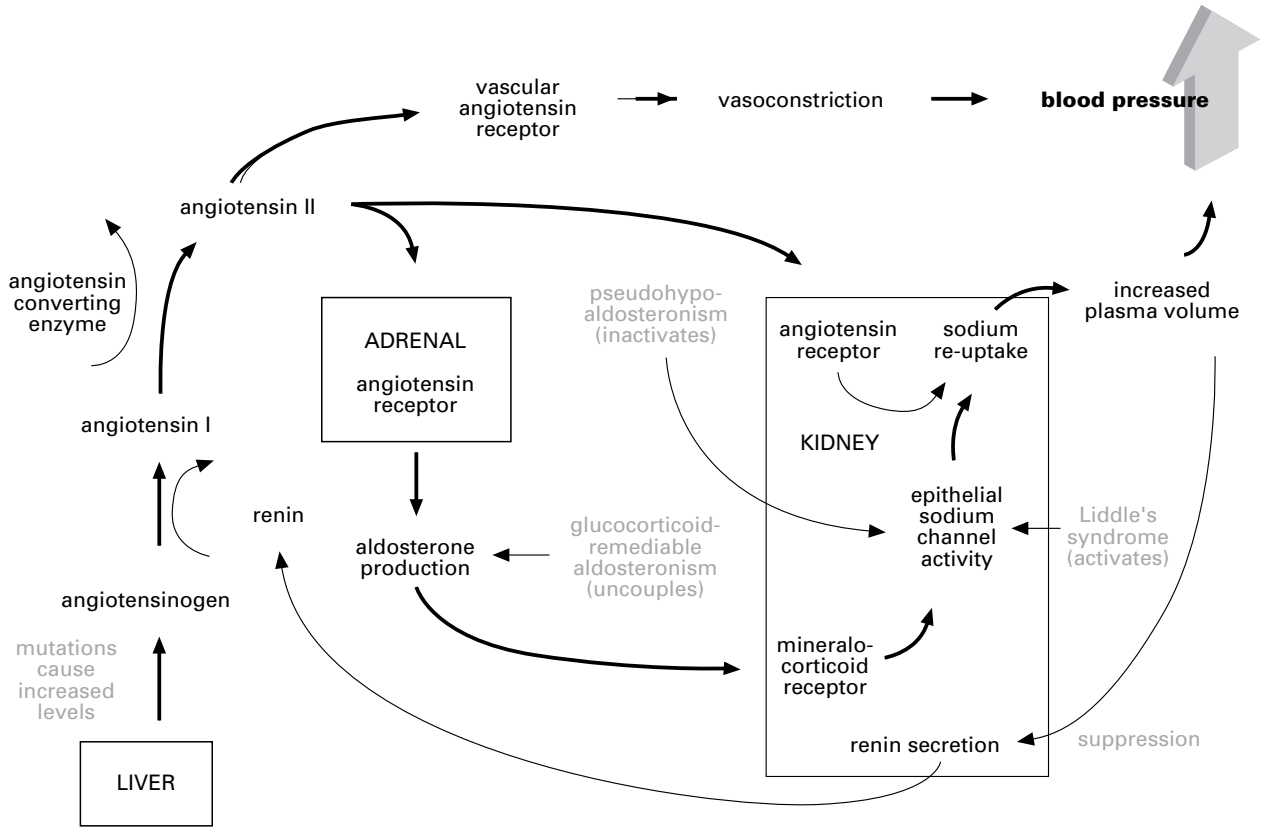

Figure 2 Some of the major pathways involved in blood pressure control

Variants of the angiotensinogen gene are associated with increased serum angiotensinogen that is converted into angiotensin II, causing elevated blood pressure. In glucocorticoid-remediable aldosteronism, a gene duplication event results in regulation of aldosterone production by ACTH, such that feedback control involving reduction in renin secretion, and thus ultimately angiotensin II, is ineffective. Pseudohypoaldosteronism is caused by mutations that inactivate the amiloride-sensitive sodium channel, leading to salt wastage and hypotension. Mutations that activate this sodium channel cause hypertension in Liddle's syndrome.

tubules. A feedback mechanism in the kidney responds to increased plasma volume by suppressing renin secretion, resulting in a reduction in circulating A-II levels, and consequent fall in aldosterone production [reviewed in 1].

However, in glucocorticoid-remediable aldosteronism (GRA), a rare Mendelian trait, hypertension is caused by a gene duplication event that places an additional copy of the aldosterone synthase gene under the control of promoter elements from the 11β-hydroxylase gene, which is regulated by adrenocorticotropic hormone (ACTH), not by A-II. Hypertension in individuals carrying this mutation is caused by sodium reabsorption driven by excess aldosterone. The feedback mechanism involving the RAS is ineffective because reduction in circulating AII has no effect on production of aldosterone from the mutant fusion gene [1].

The kidney amiloride-sensitive sodium channel is a key regulator of sodium balance and mutations affecting the activity of this complex can cause opposite phenotypes. In Liddle's syndrome, constitutive *activation* of the channel by mutations in the genes encoding the β or γ subunits results in excessive sodium reabsorption, leading to hypertension [35, 36]. Different mutations in the α or β subunits cause *loss* of sodium channel function, resulting in salt wastage and *hypo*tension, in patients suffering from pseudohypoalderosteronism type I (PHA-1) [37, 38]. In the latter disease, treatment includes sodium chloride supplementation and dialysis to reduce hyperkalaemia. Death is common in the neonatal period if the disease is not diagnosed.

The different mechanisms causing hypertension in these patients have different therapeutic consequences.

Hypertension resulting from overproduction of angiotensinogen should respond to agents active against the RAS, to reduce the effects of A-II at vascular, adrenal and kidney receptors. Although specific antagonists of the amiloride-sensitive sodium channel should reduce sodium reabsorption in these patients, this would not control vasoconstriction mediated by vascular A-II receptors.

Although RAS inhibitors should reduce sodium reuptake mediated by kidney angiotensin receptors, this has an insignificant effect in GRA patients, in whom sodium absorption is mediated primarily by ACTH-induced overproduction of aldosterone. Instead, GRA may be treated with glucocorticoids, which suppress ACTH secretion; by specific antagonists of the kidney mineralocorticoid (aldosterone) receptor; or by antagonists of the kidney epithelial sodium channel, that controls sodium re-uptake in response to aldosterone [1].

Patients with Liddle's syndrome, unlike those suffering from GRA, respond only to the last of these three therapeutic options, since the sodium channel is constitutively active and unresponsive to aldosterone.

Genetic tests may now be used to identify patients suffering from hypertension due to the different mechanisms described above, allowing early selection of the most appropriate therapeutic options currently available. This approach will not have a significant effect on the overall management of hypertension in the short term since the Mendelian disorders account for only a small proportion of all hypertensive patients. However, as more of the genes contained in known susceptibility loci for hypertension and other common diseases are identified, DNA-based tests will become an increasingly powerful means of identifying sub-groups of patients for whom particular therapies are more likely to be successful.

Commercial implications of mechanism-based therapies

The most immediate application of genetic information will be in clinical trials of compounds already under development, to determine empirically if efficacy or adverse reactions correlate with any mutations in known genes. This will evolve into directed development of drugs targeted against specific mechanisms identified following the discovery of common disease-susceptibility genes. The problem for the pharmaceutical industry is that some diseases may be caused by a number of essentially independent mechanisms that, unlike the RAS and steroid pathways in hypertension, do not interact or converge. If this proves to be common, the logical consequence is that targeted therapies may have to be developed for each distinct mechanism.

On the face of it, the prospect of developing a series of novel drugs targeted at the different mechanisms leading to a disease, instead of having a single therapy aimed at treating the symptoms, does not appear an attractive proposition, because the increased development costs will not be matched by a corresponding increase in market size. However, there are advantages. Since genetic tests can be used to identify those patients most likely to respond to a novel mechanism-specific drug, it should be considerably more straightforward to demonstrate efficacy. Therefore clinical trials could be smaller, with corresponding cost savings. Novel, effective, mechanism-based drugs should appear attractive both to the regulatory authorities, and to the medical community. Patients unsuited to a particular treatment would be spared unnecessary risk of side-effects. Furthermore, the availability of reliable genetic diagnostic tests may allow identification of previously undiagnosed or mis-diagnosed sections of the population that would stand to benefit from new therapies, maximising effective market size. All of these factors should help to ameliorate the negative commercial implications of genetically subdividing patient populations.

Pharmacogenetics

Genetic tests for mutations in disease-susceptibility genes can also be used to screen patients for variants of cytochrome P450 genes associated with adverse reactions, or failure to respond to therapeutic doses of certain drugs. Known examples include alleles of the CYP2D6 and CYP2C19 genes that correlate with altered metabolism of a wide range of compounds. In affected individuals, standard doses may result in plasma drug levels that are too low to produce the desired therapeutic effect or that reach concentrations more likely to produce side-effects [reviewed in 39].

CYP2D6 catalyses metabolism of a wide range of lipophilic drugs, including β-adrenoceptor blockers (e.g. propranolol), antiarrhythmics (e.g. propafenone), antidepressants (e.g. imipramine, fluoxetine), neuroleptics (e.g. haloperidol) and others such as codeine and debrisoquine.

The B allele of this gene, which causes a pronounced poor metaboliser (PM) phenotype, is found at a frequency of around 7% in Caucasians. This allele is much rarer in Oriental populations where the frequency of poor metabolisers is only 1%. Poor metabolisers are more likely to develop adverse reactions to drugs such as neuroleptics, relative to extensive metabolisers.

In contrast, as many as 50% of CYP2D6 alleles found in Oriental populations are accounted for by a gene variant encoding a partially-deficient 2D6 enzyme in which proline residue 34 is mutated to serine [40]. This results in a lower mean CYP2D6 activity in Oriental people compared with Caucasians, which accounts for their relatively slower metabolism of antidepressants and neuroleptics and the lower doses that are used as a consequence [39].

Gene amplification events can produce the opposite effect i.e. the rapid metaboliser phenotype. A patient who required 3–5 times the normal dose of the antidepressant nortriptyline was found to have an extra, active, copy of the CYP2D6 gene [41]. Amplification events resulting in as many as 12 functional copies of CYP2D6 have been reported, producing an even more dramatic rapid metaboliser phenotype [42].

Intra- and inter-ethnic differences in metaboliser phenotypes have also been reported for CYP2C19 [43], which is involved in the metabolism of drugs such as omeprazole (proton pump inhibitor), diazepam (anxiolytic), imipramine and propranolol.

Genetic testing for disease-susceptibility alleles in clinical trials can be coupled with tests for alleles associated with altered drug metabolism, or allergic responses, to account for individuals who experience adverse reactions or failure to respond to the standard dosage.

Logically, this should be extended into the general population, so that patients can be offered tests to identify the treatments most likely to be effective and to predict their expected rate of metabolism of any drug with a narrow therapeutic window. This would seem particularly appropriate in the ethnically-diverse populations typical of modern societies where pronounced differences in metaboliser phenotypes are to be expected. Such tests need be performed only once, and would become part of the patients' records. Genotypes known to be associated with poor metaboliser phenotypes should encourage caution on the part of the medical practitioner, in selecting the initial dosage. A rapid metaboliser genotype would indicate that failure to

achieve a therapeutic effect at standard dosages might be remedied by increasing the dose rather than by trying another drug.

Current molecular genetics and cDNA sequencing programmes are providing an extensive collection of genes involved in drug uptake, metabolism, and clearance, as well as novel genes associated with allergic reactions. It seems inevitable that increased application of genetic testing in clinical trials, using such information, will establish strong correlations between genotype, efficacy and adverse reactions. This will strengthen the case for personalised dosing regimes since it should become possible to predict more accurately the dose required to achieve safe, therapeutic, serum drug levels in individuals carrying particular combinations of genetic mutations. Clearly there is great potential for achieving marked improvements in patient care but the implications for general medical practice are considerable. It is recognised that providing genetic testing on the scale required will be a major challenge, and a number of different approaches are being developed in an attempt to meet the expected demand.

Genetic testing

Trends in genetic testing technology

Until recently, application of genetic testing was restricted largely by the relative lack of information about genetic polymorphisms, but also by technological limitations. Development of the polymerase chain reaction (PCR) was a major advance that made it possible to isolate fragments of human genomic DNA in reasonably large quantities, without the need for molecular cloning [44, 45]. PCR continues to be the most commonly used amplification procedure, although alternative methods that avoid the requirement for thermal cycling have been developed subsequently.

Following amplification of the region of interest from patients' DNA, screening for variation at known polymorphic sites can be performed by a number of different methods, two of which are illustrated in Figure 3. The upper panel depicts a PCR-based method where one of the pair of PCR primers corresponds to a specific, known allele of the polymorphic site. Additional primers, corresponding to other possible variants at this site, are used in parallel reactions. Amplified products of the expected size are obtained only if the allele-specific primer corresponds exactly to a sequence present in the test sample. Production of amplified fragments from two different primers indicates that the individual tested was heterozygous at the test site. Successful priming events can also be detected by measuring incorporation of radiolabelled dATP into the PCR products.

In the second system shown in Figure 3 an oligonucleotide primer is enzymatically extended by one base across a polymorphic site, in four parallel reactions containing individual dideoxynucleotide chain terminators (ddA, ddC, ddG, or ddT). No additional bases can be added once the chain terminator has been incorporated into the primer, since these dideoxy derivatives lack the 3′-hydroxyl group required for formation of phosphodiester bonds. Template-directed incorporation of chain terminator in any of the reactions is detected using a conjugated antibody detection system that gives a colourimetric readout.

Other methods for detecting known mutations include hybridising allele-specific oligonucleotides (ASOs) to immobilised samples of test DNA, and the oligonucleotide-ligation assay (OLA) [46].

Substantial numbers of patients can be screened for single, known genetic polymorphisms by such methods. However, the large number of mutations that have been identified in disease susceptibility genes poses a major problem for large screening programmes. More than 500 mutations are known to occur in the cystic fibrosis transmembrane conductance regulator (CFTR) gene [47] and over a hundred have been found in BRCA1, one of the genes conferring susceptibility to breast cancer [48]. Most of the mutations in CFTR are relatively rare and are usually ignored in genetic testing programmes. Nevertheless, screening large numbers of individuals only for the 'important' mutations remains a daunting task.

An alternative approach is to completely resequence the test gene. This should identify known *and novel* mutations, and will be particularly important if previously undetected mutations modify the effects of known sequence variants. However, complete resequencing is a relatively slow, laborious and expensive procedure using conventional equipment. Although commercially-available automated DNA sequencers can generate around 200–300 bases of sequence data from each of 36 samples in a 2 h run, it may be necessary to amplify and analyse several fragments from each gene to determine the complete sequence, particularly if the coding region is distributed across a large number of exons. Improved automation of the steps involving DNA extraction, template generation, enzymatic sequencing reactions and automated sample loading would make this a more attractive option, but still relatively expensive.

Microchip technology may provide an acceptable solution. A number of commercial organisations are developing microchips capable of resequencing known genes completely [49]. Each glass or silicon chip contains a gridded array of oligonucleotides corresponding to overlapping sequences covering the length of a target gene. Fluorescently-labelled test samples amplified from blood or saliva are incubated with the chips to produce a pattern of hybridisation that can be measured using a laser chip-reader. Mutations are detected by comparing the hybridisation patterns obtained from test samples to that expected from the wild-type gene (Figure 4a). Each chip has sufficient capacity to resequence several entire genes simultaneously in a single individual or to screen for major known mutations only (Figure 4b).

Resequencing genes by hybridisation to gene-specific oligonucleotides can also be performed by the reverse procedure where the PCR-amplified gene fragments are arrayed on nylon membranes, followed by sequential

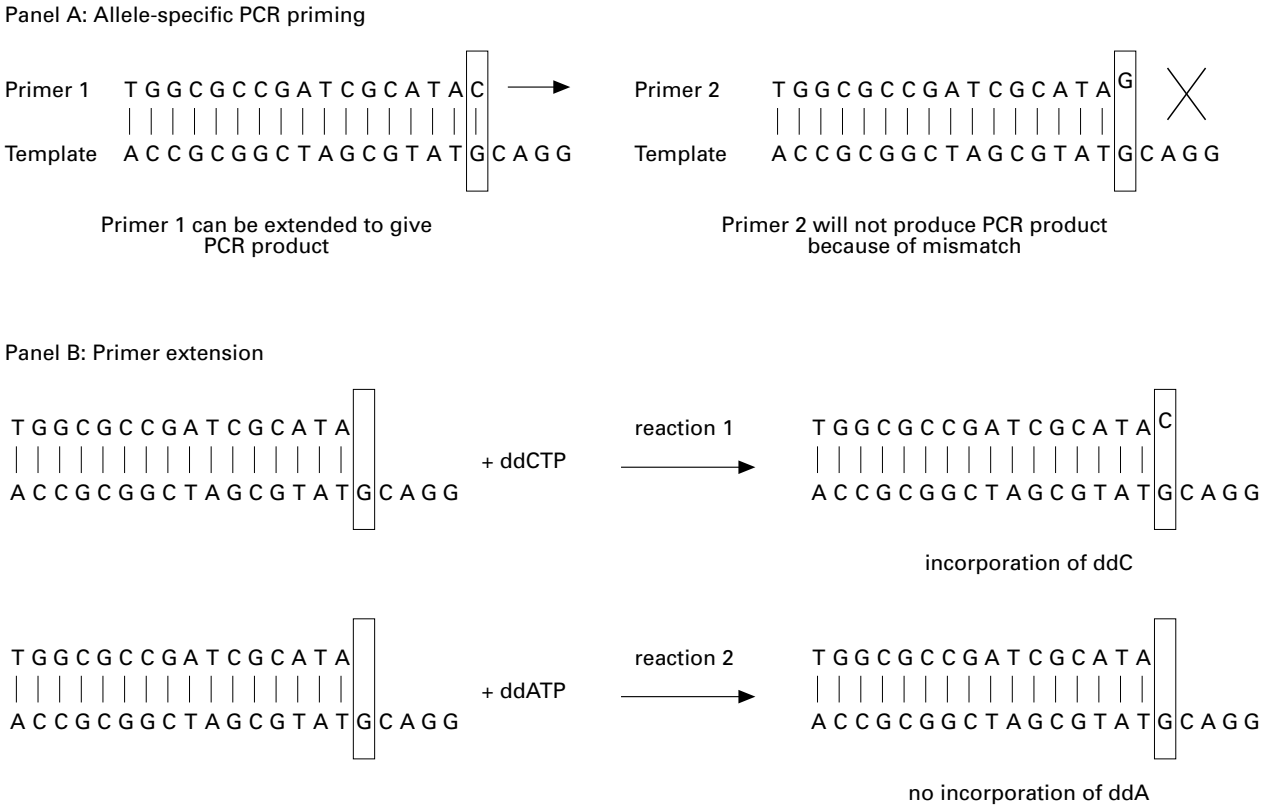

Figure 3 Methods for detecting sequence polymorphisms

Panel A PCR-based polymorphism screening on genomic DNA. Primer 1 (top strand) can be extended and will yield a PCR product on this template, in combination with a suitable second primer. Primer 2 is extended very inefficiently because of the mismatch at the 3'-terminal residue and will not yield a PCR product. Detection of a product with Primer 1 but not with Primer 2 identifies the residue present in the template DNA at this position. Detection of products with *both* primers indicates that the individual is heterozygous at this position.
Panel B Enzymatic extension of the single primer is performed in four parallel reactions containing one of the four dideoxynucleotide chain terminators. Incorporation of one of the four bases, determined by the template sequence, is detected using a conjugated antibody system giving a colourimetric readout.

hybridisation of radio-labelled, or fluorescently-labelled oligonucleotides to the arrays [50]. Over 20 000 PCR fragments can be gridded and sequenced in parallel on a single 22×22 cm membrane. Hybridisation patterns are recorded and decoded to produce the sequence information.

The US Department of Commerce has awarded grants totalling over $50 million to a number of biotechnology companies for the development of genetic diagnostics based on microchip and other advanced technologies. The goal is to develop systems that are at least ten times cheaper than those now in use (e.g. currently $100–$200 per test for cystic fibrosis) and which are convenient enough to encourage widespread use.

Already, prototypes of miniaturised, high-speed PCR instruments are reported to have achieved cycle times of less than 25 s, reducing the time taken for a typical amplification from around 2 h to less than fifteen minutes. Devices that allow PCR-amplification and analysis of the reaction products to be performed directly on microchips are in development. One such system is planned to incorporate DNA extraction, isothermal amplification, and 'dip-stick' detection of target sequences, in a self-contained, disposable cartridge, all for $5.00 per assay. If reduced to practice, devices such as these would be fast enough to allow point-of-care testing for a wide range of mutations. A diagnostic technology that can simultaneously analyse DNA from 500 patients for 106 different mutations in seven genes was described recently [51]. The fact that some of these systems are about to become available, and could be used to select patients most likely to respond to particular treatments, will create pressure for pre-prescription genetic testing to be used for an increasing number of diseases, with profound consequences for medical practice.

Ethical concerns

Offering genetic tests to patients already displaying disease symptoms, should be relatively uncontroversial. Test results would be used to support or refine a medical diagnosis, and to provide additional data about the suitability of different therapies. In the absence of the test, diagnosis and prescription would have to be made anyway, but with less information available.

Presymptomatic testing is rather different. Prophylactic treatment or early therapeutic intervention can be crucially important in many common diseases and the benefit from early diagnosis is obvious. For

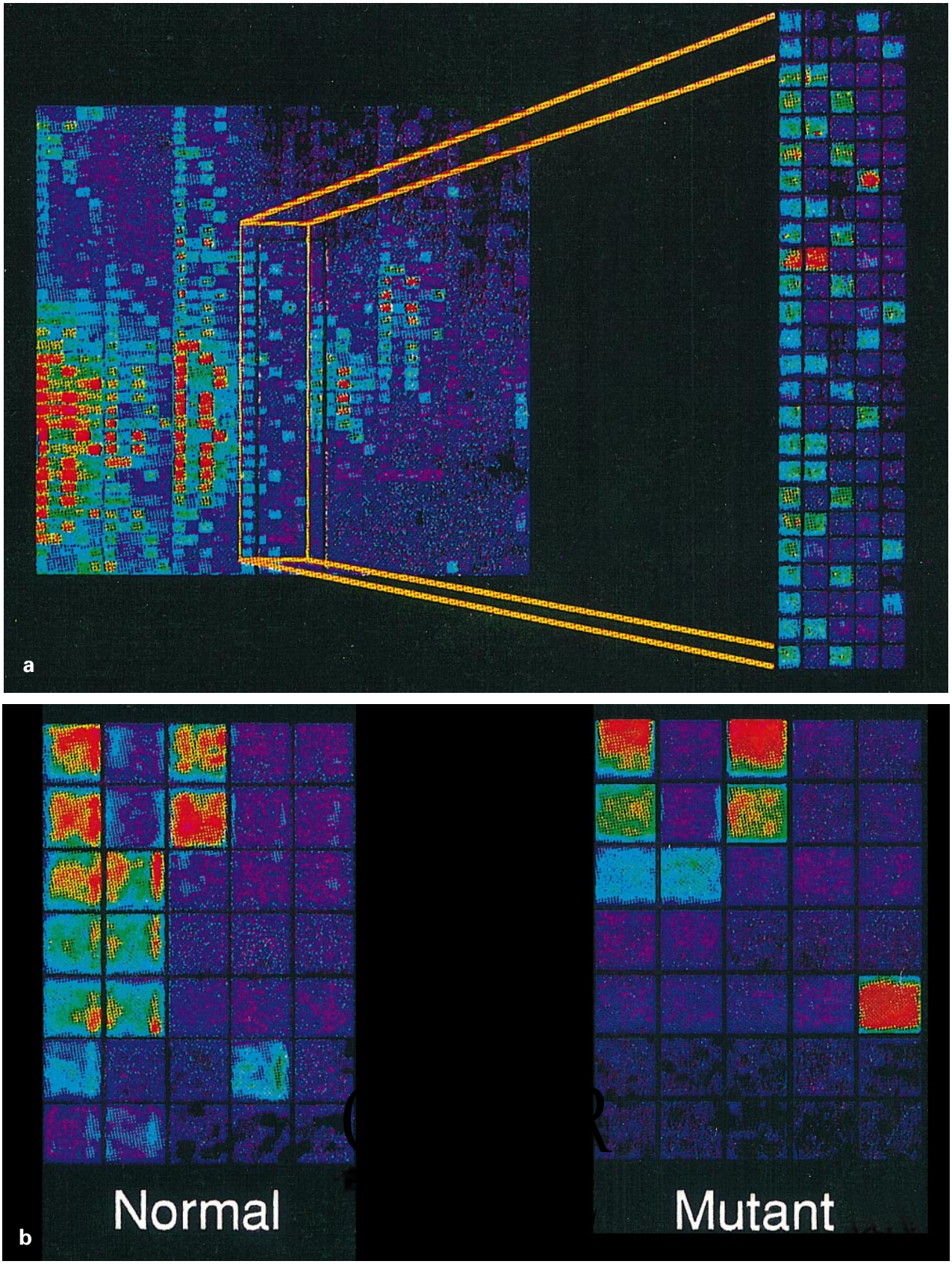

Figure 4 Mutation detection using high-density oligonucleotide arrays

A set of overlapping oligonucleotides, completely spanning the sequence to be tested, are synthesised directly on glass or silicon chips (each small square in the figure above corresponds to a cell containing one oligonucleotide sequence). A fragment to be resequenced is amplified from genomic DNA, and used to prepare a fluorescent probe that is hybridised to the chip. Strength of hybridisation of the probe to each oligonucleotide is visualised as fluorescence intensity using a laser 'chip-reader', and the pattern of fluorescence is decoded to reconstitute the sequence of the original fragment (top panel). Chips can also be designed specifically to detect major known mutations such as the ΔF508 mutation in cystic fibrosis transmembrane conductance regulator (bottom panel). Photographs courtesy of Affymetrix, Palo Alto, CA.

some diseases, an early positive diagnosis may require nothing more than alteration of lifestyle, so that individuals who find that they have a genetic predisposition to hypercholesterolaemia, or emphysema, would be well advised to avoid risk factors such as smoking. However, a positive test result for an allele known to confer a high risk of cancer, or Alzheimer's disease has far more serious implications.

As more disease susceptibility genes, and their causative mutations, are identified, the pressure from some concerned individuals to have such tests performed will increase. Advertisements offering genetic testing directly to private individuals have already appeared in the national press in the UK and USA.

Unfortunately, our ability to perform these tests is already outstripping our ability to understand the biological relevance of the results. For predictions of disease risk based on pre-symptomatic genetic testing to be acceptable there must be clear correlation between known mutations and clinical consequence. Yet this is rarely the case. Even in cystic fibrosis, which is inherited as a Mendelian autosomal recessive trait, there is evidence for heterogeneity in the phenotype expressed by individuals carrying identical CFTR mutations [52]. The vas deferens is functionally disrupted in nearly all males suffering from CF, whereas the degree of pancreatic insufficiency is determined by the patient's genotype. However, the severity of pulmonary symptoms correlates poorly with genotype, presumably because of other genetic or environmental factors that influence lung function.

Similarly, the likelihood of developing a complex disease, and the severity of the symptoms, may vary enormously in individuals carrying identical mutations in known susceptibility genes. Family pedigrees selected for genetic linkage analysis tend to be those that show the clearest predisposition to the disease. It may be that because of their genetic background, these individuals are far more susceptible to the effects of the causative mutation than members of the general population. Therefore it may be necessary to restrict the use of pre-symptomatic genetic tests initially to those families in which a clear correlation between mutations and clinical outcome has been established. Eventually, it will become possible to identify combinations of mutations in groups of genes, that do accurately predict development, progression and severity of inherited disorders, but that lies years in the future.

There is considerable debate about whether pre-symptomatic genetic testing should be restricted to those diseases for which an effective therapy exists. There have been well-documented cases where positive tests for Alzheimer's and Huntington's diseases, for which no effective treatment exists, have driven some patients to the brink of suicide. Yet, in one study involving 135 patients it was found that, a year after testing, the group that had had a clear result in a genetic test for Huntington's disease (37 positive plus 58 negative) scored higher on psychological tests of well-being and lower on tests for depression, than the 40 patients whose test results were ambiguous [53]. For many individuals, uncertainty may have a more debilitating effect than a positive test result. It may be possible to derive greater benefit from testing for such diseases by using psychological profiling to select those patients most able to cope with the trauma of a positive diagnosis.

Nevertheless, there is now a strong consensus against the use of genetic screening for diseases where no effective therapy exists. There is also agreement that genetic testing should not be performed unless it is accompanied by counselling on the significance of test results. However, many commentators doubt the feasibility of providing suitable counselling on the scale required should screening become widespread. These issues will remain at the forefront of the debate on presymptomatic genetic testing, for the foreseeable future.

Genetic discrimination

The potential for misuse of genetic information continues to give cause for concern. A number of surveys have confirmed that families affected by diseases with obvious patterns of inheritance (e.g. Huntington's disease) have difficulty obtaining health insurance [51]. One study reported that 22% of a group of 332 people who had a genetic illness in their families had been refused health insurance. Insurance companies can already identify some people at risk from serious inherited diseases by requesting information about the health, or cause of death, of parents and siblings. Availability of genetic tests for diseases with more complex inheritance patterns may exacerbate this problem. Some insurers have regarded disclosure of a susceptible genotype as being 'evidence of pre-existing disease'. Over half of a group of 917 people who were at risk from an inherited disease reported that they had suffered discrimination after revealing a genetic diagnosis [51].

Early diagnosis and appropriate monitoring or prophylactic treatment can have a considerable impact on morbidity and mortality associated with many common diseases, and also considerable potential for reductions in the cost of healthcare provision. Patients who know that they have inherited susceptibility to colon cancer, for example, can ensure regular screening, and can opt for removal of the colon at the first sign of pre-cancerous changes. Paradoxically, insurance companies may miss an opportunity to cut their healthcare costs, since many individuals who could derive benefit from early diagnosis, and avoid the need for extensive surgery and hospitalisation, may decline to take appropriate tests if disclosure of a positive result may result in subsequent refusal of insurance coverage.

Concerns such as these have persuaded scientists, doctors, patients and other groups to press for laws preventing insurers from denying coverage on the basis of genetic information. Legislation has already been enacted by several US states. US Federal law prohibits genetic discrimination in employment and recently both the US Senate and House of Representatives passed health reform bills including 'genetic information' as a factor that may not be used to deny insurance coverage. Similarly, among European countries, France, Belgium

and Norway have laws preventing the use of genetic information in life and medical insurance.

Conclusion

Information from genomic research offers the potential for enormous improvements in medical practice, not only by providing new therapeutic targets and treatments, but also through the use of genetic testing to target therapies to appropriate patients.

Realisation of this potential may be delayed, or even prevented, by concerns about the exploitation of genetic information and by premature application of genetic testing methodologies before the consequences have been fully considered. Open and informed debate including scientific, medical, commercial, lay, ethical and legislative representation is urgently required to ensure that safeguards agreed to be necessary are put in place. Provision of extra resources for counselling on the significance of genetic test results, and strong legislation to reassure the general public that genetic information will not be abused, would appear to be the minimum required and it is encouraging that a number of governments have already indicated their intention to give this high priority.

Glossary

Allele: Alternative forms of a gene that occupy the same locus.
Centimorgan: A statistical measure of genetic distance, equal to a 1 in 100 chance that two chromosomal markers will be separated due to recombination in a single generation. In humans this averages one million basepairs.
cDNA: (Complementary DNA) Condensed form of a gene, normally lacking introns, produced by *in vitro* reverse transcription of messenger RNA.
DbEST: Computer database of partial cDNA sequences.
EST: (Expressed Sequence Tag) Partial sequence of a cDNA clone, that defines a gene product.
Exons: Regions of a gene containing the protein coding and untranslated mRNA sequences.
Genetic (linkage) map: A map showing the relative positions of traits, genes and other markers. Unlike the physical map, this is a statistical measure.
Introns: Intervening sequences that interrupt the protein coding sequences of a gene. These are usually removed from RNA transcripts before mRNA is translated into protein and therefore are also usually absent from cDNAs.
Marker: An identifiable physical location on a region of a chromosome, whose inheritance can be monitored.
Microsatellite: A stretch of di-, tri-, or tetra- nucleotide repeats in the genome, flanked by unique sequences. The length of the repeat sequence may vary between individuals and can be used to trace the inheritance of regions of specific chromosomes.
PCR (Polymerase Chain Reaction):–a method for amplifying DNA fragments using repeated enzymatic extension of a pair of specific primers.
Phenotype: Physical manifestation (e.g. hypertension) of genetic traits.
Physical map: Figure showing the actual physical locations of cloned genomic DNA fragments relative to chromosomes. There are sets of overlapping cloned genomic DNA that entirely cover some chromosomes, such as 21.
Polymorphism: Inherited DNA sequence differences between individuals.
Positional cloning: Isolation of genes according to their position on a chromosome, rather than by sequence or function.
Transcript map: A table giving the chromosomal locations of genes that are expressed i.e. transcribed into mRNA.

References

1 Lifton R. Molecular genetics of human blood pressure variation. *Science* 1996; **272**: 676–680.
2 Aitman TJ, Todd JA. Molecular genetics of diabetes mellitus. *Baillieres Clin Endocrinol Metab* 1995; **9**: 631–56.
3 Thomson G. HLA disease associations: models for the study of complex human genetic disorders. *Crit Rev Clin Lab Sci* 1995; **32**: 183–219.
4 Sandberg AA, Chen Z. Cancer cytogenetics and molecular genetics: detection and therapeutic strategy. *In Vivo* 1994; **8**: 807–18.
5 Klein J, O'Huigin C. MHC polymorphism and parasites. *Philos Trans R Soc Lond B Biol Sci* 1994; **346**: 351–7.
6 The Huntington's disease Collaborative Research Group. A novel gene containing a triplet repeat that is expanded and unstable on Huntington's Disease chromosomes. *Cell* 1993; **72**: 971–983.
7 Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.
8 Goate A, Chartier HM, Mullan M, *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 1991; **349**: 704–706.
9 Sherrington R, Rogaev E, Liang Y, *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease *Nature* 1995; **375**: p754–760.
10 Levy-Lahad E, Wasco W, Poorkaj P, *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 1995; **269**: 973–977.
11 Miki Y, Swensen J, Shattuck-Eidens DA, *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; **266**: 66–71.
12 Wooster R, Bignell G, Lancaster J, *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995; **378**:789–792.
13 Collins F, Galas D. A new five-year plan for the U.S. Human Genome Project. *Science* 1993; **262**: 43.
14 Murray JC, Beutow KH, Weber JL, *et al.* A comprehensive human linkage map with centimorgan density. *Science* 1994; **265**: 2049–2054.
15 Marshall E. A strategy for sequencing the genome 5 years early. *Science* 1995; **267**: 783–784.
16 Anon. Capitalizing on the genome. *Nature Genet* 1996; **13**: 1–5.

17 Boguski MS, Schuler GD. Establishing a human transcript map. *Nature Genet* 1995; **10**; 369–71.
18 Anon. Biopharmaceuticals market growing fast. *Marketletter* 1994, Issue of June 13th, 19.
19 Shamel RE, and Keough M, Sales of US biopharmaceutical products expected to triple by 2004. *Genet Eng News* 1995; **15**: 6.
20 Fleischmann RD, Adams MD, White O, *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; **269**: 496–512.
21 Miller, A.D. Human gene therapy comes of age. *Nature* 1992; **357**: 455–460.
22 Friedmann T. Human gene therapy–an immature genie, but certainly out of the bottle. *Nature Medicine* 1996; **2**: 144–147.
23 Crystal RG. Transfer of genes to humans: Early lessons and obstacles to success. *Science* 1995; **270**: 404–410.
24 Hodgson CP. The vector void in gene therapy. *Bio/technology* 1995; **13**: 222–225.
25 Treco DA, Selden RF. Non-viral gene therapy. *Molecular Medicine Today* 1995; **1**: 314–332.
26 Gao X, Huang L. Cationic liposome-mediated gene transfer. *Gene Therapy* 1995; **2**: 710–722.
27 Gunzburg WH, Salmons B. Virus vector design in gene therapy. *Molecular Medicine Today* 1995; **1**: 410–417.
28 Blau HM, Springer ML. Gene therapy-a novel form of drug delivery. *New Eng J Med* 1995; **333**: 1204–1207.
29 Wagner JA, Chao AC, Gardner P. Molecular strategies for therapy of cystic fibrosis. *Ann Rev Pharmacol Toxicol* 1995; **35**: 257–276.
30 Sisodia SS, Price DL. Role of the β-amyloid protein in Alzheimer's Disease. *FASEB J* 1995; **9**: 366–370.
31 Strittmatter WJ, Roses AD. Apolipoprotein E and Alzheimer's Disease. *Ann Rev Neurosci* 1996; **19**: 53–77.
32 Levitan D, Greenwald I. Facilitation of *lin-12*-mediated signalling by *sel-12*, a *Caenorhabditis elegans* S182 Alzheimer's disease gene. *Nature* 1995; **377**: 351–4.
33 Jeunemaitre X, Soubrier F, Kotelevtsev YV, *et al*. Molecular basis of human hypertension: Role of Angiotensinogen. *Cell* 1992; **71**: 169–180.
34 Kim S, Krege JH, Kluckman KD, *et al*. Genetic control of blood pressure and the angiotensinogen locus. *Proc Natl Acad Sci U.S.A*. 1995; **92**: 2735–2739.
35 Shimkets RA, Watnock DG, Bositis CM, *et al*. Liddle's syndrome: heritable human hypertension caused by mutations in the β-adrenoceptor subunit of the epithelial sodium channel. *Cell* 1994; **79**: 407–414.
36 Hannson JH, Nelson-Williams C, Suzuki H, *et al*. Hypertension caused by a truncated epithelial sodium channel gamma subunit: genetic heterogeneity of Liddle syndrome. *Nature Genet* 1995; **11**: 76–82.
37 Strautnieks SS, Thompson RJ, Hanukoglu A, *et al*. Localisation of pseudohypoaldosteronism genes to chromosome 16p12.2–13.11 and 12p13.1-pter by homozygosity mapping. *Hum Mol Genet* 1996; **5**: 293- 299.
38 Chang SS, Grunder S, Hanukoglu A, *et al*. Mutations in subunits of the epithelial sodium channel cause wasting with hyperkalaemic acidosis, pseudohypoaldosteronism type 1. *Nature Genet* 1996; **12**: 248–53.
39 Bertilsson L. Geographical/interracial differences in polymorphic drug oxidation. *Clin Pharmacokinet* 1995; **29**: 192–209.
40 Johansson I, Oscarsson M, Yue QY, *et al*. Genetic analysis of the Chinese CYP2D6 locus. Characterisation of mutant genes present in subjects with diminished capacity for debrisoquine hydroxylation. *Mol Pharmacol* 1994; **46**: 452–459.
41 Bertilsson L, Aberg-Wistedt A, Gustafsson LL, *et al*. Extremely rapid hydroxylation of debrisoquine–a case report with implication for treatment with nortryptyline and other tricyclic antidepressants. *Ther Drug Monit* 1985; **7**: 478–480.
42 Johansson I, Lundqvist E, Bertilsson L, *et al*. Inherited amplification of an active gene in the cytochrome P450 CYP2D6 locus as a cause of ultrarapid metabolism of debrisoquine. *Proc Natl Acad Sci U.S.A*. 1993; **90**: 11825–11829.
43 De Morais SMF, Wilkinson GR, Blaisdell J, *et al*. Identification of a new genetic defect responsible for the polymorphism of S-mephentoin metabolism in Japanese. *Mol Pharmacol* 1994; **46**: 594–598.
44 Mullis K, Faloona F, Scharf S, *et al*. Specific amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 1986; **51**: Molecular Biology of Homo Sapiens XXV + 702P.(Part 1); XV + 527P.(Part 2).
45 Saiki RK, Gelfand DH, Stoffel S, *et al*. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988; **239**: 487–491.
46 Nickerson DA, Kaiser R, Lappin S, *et al*. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc Natl Acad Sci U.S.A*. 1990; **87**: 8923–8927.
47 Patrizio P, Zielenski J. Congenital absence of the vas deferens: a mild form of cystic fibrosis. *Mol Med Today* 1996; **2**: 24–31.
48 Hubbard R, Lewontin RC. Pitfalls of genetic testing. *N Engl J Med 1996;* **334**; 1192–1193.
49 Lipshutz RJ, Morris D, Chee M, *et al*. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* 1995; **19**: 442–447.
50 Drmanac R, Drmanac S, Strezoska Z, *et al*. DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. *Science* 1993; **260**: 1649–1652.
51 Beardsley T. Trends in human genetics: Vital Data. *Scientific American* 1996: **274**: 76–81.
52 Dean M, Santis G. Heterogeneity in the severity of cystic fibrosis and the role of CFTR gene mutations. *Hum Genet* 1994; **93**: 364–368.
53 Nowak, R. Genetic testing set for takeoff. *Science* 1994; **265**: 464–467.

(*Received 20 June 1996, accepted 2 July 1996*)