Pharmacogenetics and the practice of medicine

Allen D. Roses

Genetics Directorate, Glaxo Wellcome plc, Greenford, Middlesex UB6 0HE, UK, and Duke University Medical Center, Durham, North Carolina 27710, USA

"If it were not for the great variability among individuals medicine might as well be a science and not an art." The thoughts of Sir William Osler in 1892 reflect the view of medicine over the past 100 years. The role of physicians in making the necessary judgements about the medicines that they prescribe is often referred to as an art, reflecting the lack of objective data available to make decisions that are tailored to individual patients. Just over a hundred years later we are on the verge of being able to identify inherited differences between individuals which can predict each patient's response to a medicine. This ability will have far-reaching benefits in the discovery, development and delivery of medicines. Sir William Osler, if he were alive today, would be re-considering his view of medicine as an art not a science.

Every individual is a product of the interaction of their genes and the environment. Pharmacogenetics is the study of how genetic differences influence the variability in patients' responses to drugs. Through the use of pharmacogenetics, we will soon be able to profile variations between individuals' DNA to predict responses to a particular medicine. The medical significance and economic value of a simple, predictive medicine response profile, which will provide information on the likelihood of efficacy and safety of a drug for an individual patient, will change the practice and economics of medicine. The ability to rapidly profile patients who are likely to benefit from a particular medicine will also streamline drug development and provide opportunities to develop discrete medicines concurrently for different patients with similar disease phenotypes. Other than relatively rare and highly penetrant diseases related to mutations of a single gene inherited in families (Box 1), science has never before had the tools to characterize the nuances of inherited metabolic variations that interact over time and lead to common diseases. Powerful pharmacogenetic research tools are now becoming available to classify the heterogeneity of disease as well as individual responses to medicines.

An ongoing ethical debate concerning potential genetic applications and the impact on individuals and families accompanies scientific advances. Clearly defined terminology should form the basis for informative discussions so that the word 'genetics' is not demonized. For example, tests that are specific to disease genes can help diagnose disease, determine the carrier status of an individual or predict the occurrence of disease. These are quite distinct from profiles that, for example, are specific for genes involved in drug metabolism, which provide information on how a medicine will be metabolized in an individual. In the near future (1–3 years) there will be non-disease- and non-gene-specific pharmacogenetic profiles developed to determine whether an individual is likely to respond to a medicine and/or to not experience serious side effects. Language needs to be more precise so that there can be clarity, especially for public policy debates. Pharmacogenetics is not gene therapy, not genetically modified foods, not genetic engineering, and not cloning of humans or their organs. Ethical, legal and social implications for 'genetic tests' of single-gene mutational diseases should

Box 1
Genes and disease

Single-gene mutations and rare diseases (mendelian inheritance)

Causally related to rare inherited diseases (high penetrance)
Examples:
- Cystic fibrosis (*CFTR* gene)
  - Inheritance: autosomal recessive
  - Location: chromosome 7 (7q31)
  - Mutation: deletion of 3 bp at codon 508 accounts for 70% of mutations
- Huntington disease:
  - Inheritance: autosomal dominant
  - Location: chromosome 4 (4p16.3)
  - Mutation: cytosine/adenine/guanine repeat >35 times

Gene polymorphisms and common disease susceptibility (polygenic; complex inheritance)

Examples of susceptibility loci:
- Common late-onset Alzheimer's disease
  - Susceptibility gene (*ApoE*) on chromosome 19 (19q13)
  - Susceptibility gene locus on chromosome 12 (12q)
- Migraine
  - Susceptibility gene loci on chromosome 19 (19p13); chromosome X (Xq24)
- Non-insulin-dependent diabetes mellitus
  - Susceptibility gene loci on chromosome 12 (12q); chromosome 2 (2q)
- Psoriasis
  - Susceptibility gene locus on chromosome 3 (3q21)

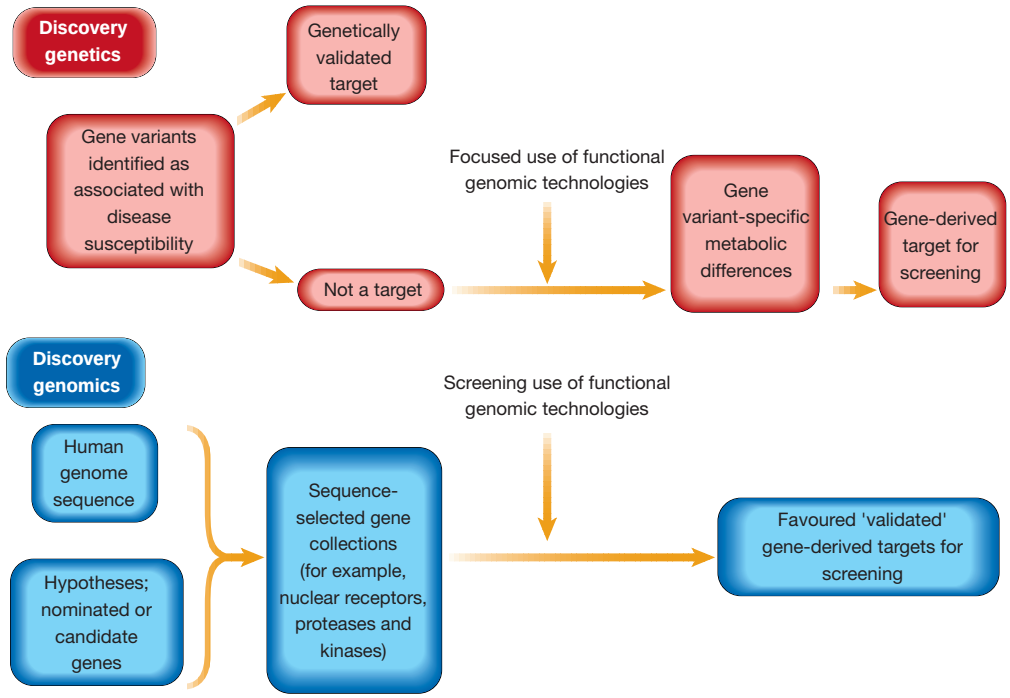


Figure 1 Genetics and genomics to identify drug targets. Two general strategies are used to identify genes and find new targets for drugs: genetics and genomics. Each approach shares technologies, like functional genomics, but as a part of different experimental designs. Genetics identifies disease-related susceptibility genes and genomics identifies genes that belong to similar families based on their sequence homologies. The goal of most genomic strategies is to collect genes that may be expressed and used for high-throughput screening targets. Any one of the identified genes may or may not have a connection to any disease process, with a high probability that it does not. Focused uses of functional genomic technologies include, for example, study of lines of transgenic mice that differ only in the specific polymorphisms defined in the susceptibility gene that relates to disease expression in humans. Understanding isoform-specific metabolic functions can lead to the identification of new metabolic targets for drug screening. Screening use of functional genomic technologies are used to imply validation for targets derived from discovery genomics, such as higher gene expression in a tissue of a subset of genes or the expression of protein observed in disease tissues but not seen in comparable tissue from controls.

not automatically be assumed for other non-disease-specific applications simply because they are labelled imprecisely as 'genetic tests'. Use of inaccurate terminology may hinder and delay the significant health-care benefits that will accrue from pharmacogenetics.

It is important to discuss how the benefits of pharmacogenetics can be applied to drug development and the provision of better health care today — 3–5 years before the widespread application of pharmacogenetics. This will enable the maximum benefits for patients to be obtained as rapidly as possible. In this review I begin with a brief discussion of how genetics and genomics are used in the pharmaceutical industry to identify targets and discover new medicines that will stop or prevent disease processes and then discuss how pharmacogenetics will impact the pharmaceutical industry and the provision of health care.

Target selection

Target validation that will predict a well-tolerated and effective medicine for a clinical indication in humans is a widely perceived problem; but the real challenge is target selection[1–3]. A limited number of molecular target families have been identified, including receptors and enzymes, for which high-throughput screening is currently possible. A good target is one against which many compounds can be screened rapidly to identify active molecules (hits). These hits can be developed into optimized molecules (leads), which have the properties of well-tolerated and effective medicines. Selection of targets that can be validated for a disease or clinical symptom is a major problem faced by the pharmaceutical industry. The best-validated targets are those that have already produced well-tolerated and effective medicines in humans (precedented targets). Many targets are chosen on the basis of scientific hypotheses and do not lead to effective medicines because the initial hypotheses are often subsequently disproved.

Two broad strategies are being used to identify genes and express their protein products for use as high-throughput targets. These approaches of genomics and genetics share technologies but represent distinct scientific tactics and investments. Discovery genetics uses human disease populations to identify disease-related susceptibility genes. Discovery genomics uses the increasing number of databases of DNA sequence information to identify genes and families of genes for tractable or screenable targets that are not known to be genetically related to disease.

The advantage of information on disease-susceptibility genes derived from patients is that, by definition, these genes are relevant to the patients' genetic contributions to the disease. However, most susceptibility genes will not be tractable targets or amenable to high-throughput screening methods to identify active compounds[1,3]. The differential metabolism related to the relevant gene variants can be studied using focused functional genomic and proteomic technologies to discover mechanisms of disease development or progression. Critical enzymes or receptors associated with the altered metabolism can then be used as targets. Gene-to-function-to-target strategies that focus on the role of the specific susceptibility gene variants on appropriate cellular metabolism become important (Fig. 1).

Data mining of sequences from the Human Genome Project and similar programmes with powerful bioinformatic tools has made it possible to identify gene families by locating domains that possess similar sequences. Genes identified by these genomic strategies

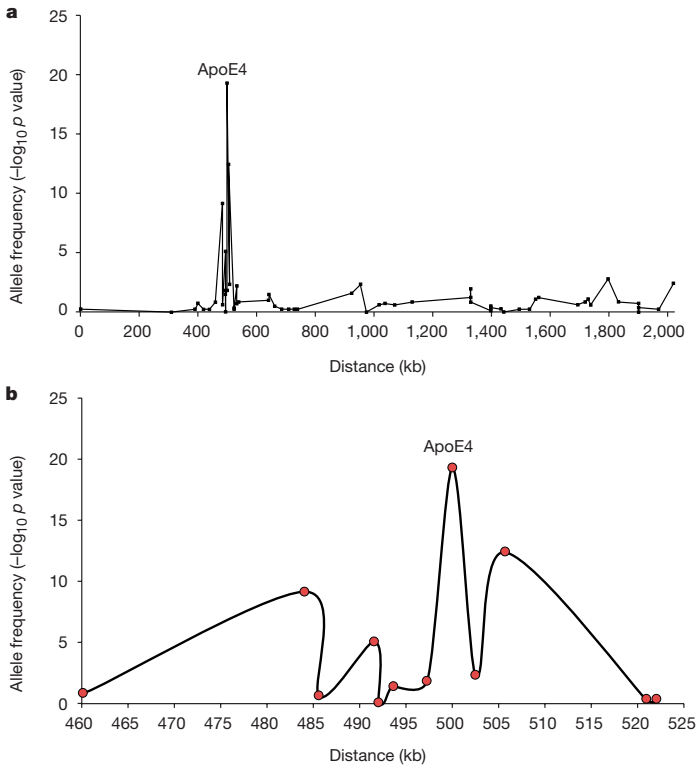


Figure 2 Significance of SNP allele frequency differences in an affected Alzheimer's disease population and age-matched controls. **a**, The association data for dozens of ordered SNPs from a region of 2 million bases on either side of *ApoE*. When the allele frequencies of each SNP are compared in large series of Alzheimer's disease patients and controls, a sharp peak of several SNPs can be readily observed in linkage disequilibrium, with no significant difference in the frequencies of background alleles. (From ref. 45; published with permission.) **b**, If the peak is enlarged to illustrate a region of only ~60,000 bases around *ApoE*, three SNPs from the map that are each highly significantly associated with Alzheimer's disease can be identified. Only two genes, *ApoC1* and *ApoE*, are coded in the physical DNA segment defined by the SNPs associated with Alzheimer's disease. The association data from the SNP defining the *ApoE4* polymorphism, known to be associated with earlier onset of disease, is also illustrated. Not illustrated is the lack of association of another defined *ApoE* polymorphism, that for *ApoE2*. *ApoE2* is associated with protection, or later onset of the phenotype of Alzheimer's disease. Although *ApoE2* is in linkage disequilibrium with *ApoE4*, there is no association with the disease[7,59]. Thus whereas SNPs may be in linkage disequilibrium, as are those for *ApoE2* and *ApoE4*, the association with Alzheimer's disease is found only for several SNPs in linkage disequilibrium with *ApoE4* and Alzheimer's disease. It is the presence of these SNPs that allow rapid recognition of the region within which *ApoE4* is located. (From ref. 45; published with permission.)

generally require some sort of functional validation or relationship to a disease process. Technologies such as differential gene expression, transgenic animal models, proteomics, *in situ* hybridization and immunohistochemistry are used to imply relationships between a gene and a disease process. Over the next five years there will be many opportunities to identify the full complement of gene families. Some of these families can provide tractable targets for high-throughput screening of molecules.

The difference between the genomic approach and the genetic approach is that the former creates a need to functionally validate the tissue distribution and other aspects of each identified gene and find a relevant disease or clinical indication. In contrast, once the disease-related variants of susceptibility disease genes are identified, a single susceptibility gene is automatically validated in human disease. The major distinction between the genomic and genetic approaches is target selection, with genetically defined genes and variant-specific targets already known to be involved in the disease process. The current vogue of discovery genomics for nonspecific, wholesale gene identification, with each gene in search of a relationship to a disease, creates great opportunities for development of medicines. However, there are also enormous economical costs associated with searching huge lists of genes for 'the right disease for the available gene'. It is correct to state that target validation is a major challenge to the pharmaceutical industry, but it is also critical to realize that the core problem for drug development is poor target selection. The screening use of unproven technologies to imply disease-related validation, and the huge investment necessary to progress each selected gene to proof of concept in humans, is based on an unproven and cavalier use of the word 'validation'. Each failure is very expensive in lost time and money.

For example, differential gene expression (DGE) and proteomics are screening technologies that are widely used for target validation. They detect different levels and/or patterns of gene and protein expression in tissues, which may be used to imply a relationship to a disease affecting that tissue[4–6]. Screening with these powerful tools has yet to lead to a specific target for a drug candidate with proven efficacy in humans or to a marketed drug. In fact, the proof-of-concept experiments to demonstrate that differences in the tissue expression of a particular gene are related to disease expression (two very different meanings to 'expression') have not been performed in any common disease with known susceptibility genes. Neither have functional genomic screening methods yet been applied to rare mutational diseases for proof of principle. Rather there has been a tacit and widespread assumption that differentially expressed genes will be related causally to disease progression, rather than as a consequence of disease-related processes. Selecting the right gene using large-scale screening technologies is a significant and expensive problem.

There at least are two common disease examples in which the expression of genetic differences identified by DGE technologies would not have led to target definition. The gene encoding apolipoprotein E (ApoE) is a known susceptibility gene for common, late-onset Alzheimer's disease. Specific allelic variants that are inherited determine the risk and age of onset distribution of the disease[7,8]. Traditional tissue immunohistochemical and *in situ* hybridization studies of the distribution of ApoE have been more revealing than functional genomic screening methods, showing that ApoE is expressed in human neurons under normal conditions, but not in rodent neurons, which are used to model characteristics of Alzheimer's disease[9–12]. Differential expression of total brain ApoE in patients with Alzheimer's disease has not led to the identification of tractable targets. It is highly unlikely (and to date untested as a proof of principle) that β-amyloid precursor protein (APP) or presenilin mutations, each causing rare, early-onset, dominantly inherited Alzheimer's disease, would have been identified using these methods. Yet DGE and proteomic screening methods are currently major investments in several research programmes that work on Alzheimer's disease.

The converse experiment was published recently for an already validated target. Peroxisome proliferator-activated receptor-γ (PPAR-γ) is a nuclear receptor with documented involvement in glucose metabolism and homeostasis[13,14]. PPAR-γ can be considered a precedented target molecule, which can be screened using high-throughput methods for molecules that are effective in treating diabetes mellitus. In this case, there was no previous evidence for PPAR-γ as a susceptibility gene for diabetes mellitus, nor was there any

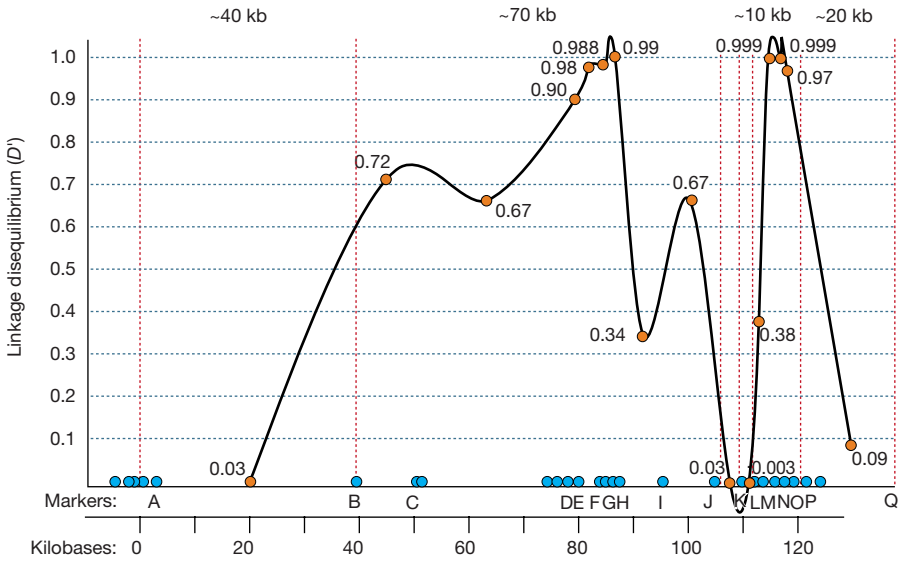


Figure 3 Linkage disequilibrium data for 12 adjacent SNPs that are located and ordered within the 120-kilobase region encoding a migraine susceptibility gene (in this instance, a *D'* value above 0.30 is indicative of highly significant linkage disequilibrium). Five of these 12 SNPs also demonstrated significant association with migraine, illustrating the use of linkage disequilibrium mapping to identify disease-associated polymorphisms.

abnormality in differential PPAR-γ expression. But a rare and severe form of diabetes mellitus has now been shown to be related to specific mutations of the PPAR-γ molecule[15], thus providing further validation of PPAR-γ as a target. There is, however, no indication that DGE screening or proteomic analyses of comparative tissues from common diabetic patients would have identified the precedented molecular target, PPAR-γ. In this case, the genetic data followed validation of a target in humans, and not from differential genomic screening techniques.

The identification of disease-susceptibility genes and study of the function of the susceptibility gene variants will lead to targets that, by definition, will be related to the disease in patients and will therefore be validated. This process identifies few targets compared with the approach used in discovery genomics of data-mining human sequence information. It is therefore practical to use both genetic and genomic strategies and to focus screening technologies to 'pick the winners'.

Pharmacogenetics and medical practice

Diagnosis

When we go to see our doctor, our symptoms and physical signs are evaluated, and appropriate tests (for example, blood, urine, X-ray and magnetic resonance imaging) are undertaken. To the non-physician, this process of disease diagnosis seems straightforward. However, for a patient to have all the classical symptoms and signs of a particular disease is the exception rather than the rule. How these diagnoses relate to the underlying mechanism of disease is often unknown. For example, patients with mutations in different genes may present as clinically identical. Mutations of APP, presenilin 1 and presenilin 2 lead to clinically indistinguishable forms of Alzheimer's disease[16–19]. It is also important to note that mutations at different sites along the *APP* gene can lead to two distinct diseases, early-onset Alzheimer's disease and recurrent intracerebral haemorrhages[20]. For many common diseases, the situation may be assumed to be even more complicated, with many contributing molecular variants of several interacting susceptibility genes leading to multiple clinical effects over varying time frames[7,21–25] (Box 1). Thus many of the diseases that we classify clinically may be syndromes with several distinct contributing pathogenic mechanisms. With all this clinical and genetic heterogeneity we should not lose sight of the fact that the major objective is to treat, cure or prevent disease. It is significant that a medicine works; does it matter whether it is effective in patients who may have different diagnoses? The goal of medicine is to relieve pain and suffering. Similar mechanisms may exist for quite diverse clinical diseases. As the targets and mechanisms are validated in humans, additional clinical indications may become more obvious because of shared mechanisms rather than similar clinical presentations.

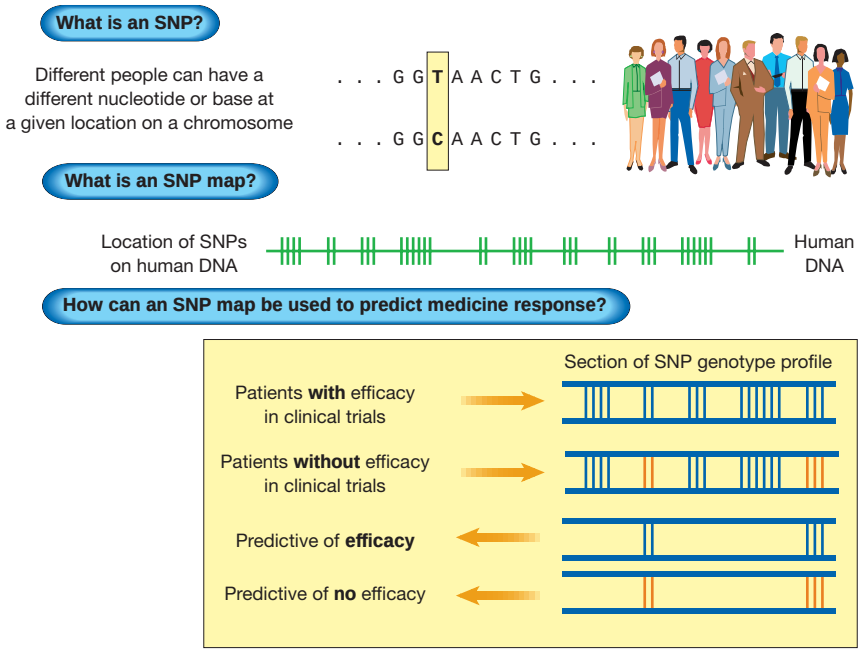
How does your doctor know when making the diagnosis that medicines that are effective for you have not been precluded? Pharmacogenetics will enable individuals to be classified according to their likely response to a medicine. This is not a new concept as clinical subtypes are often classified by drug responsiveness (for example, steroid-sensitive and steroid-resistant asthma). Application of pharmacogenetics will expand the population to those who can be helped but might have otherwise been missed because their clinical syndrome did not fit neatly into a traditional disease category. Alosetron is a recently approved medicine in the United States for the treatment of female patients with diarrhoea-predominant irritable bowel syndrome (IBS)[26,27]. Most physicians will acknowledge that the diagnosis of IBS can be imprecise — in fact, the 'disease' is truly a syndrome. The value of a diagnostic test to sub-classify IBS into different types may be limited, but a simple medicine response profile to determine whether the patient's symptoms will be alleviated by alosetron could have considerable value[28]. Pharmacogenetic approaches will no doubt confirm what clinicians already know — disease diagnosis is not easy nor necessarily homogeneous and accurate.

Apparently distinct diseases may have similar underlying mechanisms. A medicine developed for a specific indication could have value in treating other related or non-related conditions. This is also not a new concept. There are many medicines that were initially registered with a single indication, which have then been expanded as more clinical research is conducted. For example, carbamazepine was initially registered as a treatment for trigeminal neuralgia, a syndrome with intermittent severe lightning-like bursts of facial pain, but was later extended to treat various forms of epilepsy. By understanding the genetic basis of patient responses to medicines, and perhaps also by having a better understanding of how the medicine works, we will be able to identify additional clinical indications more quickly.

Treatment

How does a physician know if the medicine and the dose prescribed will be effective and whether or not the patient will experience adverse effects? Information is available from clinical trials in the medicine's data sheet/label in which similar patients were included and the physician may use experience of treating previous patients. On many occasions, the prescribed medicine will be effective and not cause serious side effects. Other patients may not respond or suffer adverse reactions. By applying the results of pharmacogenetic

Box 2
SNPs and pharmacogenetics



research to clinical practice, physicians will be able to use information from patients' DNA to determine how patients are likely to respond to a particular medicine. The clinical fact that the drug dose for some patients must be individualized has been accepted for years. Polymorphisms in genes encoding P450 enzymes, *N*-acetyltransferase and other key enzymes in drug metabolism account for the concentration variation of certain drugs in patients' blood[29,30]. It is also well established that some patients can be slow in activating drugs and respond inadequately to some prodrugs, or exhibit reduced clearance and increased effects from some pharmacologically active agents[31–33]. Enzyme tests that measure those variants have, in some cases, already been replaced with genetic variants on chips. In the future, metabolic screens of genetic variants will be standardized so that automated read-outs of each person's predicted response to each medicine could be generated. These DNA-based screens will not provide disease-specific diagnosis, but useful information to aid in individual dosing of medications or avoidance of side effects.

SNP mapping: a tool for personalized genetic profiling

Single nucleotide polymorphisms (SNPs) are single-base differences in the DNA sequence that can be observed between individuals in the population[34–36] (Box 2). A polymorphism has been defined as the least common allele occurring in 1% or greater of the population[37], whereas mutations are rare differences which occur in less than 1% of the population (usually much less than 1%). Typically, mutations have been discovered in coding sequences of genes causing rare inherited diseases[38]. SNPs are present throughout the human genome with an average frequency of approximately 1 per 1,000 base pairs (bp)[35]. 'The SNP Consortium' (a consortium of pharmaceutical and bio-informational companies, five academic centres and a charitable trust) is currently producing an ordered high-density SNP map of the human genome (Box 2). Mapped SNPs are being placed regularly into public domain websites (http://snp.cshl.org). The original target was to produce an SNP map with 200,000–300,000 SNPs evenly distributed throughout the human genome. In fact, this initiative is ahead of schedule and will probably provide 600,000–800,000 SNPs by the end of year 2 (April 2001). This map will enable disease and drug response phenotypes to be mapped by linkage disequilibrium. Linkage disequilibrium occurs when haplotype combinations of alleles at different loci occur more frequently than would be expected from random association; it decays with time (generations) in proportion to the recombination fraction between the loci. When alleles are physically close, they are more likely to be inherited together than are alleles that are further apart. Therefore, variations of several ordered SNP markers that are close to, or within, a particular gene variant on a chromosome are likely to be inherited together with that gene variant when they are in linkage disequilibrium. So consecutive SNP variations that are in linkage disequilibrium and associated with a disease phenotype can 'mark' the position on the chromosome where a susceptibility gene is located.

Recent data show the utility of using high-density SNP linkage disequilibrium mapping to find disease-susceptibility genes. Before these SNP mapping experiments, individual testing of multiple candidate genes found to be located within a linkage region was a long, expensive and relatively unproductive way of searching for disease-susceptibility genes.

Polymorphisms of the *ApoE* gene provided the first proof of principle for the detection of a linkage disequilibrium locus around a known susceptibility gene for Alzheimer's disease. In 1997, a high-density SNP map for a region of 4 million bases (0.1% of the human genome) around the *ApoE* locus on chromosome 19 was constructed[39]. The goal of the experiment was to determine whether the *ApoE* gene could be detected as a susceptibility locus associated with Alzheimer's disease using high-density SNP mapping to detect a small region of linkage disequilibrium (Fig. 2a,b)[40]. These studies showed that by using DNA from patients with Alzheimer's disease and controls, it is possible to detect those SNPs in linkage disequilibrium that are associated with the disease.

This methodology has been used to identify susceptibility genes for other diseases such as migraine with aura, which is also localized on chromosome 19 (Fig. 3). In this case a linkage region of approximately 1 million bases was reduced to a 70,000–120,000-bp locus (C.-F. Xu *et al.*, unpublished results). Although this linkage disequilibrium segment of DNA is larger than that found for *ApoE* and

Alzheimer's disease, it contains the coding sequences of a single gene. Similar data have been collected and tested for psoriasis on chromosome 3 (C.-F. Xu *et al.*, unpublished results) and non-insulin-dependent diabetes mellitus on chromosome 12 (E. Lai *et al.*, unpublished results). Thus it is now possible to rapidly reduce the size of the DNA region which contains disease-susceptibility genes by two to three orders of magnitude from millions of base pairs to thousands of base pairs. In practical terms, this accelerates the identification of susceptibility genes within the relatively large regions of DNA that are found by traditional linkage using typical 400 marker screens. Using theoretical, simulated data, some researchers had suggested that one SNP per 6,000 bp would be necessary to locate disease-susceptibility genes[41]. These simulations have been questioned and are not supported by published data or data from mapping of susceptibility genes[42,43]. In particular, this research in these disease areas is a practical demonstration that a density of SNPs of one every 10,000–30,000 bp can rapidly narrow the search for susceptibility genes. After the high-density SNP map of the whole genome is completed, it will no longer be necessary to create SNP maps at each disease locus as they will already exist. Thus the rate of discovery of susceptibility genes will depend on the quality of the patient and control populations, rather than being limited by the technical capacity to construct new, ordered, limited SNP maps. The next technical hurdle will be the development of inexpensive high-throughput methods for scoring large numbers of SNPs from hundreds of patients and controls. Considerable efforts are now underway within the biotechnical community to establish low-cost, high-throughput, accurate SNP scoring technologies.

Determining abbreviated SNP linkage disequilibrium profiles

SNPs are the simplest form of DNA polymorphism. Using currently available DNA analysis systems, such as chip-based resequencing or microsphere-based analytical methodologies, thousands of SNPs can be read out automatically and rapidly[36,44]. By applying whole-genome SNP linkage disequilibrium mapping to patients during phase II clinical trials of a medicine, it may be possible to select multiple small regions from the whole-genome SNP map where SNPs are in linkage disequilibrium and associated with efficacy and common adverse event phenotypes[45]. Selecting only these small regions of SNP linkage disequilibrium into abbreviated SNP linkage disequilibrium profiles (Box 2) will enable more rapid and inexpensive screening of patients who are likely to experience efficacy or adverse events in response to that medicine[46]. Thus whereas the phase II SNP scan might genotype 200,000 SNPs for each patient, the critical data used for identifying markers for efficacy for subsequent phase III clinical trials may use only several hundred SNPs from multiple small regions in linkage disequilibrium and associated with efficacy or adverse events. The abbreviated patterns for efficacy could be extended during large-scale post-approval drug surveillance (see below) to include further efficacy phenotypes and adverse event profiles without providing any significant collateral disease information for relatives regarding inheritance of any specific disease-associated gene allele.

Chip technologies are already available for accurately genotyping hundreds to a few thousand SNPs[36]. The cost of chips as a platform for medicine response profiling is likely to be reduced when analyses of hundreds of thousands of patients are performed once the medicine is marketed. In fact, each chip could contain a panel of abbreviated SNP linkage disequilibrium profiles for several drugs with the same clinical indications so that the most appropriate medicine with that indication for that patient can be determined from a single blood sample.

Similar analyses of patients with identical disease phenotypes could be used to determine disease heterogeneity. Different SNP linkage disequilibrium profiles of patients with the same disease phenotype could define patterns of disease heterogeneity without necessarily identifying the actual genes and alleles involved[36,47]. Genetic research conducted during phase II clinical trials of investigational medicines could use the high-density SNP map of the human genome to identify the sub-type of the disease as well as SNP markers in linkage disequilibrium that correlate with specific responses to the medicine.

Pharmacogenetics and drug development

More efficient clinical trials and enhanced drug surveillance

Application of SNP mapping technologies will enable effective medicines to be developed and made available for clinical use more rapidly. Using abbreviated SNP linkage disequilibrium mapping, medicine response profiles could be identified during phase II clinical trials. These could be used in the selection of patient groups enriched for efficacy in phase III studies. This is likely to make these trials smaller, faster and more efficient[48].

Regulatory agencies would correctly be concerned that there were not enough patients in these streamlined phase III trials to evaluate adverse events, although larger clinical trials that do not select 'efficacy' patients are also unlikely to detect rare adverse events (less than 1 in 1,000). Regulatory authorities would also be apprehensive that, when the drug is marketed, patients who did not meet the pharmacogenetic criteria for prescription may be prescribed the drug without study of their potential benefits or adverse events. However, the risk–benefit ratio for patients with poor efficacy predictions may exclude them from phase III studies on ethical grounds as they would now knowingly be included solely to experience potential adverse events. Furthermore, in clinical practice, access to the medicine could be determined by prescriptions based on pharmacogenetic profiles.

In fact, pharmacogenetic technology may enable a significantly enhanced post-approval surveillance system to be established for approved medicines. Regulatory agencies, pharmaceutical companies and the public recognize the need to improve strategies for drug surveillance[49,50]. In this proposed concept of regulated surveillance, hundreds of thousands of patients who receive the medicine would have blood spots taken and stored on filter papers in an approved location using the original blood sample screened for the initial medicine response profile for efficacy. As rare, serious adverse events are documented and characterized, DNA from patients who experienced the adverse event could be extracted and compared with DNA from control patients who received the drug but did not experience the adverse event. This would enable abbreviated SNP profiles for patients susceptible to the adverse event to be determined. These adverse event profiles would be combined with efficacy profiles to produce a comprehensive medicine response profile. This would allow selection of patients for both efficacy and lower complications of therapy (Fig. 4).

A predictive pharmacogenetic adverse event profile derived from hundreds of thousands of patients taking the drug would be a major advance on the present system of documenting reported serious adverse events during the use of the medicine in clinical practice, as this current system often obtains little or no predictive information to help subsequent patients, other than broad warnings.

Over the next few years, as we approach the ability to differentiate patients by their therapeutic responses, regulatory agencies and pharmaceutical companies will need to work together to pilot and examine methods to evaluate fewer total patients in faster, more efficient clinical trials while enhancing drug surveillance systems. Initial studies using medicine response profiles would no doubt use nested populations of patients within trials designed to meet current guidelines and regulations in order to demonstrate proof of concept.

Medicines for all

The application of pharmacogenetics will not diminish the population in whom a drug is effective, but simply allow prediction of patient response rather than prolonged and expensive prescribing by trial and error. Just as it will be possible to identify patients with drug efficacy, it will also be possible to identify those patients who do not respond early in the process of drug development. The ability to target heterogeneous groups of patients for parallel drug development early, rather than waiting years for non-responsive populations to emerge after

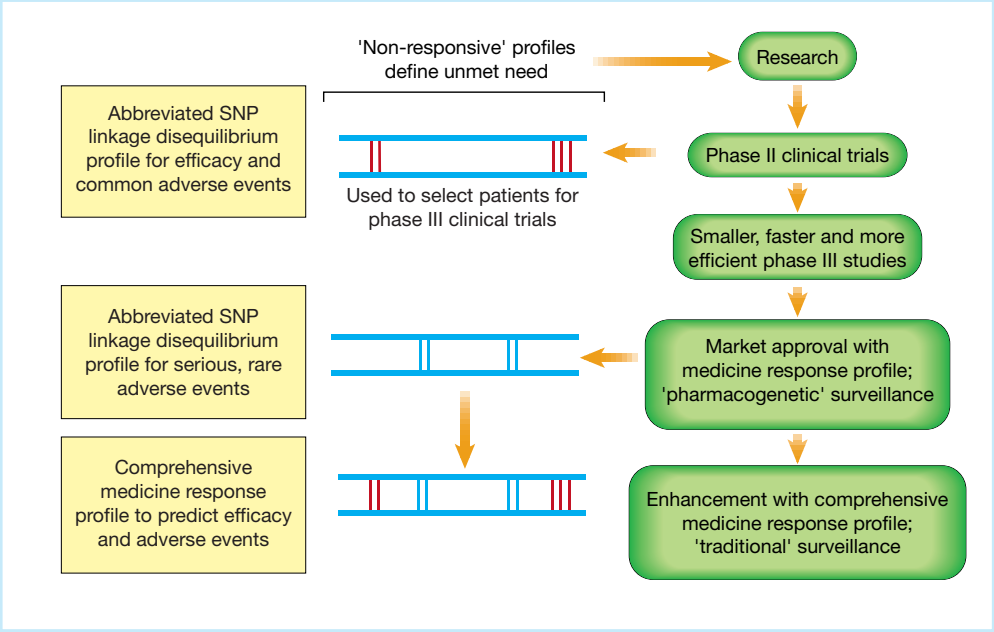


Figure 4 The development of a pharmacogenetic medicine response profile. An abbreviated SNP profile to predict efficacy could be identified in phase II clinical trials by detecting those SNPs along the genome that are in linkage disequilibrium when patients with efficacy are compared with patients who did not respond to the drug candidate. An abbreviated profile of these small regions of linkage disequilibrium that differentiate efficacy can then be used to select patients for larger phase III studies. This could make many of these phase III studies smaller and therefore more efficient. Pharmacogenetics could also be used during the initial post-marketing surveillance period to identify SNP markers associated with serious but rare adverse events. These markers could be added to the SNP markers for efficacy and common adverse events identified during development to produce a comprehensive medicine response profile, and to identify which patients respond to the drug and which patients will be at high risk for an adverse event.

extensive clinical use of the medicine, will be a significant benefit (Fig. 4). For example, SNP profiling of different medicine-responsive association groups during phase II trials will enable identification of the location of genes contributing to heterogeneous forms of the disease, leading to the discovery of new medicines and additional susceptibility targets.

By focusing clinical trials on patients who are most likely to respond, drug development resources could be targeted to those patients with continued unmet medical need. In particular, molecules that show less than a 30% response rate in a large population, but have clear efficacy in an identifiable smaller population of patients, would become viable as they could be readily identified for development and clinical practice.

As a result of disease heterogeneity, there may be large, definable sub-groups of patients suffering with a common phenotype, for example Alzheimer's disease, which represent only 10–15% of patients with that diagnosis. Focusing drug development on sub-groups of patients selected by either a disease-specific diagnostic or a medicine response profile will provide opportunities to develop more medicines for a larger proportion of patients with heterogeneous diseases. Similarly, patient groups who have vaguely defined phenotypes that are more difficult to categorize by objective criteria, such as depression, could be studied more efficiently using medicine response profiles as selection variables.

Value of pharmacogenetics to health-care delivery

The cost-effectiveness of new medicines (which are the product of considerable investment in research and development) is a significant concern to patients, funding bodies and governments[46,51,52]. The application of pharmacogenetics to the delivery of medicines will maximize the value of each medicine. Medicines would be prescribed to only those patients where a high probability of efficacy without significant adverse events is expected[45,46]. This is a much-preferred scenario than the problems facing funding agencies and governments at the present time. Medicines that might be prescribed to 100 patients to achieve an effect in 20 are becoming more difficult for sponsors of medical care to consider. However, selection of predicted responders offers a more efficient and economical solution to a growing problem that is leading governments and health-care providers to deny effective medicines to the few, because a proportion of patients do not respond to the treatment[52]. The economy of predictable efficacy, limited adverse events, lower complications owing to targeted delivery, and increased cost-effectiveness of medicines will improve health-care delivery and eliminate the need for rationing. Effective and well-tolerated medicines with predictive medicine response profiles will obviate the need for formulary restrictions on prescribing and new policies to mandate cost-effectiveness to be proved in a broad population of patients.

Pharmacogenetics will impact medical care at multiple levels. As well-tolerated and effective medicines that treat, cure or prevent common diseases become a greater proportion of the medical care bill, the costs of chronic debilitating illnesses will be significantly reduced. As treatment and prevention of chronic and common diseases improves, a significant proportion of money saved by reducing hospitalization and long-term care costs could be transferred to well-tolerated and effective medicines.

Understanding the differences in 'genetic testing'

The term 'genetic testing' is currently used indiscriminately to refer to very different applications of genetic science. It has entered into common vocabulary with very little specificity surrounding the wide diversity included in this shorthand term. Figure 5 illustrates some of the differences in using the term 'genetic testing'. Until now, government-sponsored committees convened to address 'genetic testing' have generally limited their definition and their reports to concerns regarding diseases caused by single-gene mutations. For example, the US National Institutes of Health Task Force on Genetic Testing and the SACGT (Secretary's Advisory Committee on Genetic Testing), its successor, have dealt mainly with mutational genetics and the need for government oversight in this area. While this objective has considerable merit, it represents only part of the spectrum of 'genetics tests'. Unfortunately, subsequent references to the Task Force conclusions, particularly by ethics commentators, have broadened the limited scope of the Task Force report[53]. Quite distinct differences in recommendations for patients and relatives of patients with complex diseases are frequently miss-stated with authority by authors whose only experience is in mutational diseases[53,54].

Another class of 'genetic tests' is related to pharmacogenetics, including polymorphic detoxifying enzymes, drug-receptor variants or other inherited polymorphic traits that are not diagnostic of

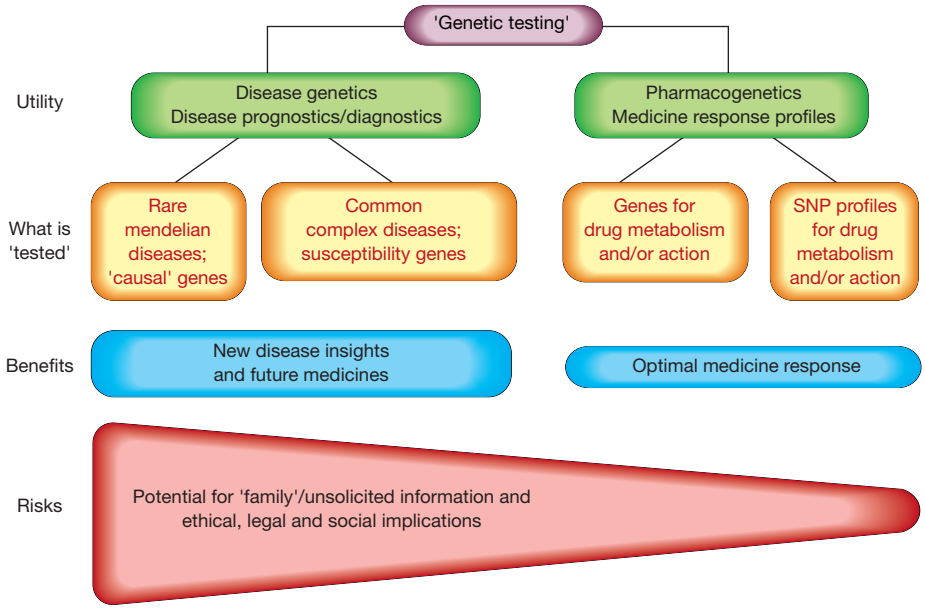


Figure 5 'Genetic testing' needs to be defined carefully. The magnitude of the ethical, legal and social implications of genetic testing is dependent on the information derived from the test. Genetic tests for mutations in single genes that are causally related to rare diseases and are inherited in a simple mendelian fashion can have profound implications for the individual and family members. Genetic tests for disease-susceptibility gene polymorphisms — which are risk factors for the disease — have the added complication of uncertainty. In both cases the lack of effective intervention drives many of the issues. Pharmacogenetic profiles, on the other hand, will predict if an individual patient is likely to benefit from a medicine and be free of serious side effects. These profiles will not be designed to provide any other information, as the profile data are derived from the patients who respond with efficacy or adverse event when taking the drug, compared with patients who did not respond. It does not differentiate disease. Should a polymorphism that is found to be related to disease association be included in a profile, it can be removed and replaced by another SNP that is in linkage disequilibrium, thus avoiding any disease-specific association, even if inadvertent. This would be similar to replacing the *ApoE4* SNP by one or more of the others in linkage disequilibrium with *ApoE4* but not specifically associated with Alzheimer's disease. The ethical, legal and social implications of pharmacogenetic profiles are therefore of a lower magnitude of societal concern compared with specific genetic tests for disease. (From ref. 46; published with permission.)

disease[29–31,55]. In fact, when terms such as 'genetic testing' are applied, differentiation between tests that are specific to disease genes and profiles that are specific to genes involved in drug metabolism are often not well appreciated. Greater specificity of language is required to differentiate tests for disease genes from profiles for non-disease genes. Similarly, distinctions exist between non-disease gene polymorphisms associated with metabolic and drug-target characteristics and extended genomic profiles (for example, abbreviated SNP linkage disequilibrium profiles or medicine response profiles) that simply describe the phenotypic response (efficacy or adverse events) in response to a medicine.

Specificity of language use can be clarified in a hypothetical example. Assume that a 62-year-old man presents with symptoms of dementia that, after a thorough evaluation for other causes of dementia, is diagnosed as 'probable Alzheimer's disease'. If that patient carried an *APP717* mutation or an *ApoE4/4* homozygous genotype, the probability of accurate diagnosis of Alzheimer's disease, defined by subsequent autopsy neuropathologic confirmation, goes from 60–70% at clinical diagnosis to >97%[56,57]. Both are disease-specific diagnostic 'genetic tests' and both provide predictive value in a symptomatic patient, although the *APP717* mutation is generally (but incorrectly) interpreted as being 100% predictive before symptoms begin. It should be noted that there are only two dozen families carrying autosomal dominant *APP* mutations associated with early-onset Alzheimer's disease. Most of these families segregate the *APP717* mutation. Thus there are less than 100 known *APP717* individuals carrying the *APP717* mutation. There are, however, three asymptomatic individuals who are at least one, or two, standard deviations over the mean age of onset for *APP717* mutations. All carry the *ApoE2/3* genotype. To date, no patient with the *APP717* mutation who developed clinical Alzheimer's disease has carried the *ApoE2/3* (or *ApoE2/2*) genotype. *ApoE2/3* seems to protect from the *APP717* mutation. Thus, genetic counselling predictions made from measuring the *APP717* allele should not be made without also considering concomitant *ApoE* genotyping. These data and their significance have been either unknown to or perhaps unappreciated by 'ethics' commentators. *APP717* provides predictive information before any symptoms because it is very rare and disease begins in the 40–60-year age range[17,58]. Carrying two *ApoE4* alleles does not predict Alzheimer's disease, only an increased susceptibility for the development of the disease as a function of age compared with other *ApoE* genotypes. Both are examples of disease gene-specific tests with very different implications for asymptomatic individuals, family members and societal risks of medical-care burden (Fig. 5).

Assume that a hypothetical drug exists for Alzheimer's disease that has two properties. The first is that the half-life of the drug in people varies as a function of a cytochrome P450 drug metabolizing polymorphism. The second property is a greater probability of efficacy in patients with a particular pharmacogenetic profile, that is, there is an abbreviated SNP profile using a panel of 400 SNPs from a map of 200,000 SNPs. 'Genetic testing' using the abbreviated SNP profile could select this particular drug for this patient, whereas a P450 'genetic test' might indicate a higher or more frequent dosing schedule. Neither provides any information about Alzheimer's disease. Neither provides any significant negative collateral information to relatives about Alzheimer's disease. Neither profile has the same ethical implications as measuring a mendelian mutation (*APP717*) or disease-specific susceptibility genotype (*ApoE4/4*). However, all of these are referred to as 'genetic tests'.

The abbreviated SNP linkage disequilibrium profiles will predict patients' responses to medicines, but they will not specifically 'test' the patient for the presence or absence of a disease gene-specific mutation, nor will they provide any other significant disease-specific predictive

information about the patient or family members. For practical purposes they would be anonymous laboratory profiles providing a read-out of predicted efficacy and adverse events. Medicine response profiles will simply measure phenotypic responses to a medicine based on a pattern of inherited factors detected as small regions of linkage disequilibrium. Thus, 'genetic' methods would be used to differentiate those patients who experience good efficacy and lower significant adverse events in response to a medicine from other patients who fail to respond or develop serious adverse events. The genetics of response to the medicine will be the only data generated using an abbreviated SNP linkage disequilibrium profile and, practically, could be easily designed, edited and safeguarded to be totally meaningless with respect to any known disease-specific gene information.

Traditional genetic counselling regarding education about disease inheritance would be of little value to an individual or a relative because no predictive information about disease risk is identified in the SNP linkage disequilibrium profile. Thus, as a practical matter, ethical and legal considerations of disease-specific gene tests, drug target or metabolic gene profiles, and abbreviated SNP linkage disequilibrium profiles for drug response deserve to be considered independently. As the scientific base shifts over the next decade from rare mutational to common diseases affecting millions of people, the rules governing 'genetic testing' should accurately reflect these distinctions. It is therefore incumbent that medical guidelines for mendelian- or susceptibility-gene testing do not extend automatically to discussions of other types of genetically based profiles in pharmacogenetics. Clear language and differentiation of respective ethical, legal and societal issues are required to prevent inaccurate vernacular usage creating a confused public perception of 'genetic testing'. □

1. Curran, M. E. Potassium ion channels and human disease: phenotypes to drug targets? *Curr. Opin. Biotechnol.* **9,** 565–572 (1998).
2. Marton, M. J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4,** 1293–1301 (1998).
3. Wiley, S. R. Genomics in the real world. *Curr. Pharmaceut. Des.* **4,** 417–422 (1998).
4. Blackstock, W. P. & Weir, M. P. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17,** 121–127 (1999).
5. Kozian, D. H. & Kirschbaum, B. J. Comparative gene-expression analysis. *Trends Biotechnol.* **17,** 73–78 (1999).
6. Rockett, J. C., Esdaile, D. J. & Gibson, G. G. Differential gene expression in drug metabolism and toxicology: practicalities, problems and potential. *Xenobiotica* **29,** 655–691 (1999).
7. Roses, A. D. Apolipoprotein E affects the rate of Alzheimer disease expression: beta-amyloid burden is a secondary consequence dependent on APOE genotype and duration of disease. *J. Neuropathol. Exp. Neurol.* **53,** 429–437 (1994).
8. Saunders, A. M. Apolipoprotein E and Alzheimer's disease: an update on genetic and functional analyses. *J. Neuropathol. Exp. Neurol.* (in the press).
9. Boschert, U., Merlo-Pich, E., Higgins, G., Roses, A. D. & Catsicas, S. Apolipoprotein E expression by neurons surviving excitotoxic stress. *Neurobiol. Dis.* **6,** 508–514 (1999).
10. Xu, P. T. *et al.* Human apolipoprotein E2, E3 and E4 isoform-specific transgenic mice: human-like pattern of neuronal immunoreactivity in central nervous system not observed in wild type mice. *Neurobiol. Dis.* **3,** 229–245 (1996).
11. Xu, P. T. *et al.* Regionally specific neuronal expression of human APOE gene in transgenic mice. *Neurosci. Lett.* **246,** 65–68 (1998).
12. Xu, P. T. *et al.* Specific regional transcription of apolipoprotein E in human brain neurons. *Am. J. Pathol.* **154,** 601–611 (1999).
13. Huang, J. T. *et al.* Interleukin-4-dependent production of PPAR-γ ligands in macrophages by 12/15-lipoxygenase. *Nature* **400,** 378–382 (1999).
14. Kliewer, S. A. Lehmann, J. M. & Wilson, T. M. Orphan nuclear receptors: shifting endocrinology into reverse. *Science* **284,** 757–760 (1999).
15. Barroso, I. *et al.* Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402,** 880–883 (1999).
16. Roses, A. D. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* **47,** 387–400 (1996).
17. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349,** 704–706 (1991).
18. Levy-Lehad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269,** 973–977 (1995).
19. Sherrington, R. *et al.* Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375,** 754–760 (1995).
20. Van Nostrand, W. E., Wagner, S. L., Haan, J., Bakker, E. & Roos, R. A. Alzheimer's disease and hereditary cerebral hemorrhage with amyloidosis-Dutch type share a decrease in cerebrospinal fluid levels of amyloid beta-protein precursor. *Ann. Neurol.* **32,** 215–218 (1992).
21. Aitman, T. J. & Todd, J. A. Molecular genetics of diabetes mellitus. *Baillieres Clin. Endocrinol. Metabol.* **9,** 631–656 (1995).
22. Sandford, A., Weir, T. & Pare, P. The genetics of asthma. . *Am. J. Respir. Crit. Care Med.* **153,** 1749–1765 (1996).
23. Schork, N. J. Genetics of complex disease: approaches, problems, and solutions. *Am. J. Respir. Crit. Care Med.* **156,** S103–S109 (1997).
24. Concannon, P. *et al.* A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nature Genet.* **19,** 292–296 (1998).
25. Howard, T. D. *et al.* Genetics of allergy and bronchial hyperresponsiveness. *Clin. Exp. Allergy* **29**(Suppl.), 86–89 (1999).
26. Talley, N. J. Irritable bowel syndrome: disease definition and symptom description. *Eur. J. Surg.* **583**(Suppl.), 24–28 (1998).
27. Paterson, W. G. *et al.* Recommendations for the management of irritable bowel syndrome in family practice. IBS Consensus Conference Participants. *Can. Med. Assoc. J.* **161,** 154–160 (1999).
28. Hamm, L. R. *et al.* Additional investigations fail to alter the diagnosis of irritable bowel syndrome in subjects fulfilling the Rome criteria. *Am. J. Gastroenterol.* **94,** 1279–1282 (1999).
29. Evans W. E. & Relling, M. V. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286,** 487–491 (1999).
30. Meyer, U. A. & Zanger, U. M. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **37,** 269–296 (1997).
31. Cohen, L. J. & DeVane, C. L. Clinical Implications of antidepressant pharmacokinetics and pharmacogenetics. *Ann. Pharmacotherapy* **3,** 1471–1480 (1996).
32. Rudorfer, M. V., Lane, E. A., Chang, W. H., Zhang, M. & Potter, W. Z. Desipramine pharmacokinetics in Chinese and Caucasian volunteers. *Br. J. Clin. Pharmacol.* **17,** 433–440 (1984).
33. Mueller, R. F. & Young, I. D. in *Emery's Elements of Medical Genetics* (ed. Kalow, W.) 169–175 (Churchill Livingstone, Edinburgh, 1997).
34. Zhao, L. P., Aragaki, C., Hsu, L. & Quiaoit, F. Mapping of complex traits by single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **63,** 225–240 (1998).
35. Brookes, A. J. The essence of SNPs. *Gene* **234,** 177–186 (1999).
36. Hacia, J. G. *et al.* Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nature Genet.* **22,** 164–167 (1999).
37. Marez, D. *et al.* Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution. *Pharmacogenetics* **7,** 193–202 (1997).
38. McKusick, V. A. Foreword to *Mendelian Inheritance in Man* 10th edn, vol. 1, xv–xxxvii (The Johns Hopkins University Press, Baltimore, 1992).
39. Lai, E., Riley, J., Purvis, I. & Roses, A. A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics* **54,** 31–38 (1998).
40. Martin, E. R. *et al.* Analysis of association at single nucleotide polymorphisms in the APOE region. *Genomics* **63,** 7–12 (2000).
41. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22,** 139–144 (1999).
42. Collins, A., Lonjou, C. & Morton, N. E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl Acad. Sci. USA.* **96,** 15173–15177 (1999).
43. Ott, J. Predicting the range of linkage disequilibrium. *Proc. Natl Acad. Sci. USA.* **97,** 2–3 (2000).
44. Chen, J. *et al.* Microsphere-based assay for single-nucleotide polymorphism analysis using single base chain extension. *Genome Res.* **10,** 549–557 (2000).
45. Roses, A. D., Manasco, P. & Freeman, A. J. Pharmacogenetics and genetics to discover new medicines. *J. Comm. Biotech.* (in the press).
46. Roses, A. D. Pharmacogenetics and the future of drug development and delivery. *Lancet* **355,** 1358–1361 (2000).
47. Alon, U. *et al.* Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **96,** 6745–6750 (1999).
48. Bonnie, A., Fijal, M. S., Hall, J. M. & Witte, J. S. Clinical trials in the genomic era. Effects of protective genotypes on sample size and duration of trial 2000. *Controlled Clin. Trials* **21,** 7–20 (2000).
49. Lazarou, J., Pomeranz, B. H. & Corey, P. N. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *J. Am. Med. Assoc.* **279,** 1200–1205 (1998).
50. Wood, A. J., Stein, C. M. & Woosley, R. Making medicines safer—the need for an independent drug safety board. *N. Engl. J. Med.* **339,** 1851–1854 (1998).
51. White, J. Targets and systems of health care cost control. *J. Health Politics Policy Law* **24,** 653–696 (1999).
52. Mather, D. B., Sullivan, S. D., Augustein, D., Fullerton, P. & Atherton, D. Incorporating clinical outcomes and economic consequences into drug formulary decision: a practical approach. *Am. J. Man. Care* **5,** 277–285 (1999).
53. McConnell, L. M. B. *et al.* Genetic testing and Alzheimer disease: recommendations of the Stanford Program in Genomics, Ethics, and Society. *Genetic Testing* **3,** 3–12 (1999).
54. Koenig, B. A., Greely, H. T., McConnell, L. M., Silveberg, H. L. & Raffin, T. A. Genetic testing for BRCA1 and BRCA2: recommendations of the Stanford Program in Genomics, Ethics, and Society. Breast Cancer Working Group. *J. Womens Health* **7,** 531–545 (1998).
55. Taningher, M., Malacarne, D., Izzotti, A., Ugolini, D. & Parodi, S. Drug metabolism polymorphisms as modulators of cancer susceptibility. *Mutation Res.* **436,** 227–261 (1999).
56. Saunders, A. M. *et al.* Specificity, sensitivity, and predictive value of apolipoprotein-E genotyping for sporadic Alzheimer's disease. *Lancet* **348,** 90–93 (1996).
57. Mayeux, R., *et al.* Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's disease Centers Consortium on Apolipoprotein E and Alzheimer's disease. *N. Engl. J. Med.* **338,** 506–511 (1998). [Published erratum appears in *N. Engl. J. Med.* **338,** 1325 (1998).]
58. Roses, A. D. Genetic testing for Alzheimer disease. Practical and ethical issues. *Arch. Neurol.* **54,** 1226–1229 (1997).
59. Roses, A. D. Apolipoprotein E is a relevant susceptibility gene that affects the rate of expression of Alzheimer's disease. *Neurobiol. Aging* **15**(Suppl 2), S165–S167 (1994).

Acknowledgements

I acknowledge my colleagues at Glaxo Wellcome in the development of the strategies outlined in this review; in particular, the vision and support of J. Niedel and the commitment of the Glaxo Wellcome R&D Executive Committee — A. Baxter, T. Eaves, A. Hennah and J. Palmer — in using their wealth of experience to help make this vision become a reality. I also thank the Genetics Directorate at Glaxo Wellcome for their expertise and continued hard work. Some of the recent pre-publication work carried out by Glaxo Wellcome scientists is reviewed in this paper. I especially thank A. Freeman, A. Saunders and A. Kidgell for expert editorial assistance.