DDT Vol. 6, No. 4 February 2001

Exploiting human genetic variation in drug discovery and development

Elizabeth C. Jazwinska

'The right drug for the right patient at the right time'

Is this a realistic goal for today's pharmaceutical industry and tomorrow's medical practitioner? Or merely an over-simplistic refrain that can only ever be an unfulfilled dream? Here we discuss the reality behind the dream and illustrate how the analysis of genetic variation is a complex science that has the capacity to make significant contributions to drug discovery and development strategies. An understanding of the impact of human variation must be a central consideration in the future practice of pharmaceutical R&D.

Elizabeth C. Jazwinska
Business Liaison Office
University of Sydney
NSW 2006
Australia
tel: +61 2 9351 7912
fax: +61 2 9351 3636
e-mail:
liz.jazwinska@bio.usyd.edu.au

▼ In its broadest sense, human genetic variation refers to the differences in DNA sequence between individuals. These differences occur most commonly as single nucleotide polymorphisms (SNPs) but they can also be duplications or deletions of single or multiple bps. This review will focus on the SNP and will describe how this form of genetic variation is being exploited to support the drive for novel, safer and more effective therapeutics.

The importance of human genetic variation

It has been estimated that any two unrelated individuals differ by approximately one bp change in every 1000 bp; given that there are approximately 3×10^9 bp in the human genome, this frequency equates to 3×10^6 differences between any two unrelated individuals[1]. Many of the changes encode the visible differences between unrelated individuals; but it is the identification of the genetic factors that contribute to our health and clinical outcome that has become an important challenge for the scientist and a significant debate for society.

Historically, the analysis of human genetics has focussed on the identification of the genes and gene changes responsible for monogenic disorders. The task has been arduous, but there have been many successes and we are now in possession of information on the identification of the molecular defect(s) in over 100 of these disorders (http://www.ncbi.nlm.nih.gov/Omim). Although many monogenic disorders are rare, the identification of the molecular defect is of significant medical importance, offering potential for early presymptomatic diagnosis, an improved understanding of the basis of the disease process, and hope for new cures for debilitating and often fatal diseases. Following the success of the cloning of genes responsible for monogenic disorders, attention has focussed on the identification of the genes responsible for more common polygenic (complex) disorders. This work has revealed that eight out of the ten most common causes of mortality in developed countries have substantial genetic input (http://www.cdc.gov/nchs/fastats/deaths.htm). This statistic implies that the genetic factors responsible for common diseases must themselves be common. It is the identification of the common genetic factors contributing to complex diseases, and also to medically significant complex phenotypes such as drug response, which is proving to be the greatest challenge of human genetics.

The impact of the Human Genome Project

One of the most significant scientific endeavours of recent years is the Human Genome Project – the ambitious project to identify the sequence of the entire human genome[2,3]. The first broad draft has already been completed[4,5], and this landmark endeavour is on track to reach its full completion in the year 2003. Because of the Human Genome Project, there is free access to substantial amounts of sequence information, and this has had a significant impact on the tools and expertise

1359-6446/01/$ – see front matter ©2001 Elsevier Science Ltd. All rights reserved. PII: S1359-6446(00)01642-1

required by human geneticists. There is now a move away from the laborious, costly and time-consuming experimental approach to gene identification, towards bioinformatics-based gene definition. This simple move has dramatically increased the speed with which genes can be identified[6–8]. Before becoming complacent about the resounding success of this project, it is important to recognize its limitations, specifically on the understanding and practice of human genetics. One of the key limitations comes from the origin of the sequence information, much of which has been derived from the analysis of the DNA from just one anonymous individual. In some cases, sequence information has been pieced together from a patchwork of up to perhaps a dozen individuals. This implies that the Human Genome Project is primarily significant in providing the genetic template that defines a human; it will define the gene organization and broad sequence identity, but it will stop short of defining the genetic factors behind human variation.

The impact of SNPs

SNPs, the simple changes of one bp at any point in the DNA molecule, comprise the most common form of genetic variation. Although these single bp changes can occur anywhere in the genome, their position and type defines much of their biological consequence[9–11]. SNPs in non-coding DNA are unlikely to have a discernable impact on the structure and/or function of the expressed protein or its level of expression; by contrast, SNPs within the coding region or control elements of a gene can have a substantial effect on the encoded protein (Fig. 1). SNPs within coding regions (cSNPs) result in either a change to an amino acid in the expressed protein (non-synonymous SNPs) or a codon change with no effect on an amino acid and no change to the expressed protein (synonymous SNPs). The frequency of SNPs varies according to their position and type (Table 1).

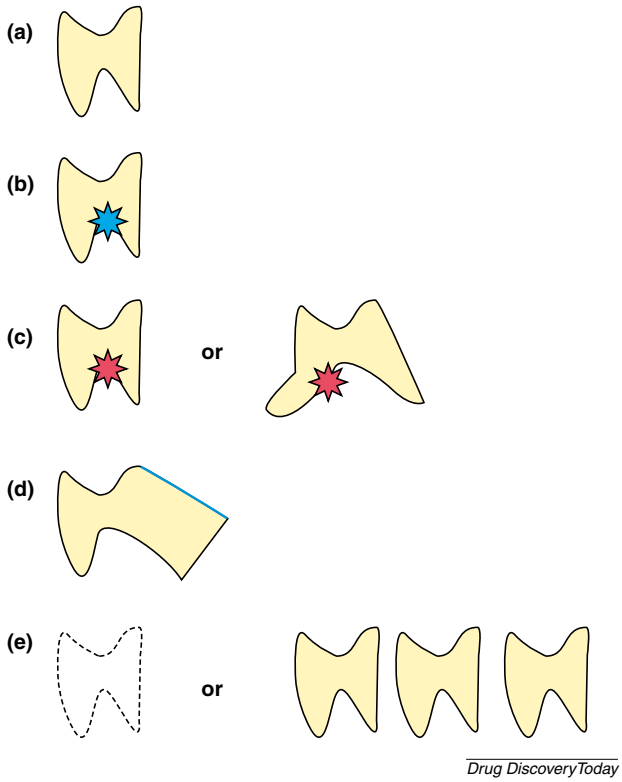

Figure 1. The effects of cSNPs (single nucleotide polymorphisms within coding regions) on proteins. SNPs that do not affect the coding sequences of genes or their control elements are genetically 'silent'. By contrast, cSNPS, which occur within genes or their promoter regions, can have several different effects on the protein: **(a)** the protein encoded by the non-variant or wild-type gene; **(b)** a cSNP resulting in a synonymous codon change results in no change to the amino acid (blue star); **(c)** a cSNP resulting in a non-synonymous codon change results in a change to an amino acid (red star), which can have negligible effect (left) or cause an alteration to the three-dimensional properties of the protein and therefore its function; **(d)** a SNP in a splicing control sequence within a gene can result in a novel mRNA splice product giving rise to a protein with the insertion (or loss) of amino acids (blue); and **(e)** SNPs in promoter regions can alter the transcriptional control of the gene resulting in decreased (left) or increased (right) quantities of protein.

The SNP Consortium

The second part of the Human Genome Project will be the characterization of the range, extent and significance of human genetic variation. This work has already begun and, given the frequency and importance of single bp changes, it is no surprise that the most significant project in this area, being driven by the SNP Consortium (http://snp.cshl.org), has begun by focussing on the characterization of random SNPs in the human genome[12]. This project aims to identify at least one million SNPs with the ultimate goal of providing a densely spaced map of variant markers that will be of value in the mapping and identification of human disease genes. Whether or not the density and position of the SNPs identified will be adequate for the construction of maps to enable gene identification is still the subject of debate[13,14]. In contrast to the sequence data from the Human Genome Project, which has been predominantly derived from one human, the SNP project is scanning for variation in DNA from 10 individuals from different populations and is one of the first significant attempts to identify and exploit wide-scale human genetic variation.

Similar to the Human Genome Project, the SNP Consortium is a major scientific endeavour and an important

Table 1. Estimated frequency of single nucleotide polymorphisms in the human genome according to their position and type

SNP position or type	Approximate no. bp per SNP
Coding	1250
Synonymous	662
Non-synonymous	1754
Non-coding	1176
5′UTR[a]	1470
Intron	952
3′UTR	1190

[a]Data taken from Ref. 1.
Abbreviations: SNP, single nucleotide polymorphism; UTR, untranslated region; bp, base pair.

step towards our understanding of human genetics; but it is important also to realize its limitations. For example, although data from the SNP Consortium will provide the first broad definition of the pattern of genetic variation, it will not provide a full analysis of all the common variation present in humans. Furthermore, although it will identify changes that will provide a tool for the conduct of disease-gene analysis, it does not aim to directly identify the genetic variation that contributes to disease. To date, early analysis of the SNPs identified through the Consortium indicates that out of the 100,000 SNPs documented, only ~250 are non-synonymous cSNPs. If this average is maintained, it is expected that a total of 2500 non-synonymous cSNPs will be delivered; this is a fraction of the expected number of gene changes in humans when taken against the backdrop of 40,000–100,000 genes in the human genome.

Genetic variation in populations

In addition to understanding that genetic variation defines us as individuals, it is important to understand that genetic variation also defines populations[15,16]. This understanding is of key importance to the pharmaceutical industry; if 'the right drug for the right patient' is unrealistic, the right drug for the right population might be a more achievable goal.

How is the analysis of genetic variation being used to discover the best drugs?

It is well recognized that some of the most effective drugs currently on the market were developed through a serendipitous approach to compound identification[17]. The successful treatment of many disorders clearly owes a great deal to the success of this approach and the profits of many companies have been built on these foundations. The use of genetics to identify the molecular defect responsible for disease has led to a quantum change in drug discovery[18].

Target identification

The central hypothesis is that by the use of genetics, an understanding of the underlying molecular defect in common clinical disorders will be achieved, and therefore novel and improved therapeutics will be discovered that will target these defects. To many, particularly those outside the industry, the conventional view of the role of genetics is that it is focussed on the identification of new targets; the reality is somewhat different. First, the genomic revolution has already had a substantial impact on target identification and many companies have been exploiting genomic information successfully for several years. As a result, early discovery portfolios have a plentiful supply of new targets, and the need to identify more novel targets has been superseded by an urgent need not only to trim the portfolio to a manageable size, but also to identify the targets with the greatest potential for success[19]. Further, it is important to realize that although the identification of the molecular defect can improve the early diagnosis of disease and can give clues to the biochemical pathways disrupted in the disease, it does not automatically, or easily, translate into new therapeutics. For example, the genetic basis of sickle cell anaemia has been understood since 1954 but there is still no effective therapeutic strategy.

Given the current target-rich situation 'plaguing' many of today's pharmaceutical companies, the most significant application for the analysis of genetic variation is to contribute information that will assist in identifying the targets, and subsequently compounds, most likely to translate their theoretical potential into practical success.

The time period from the early identification of a potentially important therapeutic molecule through to the successful launch of a new compound on the market is lengthy, often taking at least 10 years. It is arduous, with a dropout rate of 98%, and is extremely costly. The average estimated cost is US$500 million for each compound successfully launched on the market. A key challenge for the industry is to identify potential as early as possible and to invest accordingly. There are two main ways in which analysis of genetic variation can assist in the discovery selection process: (1) target characterization and (2) target validation.

Target characterization

In the context of this article, the term target characterization refers to the use of genetic analysis to define the degree of variation within a gene encoding a molecule identified as a potential drug target.

Historically, pharmaceutical discovery was compound focussed and in many cases the gene encoding the drug target was never identified. Currently, with genomics driving the discovery of novel targets, a single gene forms the foundation for 12 years of R&D. To appreciate the potential impact of genetic variation within a target gene, it is worthwhile reiterating that variation within genes is common; a recent survey of genetic variation within 75 hypertension candidate genes identified SNPs in 74 of the genes, with an average of nine SNPs per gene, two of which were cSNPs[1]. Given these statistics, there is a high probability that variation will be found in any gene selected as a potential drug target, and an integral part of the drug discovery process should thus involve the analysis of the impact of this variation on the selection and subsequent delivery of a compound. The two main aspects relating to genetic characterization of a target gene are:

(1) Definition of variants – this requires that the target gene is scanned for variants, typically non-synonymous cSNPs. The key outcome is the confidence that the most common form of the gene is used for compound selection as a safeguard that the experimental activity of the selected compound will be reflected with actual activity in the population. This can be problematic because the common variant in one population might be the minor variant in another and yet both populations might be significant markets for the compound[20,21].

(2) Definition of impact of variants – one of the most challenging issues facing the discovery scientist is predicting the probable impact of drug target variation on the selection and use of the resulting compound. Currently, the predictive modelling of the impact of variation within a protein partly relies on the crystal structure for the protein, or a highly related protein molecule, being defined. To date, the crystal structure of only a few key proteins have been defined and until more are available, predictive modelling is of limited value. The discovery scientist is faced with deciding whether or not the impact of target gene variation should be tested experimentally; or indeed, whether the potential impact of target variation can ever afford to be ignored.

In addition to the consideration of non-synonymous cSNPs within the target gene, it is important to be aware of the potential impact of variation within the control elements of the gene. For example, variation within the promoter of the gene *ALOX5* is thought to affect the response to anti-asthma treatment by having a direct effect on the amount of protein produced for therapeutic intervention[22]. Definition of the promoter region of a gene is a complex process; currently, there is no easy route and although DNA surrounding the gene can be scanned for probable control elements, this requires analysis of large tracts of DNA; for example, some of the control elements for β-globin are up to 60 kb in distance from the gene.

In summary, the characterization of target variation should be regarded as an important element in early discovery. At least, this analysis will offer the reassurance that the most suitable variants are used for compound selection, an important foundation for successful future investment.

Target validation

The term 'target validation' refers to the process by which genetic variation can be exploited to show the association between a particular target and a disease process.

Evidence supporting the therapeutic value of a compound is typically gathered through clinical trials and requires a substantial investment of time and money. To provide this information before the clinical trial process would be advantageous to a pharmaceutical company[18]; such information can provide reassurance that future investment in a compound will yield a suitable return and can define the most appropriate clinical entities to be tested for in the development process. In its simplest form, the analysis of genetic variation can assist in this process by demonstrating that variation within the target gene is either directly contributing to, or is associated with, a disease or clinical process suggesting that pharmaceutical intervention might have a useful end-point. First, large well-characterized clinical populations that are of specific relevance to the predicted therapeutic value of the test compound should be identified. Variants within, or close to, the test gene are identified and the association between these variants and the presence of a specific clinical marker is analysed. A positive association between a target gene variant and a clinical measure can provide important support for the clinical relevance and therapeutic value of the compound. The key feature of this analysis is that it does not require that patients are treated with the compound; it can therefore be performed at an early stage in discovery and, if successful, can provide more support for the decision to continue with a specific target through the discovery process.

How is the analysis of genetic variation being used to help in the drug development process?

Drug response, efficacy, safety and the study of pharmacogenetics

The analysis of genetic variation in the drug development process has attracted a great deal of attention in the past few years and much has been written about the promise of pharmacogenetics. In its simplest interpretation, pharmacogenetics can be defined as the analysis of inherited

factors that define an individual's response to a drug. Differences in drug response between individuals can be attributed to factor(s) affecting drug absorption, metabolism or excretion, which, in turn, affect the safety and efficacy of a specific drug in a specific individual. To date, the most well-characterized genes affecting drug response are the drug metabolism genes, in particular those in the phase I cytochrome P450 gene family, such as CYP2D6. CYP2D6 (debrisoquine hydroxylase) metabolizes ~25% of prescribed drugs and is inactive in 6% of the Caucasian population[23]. The analysis of the genetic basis of drug response has potential in both the development process and in healthcare delivery[24–29]. In reality, there are still major challenges impeding the progress of pharmacogenetics.

At first glance, pharmaceutical companies appear to be ideally placed to carry out pharmacogenetic analysis; they are uniquely positioned in that they have access to the DNA and clinical data relating to subjects being treated with a specific drug. However, as yet, this has not translated into 'hard' data from the industry on the true value of pharmacogenetics; instead, most studies of the genetic basis of drug response have been driven by academic groups[30–32]. A search for the reasons behind the apparent reluctance of pharmaceutical companies to sponsor such studies reveals numerous and complex issues impeding the conduct of pharmacogenetics within the pharmaceutical industry, in particular, the major societal and ethical considerations of DNA analysis and the regulatory impact of genetic analysis. Many of the ethical and societal challenges inherent in the analysis of human DNA are well known and affect all genetic researchers[33]; the protection of medical privacy is an essential consideration and, as such, is being carefully dealt with by some of the major scientific advisory bodies, such as the National Institutes of Health–Department of Energy's Working Group on Ethical, Legal and Social Implications of Human Genome Research (http://www.nhgri.nih.gov/ELSI/TFGT_final) and the Human Genome Organization Ethics Committee (http://www.gene.ucl.ac.uk/hugo/sampling.html). The issue of the regulatory impact of DNA analysis, however, has not yet been fully explored.

Regulatory impact

It is crucial that regulatory bodies develop the paradigms to assist their interpretation of how genetic analysis of drug response might relate to drug licensing. A popular misconception is that genetic analysis gives deterministic results; in reality, genetic analysis often yields associations between genes and clinical phenotypes, and in many cases the best it can offer is a series of probabilities of a given clinical outcome. The way forward is to encourage the collection of data regarding the genetic basis of drug response and to make this publicly available for scrutiny, further testing and, ultimately, informed interpretation. Until the regulatory authorities know more about the impact of this information, it might be problematic for the industry to be involved in the early information-gathering exercise. The first real data pertaining to the genetics of drug response will probably continue to be gathered by academic groups and biotechnology companies, either independently or in collaboration with industry. An important goal in this work will be to enable both industry workers and the regulators to gather an understanding of the impact of genetics on drug development, such that this initial understanding can then be used as a framework for defining the application of genetics to the future of drug development and therapeutic delivery.

The right drug for the right patient at the right time

In the context of genetic analysis, 'the right drug for the right patient at the right time' refers to the future hope of individualized medicine. Historically, the treatment of a disorder was empirical and the only way of knowing whether or not a drug would work was to try it. By a process of trial and error, the best drug and the best dose was defined for the patient[34]. The future hope and promise of pharmacogenetics is that by understanding the molecular basis of individual variation in drug response, knowledge will be gained on how to focus on the patient as an individual, defining the medicine and dose most suited to a patient before prescription. In reality, it is known that environmental factors will also play a significant role in defining drug response[35] and the full definition of the environmental and molecular basis of the complex phenotype of drug response will be a long and complicated process. The potential advantages are clear, but given that there is currently little knowledge of the genetic factors defining the complex phenotype of drug response, can we be confident that we can realize the dream of using genetics for individualized medicine?

SNP haplotypes and individualized medicine – dream or reality?

To proceed towards individualized medicine in the absence of the definition of the specific genes and gene changes directly contributing to drug response, it has been suggested that a series of anonymous markers could be used to identify a drug response SNP haplotype, or SNP profile, which would have potential for use in clinical trials and medical practice[25]. The crucial issues to be explored here are the feasibility of identifying these haplotypes, the feasibility of integrating haplotype analysis into the process of

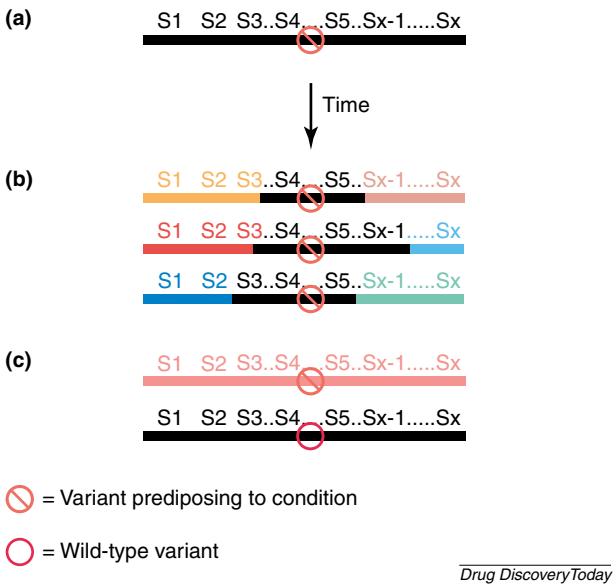

Figure 2. Using haplotypes to detect variant genes. **(a)** A haplotype and a variant gene associated with a disorder. DNA variation within a gene first occurs in a single 'founder' individual within a population. In this case, the variation has occurred upon a single DNA strand with a set of existing SNPs at locations S1 to Sx (black). This specific set of SNPs, associated with the founder variation, is called an ancestral haplotype. **(b)** The population history of the predisposing variant. The ancestral haplotype is passed down through the descendants of the founder individual and the original DNA (black) will be subject to genetic recombination, which will gradually result in replacement of the flanking (black) DNA with new DNA (coloured) containing potentially different alleles of S1 to Sx and new SNPs. The extent to which this occurs is dependent on the number of generations, population expansion, recombination frequencies within the region, and selection pressure. The result is that the ancestral haplotype is increasingly eroded. **(c)** Confounding problems of using the ancestral haplotype as an indicator of the presence of a variant gene. First, depending on the population history of the disorder, the variation contributing to a disorder might become associated with a completely different haplotype (pink), a situation that would also result if there is more than one origin of the disorder in the population. Second, the extent to which the haplotype (in black) is found to be associated with the non-variant or wild-type gene under study, is dependent on the frequency of the black haplotype in the population at the time the initial disorder arose; if this black haplotype was common, many individuals in the current population would possess the black haplotype, but not the disorder.

therapeutic choice and, finally, the proposed advantages of this analysis to existing prescription practices.

The feasibility of defining haplotypes linked to drug response

First, it is important to recognize that drug response is a complex phenotype to which it is probable that a series of genetic and environmental factors will contribute[35]. To date, the relative value of these components and the importance of their interaction is not known. There is likely to be significant truth in the statement that if we are to define the genetics of drug response we need first to define the gene for drug compliance. Suffice to say, there is currently a great deal of debate regarding the feasibility of identifying genetic associations with complex phenotypes, but much effort is being invested to identify tools that will improve the chance of success[36,37].

Second, there are several issues affecting the practicability of integrating haplotype analysis into the process of therapeutic choice. The use of haplotypes comprising markers that are closely linked to, but not directly contributing to, a disorder have been used as a diagnostic tool in the past, but their greatest application is for analysis within a family unit. In this capacity, haplotype sharing with the proband is used to define the risk of inheritance of a disorder in a family, for example, haemochromatosis[38], rheumatoid arthritis[39] and Batten disease[40]. In the case of haemochromatosis, the human leukocyte antigens (HLA) A and B were found to be closely linked to the disorder in 1976 and over the next few years common HLA-A and B haplotypes in patients were identified. Haplotype analysis supported the premise that the gene defect arose on a chromosome carrying the HLA haplotype A3,B7; subsequently, following recombination events between the causative gene and HLA-A and B, several other HLA haplotypes became associated with haemochromatosis in different populations[41]. Consequently, a specific HLA haplotype was never used as a diagnostic, instead, the inheritance pattern of HLA-A and HLA-B alleles within an affected family were used to track the probable inheritance of the disorder[38]. The key message from this is that although a specific SNP haplotype might be associated with certain measures of drug response, it is not the cause of drug response; therefore it is not clear how accurately this same haplotype will define drug response in a series of unrelated individuals from different populations (Fig. 2).

Feasibility of integrating haplotype analysis into the process of therapeutic choice

The lack of certainty with which a haplotype can be taken to mark a 'disorder', in this case drug response, will naturally affect how widely haplotype analysis can be applied to medical practice. In addition, the use of genetics in prescription medicine will require that a patient undergoes genetic testing to define a drug response profile. This will have a substantial impact on the current healthcare delivery systems, specifically, a significant increase in the diagnostic load and cost of healthcare delivery. The challenges

inherent in the delivery of such a system are the subject of lively debate both within and outside the pharmaceutical industry.

Proposed advantages of using genetic analysis in prescription practices

An even more fundamental issue affecting the acceptance of individualized medicine is the issue of proposed benefit. If we propose that existing prescription practices will, on average, be successful in 70% of cases, then given the increased load on the healthcare system introduced by wide-scale genetic testing we must expect, at the very least, that genetics will offer a greater level of success in therapeutic delivery, and that any increase is clearly cost effective. Although, at present, we cannot estimate the realistic impact of pharmacogenetics on prescription practice, the acceptance of genetic testing into the delivery of healthcare is crucial for the future of pharmacogenetics.

Conclusions

The analysis of genetic variation is clearly gaining acceptance in today's pharmaceutical industry, but if genetic analysis is to have a permanent future in both industry and medical practice, there needs to be a rationalization of the current aspirations and a focus on realistic delivery. In the first instance, sufficient investment in early discovery is required to support the identification and functional assessment of variants within targets; the payback for this investment will be in the early definition of the targets most likely to succeed. Second, routine collection of DNA in clinical trials should be considered mandatory, to facilitate the testing of the hypothesis that there are real and identifiable genetic factors affecting drug response. Related to this last point, a clear understanding must be promoted within the ethical and regulatory bodies, that genetic information must be generated during the development process and, further, during this early information-gathering stage, such analysis should be interpreted as research, the outcomes of which cannot and should not impact on medical practice or result in prohibitive demands on the labelling of drugs.

The pharmaceutical industry needs to be cognisant of the rapid and significant developments in genetic analysis outside its own walls; specifically, the biotechnology companies and academic research groups eager to understand how a compound works and why a cohort of non-responders can exist. A company not undertaking that research itself at an early stage in compound development, in the hope of protecting a compound's future, will soon find that other groups have nothing to lose and everything to gain by trying to identify the genetic factors governing the safety and efficacy of medicines. In summary, taken together, the current developments in academic, biotechnology and pharmaceutical genetic research present the most radical hope for significant change in patterns of medical treatment.

Acknowledgement

I am grateful to Peter Little (University of New South Wales, Sydney, Australia) for critical review and discussion of this manuscript.

References

1 Halushka, M.K. *et al.* (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* 22, 239–247
2 Green, E.D. and Waterston, R.H. (1991) The human genome project. Prospects and implications for clinical medicine. *J. Am. Med. Assoc.* 266, 1966–1975
3 Hoffman, E.P. (1994) The evolving genome project: current and future impact. *Am. J. Hum. Genet.* 54, 129–136
4 Pennisi, E. (2000) Human genome. Finally, the book of life and instructions for navigating it. *Science* 288, 2304–2307
5 Butcher, J. (2000) 'Working draft' of human genome completed. *Lancet* 356, 47
6 Schultz, J. *et al.* (2000) More than 1,000 putative new human signalling proteins revealed by EST data mining. *Nat. Genet.* 25, 201–204
7 Singh, G.B. (2000) Computational approaches for gene identification. *Methods Mol. Biol.* 132, 351–364
8 Bailey, L.C., Jr *et al.* (1998) Analysis of EST-driven gene annotation in human genomic sequence. *Genome Res.* 8, 362–376
9 Brookes, A.J. (1999) The essence of SNPs. *Gene* 234, 177–186
10 McCarthy, J.J. and Hilfiker, R. (2000) The use of single-nucleotide polymorphism maps in pharmacogenomics. *Nat. Biotechnol.* 18, 505–508
11 Bentley, D.R. (2000) The Human Genome Project – an overview. *Med. Res. Rev.* 20, 189–196
12 Masood, E. (1999) As consortium plans free SNP map of human genome. *Nature* 398, 545–546
13 Terwilliger, J.D. and Weiss, K.M. (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.* 9, 578–594
14 Weiss, K.M and Terwilliger, J.D. (2000) How many diseases does it take to map a gene with SNPs? *Nat. Genet.* 26, 151–156
15 Underhill, P.A. *et al.* (2000) Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26, 358–361
16 Chakravarti, A. (1999) Population genetics – making sense out of sequence. *Nat. Genet.* 21 (Suppl.), 56–60
17 Mueller, R.L. and Scheidt, S. (1994) History of drugs for thrombotic disease. Discovery, development, and directions for the future. *Circulation* 89, 432–449
18 Dollery, C.T. (1999) Drug discovery and development in the molecular era. *Br. J. Clin. Pharmacol.* 47, 5–6
19 Kuhlmann, J. (1999) Alternative strategies in drug development: clinical pharmacological aspects. *Int. J. Clin. Pharmacol. Ther.* 37, 575–583
20 Weber, W.W. (1999) Populations and genetic polymorphisms. *Mol. Diagn.* 4, 299–307
21 Gelernter, J. *et al.* (1998) Population studies of polymorphisms at loci of neuropsychiatric interest (tryptophan hydroxylase (TPH), dopamine transporter protein (SLC6A3), D3 dopamine receptor (DRD3), apolipoprotein E (APOE), μ opioid receptor (OPRM1), and ciliary neurotrophic factor (CNTF)). *Genomics* 52, 289–297
22 Drazen, J.M. *et al.* (1999) Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat. Genet.* 22, 168–170

23 Wolf, C.R. and Smith, G (1999) Pharmacogenetics. *Br. Med. Bull.* 55, 366–386
24 Ledley, F.D. (1999) Can pharmacogenomics make a difference in drug development? *Nat. Biotechnol.* 17, 731
25 Rioux, P.P. (2000) Clinical trials in pharmacogenetics and pharmacogenomics: methods and applications. *Am. J. Health Syst. Pharm.* 57, 887–898
26 Roses, A.D. (2000) Pharmacogenetics and the future of drug development and delivery. *Lancet* 355, 1359–1361
27 Roses, A.D. (2000) Pharmacogenetics and the practice of medicine. *Nature* 405, 857–865
28 Stephens, J.C. (1999) Single-nucleotide polymorphisms, haplotypes, and their relevance to pharmacogenetics. *Mol. Diagn.* 4, 309–317
29 Pfost, D. *et al.* (2000) A SNPshot: pharmacogenetics and the future of drug therapy. *Trends Biotechnol.* 18, 334–338
30 Evans, W.E. and Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491
31 Freemantle, S.J. *et al.* (1995) Molecular characterization of two cell lines selected for resistance to the folate-based thymidylate synthase inhibitor, ZD1694. *Br. J. Cancer* 71, 925–930
32 Krynetski, E.Y. and Evans, W.E. (1999) Pharmacogenetics as a molecular basis for individualized drug therapy: the thiopurine *S*-methyltransferase paradigm. *Pharm. Res.* 16, 342–349
33 McQueen, M.J. (1998) Ethical and legal issues in the procurement, storage and use of DNA. *Clin. Chem. Lab. Med.* 36, 545–549
34 Patsalos, P.N. (2000) Antiepileptic drug pharmacogenetics. *Ther. Drug. Monit.* 22, 127–130
35 Poolsup, N. *et al.* (2000) Pharmacogenetics and psychopharmacotherapy. *J. Clin. Pharm. Ther.* 25, 197–220
36 Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22, 139–144
37 Schork, N.J. (1997) Genetics of complex disease: approaches, problems, and solutions. *Am. J. Respir. Crit. Care Med.* 156, S103–S109
38 Conte, W.J. and Rotter, J.I. (1984) The use of association data to identify family members at high risk for marker-linked diseases. *Am. J. Hum. Genet.* 36, 152–166
39 Rigby, A.S. *et al.* (1998) HLA haplotype sharing in rheumatoid arthritis sibships: risk estimates subdivided by proband genotype. *Genet. Epidemiol.* 15, 403–418
40 Taschner, P.E. *et al.* (1995) Carrier detection of Batten disease (juvenile neuronal ceroid-lipofuscinosis). *Am. J. Med. Genet.* 57, 333–337
41 Jazwinska, E.C. (1998) Hemochromatosis: a genetic defect in iron metabolism. *Bioessays* 20, 562–568

In the 1st March issue of *Drug Discovery Today*…

Editorials High throughput purification: walk-up versus dedicated support
by Mark Jones

Update latest news and views including:

- Discussions on academia–industry collaborations and on training of chemists
- 'Private prescription' will discuss how to keep your audience alert during presentations
- Reports on IIR's *Embracing New Tools for Parallel Automated Chemistry* conference and on the *Clinical Trials – The Next Phase!* conference organized by Access Conferences International
- Up-to-date News, News in brief and People

Reviews **Targeting protein ubiquitination for drug discovery. What is in the drug discovery toolbox?**
by David Swinney

The design of combinatorial libraries using properties and 3D pharmacophore fingerprints
by Brett Beno and Jonathan Mason

Direct gene delivery strategies for the treatment of rheumatoid arthritis
by Steven Ghivizzani, Thomas Oligino, Joseph Glorioso, Paul Robbins and Christopher Evans

Monitor new bioactive molecules and combinatorial chemistry