Making drug discovery a SN(i)P

David A. Campbell, AnaMaria Valdes and Nigel Spurr

The pharmaceutical world is coming to terms with the fact that genetic variation is the key to identifying the biological basis behind both susceptibility to disease and the response to drugs. Genetic variation within drug targets is common and we can safely predict that genetic factors influence almost every human disease. Thus, the study of single-nucleotide polymorphisms (SNPs) is crucial for characterizing molecular targets and can also validate the role of these targets in disease. This article reviews the progress to-date and outlines some of the key areas in which SNPs will impact on the drug discovery process.

Genetics is the study of inherited variation. Since 1980, DNA sequence variation has become the hard currency of modern genetics. The first DNA-based variants were identified by differences in the length of DNA fragments generated following restriction enzyme digestion, termed restriction fragment length polymorphisms (RFLPs)[1]. This was followed rapidly by the discovery of variable number tandem repeats (VNTRs), mini-satellites that form the backbone of our current forensic approach to DNA fingerprinting[2]. RFLPs and VNTRs formed the core of the first human genetic-linkage maps[3]. In the late-1980s, a new class of variable genome markers was identified, based on dinucleotide repeats[4]. These appear to be distributed randomly within the human genome and their length varies between individuals. Dinucleotide repeats are usually multi-allelic and form the basis of the recent generation of whole human genome genetic linkage maps published in the 1990s (Ref. 5). The majority of these markers occur outside the coding regions of genes. There are several notable exceptions, however, for example an RFLP within the β-globin gene can identify individuals susceptible to sickle cell anaemia and has been used as a diagnostic test for many years[6].

In the past few years, a new class of DNA marker has come to prominence, termed single-nucleotide polymorphism (SNP). A SNP occurs when a different nucleotide occurs at a given position in different individuals. SNPs are the most abundant DNA marker present in the human genome. The basic premise that there is one SNP in every 1000 bases[7–14] leads to an estimate that any two individuals differ by up to three million SNPs. However, only a few hundred thousand are currently identified. Many of these are in the coding regions of genes and have been termed cSNPs. cSNPs can have several effects, including:

- Neutral alteration, a base substitution with no effect on the amino acid residue encoded.
- Conservative change, a base substitution that results in an altered amino acid, but has minimal effect on protein structure or function.
- Nonconservative alteration leading to a change in the encoded amino acid residue that has dramatic effects on protein structure or function. Included in this group are base changes leading to, for example, frame-shifts that result in premature termination or other nonsense events.

Although nonconservative cSNPs are the most likely to cause marked effects on protein function, other SNPs (such as those within noncoding regions of a gene) can influence, for example, mRNA stability and rate of transcription. This will therefore have an overall effect on protein levels. These markers are the tools of the future and our ability to detect associations between SNPs and hereditary phenotypes will drive a new revolution in medicine over the next ten years.

SNPs in drug discovery

Genetic factors influence virtually every human disease by determining susceptibility or resistance to the disorder and

David A. Campbell, AnaMaria Valdes and ***Nigel Spurr**, Genetic Technologies, SmithKline Beecham Pharmaceuticals, New Frontiers Science Park, Harlow, Essex, UK CM19 5AW. *tel: +44 1279 622000, fax: +44 1279 622500, e-mail: Nigel_K_Spurr@sbphrd.com

 1359-6446/00/$ – see front matter PII: S1359-6446(00)01546-4

Table 1. Putative susceptibility loci for complex polygeneic disorders

Cardiovascular disease	Obesity	Schizophrenia	Osteoporosis	Asthma
ACE	Pro-opiomelanocortin	1q42-1q41	Cathepsin K	TNF
TGFβ	MC_4	13q14-q32	ER_α	FCεR1
ApoB		22q12-q13	β_3-adrenoceptor	IL-4
			Vitronectin receptor a and b subunits	

Abbreviations: ACE, angiotensin 1 converting enzyme; ER_α, oestrogen receptor α; FCεR1, high-affinity IgE receptor; IL-4, interleukin 4; MC_4, melanocortin receptor 4; TGFβ, transforming growth factor β; TNF, tumour necrosis factor; ApoB, apolipoprotein B.

by interacting with environmental factors. In the past decade, it has been possible to identify the genetic location, clone and determine the molecular basis of monogenic disorders including Huntington's disease[15] and cystic fibrosis[16].

However, many of the diseases targeted by the pharmaceutical industry, including cardiovascular disease, obesity, schizophrenia, osteoporosis and asthma, involve multiple genes and are subject to environmental effects (Table 1). Genetic susceptibility to complex traits is not caused by specific, relatively rare mutations, rather, the inheritance pattern of these common disorders results from genetic variations that are relatively common in the general population, with many genes each contributing a small quantity of the overall disease susceptibility.

Biological variation is important in the development of new pharmaceuticals, with an impact on the following areas:

- Target identification
- Target characterization
- Target validation
- Pharmacogenetics.

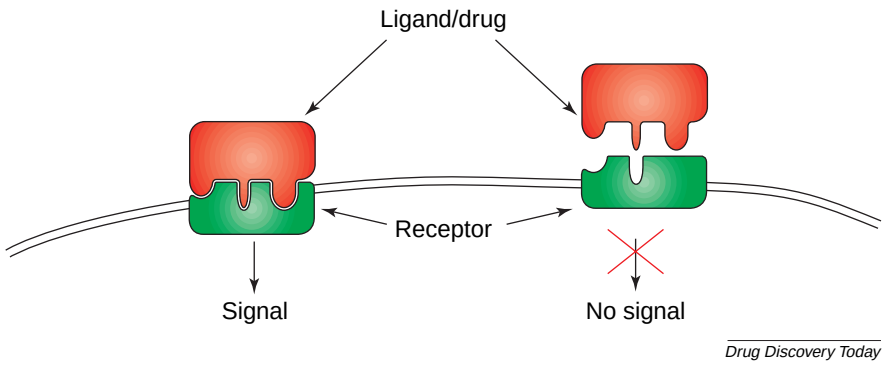

Figure 1. *Major changes in protein structure can result from subtle variation in the genetic sequence encoding it. These changes can completely block the function of the protein* in vivo *and can also affect the development of high-throughput assays for screening therapeutic compounds.*

Target identification

Target identification is the identification of regions of the genome linked to clinical end-points. This approach, also termed positional cloning, looks for a disease-susceptibility gene near markers that have an inheritance pattern most similar to that of the disease. In this case, SNPs are used as simple genetic markers in the same way that microsatellites are currently used in pedigree (family-based) studies[17,18]. The polymorphisms used in this type of study are unlikely to be involved directly with the clinical trait. There are excellent reviews on the genetic mapping of complex diseases in general[19] and using SNPs in particular[20].

One of the most common forms of target identification involves linkage studies[21]. These involve cohorts of individuals, either affected sibling pairs in which both siblings have the disease of interest, or more extended, multiplex families in which there are many affected individuals. Data from such studies can be used to demonstrate the Mendelian inheritance of a given phenotype and to perform full-scale genome mapping to identify regions of the genome containing disease-associated genes. Having performed these analyses, classical positional-cloning techniques can identify the disease-associated mutations.

A recent study by Detera-Wadleigh *et al.*[22] has identified a susceptibility locus for bipolar disorder on human chromosome 13q32, while a study by Pericak-Vance *et al.*[17] has led to the identification of apolipoprotein E4 as a major risk factor in Alzheimer's disease[23].

Target characterization

Target characterization is the identification of variants within genes of interest. There are several reasons why it is vital to establish the degree of genetic variation within a

Table 2. Examples of pharmacologically relevant polymorphisms

Gene	Mutation/variant	Effect	Refs
Dopamine D5 receptor *DRD5*	Asn351Asp in the 7th transmembrane domain	Approximately tenfold decrease in dopamine and threefold decrease in *R*(+)-SKF38393 binding affinities.	24
Dopamine D2 receptor *DRD2*	Ser311Cys, Pro310Ser and Val96Ala	Several typical and atypical neuroleptics commonly used in the treatment of psychotic disorders have differences in binding affinities and potencies for the dopamine D2 receptor.	25
Apolipoprotein E *ApoE*	Apo E2, E3 and E4 alleles	Response to the cholinesterase inhibitor tacrine, measured by either Alzheimer's disease (AD) assessment scores, or by the quantity of cortical electrical arousal, is significant only among AD patients without the Apo E4 allele. The Apo E4 allele is a prognostic indicator of poor response to therapy with acetylcholinesterase inhibitors in AD patients.	26–29
Cholesteryl ester transfer protein *CETP*	B1 and B2 alleles	The lipid-lowering agent pravastatin slowed the progression of coronary atherosclerosis in individuals with the B1B1, but not B2B2, haplotype (representing 16% of the patients taking pravastatin).	30
Herceptin *HER2*	Normal expression/overexpression variants	*HER2*/neu overexpression results in a more aggressive, less responsive breast cancer. *HER2* overexpression is linked to sensitivity and/or resistance to hormone therapy and chemotherapeutic regimens, including CMF (cyclophosphamide, methotrexate and fluorouracil) and anthracyclines.	31
5-Lipoxygenase *ALOX5*	Gene-promoter sequence variants modifying gene transcription	Carriers of the promoter mutation are all non-responders to asthma treatments that specifically interfere with the 5-lipoxygenase.	32

drug target. Figure 1 shows an example of how a drug interaction can be influenced by variations in a target protein and there are numerous examples in the literature of single-base variants that alter gene function (Table 2). Modern methods of drug discovery rely on the use of HTS of vast chemical libraries produced using combinatorial chemistry, and it is important that the target of this screen is representative of the majority of the target population. This is vital when SNPs are known to affect the amino acid structure and function of a protein. For example, studies demonstrate that an asparagine to aspartic acid mutation in the human dopamine D5 receptor increases the affinity of the receptor for dopamine but decreases the affinity for the agonist *R*(+)-SKF38393 (Ref. 24). Similarly, studies of a naturally occurring variant of the human serotonin 5-HT_{1B} receptor cystine at position 124 indicates a difference in the affinity of this variant for a number of ligands when compared with the affinity profile of the protein containing phenylalanine at position 124 (Ref. 33).

Changes within the coding region of a gene are not the only variants known to alter gene function. Polymorphisms in promoter regions, such as the cytosine to adenine substitution at position −863 relative to the translation initiation start site of the tumour necrosis factor (TNF) gene, can alter the level of expression of the gene and, thus, the quantity of protein produced[34].

Although every effort is now made to screen drug targets at the nucleic acid level to determine their degree of genetic variance, identification of these variants is often not enough. State-of-the-art bioinformatic support is necessary to determine whether a genetic variant alters the protein that the gene encodes. For example, it is possible to use structural data to predict the effect of an amino acid substitution, based on amino acid preference tables[35]. Classical biochemistry and molecular biology also have roles here.

Target validation

Target validation is the ability to assign a drug or class of drugs to a particular clinical indication and to extend the therapeutic indications of existing drugs. The extensive use of genomic information by pharmaceutical companies and the biotechnology industry mean that there are now a plethora of novel drug targets, often with little indication of their role in disease. Although the ability to identify chemical compounds suitable for use by patients is paramount to the drug discovery process, the ability to identify the biological rational for the use of these compounds remains vital. The use of classical genetics has successfully been used to support this biological effort. For example, calcitonin has been used in the treatment of osteoporosis for many years. Recent data from Taboulet *et al.*[36] suggest that a mutation resulting in the substitution of proline to leucine at residue 447 of this protein could be associated with an altered risk of osteoporotic fractures in postmenopausal women. Studies of patients with early-onset, type 2 diabetes (maturity-onset diabetes of the young) suggest a possible role for polymorphisms in the glucokinase gene[37], whilst deletion of nucleotide 32 in the β-chemokine receptor 5 (*CCR5*) gene results in a protective effect on exposure to HIV (Ref. 38).

Knowing that genetic variation plays a crucial role in disease, genetic epidemiology can be used to show functional involvement of a particular drug target in the disease of interest. This approach can also be used to identify new therapeutic targets for existing drugs.

Pharmacogenetics

Pharmacogenetics is the study of how genetic variation affects drug response and metabolism. A list of examples of pharmacologically relevant polymorphisms is shown in Table 2 (for review, see Ref. 39).

Polymorphisms in enzymes that metabolize drugs are responsible for unexpected toxicity after administration of a normal drug dose[40]. SNPs in the coding regions of genes of the enzymes necessary for the biotransformation of endogenous and exogenous compounds can decrease or even prevent drug metabolism. For instance, mutation of human cytochrome P450 genes that encode drug metabolizing enzymes, including *CYP2C19*, *CYP1A1*, *CYP2A6*, *CYP2C9* and *CYP2E1* (Ref. 41), increases the side effects of a wide variety of drugs, including antidepressants, amphetamines and many β-adrenoceptor antagonists[42,43].

Another case where common polymorphisms have important pharmacological consequences are the methylation enzymes[44]. For example, polymorphism within the gene encoding thiopurine methyltransferase (TPMT) is a major determinant of individual differences in toxicity or therapeutic efficacy of thiopurine drugs[45]. To-date, at least ten separate mutations associated with very low TMPT activity have been described[46]. Because the drugs metabolized by this enzyme have relatively narrow therapeutic indices, and because they are used to treat life-threatening situations, individual variation in either drug toxicity or therapeutic efficacy can have profound consequences for the patients. For this reason, in many medical centres the TPMT genotype is determined prior to the initiation of therapy with these drugs[47].

How common are SNPs?

Estimates differ widely as to how common these single base changes are within the human genome. Three recent studies[48–50] that have screened several hundred genes potentially involved in cardiovascular, neuropsychiatric and endocrine disorders found, on average, between 2.2 and 5.3 cSNPs per gene. This includes all polymorphisms with a frequency of at least 5% in normal individuals and, therefore, could be relevant for HTS considerations as well as target validation. However, there is evidence to suggest that not every gene contains cSNPs.

The technology requirements for genotyping

The following section describes the many uses of SNPs in the drug discovery process.

DNA samples

Without the appropriate populations, it is impossible to use SNPs as biological tools. The pharmaceutical industry, like the academic and biotechnology communities, is making enormous efforts to obtain DNA samples from patient cohorts appropriate for the analysis already described. The issues involved in obtaining appropriate cohorts for genetic analysis have been reviewed previously[51,52]. Suffice it to say that most pharmaceutical companies are of the opinion that material should be collected wherever possible and that the usefulness of this material should be a matter of debate later. For example, many companies use pharmacogenetic-based studies of the response of patients to new drugs to understand adverse effects or the reasons for non-responders. This has led to the collection of DNA samples from clinical trials, often as a precaution should problems be identified during or after a trial.

More fundamental questions concerning the mechanisms underlying multigenic diseases are also being addressed. These studies usually require blood samples from a population of patients for either population-association studies or quantitative trait locus (QTL) analysis. The former requires the parallel collection of a control group of unaffected individuals, matched for race, age and gender.

QTL studies do not necessarily need a control group, but must have a well-defined phenotype that can be studied across the population. Examples of this are body mass index in the study of obesity and diabetes, and bone mineral density to understand the development of osteoporosis.

All population-based studies need careful selection criteria if the study is to succeed. These criteria include a well-defined phenotype (usually a component of the disease under study), a homogenous population with similar background and origin, and matching for gender and age.

Two major classes of markers are genotyped routinely in population-association studies, dinucleotide repeats and SNPs. Genotyping can be carried out using a range of methods, many of which are commercially available, the commonest methods will now be summarized. The goal of all new genotyping developments is to minimize cost and, where appropriate, to maximize throughput. These goals are necessary to analyze the components of complex traits, such as response to drug treatment or predisposition to diabetes and schizophrenia, because of the large number of markers involved in this type of study.

De novo *DNA sequencing*

Throughout the 1990s, DNA sequencing has been dominated by the use of gel-based fluorescent methods that incorporate automated laser detection of a range of fluorochromes. Recently, several companies have produced sequencing instruments based on the same chemistries, but using capillaries to replace the traditional slab gels. This fluorescent-based sequencing technology has been used to analyze a wide range of viral, bacterial and, more recently, eukaryotic genomes and remains the gold standard by which genetic variation is assessed.

SNPs can be identified and verified directly using DNA sequencing. Approaches such as that taken by Taillon-Miller *et al.*[53] enabled the use of direct sequencing not only to detect novel SNPs but also to make estimations of allele frequency using a pooled DNA-sequencing approach.

RFLP analysis

In the 1980s, the first genetic mutations were detected using alterations in restriction enzyme recognition sites to detect RFLPs (Ref. 1). This methodology is of limited application as it is restricted by the frequency of suitable enzyme recognition sites. However, such assays remain a popular method for detecting variants when combined with PCR amplification of specific regions of DNA (amplified restriction fragment length polymorphism or AFLP) rather than the restriction enzyme digestion of genomic DNA. Using AFLP analysis, both alleles can be identified in one reaction. Although considered to be a reasonably satisfactory method, the overall costs are high (*Taq* DNA polymerase and a suitable restriction enzyme are required) and it is both labour intensive and time consuming to set up and analyze.

Allele-specific oligonucleotides

Although many different technologies are being developed for automated SNP assays, one of the easiest to use is the amplification refractory mutation system (ARMs), developed by Newton *et al.* in 1988 (Ref. 54). ARMs, also known as allele-specific amplification (ASA), relies on the ability of *Taq* polymerase to extend a primer bound to DNA template, where the 3′ base of the primer matches its complementary base on the template. This generates +/− scores for each allele, so that two PCR reactions are required for each DNA sample. In these cases, the PCR products are analyzed on acrylamide or agarose gels. Some fluorescent technologies, such as TaqMan[55], fluorescence polarization[56] and free phosphate accumulation[57], eliminate the need for gel analysis. Other fluorescence-based genotyping techniques that do not require gel analysis often require extra steps, for example, fluorescence resonance energy transfer[58] and dynamic allele-specific hybridization[59]. An additional genotyping technique that is PCR-independent and shows great promise is the homogeneous isothermal technique, which involves using a cleavase enzyme[60]. The cleaved fragment can be visualized on gels or by fluorescence. Currently, two assays are required for each genotype.

Mass spectroscopy

A major area of emerging technology is the use of the mass spectrometer in the analysis of DNA. This has the advantage of eliminating the need for the gel-based analysis of DNA fragments in both DNA sequencing and genotyping. The main type of MS used for genotyping is matrix-assisted laser desorption ionisation time-of-flight spectrometry (MALDI-TOF)[61].

There are several problems in the application of MS for genome analysis, particularly in DNA sequencing, as only short fragments (less than 100 bases) can be measured accurately. However, an alternative form of MS using IR-MALDI has been used recently to analyze fragments of DNA more than 2000 bases in length[62]. This new development opens up the possibility of analyzing longer sequence fragments and supports the potential utility of this tool for the detection of multiple variants in a single analysis.

DNA chips and microarrays

Since 1996, the two competing technologies of DNA chips and microarrays have become the mainstay of high-resolution genetic analysis. The extraordinary rise of this

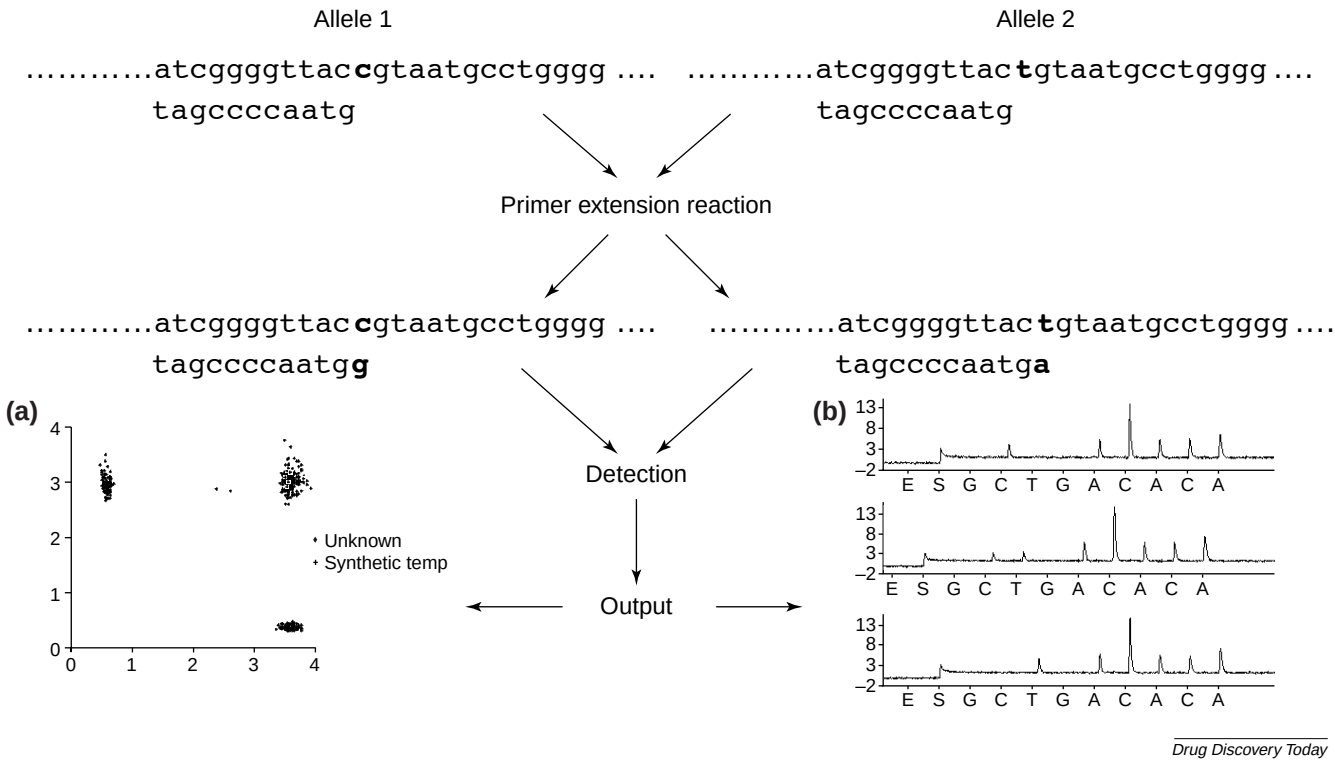

Figure 2. *Advances in SNP genotyping technology have meant a considerable increase in throughput and sensitivity and a subsequent decrease in cost. The method of choice for genotyping remains with primer extension. Although there are a number of methods available for performing this assay, two systems are worthy of mention. The ultra-HTS offered by Orchid Bioscience (Princeton, NJ, USA) and the lower-throughput system offered by Pyrosequencing AB (Uppsala, Sweden). The output of each system is shown as (a) Orchid's Scatterplots and (b) Pyrosequencing's Pyrogram. The Orchid system clearly shows the three clusters representing the two homozygous groups (top left and bottom right) and the heterozygous group (top right). The pyrograms represent the three possible outputs for a CC, CT or TT genotype.*

technology and its importance in the new area of genetic analysis was recently illustrated in the publication of a supplement to *Nature Genetics*[63], which describes the latest technology developments and their many applications.

Although the two technologies are based on different methods of preparing DNA arrays, they both rely on the hybridization of complementary strands of DNA, one of which is labelled with a fluorescent tag to enable the automated detection of hybridizing pairs of sequences. The results are interpreted using a computer database. Indeed, one of the key features of all such DNA array-based technologies is that high-quality informatics is essential for the correct interpretation of data.

Genomic analyses using DNA chips have focused on the identification of mutations, the analysis of SNPs and the detection of changes in gene expression. In the first two applications, genomic DNA is fluorescently labelled and hybridized to the arrayed oligonucleotides, whereas for gene expression analysis, mRNA is isolated from a tissue of interest and single-strand cDNA is synthesized incorporating a fluorescent label. This enables the comparison of different cDNA samples in the same experiment (for example a normal tissue versus a tumour) using dual colour laser scanning and two fluorescent labels with different excitation wavelengths.

Wang and colleagues describe the identification, mapping and genotyping of 3241 new SNPs involving the use of oligonucletotide chips covering 2.3 megabases of genomic DNA (Ref. 10). Oligonucleotide arrays have been used to examine gene expression in a range of genes from both humans and from other species[55].

As genotyping tests become available it will be possible to design prospective clinical studies of adverse drug reactions and drug toxicity and to adjust drug doses to suit the metabolic capacity of each patient.

Primer extension

Primer extension is rapidly becoming the method of choice for high-throughput genotyping using SNPs, and

there are several technologies available. These range from the relatively low-throughput system Luc 96 (Pyrosequencing AB, Uppsala, Sweden) to the ultra-high-throughput SNPstream system (Orchid Bioscience, Princeton, NJ, USA). Both technologies use a slightly different approach to the detection of extension products and, thus, genotyping, but both rely on the basic principle of primer extension (Fig. 2, see Refs 64 and 65 for a review of these systems).

There is no doubt that there is a market for both systems. For target characterization, where a limited number of genotypes are required, low-throughput systems offer the best, most cost-effective approach. For target identification and validation studies, where significantly large sample sizes are required and a number of markers are analyzed within each experiment, more automated processes are more useful.

Marker identification

By definition, without markers there is no study. Although public efforts such as The SNP Consortium (http://snp.cshl.org/) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) offer access to large numbers of markers, there is still a need for in-house, project-driven, SNP discovery.

Several methods are used to detect single nucleotide variations in genomic DNA. Some of these, such as direct sequencing of DNA (Refs 10,14), denaturing HPLC (Ref. 66), single-strand conformation polymorphism analysis[67] or high-density DNA arrays[68], are well-suited to detecting variants but are only applicable for limited numbers of samples or pools of samples. Other methods, such as that offered by Third Wave Technologies' Invader® assay (Madison, WI, USA)[69], exist for the parallel detection of variants in large numbers of independent samples.

Will drug discovery become a SN(i)P?

The recent data suggests that coding SNPs are common in drug targets. Indeed, the average target will have 2–5 nucleotide changes in the coding region[48–50]. This is a reality to be harnessed in drug discovery. The pharmaceutical industry is determined that this new genomics era should benefit us all. Indeed, the year 2000 should see the first genomics-derived drug going into clinical trials. The pressure is on to utilize this once-in-a-lifetime chance to obtain the blueprint of life and to make it work for the benefit of the overall health of the Nation. As the costs of this work decrease we will be able to maximize the use of the resources we already have. We should be able to design better, although not necessarily more complex, experiments to both identify and validate new drug targets. In the next few years, we will finally see the dawn of the pharmacogenetics era and the potential of genetic testing being offered hand-in-hand with classical biochemistry towards the ultimate aim of 'designer drugs'.

Thankfully, the SNP Consortium will reduce the burden of discovering SNPs. SNPs identified by the Consortium will be available royalty-free and, as such, both academic and commercial enterprises will benefit from the intellectual property generated by using these markers. The Consortium is on target to identify 300 000 markers over a two-year period. We will be able to utilize these markers, along with SNPs discovered in-house, to fulfil the promise of pharmacogenetics. This should improve clinical trials by enabling better selection of patient groups, to better characterize adverse events by understanding the genetics of drug metabolism, and to better target new or existing drugs according to individual genetic profiles.

This will be accompanied by a reduction in the cost and, hopefully, increased efficacy of drugs. The cost–benefit ratio of designing new drugs is of major concern to the Nation as a whole. The UK Government has recently set up the National Institute of Clinical Excellence (http://www.nice.org.uk/nice/nic_ind.htm) to make recommendations on new technologies and treatments for specific conditions to the National Health Service (NHS). This group has already ruled on the use of taxanes in the treatment of breast and ovarian cancer. It has also recently recommended that the NHS should not offer zanamivir (Relenza) on general prescription during the 1999/2000 influenza season because of lack of data on its effectiveness in reducing secondary complications. It is clear that the British Government is keen to get value for money for the NHS.

As we enter the new millennium, the pharmaceutical industry is undergoing one of the largest shake-ups since its conception. The industry has to balance the increased cost of discovering new drugs with the market pressure to reduce the costs of prescribing drugs. As we move into a new decade, we put our faith in the genomics era to facilitate this. SNPs will play a major role!

REFERENCES

1 Botstein, D. *et al.* (1980) Construction of a genetic map in man using RFLP's. *Am. J. Hum. Genet.* 32, 314–331

2 Jeffreys, A.J. *et al.* (1985) Individual-specific 'fingerprints' of human DNA. *Nature* 316, 76–79

3 O'Connell, P. *et al.* (1988) Twelve loci form continuous linkage map for human chromosome 18. *Genomics* 3, 367–372

4 Litt, M. and Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* 44, 397–401

5 Reed, P.W. *et al.* (1994) Chromosome-specific microsatellite sets for fluorescent-based, semi-automated genome mapping. *Nat. Genet.* 7, 390–395

6 Kan, Y.W. and Dozy, A.M. (1978) Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. U. S. A.* 75, 5631–5635

7 Collins, F.S. *et al.* (1997) Variations on a theme: cataloguing the human DNA sequence variation. *Science* 278, 1580–1581

8 Collins, F.S. *et al.* (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8, 1229–1231

9 Li, W. and Sadler, L.A. (1991) Kow nucleotide diversity in man. *Genetics* 129, 513–523

10 Wang, D.G. *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082

11 Lai, E. *et al.* (1998) A 4 Mb high-density single nucleotide polymorphism-based map around human APOE4. *Genomics* 54, 31–38

12 Nickerson, D.A. *et al.* (1998) DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. *Nat. Genet.* 19, 233–240

13 Harding, R.M. *et al.* (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* 60, 772–789

14 Taillon-Miller, P. *et al.* (1998) Overlapping genomic sequences: A treasure trove of single-nucleotide polymorphisms. *Genome Res.* 8, 748–754

15 Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983

16 Riordan, J.R. *et al.* (1989) Identification of the cystic fibrosis gene: cloning and characterisation of complementary DNA. *Science* 245, 1066–1073

17 Pericak-Vance, M.A. *et al.* (1988) Genetic linkage studies in Alzheimer's disease families. *Exp. Neurol.* 102, 271–279

18 Davies, J.L. *et al.* (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371, 130–136

19 Whittemore, A.S. (1996) Genome scanning for linkage: an overview. *Am. J. Hum. Genet.* 59, 704–716

20 Zhao, L.P. *et al.* (1998) Mapping of complex traits by single-nucleotide polymorphisms. *Am. J. Hum. Genet.* 63, 225–240

21 Todd, J.A. and Farral, M. (1996) Panning for gold: genome-wide scanning for linkage in type 1 diabetes. *Hum. Mol. Genet.* 5, 1443–1448

22 Detera-Wadleigh, S.D. *et al.* (1999) A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5604–5609

23 Corder, E.H. *et al.* (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923

24 Cravchik, A. and Gejman, P.V. (1999) Functional analysis of the human D5 dopamine receptor missense and nonsense variants: differences in dopamine binding affinities. *Pharmacogenetics* 9, 199–206

25 Cravchik, A. *et al.* (1999) Analysis of neuroleptic binding affinities and potencies for the different human D2 dopamine receptor missense variants. *Pharmacogenetics* 9, 17–23

26 Farlow, M.R. *et al.* (1998) Treatment outcome of tacrine therapy depends on apolipoprotein genotype and gender of the subjects with Alzheimer's disease. *Neurology* 50, 669–677

27 Farlow, M.R. *et al.* (1996) Apolipoprotein E genotype and gender influence response to tacrine therapy. *Ann. New York Acad. Sci.* 802, 101–110

28 Riekkinen, P. Jr *et al.* (1997) The ability of THA treatment to increase cortical alpha waves is related to apolipoprotein E genotype of Alzheimer disease patients. *Psychopharmacology* 129, 285–288

29 Poirier, J. *et al.* (1995) Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 92, 12260–12264

30 Mitchell, M.S. and Press, M.F. (1999) The role of immunohistochemistry and fluorescence *in situ* hybridization for HER2/neu in assessing the prognosis of breast cancer. *Semin. Oncol.* 4 (Suppl. 12), 108–116

31 Baselga, J. *et al.* (1997) HER2 overexpression and paclitaxel sensitivity in breast cancer: therapeutic implications. *Oncology (Huntingt.)* 3 (Suppl. 2), 43–48

32 Drazen, J.M. *et al.* (1999) Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat. Genet.* 22, 168–170

33 Bruss, M. *et al.* (1999) Modified ligand binding to the naturally occurring Cys-124 variant of the human serotonin 5-HT_{1B} receptor. *Pharmacogenetics* 9, 95–102

34 Skoog, T. *et al.* (1999) A common functional polymorphism (C-A substitution at position –863) in the promoter region of the tumour necrosis-α (TNF-α) gene associated with reduced circulating levels of TNF-α. *Hum. Mol. Genet.* 8, 1443–1449

35 Livingstone, C.D. and Barton, G.J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comp. Appli. Biosci.* 9, 745–756

36 Taboulet, J. *et al.* (1998) Calcitonin receptor polymorphism is associated with a decreased fracture risk in post-menopausal women. *Hum. Mol. Genet.* 7, 2129–2133

37 Ng, M.C. *et al.* (1999) Molecular genetics of diabetes mellitus in Chinese subjects; identification of mutations in glucokinase and hepatocyte nuclear factor-1 alpha genes in patients with early-onset type 2 diabetes mellitus/MODY. *Diabetic Med.* 16, 956–963

38 Huang, Y. *et al.* (1997) The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat. Med.* 2, 1240–1243

39 Roses, A.D. (2000) Pharmacogenetics and the practice of medicine. *Nature* 405, 857–865

40 Meyer, U.A. and Zanger, U.M. (1997) Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* 37, 269–296

41 Brosen, K. *et al.* (1992) Pharmacogenetics of tricyclic and novel

antidepressant:recept developments. *Clin. Neuropharmacol.* 15 (Suppl. 1), 80A–81A

42 Eichelbaum, M. and Gross, A.S. (1990) The genetic polymorphism of debrisoquine/sparteine metabolism – clinical aspects. *Pharmacol. Ther.* 46, 377–394

43 Daly, A.K. (1995) Molecular basis of polymorphic drug metabolism. *J. Mol. Med.* 73, 539–553

44 Weinshilboum, R.M. *et al.* (1999) Methylation pharmacogenetics: catechol O-methyltranferase. The thiopurine methyltransferase and the histamine N-methyltransferase. *Annu. Rev. Pharmacol. Toxicol.* 39, 19–52

45 Weinshilboum, R.M. (1992) Methylation pharmacogenetics. thiopurine methyltranferase as a model system. *Xenobiotica* 22, 1055–1071

46 Alves, S. *et al.* (2000). Screening of thiopurine *S*-methyltransferase mutations by horizontal conformation-sensitive gel electrophoresis. *Hum. Mutat.* 15, 246–253

47 Lennard, L. (1992) The clinical pharmacology of 6-mercaptopurine. *Eur. J. Clin. Pharmacol.* 43, 329–339

48 Halushka, M.K. *et al.* (1999) Patterns of single nucleotide polymorphisms in candidate genes for blood pressure homeostasis. *Nat. Genet.* 22, 329–339

49 Cargill, M. *et al.* (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22, 231–238

50 Cambien, F. *et al.* (1999) Sequence diversity in 36 candidate genes for cardiovascular disorder. *Am. J. Hum. Genet.* 65, 183–191.

51 Lander, E.S. and Schork, N.J. (1994) Genetic dissection of complex traits. *Science* 265, 2037–2048

52 Wright, A.F. *et al.* (1999) Population choice in mapping genes for complex diseases. *Nat. Genet.* 23, 397–404

53 Taillon-Miller, P. *et al.* (1999) Efficient approach to unique single-nucleotide polymorphism discovery. *Genome Res.* 9, 499–505

54 Newton, C.R. *et al.* (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17, 2503–2516

55 Whitcombe, D. *et al.* (1998) A homogeneous fluorescence assay for PCR amplicons: its application to real-time, single-tube genotyping. *Clin. Chem.* 44, 918–923

56 Gibson, N.J. *et al.* (1997) Colorimeteric assay for phosphate to measure amplicon accumulation in polymerase chain reaction. *Anal. Biochem.* 254, 18–22

57 Gibson, N.J. *et al.* (1997) A homogeneous method for genotyping with fluorescence polarization. *Clin. Chem.* 8, 1336–1341

58 Chen, X. *et al.* (1998) A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res.* 8, 549–556

59 Howell, W.M. *et al.* (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat. Biotechnol.* 17, 87–88

60 Marshall, D.J. *et al.* (1997) Determination of hepatitis C virus genotypes in the United States by cleavase fragment length polymorphism analysis. *J. Clin. Microbiol.* 35, 3156–3162

61 Griffin, T.J. and Smith, L.M. (2000) Single nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* 18, 77–84

62 Berkenkamp, S. *et al.* (1998) Infrared MALDI mass spectrometry of large nucleic acids. *Science* 281, 260–263

63 The Chipping Forecast (1999) *Nat. Genet.* 21 (Suppl.)

64 Nyren, P. *et al.* (1997) Detection of single-base changes using bioluminometric primer extension assay. *Anal. Biochem.* 244, 367–373

65 Boyce-Jacino, M.T. (1994) High volume molecular genetic identification of single nucleotide polymorphisms using genetic bit analysis: application to human genetic disease. *Am. J. Hum. Genet.* 55, A2088

66 Kuklin, A. (1998) Detection of single-nucleotide polymorphisms with the WAVE™ DNA fragment analysis system. *Genet. Test.* 1, 201–206

67 Orita, M. *et al.* (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* 86, 2766–2770

68 Chee, M. *et al.* (1996) Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614

69 Lyamichev, V. *et al.* (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* 17, 292–296

Collaboration…

ICOS Corporation (Houston, WA, USA) and **Texas Biotechnology Corporation** (Bothwell, WA, USA) have entered into a joint 50/50-owned venture to develop and globally commercialize endothelin A (ET_A)-receptor antagonists. Initially, this venture will focus on moving the selective ET_A-receptor antagonist, Sitaxsentan (TBC11251), into Phase IIb/III trials from its current evaluation in a Phase IIa trial for pulmonary hypertension and continue its clinical development for chronic heart failure. Applications for the second-generation endothelin-receptor antagonist, TBC3711, will also be explored.

[For more information on endothelin-receptor antagonists, see the October issue of *Drug Discovery Today*.]