Malaysian Journal of Medical Sciences, Vol. 12, No. 2, July 2005 (4-12)

REVIEW ARTICLE

THE USE OF SNPS IN PHARMACOGENOMICS STUDIES

Zilfalil Bin Alwi

Department of Paediatrics,
School of Medical Sciences, Universiti Sains Malaysia
16150 Kubang Kerian, Kelantan, Malaysia

Pharmacogenomics is the study of how genetic makeup determines the response to a therapeutic intervention. It has the potential to revolutionize the practice of medicine by individualisation of treatment through the use of novel diagnostic tools . This new science should reduce the trial-and-error approach to the choice of treatment and thereby limit the exposure of patients to drugs that are not effective or are toxic for them. Single Nucleotide Polymorphisms (SNPs) holds the key in defining the risk of an individual's susceptibility to various illnesses and response to drugs. There is an ongoing process of identifying the common, biologically relevant SNPs, in particular those that are associated with the risk of disease. The identification and characterization of large numbers of these SNPs are necessary before we can begin to use them extensively as genetic tools. As SNP allele frequencies vary considerably across human ethnic groups and populations, the SNP consortium has opted to use an ethnically diverse panel to maximize the chances of SNP discovery. Currently most studies are biased deliberately towards coding regions and the data generated from them therefore are unlikely to reflect the overall distribution of SNPs throughout the genome. The SNP consortium protocol was designed to identify SNPs without any bias towards these coding regions. Most pharmacogenomic studies were carried out in heterogeneous clinical trial populations, using case-control or cohort association study designs employing either candidate gene or Linkage disequilibrium (LD) mapping approaches. Concerns about the required patient sample sizes, the extent of LD, the number of SNPs needed in a map, the cost of genotyping SNPs, and the interpretation of results are some of the challenges that surround this field. While LD mapping is appealing in that it is an unbiased approach and allows a comprehensive genome-wide survey, the challenges and limitations are significant. An alternative such as the candidate gene approach does offer several advantages over LD mapping. Ultimately, as all human genes are discovered, the need for random SNP markers diminishes and gene-based SNP approaches will predominate. The challenges will then be to demonstrate convincing links between genetic variation and drug responses and to translate that information into useful pharmacogenomic tests.

***Key words** : Single Nucleotide Polymorphism (SNP), Pharmacogenomics, Pharmacogenetics*

Introduction

Pharmacogenomics is one of the most promising sciences for the pharmaceutical industry to emerge in the post-genomic era.

Pharmacogenomics, can be broadly defined, as the study of the impact of genetic variation on the efficacy and toxicity of drugs, or the study of how genetic makeup determines the response to a therapeutic intervention.[1] As the volume of high quality genetic and genomic information for predicting the response to drugs become available, better clinical trials and more targeted drug development will then follow.

Pharmacogenomics has the potential to revolutionize the practice of medicine by

individualisation of treatment through the use of novel diagnostic tools . This new science should reduce the trial-and-error approach to the choice of treatment and thereby limit the exposure of patients to drugs that are not effective or are toxic for them.

Pharmacogenetics

The term pharmacogenetics is often used interchangeably with pharmacogenomics, but is used more generally to describe the study the effect of genetic factors on drug response (1).

Most of the variations or polymorphisms describe to data occur in the drug metabolizing enzymes (DMEs) or cytochrome P450 enzymes. However, polymorphisms of drug transportation genes and genes that encode protein receptors and other effectors also lead to variations in drug response.

Pharmacogenomics vs. Pharmacogenetics

The many interactions involving cytochrome P450 illustrate the importance of pharmacogenomics (2). Likewise, we now have extensive compilations of clinically relevant polymorphisms (SNPs) that influence other drug metabolism pathways. For example, it is estimated that 2-10% of the population is homozygous for non-functional CYP2D6 mutant alleles, leading to an inability to activate opioid analgesics. This would explain why there is great variability in pain relief experienced by patients receiving the same dose of codeine.

Functional genomics

Functional genomics is the study of the relationships between particular genotypes and specific phenotypes.

Pharmacoproteomics

Pharmacoproteomics is the subtyping of patients on the basis of protein analysis. This mode of characterization is a more functional representation of patient-to-patient variation than is provided by genotyping, and includes the added effects of post-translational modification. Thus, pharmacoproteomics connects the genotype with the phenotype.

Single Nucleotide PolymrophISMS (SNP)

Every individual carries two copies of each gene. Copies of a specific gene present within a population may not have identical nucleotide sequences. These single nucleotide changes are scattered throughout the genome of all species and forms the basis for human diversity. SNP occur in humans every 300-2000 base pairs along the genome (3) In principle, they may occur at any nucleotide, and for genetic epidemiological study, those that are relatively common will be of greatest interest.

The vast majority of SNPs are functionally silent, occurring in non-coding or non-regulatory regions of the genome. However, some of the SNPs lead to altered protein structure or expression. These biologically functional SNPs are considered the essence and substrate of human diversity in both health and disease.

There is an ongoing process of identifying these common, biologically relevant SNPs, in particular those that are associated with the risk of disease. Once identified and characterized, this SNP-based 'genetic profile', may be viewed as a 'fingerprint', useful in defining the risk of an individual's susceptibility to various illnesses and response to drugs.

History

In the 1980s, single nucleotide polymorphisms (SNPs) were detected using restriction enzymes to identify the presence or absence of cutting sites and scored by observing the resulting fragment length variation (4). In the 1990s, the SNP was largely replaced by the simple tandem repeat (STR) as the marker of choice for linkage studies. STRs (di-, tri- or tetranucleotide repeats) show high levels of allelic variation in the number of repeat units, are widely and evenly distributed across the human genome and can be typed using Polymerase Chain Reaction (PCR) amplification. The combination of a highly polymorphic marker set and rapid typing technology led to the development of high-throughput semi-automated systems for STR genotyping during the 1990s (5).

The late 1990s saw a reversal from the use of STRs back to SNPs, which regained favour amongst molecular geneticists. Following the completion of the Human Genome Project in 2001, there has been further increase on the number of studies as well as interests on SNPs. The main driving force behind the switch back to SNPs was a change in the type of genetic studies undertaken by the various research groups. STRs are ideal for linkage studies involving pedigree analysis to identify single genes responsible for monogenic disorders. However, more recently the need to study diseases with more complex inheritance pathways, but with a higher prevalence and hence higher social burden, such as osteoporosis,

diabetes, cardiovascular and inflammatory diseases, psychiatric disorders and most cancers, has led to a refocus on SNPs. Moreover, there has also been increasing interest in the genetics of drug response (pharmacogenetics), an understanding of which may allow the 'tailoring' of therapies on an individual basis.

The broadly familial nature of complex diseases clearly indicates a significant genetic component. However, in contrast to monogenic conditions, this genetic element is comprised of multiple gene variants each contributing a small effect. This genetic complexity may also be compounded by heterogeneity, with different combinations of gene variants giving rise to a similar phenotype. The extent of this problem is likely to be so great that the frequency of any polymorphism contributing to a disease phenotype may be only slightly elevated in a disease group when compared with unaffected controls. Unfortunately, linkage analysis has limited power to detect such small effects, and attempts to identify genes involved in complex disease using linkage-based approaches have generally proved disappointing. Association studies with large sample sizes, and involving comparisons of cases of disease with matched controls from the same population, are likely to give a greater chance of detecting small effects.

Until now, population-based case–control studies have been limited to attempts to associate one or a few 'candidate genes' with disease. This restricted approach has been due largely to the lack of appropriate genetic markers and the inadequacy of the available genotyping tools for the high-throughput approaches required for large-scale genome-wide experiments. Ironically, the STR markers that have been so successful in the study of monogenic disease will probably be of limited value in population-based studies. The high level of variation reflects high mutation rates, which are likely to confound population-based approaches (6). Furthermore, due to the large number of markers required, STR loci may be too sparse for association-based approaches.

In contrast, SNPs are abundant and are more stable than STRs due to lower mutation rates. Moreover, STR loci suffer from being 'surrogate' markers in the sense that polymorphism in the STR is used to locate an adjacent functional variant that contributes to the disease state. Variation at the STR itself rarely contributes to the phenotype. While SNPs may also act as surrogate markers, many SNPs have functional consequences if they occur in the coding or regulatory regions of a gene. Therefore, by using SNP markers, it is often possible to test for association between a phenotype and a functional variant directly. For these reasons, SNPs are preferred for drawing the high-density genetic marker maps required for one of the major thrusts in human genetics research: the unraveling of complex genetic traits.

Methods for Identification of SNPs

The identification and characterization of large numbers of SNPs are necessary before we can begin to use them extensively as genetic tools. A pool of several hundred thousand SNPs will be required as a resource for the construction of optimized marker sets for association studies.

There are five commonly used methods for SNP (or mutation) detection (7 - 11).

(1) Single strand conformation polymorphisms (SSCPs)

(2) Heteroduplex analysis

(3) Direct DNA sequencing

(4) Variant detector arrays (VDAs).

(5) DNA microarray technology.

(i) SSCP detection

For SSCP detection, the DNA fragment spanning the putative SNP is PCR amplified, denatured and run on a non-denaturing polyacrylamide gel. During the gel run, the single-stranded fragments adopt secondary structures according to their nucleotide sequences. Fragments bearing SNPs are identified as a result of their aberrant migration patterns and confirmed by sequencing.

Although SSCP is a widely used and relatively simple technique, its success rate for SNP detection, has been variable, typically ranging from 70 to 95%[7]. It is a labour-intensive method and has a relatively low throughput, although higher capacity methods using capillary- rather than gel-based detection are under development (9).

(ii) Heteroduplex analysis

Heteroduplex analysis relies on the detection of a heteroduplex formed during reannealing of the denatured strands of a PCR product derived from an individual heterozygous for the SNP. The heteroduplex can be detected as a band shift on a

gel, or by differential retention on a high-performance liquid chromatography (HPLC) column.

HPLC has rapidly become a popular method for heteroduplex-based SNP detection due to its simplicity, low cost and high rate of detection (95-100%) (12) Reasonable throughput at 10 min per sample can be achieved with commercially-available systems such as the Transgenomic Wave (13).

(iii) Direct DNA sequencing

Currently, the favoured high-throughput method for SNP detection is direct DNA sequencing. Once the sequencing reactions have been completed, a single capillary system (e.g Applied Biosystems 3700) can generate sequences from more than 1500 DNA fragments of 500 bp in 48 h with minimal human intervention. Dye-terminator sequencing chemistry will detect 95% of heterozygotes and the more expensive and labour-intensive dye-primer chemistry may identify all sequences (8).

The recently formed SNP consortium (TSC), a non-profit foundation sponsored by 10 major pharmaceutical companies and the UK Wellcome Trust, has used dye-terminator sequencing to identify and has succeeded in mapping more than 100 000 SNPs by 2001.

SNPs may also be detected *in silico* at the DNA sequence level. The wealth of sequence data deposited in public databases in recent years, in particular expressed sequence tag (EST) sequences, allows SNPs to be detected by comparing multiple versions of the same sequence from different sources.

(iv) VDA technology

VDA technology is a relatively recent addition to the high-throughput tools available for SNP detection.

This technique allows the identification of SNPs by hybridization of a PCR product to oligonucleotides arrayed on a glass chip and measuring the difference in hybridization strength between matched and mismatched oligonucleotides.

The VDA detection rate is comparable to that of dye-terminator sequencing and allows rapid scanning of large amounts of DNA sequences. For example, Wang et al[7] used this technique to identify 2500 SNPs in 2 Mb of human DNA and, more recently, Halushka et al (13) have used the same method to identify 874 SNPs in 75 candidate genes for hypertension.

Pyrosequencing

Pyrosequencing, described by Ahmadian et al (14) is a sequencing-by-synthesis method in which a cascade of enzymatic reactions yields detectable light radiation, characteristic of the incorporated nucleotides. One feature of typing SNPs with pyrosequencing is that each allelic variant, being unique in sequence, can easily be distinguished by pattern-recognition software. The software displays the allelic alternatives and allows for direct comparison with the pyrosequencing raw data.

For optimal determination of SNPs, various protocols for the order of dispensing of the nucleotides should be used. Ahmadian et al demonstrated that suitability of the technique for large-scale screening and typing of SNPs by pyrosequencing 96 samples in approximately 5 min using an automated system for parallel analysis (14).

(v) DNA Microarray Technology

The DNA microarray technology is the latest, cutting edge technology for the studies on SNPs (10, 11). It offers a biotechnological revolution with the help of DNA chemistry, silicon chip technology and optics to be used to monitor gene expression for thousands of genes in one single experiment. Briefly, 20,000 to 100,000 unique DNA molecules get applied by a robot to the surface of silicon wafers (approximately the size of a microscope slide). Using a single microarray experiment, the expression level of 20,000 to 100,000 genes could be examined in one single experiment. Microarray tools are now used on regular basis for monitoring gene expression of large number of genes and also frequently applied to DNA sequence analysis, genotyping, and molecular diagnosing. These tools can be used to distinguish and differentiate between different DNA fragments that differ by as little as a single nucleotide polymorphism (SNP), making it a powerful tool for identifying novel molecular drug targets and for elucidating mechanisms of drug action. Furthermore, microarrays can monitor the global profile of gene expression in response to specific pharmacologic agents, providing information on drug efficacy and toxicity (11).

Sample population

In addition to choosing a method for SNP detection, the population in which the SNPs are to be detected must be defined. SNP allele frequencies vary considerably across human ethnic groups and populations. The SNP consortium has opted to use

an ethnically diverse panel to maximize the chances of SNP discovery.

In their study of hyertension, Halushka et al. (15) analyzed African and Northern European populations due to known differences in prevalence and disease phenotype in these two ethnic groups. Other studies use populations with a target disease for SNP discovery, on the logic that variants contributing to the disease state should occur with higher frequencies in such cohorts (16).

Given that any polymorphism is likely to make only a small contribution to a disease phenotype and that it will be found at only a slightly higher frequency in the disease cohort compared with a control group studies with a matched non-diseased population is necessary.

Different SNP panels will be required for different studies. However, a diverse approach is necessary for the generation of a large pool of SNPs from which to draw the most appropriate panel for any given study.

Recent advances

By 1999, nearly 300 genes have undergone detailed analysis for SNP content (15, 17). Although the methods and populations used in each study were different, several useful inferences can be made. Changes in non-coding sequence and synonymous changes in coding sequence are generally more common than non-synonymous changes. This reflects greater selective pressure for reducing diversity at positions dictating amino acid identity. Transitional changes are more common than transversions, with CpG dinucleotides showing the highest mutation rate, presumably due to deamination. There is enormous diversity in SNP frequency between genes, reflecting different selective pressures on each gene as well as different mutation and recombination rates across the genome. The degree of linkage disequilibrium varies widely across different genes, again reflecting different recombination and mutation rates.

SNP consortium (TSC)

The identification and study of SNPs in specific genes has provided useful confirmation of hypothesized models for gene and genome dynamics. However, as such studies are usually biased deliberately towards coding regions, the data generated from them are unlikely to reflect the overall distribution of SNPs throughout the genome.

In contrast, the protocol used by the SNP consortium protocol was designed to identify SNPs with no bias towards the coding regions, and the 100 000 TSC SNPs mapped should generally reflect sequence diversity across the human chromosomes. However, the data set will not be completely free of bias. For example, selection will occur against sequences that are unclonable using the TSC protocol.

The TSC aimed to expand the number of SNPs identified across the genome to 300 000 by the end of 2001. Data are released quarterly via both the TSC's own web page and the SNP database dbSNP, hosted by the National Center for Biological Information (NCBI; http://www.ncbi.nlm.nih.gov/SNP/index.html). By December 2002, 2,536,021 dbSNP had been identified.

These initiatives come in response to efforts in the biotechnology industry to identify and patent large numbers of SNPs. Most notable are the efforts by Celera Genomics (Rockville, MD), Genset (Paris, France), CuraGen (New Haven, CT), and Incyte Genomics (Palo Alto, CA).

The Use of SNP Maps in Pharmacogenomics

There are two approaches (18) for the use of SNP maps in pharmacogenomics: the candidate gene approach and linkage-disequilibrium mapping.

(i) Candidate gene approach

The candidate gene approach uses biological paradigms or a prior knowledge of disease pathogenesis to identify genes relevant to disease. SNPs found in these genes are tested for statistical association with disease in patients enrolled in family, case-control, or cohort studies. These "susceptibility genes" are hypothesized to directly influence an individual's likelihood of developing the disease.

This approach has already been extended to identifying candidate genes affecting drug response. For example, gene variants in a drug-metabolizing enzyme (thiopurine methyltransferase; *TPMT*) have been linked to adverse drug reactions (19). Gene variants in a drug target (5-lipoxygenase; *ALOX5*) have been associated with variation in drug response (20) and variants in a disease susceptibility gene (apolipoprotein E; *APOE*) have been correlated with response to a cholinesterase inhibitor in Alzheimer's patients (21)

(ii) Linkage disequilibrium mapping

An alternative to the candidate gene approach

is linkage disequilibrium mapping. This approach relies on linkage disequilibrium (LD) or nonrandom association between SNPs in proximity to each other.

Tens to hundreds of thousands of anonymous SNPs need to be identified and their location in the genome mapped. Although these anonymous SNPs may fall within genes and may in fact be susceptibility SNPs, most are located in the vast noncoding DNA regions between genes and play no obvious role in drug response. Through LD, associations found, the anonymous markers can be used to identify a region of the genome that may harbor a susceptibility gene without any a priori assumptions about what or where the susceptibility gene is. Additional significant efforts using positional cloning are then required to find the specific gene and the SNPs within it that confer the underlying association.

Linkage disequilibrium mapping has been employed successfully on families with multiple affected individuals to uncover genes for monogenic diseases (22). Even though this mapping technique is now being considered in the context of association studies, it has not been successful for identifying genetic predictors of either disease or drug response in unrelated individuals.

Limitationns of SNPs as A Tool in Pharmacogenomics Analyses

Studies of the genetic basis of disease can take advantage of characteristics of familial inheritance, use of homogeneous populations and relatively straightforward case-ascertainment of affected individuals. In contrast, pharmacogenomic analyses are more complicated:

(i) Drug response is a trait whose expression can only be gauged after administration of the therapeutic compound under study. Ascertainment of responders cannot be made from the non-exposed general population and use of families, is generally precluded except in the rare instance where multiple family members are given the drug.

(ii) Clinical trials are the main source of patients for pharmacogenomic studies and these are limited in size, making estimation of linkage-disequilibrium, often impossible, and usually imprecise.

Moreover, although drugs on the market may be sold worldwide, most clinical trials of new therapies are performed in Caucasian Americans or Europeans. Pharmacogenomic studies are therefore usually limited to these genetically heterogeneous clinical trial populations. Case-control or cohort association studies are usually employed to identify candidate gene or for LD mapping.

(iii) The cost of genotyping SNPs, and the interpretation of results are also problems.

Sample size:

The number of patients required to find a statistically significant association between an SNP and an abnormal drug response depends on a number of factors, including the frequency of the drug response, the proportion of patients having the SNP allele, the minimum detectable drug effect using existing diagnostic criteria, the level of statistical significance (*p* value or probability of missing a true difference), and the power (the probability of not missing a true association) required.

The SNPs most likely to have a direct impact on the protein product of a gene are coding region SNPs (cSNPs) that change amino acid sequence, and SNPs in gene regulatory regions, which control protein levels. Coding SNPs that confer association in a recessive fashion (where two copies of the cSNP are required) may occur in too few patients to be useful as pharmacogenomic markers (17, 20).

The extent of LD

(Estimating the number of markers needed in a SNP map)

Genome-wide SNP LD mapping is predicated on the assumption that LD exists between SNPs. The extent of LD occurs as a consequence of many factors, including population admixture, genetic drift, mutation, and natural selection (23). For genetic distances measured in kilobases (kb) of DNA, LD tends to decline with larger distance between SNPs in the range of 10–100 kb. Over shorter genetic distances the degree of LD is highly variable from one genomic region to the next. In some genomic regions, LD extends over several thousands of kilobases, whereas in other genomic regions surrounding single genes, LD can be quite small. Theoretical estimates of the average extent of LD in the human genome vary widely, ranging from <100 kb to <3 kb (24, 25).

As increasing amounts of genetic data become available, the true extent of LD throughout the genome can be tested empirically.

Understanding the average extent of LD is useful for estimating the number of markers needed in a SNP map and the strength of the association that the markers are capable of detecting.

Linkage disequilibrium mapping requires that a susceptibility allele be detectable with a marker that lies within the interval afforded by the SNP map density. Given the estimated 3 billion bp size of the human genome, a minimum of 30,000 to 500,000 evenly spaced SNP markers would be needed to have a marker every 100 to 6 kb (i.e., within the range of LD). Maps currently under construction range in size from 60,000 to 300,000 SNPs (26) resulting in a SNP mapped on average every 50 to 10 kb, respectively. Whereas this density may be useful for uncovering SNPs in genomic regions with extensive LD, genes in regions where LD is less extensive may be missed. To improve the chance of successfully identifying susceptibility SNPs through LD, high-density maps will be required.

The strength of LD

The strength of LD will also affect the magnitude of an association. A marker in LD with a susceptibility SNP will yield a relative risk that is smaller than if the susceptibility SNP were tested directly. D' is a measure of linkage disequilibrium (27) that ranges in value from 0 (no disequillibrium) to 1.0 (complete disequilibrium). The weaker the LD between marker and susceptibility SNPs, the smaller the relative risk, and the more difficult the association is to detect unless the sample size is increased proportionately. This situation becomes even more complex when large differences in allele frequencies exist between markers and susceptibility SNPs. If marker allele frequencies are substantially different from the susceptibility allele frequency, then the required sample size, the number of markers, or both will need to be dramatically increased.

Sample sizes for LD mapping

As LD mapping requires testing of hundreds of thousands of markers, the chance of producing false positive results is high. To reduce the number of false positive results, a correction can be applied whereby a more stringent cutoff is used for establishing statistical significance. To achieve an overall 5% false positive rate when 100,000 independent markers are tested, a *p* value of 0.0000005 should be used (18)

Cost-of-genotyping

One of the major challenges of LD mapping is the need to genotype each person in the study for every one of the 60,000–500,000 SNPs in the map. Several genotyping platforms are available today, including nucleic acid hybridization on filters or chips, single-strand conformational polymorphism (SSCP), and primer extension-based methods. Although these genotyping technologies are robust, at a current average price of one dollar per genotype, their use in large-scale SNP genotyping studies may be very expensive. Even at one cent per genotype, the cost per person in a typical association study testing 100,000 SNPs will be about $1,000, possibly adding $1 million to the cost of a clinical trial. A method that involves pooling of patient DNA samples has been suggested to reduce the overall number of genotypes needed (28). However, pooling presents technical challenges and drawbacks because it prohibits subgroup and haplotype analysis. Significant advances are required to make extensive genotyping a standard part of clinical trials.

Interpretation of results

Interpretation of data from pharmacogenomic association studies is challenging. Two questions should be asked: "Is any association detected real" and "are the results useful?" Studies yielding statistically significant results are often considered real even though the results are rarely replicated. Consistency of results between studies is a major issue for both candidate gene studies and studies employing genome-wide LD mapping. Furthermore, the use of low *p* values is recommended to allow for the vast number of genetic hypotheses that will be tested collectively in the field. Very low *p* values and reproducibility should be part of any set of criteria for judging the reality of associations, especially when the results trigger further investment in positional cloning (for LD mapping studies) or are the basis of diagnostic tests used to convey risks to patients or direct a course of therapy.

How strongly will SNPs be associated with drug response?

The use of a SNP associated with drug response can be judged in terms of how well it predicts drug response in patients, and the proportion of patients who will benefit from the test. In the recent paper by Drazen and coworkers (20), an *ALOX5* genotype was associated with response to an antiasthmatic compound. The genotype had a 100% positive predictive value for nonresponse to the drug. However, because the susceptibility genotype is uncommon (6–9% of patients), less than 10% of the nonresponse can be attributed to this

genotype. Therefore, if patients with the susceptibility genotype avoided taking the drug, the efficacy would only improve from 46% to 51% in the remaining patients. Whereas the test may benefit a few patients who would otherwise receive unnecessary and ineffective therapy, it will not identify the majority of nonresponders. In order to have practical utility, additional SNPs will need to be identified and used together with the *ALOX5* SNP as a battery where each individual SNP explains a small portion of drug response. There may of course also be non-genetic causes of non-response.

The"effect-size"

The "effect size" of an association in pharmacogenomics is the likelihood of response to a drug in individuals with the susceptibility allele compared with those without the allele. Depending on whether the study design is a cohort or case-control, the magnitude of the effect is usually expressed as a relative risk (RR) or as an odd ratio (OR), respectively. Deviations from the baseline value of 1.0 for either measure indicate increased or decreased likelihood of response. Relative risk estimates for common diseases are expected to be low, in the range of 1.5 to 3.0, owing to their multifactorial nature.

Drug response is just as complex as disease genetics, resulting not only from underlying genotypic variation at many loci, but also from variation at the level of gene expression, post-translational modification of proteins, drug dose, drug interactions, diet, and other nongenetic factors. Therefore, usually individual genes will be associated with relatively small effects on drug response. In fact, pharmacogenomic markers reported to date confer only about a twofold increased likelihood of response (e.g. RR = 2.0) (18)

Future-prospects

While LD mapping is appealing in that it is an unbiased approach and allows a comprehensive genome-wide survey, limitations are significant. A number of challenges need to be overcome before the value of a high-density SNP map can be maximised for practical use. Concerns about the high price of genotyping are being addressed but it may be several years before the price of genotyping large populations becomes acceptable. In addition, availability of large patient populations will be crucial for discovering and validating SNPs. The extent of LD and success in detecting associations with small effect that will identify situations for which SNP LD mapping could work.

The candidate gene approach is an appealing alternative to the LD mapping. It is a proven method. Genotyping a limited number of candidate SNPs is already economically feasible with this method and no assumptions are made about LD. The required sample sizes are consistent with the number of subjects recruited in current clinical trials.

As more human genes are discovered, the need for random SNP markers will diminish and gene-based SNP approaches will predominate. The real challenges will then be to demonstrate convincing links between genetic variation and drug responses and to translate that information into useful pharmacogenomic tests.

How likely will individualized pharmacotherapy, within a pharmacogenetic framework, become a reality? While the technical basis for these developments is in place today and appears quite logical, two tasks are needed:

First, the creation of the necessary knowledge base for genetic risk profiling, an enormous and dauntingly difficult task and second, the acceptance of these new approaches by the general public. The experience with patient advocacy groups for single gene disorders shows that efforts to find any causative gene usually find strong support, being recognized as the first, essential step towards treatment, cure or prevention.

Indeed patients may well become the driving force behind the development of this vision of integrated and individualized medicine as ultimately, it is they who stand to gain most. However, as corporations stand to gain much too, vigilance is required to ensure that societal agenda of maximum health gain is not hijacked to be transformed into one aimed at maximizing profits. Some would argue that this is the major challenge to genomic medicine.

Acknowledgements

I would like to thank Dr. Maureen Boxer of Glasgow University for all her help and support.

Correspondence :

Dr. Zilfalil Bin Alwi, MBBS (Dac), MSc (Glasgow), MMed (USM).
Department of Paediatrics,
School of Medical Sciences,
Universiti Sains Malaysia, Health Campus,
16150 Kubang Kerian, Kelantan, Malaysia
Tel: +609-766 4151 Fax: +609-765 8914
Email: zilfalil@kb.usm.my

References

1. Norton P. Peet and Philippe Bey Pharmacogenomics: challenges and opportunities *Drug Discovery Today* 2001, **6**:495-498
2. Cytochrome P450 Drug interaction Table Georgetown University Medical Center Web Page, (http://www.georgetown.edu/departments/pharmacology/davetab.html)
3. Klaus Lindpaintner Genetics in drug discovery and development; challenge and promise of individualizing treatment in common complex diseases. *British Medical bulletin* 1999; **55(2)**: 471-491
4. Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*,1980; **32**: 314–331.
5. Livak KJ, Marmaro J, Todd JA Towards fully automated genome-wide polymorphism screening *Nature Genetic* 1995; **9:** 341-342
6. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc.Natl Acad.Sci.USA* 1997; **94:** 1041-1046
7. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N et al. Large scale identification, mapping, and genotyping of SNP in the human genome. *Science* 1998; **280:** 1077-1082
8. Gross E, Arnold N, Geotte J, Schwarz-Boeger U, Kiechle M A comparison of BRACA1 mutation analysis by direct sequencing, SSCP and DHPLC. *Human Genetics* 1999; **104:** 29-35
9. Wenz HM, Baumhueter S, Ramachandra S, Worwood M. A rapid automated SSCP multiplex capillary electrophoresis protocol that detects the two common mutations implicated in hereditary hemochromatosis. *Human Genetics* 1999; **104:** 29-35
10. Khan AS Genomics and microarray for detection and diagnostics. *Acta Microbiol Immunol Hung*. 2004; **51(4):** 463-7.
11. Nees M, Woodworth CD Microarrays: spotlight on gene function and pharmacogenomics. *Curr Cancer Drug Targets.* 2001; **1(2):** 155-75.
12. O'Donovan MC, Oefner PJ, Roberts SC, Austin J, Hoogendoorn B, Guy C, Speight G, Upadhyaya M, Sommer SS, McGuffin P. Blind analysis of denaturing high performance liquid chromatography as a tool for mutation detection. *Genomics* 1998; **52:** 44-49
13. Buyse IM, Fang P, Hoon KT, Amir RE, Zoghbi HY, Roa BB Diagnostic testing for Rett Syndrome by DHPLC and direct sequencing analysis of the MECP2 gene: identification of several novel mutations and polymorphisms. *Am.J.Hum. Genet* 2000; **67:** 1428-1436
14. Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, LundebergJ SNP analysis by pyrosequencing. *Anal. Biochem* 2000 10; **280(1):** 103-10
15. Halushka MK, Fan JB, Bently K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of SNPs in candidate genes for blood pressure homeostasis. *Nature Genetics* 1999; **22:** 239-247
16. Ohnishi Y, Tanaka T, Yamada R, Suematsu K, Minami M, Fujii K, Hoki N, Kodama K, Nagata S. Identification of 187 SNPs among 41 candidate genes for ischaemic heart disease in the Japanese population. *Hum Genet* 2000; **106:** 288-292.
17. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N. Characterization of SNPs in coding regions of human genome. *Nature Genetics* 1999; **22:** 231-238
18. Jeanette J, McCarthy, Rolf Hilfiker The use of SNP maps in pharmacogenomics *Nature Biotechnology* 2000, **18 (5):** 505-508
19. Krynetski EY, Evans WE. Pharmacogenetics as a molecular basis for individualised drug therapy: the thiopurine S- methyltransferase paradigm. *Pharm. Res* 1999; **16:** 342-349
20. Drazen JM, Yandava CN, Dube L, Szczerback N, Hippensteel R, Pillari A, Israel E, Schork N, Silverman ES, Katz DA, Drajesk J. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat.Genet* 1999; **22:** 168-170
21. Poirier J, Delisle MC, Quirion R, Aubert I, Farlow M, Lahiri D, Hui S, Bertrand P, Nalbantoglu J, Gilfix BM. Apolipoprotein E4 allele as a predictor of cholinergic deficits and treatment outcome in Alzheimers disease. *Proc Natl Acad Sci U S A.* 1995, 19; **92(26):** 12260-4.
22. Friedman TB, Liang Y, Weber JL, Hinnant JT, Barber TD, Winata S, Arhya IN, Asher JH Jr. A gene for congenital recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nat Genet.* 1995; **9(1):** 86-91.
23. Hartl DI, Clark AG Principles of population genetics. Sinauer Associates, Sutherland, MA; 1990.
24. Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M. Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am J Hum Genet.* 1994; **54:** 884-898

25 Kruglyak L Prospects for whole genome linkage disequilibrium mapping of common disease genes. *Nat Genet.* 1999 ; 22: 139-144

26. Marshall E Drug firms to create public database of genetic mutations. *Science.* 1999, 16; **284(5413):** 406-7.
27. Lewontin RC The detection of linkage disequilibrium in molecular sequence data. *Genetics.* 1995; **140(1):** 377-88.
28. Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. Association mapping of disease loci by use of a pooled DNA genomic screening. *Am J Hum.Genet.* 1997; **61:** 734-747