The emergence of
pharmacogenomics
within the pharmaceutical industry

The sequencing of the human genome represents one of the most significant scientific advances of the 20th century that will shape the foundation of medical research well into the 21st. This accomplishment was enabled by remarkable technological advances – high throughput sequencing, increased computing power, automated methods of analysis – that 25 years ago seemed unimaginable. Through this project, we have gained the understanding that human beings are an estimated 99.9% identical at the genetic level. Yet, it is the 0.1% of variation among individuals that serves as the foundation for the emerging discipline of pharmacogenomics. It is this variation that contributes to physical diversity in the human population as well as differences in disease susceptibility and response to pharmacological therapies. This contribution of the Human Genome Project offers the opportunity to shape the face of drug discovery and development in this century.

P harmacogenetics was first described in the 1950s and can be defined, with broad agreement, as the study of DNA sequence variation or genotype as it relates to differential drug response. No general agreement on a definition of pharmacogenomics has yet been reached. The Pharmacogenetics Working Group (an informal collection of representatives from pharmaceutical companies that meet to discuss non-competitive issues related to pharmacogenetics) defines pharmacogenomics as 'the study of the genome and its products (including RNA and protein) as they relate to drug discovery and development' (see Drug Information Association website at www.dia-home.org). This is the sense in which most companies use the term although recent authors have attempted to define pharmacogenomics more narrowly as the "study of differences in drug response due to variation in the expression of the individual genes in the cells of particular tissues"[1].

At first pharmacogenetics involved the characterisation of interindividual variation in drug metabolising enzymes. It was determined by traditional phenotype measurements such as urinary metabolite response to a probe drug. During the 1980s and 90s phenotypic variation began to be understood at the level of gene mutations (defined as <1% frequency in the population) or polymorphisms (≥1% in the general population) within major classes of drug metabolising enzymes. Today the phenotype of interest is more diverse. It may be a biochemical measurement (eg fasting glucose level), a physical measurement (eg VO_2), demographic data (eg gender, smoking status, age) or outcomes data (relapse, remission).

The simplest correlation involves DNA mutations or polymorphisms within a gene that disrupt the normal activity, structure, or function of that gene's product and result in an easily observable phenotype. The earliest example of this was

**By Aidan Power,
Suzin Webb,
Dr Albert Seymour
and Dr Patrice Milos**

the discovery of a molecular basis for the ABO blood groups[2]. A more current example is the surprisingly large interindividual variability observed in the activity of critical drug metabolising enzymes, particularly the cytochrome P450 enzyme family. There are numerous reports of genetic contributions to this inherent variability. In fact, for some genes (CYP2D6, CYP2C19, and others) the genotype-phenotype correlation approaches 100%[3-6].

Pharmacogenetics relies on the correlation of genetic polymorphisms with an observable phenotype. The emergence of new information on genetic variation across the entire genome and the extensive cataloguing of this variation through the efforts of The SNP Consortium have provided us with a new foundation to further explore the science of pharmacogenomics. Through scientific and technological advances in genetics and genomics we see a future with a much more comprehensive understanding of human disease, adding valuable knowledge to putative therapeutic targets and emerging insight into how individuals respond to drug therapies.

The challenge for the pharmaceutical industry

There are a number of human diseases for which a specific causative genetic mutation in a particular gene is known, such as cystic fibrosis, Huntington disease, Tay-Sachs disease and sickle cell anaemia[2]. The more common or ethnic specific mutations causing these diseases can be detected, and families with a history of these disorders often use genetic information for family planning and counselling purposes. Fortunately, these diseases are relatively rare in most communities. However, the diseases that are most prevalent in all communities have a complex etiology that undoubtedly involves multiple genes and environmental effects. Overcoming this complexity is the key challenge the pharmaceutical industry faces in designing and developing therapeutic agents to treat diseases such as obesity, diabetes, cardiovascular disease, atherosclerosis, osteoporosis, rheumatoid arthritis, osteoarthritis, infectious diseases, chronic pain, neurodegenerative diseases, depression, schizophrenia, immunodeficiencies, allergies, respiratory disorders and cancer.

For most (if not all) of these diseases there is no known cure, only therapeutic intervention. And these therapeutic interventions are not effective in all patients. Recently a common theme has evolved among pharmaceutical companies – that of 'personalised medicine'[7]. The goal is to discov-

er and prescribe therapies that are tailored to an individual's needs. The potential impact of personalised medicine may be large: safer medicines, faster relief and, ultimately, cost savings for healthcare. This is very much a vision for the future and we are just beginning to understand how to approach this goal. Pharmacogenetics could play a very important role in realising personalised medicine as we begin to identify genetic markers associated with superior efficacy. In the future, individuals with a particular disease may be genetically screened to determine which therapeutic agent will have the highest probability of delivering efficacy. But this future is quite distant and the industry faces many hurdles.

The requirements for pharmacogenomic studies

Pharmacogenomics moves the pharmaceutical industry into new avenues of research that require technologies and methods of data analysis that are not traditionally part of the core business. Execution of pharmacogenomic studies is dependent on many diverse pieces of information that must be brought together to integrate genotype and phenotype. These include access to: accurate clinical and demographic data; DNA samples from well designed studies; single nucleotide polymorphisms; genotyping technologies; informatics technologies to handle large quantities of data; statistical methodologies for data analysis and interpretation as well as general education within the pharmaceutical setting in the area of pharmacogenomics. Each of these is explored in more detail below.

Accurate clinical and demographic data
The single most important requirement for pharmacogenomics is well-defined information for the disease or phenotype being studied. Without these data one can never expect to be successful in exploring the relationship between subject differences and disease state or drug response. In the pharmaceutical industry we have the opportunity to conduct well-defined clinical studies in specific disease populations. These studies often provide the opportunity to evaluate genetic contributions to disease within the parameters of the entry criteria for that particular trial, for example subjects with LDL levels within certain limits. However, it must be realised that these are selected populations for the purpose of a clinical trial and not always representative of the wider population with disease. Many trials also offer the opportunity to evaluate longer-term outcomes of morbidity and
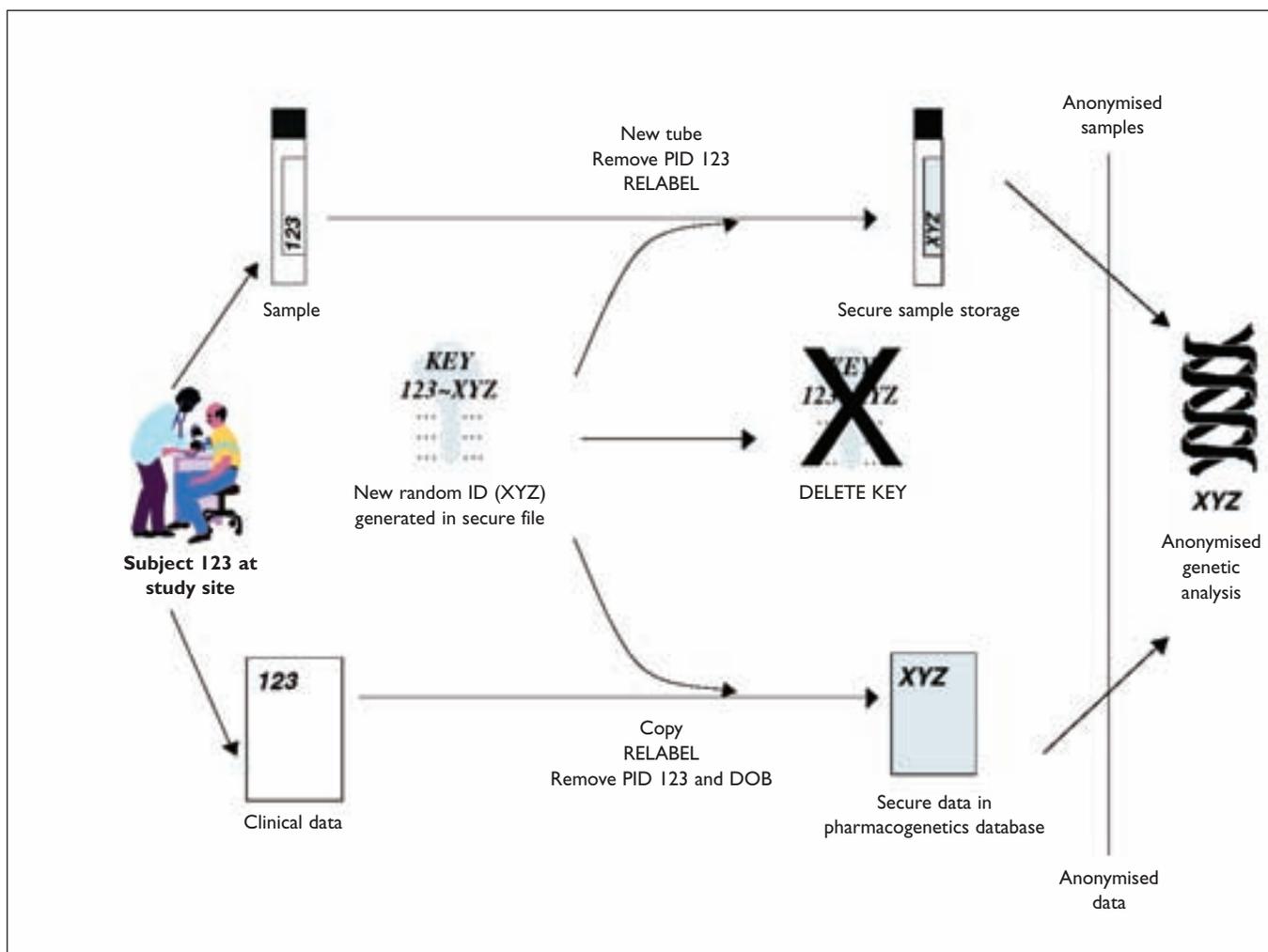
mortality – data that are not easily obtainable through other kinds of studies.

DNA samples
Currently our efforts in Pfizer, as in many other companies, focus on the collection of DNA from clinical trial subjects who voluntarily agree to donate a portion of their blood for research purposes. Informed consent is a critical part of this process and a separate informed consent is provided to patients who agree to participate, informing them of the purpose of the research as well as the process used to anonymise their sample for patient protection (**Figure 1**). This allows the subsequent investigation of genes related to drug response and underlying disease state through the association of genetic variations with observed phenotypes.

Polymorphic genetic markers
Several types of genetic polymorphisms can be found throughout the human genome, including single nucleotide polymorphisms (SNPs), insertions and deletions of one or more bases and variable numbers of tandem repeats (VNTRs). Some genetic polymorphisms have functional consequences, while others are benign on their own but still useful from a genetic standpoint, as they serve as 'markers' for particular regions of a chromosome. It appears that SNPs are the most common polymorphisms and, due to their common frequency and binary nature which enable high-throughput analysis, they have predominated in large scale pharmacogenomic studies. SNP discovery used to be a long and arduous task, but several new technologies, companies and collaborations have recently made SNPs much more accessible.

In 1999 The SNP Consortium Ltd (TSC) was formed as a non-profit foundation with the mission of providing public genomic data in the form of SNPs distributed evenly throughout the human genome and to make the information related to

Pharmacogenomics
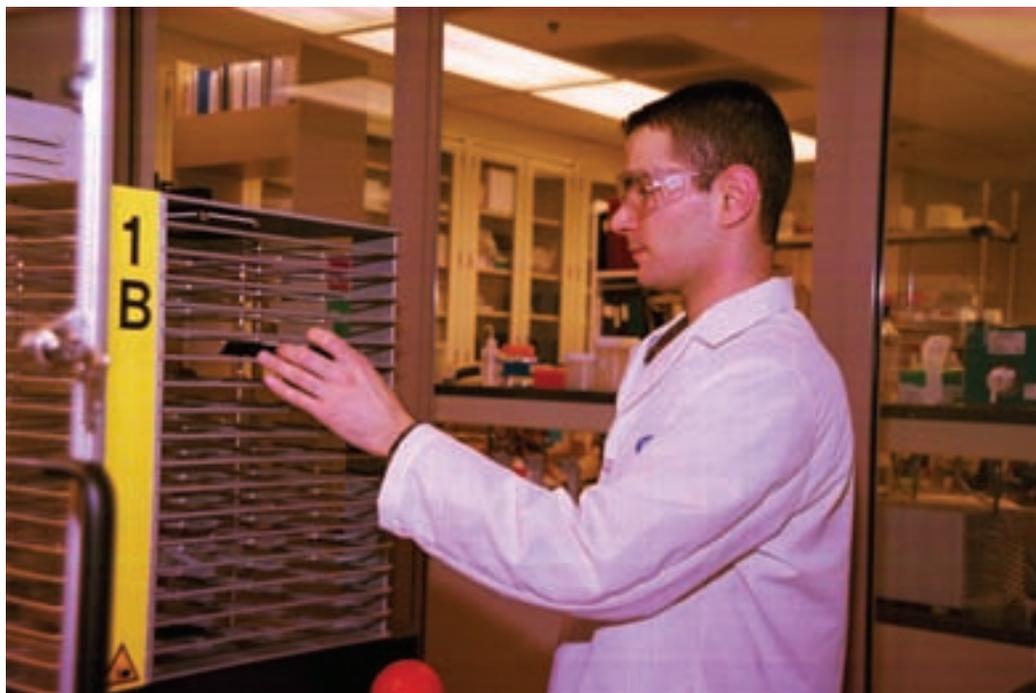
these SNPs available to the public without intellectual property restrictions (http://snp.cshl.org). The project started in April 1999 and was expected to complete delivery of 300,000 SNPs by the end of 2001. During this period the TSC delivered 1.5 million SNPs into the public domain for academic and industry scientists throughout the world to use. Complementing this genome-wide coverage of SNPs, several biotechnology companies have also applied their technologies to discover novel SNPs throughout the genome or to provide deep coverage of SNP discovery within specific candidate genes. Examples of these include Celera, Perlegen and Genaissance, respectively.

It is estimated that there are between 3 and 10 million SNPs within the human genome, with an average spacing of one SNP every 500 to 1,000 base pairs. Lack of knowledge of these SNPs once represented a major limitation to advances in pharmacogenomics. Now, web-based access to comprehensive SNP information is routine and there are few regions in the genome where SNP data are lacking. Current limitations are the lack of data relating to SNP frequency, population distribution and inheritance patterns and the challenge of assay development for large numbers of SNPs.

Genotyping technologies

With the hurdle of SNP discovery disappearing, the next challenge is the ability to perform sufficient genotyping to address specific questions related to patient disease state and therapeutic response. Technological advances in genotyping for SNPs have resulted in rapid, gel-free scoring systems that rely on either fluorescence-based or mass-based scoring of a particular SNP (**Figure 2**). These advances have driven the cost of genotyping far below the $1 mark and costs continue to fall as the possibility of multiplexing becomes a reality. In addition, many of these technologies enable pooling of large populations of subjects (100-500), thus further increasing data generation while decreasing overall cost.

The design of a pharmacogenomics study is a major factor in deciding which genotyping technology is the most appropriate. The majority of genotyping platforms are acceptable for analysis of one or a few genes, interrogating data sets of moderate size. For example, investigation of a candidate gene potentially involved in obesity may involve an association study design of 500 cases and 500 controls and the interrogation of 10 SNPs across the candidate gene locus. Thus, the required 10,000 genotypes could realistically be delivered in most genetics laboratories within a reasonable period of time. However, to interrogate multiple SNPs in 50 candidate genes more sophisticated methods for sample handling, PCR preparation and genotyping are required and capacity quickly becomes an issue for many human genetics laboratories. For whole genome analysis the scale increases dramatically. A 3,000 patient Phase

III study without an *a priori* hypothesis related to drug response could well require 300,000 or more genotypes across the genome for each patient! Several companies, such as Perlegen and Sequenom, are developing whole genome methodologies and genotyping platforms capable of handling studies of this magnitude. Each uses proprietary technology although their ability to interrogate the whole genome for statistically relevant genetic associations with phenotypes has yet to be demonstrated. While polymorphism discovery and genotyping used to be rate-limiting, the amount of data that can now be generated has pushed the bottleneck further downstream to data management, analysis and translation into disease relevant information.

Informatic platforms

The ability to integrate disparate data sources into an organised data repository remains a significant challenge for pharmacogenomics research. Consider the volume of data collected in clinical trials; the anonymisation of the data to remove certain patient identifiers (**Figure 1**) and the requirement of a separate database to retain this information; the generation of massive numbers of genotypes associated with clinical data; the statistical tools needed to establish relationships between genotype and phenotype; the mining of the genotype-phenotype data in the hypothesis-generating stage and the desire to replicate initial findings; and one quickly understands the challenges to delivery of informatic solutions. The pharmaceutical industry has developed many data repository platforms to enable collection and audit of clinical data, so we stand poised to address information platforms for such analysis. A number of biotech companies have sprung up with the goal of providing an integrated platform for such diverse data sources.

Statistical methodologies and data analysis

Technological advances in SNP discovery, genotyping and informatics platforms to manage vast amounts of information have provided the genetics community with the tools to better understand how genetic variation contributes to disease susceptibility and therapeutic efficacy and safety. However, with these advances comes the need to better understand the patterns of SNP distributions and environmental factors and the novel quantitative clinical measurements that are used to phenotype subjects in a clinical trial setting. Advances in statistical analysis of these complex datasets are currently being tested and implemented.

The statistical analysis required for pharmacogenomic studies can be broken down into two categories: population genetics and genotype/phenotype analyses. The extent of each depends on the study design.

A study may test for a direct association between a known functional SNP and a particular phenotype. Alternatively, the study design may depend on linkage disequilibrium (LD) – a measure of relatedness between neighbouring SNPs that can be used to detect association, indirectly, to an unknown functional variant. Pharmacogenomic studies routinely rely on a case/control design using unrelated subjects thus requiring statistical algorithms to quantitate LD and estimate haplotypes (SNPs along a common chromosome of a single parental origin) in the absence of parental data. This information can then be used to select SNPs for study in both a candidate gene or a high-density genome scan study design. In addition, several groups are providing very important clues to the extent of LD across the genome, the influences of genetic variation and how this information can be applied to improve the design of studies for pharmacogenomics and human genetics in general[8-9].

The success of each pharmacogenomic study depends on how much of the variability in phenotype is caused by variation within the candidate gene(s) (referred to as the effect size), how well characterised the phenotype is (quantitative vs subjective measurements) and the overall power of the study as determined by sample size. The complexity of the diseases that the pharmaceutical industry studies means that specific statistical analysis methods and study design may be required to provide the best chance of detecting a genetic signal in the presence of confounding effects such as environmental factors, variation in phenotypic measurement and population specific effects such as admixture. This is particularly the case when the pharmacogenomics study entails multiple biologically relevant candidate genes – with multiple SNPs within each gene – being tested for association with a complex phenotype which may be the result of several genes interacting with each other and with environmental factors. Thus statistical approaches to the analysis of these multivariate data sources, such as regression and recursive portioning methodologies, are currently being employed.

In addition, replication of the finding in a prospective study may be required to provide the confidence needed to assess the impact of pharmacogenomics in that particular disease or response variable.

References

1 Buchanan, A, Califano, A, Kahn, J, McPherson, E, Robertson, J, Brody, B. Pharmacogenetics: ethical issues and policy options. Kennedy Institute of Ethics Journal 2002;12:1-15.
2 Nussbaum, RL, McInnes, RR, Willard, HF. Genetic variation in populations. Thompson and Thompson Genetics in Medicine. 6th Edition. Philadelphia: W.B. Saunders Company. 2001; pp95-109.
3 Bertilsson, L. Geographical/ Interracial differences in polymorphic drug oxidation – current state of knowledge of cytochromes P450 (CYP) 2D6 and 2C19. Clin Pharmacokinet 1995;29:192-209.
4 Kroemer, HK and Eichelbaum, M. Molecular basis and clinical consequences of genetic cytochrome P450 2D6 polymorphism. Life Sciences 1995;56:2285-98.
5 Chen, S, Chou, W-H, Blouin, RA, Mao, Z, Humphries, LL, Meek, QC et al. The cytochrome P450 2D6 (CYP2D6) enzyme polymorphism: screening costs and influence on clinical outcomes in psychiatry. Clin Pharm Ther 1996;60:522-34.
6 McElroy, SM, Sachse, C, Brockmoller, J, Richmond, JL, Lira, ME, Friedman, DL, Roots, I, Silber, BM, Milos, PM. CYP 2D6 genotyping as an alternative to phenotyping for determination of metabolic status in a clinical trial setting. AAPS PharmSci 2000;2(4) article 33.
7 Mancinelli, L, Cronin, M, Sadee, W. Pharmacogenomics: the promise of personalized medicine. AAPS PharmSci. 2000;2(1):E4.

8 Patil, N, Berno, AJ, Hinds, DA, Barrett, WA, Doshi, JM, Hacker, CR, Kautzer, CR, Lee, DH, Marjoribanks, C, McDonough, DP, Nguyen, BT, Norris, MC, Sheehan, JB, Shen, N, Stern, D, Stokowski, RP, Thomas, DJ, Trulson, MO, Vyas, KR, Frazer, KA, Fodor, SP, Cox, DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science,. 2001;294:1719-23.
Reich, DE, Schaffner, SF, Daly, MJ, McVean G, Mullikin, JC, Higgins, JM, Richter, DJ, Lander ES, Altshuler, D. Human genomic sequence variation and the influence of gene history, mutation and recombination. Nature Genetics 2002;32:135-142

Education within the pharmaceutical industry

Finally, there is a need to provide a comprehensive understanding of the science of human genetics and the use of this information for the development of pharmacogenomic approaches. Through internal seminar series for clinicians, regulatory staff, discovery scientists and management, together with external education of physicians and patients, we hope to develop a finer appreciation for the fruits emerging from the human genome project. This education will continue to provide insight into potential research initiatives, as well as clinical trial design and interpretation.

Summary

The concept that genetic background can contribute to interindividual differences in disease susceptibility and genetic variation and how it influences variation in response to medicines has led to significant interest in pharmacogenomics within the pharmaceutical industry and biotechnology world. Combined with proteomics, expression profiling, bioinformatics and animal models, it provides a unique opportunity to potentially influence decision-making in the discovery and development of new medicines to better understand the relationship of targets and human disease, improve clinical trial design and select optimal doses for medicines. The field of pharmacogenomics has evolved dramatically over the last five years and will continue to do so. We have become adept at identifying polymorphic genetic markers and using genotyping assays to detect them in large, well-characterised patient populations. We have begun to develop bioinformatic tools, data management systems and statistical approaches to handle the wealth of data coming out of the genomics revolution. We are only just beginning to understand what impact pharmacogenomics studies will have on the future of drug discovery and development. Success will need to be measured by scientific and clinical researchers, physicians, business leaders and the population at large. **DDW**

Aidan Power *is Worldwide Head of Clinical Pharmacogenomics for Pfizer Global Research and Development based in New London, CT. He received his medical degree from University College Cork in Ireland, an MSc from University College London, UK and has post-graduate qualifications in Psychiatry.*

Suzin Webb *received a BS in Genetics and Developmental Bio from Cornell University and her MS degree in Biology and Medicine from Brown University. She has used molecular genetic diversity to address questions in Australian prehistory forensics, conifer genomic evolution, marker-assisted breeding and pharmacogenomics.*

Dr Albert Seymour *is a Senior Research Scientist in Discovery Pharmacogenomics at Pfizer Global Research and Development. He received a MS from The Johns Hopkins University in Molecular Biology and subsequent to that received his PhD from The University of Pittsburgh in human genetics.*

Dr Patrice Milos *currently oversees the Discovery Pharmacogenomics, Clinical Biochemical Measurements and DNA Sequencing Core for Pfizer Global Research and Development in Groton, CT. Her previous position with Pfizer involved molecular biology research in the Atherosclerosis disease area. She received her PhD from Rensselaer Polytechnic Institute and completed post-doctoral fellowships are Harvard and Brown Universities.*