CMLS, Cell. Mol. Life Sci. 60 (2003) 1636–1646
1420-682X/03/081636-11
DOI 10.1007/s00018-003-2369-4

CMLS Cellular and Molecular Life Sciences

Review

Pharmacogenetics and disease genetics of complex diseases

V. D. Schmith[a,*], D. A. Campbell[b], S. Sehgal[a], W. H. Anderson[a], D. K. Burns[a], L. T. Middleton[b] and A. D. Roses[a]

[a] Genetics Research, 5 Moore Drive, GlaxoSmithKline, PO Box 13398, RTP, North Carolina 27709 (USA), Fax: +1 919 483 6380, email: ginny.d.schmith@gsk.com
[b] Genetics Research, GlaxoSmithKline, Greenford, Middlesex (United Kingdom)

Received 19 December 2002; received after revision 14 February 2003; accepted 20 February 2003

Abstract. Advances in technologies and the availability of a single nucleotide polymorphism (SNP) map are beginning to show the true potential for the human genome project to affect patient healthcare. A whole genome scan, the use of 100,000–300,000 SNPs across the genome, is now possible. Use of traditional approaches and the whole genome scan will result in identification of disease susceptibility genes and development of many new treatments in the longer term. In the shorter term, the goal will be to predict those patients at risk to experience an adverse reaction or those with a high probability for improved efficacy (i.e. pharmacogenetics). As progress is made in the area of disease genetics and pharmacogenetics, our understanding of disease susceptibility and its interrelationship with drug response will improve, making targeted therapy (i.e. the right drug to the right patient) a reality.

Key words. Pharmacogenetics; disease susceptibility; SNP; bioinformatics.

Introduction

Rapid advances in technologies (e.g. genetic and genomic technologies as well as bioinformatics) coupled with a unique collaboration between industry and academia (the SNP Consortium) are beginning to show the true potential for the human genome project to affect patient healthcare. By knowing the sequence of the human genome and beginning to unravel the location and sequence of all genes and their variants, scientists can establish a better understanding of the mechanisms for diseases, with subsequent availability of new treatments. This process will take a significant investment and many years (estimated at 10–15 years from identification of a disease susceptibility gene to availability of a new medicine) to come to fruition. The shorter-term impact of the Human Genome Project and the SNP Consortium will be identification of the genetic basis for differential drug response (i.e. pharmacogenetics). An excellent overview of these developments and their impact on the development of new treatments has been provided previously [1–3].

Terminology in this rapidly growing field has not always been consistent across publications, with the terms 'pharmacogenetics' and 'pharmacogenomics' being used interchangeably. For this paper, the definitions described by the Pharmacogenetics Working Group (PWG, a group with international representation from the pharmaceutical industry) [4] will be used. The PWG defines pharmacogenetics as the study of DNA sequence variation as it relates to differential drug response. The PWG defines pharmacogenomics as a broader term, with no absolute description, as various writers use it in different ways. However, in its broadest sense PWG defines pharmacogenomics as the study of the genome and its products (including RNA and protein) as they relate to drug discovery and development. This review will discuss pharmacogenetics and only briefly mention pharmacogenomics.

The goal of pharmacogenetics is to improve our understanding of the genetic basis for differential drug response and ultimately to have the ability to identify indi-

* Corresponding author.

viduals at risk for an adverse event or those individuals with a high probability for enhanced efficacy, prior to treatment (i.e. targeted drug therapy). In turn, as our understanding of the pharmacogenetic factors affecting drug response increases, knowledge of the disease itself also increases, leading to better characterization (and possible subclassification) of the actual disease. Conversely, increased characterization of the disease can lead to better understanding of the genetic factors that may affect drug response. This important interplay between pharmacogenetics and disease genetics may lead to improved treatment alternatives and improved selection of particular drugs in selected patients (i.e. the right medicine for the right patient).

This paper will cover reviews of disease genetics and pharmacogenetics. Initially, a summary of important background definitions and technology advances will be summarized.

Background

The DNA sequence of any two people is some 99.9% identical, with the 0.1% variability in DNA likely affecting an individual's risk for disease and/or differential drug response. The most common type of genetic variation is a single nucleotide polymorphism (SNP), where the DNA sequence among individuals differs by a single DNA base. Sets of nearby SNPs on the same chromosome are inherited in blocks or clusters, called a haplotype. These blocks may contain a large number of SNPs, but a haplotype may be uniquely identified using relatively few SNPs. Other genetic variants that may also be important predictors of drug response or disease susceptibility include insertions (the insertion of one or more base pairs) deletions (deletion of one or more base pairs) and a variable number of tandom repeats (VNTRs, sometimes called microsatellites). A VNTR or microsatellite that consists of a repeat of two bases (e.g. TA_n) is called a dinucleotide repeat.

Results from the human genome project made it evident that there was a need for a standardized SNP map. Therefore, the SNP Consortium (referred to hereafter as TSC), a large-scale public effort was undertaken in 1999 to complete a standardized SNP map. TSC was a collaboration of 10 large pharmaceutical companies, 3 technology companies, 5 academic centers and the Wellcome Trust (*http://snp.cshl.org/*). TSC was an extraordinary success as a public-private consortium, with the identification and mapping of more than 2 million SNPs in less than 2 years. Furthermore, the identification and location of these SNPs were deposited in the public domain, available to all investigators at no cost. Additional projects, such as the determination of allele frequencies in multiple ethnic populations and quality control experiments, have continued past the March 2001 completion of the initial mapping objectives. In addition, TSC is collaborating with the National Institutes of Health (NIH) in the United States in the creation of a haplotype-based SNP map (referred to as the HapMap).

Technology advances

With the unprecedented output from TSC, we have seen a major drive within the biotechnology industry to develop cheaper, faster and more accurate genotyping technologies that will allow application of the data from the human genome project. In the past, most disease genetics studies used 400–500 dinucleotide repeat or microsatellite markers spread over the genome at 10 cM intervals (~10 million base pairs depending where the markers are in the genome) in a labor-intensive, time-consuming, expensive gel-based or sequencing methodology. Most pharmacogenetic studies have used the candidate gene approach, in which variants in specific genes related to the metabolism of the drug, the drug target and any associated biological pathways are evaluated. Several new genotyping technologies are now available, making it possible to look at 100,000–300,000 SNPs evenly spread (and/or as haplotypes) across the genome (referred to as the whole genome scan approach) over a very short period of time. These technologies have been described in detail by others [5,6] and include (but are not limited to): solid-phase microsequencing or SBE (e.g. Affymetrix chip technology), oligonucleotide ligation assays, allele-specific hybridization (e.g. TaqMan) and allele-specific discrimination by DNA polymerase (e.g. apyrase-mediated allele specific extension, AMASE). The key driver now is to dramatically reduce the cost and time of genotyping. Although there are numerous companies working toward this aim, table 1 gives a breakdown of a number of the key players in this field.

Bioinformatics tools

Because of the vast amount of data coming out of the Human Genome Project and TSC, bioinformatics tools and databases have become an integral part of pharmacogenetic and disease susceptibility gene research. They play an important role in candidate gene identification, gene finding, SNP detection, genotyping and genetic analysis. Public sources of databases and tools abound, although it is sometimes difficult to determine the quality, consistency and sustainability of these sources. An overview of the history along with current state of affairs of bioinformatics tools and resources is provided by Searls [7]. Several resources available through public web sites are also described in appendix 1.

Table 1. Companies leading the development of more efficient genotyping technologies.

Company	Platform	Technology	Website
Affymetrix	Genechip arrays	DNA probes on glass chips	www.affymetrix.com
Applied Biosystems	7900 HT, 7700	Taqman	www.appliedbiosystems.com
Illumina	BeadArray Scanner	BeadArray	www.illumina.com
Orchid	SNPstream UHT	SNP-IT taq array	www.orchidbio.com
Perlegen Sciences	based on Affymetrix chip	high-density DNA probes on glass wafers	www.perlegen.com
Pyrosequencing	PSQHS 96A	Pyrosequencing	www.pyrosequencing.com
Sequenom	MassARRAY	MALDI-TOF mass spec	www.sequenom.com
Third Wave	Invader OS	Cleavase	www.twt.com

Bioinformatics data integration and tool standardization are critical to the success of association and linkage studies. A generic model for such an infrastructure is given in figure 1. The underlying data models accommodate the variability inherent in subject collections, the ability to trace the data source, and the automation and archival storage of analysis results. A fully traceable data source is important, as we are often faced with anomalies in data at a late stage that can be very time consuming to resolve in an infrastructure that does not facilitate data integration. The polymorphism database component includes data from public and proprietary sources. The subject phenotypes (a relevant measure of disease severity, disease progression and/or disease subclassification for disease genetics or a relevant measure of drug response for pharmacogenetics) and genotype components are fully integrated with the source databases. The subject database component also includes reference collections and allele frequency information needed for analysis. This model has proven useful in analyzing reasonably large datasets. The model is scalable to variations in volumes and expandable to accommodate a variety of markers. The performance for very high volumes (e.g. genome-wide scans of a large population) is currently being investigated.

SNPs are the most common markers for disease-gene and drug-response associations. However, to detect association at a SNP near a complex disease gene, the appropriate SNPs must be chosen for analysis. In addition, the order and relationship of SNP markers is extremely important. The cost of doing high-density genome-wide association scans is still quite high, so, using a haplotype-based SNP map would maximize the information content and reduce the resource needs. The use of haplotypes has been discussed in great detail, including their benefits and limitations [8–10]. One limitation of haplotypes that needs to be considered is the fact that frequencies of most clinically significant AEs are low (< 5–10%) so the use of commonly occurring haplotypes (those with frequencies of at least 10%) may overlook important genetic associations [9].

Another approach that has been advocated to reduce the cost of genotyping is DNA pooling. Instead of analyzing SNPs from individual subjects, DNA from responders is pooled and compared with pooled DNA from control subjects. The advantages and disadvantages of this approach are reviewed in detail elsewhere [11].

Disease genetics and pharmacogenetics

Genotypic data can be combined with accurate phenotypic data and analyzed to determine the SNPs and/or haplotypes associated with disease susceptibility and/or drug response. A high-density genome association scan can be used to thoroughly evaluate the genes that modify a patient's response to medications (i.e. pharmacogenetics) and to push the limits of disease gene identification in appropriate populations (i.e. disease genetics). Examples of the use of the candidate gene approach and/or the whole genome scan approach are described below as they relate to disease genetics and pharmacogenetics.

Disease genetics

In the past, disease genetics has focussed on monogenic diseases such as Huntington's disease in which the expression of a particular variant of a single gene will, in the vast majority of cases, lead to disease. There are innumerable monogenic diseases, each of which affect only a small number of patients. In contrast, disease genetics research is now focussed on identification of genes associated with common diseases (diseases affecting thousands or millions of people). These common diseases are multifactorial [i.e. dependent on complex interactions between numerous environmental factors and a number of alternative forms (alleles) of genes called disease susceptibility genes] and polygenic (involving more than one gene in their multifactorial pathogenesis) [12]. The overall goal of disease genetics is to identify how genetic variation can influence disease susceptibility and to improve our understanding of the molecular processes resulting in

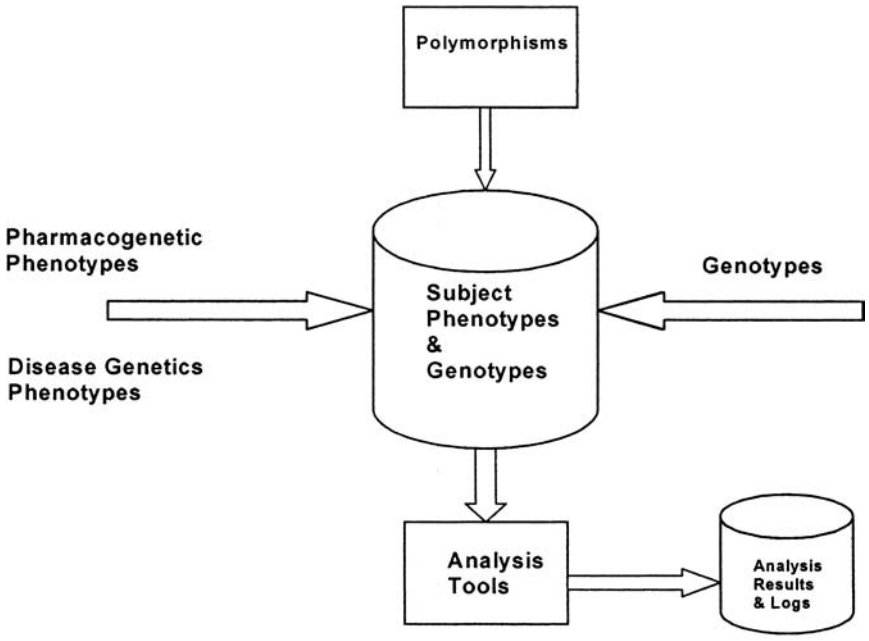


Figure 1. Bioinformatics data integration.

clinically overt disease. New treatments can then be designed to target these molecular processes to prevent and/or treat the disease.

Typically, new disease susceptibility genes have been identified using a combination of linkage and association studies. The linkage studies involve collection of DNA samples and extensive clinical phenotypic data from multiple members of affected families. Markers are typed throughout the genome, and, using linkage analysis algorithms, chromosomal regions harboring disease genes are identified. The regions are identified using highly informative markers on the basis of their chromosomal location by taking advantage of the meiotic process of recombination as apparent in families segregating for the disease. Markers closest to the disease gene show the strongest correlation with disease patterns in families. These linkage studies allow identification of a region on a chromosome and large portions (1–20 cM) of the DNA (which may include 10–1000 genes) that may be linked to a specific disease. Candidate genes within the region can sometimes be inferred from the genome-wide databases that are currently available (see appendix 1). Unfortunately, most of the few validated disease genes were not obvious candidates. Association studies are then conducted to identify the causative mutation responsible for the disease either using family-based association studies or unrelated case-control association studies. The key to success for linkage and association studies is the availability of high-quality clinical information, available appropriate genotypic data and the ability to link such data (see above).

Linkage and/or association studies have been reported to identify suceptibility genes for many therapeutic areas. These have been extensively reviewed for heart failure [13], neurodegenerative diseases [14], epilepsy [15], diabetes mellitis [16], obesity [17, 18], schizophrenia [19, 20], osteoarthritis [21], COPD [22, 23] and rheumatoid arthritis [24, 25].

Most of these studies have used the more traditional dinucleotide repeat approach. However, a proof-of-concept study has shown that high-density SNP mapping can be successfully applied in whole genome scan studies for linkage and association [26]. This experiment confirmed the previous discovery of APOE4, the susceptibility gene variant responsible for common, late-onset Alzheimer disease. The APOE4 variant of the APOE gene had been widely confirmed to increase the risk and lower the mean age of onset of Alzheimer disease. A high-density SNP map was constructed across a 4-Mb region encompassing the APOE gene on chromosome 19, with an average SNP distance of 15 Kb. A small region of highly significant linkage disequilibrium (LD) was clearly identified, confirming the already established chromosomal region and providing proof of concept for the identification of susceptibility variants using high-density SNP mapping. The LD region contained only two genes, APOE and APOC1 [26, 27]. The conclusion was that if these methods had been available in 1988, APOE might have been identified in less than a year after the original linkage was found [1–3].

The high-density SNP mapping approach has since been used to identify susceptibility genes in regions of linkages for several other diseases, including migraine with aura [28], psoriasis [29] and schizophrenia [30].

The ability to identify and study susceptibility genes for common diseases is now considerably expanded. Using

the whole genome scan approach as a tool, it is technically possible and feasible to study multiple genetic factors for common diseases simultaneously, eliminating the need for knowledge about molecular mechanisms. This is a significant improvement over the previous historic approach of searching for highly penetrant single-point mutations associated with uncommon or rare diseases.

Pharmacogenetics

The goal of pharmacogenetics is to predict the genetic basis for differential drug response. Genetics is one of many factors that may affect drug response. The probability of experiencing an adverse drug reaction or altered efficacy can depend on the concentration of drug at the site of action [based on absorption, distribution, metabolism and elimination (ADME), dose, compliance, drug-drug or food-drug interactions], the receptors or mediators involved in response (e. g. drug target, related pathways, endogenous mediators) and/or other factors (e. g. environment). Investigators have known for many years that there were genetic differences in the ability of individuals to metabolize certain drugs. For some drug-metabolizing enzymes (DMEs) like cytochrome P450 (CYP) CYP2D6, CYP2C19, CYP2C9 and *N*-acetyltransferase 2 (NAT-2), some patients have little or no enzyme, resulting in higher concentrations of drugs predominantly metabolized by these pathways. This may potentially lead to a higher probability of experiencing pharmacologically predictable adverse events or greater response. Over the past 10 years, the genetic basis for these differences in metabolism have been identified and reviewed extensively [31–39].

Genetic variants in specific drug targets have been identified that may also contribute to variability in drug response. Some examples include variants in the ryanodine receptor (RYR1), which have been implicated in the development of malignant hyperthermia following suxamethonium or halogenated inhalational anesthesia [40, 41]; variants in the ALOX-5 gene associated with differential response to leukotriene antagonists [42]; and the Ser9Gly polymorphism in the dopamine D3 receptor, associated with an increased risk for drug-induced tardive dyskinesias [43–45].

Until recently, the focus of pharmacogenetics has been to identify variants in genes related to the metabolism of the compound, the drug target and any associated pathways using the candidate gene approach. A new approach, referred to as a whole genome scan approach, allows for the characterization of the set of SNPs and/or haplotypes (i. e. the SNP Print or SNP profile) across the genome in individual patients who suffer adverse events (AEs) or to those who exhibit efficacy, regardless of the mechanism for the event. The use of the whole genome scan in this context provides an unbiased approach to identifying 'at risk' subjects and an alternative approach that is not dependent on a molecular understanding of response. In addition, this approach accounts for the fact that the genetic basis for adverse events and/or increased response is likely to involve multiple genes. The use of a whole genome scan can, therefore, lead to a more comprehensive assessment of the genetic basis for an adverse event or enhanced response [1–3].

Whether using the whole genome scan or the candidate gene approach, a sensitive and specific test can be developed to predict who would have a high probability of experiencing an AE or enhanced efficacy. In the future, it may be possible to make this test available prior to exposing the patient to this therapeutic agent. A regulatory perspective on the use of pharmacogenetic tests has recently been published [46].

Safety

The occurrence of adverse drug reactions is a significant risk in the development and subsequent clinical use of virtually all medicines. AE pharmacogenetics can add value to the evaluation of AE in several ways [3]. First, during early drug development, the addition of pharmacogenetics to clinical trials can improve the understanding of pharmacologically predictable AEs and the subsequent risk/benefit ratio. This understanding can lead to development of a medicine response test and can potentially reduce compound attrition. Second, pharmacogenetics can add value by explaining less common AEs observed during later-phase clinical development and post-marketing. The less common AEs uncovered during later-phase drug development and post-marketing tend to be idiosyncratic reactions; reactions in which the mechanisms are not fully understood. The use of pharmacogenetics in these two contexts (pharmacologically predictable, idiosyncratic) is described in more detail below.

Pharmacologically predictable AEs are typically dose dependent. Therefore, an increase in drug exposure due to low or absent levels of drug-metabolizing enzymes could lead to increases in the incidence of a specific AE. Phillips et al. [47] recently reported that of the 27 drugs frequently cited in adverse drug reaction studies, 69% were metabolized by at least one enzyme known to have genetic variants resulting in a functional change of the enzyme. Genetic variants in drug-metabolizing enzymes that have resulted in enzyme deficiencies and drug toxicity have been reviewed in general elsewhere [31, 34]. Specific examples have been reported in oncology [48], cardiovascular disease [49] and in inflammatory bowel disease [50], some of which are used in clinical practice.

Other than pharmacokinetic differences, it is likely that genetic variability in the drug target, related pathways or

endogenous mediators may contribute to pharmacologically predictable adverse drug reactions. An excellent example of this type of pharmacogenetic effect was reported by several investigators [51–53]. These investigators reported an increased likelihood of experiencing drug-induced long QT when subjects had specific variants in genes for potassium or sodium channels.

Idiosyncratic drug reactions are typically less common and less well understood than pharmacologically predictable adverse drug reactions. With the availability of a whole genome scan approach, our understanding of the contribution of genetics to these AEs is likely to increase substantially. Two examples of idiosyncratic adverse event pharmacogenetics are described in detail.

Hetherington et al. [54] described how pharmacogenetics can play a role in understanding patient response to abacavir. Abacavir is a nucleoside reverse transcriptase inhibitor for use in combination therapy for human immunodeficiency virus (HIV) infection. It is generally well tolerated in most patients. In clinical trials, ~4% of patients receiving abacavir developed a hypersensitivity reaction (HSR) to the drug. This HSR typically occurs within the first 6 weeks of treatment and resolves upon discontinuation of abacavir. A risk management program for abacavir has effectively reduced the morbidity in clinical practice, but the situation lent itself to a proof of principle clinical trial to test the utility of pharmacogenetics.

In summary, over 100 variants from 12 candidate gene families, as well as three human leucocyte antigen (HLA) loci, were chosen for analysis. Statistical analyses of data from patients with HSR and control patients (patients who received abacavir but did not experience HSR) indicated that alleles from two genes [tumor necrosis factor α (TNFα) and HLA-B] showed highly significant association with HSR. Independent investigators confirmed implication of the same genomic region [55]. The fact that both genes are involved with immune response is consistent with the symptoms of the AE. However, the fact that they were ~200 Kb apart on the same chromosome opened the question of an underlying haplotype effect and the possibility of neither one being the true causal variant.

Lai et al. [9] describe how the data were analyzed to evaluate whether the risk-associated alleles were in linkage disequilibrium (LD) with each other or contributing independently to risk of HSR. The analysis showed that HLA-B57 and TNFα-238A were in high LD with each other and that the intervening region between the SNPs can be divided into approximately three haplotype bins/blocks. This experiment showed that association is not necessarily limited to small regions, a result that cautions against inferring causality from association results that define a small chromosomal segment. In addition, while complex disease theory suggests that large numbers of patients are required for gene discovery, at least this one pharmacogenetic effect is of sufficient magnitude to be detectable with a very modest sample size (as little as 18 cases and 18 controls) [9].

Overall, the abacavir study provides a proof of principle that pharmacogenetics can play a role in understanding patient response to medicines. Multiple markers within the HLA-B region were associated with hypersensitivity reactions. However, the high false-negative rates do not accurately predict individuals who will develop a hypersensitivity reaction. It is, therefore, premature for any recommendation to be made regarding the use of a test for hypersensivitiy to abacavir in a clinical setting [54]. The whole genome scan approach is ongoing and may lead to a more comprehensive assessment of the genetic basis for this AE [3].

Roses [3] described how safety pharmacogenetics could be applied to Tranilast. Tranilast is a drug that was in phase III clinical trials for the treatment of restenosis after percutaneous transluminal coronary revascularisation. During the course of the double-blind, placebo-controlled trial involving 11,500 patients, ~4% of patients in the trial developed hyperbilirubinemia. In these patients, bilirubin levels were transiently increased significantly with only variably mild elevations of other liver function enzymes, and no clinically overt signs of hepatic disease. DNA from subjects who had developed hyperbilirubinemia following Tranilast administration and control patients (i.e. those who received Tranilast but did not develop hyperbilirubinemia) were studied to address the question, 'Did patients with hyperbilirubinemia following Tranilast administration have a genetic predisposition to this event?' If so, a test could potentially be developed that would predict which patients a at risk. This question was addressed while the phase III study was being completed so the data could be included in a regulatory package to address the risks/benefits of the compound [3].

Whole genome SNP mapping was still unavailable in 2001, so variants from multiple candidate genes were chosen for screening patients who developed hyperbilirubinemia in earlier clinical trials compared with matched control patients in the trials who did not develop hyperbilirubinemia. The UDP-glucoronosyltransferase-1 (UGT1) gene had been included as one of the candidate genes associated with hyperbilirubinemia. Mutations in the UGT1A1 gene had been shown to be associated with Gilbert disease, a spontaneously occurring benign form of hyperbilirubinemia. A VNTR [$(TA)_n$ where $N=6$ or 7] in the TATA box of the UGT1A1 gene has been reported; patients homozygous for the 7 repeats (i.e. genotype of 7,7) have a higher likelihood of Gilbert syndrome [56].

Figure 2 shows the proportion of patients with normal or increased total bilirubin following Tranilast administration by genotype. Eighty percent of the patients with increased total bilirubin were homozygotes for the seven repeats. A small number of patients homozygotes for the

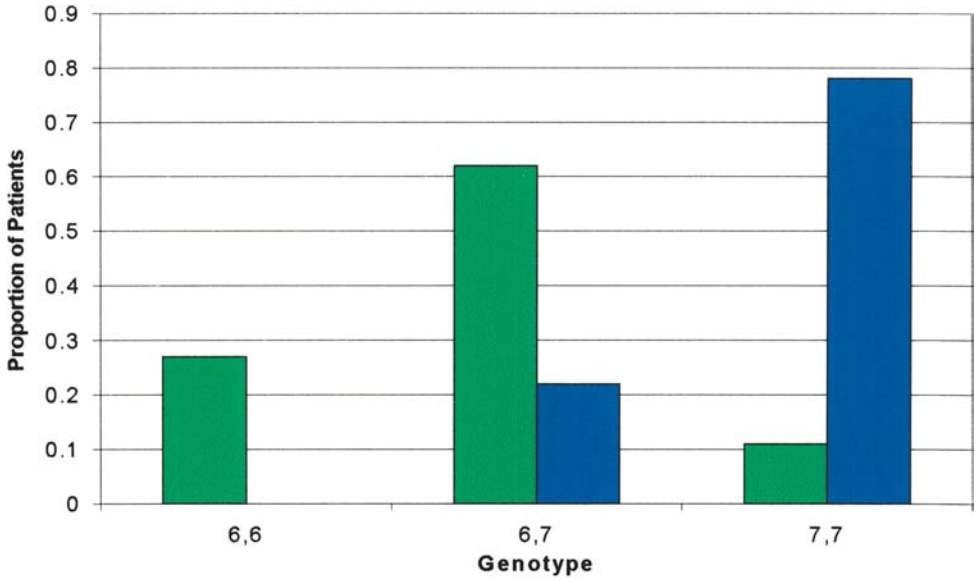


Figure 2. Proportion of patients with hyperbilirubinemia following Tranilast administration by genotype. Blue bars represent patients with hyperbilirubinemia following Tranilast administration; green bars represent patients with normal bilirubin following Tranilast administration.

seven repeats did not develop hyperbilirubinemia following Tranilast. This is consistent with a 14% incidence of the 7,7 genotype in patients with no evidence of Gilbert syndrome. These data provided the opportunity to predict which patients were susceptible to Tranilast-induced hyperbilirubinemia. Having such a test that could predict who would develop hyperbilirubinemia would be a great asset to the company and patients. By further determining that Tranilast could induce Gilbert syndrome in identifiable susceptible patients, it was possible to define a benign condition associated with hyperbilirubinemia, rather than patients susceptible to more severe hepatic complications [3].

From a pharmacogenetic perspective and under the time limits required, this experiment was able to show that pharmacogenetics can change the perceived risk of the drug. Unfortunately, Tranilast was not pursued in this indication, unrelated to this bilirubin issue. However, this experiment clearly demonstrated the power of pharmacogenetics screening to rapidly identify susceptibility to a well-defined AE arising during clinical development [3].

Efficacy

Genetic variants in the drug target, related pathways or drug-metabolizing enzymes have the potential to predict those patients with a high probability of enhanced efficacy. As discussed previously, variants in the genes for polymorphic drug metabolizing enzymes can result in some patients having little or no enzyme (i.e. poor metabolizers). For drugs predominantly metabolized by these pathways, drug concentrations are much higher in poor metabolizers. These higher concentrations can result in enhanced efficacy. For example, Furuta et al. [57] reported that the cure rates following lansoprazole for gastroesophageal reflux disease were much greater in poor metabolizers of CYP2C19 (85%) than in extensive metabolizers (46%). Similar findings have also been reported for omeprazole [58]. Genetic variants in the drug target (or related pathways) can also lead to enhanced efficacy. An example of the latter is the use of trastuzamab (Herceptin). Trastuzamab is a biologic (antibody) approved for therapeutic use only in patients whose tumors test positive for the HER-2/neu protein using the DAKO Hercep Test (a semiquantitative assay for testing overexpression of the HER-2/neu protein) or *HER2* FISH pharmDx Kit [a complete fluorescence in situ hybridization (FISH) assay that is designed to quantitatively determine interphase *HER2/neu* gene amplification] [59]. Clinical trials showed that trastuzamab significantly improved survival in patients with tumors that overexpressed HER-2/neu (~1/3 of patients), while the clinical effect may have seemed marginal if all patients were analyzed as a whole [60]. The test is now widely used in clinical practice, prior to a patient receiving trastuzamab.

Other examples of efficacy pharmacogenetics have also been reported for the treatment of asthma [42, 61, 62], hypertension [63], response to antipsychotics [64, 65], and response to immunosuppressive therapy [66]. These examples show that pharmacogenetics has the potential to identify patients with improved efficacy.

Efficacy pharmacogenetics has the potential to change the traditional paradigm of drug development, although it is not embraced by all. If pivotal studies are conducted in a subgroup of patients with improved efficacy, the number of subjects needed to have adequate power to show efficacy would be reduced, thereby improving the effi-

ciency of drug development. Discovery of novel therapeutics could potentially focus on the nonresponder population, identifying new targets and eventually new therapies [1–3].

Figure 3 summarizes the potential impact of pharmacogenetics (efficacy and safety) on the drug development process, including the advantages of this approach to patients, clinicians and the pharmaceutical industry [67].

Interrelationship between pharmacogenetics and disease genetics

As progress is made in the areas of disease genetics and pharmacogenetics, it is becoming increasingly clear that today's definition of a disease may actually be a combination of multiple subtypes of a disease with different etiologies. For most drug treatments, there is a range of response to treatments by different agents for the same disease. An understanding of the genetic factors affecting variability in response to treatment could potentially lead to an increase in the understanding of the disease and vice versa. For example, figure 4 shows the distribution of treatment responses following beclomethasone and montelukast in asthmatic patients, demonstrating a wide range of variability in response [61]. As our understanding of the factors explaining this variability in response increases, the knowledge of the disease itself (in this case, asthma) increases, leading to a better characterization (and possible subclassification) of the actual disease. The increased understanding of the disease can lead to better treatment alternatives and to a better understanding of factors that may affect drug response. Thus, the interplay between pharmacogenetics and disease genetics may ultimately lead to more appropriate selection of particular drugs in selected patients (i.e. the right drug to the right patient).

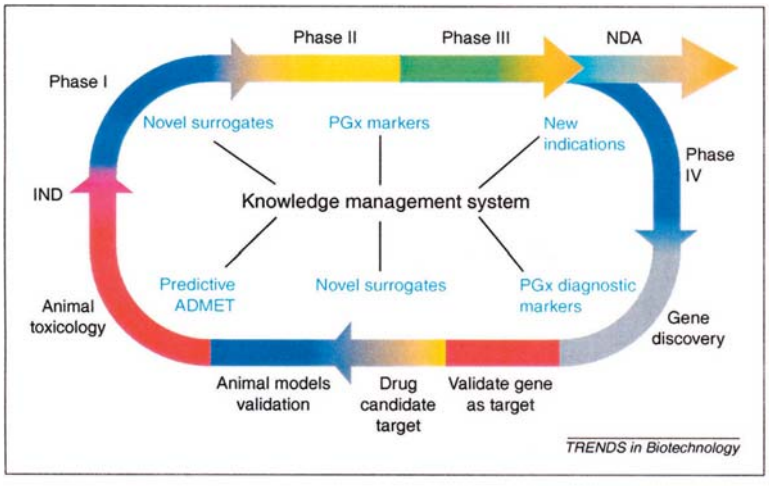
As associations are uncovered in disease genetics and pharmacogenetics, pharmacogenomic approaches may add to the knowledge and confirm the mechanisms involved in disease susceptibility or differential drug response. These pharmacogenomic approaches (e.g. RNA and DNA microarrays, proteomics) are reviewed elsewhere [68–73].

Overall summary

The potential benefits of the human genome project are beginning to be realized with the availability of technology advances and bioinformatics tools. The identification of disease susceptibility genes and the development of many new treatments are the longer-term benefits. In the shorter term, the benefits will be the ability to predict those patients at risk for experiencing adverse reactions or patients with a high probability of experiencing improved efficacy (i.e. pharmacogenetics). As progress is made in the area of disease genetics and pharmacogenetics, our understanding of disease susceptibility and its interrelationship with drug response will improve, making targeted therapy (i.e. the right drug to the right patient) a reality.

Appendix 1
Bioinformatics Tools

Useful web sites that are critical in the conduct of disease genetics and pharmacogenetics are summarized below.



Box 1. Personalized medicine: advantages to the pharmaceutical industry

Increased efficiency and reduced costs of target and lead discovery
Reduced timelines and costs of clinical trials
Emergence of new gene targets for drug discovery
Product differentiation in the market place

Box 2. Personalized medicine: advantages to patients and clinicians

Higher probability of desired outcome with a drug
Low probability of untoward side effects
Preventive strategies
Focused therapies
Reduced costs
Better health and better healthcare

Figure 3. Impact of pharmacogenetics and disease genetics on the drug discovery and development process. Genetic/genomic information and markers emerging at each stage of the discovery process will be used as tools upstream and downstream, resulting in better pharmaceuticals and personalized medicine products. A knowledge warehouse will store information, enabling continued process and product improvements. Abbreviations: ADMET (absorption, distribution, metabolism, excretion, toxicity); IND (investigational new product); NDA (new drug application); PGx (pharmacogenetics and pharmacogenomics). Reprinted from Trends in Biotechnology, Vol. 19, Ginsberg and McCarthy, Personalized medicine: revolutionizing drug discovery and patient care, Pages 491–496, Copyright (2001), with permission from Elsevier Science.

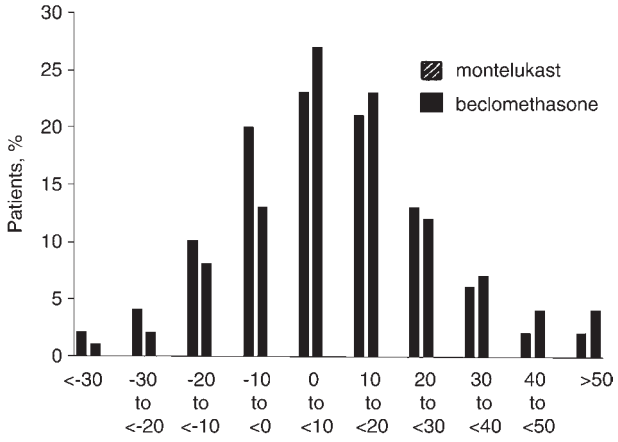


Figure 4. Distribution of treatment response for FEV_1. Response distributions (percentage of patients) are shown as histograms for predifined intervals of percent change in FEV_1. For each predefined interval, the left most bar represents patients treated with montelukast, and the right most bars represent patients treated with beclomethasone. Note wide variation in therapeutic response among patients. Reprinted with permission from the Pharmacogenomics Journal **1:** 27–37, 2001.

A notable resource, Ensemble [74] (http://www.ensembl.org), is a comprehensive source of human genome annotations with links to many useful sites. It is regularly updated, and the latest version is constructed on the National Center for Biotechnology Information's (NCBI) human build 30. As an open source project, the entire system is portable, including the Web site and analysis pipeline. This allows users to install the system to process their own genome data, a highly desirable feature given the frequent need to view public and proprietary data together. LocusLink is another database of interest to clinical genetics. LocusLink provides a single query interface to curated sequence and descriptive information about genetic loci. It links to GenBank, UniGene, PubMed, dbSNP and several other resources of biological data of interest to geneticists. dbSNP is a major public source of data on candidate SNPs from TSC, the Sanger Center and Washinton University. dbSNP is a candidate SNP database, and SNPs need to be validated and characterized before they can be turned into genetic markers. A good overview of dbSNP is provided by Marth et al. [10]. LocusLink, dbSNP and other NCBI resources are available from the NCBI Web site (http://www.ncbi.nlm.nih.gov/) and covered in detail in by Wheeler et al [75,76]. PharmGKB (http://pharmgkb.org/do/serve?id=home.welcome) is a NIH funded resource on pharmacogenetics. Genomic data, molecular and cellular phenotype data, and clinical phenotype data are accepted from the scientific community and made available to the public.

A number of bioinformatics tools for sequence analysis are available to discover functional, structural and evolutionary information in biological sequences. Several commercial and noncommercial packages include access to public sequence databases through the Internet. BLAST is the most widely used sequence-searching program and is available from the NCBI Web site. A common reason for performing a database search with a query sequence is to find a related gene in another organism with known function. If a query sequence aligns with a database sequence with known function, the query sequence is predicted to have the same function. Several types of sequence search and alignment algorithms, methods and tools are used for determining location of SNPs and genes, gene structure and gene function. An exhaustive account of biological sequence analysis is given in by Mount et al. [77].

Acknowledgements. The authors would like to acknowledge David Searls, Michael Lutz and Sreekumar Pillai for their insightful comments.

1 Roses A. D. (2000) Pharmacogenetics and the practice of medicine. Nature **405:** 857–865
2 Roses A. D. (2002) SNPs, where's the beef? Pharmacogenomics J. **2:** 277–283
3 Roses A. D. (2002) Pharmacogenetics place in modern medical science and practice. Life Sc **70:** 1471–1480
4 Ray S.C. (2001) Pharmacogenetics Working Group. Deutsche Zeitschrift fur Klinische Forschung **1:** 48–49
5 Ahmadian A. and Lundeberg J. (2002) A brief history of genetic variation analysis. Biotechniques **32:** 1122–1137
6 Brennan M. D. (2001) High throughput genotyping technologies for pharmacogenomics. Am. J. Pharmacogenomics **1:** 295–302
7 Searls D. B. (2000) Bioinformatics tools for whole genomes. Annu. Rev. Genomics Hum. Genet. **1:** 251–279
8 Stephens J. C. (1999). Single-nucleotide polymorphisms, haplotypes and their relevance to pharmacogenetics. Molecular Diagnosis **4:** 309–317
9 Lai E., Bowman C., Bansal A., Hughes A., Mosteller M. and Roses A. D. (2002) Medical applications of haplotype-based SNP maps: shouldn't we learn how to walk before trying to run? Nat. Genet. **32:** 353
10 Marth G., Yeh R., Minton M., Donaldson R., Li Q., Duan S. et al. (2001) Single-nucleotide polymorphisms in the public domain: how useful are they? Nat. Genet. **27:** 371–372
11 Chanock S. (2001) Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. Disease Markers **17:** 89–98
12 Middleton L., Freeman A., Brewster S., Foster C. and Roses A. D. (2000) From gene-specific tests to pharmacogenetics. Community Genetics **3:** 198–203
13 Komajda M., Charron P. and Tesson F. (1999) Genetic aspects of heart failure. Eur. J. Heart Failure **1:** 121–126
14 Maimone D., Dominici R. and Grimaldi L. M. E. (2001) Pharmacogenomics of neurodegenerative diseases. Eur. J. Pharmacol. **413:** 11–29
15 Gurrieri F. and Carrozzo R. (2001) The genetics of epilepsy. Am. J. Med. Genet. (Semin. Med. Genet.) **106:** 117–118
16 Bartsocas C. S. and Leslie R. D. G. (2002) Genetics of diabetes mellitus. Am. J. Med. Genet. (Semin. Med. Genet.) **115:** 1–3
17 Clement K., Boutin P. and Froguel P. (2002) Genetics of obesity. Am. J. Pharmacogenomics **2:** 177–187
18 Sharma A. M. (2002) Second international symposium on obesity and hypertension: genetics and molecular mechanisms, 25–27 October 2001, Berlin Germany. Int. J. Obesity **26:** 1283–1288

19 Joober R., Boksa P., Benkelfat C. and Rouleau G. (2002) Genetics of schizophrenia: from animal models to clinical studies. J. Psychiatr. Neurosci. **27:** 336–347
20 Riley B. P. and McGuffin P. (2000). Linkage and associated studies in schizophrenia. Am. J. Med. Genet. (Semin. Med. Genet.) **97:** 23–44
21 Simonet W. S. (2002) Genetics of primary generalized osteoarthritis. Mol. Genet. Metab. **77:** 31–34
22 Silverman E. K. (2001) Genetics of chronic obstructive pulmonary disease. Novartis Found. Symp. **234:** 45–58; discussion 58–64
23 Lomas D. A. and Silverman E. K. (2001). The genetics of chronic obstructive pulmonary disease. Respiratory Res. **2:** 20–26
24 Buckner J. H. and Nepom G. T. (2002) Genetics of rheumatoid arthritis: is there a scientific explanation for the human leukocyte antigen association? Curr. Opin. Rheumatol. **14:** 254–259
25 Corr M. and Firestein G. S. (2002) The genetics of the target tissue in rheumatoid arthritis. Rheumatic Diseases Clinics of North America **28:** 79–94
26 Lai E., Riley J., Purvis I. and Roses A. D. (1998) A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics **54:** 31–38
27 Martin E. R., Lai E. H., Gilbert J. R., Rogala A. R., Afshari A. J., Riley J. et al. (2000) SNPing away at complex diseases: analysis of single nucleotide polymorphisms around APOE in Alzheimer disease. Am. J. Hum. Genet. **67:** 383–394
28 McCarthy L. C., Hosford D. A., Riley J. H., Bird M. I., White N. J., Hewett D. R. et al. (2002) Single nucleotide polymorphism alleles in the insulin receptor gene are associated with typical migraine. Genomics **79:** 271
29 Hewett D., Samuelsson L., Polding J., Enlund F., Cantone K., See C. G. et al. (2002) Identification of a psoriasis susceptibility candidate gene by linkage disequilibrium mapping with a localised single nucleotide polymorphism map. Genomics **79:** 305–314
30 Wilcox M. A., Faraone S. V., Su J., Van Eerdewegh P. and Tsuang M. T. (2002). Genome scan of three quantitative traits in schizophrenia pedigrees. Biol. Psychiatr. **52:** 847–854
31 Evans W. E. and Relling M. V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. Science **286:** 487–491
32 Broder S. and Venter J. C. (2000) Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. Annu. Rev. Pharmacol. Toxicol. **40:** 97–132
33 Tribut O., Lessard Y., Reymann J. M., Allain H. and Bentue-Ferrer D. (2002) Pharmacogenomics. Med. Sci. Monit. **8:** RA152–163
34 Bertz R. J. and Granneman G. R. (1997) Use if in vitro and in vivo data to estimate the likelihood of metabolic drug interactions. Clin. Pharmacokinet. **32:** 210–258
35 Wormhoudt L. W., Commandeur J. N. M. and Vermeulen N. P. E. (1999) Genetic polymorphisms of human N-acetyltransferase, cytochrome P450, glutathione-S-transferase and epoxide hydrolase enzymes: relevance to xenobiotic metabolism and toxicity. Crit. Rev. Toxicol. **29:** 59–124
36 Goldstein J. A. (2001) Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. Br. J. Clin. Pharmacol. **52:** 349–355
37 Hayes J. D. and Strange R. C. (2000) Glutathione S-transferase polymorphisms and their biological consequences. Pharmacology **61:** 154–166
38. Mackenzie P .I., Miners J. O. and McKinnon R. A. (2000) Polymorphisms in UDP glucuronosyltransferase genes: functional consequences and clinical relevance. Clin. Chem. Lab. Med. **38:** 889–892
39 deWildt S. N., Kearns G. L., Leeder J. S. and van den Anker J. N. (1999) Glucuronidation in humans: pharmacogenetic and developmental aspects. Clin. Pharmacokin. **36:** 439–452
40 Gronert G. A., Mott J. and Lee J. (1988) Aetiology of malignant hyperthermia. Br. J. Anaesth. **60:** 253–627
41 MacLennan D. H., Otsu K., Fujii J., Zorzato F., Phillips M. S., O'Brien P. J. et al. (1992) The role of the skeletal muscle ryanodine receptor gene in malignant hyperthermia. Symp. Soc. Exp. Biol. **46:** 189–201
42 Drazen J. M., Yandava C., Dube L. M., Szxerback N., Hippensteel R., Pillari A. et al. (1999) Pharmacogenetic association between ALOX5 promotor genotype and the response to anti-asthma treatment. Nat. Genet. **22:** 168–170
43 Steen V. M., Løvlie R., MacEvan T. and McCreadie R. G. (1997) Dopamine D3-receptor gene variant and susceptibility to tardive dyskinesia in schizophrenia. Mol. Psychiatry. **2:** 139–145
44 Basile V. S., Masellis M., Badri F., Paterson A. D,. Meltzer H. Y., Lieberman J. A. et al. (1999) Association of the MscI polymorphisms of type dopamine D3 receptor gene with tardive dyskinesia in schizophrenia. Neuropsychopharmacology **21:** 17–27
45 Segman R., Neeman T., Heresco-Levy U., Finkel B., Karagichev L., Schlafman M. et al. (1999) Genotype association between the dopamine D3 receptor and tardive dyskinesia in chronic schizophrenia. Mol. Psychiatry. **4:** 247–253
46 Lesko L. J. and Woodcock J. (2002) Pharmacogenomic-guided drug development: regulatory perspective. Pharmacogenomics J. **2:** 20–24
47 Phillips K. A., Veenstra D. L., Oren E., Lee J. K. and Sadee W. (2001) Potential role of pharmacogenomics in reducing adverse drug reactions: a systemic review. JAMA **286:** 2270–2279
48 Innocenti F., Iyer L. and Ratain M. J. (2000) Pharmacogenetics: a tool for individualizing antineoplastic therapy. Clin. Pharmacokinet. **39:** 315–325
49 Wuttke H., Rau T., Heide R., Bergmann K., Böhm M., Weil J. et al. (2002) Increased frequency of cytochrome P450 2D6 poor metabolizers among patients with metoprolol-associated adverse events. Clin. Pharmacol. Ther. **72:** 429–437
50 Ansari A., Hassan C., Duley J., Maranaki A., Shobowale-Bakre E.M., Seed P. et al. (2002) Thiopurine methyltransferase activity and the use of azothioprine in inflammatory bowel disease. Aliment. Pharmacol. Ther. **16:** 1743–1750
51 Sesti F., Abbott G. W., Wei J., Murray K. T., Skasena S., Schwartz P. J. et al. (2000) A common polymorphism associated with antibiotic-induced cardiac arrhythmia. PNAS **97:** 10613–10618
52 Roden D. M. (2001 Pharmacogenetics and drug-induced arrythmias. Cardiovasc. Res. **50:** 224–231
53 Escande D. (2000) Pharmacogenetics of cardiac K^+ channels. Eur. J. Pharmacol. **410:** 281–287
54 Hetherington S., Hughes A. R., Mosteller M., Shortino D., Baker K. L., Spreen W. et al. (2002) Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet **359:** 1121–1122
55 Mallal S., Nolan D., Witt C., Masel A. M., Martin A. M., Moore C. et al. (2002) Association between the presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV reverse-transcriptase inhibitor abacavir. Lancet **359:** 727–732
56 Bosma, P. J., Chowdhury J. R., Bakker C., Gantla S., de Boer A., Oostra B. A. et al. (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N. Engl. J. Med. **333:** 1171–1175
57 Furuta T., Shirai N., Watanabe F., Honda S., Takeuchi K., Iida T. et al. (2002) Effect of cytochrome P4502C19 genotypic differences on cure rates for gastroesophageal reflux disease by lansoprazole. Clin. Pharmacol. Ther. **72:** 453–460
58 Kita T., Tanigawara Y., Aoyama N., Hohda T., Saijoh Y., Komada F. et al. (2001) CYP2C19 genotype related effect of omeprazole on intragastric pH and antimicrobial stability. Pharmaceut. Res. **18:** 615–621
59 Wang S., Saboorian M. H., Frenkel E., Hynan L., Gokaslan S. T. and Ashfaq R. (2000) Laboratory assessment of the status of

Her-2/neu protein and oncogene in breast cancer specimens: comparison of immunohistochemistry assay with fluorescence in situ hybridisation assays. J. Clin. Pathol. **53:** 374–381
60 Baselga J. (2000) Current and planned clinical trails with trastuzumab. Semin. Oncol. **27(5 Suppl. 9):** 27–32
61 Silverman E. S., Liggett S. B., Gelfand E. W., Rosenwasser L. J., Baron R. M., Bolk S. et al. (2001) The pharmacogenetics of asthma: a candidate gene approach. Pharmacogenomics J. **1:** 27–37
62 Hall I. P. (2000) Pharmacogenetics of asthma. Eur. Respir. J. **15:** 449–451
63 Bianchi G. (2001) Pharmacogenomics and genetics of primary hypertension. Curr. Hypertension Reports **3:** 441–443
64 Scharfetter J. (2001) Dopamine receptor polymorphisms and drug response in schizophrenia. Pharmacogenomics **2:** 251–261
65 Cichon S., Nöthen M. M., Rietschel M. and Propping P. (2000) Pharmacogenetics of schizophrenia. Am. J. Med. Genet. **97:** 98–106
66 Danesi R., Mosca M., Boggi U., Mosca F. and Del Tacca M. (2000) Genetics of drug response to immunosuppressive treatment and prospects for personalized therapy. Mol. Med. Today **6:** 475–482
67 Ginsberg G. S. and McCarthy J. J. (2001) Personalized medicine: revolutionizing drug discovery and patient care. Trends Biotechnol. **19:** 491–496
68 Quinn-Senger K. E., Ramachandran R., Rininger J. A., Kelly K. M. and Lewin D. A. (2002) Staking out novelty on the genomic frontier. Curr. Opin. Chem. Biol. **6:** 418–426
69 Pociot F. and Karlsen A. E. (2002) Combined genome and proteome approach to identify new susceptibility genes. Am. J. Med. Genet. (Semin. Med. Genet.) **115:** 55–60
70 Zanders E. D. (2000) Gene expression analysis as an aid to the identification of drug targets. Pharmacogenomics **1:** 375–384
71 Harris S. and Foord S. M. (2000) Transgenic gene knock-outs: functional genomics and therapeutic target selection. Pharmacogenomics **1:** 433–443
72 Debouck C. and Metcalf B. (2000) The impact of genomics on drug discovery. Ann. Rev. Pharmacol. Toxicol. **40:** 193–207
73 Debouck C. and Goodfellow P. N. (1999) DNA microarrays in drug discovery and development. Nat. Genet. **21(1 Suppl.):** 48–50
74 Hubbard T., Barker D., Birney E, Cameron G., Chen Y., Clark L. et al. (2002) The Ensembl genome database project. Nucleic Acids Res. **30:** 38–41
75 Wheeler D. L., Church D. M., Lash A. E., Leipe D. D., Madden T. L., Pontius J. U. et al (2002) Database resources of the National Center for Biotechnology Information: 2002 update. Nucleic Acids Res. **30:** 13–16
76 Wheeler D. L., Church D. M., Lash A. E., Leipe D. D., Madden T. L., Pontius J. U. et al. (2001) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **29:** 11–16
77 Mount D. W. (2001) Bioinformatics: Sequence and Genome Analysis, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY