



Published in final edited form as:

Per Med. 2009 November 1; 6(6): 643–652. doi:10.2217/pme.09.55.

Pharmacogenetics and personal genomes

Michael J Wagner

Institute for Pharmacogenomics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599–27361, USA

Abstract

While pharmacogenetics - the correlation of genotype and response to medicines - currently has a small but measurable impact on the prescribing practice of clinicians, the advent of the 'personal genome' is likely to change this significantly. Advances in high-throughput technologies aimed at characterizing human genetic variation, including chip-based genotyping and next-generation sequencing, are poised to provide a flood of information that will affect both pharmacogenetic discovery and pharmacogenetic application in clinical practice. In order for this flood of information to not overwhelm both researchers and clinicians alike, a variety of new and expanded information management tools will be needed, including electronic medical records, bioinformatic algorithms for analyzing sequence data, information management systems for storing, retrieving and interpreting whole-genome sequence data, and pharmacogenetic decision tools for prescribers.

Keywords

genome sequence; genome-wide association study; next-generation sequencing; personalized medicine; pharmacogenetics; rare variant; SNP

The practice of modern medicine employs a broad array of drugs to cure acute conditions, such as infections, and to manage chronic conditions, such as diabetes and heart disease. To be approved for use for a given indication, a drug must be demonstrated to be both safe and effective. However, this determination is based on aggregate measures, and individual response to drugs can vary substantially. Some patients may show no response, some may have the desired therapeutic response, while others may have adverse responses ranging from the merely annoying to life threatening. This variation in response may be explained in part by diagnostic uncertainty, environmental factors such as diet and interacting drugs, and by clinical factors such as age and comorbid conditions. However, genetics has also been demonstrated to play a major role in patient–patient variation in drug response. Individual genetic variation in enzymes of drug metabolism and transport may lead to inadequate therapeutic responses owing to the inability to absorb a drug, the inability to activate a prodrug, or excessive metabolism and/or excretion of an active drug. Alternatively, the form of the drug target encoded in an individual's genome may be insensitive to the drug [1]. Similarly, an individual may experience adverse reactions to a drug if their genome encodes defective enzymes of drug metabolism, resulting in an abnormally high exposure to the drug despite normal dosing. Pharmacogenetics is the study of these effects of the 'personal genome' on drug response.

© 2009 Future Medicine Ltd

Tel.: +1 919 966 9590 Fax: +1 919 966 5863 michael_wagner@unc.edu.

Financial & competing interests disclosure The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

The field of pharmacogenetics had its beginnings in the 1950s, with discoveries demonstrating that individual variation in the levels of activity of enzymes involved in the metabolism of a number of drugs correlated with adverse reactions to those drugs [2]. The advent of molecular cloning and sequencing technologies enabled the identification of sequence variants in the genes for these enzymes that are responsible for the variation in activity, and catalogs detailing large numbers of these variants are now available online [101]. Chip-based genotyping technologies have now made practical the testing of panels of genetic variants in large numbers of individuals, effectively making available a limited but focused view of the 'personal genomes' of those individuals. Next-generation sequencing technologies are now coming into widespread use and offer the prospect of complete, genome-wide information on an individual basis, but raise new issues relating to the interpretation of that information in relation to pharmacogenetics. This rapidly changing landscape has implications for both the discovery of genetic variants with pharmacogenetic effects and the application into clinical practice.

Cataloging variation in the human genome: a prerequisite for pharmacogenetic studies

Characterizing the extent of genetic variation in populations of interest and understanding its distribution across the genome is a prerequisite for any genetic study, and this is certainly true for pharmacogenetic studies. Discovery of most types of genetic variation depends on DNA sequencing. Discovery can be focused on a single gene or a limited number of genes by using molecular cloning or PCR to isolate those genes individually from the DNA of multiple subjects. Comparison of the DNA sequences derived from different subjects allows for the identification of sequence variants, which are positions in the genome where the DNA sequence differs between individuals. Alternatively, an automated pipeline for random, 'shotgun' sequencing can identify variants distributed across the entire genome, as was done for the HapMap project, which identified and validated more than 3.1 million SNPs [3,4]. The dbSNP database [102] currently lists more than 6.5 million validated DNA sequence variants in humans. The large majority of these are common variants, defined as having minor allele frequencies (MAF) greater than 5%, but this is still only a fraction of the 15–20 million sequence variants estimated to occur in human populations, approximately half of which are expected to be uncommon variants (MAF < 5%) [3]. To identify 99% of both common and uncommon variants (MAF > 0.1%) in the Caucasian population would require the sequencing of approximately 1000 subjects [5], and a publicly funded project to obtain complete genome sequences for this many subjects is now underway [103].

For the last 30 years, the sequencing methodology of choice has been the Sanger method [6]. Constant improvements in sequencing chemistry, gel-based sequencing equipment, automation and software have resulted in the dramatic increases in sequencing throughput, culminating in the determination of two versions of a consensus sequence for the human genome [7,8]. These sequences now serve as the reference human genomes against which other sequences can be compared for the identification of variants. However, the use of this technology base for the determination of the complete sequences of multiple, individual human genomes remains impractical owing to issues of cost and capacity. Novel technologies for sequencing using highly multiplexed arrays of single DNA molecules or single molecule-derived clusters of DNA (so-called next-generation sequencing) have emerged in the last few years and represent a transformational advance for DNA sequencing, making the prospect of individual, 'personal genome sequences' conceivable [9]. Several complete human genome sequences have been published using these technologies, and have identified a treasure trove of novel genetic variation. For example, complete genome sequences of a Caucasian, three Asians and an African individual yielded approximately 600,000, 420,000–590,000, and 1,000,000 novel variants, respectively [10–14]. The 1000 Genomes Project [103], currently underway, is expected to catalog the vast majority of the remaining common genetic variation

in the human genome. Given the means to comprehensively catalog human genetic variation, the challenge for the field of pharmacogenetics is now twofold: first, to discover which of the known or novel variants in the human genome are responsible for differences in individual responses to drugs, and second, to apply these discoveries to clinical practice to allow for the prescription of the safest and most effective drug for the individual patient.

Pharmacogenetic discovery

Patient cohorts

Demonstrating the pharmacogenetic effects of DNA variants requires the collection of a patient cohort with documented medication and outcomes data as well as available DNA samples for testing. Clinical trials and observational studies of patients in specialty clinics often provide the best opportunity to collect the requisite data and samples. However, the number of subjects with the phenotype of interest that can be collected in this way is often small, severely limiting the power to detect pharmacogenetic effects. This is particularly true when the phenotype is uncommon, as is often the case for adverse events. Postmarketing surveillance would seem to be a logical source of patients experiencing adverse events. However, the current system in the USA, the US FDA MedWatch system, with its reliance on voluntary submissions of adverse event reports, is inadequate to provide the high quality patient data required in sufficient numbers. Large epidemiological cohorts, such as the Framingham study in the USA [15] or birth cohorts collected in a number of European countries, may provide a useful number of subjects, particularly when these studies are linked to electronic medical and prescribing records. Health maintenance organizations (HMOs) and pharmacy benefits managers (PBMs) have large patient databases with electronic records, and a number of projects exploiting these resources are already underway. For example, the Marshfield Clinic's personalized medicine research project combines a biobank of DNA, plasma and serum samples from 20,000 consented individuals with access to the Clinic's medical records [16]. A number of pharmacogenetic studies based on this resource have been completed or are underway, including studies of statins, metformin, warfarin and tamoxifen. Many of these studies were undertaken in collaboration with academic researchers, demonstrating that academic/HMO partnerships are feasible and can be very productive. Interest in pharmacogenetics by PBMs is exemplified by Medco's (NJ, USA) project on the pharmacogenetic dosing of warfarin, where patients initiating warfarin therapy are recruited in real-time from Medco's covered population [17]. The National Human Genome Research Institute (MD, USA) has organized the eMERGE Network, a consortium of academic and private health-care systems, designed to 'develop, disseminate and apply approaches to research that combine DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research' [104]. The success of these efforts demonstrates that the recruitment of subjects based on electronic medical and prescribing records represents an important source of patient cohorts for pharmacogenetic discovery in the future.

Testing for pharmacogenetic associations: the candidate gene approach

Once a cohort of patients has been identified, a choice must be made as to which variants to genotype, how many variants to genotype and the appropriate genotyping technology platform to be used. The traditional approach of testing variants one at a time with individual genotyping assays may severely limit the number of variants that can be typed due to considerations of cost, time and the amount of DNA available. It is then crucial to choose variants with high *a priori* probability of involvement in the phenotype to be tested. For this reason, pharmacogenetic discovery has, to date, largely relied on the so-called 'candidate gene' approach. Candidate genes for a particular drug are selected for evaluation based on pre-existing knowledge relating to the drug target or to pathways of drug metabolism. A typical candidate gene study usually involves one to a few genes. Variants in the candidate genes may

be selected from databases or from published reports, based on their presumed or established functional effects, or simply be distributed across the genes at some prespecified density. Typically, one to a few variants are tested per gene. If sufficient variants are not available, the candidate genes may be resequenced in a limited number of subjects (either randomly selected controls or patients with the drug-response phenotype of interest) to identify genetic variants, which may subsequently be genotyped in larger collections of patients with measured drug-response phenotypes. A significant association of a drug-response phenotype with specific alleles or genotypes for a given genetic variant is taken as evidence of an effect of the gene containing that variant on drug response.

The candidate gene approach can be very successful when the correct candidates are chosen and genetic variation within those candidates has been adequately surveyed. However, the candidate gene approach will fail when the state of knowledge about metabolic pathways or drug targets or target pathways for the drug under study is inadequate to allow the correct choice of candidates for testing. For example, selective serotonin reuptake inhibitors are widely used in the treatment of depression, and show widely varying efficacy and side-effect profiles in individual patients. They have also been the subject of many pharmacogenetic studies focusing on the drug target (serotonin reuptake transporter), downstream target pathways (serotonin receptors) and genes involved in drug metabolism (*CYP2D6* and *CYP2C19*). However, results of most of these studies have been either negative or inconsistent [18], suggesting that genes with significant pharmacogenetic effects must be sought outside of the primary pharmacokinetic and pharmacodynamic pathways. The candidate gene approach may also fail when the state of knowledge concerning genetic variation in the candidate genes is inadequate. Variant discovery may have been carried out in an inadequate number of subjects to successfully identify all relevant variants, particularly those occurring at a low frequency in the discovery population, or may not have covered critical regions of the gene, such as regulatory elements in introns or promoter regions. Patterns of genetic variation differ between racial groups [19,20], and variants may have been described in subjects from a different racial group than those being tested for pharmacogenetic effects.

Genome-wide association studies

Recent advancements in our understanding of the patterns of genetic variation across the human genome, provided in large part by the HapMap project, as well as advances in highly multiplexed genotyping technologies, have made possible an alternative to the candidate gene approach: the genome-wide association study (GWAS). No knowledge of candidate genes is required; the GWAS approach only assumes that genetic polymorphisms affecting the trait of interest exist somewhere in the genome, and conducts an unbiased search across the entire genome to find them. This is done by utilizing chip-based technologies to simultaneously (and cost-effectively) genotype 500,000–1 million SNPs (the most common type of genetic variant in the human genome) approximately evenly distributed across the genome. Even this enormous number of variants is only a small fraction of the total number of variants known or predicted to exist in the human population. However, genotypes for common SNPs in close proximity to each other in the genome often tend to be highly correlated, due to the coinheritance through human evolutionary history of the ancestral genotype combinations present on small chromosomal segments, a phenomenon known as linkage disequilibrium. Linkage disequilibrium between the SNPs tested on the chips and nearby, untested SNPs allows a GWAS to capture the majority of genetic information at a much larger number of SNPs, with the tested SNPs serving as proxies, or 'tags' for the information contained in the untested SNPs. Patterns of linkage disequilibrium differ between racial groups and estimates of the proportion of common variants for which the current generation of GWAS chips will yield more than 80% of the total available information range from nearly 90% in Caucasians to less than 70% in individuals of African descent [4].

A GWAS will largely test for only the effects of common polymorphisms (MAF > 0.05). Because the power to detect association decreases as the MAF decreases, the vast majority of SNPs included on the commercially available genotyping chips are common SNPs. Furthermore, of all the untested variants in the genome, only common polymorphisms will show substantial linkage disequilibrium with the common SNPs tested on the chip. The large number of association tests performed in a GWAS requires the use of stringent levels of statistical significance (p-values of 10^{-7} or better) to avoid false-positive associations, and consequently the power to detect associations at these genome-wide levels of significance for variants with modest effects on phenotype may be quite low unless large numbers of subjects are studied. Regions of segmental duplication in the human genome are typically under-represented on GWAS genotyping chips, and many genes of potential pharmacogenetic interest, such as cytochrome P450 gene families, tend to fall in these regions and may therefore be inadequately covered in a GWAS. Despite these caveats, hundreds of GWAS studies have been conducted in the last few years and have successfully identified scores of genes contributing to the risk of many common human diseases and conditions, such as diabetes, heart disease, asthma and obesity [105]. Many of the genes identified would not have been considered candidates for the trait in question, demonstrating the value of the unbiased, genome-wide approach for making novel discoveries.

The GWAS era in pharmacogenetics is just beginning, with only a small number of published studies to date. Investigation of variation in therapeutic response to drugs include studies of iloperidone treatment of schizophrenia [21], thiazide diuretic treatment of hypertension [22], anti-TNF treatment in rheumatoid arthritis [23] and warfarin dosing [24,25]. The investigation of adverse events includes studies of the elevation of serum alanine aminotransferase by ximelagatran [26], QT interval prolongation by iloperidone [27], bisphosphonate-related osteonecrosis of the jaw [28], statin-induced myopathy [29] and drug-induced liver injury due to flucloxacillin [30]. As in the common disease studies, many of the pharmacogenetics associations detected are in genes that would not necessarily have been considered candidates for the effect studied, providing novel insights into biology. The odds ratios for the variants detected in pharmacogenetic GWAS have mostly been relatively high (>2), in contrast to the variants found in common disease GWAS, where the odds ratios are usually quite low (<1.5). However, this is not surprising in that most of the cohorts used in the pharmacogenetic studies (several hundred to a thousand subjects) are quite small by GWAS standards, and therefore were not sufficiently powered to detect variants with low odds ratios [31]. It is to be expected that expanding the size of pharmacogenetics cohorts, perhaps by exploiting large-scale epidemiological cohorts or HMO or pharmacy databases, will allow the detection of additional variants contributing to drug response phenotypes at lower odds ratios.

An intermediate between the traditional candidate gene approach and GWAS is the use of chip-based technology to genotype panels of variants from large numbers of potential candidate genes. For example, the commercially available DMET™ plus chip (Affymetrix, CA, USA) assays nearly 2000 variants in 225 genes thought to play a prominent role in adsorption, distribution, metabolism and excretion (ADME) for a wide range of drugs [32]. By assaying a number of variants several orders of magnitude smaller than a GWAS chip, the statistical penalty for multiple testing is reduced, increasing the power to detect pharmacogenetic effects. More alleles of each gene are assayed, over a broader range of MAF, providing a more comprehensive view of the effects of these genes. DMET chips have been used in the discovery of the pharmacogenetics effect of polymorphisms in the *CYP4F2* gene on warfarin dose [33], and polymorphisms in the *CYP2C19* gene on clopidogrel efficacy [34]. It is likely that the use of this and other custom chip designs will have a prominent place in pharmacogenetics studies in the future.

Common versus rare variants in pharmacogenetics

Given the fact that both the candidate gene approach and GWAS typically focus primarily on common variants, it is fair to ask if rare variants should also be investigated for pharmacogenetics effects. In studies of susceptibility to common, complex diseases, the effect sizes of common polymorphisms are often small and account for a small percentage of total variation in the phenotype [31,35]. Rare variants of high penetrance are also likely to contribute to phenotypes of interest, and multiple rare variants in the many genes involved in pathways affecting a given phenotype may, in aggregate, account for a larger fraction of the total variation in the phenotype than common polymorphisms [36]. For example, Fearnhead and colleagues identified 13 potentially deleterious, rare variants in five candidate genes involved in Wnt signaling and mismatch repair from a series of patients with multiple colorectal adenomas and a series of random controls [37]. A variant was present in 24.9% of patients compared with only 12% of controls. The difference was highly significant [37]. Similarly, Sandilands and colleagues showed that multiple null alleles of the filaggrin gene, including uncommon polymorphisms (MAF ~ 1%) and rare variants, were present in 47% of a series of Irish patients with atopic dermatitis, compared with approximately 8% of controls [38]. In a series of papers, Cohen and colleagues identified rare, nonsynonymous variants in six candidate genes related to cholesterol metabolism by comprehensive sequencing of the exons of these genes in a series of patient samples, and showed higher frequencies of these variants in individuals at the extremes of the trait distributions [39–42], suggesting a causal role. Ji and colleagues screened 3000 patients from the Framingham Heart Study offspring cohort for rare variants in three genes known to be involved in recessive diseases exhibiting extremely low blood pressure. They found 30 rare variants of known or inferred functional effect in the heterozygous state in 49 subjects. These subjects had an average systolic blood pressure of 6.3 mmHg lower than that of the entire cohort, and a reduced prevalence of hypertension [43].

In pharmacogenetics research, in contrast to disease susceptibility research, there are many examples of common variants with high penetrance that account for a substantial portion of the variation in response to certain drugs in human populations. However, there are also many examples of uncommon to rare variants that have similar values of penetrance. For example, the *CYP2D6* gene, responsible for the metabolism of 25% of commonly used drugs, has been extensively studied over the years, and over 75 variants have been described [101]. Most are low frequency variants in any given population (though frequencies vary geographically), and novel variants continue to be discovered as larger and more diverse groups of subjects are examined. Not all of these *CYP2D6* variants have been characterized functionally, but a substantial fraction have been demonstrated to have little or no enzyme activity and would be expected to have an impact on phenotypes related to *CYP2D6* activity. Despite this evidence of allelic diversity, most candidate gene studies will choose to genotype only the few variants known to be common in the population being studied for reasons of cost and convenience, and GWAS will miss the effects of all of the rare variants due to the fact that they will not be adequately 'tagged' by any of the common SNPs on the genotyping chip. Thus, in a discovery setting, part of the effect of *CYP2D6* on a pharmacogenetic phenotype in a population will be missed, and in a clinical setting, individual patients carrying the rare, but untyped, variants will be given incorrect information about their *CYP2D6* status and their consequent risk for drug reactions. The situation could be improved by using a multiplexed genotyping assay to allow cost-effective coverage of a larger range of genetic variation. For example, the DMET Plus chip assays 31 variants in the *CYP2D6* gene, a substantial improvement over the handful of common variants typically tested by individual assays. However, novel variants will still be missed, and for some applications, multiplexed assays with the required content will not be available.

Resequencing to detect both common & rare variants in pharmacogenetic studies

The best way to insure the comprehensive identification and genotyping of all variants of a gene, both common and rare, is to resequence that particular gene in each subject. For example, complete sequencing of the *BRCA1* and *BRCA2* genes is routinely employed in a diagnostic setting to identify variants conferring a risk for breast cancer in subjects with a family history of the disease [44]. Traditional Sanger sequencing is adequate for resequencing of one or two genes, but if large panels of genes must be resequenced for pharmacogenetic discovery or pharmacogenetic risk prediction, this approach becomes impractical.

A variety of next-generation sequencing technologies are now commercially available [9]. All rely on sequential fluorescent imaging of base incorporations in large scale arrays of single DNA molecules or single molecule-derived clusters of DNA, though the enzymology and chemistry of cluster generation and base incorporation differ between technologies. All these technologies currently produce shorter individual sequences with higher per base error rates than traditional Sanger sequencing, but rely on the enormous number of individual sequences generated by the massively parallel nature of the approach to produce highly-redundant coverage of each base, resulting in an accurate consensus sequence and accurate calling of sequence variants. Next-generation sequencing technologies have recently been used to determine the essentially complete genome sequences of five individuals [10–14], and both tumor and constitutional genome sequences of a sixth individual [45]. However, despite the enormous increases in throughput and reductions in cost per consensus base pair of these approaches, applying whole-genome sequencing to the large number of individuals needed to identify rare variants and demonstrate associations of those variants with phenotypes of interest is still a large project, and is still prohibitively expensive. Inexpensive, complete, 'personal genomes' may be on the horizon, but that horizon is still a few years away.

An alternative to whole-genome sequencing is to apply next-generation technology to targeted resequencing of specific genes. Two main approaches have been employed for targeting specific regions of the genome for sequencing: hybridization-based sequence capture [46–49] and multiplexed PCR amplification [50–53]. Each of these approaches has advantages and disadvantages. Both are still in the early stages of development, and it is likely that improvements in one or more of these methods in the near future will allow for the targeting of larger regions with improved specificity and uniformity of coverage, and at reasonable cost. Targeted resequencing using next-generation technology can easily produce megabases of sequence covering extensive lists of candidate genes in a cost-effective and timely manner, providing a comprehensive but focused view of the portion of an individual genome where variants believed to have high *a priori* probabilities of affecting a phenotype of interest may be found. Alternatively, a gene-targeted but genome-wide approach can be taken by targeting the entire exome (the protein coding sequences for all genes, comprising approximately 1–2% of the genome) [54].

In a recent paper, the exome was extracted from the complete genome sequence of Craig Venter (one of a handful of complete individual genome sequences currently in the public domain) and analyzed for potentially functional variants [55]. Venter's exome contained approximately 10,400 nonsynonymous single nucleotide variants ([SNVs] i.e., single nucleotide changes that alter the amino acid encoded at a particular position in a protein), and about 15–20% of these SNVs are rare in the human population. Bioinformatic analysis of the amino acid changes produced by these variants indicated that about 1500 are likely to be functional. Only seven nonsynonymous SNPs had annotations in the literature linking them to disease states, and all of these had low odds ratios. No obvious correlations of any of these SNPs with the phenotype of Dr Venter could be inferred. This study points out some of the difficulties of interpretation that will be encountered in analyzing large-scale individual sequence data. In any whole-genome sequence, whole-exome sequence, or even moderately large set of candidate gene

sequences, a large number of variants will be found. Most of the common variants and essentially all of the rare variants will have no annotation in the literature linking them with disease states or response to medicines. Functional consequences can be assumed for SNVs that result in protein truncation, or indels (insertion/deletion polymorphisms) that result in frameshifts, or deletions that remove large portions of a gene. Bioinformatic analysis can infer function for some of the nonsynonymous SNVs [56,57] and an even smaller portion of regulatory variants. However, many variants will remain uncharacterized as to function, unless high-throughput systems for functional screening can be devised. Even if a variant can be demonstrated to have an effect on protein function, it remains to be shown whether this results in an effect on phenotype. Moderately common to common variants discovered through sequencing can be tested for association with a phenotype in a suitable patient cohort, as has traditionally been done in both candidate gene studies and GWAS. For rare variants, this approach will have insufficient power even in large cohorts, and new approaches must be employed that will aggregate multiple rare variants of similar functional consequences for analysis. For example, Cohen *et al.* [39] showed that multiple nonsynonymous variants in candidate genes for high-density lipoprotein (HDL) cholesterol levels, taken together, were enriched in patients at one extreme of the population distribution of HDL levels, suggesting that variants that affect the function of these genes also affect HDL levels. Focusing resequencing efforts on subjects collected from the extremes of the population distribution of pharmacogenetic traits may prove to be a powerful approach to pharmacogenetic discovery.

Application of pharmacogenetics in clinical practice

Testing for individual genetic variants with known effects on drug response is already available to clinicians and is being used to guide therapy to a limited extent. For example, testing for common variants of *CYP2C9* and *VKORC1*, which affect warfarin dose requirements, is available from many clinical laboratories. Algorithms for calculating a starting dose of warfarin based on genotype plus clinical factors have been developed [58] and are available online [106], facilitating the interpretation and use of pharmacogenetic information by clinicians. Testing for *HLA-B*5701* in patients initiating abacavir therapy has become widespread, and essentially eliminates cases of abacavir hypersensitivity [59]. Cetuximab, a monoclonal antibody directed against the EGF receptor, is ineffective in treating colorectal cancers containing somatic mutations in the *KRAS* oncogene, and *KRAS* genotyping is becoming common in selecting an appropriate chemotherapy regimen for colorectal cancer patients [60]. Many other single gene tests are available from both clinical laboratories and direct-to-consumer genetics companies, with varying degrees of proof regarding clinical utility and cost-effectiveness, and clinicians are generally left to their own devices in interpretation and incorporation of the genetic results into their prescribing practices. Improvements in the evidence base and development of widely accepted, detailed pharmacogenetic-based prescribing guidelines for clinicians are needed to further the goal of personalizing prescribing practices.

Future perspective

The field of pharmacogenetics is evolving rapidly today. While traditional candidate gene studies are still being pursued, GWAS are becoming much more common, and deep resequencing is seen as the next logical step. In the near term, as larger numbers of clinically validated pharmacogenetic effects are identified, multiplexed, chip-based assays for panels of variants are likely to become the preferred method for delivering personalized prescribing based on individual genetic variation to large numbers of patients. The AmpliChip® (Roche, Basel, Switzerland) was the first such device to gain US FDA approval for testing for polymorphisms in the *CYP2D6* and *CYP2C19* genes. However, the AmpliChip has not come into widespread use owing to its relatively high cost and limited coverage of drug metabolism

genes. The DMET Plus chip, though not an FDA approved device, is a promising development in this area, covering over 200 genes involved in drug metabolism and transport. Additional, custom chips or alternative multiplexed genotyping technologies will undoubtedly be developed for specific diagnostic needs and updated as novel pharmacogenetic variants are discovered. These types of highly multiplexed assays for individual variants provide a view into the 'pharmacogenome' of an individual, and thus represent a focused version of the 'personal genome'.

The holy grail of the genome sequencing technology world is the US\$1000 genome. As the cost of genome sequencing declines towards this level over the next few years, clinicians face the prospect of patients arriving in their offices with copies of their personal genome in hand. What will be required to translate this overwhelming amount of information into clinical decision making that takes into account the patient's inherited susceptibility to disease and their likely reactions to the drugs that may be used to treat those diseases? First, an enormous infrastructure to support the actual sequencing itself, the storage of the resulting sequences, and the ability to access those sequences, will have to be developed. A single run of a next-generation sequencing instrument produces approximately one terabyte of raw data, which must be transferred from the sequencer to temporary storage space over high-speed networks, analyzed on dedicated servers to extract sequence files, and assembled into completed genome or target-gene sequences using special purpose software [61]. This infrastructure exists today only within a few genome centers, but will have to be replicated and extended enormously to allow worldwide dissemination of the technology. Given that the finished sequences themselves can be generated, EMR systems adapted to handling sequence data will have to be widely adopted across the healthcare system [62]. Second, improvements in bioinformatics tools will be needed to assess the functional significance of rare variants identified in genes of potential pharmacogenetic impact in individual patients' genome sequences. Third, major advances will be needed in our understanding of how genetic variation affects responses to commonly used drugs and other drugs with potentially severe side effects, and in the clinical utility of using genetic information to guide prescribing of these drugs. Assuming that suitable protections are put into place to protect patient privacy, the mining of large databases of individual genome sequences coupled to drug response data in EMRs will lead to novel pharmacogenetic discoveries and provide the evidence base for the clinical utility of pharmacogenetic markers [63]. Fourth, as the pharmacogenetics knowledge base expands, application of sequence data to clinical decision making will follow. Informatics-based decision tools will be needed to help guide clinicians in interpretation of genome sequences and personalization of prescribing practices based on those interpretations. Finally, outreach and education programs will be needed to facilitate clinician understanding of the new tools available to them to bring the benefits of personalized, genome-based medicine to their patients.

Executive summary

- Pharmacogenetics is the study of how individual genetic variation affects responses to medicines.
- The field is rapidly evolving owing to the development of novel technologies for detecting human genetic variation.

Variation in the human genome

- A total of 10–20 million sequence variants are believed to exist in the human population.
- Roughly half are considered to be common variants (minor allele frequency > 0.05) and half uncommon to rare variants.

- Next-generation DNA sequencing technology now makes it feasible to catalog all moderately common to common variants and to identify all variants in an individual human genome.

Pharmacogenetic discovery

- The discovery of pharmacogenetic effects requires the assembly of a patient cohort of suitable size with measurements of drug response.
- Most pharmacogenetic discovery to date has relied on testing the effects of variants in candidate genes for drug metabolism or drug action.
- Genome-wide association studies represent a genome-wide, essentially unbiased approach to discovering novel pharmacogenetic effects, and does not rely on the adequacy of current knowledge relating to drug metabolism or mechanism of action.
- While most pharmacogenetic studies to date have focused on the effects of common variants, rare variants are also likely to play a role, and understanding the role of rare variants will be important for making accurate predictions of risk for individual patients.
- Resequencing using next-generation sequencing technology is the most practical way to detect both common and rare variants in large numbers of patients.
- Resequencing can be applied to whole genomes, or can be focused on panels of candidate genes.
- Interpretation of sequencing data can be challenging due to the large number of variants discovered and uncertainty about the functional consequences of many variants.

Clinical applications

- Testing for individual variants represents the state of pharmacogenetic application in clinical practice today, and is having a small, but significant impact.

Future perspective

- The rapid development of genome technology is changing the pace of pharmacogenetic research, and will change the way pharmacogenetics impacts the practice of medicine in both quantitative and qualitative ways.
- New and expanded information management tools will be needed to capture the benefits expected from the intersection of genome technology and pharmacogenetics.

Acknowledgments

The author would like to acknowledge Howard McLeod, Sarah McWhinney Glass and Kevin Long for their helpful comments.

Bibliography

Papers of special note have been highlighted as:

- of interest
- of considerable interest

1. Evans WE, McLeod HL. Pharmacogenomics – drug disposition, drug targets, and side effects. *N. Engl. J. Med* 2003;348:538–549. [PubMed: 12571262]
2. Motulsky AG. Drug reactions enzymes, and biochemical genetics. *J. Am. Med. Assoc* 1957;165:835–837. [PubMed: 13462859]
3. The International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]
4. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861. [PubMed: 17943122]
5. Ionita-Laza I, Lange C, M Laird N. Estimating the number of unseen variants in the human genome. *Proc. Natl Acad. Sci. USA* 2009;106:5008–5013. [PubMed: 19276111]
6. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* 1977;74:5463–5467. [PubMed: 271968]
7. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]
8. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304–1351. [PubMed: 11181995]
9. Mardis ER. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet* 2008;9:387–402. [PubMed: 18576944] ▪ A recent review of current methods in next-generation sequencing.
10. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 2008;452:872–876. [PubMed: 18421352]
11. Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual. *Nature* 2008;456:60–65. [PubMed: 18987735]
12. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53–59. [PubMed: 18987734]
13. Ahn SM, Kim TH, Lee S, et al. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* 2009;19:1622–1629. [PubMed: 19470904]
14. Kim JI, Ju YS, Park H, et al. A highly annotated whole-genome sequence of a Korean individual. *Nature* 2009;460:1011–1015. [PubMed: 19587683]
15. Govindaraju DR, Cupples LA, Kannel WB, et al. Genetics of the Framingham Heart Study population. *Adv. Genet* 2008;62:33–65. [PubMed: 19010253]
16. McCarty CA, Peissig P, Caldwell MD, Wilke RA. The Marshfield Clinic Personalized Medicine Research Project: 2008 scientific update and lessons learned in the first 6 years. *Per. Med* 2008;5:529–542. ▪ Describes the experience to date in a pioneering effort to utilize the patient resources of a health maintenance organization, including the use of electronic medical records, in pharmacogenetic research.
17. Frueh FW, Salerno RA, Lesko LJ, Hockett RD. 4th US FDA-Drug Information Association pharmacogenomics workshop, held 10–12 December, 2007. *Pharmacogenomics* 2009;10:111–115. [PubMed: 19102720]
18. Thomas KL, Ellingrod VL. Pharmacogenetics of selective serotonin reuptake inhibitors and associated adverse drug reactions. *Pharmacotherapy* 2009;29:822–831. [PubMed: 19558256]
19. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229. [PubMed: 12029063]
20. Hinds DA, Stuve LL, Nilsen GB, et al. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005;307:1072–1079. [PubMed: 15718463]
21. Lavedan C, Licamele L, Volpi S, et al. Association of the *NPAS3* gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol. Psychiatry* 2008;14(8):804–819. [PubMed: 18521090]
22. Turner ST, Bailey KR, Fridley BL, et al. Genomic association analysis suggests chromosome 12 locus influencing antihypertensive response to thiazide diuretic. *Hypertension* 2008;52:359–365. [PubMed: 18591461]

23. Liu C, Batliwalla F, Li W, et al. Genome-wide association scan identifies candidate polymorphisms associated with differential response to anti-TNF treatment in rheumatoid arthritis. *Mol. Med* 2008;14:575–581. [PubMed: 18615156]
24. Cooper GM, Johnson JA, Langaee TY, et al. A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* 2008;112:1022–1027. [PubMed: 18535201]
25. Takeuchi F, McGinnis R, Bourgeois S, et al. A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet* 2009;5:E1000433. [PubMed: 19300499]
26. Kindmark A, Jawaid A, Harbron CG, et al. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J* 2008;8:186–195. [PubMed: 17505501]
27. Volpi S, Heaton C, Mack K, et al. Whole genome association study identifies polymorphisms associated with QT prolongation during iloperidone treatment of schizophrenia. *Mol. Psychiatry*. 2008
28. Sarasquete ME, Garcia-Sanz R, Marin L, et al. Bisphosphonate-related osteonecrosis of the jaw is associated with polymorphisms of the cytochrome P450 *CYP2C8* in multiple myeloma: a genome-wide single nucleotide polymorphism analysis. *Blood* 2008;112:2709–2712. [PubMed: 18594024]
29. The Search Collaborative Group. *SLCO1B1* variants and statin-induced myopathy – a genomewide study. *N. Engl. J. Med* 2008;359:789–799. [PubMed: 18650507]
30. Daly AK, Donaldson PT, Bhatnagar P, et al. *HLA-B*5701* genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet* 2009;41(7):816–9. [PubMed: 19483685]
31. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008;322:881–888. [PubMed: 18988837] • A recent review of current methods in genetics for identifying genes associated with medically relevant phenotypes.
32. Daly TM, Dumauval CM, Miao X, et al. Multiplex assay for comprehensive genotyping of genes involved in drug metabolism, excretion, and transport. *Clin. Chem* 2007;53:1222–1230. [PubMed: 17510302]
33. Caldwell MD, Awad T, Johnson JA, et al. *CYP4F2* genetic variant alters required warfarin dose. *Blood* 2008;111:4106–4112. [PubMed: 18250228]
34. Mega JL, Close SL, Wiviott SD, et al. Cytochrome p-450 polymorphisms and response to clopidogrel. *N. Engl. J. Med* 2009;360:354–362. [PubMed: 19106084]
35. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet* 2009;10:241–251. [PubMed: 19293820]
36. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet* 2008;40:695–701. [PubMed: 18509313] •• An analysis of the medical genetics literature that argues for the importance of including rare and uncommon variants in the study of common human traits and diseases.
37. Fearnhead NS, Wilding JL, Winney B, et al. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc. Natl Acad. Sci. USA* 2004;101:15992–15997. [PubMed: 15520370]
38. Sandilands A, Terron-Kwiatkowski A, Hull PR, et al. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat. Genet* 2007;39:650–654. [PubMed: 17417636]
39. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305:869–872. [PubMed: 15297675]
40. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat. Genet* 2005;37:161–165. [PubMed: 15654334]
41. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in *NPC1L1* associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc. Natl Acad. Sci. USA* 2006;103:1810–1815. [PubMed: 16449388]

42. Romeo S, Pennacchio LA, Fu Y, et al. Population-based resequencing of *ANGPTL4* uncovers variations that reduce triglycerides and increase HDL. *Nat. Genet* 2007;39:513–516. [PubMed: 17322881]
43. Ji W, Foo JN, O’Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet* 2008;40:592–599. [PubMed: 18391953]
44. Frank TS, Manley SA, Olopade OI, et al. Sequence analysis of *BRCA1* and *BRCA2*: correlation of mutations with family history and ovarian cancer risk. *J. Clin. Oncol* 1998;16:2417–2425. [PubMed: 9667259]
45. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66–72. [PubMed: 18987736]
46. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* 2007;4:907–909. [PubMed: 17934469]
47. Hodges E, Xuan Z, Balija V, et al. Genome-wide *in situ* exon capture for selective resequencing. *Nat. Genet* 2007;39:1522–1527. [PubMed: 17982454]
48. Bashiardes S, Veile R, Helms C, Mardis ER, Bowcock AM, Lovett M. Direct genomic selection. *Nat. Methods* 2005;2:63–69. [PubMed: 16152676]
49. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol* 2009;27:182–189. [PubMed: 19182786]
50. Dahl F, Stenberg J, Fredriksson S, et al. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc. Natl Acad. Sci. USA* 2007;104:9387–9392. [PubMed: 17517648]
51. Fredriksson S, Baner J, Dahl F, et al. Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Res* 2007;35:E47. [PubMed: 17317684]
52. Porreca GJ, Zhang K, Li JB, et al. Multiplex amplification of large sets of human exons. *Nat. Methods* 2007;4:931–936. [PubMed: 17934468]
53. Varley KE, Mitra RD. Nested Patch PCR enables highly multiplexed mutation discovery in candidate genes. *Genome Res* 2008;18:1844–1850. [PubMed: 18849522]
54. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461(7261):272–276. [PubMed: 19684571] ▪ One of the first uses of complete exome sequencing in multiple individuals to identify a gene for a human inherited disease.
55. Ng PC, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet* 2008;4:E1000160. [PubMed: 18704161]
56. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–3814. [PubMed: 12824425]
57. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;30:3894–3900. [PubMed: 12202775]
58. The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med* 2009;360:753–764. [PubMed: 19228618]
59. Phillips E, Mallal S. Successful translation of pharmacogenetics into the clinic: the abacavir example. *Mol. Diagn. Ther* 2009;13:1–9. [PubMed: 19351209]
60. Gill S, Goldberg RM. Cetuximab, chemotherapy and KRAS status in mCRC. *Nat. Rev. Clin. Oncol* 2009;6:379–380. [PubMed: 19561631]
61. Richter BG, Sexton DP. Managing and analyzing next-generation sequence data. *PLoS Comput. Biol* 2009;5:E1000369. [PubMed: 19557125]
62. Hoffman MA. The genome-enabled electronic medical record. *J. Biomed. Inform* 2007;40:44–46. [PubMed: 16616698]
63. Frueh FW. Back to the future: why randomized controlled trials cannot be the answer to pharmacogenomics and personalized medicine. *Pharmacogenomics* 2009;10:1077–1081. [PubMed: 19604080] ▪ A provocative discussion of the type of evidence necessary for demonstrating pharmacogenetic effects, and the direction future studies may take for amassing that evidence.
101. Human Cytochrome P450 (CYP) Allele Nomenclature Committee. www.cypalleles.ki.se/
102. dbSNP: Single Nucleotide Polymorphism database at NCBI. www.ncbi.nlm.nih.gov/projects/SNP/

103. The 1000 Genomes Project. www.1000genomes.org/page.php
104. The eMERGE Network. www.mc.vanderbilt.edu/victr/dcc/projects/acc/index.php/Main_Page
105. Hindorff, LA.; Junkins, HA.; Mehta, JP.; Manolio, TA. A catalog of published genome-wide association studies. www.genome.gov/gwastudies
106. Online calculator for initiation of warfarin dosing based on pharmacogenetic algorithms. www.WarfarinDosing.org