Polymorphisms affecting gene regulation and mRNA processing: Broad implications for pharmacogenetics

Andrew D. Johnson\textsuperscript{a,\*}, Danxin Wang\textsuperscript{b}, Wolfgang Sadee\textsuperscript{c}

\textsuperscript{a}Integrated Biomedical Sciences Graduate Program, Program in Pharmacogenomics, Department of Pharmacology, College of Medicine and Public Health, The Ohio State University, 5072 Graves Hall, 333 W. 10th Avenue, Columbus, OH 43210-1239, United States

\textsuperscript{b}Department of Pharmacology, College of Medicine and Public Health, The Ohio State University, Columbus, OH, United States

\textsuperscript{c}Felts Mercer Professor of Medicine and Pharmacology, Chair, Department of Pharmacology, Director, Program in Pharmacogenomics, College of Medicine and Public Health, The Ohio State University, Columbus, OH, United States

Abstract

Functional polymorphisms that alter gene expression and mRNA processing appear to play a critical role in shaping human phenotypic variability. Intensive studies on polymorphisms affecting drug response have revealed multiple modes of altered gene function, frequently involving \textit{cis}-acting regulatory sequence variants. Experimental and in silico methods have advanced the search for such polymorphisms, but considerable challenges remain here. A survey of polymorphisms in drug-related genes indicates that: (a) a substantial proportion of genetic variability still remains unaccounted for; (b) a majority of these genes harbors known regulatory polymorphisms; and (c) a portion of polymorphisms affect splicing and mRNA turnover. Pharmacogenetic optimization of individual drug therapy may require a complete understanding of all functional sequence variants in key genes. This review surveys known noncoding polymorphisms in genes encoding cytochrome \textit{P}450s and other drug-metabolizing enzymes, drug transporters, and drug targets and receptors. Current methods and challenges associated with the identification and characterization of functional polymorphisms are also discussed.

© 2004 Elsevier Inc. All rights reserved.

Keywords: \textit{Cis}-Regulatory; Polymorphism; Pharmacogenetics; Pharmacogenomics; SNP; RNA; Allele-specific; Genotype-phenotype; Cytochrome; Review

Abbreviations: 5-FU, 5-fluorouracil; 5-HT, serotonin transporter; 5-HTTLPR, serotonin gene-linked polymorphic region; 6-MP, 6-mercaptopurine; 6-TG, 6-thioguanine; ABCB, ATP-binding cassette (ABC), subfamily B; ACE, angiotensin I converting enzyme (peptidyl-dipeptidase A) 1; ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; ARE, adenylate-uridylate-rich element; ARED, adenylate-uridylate-rich element database; bp, base pair; BS, binding site; C/EBP, CCAAT/enhancer binding proteins; CAR, constitutive androstane receptor; CGAP, cancer genome anatomy project; CREATE, comprehensive research on expressed alleles in therapeutic evaluation; CRM, \textit{cis}-regulatory module; CYP, cytochrome; DBP, albumin D-site binding protein; DPYD, dihydropyrimidine dehydrogenase; \textit{Drd}2, dopamine D2 receptor; EM, extensive metabolizer; ESE, exonic splicing enhancer; Ets, E twenty six; FXR, farnesol X receptor; haploCHIP, haplotype-specific chromatin immunoprecipitation; hGR\text{a}, human glucocorticoid receptor alpha; hGR\text{b}, human glucocorticoid receptor beta; HGVbase, human genome variation database; HMGCR, 3-hydroxy-3-methylglutaryl-coenzyme A reductase; HNF, hepatocyte nuclear factor; hRNA, heterogeneous nuclear RNA; \textit{hPepT}2, human solute carrier family 15 (H+/peptide transporter), member 2 gene; HTR2A, 5-hydroxytryptamine (serotonin) receptor 2a; IM, intermediate metabolizer; IVS, intrinsic splice site variation; kb, kilobases; LDL, low-density lipoprotein; LIPC, hepatic lipase; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight spectroscopy; MMP, matrix metalloproteinase; NAT, \textit{N}-acetyltransferase; NF-Y, nuclear transcription factor Y; OATP, organic anion-transporting polypeptide; PABA, para-aminobenzoic acid; PAG, program for genomic applications; PharmGKB, the pharmacogenetics and pharmacogenomics knowledge base; PM, poor metabolizer; PPAR\text{a}, peroxisome proliferator-activated receptor alpha; PTC, premature termination codon; PTP1N, protein tyrosine phosphatase, nonreceptor type 1; PXR, pregnane X receptor; SIFT, sorting tolerant from intolerant; SLC, solute carrier; SLCO, solute carrier organic anion transporter family; SM12502, (+)-\textit{cis}-3,5-dimethyl-2(3-pyridyl) thiazolidin-4-one hydrochloride; SNP, single nucleotide polymorphism; SP-1, Sp1 transcription factor; SULT, sulfotransferase; TPMT, thiopurine (S)-methyltransferase; TSER, thymidylate synthase enhancer region; TYMS, thymidylate synthase; UGT1, UDP-glucuronosyltransferase-1; USF1, upstream transcription factor 1; UTR, untranslated region; VDR, vitamin D receptor; VNTR, variable number tandem repeat.

* Corresponding author. Tel.: 614 292 6680; fax: 614 292 7232.
E-mail address: johnson.2060@osu.edu (A.D. Johnson).

0163-7258/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.pharmthera.2004.11.001
1. Introduction

Regulation at the level of transcription initiation and RNA processing defines downstream biological effects. Such regulation occurs in cis, directly affecting the regulated gene, but it can also act in trans by altering activity of downstream genes (Fig. 1). Significant interindividual differences in gene expression patterns are common and may result from both environmental factors and cis- or trans-mediated genetic effects (Singer-Sam et al., 1992; Enard et al., 2002; Whitney et al., 2003; Pastinen & Hudson, 2004). There is growing evidence for abundant polymorphisms in cis-acting sequences that influence gene expression (Rockman & Wray, 2002; Yan et al., 2002; Bray et al., 2003; Lo et al., 2003) and indication that a significant portion of functional polymorphisms affect cis-regulatory elements (Stamatoyannopoulos, 2004; Wittkopp et al., 2004). Identifying the functional alleles that account for interindividual differences remains difficult (Ioannidis, 2003; Page et al., 2003; Sun et al., 2004). The genetic components of complex interindividual differences may require resolution of multiple modest variations in genotype which collectively yield a recognizable phenotype such as disease susceptibility or drug response.

Phenotypic differences can arise from genetic polymorphisms acting in cis by changing the protein coding sequence or at the level of RNA (Day & Tuite, 1998): affecting transcription (activation or inhibition through regulatory sites or structure of regulatory elements), mRNA processing, pre-mRNA splicing, exonic splicing enhancers (ESEs), exon skipping (Cartegni et al., 2003), mRNA stability (Sheets et al., 1990; Conne et al., 2000; Di Paola et al., 2002; Tebo et al., 2003), mRNA trafficking, or regulatory RNAs (Fig. 1). The most commonly studied polymorphisms, nonsynonymous changes that alter amino acid coding, appear in many cases insufficient to account for interindividual differences in disease aetiology and response to therapies. Further, it is estimated that functional polymorphisms that are cis-regulatory in the human genome outnumber those that alter protein sequence, and that the bulk of regulatory polymorphisms remain to be discovered (Ng & Henikoff, 2002; Rockman & Wray, 2002; Stamatoyannopoulos, 2004; Yan & Zhou, 2004). On the other hand, genomewide linkage analysis with mRNA expression as the quantitative trait demonstrates that interindividual differences in mRNA profiles appear to be largely caused by trans-acting factors (Morley et al., 2004). These statements are not incompatible since a single cis-acting polymorphism in a transcription factor or receptor could affect the expression of numerous other genes (Fig. 1). Therefore, to

Fig. 1. Human genetic variability, involving cis- and trans-acting polymorphisms. Cis-acting regulatory polymorphisms appear to outnumber functional polymorphisms in coding regions affecting protein sequence. If cis-acting polymorphisms alter signaling or transcription factor activity, multiple trans-acting changes ensue. Epigenetic changes can mimic cis-acting polymorphisms. Lastly, epistatic effects (multiple interacting polymorphisms) are likely to play a role as well. Not shown are regulatory effects exerted by small RNAs, also subject to genetic variability.
understand consequences of genetic variations, we must first determine whether interindividual differences of a protein’s activity are caused by polymorphisms in \textit{cis} or \textit{trans} (or both) or by environmental factors. If in \textit{cis}, we must then find the functional polymorphism(s) in the candidate gene that can account for the observed variations and epistatic interactions among them if several are present. If in \textit{trans}, we search for polymorphisms in \textit{trans}-acting factors (e.g., transcription factors). Lastly, we must also consider epigenetic factors such as methylation, imprinting, and chromatin structure modulation that can be transmitted through the germline or observed in somatic cells without alterations in the primary DNA sequence (Grewal & Moazed, 2003; Yan & Zhou, 2004).

While many studies have addressed \textit{cis}-regulatory variations, it is likely that a majority of functional variants are yet to be discovered. Novel techniques now enable broad investigation of this type of variation, which is likely to contribute substantially to knowledge of phenotypic variability of pharmagenetic relevance. Previous extensive studies on \textit{cis}-regulatory variations affecting disease susceptibilities (e.g., Loktionov, 2004) inform our thinking about functional variations with pharmacological-pharmagenetic implications. Here, we review advances in the discovery of \textit{cis}-regulatory variations within genes encoding drug-metabolizing enzymes, drug transporters, and drug targets and receptors (Table 1).

2. Modes of \textit{cis}-acting polymorphisms and methods for discovery

2.1. Experimental methods for discovering \textit{cis}-acting polymorphisms

Measurement of sequence variants, primarily single nucleotide polymorphisms (SNPs), provides the fundamental units for linking genetic sequence to traits. Most genes harbor multiple sequence variations (e.g., SNPs, repeats, indels) showing a broad range of frequencies and linkage disequilibrium among them. In clinical genetic association studies with goals to pinpoint candidate genes, selection of polymorphisms yielding maximum information is difficult. Confounding factors, such as their relative frequencies in the targeted populations, population admixture, and the effects of age and sex (Pinsonneault & Sadée, 2003), account in part for the failure to replicate many association studies. Most polymorphisms are nonfunctional and thus serve as markers for functional alleles. Rather than using single polymorphisms, associations are now often made with the use of haplotypes, blocks of linked polymorphisms, that may demarcate trait-significant \textit{cis}-regions of sequence. High-throughput SNP genotyping methods are now coming online, such as SNPlx, capable of screening thousands of SNP in many samples (Wenz, 2004). Such methods have been used to establish haplotype maps on a genomewide basis, including genes involved in drug metabolism, at significant marker density (Kamatani et al., 2004). Despite improvements in throughput for identification and association of sequence variants, the search for the specific identity of key regulatory variations is a difficult problem. Linking polymorphisms to transcriptional regulation has traditionally employed gene reporter constructs and in vitro DNA protein factor binding assays. For example, haplotype-specific chromatin immunoprecipitation (haploCHIP) takes advantage of the relation between the amount of phosphorylated RNA polymerase II and transcriptional activity (Knight et al., 2003). Via haploChIP, SNPs in regulatory regions can be investigated in conjunction with variation in transcription levels. However, these approaches provide incomplete pictures because they lack the physiological and structural context of a target tissue, or they currently lack high-throughput capability.

More broadly, mRNA expression measured by microarrays has been combined with genomewide linkage analysis, taking the expression level of each gene as the measured phenotype (Morley et al., 2004). Heritability of gene expression phenotypes can be explored through familial genotyping and transmission disequilibrium testing in nuclear families (Spielman & Ewens, 1996) or pedigree disequilibrium testing in larger pedigrees (Martin et al., 2000). Using target tissues (such as immortalized lymphocytes) from family members, this type of analysis is capable of distinguishing between \textit{cis}- and \textit{trans}-acting genetic factors and shows an abundance of functional genomic loci and a preponderance of \textit{trans}-acting effects, as expected (see Fig. 1). However, the technology suffers from low sensitivity and therefore may limit the detection of functional variations in target genes.

An alternative approach involves the analysis of allele-specific expression in a relevant target tissue; each allele experiences its own regulation in the same cellular environment, with the other allele (for autosomal genes) serving as an internal control. As a result, the method controls for tissue conditions, \textit{trans}-acting factors, and other environmental influences. Thus, SNPs in exonic and untranslated regions of message, can serve as markers for allele expression levels in individuals heterozygous for these markers. Taking the human solute carrier family 15 (H+/peptide transporter), member 2 gene (hPepT2) as one example, our laboratory has recently described a method for allele-specific measurement of mRNA expression through primer extension incorporation of fluorescent dideoxy-nucleotide terminating probes after RT-PCR amplification (Pinsonneault et al., 2004). Significant differences in the relative abundance of each allele in mRNA from kidney tissues demonstrated the presence of functional \textit{cis}-acting factors. The primer extension reaction can be multiplexed (Bray et al., 2004) so that it will be possible to search for functional \textit{cis}-acting polymorphisms in a large number of genes (Yan et al., 2002). Similar results can be achieved through methods employed on other platforms.
## Table 1
### Polymorphisms affecting gene regulation and mRNA processing

<table>
<thead>
<tr>
<th>Gene*allele</th>
<th>Type</th>
<th>Functional reports</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A2*1C</td>
<td>Enhancer (−3860G&gt;A)</td>
<td>↓ enzyme activity</td>
<td>Nakajima et al., 1999</td>
</tr>
<tr>
<td>CYP1A2*1F</td>
<td>Intron 1 (−163C&gt;A)</td>
<td>↑ enzyme induction</td>
<td>Chida et al., 1999; Sachse et al., 1999; Shimoda et al., 2002; Nordmark et al., 2002</td>
</tr>
<tr>
<td>CYP1A2*1K</td>
<td>Intron 1</td>
<td>Disrupt Ets BS, ↓ RNA, metabolism</td>
<td>Akilli et al., 2003</td>
</tr>
<tr>
<td>CYP1A2*7</td>
<td>Intron 6 splice donor SNP (5354G&gt;A)</td>
<td>PM to clozapine (single individual)</td>
<td>Alorge et al., 2003</td>
</tr>
<tr>
<td>CYP2A6*9</td>
<td>TATA Box</td>
<td>↓ RNA, protein, enzyme activity</td>
<td>Pitarch et al., 2001; Kiyotani et al., 2003; Yoshida et al., 2003</td>
</tr>
<tr>
<td>CYP2A6*1D</td>
<td>Enhancer (−1013A&gt;G)</td>
<td>↓ transcription (reporter assay)</td>
<td>Pitarch et al., 2004</td>
</tr>
<tr>
<td>CYP2A6*1H-J</td>
<td>Enhancer (−745A&gt;G)</td>
<td>Disrupt NF-Y BS, ↓ transcription</td>
<td>von Richter et al., 2004</td>
</tr>
<tr>
<td>CYP2A6*12</td>
<td>Intron 1 2A6/7 crossover</td>
<td>↓ enzyme activity in vitro and in vivo</td>
<td>Oscarson et al., 2002</td>
</tr>
<tr>
<td>CYP2B6*9</td>
<td>Splice variant</td>
<td>Skip exons 4–6</td>
<td>Lamba et al., 2003</td>
</tr>
<tr>
<td>CYP2B6*1G</td>
<td>Promoter (−750C&gt;T)</td>
<td>↓ RNA</td>
<td>Lamba et al., 2003</td>
</tr>
<tr>
<td>CYP2B6*1B</td>
<td>Enhancer (−2320T&gt;C)</td>
<td>↓ protein in Caucasian females</td>
<td>Lamba et al., 2003</td>
</tr>
<tr>
<td>CYP2B6*1C</td>
<td>ESE synonymous SNP</td>
<td>↓ protein in Caucasian females</td>
<td>Lamba et al., 2003</td>
</tr>
<tr>
<td>CYP2C9*6</td>
<td>Frameshift</td>
<td>Null, severe toxicity</td>
<td>Kidd et al., 2001</td>
</tr>
<tr>
<td>CYP2C9*1D</td>
<td>Exon 5 splice variant</td>
<td>PTC, PM phenotype</td>
<td>De Morais et al., 1994a</td>
</tr>
<tr>
<td>CYP2C9*3</td>
<td>Exon 4 premature stop</td>
<td>PTC, PM phenotype</td>
<td>De Morais et al., 1994b</td>
</tr>
<tr>
<td>CYP2C9*4</td>
<td>Initiation codon</td>
<td>Transcription ablation, PM</td>
<td>Ferguson et al., 1998</td>
</tr>
<tr>
<td>CYP2C9*7</td>
<td>Intron 5 splice donor SNP (IVS5+2 T&gt;A)</td>
<td>PM phenotype</td>
<td>Ibeau et al., 1999</td>
</tr>
<tr>
<td>CYP2D6*4A-L</td>
<td>Intron 3 splice variant</td>
<td>PTC, PM phenotype</td>
<td>Kagimoto et al., 1990</td>
</tr>
<tr>
<td>CYP2D6*11</td>
<td>Intron 1 splice acceptor</td>
<td>PTC, PM phenotype</td>
<td>Marczewski et al., 1995</td>
</tr>
<tr>
<td>CYP2D6*41</td>
<td>Promoter</td>
<td>IM phenotype, ↑ expression</td>
<td>Lovlie et al., 2001; Zanger et al., 2001; Gaedigk et al., 2003</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>Intron 6 (2988G&gt;A)</td>
<td>IM phenotype prediction</td>
<td>Raimundo et al., 2004</td>
</tr>
<tr>
<td>CYP2D6*44</td>
<td>Intron 6 splice donor SNP</td>
<td>↓ enzyme activity</td>
<td>Yamazaki et al., 2003</td>
</tr>
<tr>
<td>CYP2D7</td>
<td>138delIT</td>
<td>Pseudogene ORF, brain expression</td>
<td>Pai et al., 2004</td>
</tr>
<tr>
<td>CYP2E1*1D</td>
<td>Enhancer VNTR</td>
<td>← transcription, ↑ enzyme inducibility</td>
<td>McCarthy et al., 1998; Hu et al., 1999</td>
</tr>
<tr>
<td>CYP2J2*7</td>
<td>SP-1 BS (−76G&gt;T)</td>
<td>Disrupt SP-1 BS, ↓ transcription</td>
<td>King et al., 2002; Speckner et al., 2004</td>
</tr>
<tr>
<td>CYP3A4*1B</td>
<td>Proximal promoter (−392A&gt;G)</td>
<td>Disrupt nefidipine-specific repressor</td>
<td>Westlund et al., 1999; Spurde et al., 2002; Amirimani et al., 2003; Floyd et al., 2003</td>
</tr>
<tr>
<td>CYP3A4</td>
<td>Far upstream enhancer (−11,129−11,128 ins TGT)</td>
<td>Disrupt USF1BS, ↓ expression</td>
<td>Matsumura et al., 2004</td>
</tr>
<tr>
<td>CYP3A5*3</td>
<td>Exon 3B splice inclusion</td>
<td>PTC, ↓ protein, enzyme activity</td>
<td>Kuehl et al., 2001; Hustert et al., 2001a</td>
</tr>
<tr>
<td>CYP3A5*6</td>
<td>Exon 7 (14690G&gt;A)</td>
<td>Splice defect, Exon 7 deletion</td>
<td>Kuehl et al., 2001; Hustert et al., 2001a</td>
</tr>
<tr>
<td>CYP3A5*7</td>
<td>Exon 11 (27131 T ins)</td>
<td>Predicted PTC and ↓ protein</td>
<td>Hristov et al., 2001</td>
</tr>
<tr>
<td>CYP4F12*2v1</td>
<td>Intron 1 (146 bp del)</td>
<td>↓ transcription</td>
<td>Cauffiez et al., 2004</td>
</tr>
<tr>
<td>CYP4F12*2v2</td>
<td>9-SNP promoter allele</td>
<td>↑ transcription</td>
<td>Cauffiez et al., 2004</td>
</tr>
<tr>
<td>CYP7A1</td>
<td>Promoter (−204A&gt;C)</td>
<td>↑ response to atorvastatin</td>
<td>Kajinami et al., 2004</td>
</tr>
<tr>
<td>CYP8A1*1D-F</td>
<td>Promoter VNTRs</td>
<td>↑ number SP-1 BS, ↓ transcription</td>
<td>Chevalier et al., 2002</td>
</tr>
<tr>
<td>UGT1A1*28</td>
<td>(TA&lt;sub&gt;x&lt;/sub&gt;)TAA repeat</td>
<td>↓ transcription, protein, enz act, ↑ toxicity</td>
<td>Bosma et al., 1995; Iyer et al., 2002; Fang et al., 2004</td>
</tr>
<tr>
<td>UGT1A9*22</td>
<td>Promoter (T ins)</td>
<td>↑ transcription</td>
<td>Yamanaka et al., 2004</td>
</tr>
<tr>
<td>TPMT*V6a</td>
<td>Promoter VNTRs</td>
<td>↓ transcription, ↓ in vivo activity, conflicting</td>
<td>Spire-Vrayon de la Moureyre et al., 1999; Yan et al., 2000; Alves et al., 2001 de Leon et al., 2000</td>
</tr>
<tr>
<td>NAT1*16</td>
<td>3′ UTR AAA ins+C&gt;A</td>
<td>↓ protein, ↓ in vitro activity, disrupted predicted RNA structure</td>
<td>Gaber et al., 1999; Iida et al., 2002</td>
</tr>
<tr>
<td>ADH4</td>
<td>Promoter (−75A&gt;C)</td>
<td>↓ transcription</td>
<td>Kage et al., 2001; Hofmeyer et al., 2000; Sakaeda et al., 2001</td>
</tr>
<tr>
<td>ABCB1</td>
<td>Splice acceptor (3435C&gt;T)</td>
<td>↓ RNA, protein, drug transport activity</td>
<td>Heils et al., 1996; Heinz et al., 2000; Hranilovic et al., 2004</td>
</tr>
<tr>
<td>ABCB1</td>
<td>Promoter haplotypes</td>
<td>↑ transcription</td>
<td>Takane et al., 2004</td>
</tr>
<tr>
<td>SLC6A4</td>
<td>5′ HTTLPR (14 repeats)</td>
<td>↓ RNA, protein, transport activity</td>
<td>Takane et al., 2004</td>
</tr>
<tr>
<td>SLCO1B1*17</td>
<td>Upstream promoter</td>
<td>↑ pravastatin clearance</td>
<td>Niemi et al., 2004</td>
</tr>
<tr>
<td>TSER*3</td>
<td>Promoter VNTR</td>
<td>↑ RNA, Poor 5-FU treatment outcome</td>
<td>Horie et al., 1995; Vilasfranca et al., 2001; Marsh et al., 2001</td>
</tr>
<tr>
<td>TSER*3R</td>
<td>SNP in 2nd VNTR</td>
<td>Disrupt USF-1 binding site</td>
<td>Mandola et al., 2003</td>
</tr>
<tr>
<td>TYMS</td>
<td>3′ UTR (1494 6bp del)</td>
<td>↓ RNA, ↓ stability, ↓ intratumoral protein</td>
<td>Ulrich et al., 2000; Mandola et al., 2004</td>
</tr>
</tbody>
</table>
Table 1 (continued)

<table>
<thead>
<tr>
<th>Gene*allele</th>
<th>Type</th>
<th>Functional report(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTR2A</td>
<td>Synonymous (102C&gt;T)</td>
<td>(\downarrow) response to pravastatin</td>
<td>Arranz et al., 1998; Bray et al., 2004</td>
</tr>
<tr>
<td>HMGCRI(SNP3)</td>
<td>Intron 5</td>
<td>(\downarrow) response to pravastatin</td>
<td>Chasman et al., 2004</td>
</tr>
<tr>
<td>HMGCRI(SNP29)</td>
<td>Intron 15</td>
<td>(\downarrow) response to pravastatin</td>
<td>Chasman et al., 2004</td>
</tr>
<tr>
<td>MMP3</td>
<td>Promoter (~1171 5A&gt;6A)</td>
<td>(\downarrow) expression, (\downarrow) response to pravastatin</td>
<td>de Maat et al., 1999</td>
</tr>
<tr>
<td>LIPC</td>
<td>Promoter (~480C&gt;T)</td>
<td>Disrupt USF BS, (\downarrow) transcription, (\downarrow) enzyme activity, (\downarrow) response to treatments</td>
<td>Botma et al., 2001; Zambon et al., 2001</td>
</tr>
<tr>
<td>ACE</td>
<td>Intron 16 (287 bp ins)</td>
<td>(\downarrow) response to fluvastatin</td>
<td>Marian et al., 2000</td>
</tr>
<tr>
<td>PTP1B</td>
<td>3' UTR (1484 ins G)</td>
<td>(\uparrow) RNA, (\uparrow) mRNA stability</td>
<td>Di Paola et al., 2002</td>
</tr>
<tr>
<td>hGRB</td>
<td>3' UTR AUUUA SNP</td>
<td>(\uparrow) mRNA stability, (\uparrow) protein</td>
<td>Schaal et al., 2002</td>
</tr>
</tbody>
</table>

Polymorphisms with gene, allele denoted after **+** (if defined), the type of genetic alteration (if an allele defined by a single SNP exists, then the position and base change are given), reported functional observations, and related literature references.

(Wojnowski & Brockmoller, 2004) including the use of matrix-assisted laser desorption/ionization time-of-flight spectroscopy (MALDI-TOF; Ding et al., 2004) and allele-specific RT-PCR methodologies (Zhang et al., 2004). These techniques may be extended to unprocessed heterogeneous nuclear RNA (hnRNA) if exonic and untranslated markers are unavailable and hnRNA is abundant enough in the target samples (Hirota et al., 2004).

Taken together, these approaches allow rapid determination of the extent of cis- or trans-genetic variation in a locus and the heritability of that component. Further determination of the functional alleles is challenging because regulatory regions span across large genomic DNA segments. Therefore, in silico methods have proven helpful.

2.2. In silico methods for discovering cis-acting polymorphisms

Bioinformatics complements experimental investigations of regulatory polymorphisms, allowing investigators to interpret whether polymorphisms exist in a sequence region with predicted functional importance (Wasserman & Sandelin, 2004). Table 2 provides weblinks for tools and databases. Most tools employ phylogenetic footprinting to compare regions of sequence conservation that may highlight regulatory regions and then match these sequences against models predicting transcription factor binding sites (BSs). Pharmacogenetics-centered examples can be found in conjunction with the comprehensive research on expressed alleles in therapeutic evaluation (CREATE) website (Table 2: PromoLign, ReguLign). This general approach provides improvements over earlier methods but still fails to identify many regulatory sites (Wasserman & Sandelin, 2004). Acknowledging the combinatorial nature of factors binding regulatory regions, some tools use combinations of cis-regulatory modules (CRM) for specific tissues or gene types to make successful predictions (Liu et al., 2003).

A recent tool, PupaSNP, integrates available information on the potential for individual SNPs to alter expression or function (Conde et al., 2004). PupaSNP takes into account predicted transcription factor binding sites, intron/exon boundaries, predicted ESEs (Cartegni et al., 2003), amino acid changes in predicted motifs, and additional annotation information. ESE prediction is of interest since defects in splicing represent cis-regulatory variants, constitute a small but significant portion of known disease-causing mutations (Cooper & Mattox, 1997), and provide potential therapeutic targets (Sierakowska et al., 1996).

Posttranscriptional mRNA turnover represents another potentially important cause of genetic variability arising from cis-polymorphisms. mRNA stability can fluctuate through modulation of a number of pathways and changes in RNA structure and protein-RNA binding sites. For example, destabilizing adenylate-uridylate-rich sequence elements (ARE) are found in the 3' untranslated regions (3' UTR; Tebo et al., 2003) of 5–8% of human genes (Bakheet et al., 2003). Mutations in these elements are linked to disease pathology in human insulin resistance and have been suggested as stratifiers for administration of

Table 2

<table>
<thead>
<tr>
<th>Polymorphism databases and related tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource name</td>
</tr>
<tr>
<td>ARED (AU-rich element database)</td>
</tr>
<tr>
<td>CeleraSNPs</td>
</tr>
<tr>
<td>CGAP</td>
</tr>
<tr>
<td>CREATE</td>
</tr>
<tr>
<td>Environmental Genome DB</td>
</tr>
<tr>
<td>GeneSNPs</td>
</tr>
<tr>
<td>Human Gene Mutation DB (Cardiff)</td>
</tr>
<tr>
<td>HGvbase</td>
</tr>
<tr>
<td>Human Genome Variation Society</td>
</tr>
<tr>
<td>Hapmap Project</td>
</tr>
<tr>
<td>Innate Immunity PGA</td>
</tr>
<tr>
<td>JSNP Database</td>
</tr>
<tr>
<td>Perlegen</td>
</tr>
<tr>
<td>PharmGKB</td>
</tr>
<tr>
<td>PromoLign</td>
</tr>
<tr>
<td>PupaSNP</td>
</tr>
<tr>
<td>ReguLign</td>
</tr>
<tr>
<td>Seattle SNPs (UW-FHCRC)</td>
</tr>
<tr>
<td>SIFT</td>
</tr>
<tr>
<td>The SNP Consortium</td>
</tr>
</tbody>
</table>

Web references for tools and databases useful in finding and characterizing functional cis-regulatory polymorphisms.
protein tyrosine phosphatase, non-receptor type 1 (PTP1N) inhibitors (Di Paola et al., 2002). Human glucocorticoid receptor beta (hGR\(\beta\)) acts as a dominant negative factor for the steroid-responsive human glucocorticoid receptor alpha (hGR\(\alpha\), Bamberger et al., 1995), and increased hGR\(\beta\) expression has been associated with steroid resistance in asthmatics (Hamid et al., 1999; Sousa et al., 2000). A SNP in the 3' UTR of hGR\(\beta\) disrupts an ATTTA motif and leads to increased mRNA stability suggesting the functional cause of steroid resistance (Derijck et al., 2001; Schaaf & Cidlowski, 2002). AREs collected in a database (Bakheet et al., 2003) demonstrate how the identification of consensus motifs and novel regulatory elements in combination with integrated genomic polymorphism databases may allow parallel characterization of functional polymorphisms. Further effects of 'silent' polymorphisms in transcribed regions will likely be uncovered as methods improve for the difficult problem of predicting RNA tertiary structure (Chen et al., 1999; Zuker, 2003). Silent mutations predicted to change mRNA folding in drug-related genes have previously been supported experimentally: dopamine D2 receptor (Drd2; Duan et al., 2003) and NAT1 (de Leon et al., 2000). Newer algorithms that include evolutionary weighting schemes will likely lead to further experimentally validated examples (e.g., comRNA; Ji et al., 2004).

3. Drug-metabolizing enzymes

Numerous genes encoding proteins involved in drug metabolism and transport have been categorized, and nomenclature of polymorphic variations is being standardized (see Table 3 for references). Accounting for the largest proportion of genetic variation affecting drug therapy are the cytochrome P450s, a large, highly polymorphic family of heme-containing mono-oxygenases (http://www.imm.ki.se/CYPalleles/; Rendic, 2002; Guengerich, 2004). The human genome encodes at least 57 cytochrome P450s and contains 58 cytochrome P450 pseudogenes (Nelson et al., 2004), which are organized into 18 families (enzymes sharing >40% identity) and 43 subfamilies (enzymes sharing >55% identity). The most important cytochrome P450s involved in drug metabolism are the members of the CYP1, CYP2, and CYP3 families (Danielson, 2002). Comprehensive cytochrome genotyping assays are becoming feasible and are now applied in pharmaceutical trials, but a key question is whether we have sufficiently determined the bulk of the functional alleles in human populations. Many alleles have been reported, but there is an inherent bias toward sequencing and genotyping of coding regions. As a result, a portion of phenotypic variation remains unaccounted for by genetic factors. Large-scale sequencing projects continue to reveal noncoding alterations that could affect expression and function of drug-relevant genes (Iida et al., 2001b; Adjei et al., 2003; Aklillu et al., 2003; Allorge et al., 2003; Saito et al., 2003; Blaisdell et al., 2004; Cauffiez et al., 2004; Cauffiez et al., 2004; Fukushina-Uesaka et al., 2004; Murayama et al., 2004). Table 1 lists known functional noncoding polymorphisms, which include crucial drug-relevant functional alleles.

3.1. Cytochrome 1 family

CYP1A2 has roles in metabolism of clozapine, paracetamol, phenacetin, theophylline, imipramine, and tacrine and is generally probed with caffeine since it specifically demethylates this substrate. There is evidence for a number of functional, noncoding alleles in the sequence of CYP1A2. A moderately frequent single nucleotide change at a demonstrated protein binding site in the enhancer region (CYP1A2*1C allele) correlates with a decrease in enzyme activity (Nakajima et al., 1999). A single base change (−163C>A) in intron 1 (CYP1A2*1F), occurring frequently in a Japanese population (Chida et al., 1999), is correlated with high enzyme induction in Caucasians (Sachse et al., 1999). However, a study on the plasma concentrations of haloperidol in Japanese schizophrenics carrying CYP1A2*1F (Shimoda et al., 2002) and another on Swedish pregnant women with the same allele (Nordmark et al., 2002) dispute this connection. Another polymorphism in intron 1 (−729C>T) was found in an Ethiopian population 10 base pair (bp) upstream from one previously reported.

<table>
<thead>
<tr>
<th>Gene class</th>
<th>URL reference</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol dehydrogenases</td>
<td><a href="http://www.gene.ucl.ac.uk/nomenclature/genefamily/ADH.shtml">www.gene.ucl.ac.uk/nomenclature/genefamily/ADH.shtml</a></td>
<td>Duister et al., 1999</td>
</tr>
<tr>
<td>Aldehyde dehydrogenases</td>
<td><a href="http://www.aldh.org/">www.aldh.org/</a></td>
<td>Vasilou et al., 1999</td>
</tr>
<tr>
<td>Arylamine N-acetyltransferases</td>
<td><a href="http://www.louisville.edu/medschool/pharmacology/NAT.html">www.louisville.edu/medschool/pharmacology/NAT.html</a></td>
<td>Hein et al., 2000</td>
</tr>
<tr>
<td>Cytochrome P450s</td>
<td><a href="http://www.imm.ki.se/CYPalleles/">www.imm.ki.se/CYPalleles/</a></td>
<td>Nelson et al., 2004</td>
</tr>
<tr>
<td>Dihydropyrimidine dehydrogenase</td>
<td>(none available)</td>
<td>McLeod et al., 1998</td>
</tr>
<tr>
<td>Organic anion transporting polypeptides</td>
<td><a href="http://www.bioparadigms.org/slc/">www.bioparadigms.org/slc/</a> (see Glatt &amp; Meini, 2004 for amino acid changing alleles)</td>
<td>Hagenbuch &amp; Meier, 2004</td>
</tr>
<tr>
<td>Sulfitotransferases</td>
<td></td>
<td>Blanchard et al., 2004; Glatt &amp; Meini, 2004</td>
</tr>
<tr>
<td>UDP glucuronosyltransferases</td>
<td>som.flinders.edu.au/FUSA/ClinPharm/UGT/</td>
<td>Mackenzie et al., 1997</td>
</tr>
</tbody>
</table>

Table 3: Nomenclature guides for selected human gene families and alleles

Literature and web references for gene and allele nomenclature of families with pharmacogenetic importance. Definitive allele lists are not available for all families.
This novel allele (CYP1A2*1K) associates with decreased in vivo metabolism, decreased expression in reporter constructs, and disruption of binding within intron 1 by a member of the E twenty six (Ets) family of proteins (Aklillu et al., 2003). The results suggest that the discrepancy between previous intron 1 studies may be due to failure to completely determine haplotype structures since CYP1A2*1K (−163C>A, −729C>T, −739T>G) was shown to be functionally significant, but CYP1A2*1F (−163C>A) and CYP1A2*1J (−163C>A, −739T>G) were not. Another SNP (3534G>A) in the donor splice site of intron 6 (CYP1A2*7) was suggested to account for extremely high clozapine concentrations at normal doses in a single individual, but this has not been replicated (Allorge et al., 2003). These results illustrate difficulties in assigning functional properties to polymorphisms in regulatory regions and combinations of polymorphisms in haplotypes and, moreover, in relating them to pharmacokinetic differences in vivo where more than one factor contributes to the phenotype.

3.2. Cytochrome 2 family

CYP2A6 accounts for approximately 10% of human liver microsome CYPs and is the major player in oxidation of nicotine, cotinine, and a few pharmaceuticals (e.g., fadrozole, halothan, losigamone, letrozole, methoxyflurane, SM12502; Pelkonen et al., 2000). Interindividual expression levels vary more than 10-fold and are attributed to environmental and genetic factors, with Asian populations having a high proportion of poor metabolizers (PMs; Pelkonen et al., 2000). An uncommon allele (CYP2A6*12) results in a crossing over event between CYP2A6 and CYP2A7 in intron 2, addition of 10 amino acids and subsequently lower 7-hydroxylation activity of the enzyme (Oscarson et al., 2002). A TATA box allele (CYP2A6*1D) with high prevalence in Caucasians has been described which appears to disrupt an enhancer region affects expression and was identified so far, and the focus has been on the coding region of the gene have been extensively investigated (see review Zanger et al., 2004). More than 70 SNPs have been associated with adverse drug reactions or toxicity to drugs metabolized by CYP2C9 (Aithal et al., 1999; Lee et al., 2002; Ho et al., 2003; Schwarz, 2003). The deletion of an adenine at base pair 818 of the mRNA causes frame shift and yields a nonfunctional protein (CYP2C9*6; Kidd et al., 2001). While the allele frequency of this variant is <1%, it has been associated with toxicity after treatment with normal doses of phenytoin (Kidd et al., 2001).

CYP2D6 is the most highly expressed member of the CYP2C subfamily in hepatic tissue, and metabolizes 16% of drugs in current clinical use, including some drugs with narrow therapeutic indices such as the hypoglycemic tolbutamide, the anticonvulsant phenytoin, and the anticoagulant (S)-warfarin (Schwarz, 2003). Several SNPs that change amino acids and result in reduced enzyme activity have been identified and associated with adverse drug reactions or toxicity to drugs metabolized by CYP2C9 (Aithal et al., 1999; Lee et al., 2002; Ho et al., 2003; Schwarz, 2003). The deletion of an adenine at base pair 818 of the mRNA causes frame shift and yields a nonfunctional protein (CYP2C9*6; Kidd et al., 2001). While the allele frequency of this variant is <1%, it has been associated with toxicity after treatment with normal doses of phenytoin (Kidd et al., 2001).

CYP2C9 is the most highly expressed member of the CYP2C subfamily in hepatic tissue, and metabolizes 16% of drugs in current clinical use, including some drugs with narrow therapeutic indices such as the hypoglycemic tolbutamide, the anticonvulsant phenytoin, and the anticoagulant (S)-warfarin (Schwarz, 2003). Several SNPs that change amino acids and result in reduced enzyme activity have been identified and associated with adverse drug reactions or toxicity to drugs metabolized by CYP2C9 (Aithal et al., 1999; Lee et al., 2002; Ho et al., 2003; Schwarz, 2003). The deletion of an adenine at base pair 818 of the mRNA causes frame shift and yields a nonfunctional protein (CYP2C9*6; Kidd et al., 2001). While the allele frequency of this variant is <1%, it has been associated with toxicity after treatment with normal doses of phenytoin (Kidd et al., 2001).
Other genetic mechanisms for null alleles include frame shifts resulting from single or multiple base pair insertion or deletion (CYP2D6*3, CYP2D6*6, CYP2D6*15, CYP2D6*19, CYP2D6*20, CYP2D6*38, CYP2D6*42; Kagimoto et al., 1990; Saxena et al., 1994; Sachse et al., 1996; Marez-Allorge et al., 1999), a SNP that creates a PTC (CYP2D6*8; Broly et al., 1995), and other splicing defect mutations (CYP2D6*11, CYP2D6*44; Marez et al., 1995; Yamazaki et al., 2003), in addition to entire gene deletion (CYP2D6*5) or duplications (Gaedingk et al., 1991) and amino acid changes (CYP2D6*7,12; Evert et al., 1994; Marez et al., 1996). However, only a few SNPs in the promoter region have been identified, and phenotypic consequences have not been demonstrated. Since "normotypic" CYP2D6 carriers still display large variations in metabolic capacity, including the intermediate metabolizer (IM) phenotype, we need to discern whether this is due to yet unrecognized polymorphisms acting in cis or trans, or to enzyme induction effects. A SNP in the promoter region (−1548C>G) within the CYP2D6*41 allele has been associated with CYP2D6 expression, with the G allele correlating with higher levels of expression (Lovlie et al., 2001; Zanger et al., 2001). However, because this SNP is in linkage with other SNPs known to affect expression (Raimundo et al., 2000), it is not clear which is functional. The mutation has been used as a marker to rule out CYP2D6 PM status (Gaedingk et al., 2003). An intron 6 SNP (2988G>A) with a frequency of 8.4% in Caucasians was found to be an improved predictive marker for the IM phenotype over CYP2D6*41 (Raimundo et al., 2004).

Although most frameshift mutations in cytochrome P450 genes cause nonfunctional proteins, a frameshift mutation in the CYP2D7 pseudogene generates a functional enzyme (Pai et al., 2004). This common single base pair deletion (138delT) generates an open reading frame in the CYP2D7 pseudogene and a spliced variant containing partial inclusion of intron 6. This transcript produces a functional protein, that is expressed in the brain, but not liver or kidney. The variant CYP2D7 metabolizes codeine to morphine more efficiently than CYP2D6 in Neuro2a cells and also colocalizes with μ-opioid receptors in brain tissue, suggesting a possible role in metabolism at the drug site of action.

Less than 1% of hepatic cytochrome P450 content is due to CYP2B6, yet it is involved in the metabolism of ~70 clinically employed drugs, including alfentanil, ketamine, bupropion, verapamil, tamoxifen, efavirenz, and drugs of abuse such as methylenedioxymethyl-amphetamine and nicotine (Lang et al., 2001). CYP2B6 activity in liver microsomes varies more than 100-fold among different individuals (Yamano et al., 1989; Ekins et al., 1998), with female subjects having higher levels of mRNA, protein and enzyme activity than males (Lamba et al., 2003). Large inter racial differences are also observed for CYP2B6 with Hispanic females having higher CYP2B6 activity than Caucasian or African-American females (Lamba et al., 2003).

Many variants have been identified for CYP2B6 including mRNA splice variants (Lamba et al., 2003). These variants paint a complex picture because multiple potentially functional polymorphisms are found in linkage and thus produce epistatic interactions. The most common splice variant skips exons 4–6 (CYP2B6*9), while others lack the first 29 bp of exon 4 or contain an intron 3 insertion. Because of the presence of PTCs, all of these variants encode truncated proteins. A nonsynonymous SNP in exon 4 (15631G>T) that disrupts an ESE and a SNP (15582C>T) in an intron 3 branch site are correlated with these splicing variants. Several SNPs in the promoter region have also been reported. The −750T>C SNP (CYP2B6*1G) correlates with lower levels of expression. The −2320T>C SNP (CYP2B6*1B) in the hepatocyte nuclear factor 4 (HNF4) binding site in the promoter and the SNP in the intron 3 branch point (15582C>T) show a high degree of linkage disequilibrium and associate with low quantities of CYP2B6 protein in Caucasian females. Recently, several missense SNPs forming null alleles were identified, but their combined frequency is only 2.6% in a Caucasian population (Lang et al., 2004). In combination these genetic variants do not fully explain the phenotypic variability in CYP2B6 activity. Because CYP2B6 expression is correlated with constitutive androstane receptor (CAR) expression (Lamba et al., 2003), it is possible that polymorphisms in the CAR gene might affect CYP2B6 expression in trans. Recently, several novel tissue-specific variant isoforms of CAR have been identified (Lamba et al., 2004), but their association with CYP2B6 expression remains to be established.

### 3.3. Cytochrome 3 family

The CYP3A subfamily of proteins is highly homologous but exhibits wide tissue expression differences, implying a diversity of regulatory control. CYP3A4 is a critical enzyme because it is involved in the metabolism of over 30% of clinically used drugs including cyclosporin, erythromycin, and nifedipine. Interindividual expression differences of CYP3A4 in liver are substantial and have not been sufficiently ascribed to known functional polymorphisms (Hiroti et al., 2004). Principal among these has been a promoter mutation (CYP3A4*1B) that disrupts a nifedipine-specific repressor element (Westling et al., 1999; Amiriman et al., 2003) which has been suggested to associate with treatment-related leukemia (Felix et al., 1998). However, different reporter constructs and confounding results suggest the need for consideration of a more complex regulatory picture (Spurde et al., 2002; Floyd et al., 2003). Clarification of polymorphic trans-factors also provided insufficient explanation (Zhang et al., 2001). A recent attempt to elucidate genetic components of variation scanned up to 13 kilobase (kb) upstream of CYP3A4 and found a novel far upstream enhancer element with a polymorphism (~11,129–11,128 ins TGT) that disrupts an upstream transcription factor 1 (USF1) binding site and reduces...
expression, but is relatively uncommon (3.1%) in a French population (Matsumura et al., 2004). The complexity of the yet unfolding regulatory story of CYP3A4 underlies the importance of careful analysis of regulatory regions and the need for application of new technologies to resolve the nature and extent of genetic factors contributing to variabilities in expression and activity.

Polymorphic CYP3A5 expression may also contribute to interindividual and interracial differences in CYP3A-dependent drug clearance (Kuehl et al., 2001). CYP3A5 was formerly considered to be an extrahepatic enzyme (Wrighton et al., 1989; Schuetz et al., 1992; Murray et al., 1995; Haehner et al., 1996; Kivisto et al., 1996), only sporadically found in the liver tissue of some individuals (Wrighton et al., 1989). Kuehl et al. (2001) discovered a SNP (6986A>G) in intron 3 (CYP3A5*3) creates a cryptic splice site and an extra exon (exon 3b) and is responsible for polymorphic expression of CYP3A5 in the liver. This CYP3A5*3 allele encodes an aberrantly spliced mRNA with a PTC resulting in an unstable mRNA and no detectable CYP3A5 mRNA and protein in the liver. The allele frequency of CYP3A5*3 is reported to be 75.9% in a Japanese population (Saeki et al., 2003) and ~73% in an African-American population (Hustert et al., 2001a). CYP3A4 and CYP3A5 have overlapping substrate specificity (Bargeiri et al., 1989; Gorski et al., 1994; Guitton et al., 1997), and CYP3A5 can represent over 50% of the total hepatic CYP3A content in some individuals (Kuehl et al., 2001). Thus, genetic polymorphism of CYP3A5 may play a role in the variability in CYP3A targeted drug response, in some human populations.

3.4. Other cytochromes

Other human cytochrome genes have reported functional regulatory polymorphisms, and their diversity indicates that more will likely be uncovered. A relatively new family of cytochromes, CYP4F, contains members important for inflammation physiology. CYP4F12 is expressed in liver and intestine and co-hydroxylates eicosanoids, fatty acids, and antihistamines (e.g., ebastine; Bylund et al., 2001; Hashizume et al., 2001). Cauffiez et al. (2004) recently published the first study of the CYP4F12 promoter region and revealed 2 alleles (CYP4F12*1, CYP4F12*2) common in the French population that associate with significantly reduced expression levels in HepG2 cells (Cauffiez et al., 2004). CYP4F12*1 is a 192 bp deletion in intron 1 (21% frequency) that diminishes expression. CYP4F12*2 is a nine SNP-phased haplotype (8.5% frequency) which also associates with lower expression. In silico analysis suggests functional sites within these alleles, but further characterization is needed. The extrahepatically expressed CYP2J2 metabolizes diclofenac, bufuralol, and astemizole and contains a common mutation in a putative Sp1 transcription factor (SP-1) site (King et al., 2002). Contrary to previous reports the functional effects of this SNP were only speculated upon after discovery, while effects for less common amino acid changes were actually investigated. However, a recent article verifies that CYP2J2*7 disrupts an SP-1 binding site and reduces transcription significantly (Speckler, 2004). Other proposed functional CYP polymorphisms include variable number tandem repeats (VNTRs) in the 5' regulatory regions of CYP2E1 (McCarver et al., 1998; Hu et al., 1999) and CYP8A1 (Chevalier et al., 2002), and point mutation in the CYP7A1 promoter recently linked with reduced response to atorvastatin (Wang et al., 1998; Kajinami et al., 2004). Taken together the mutational record of the CYP superfamily argues for extensive consideration of regulatory mutations in all families critical to pharmacology and disease.

3.5. Other classes of drug-metabolizing enzymes

Drug-metabolizing enzymes other than the cytochromes also display significant interindividual allelic differences. Non-CYP genes and their variations may play increasingly important roles as pharmaceutical companies design drugs that evade the challenges of cytochrome variability (Ingelman-Sundberg, 2001). The cluster of UDP-glucuronosyltransferase-I (UGT1) genes (UGT1 superfamily) at 2q37 exhibits exon sharing and harbors identified promoter and missense mutations that have been associated with 2- to 6-fold lower conjugation activity (Kohle et al., 2003). Pathways involving members of the UGT1 superfamily act on approximately 35% of all drugs metabolized by phase II drug-metabolizing enzymes (Evans & Relling, 1999). Over and above many rare polymorphisms in the UGT1A1 gene, a common promoter region dinucleotide repeat (5–8 repeats; UGT1A1*28) accounts for reductions in expression level via alteration of transcription initiation (Bosma et al., 1995; Guillemette, 2003). The polymorphism correlates with lower protein levels and reduced conjugation activity (Fang & Lazarus, 2004), and is implicated in toxicity to treatment with irinotecan via altered glucuronidation of its active metabolite SN-38 (7-ethyl-10-hydroxycamptothecin; Iyer et al., 1999) and adverse reactions including neutropenia and diarrhea (Iyer et al., 2002). UGT1A1*28 is relatively frequent in many populations: Caucasian (~32%), Asian non-Japanese (~15%), African (~41%), Italian (~36%); Guillemette, 2003). A promoter region polymorphism in UGT1A9 (UGT1A9*22) common in Japanese, Caucasians and African-Americans has recently been associated with higher expression via HepG2 reporter constructs, but related enzyme induction or metabolism phenotypes are unreported (Yamanaka et al., 2004).

Thiopurine (S)-methyltransferase (TPMT) is a prominent example of how pharmacogenetics can impact individual treatment. The enzyme is involved in the metabolism of thiopurines: the cytotoxic drugs 6-mercaptopurine (6-MP) and 6-thioguanine (6-TG), and the immunosuppressant azathioprine, which is rapidly converted to 6-MP. Amino acid changing polymorphisms account for most of the
variability in the RBC activity of TPMT, although their frequency varies widely in different populations (Vuchetich et al., 1995; McLeod & Siva, 2002). VNTR in the promoter region of TPMT have been identified (Spire-Vayron de la Moureyre et al., 1998, 1999) and suggested as modulators of expression and thus enzyme activity (Fessing et al., 1998; Spire-Vayron de la Moureyre et al., 1998; Yan et al., 2000; Alves et al., 2001). These repeats appear to be in linkage disequilibrium with the nonsynonymous alleles making definitive declaration of their functional role difficult (Marinaki et al., 2003).

N-acetyltransferases (NAT) catalyze acetyltransfer from acetylcoenzyme A to an array of carcinogens and aryamine and hydrazine drugs (e.g., para-aminobenzoic acid [PABA], isoniazid, sulfamethazine, procainamide, nitrazepam, dapsone). The 2 human NATs, NAT1 and NAT2, are intronless and so far have mainly nonsynonymous polymorphisms that associate with enzyme activity (Hein et al., 2000; Butcher et al., 2002). Slow acetylation genotypes correlate with adverse effects after combined isoniazid and rifampicin therapy (Ohno et al., 2000) and co-trimoxazole treatment (Zielinska et al., 1999). The extent of N-acetylation also accounts for variability in toxicity to amonafide treatment in cancer patients, with rapid acetylators experiencing significantly greater toxicity (Ratain et al., 1995). A relatively uncommon allele (NAT1*16) is an AAA insertion and C to A transversion in the 3' UTR region and correlates with predicted structural variation, 2-fold lower expression and a similar reduction in N-acetylation of substrates (de Leon et al., 2000). Another 3' UTR variant NAT1*10 is suggested to correlate with higher enzyme activity (Bell et al., 1995; Payton & Sim, 1998), but this has not been supported in a number of studies perhaps due to linkage with an amino acid change not detected in some of the former assays (Bruhn et al., 1999; de Leon et al., 2000; Hein et al., 2000).

The sulfotransferase (SULT) family forms sulfate conjugates with a variety of xenobiotics and endogenous small molecules and now numbers 17 distinct genes (Freimuth et al., 2004). Multiple studies report interindividual differences in SULT expression, drug-related response, and realization of a heritable component (Reiter & Weinshilboum, 1982; Bonham Carter et al., 1983; Weinshilboum, 1990; Glatt & Meinl, 2004). Noticeably, no functional studies on noncoding variants have been reported despite a substantial number of such variants identified in sequencing studies (Freimuth et al., 2001; Iida et al., 2001; Glatt & Meinl, 2004).

A few aldehyde (ALDH) and alcohol (ADH) dehydrogenases are responsible for 90% of the metabolism of ethanol, a compound with potential for pharmacological interactions. Well-known coding variants in ADH2 and ALDH2 have been described which associate with resistance to alcoholism. ADH4 is expressed in specific portions of the GI tract (Vaglenova et al., 2003) and is essential for vitamin A metabolism (Duester et al., 2003), but may also have a role in first-pass metabolism of ethanol (Yin et al., 2003). A reporter assay indicated that a single functional promoter allele affects ADH4 expression (Edenberg et al., 1999), and no coding variations were identified in a subsequent sequencing study (Iida et al., 2002), but in vivo characterizations of allelic protein or tissue level effects are yet unreported. These examples show that further understanding of polymorphic expression in many types of genes beyond the cytochrome P450 superfamily is important to understanding the genetic basis of interindividual differences in drug responses.

4. Drug transporters

Modulation in drug transporter expression potentially affects the uptake and efficacy of many compounds. Numerous genes encode solute carriers (SLC families, >300 genes) and the class of primary active ATP-binding cassette (ABC) transporters (48 genes). Because of the size of these gene families, we only address a few examples here. The multiple drug resistance polypeptide ATP-binding cassette, subfamily B 1 (ABCB1: MDR1, Pgp) is an energy-dependent protein efflux pump that acts upon a wide range of natural and pharmacological substrates (see review Sun et al., 2004). A synonymous SNP (3435C>T) has been associated with low expression and altered pharmacokinetics in a number of studies (Hoffmeyer et al., 2000), but others have reported conflicting results (Sakaeda et al., 2001) or dismissed this as an unlikely functional SNP since it is not a coding alteration. In these cases, it is suggested that a functional allele (perhaps 2677G>A,T) must be in linkage, with 3435C>T being an indicator SNP. However, the synonymous polymorphism could affect mRNA structure, stability, or translational efficiency. Recently, Takane et al. (2004) conducted a functional analysis of ABCB1 variants. They found 10 promoter variants (7 were novel), an association of 3435C>T with lower expression, association of promoter haplotypes with transcription level differences independent of 3435C>T, a novel transcription factor binding site which is disrupted in a haplotype correlating with lower expression, and ruled out differences in methylation status as a principal cause (Takane et al., 2004). Three promoter variants found in rare haplotypes were associated with higher transcriptional expression, including one (−129T>C) that was previously reported to be associated with high transport activity in hematopoietic stem cells (Calado et al., 2002). While there are numerous studies associating ABCB1 polymorphisms with altered drug disposition and effect, these results are often not reproducible in different populations. This points to a lack of sufficient understanding of the interaction among multiple genetic factors determining ABCB1 expression and function.

Two noncoding polymorphisms of the serotonin transporter (5-HTT) gene SLC6A4 have been studied extensively based on allelic differences in expression in brain and other tissues (Lesch et al., 1994; Heils et al., 1996; Hrancilovic et al., 2004). The contribution of serotonin neurophysiology to
psychiatric disorders is great, and thus the impact of these variants may be far-reaching. In particular, the short form of the promoter region polymorphism (serotonin gene-linked polymorphic region [5-HTTLPR]) correlates with lower transcriptional and translational activity (Heinz et al., 2000; Hranilovic et al., 2004), blunted fenfluramine-induced prolactin release (Reist et al., 2001), greater amygdala neuronal activity in response to fearful stimuli (Hariri et al., 2002), poorer efficacy of antidepressants (e.g., citalopram, Eichhammer et al., 2003; fluoxetine, Rausch et al., 2002; paroxetine, Pollock et al., 2000), and antidepressant induced mania in bipolar disorder (Mundo et al., 2001). The presence of a genetic component to variability in response to antidepressants is also suggested by family studies (e.g., O’Reilly et al., 1994). Contributions of this polymorphism to variation in colonic uptake of alosetron (Scherl & Frissora, 2003) and heroin dependence (Gerra et al., 2004) have also been suggested. However, there are dissenting results, and the debate over the extent of the functional importance of these alleles is yet to be settled. There are alternative alleles, splice variants, and polyadenylation signals that might play roles (Delbruck et al., 1997; Battersby et al., 1999; Michaelovsky et al., 1999; Frisch et al., 2000; Nakamura et al., 2000; Cigler et al., 2001; Sun et al., 2002) and significant population differences in genotype to consider (Gelernter et al., 1999; Lotrich et al., 2003).

The members of the organic anion-transporting polypeptide (OATP) superfamily, encoded by the solute carrier (SLCO) genes, are broadly expressed and mediate transport of a wide range of endogenous and exogenous compounds including anions, cations, neutral compounds, and peptidomimetic agents (Tirona & Kim, 2002; Hagenbuch & Meier, 2004). Regulatory region polymorphisms have been reported in a number of SLCO genes (Iida et al., 2001a; Tirona & Kim, 2002). OATP-A (SLCO1A2) is expressed in brain capillary endothelial cells suggesting a role in blood-brain barrier permeability of solutes and this transmembrane protein transports analgesic opioid peptides (Gao et al., 2000). There is a SLCO1A2 SNP localized to a HNF1α binding site (Kullak-Ublick et al., 1997), but there are no functional reports on it or any of the identified nonsynonymous polymorphisms in the gene (Source: dbSNP). A newer far upstream polymorphism (−11187G>A) in OATP-C (SLCO1B1) was associated with 98% higher clearance (AUC) of pravastatin in Caucasian males and combined in a haplotype (SLCO1B1*17) with coding polymorphisms which exhibited a similar association (Niemi et al., 2004). Drug transporters will likely provide further examples of clinically relevant cis-regulatory polymorphisms.

5. Drug targets and receptors

Modulation of expression of drug targets is another avenue for studying interindividual differences in therapeutic response. In the field of cancer pharmacogenetics, understanding the expression patterns in patients’ tumors or their untransformed genomes in somatic cells can guide selection or administration of treatment. The principal downstream target of the common chemotherapeutic 5-fluorouracil (5-FU) is thymidylate synthase (TYMS). Three copies of a 28 bp tandem repeat (thymidylate synthase enhancer region [TSER]*3) in the promoter of TYMS have been associated with higher TYMS levels (Horie et al., 1995) and poorer outcomes with 5-FU treatment (Horie et al., 1995; Villafranca et al., 2001). Here, again, we see a complex regulatory picture that needs dissection as further polymorphisms show similar association: a SNP within the TSER*3 (TSER*3RG) that abolishes a USF1 binding site (Mandola et al., 2003) and a 3’ UTR SNP now associated with message stability (Ulrich et al., 2000; Mandola et al., 2004). Inactivation of 5-FU is principally mediated via dihydrorymidine dehydrogenase (DPYD), and variations in its activity can have fatal consequences (Van Kuilenburg et al., 2002). A splice site transition (DPYD*2A) accounts for some observed toxicity but has low frequency (<1%; Wei et al., 1996). Despite description of many other polymorphisms (Wei et al., 1996; McLeod et al., 1998; Collie-Duguid et al., 2000), 5-FU toxicity remains only partially understood. Other examples in cancer pharmacogenetics include applications of genotyping to treatment with well-known tyrosine kinase inhibitors (e.g., herceptin; Arteaga & Baselga, 2004). Functional explanations for these stratifying mutations, often involving genominc instability and high expression or loss of heterozygosity, are becoming understood (Sordella et al., 2004).

Drug targets in the brain and the degree to which their genetic variation explains differences in psychoses and their treatment have been the focus of much research. A synonymous 102C>T polymorphism of the 5-hydroxytryptamine (serotonin) receptor 2a (HTR2A) was previously proposed to associate with responsiveness to clozapine (Arranz et al., 1998), implying that it is in linkage disequilibrium with cis-acting regulatory polymorphisms. However, using this SNP as a marker for measuring relative allelic mRNA abundance, no difference in expression was detected in adult brain tissues (Bray et al., 2004). This argues against the presence of cis-acting regulatory polymorphisms. These findings do not address possible effects on translation but do argue for a shift of focus in future research on variants in this gene region.

An intriguing study in a large cohort of individuals treated with pravastatin recently revealed a significant association between reduction in low-density lipoprotein (LDL) levels after treatment and 2 intronic SNPs in the 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) gene (Chasman et al., 2004). The SNPs are in linkage disequilibrium but are not near intron-exon borders or CpG dinucleotides. The functional explanation remains an open question, but there is disequilibrium with a 3’ UTR variant present in mRNA, raising the possibility of altered
pretranslational regulation. If a functional allele is identified and generalized to other statins, it may help explain why a significant proportion of treated individuals does not respond readily to this blockbuster class of drugs (Schmitz & Drobnik, 2003). There are no definitive genetic markers that account for observed variations despite a number of non-coding variations previously associated with differential responses to statins (de Maat et al., 1999; Marian et al., 2000; Zambon et al., 2001; Schmitz & Drobnik, 2003; Kajinami et al., 2004).

These few transporter and receptor examples highlight the important contribution regulatory polymorphisms will play in defining genetic variability in pharmacokinetics and pharmacodynamics. The identification of allelic differences in key genes may also allow eventual targeting of sequence-specific functional cis-regulatory polymorphisms (Fluitert et al., 2003; Miller et al., 2003; Achenbach et al., 2004; Bruno et al., 2004). Alternatively, general classes of cis-regulatory variants may be targeted. For instance, aminoglycoside antibiotics have been shown to inhibit proofreading activity by misincorporating an amino acid and thus ameliorating nonsense mutations through increased translational read-through (Howard et al., 1996). Analysis of cystic fibrosis genotypes among over 1000 known variants has thus been used to select targeted therapy in a small, stratified subgroup of the disease population (Wilschanski et al., 2003).

6. Cis-acting polymorphisms in relevant trans-factors

While most genes harbor cis-acting changes relevant to their expression, genetic variants regulating gene expression in trans are likely to dominate interindividual variability in mRNA expression profiles (Wittkopp et al., 2004). This results from pleiotropic consequences on gene expression by cis-acting polymorphisms that alter the function of transcription factors, receptors, and other signaling molecules. We therefore need to consider cis-acting polymorphisms within the genes encoding trans-factors influencing drug-metabolizing enzymes, transporters and targets, reflecting complex regulatory networks underlying pharmacogenetics phenotypes (Rushmore & Kong, 2002; Akiyama & Gonzalez, 2003). Among the transcription factors, we can distinguish between those determining constitutive expression, on the one hand, and factors mediating enzyme induction, a common cause of temporal changes in drug-metabolizing capacity. Key players in the regulation of genes discussed in this review include members of the nuclear hormone receptor superfamily (pregnane X receptor [PXR], CAR, farnesol X receptor [FXR], HNF4α, peroxisome proliferator-activated receptor alpha [PPAR-α], vitamin D receptor [VDR]), and transcription factors—HNF1α, HNF3, HNF6, CCAAT/enhancer binding proteins [C/EBP], albumin D-site binding protein [DBP]; Akiyama & Gonzalez, 2003). Studies of human polymorphisms in these genes may have important pharmacological implications, but current understanding is immature.

For example, PXR stimulates transcription of a number of drug-metabolizing enzymes (e.g., CYP3A4; Lehmann et al., 1998; Goodwin et al., 1999; Schuetz et al., 2001; Tirona et al., 2003), as well as ABCBI in intestine (Geick et al., 2001). CYP3A4 cis-variants do not adequately account for observed drug phenotypes, thus implicating variability in other factors such as PXR (Lamba et al., 2002). PXR expression correlates with CYP3A subfamily expression (Westlund-Johnsson et al., 2003). In vitro assays of PXR variants encoding altered proteins have demonstrated correlated changes in CYP3A4 expression, particularly in response to rifampicin induction of the enzyme (Hustert et al., 2001b; Zhang et al., 2001). However, these changes are infrequent, and a heterozygous carrier of one such polymorphism showed normal CYP3A4 metabolism (Zhang et al., 2001). Three silent mutations in PXR also correlated with changes in its expression level (Zhang et al., 2001). Moreover, PXR is also transcribed from an alternative initiation codon (Bertilsson et al., 1998). A 6-bp deletion in the promoter of this alternative transcript, hPAR-2, disrupts a predicted HNF1α binding site and abolishes transcription in a liver cell line, but again no human phenotype was observed (Uno et al., 2003).

HNFs have a large role in the liver-specific enhancement of transcription of many cytochromes (Akiyama & Gonzalez, 2003). HNF1α and HNF4α polymorphisms have been widely scanned because of their relation with diabetes in many populations (Ryffel, 2001), but associated effects on drug phenotypes are not well investigated. Functional HNF1α promoter polymorphisms, including one in a putative HNF4 binding site have been described (Gragnoli et al., 1997). Amino acid changes were recently shown to result in a PTC and decreased protein stability through nonsense-mEDIATE decay (Harries et al., 2004). The documented role of HNFs in the expression of many genes of pharmacological relevance (e.g., HNF4 regulation of CYP3A4; Tirona et al., 2003) warrants further work on the effects of polymorphisms in these trans-factors. Because genes are regulated by multiple factors, often with overlapping specificity, characterization of the effects of variation in trans factors remains a challenging task. Ideally, for genes of critical pharmacological importance, we will eventually arrive at a multivariate understanding that accounts for variations at the target locus and variations effecting the contextual trans-inputs to the locus.

7. Conclusion

Accounting for genetic components of variation in all phenotypes must ultimately be done at cis-sequences with pleiotropic effects reverberating in trans (Fig. 1). The examples presented here demonstrate that as pharmacogenetics proceeds, continuing identification of novel cis-
regulatory variants and their functional effects is necessary. High-throughput genotyping for clinical pharmacological applications is increasingly feasible, and screening panels are easily developed, but the identification and selection of relevant alleles lags behind. This process remains difficult for reasons discussed here, including population-specific sources of error, limits of experimental approaches, and the challenges of investigating the many potential regulatory modes, including accounting for epistasis and epigenetics. If well-characterized genes are indicative, a few common polymorphisms within a population will cover the bulk of the genetic variation while many, less frequent polymorphisms will explain the rest. However, in genes that have accumulated many mutations, such as CYP2D6, a complete accounting of all functional polymorphisms is needed to permit prospective clinical applications including drug choice and selection of dosage. Because frequent null mutations can combine with less frequent mutations on the sister chromosome (compound heterozygosity), even low-frequency variants may be clinically important. Haplotyping may be effective in identifying key combinations of polymorphisms, but insufficient marker density and assumption of linkage over large sequences may lead to missed functional alleles. The standardization of nomenclature (Table 3) and integration of databases (Table 2) also remain ongoing challenges as the number of marker and functional alleles continues to increase. Appropriate meta-analyses can also provide a bird’s-eye view of the most important alleles within specific populations (e.g., Phillips et al., 2001). Although newer drugs may evade cytochrome metabolism, the examples given here demonstrate that many classes of genes must be pursued as potential contributors to interindividual differences in drug response, and efforts to identify their functional cis-regulatory variants will have lasting importance.

Acknowledgments

This work was supported in part by the NIH National Institute on Drug Abuse, DA018744, and an Ohio State University Distinguished University Fellowship.

References


Evert, B., Griese, E. U., & Eichelbaum, M. (1994). A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine
234 to proline exchange is associated with the poor metabolizer phenotype of sparteine. Naunyn Schmiedebers Arch Pharmacol 350(4), 434–439.


Shimoda, K., Someya, T., Morita, S., Hirokane, G., Yokono, A., Takahashi, S., et al. (2002). Lack of impact of CYP1A2 genetic polymorphism (C/A polymorphism at position 734 in intron 1 and G/A polymorphism at position-2964 in the 5’-flanking region of CYP1A2) on the


