Common CYP2D6 polymorphisms affecting alternative splicing and transcription: long-range haplotypes with two regulatory variants modulate CYP2D6 activity

Danxin Wang1,∗, Ming J. Poi2, Xiaochun Sun1, Andrea Gaedigk3, J. Steven Leeder3 and Wolfgang Sadee1

1Program in Pharmacogenomics, Department of Pharmacology, School of Medicine and 2Department of Pharmacy, The Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, The Ohio State University, Columbus, OH, USA and 3Division of Clinical Pharmacology and Therapeutic Innovation, Children’s Mercy Hospital, Kansas City, MO, USA

Received May 20, 2013; Revised August 1, 2013; Accepted August 22, 2013

Cytochrome P450 2D6 (CYP2D6) is involved in the metabolism of 20–25% of clinically used drugs. Genetic polymorphisms cause substantial variation in CYP2D6 activity and serve as biomarkers guiding drug therapy. However, genotype–phenotype relationships remain ambiguous except for poor metabolizers carrying null alleles, suggesting the presence of yet unknown genetic variants. Searching for regulatory CYP2D6 polymorphisms, we find that a SNP defining the CYP2D6∗2 allele, rs16947 [R296C, 17–60% minor allele frequency (MAF)], previously thought to convey normal activity, alters exon 6 splicing, thereby reducing CYP2D6 expression at least 2-fold. In addition, two completely linked SNPs (rs5758550/rs133333, MAF 13–42%) increase CYP2D6 transcription more than 2-fold, located in a distant downstream enhancer region (>100 kb) that interacts with the CYP2D6 promoter. In high linkage disequilibrium (LD) with each other, rs16947 and the enhancer SNPs form haplotypes that affect CYP2D6 enzyme activity in vivo. In a pediatric cohort of 164 individuals, rs16947 alone (minor haplotype frequency 28%) was associated with reduced CYP2D6 metabolic activity (measured as dextromethorphan/metabolite ratios), whereas rs5758550/rs133333 alone (frequency 3%) resulted in increased CYP2D6 activity, while haplotypes containing both rs16947 and rs5758550/rs133333 were similar to the wild-type. Other alleles used in biomarker panels carrying these variants such as CYP2D6∗41 require re-evaluation of independent effects on CYP2D6 activity. The occurrence of two regulatory variants of high frequency and in high LD, residing on a long haplotype, highlights the importance of gene architecture, likely shaped by evolutionary selection pressures, in determining activity of encoded proteins.

Cytochrome P450 2D6 (CYP2D6) is an important hepatic enzyme involved in the metabolism of 20–25% of clinically used drugs. Large inter-individual variation in CYP2D6 enzyme activity influences drug efficacy, drug dosages and adverse effects (1). Over 100 CYP2D6 allelic variants have been identified (http://www.imm.ki.se/CYPalleles), with CYP2D6 enzyme activity ranging from no activity to ultra-rapid activity. Four CYP2D6 metabolizer phenotypes have been defined among European populations, with poor metabolizers representing 5–10% of the population, intermediate metabolizer 10–15%, extensive metabolizer 75–85%, and ultra-rapid metabolizer 1–10% (2,3). Listed as a biomarker test in the US Food and Drug Administration’s Table of Pharmacogenomics Biomarkers in Drug Labels (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm), CYP2D6 genotypes are used clinically to predict CYP2D6 activity.

∗To whom correspondence should be addressed at: 333 West 10th Ave Graves Hall 5168, Columbus, OH 43210, USA. Tel: +1 6142927336; Fax: +1 6142927232; Email: wang.808@osu.edu or danxin.wang@osumc.edu

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
metabolizer phenotypes. The relevance of CYP2D6-guided therapy is highlighted by two Clinical Pharmacogenomics Implementation Consortium (CPIC) guidelines for CYP2D6 codeine and CYP2D6/tricyclic antidepressant gene/drug pairs (4,5). However, except for poor metabolizer phenotype, the relationship between CYP2D6 phenotype and genotype is still ambiguous. Because of uncertainty in phenotype prediction by genotype, individuals carrying two or only one fully active allele are all assigned the extensive metabolizer phenotype (4), in contrast to other cytochrome P450s such as CYP2C19 where individuals with only one fully active allele are considered intermediate metabolizers (6). As enzyme induction by environmental factors plays a lesser role in causing inter-individual differences in CYP2D6 activity compared with observed in liver samples (12) could have been caused by additive variants embedded in CYP2D6 (MAF) 17–60% in different populations (7), this suggests the existence of additional genetic variants in the CYP2D6 gene locus yet to be discovered. Because the phenotype affected by CYP2D6 variants can be accurately measured as drug levels, this gene provides an opportunity to study the missing portion of genetic factors that may also contribute to ambiguities at other gene loci where phenotypes are less well defined.

Among the numerous CYP2D6 variants, the frequent non-synonymous SNP rs16947 [R296C, minor allele frequency (MAF) 17–60% in different populations] defines the CYP2D6*2 allele, currently considered to convey wild-type enzyme activity (3,8). Previously, CYP2D6*2 or rs16947 has been associated in various in vitro or in vivo studies with either decreased, unchanged or enhanced CYP2D6 expression/activity (9–11). rs16947 marked CYP2D6*2 haplotypes may contain additional variants that could account for these discrepant results. The decreased expression/activity of *2 alleles was seemingly resolved by the discovery of intron 6 SNP rs28371725 (8,12), and a haplotype containing both rs16947 and rs28371725 was designated the *41 allele (see Supplementary Material, Table S1 for the combination of SNPs for each CYP2D6 star allele). The less frequent CYP2D6*4I has been consistently associated with decreased mRNA levels and enzyme activity compared with CYP2D6*1 or *2 (8,12), serving as a biomarker predicting intermediate metabolizer phenotype (4). However, except for one report (12), the underlying *41 mechanism remains uncertain, since any contribution of rs16947, always contained in *4I, is insufficiently characterized. Recombinant cDNA transfection showed the amino acid substitution (R296C) in rs16947 resulted in reduced enzyme activity (9,10), but other regulatory effects cannot be excluded. Increased mRNA expression associated with CYP2D6*2 observed in liver samples (12) could have been caused by additional variants embedded in *2. Multiple other CYP2D6 alleles also harbor rs16947, including *29, *17, *35, *45, *46. These alleles have been associated with reduced or wild-type enzyme activity or remain functionally uncharacterized.

CYP2D6 is also subject to alternative splicing, with two major splice variants (SVs) identified in human liver tissue, one lacking exon 3 (E3- SV) and another exon 6 (E6- SV) (13). CYP2D6*4I (containing both rs28371725 and rs16947, Supplementary Material, Table S1), have been associated with increased formation of the inactive E6- SV (12), but it is unclear whether the effect is caused by rs28371725 or rs16947 alone or the combination of both. Similarly, the CYP2D6 promoter SNP −1584 C>G (rs1080985) has been associated with increased CYP2D6 expression in human livers (14,15), but reporter gene assays did not support a transcriptional effect of rs1080985 (12,16). Moreover, several SNPs located over 100 kb up- or downstream of the CYP2D6 gene are significantly associated with CYP2D6 mRNA expression (17,18), but lack evidence of effects on transcription. Taken together, it remains uncertain how polymorphisms in regulatory regions impact CYP2D6 expression and function.

Here, we tested the hypothesis that frequent genetic variants have yet to be discovered that account for a portion of CYP2D6 variability. We demonstrate that rs16947 in CYP2D6*2, rather than rs28371725 in CYP2D6*4I, reduces CYP2D6 activity via increased non-productive splicing. Moreover, two SNPs located in an enhancer region >100 kb downstream of CYP2D6, rs5758550/rs133333, in high LD with rs16947, substantially increase CYP2D6 transcription. Enhancer SNPs counteract the reduced function of rs16947 in nearly half of rs16947 carrying haplotypes, while enhancer SNPs existing in the absence of rs16947 appear to convey an ultra-rapid CYP2D6 phenotype. These results could alter currently used CYP2D6 biomarker panels. They also illustrate how the architecture of an entire gene locus must be considered to characterize all main genetic effects.

RESULTS

SNP rs16947 (CYP2D6*2, R296C, in exon 6) affects CYP2D6 exon 6 splicing and decreases CYP2D6 mRNA level

Upon genotyping 12 SNPs (11 coding region SNPs and 1 promoter SNP, Supplementary Material, Fig. S1a), we measured the two major alternative SVs of CYP2D6, E3- SV lacking exon 3 and E6- SV lacking exon 6, in 58 human liver samples. rs16947 was the SNP most significantly associated with a 2-fold and a 3.3-fold increase in E3- and E6- SVs, respectively (Table 1). Since *4I has been reported to increase E6- SV formation (12), we compared E6- SV levels in livers carrying *2 (only rs16947) and *4I (rs16947 and rs28371725). Shown in Figure 1A, CYP2D6*2 carrier status revealed a significant gene dosage effect, with *2 homozygotes having highest E6- SV levels, followed by *2 heterozygotes and non-carriers. In contrast, CYP2D6*4I carrier status displayed a wider range of E6- SV levels in the non-carriers group and no significant differences between heterozygous and non-carriers (Fig. 1B). Moreover, E6- SV levels did not differ significantly between *2 and *4I carriers regardless of whether they were either homozygotes or heterozygotes (P > 0.05), a result that differs from those reported previously (12). These results suggest that the CYP2D6*2 allele (rs16947) alone accounts for altered E6- SV splicing, while the SNP characterizing CYP2D6*4I (rs28371725) may play a minor role at the most.

We constructed CYP2D6 genomic DNA expression plasmids (including all introns) with different genotypes/haplotypes (Supplementary Material, Table S2a), known to yield full-length CYP2D6 mRNA and functional CYP2D6 protein in transfected cell culture (12), for transfection into HEK293 cells. Each haplotype/allele was represented by several constructs with additional SNPs considered not to affect function (see Supplementary Material, Table S2a for sequence variations). Results obtained with four constructs representing the reference haplotype CYP2D6*1
clones 1a–d) were similar, and therefore, were combined. Consistent with the results obtained from liver samples, constructs containing rs16947 (clones 2a–c, 41a and 41b, Supplementary Material, Table S2a) produced higher levels of E3- and E6-SVs than the reference CYP2D6∗1 (Fig. 2A and B). Constructs representing CYP2D6∗4, known to cause a splicing defect of exon 3 (rs3892097, 1846G>A, clones 4a–c, Supplementary Material, Table S2a), decreased E3- SV, while increased formation of E6- SV was observed only with clone 4b, which also contains rs16947 (Fig. 2A and B). CYP2D6∗41 (also carrying rs16947) had effects similar to those observed for CYP2D6∗2, while CYP2D6∗10 (clone 10, P34S, 100C>T, rs1065852, Supplementary Material, Table S2a) had no detectable effect (Fig. 2). Similar results were obtained for E6- SV but not

<table>
<thead>
<tr>
<th>Splice variants</th>
<th>Predictor</th>
<th>Regression β (level relative to the CYP2D6∗1 reference)</th>
<th>P-value</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3- SV</td>
<td>2850T (+2, rs16947)</td>
<td>0.310 (2-fold)</td>
<td>0.011</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2988 A (+41, rs28371725)</td>
<td>0.27 (1.8-fold)</td>
<td>0.053</td>
<td>7</td>
</tr>
<tr>
<td>E6- SV</td>
<td>2850T (+2, rs16947)</td>
<td>0.512 (3.3-fold)</td>
<td>&lt;0.0001</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1661 G (rs1058164)</td>
<td>−0.392 (0.41-fold)</td>
<td>0.004</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>4180 G (rs1135840)</td>
<td>−0.392 (0.41-fold)</td>
<td>0.004</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2988 A (+41, rs28371725)</td>
<td>0.418 (2.6-fold)</td>
<td>0.007</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>−1584 G (rs1080985)</td>
<td>0.344 (2.4-fold)</td>
<td>0.014</td>
<td>11</td>
</tr>
</tbody>
</table>

Twelve frequent SNPs (rs1080985, rs1065852, rs28371706, rs1058164, rs5030655, rs3892097, rs28371720, rs5030656, rs16947, rs1135840, rs28371725 and rs59421388, see Supplementary Material, Fig. S1a SNP location) in CYP2D6 were genotyped; only those SNPs with significant associations to expression levels are listed.

Figure 1. Expression of CYP2D6 SV without exon 6 (E6-SV) grouped by CYP2D6∗2 (rs16947, panel A) and CYP2D6∗41 (rs16947 and rs28371725, panel B) haplotypes, in human liver samples. Samples with other genotypes were excluded. Heter, heterozygous carriers; homo, homozygous carriers. The differences between groups were analyzed by t-test.

Figure 2. Formation of SVs lacking exon 3 (E3-SV) (A) or exon 6 (E6-SV) (B) in HEK cells after transfection with CYP2D6 genomic DNA expression plasmids containing different genotypes/haplotypes. E3- SV and E6- SV were quantitated with PCR using fluorescently labeled primers, shown as percent of the total amount of CYP2D6 transcript. Data are mean ± SD of two independent experiments each performed with triplicates obtained from different clones of each construct: 1, combined results from four clones with CYP2D6∗1 reference sequence; 2a–2c, 3 CYP2D6∗2 clones; 4a–4b, 3 CYP2D6∗4 clones; 41a–41b, 2 CYP2D6∗41 clones; 10, 1 CYP2D6∗10 clone (Supplementary Material, Table S2a lists sequence variation). Compared with CYP2D6∗1, * P < 0.01, one-way ANOVA with Dunnett post-test.
E3- SV upon transfection into HepG2 cells (Supplementary Material, Fig. S2). These results also support the hypothesis that rs16947 rather than rs28371725 is mainly responsible for altered E6 and E3 splicing.

E3- SV is an in-frame deletion of exon 3 shortening the translated protein product by 51 amino acids, while E6- SV encodes mRNA lacking 142 bases of exon 6 which results in a frame shift and creation of a premature termination codon 17 bases downstream. To test a role for nonsense-mediated RNA decay of the out-of-frame mRNA, we suppressed nonsense-mediated RNA decay with puromycin or cycloheximide, or with knockdown of nonsense-mediated RNA decay factor UPF1 using siRNA transfection (19), in HepG2 cells endogenously expressing the CYP2D6*1 genotype. All three treatments increased E6- SV mRNA levels 2–3-fold but did not affect E3- SV (Supplementary Material, Fig. S3a), indicating that E6- SV, but not E3- SVs, undergoes nonsense-mediated RNA decay. The half-life of full-length CYP2D6 mRNA and E6- SV, measured over 24 h, was estimated to be 4.8 and 2.3 h, respectively (Supplementary Material, Fig. S3b). These results indicate that rs16947 reduces full-length CYP2D6 mRNA and increases the formation of E6- SV, which is then degraded via nonsense-mediated RNA decay.

To estimate the magnitude of the net effect of rs16947 and to compare the effects of rs16947 alone (1) versus rs16947 plus rs28371725 (41) on full-length CYP2D6 mRNA levels, we co-transfected CYP2D6*1 and CYP2D6*2 or CYP2D6*1 and CYP2D6*41 genomic DNA expression constructs into HEK293 cells and measured allelic CYP2D6 mRNA expression using rs16947 as a marker SNP. Allelic RNA expression is a superior quantitative measure than total RNA for predicting the functional consequences of allelic variation as we have demonstrated in numerous publications (20–23), controlling for transacting factors that might affect mRNA expression in tissue samples or variations of inter-well transfection efficiency and amount of DNA being transfected in cell culture. Because rs16947 is located in exon 6, which is absent in E6- SV, the measured mRNA level represents full-length CYP2D6 mRNA when using rs16947 as marker. After normalization to the transfected plasmid DNA ratio, the allelic RNA ratios (C/T) were 2.8 ± 0.1 for CYP2D6*1/CYP2D6*2 co-transfection, and 2.4 ± 0.3 for CYP2D6*1/CYP2D6*41 co-transfection. This result supports the notion that rs16947 in *2 reduces CYP2D6 full-length mRNA at least 2-fold, and the combination of rs16947 and rs28371725 in *41 has similar effects on CYP2D6 mRNA expression compared with rs16947 alone.

Enhancer region SNPs increase CYP2D6 mRNA transcription

Chromatin profiling in HepG2 cells (ENCODed data) (24) indicates that CYP2D6 has a weak promoter, while two strong enhancer regions were identified at 126 kb upstream and 114 kb downstream of CYP2D6, respectively. Using chromatin immunoprecipitation (CHIP) on p300 antibody (Fig. 3A). Physical interactions between CYP2D6 promoter and enhancers were tested with chromatin conformation capture (3C) assays (25). Screening a 300 kb region surrounding the CYP2D6 promoter, the downstream enhancer showed the highest crosslinking frequency with the CYP2D6 promoter (Fig. 3B), indicating that it physically interacts with the CYP2D6 promoter.

Two completely linked SNPs, rs5758550 and rs133333, were found within the downstream enhancer region. Promoter reporter gene assays showed the minor SNP alleles conveyed substantially higher promoter activity compared with the reference allele (Fig. 3C). SNPs rs5758550/rs133333 are in substantial LD with the promoter SNP rs1080985 (−1584C>G, Supplementary Material, Fig. S1b), which had been tentatively associated with enhanced CYP2D6 mRNA/protein expression whereas reporter gene assays failed to confirm this (12,16). Using a 4 kb upstream region (corresponding to 4–4462 bp in Ay545216) containing rs1080985 (−1584G) and additional SNPs (Supplementary Material, Table S2b), we found that a haplotype containing rs1080985 (construct m3) had similar effects on transcription as CYP2D6*1 (Supplementary Material, Fig. S4). This finding is consistent with previous results suggesting a lack of effect of rs1080985 on gene transcription (16) and indicates that the observed association with increased CYP2D6 protein levels results from LD with the upstream enhancer SNPs.
CYP2D6 mRNA expression is determined by both rs16947 and enhancer SNPs

Enhancer SNPs rs5758550/rs133333 (MAF 13–42% in different populations, Supplemental Material, Table S3) are in high LD with rs16947 (MAF 33%) in Caucasian population (Supplementary Material, Fig. S1b and Table S3). Only combined consideration of both rs16947 and rs133333/rs5758550 can fully account for the pattern of allelic CYP2D6 mRNA expression imbalance we have observed in liver samples (all were from Caucasians except one) (Fig. 4A). In livers heterozygous only for rs16947, the minor T allele was consistently associated with decreased mRNA levels (RNA ratio ≈ 2-fold compared with the CYP2D6*1 reference (showing RNA ratios > 1)). Hence, the enhancer SNPs override the reduction in CYP2D6 mRNA expression conveyed by rs16947. One liver sample that did not carry either rs16947 (CC) or enhancer SNPs (AA) showed no allelic expression imbalance (RNA ratio close to 1), while a liver homozygous for rs16947 (CT) and heterozygous for the enhancer SNPs (GA) had the highest allelic RNA ratio (3-fold). Neither rs16947 nor the enhancer SNPs alone were associated with total CYP2D6 mRNA expression (n = 54, P > 0.05). Multiple linear regression analyses showed, however, that rs16947 was significantly associated with a 2.4-fold decrease (P = 0.011) and the enhancer SNPs with a 2-fold increase in RNA levels (P = 0.036) (Fig. 4B and C). These results indicate that CYP2D6 mRNA expression is affected by both rs16947 and enhancer SNPs, in opposite directions.

In vivo effects of rs16947 and enhancer SNPs

We tested genotype/haplotype effects on metabolic activity in a cohort of 164 children with/without attention deficit-hyperactivity disorder (ADHD) (see Supplementary Material, Table S4 for demographics). CYP2D6 enzyme activity is reflected by the urinary metabolic ratio of dextromethorphan/metabolites (DM/DX). Ten individuals carrying two null alleles constituting a poor metabolizer phenotype were removed from further analysis, leaving an effective study sample of 154. CYP2D6 genotype frequencies and rs16947/enhancer SNP haplotypes are shown in Supplementary Material, Table S5a and b. Of all rs16947 containing alleles (excluding individuals with gene duplication), 68% also carried enhancer SNPs, while 28% carried enhancer SNPs, with 8% carried rs16947 only.

Figure 4. Allelic and total CYP2D6 mRNA expression in liver samples. (A) Allelic RNA or DNA ratios were measured in RNA and gDNA using markers (rs1065852, rs1058164 and rs16947), normalized to gDNA ratios, and expressed as ratio T/C (minor/major, rs16947) using a log10 scale. All samples were heterozygous for rs16947 (C/T) except two that were homozygous for reference (CC) or variant (TT) alleles as indicated. Livers carrying allele CYP2D6*4 were excluded. Data are mean ± SD, two independent experiments with duplicates. (B and C) Total CYP2D6 mRNA level grouped by rs16947 (B, adjusted by rs5758550/rs133333 genotype) or rs5758550/rs133333 (C, adjusted by rs16947 genotype) genotypes. The differences between groups were analyzed by multivariate regression.
though hepatic mRNA levels are elevated (Fig. 4A). This ∗
CYP2D6 haplotype carriers did not exceed rs16947. These results parallel the allelic mRNA ratios observed
SNPs compensated for the reduced function conveyed by
rs16947 reducing CYP2D6 enzyme activity while the enhancer
and rs16947 plus enhancer SNPs had similar ratio compared with
higher log-transformed DM/DX ratios (lower enzyme activity),
carriers without enhancer SNPs rs133333/rs5758550 had
Table S6). In model 2, we removed ∗
CYP2D6 as reported (9,10).

In individuals with a single functional CYP2D6 allele (hetero-
yzous for null alleles CYP2D6*3, *4, *5, *6 or *40), rs16947
carriers without enhancer SNPs rs133333/rs5758550 had
higher log-transformed DM/DX ratios (lower enzyme activity),
and rs16947 plus enhancer SNPs had similar ratio compared with
reference genotype carriers (Fig. 5). These results indicate that
rs16947 reduced CYP2D6 enzyme activity while the enhancer
SNPs compensated for the reduced function conveyed by
rs16947. These results parallel the allelic mRNA ratios observed
in the liver tissues of rs16947 only carriers. On the other hand, the
in vivo enzyme activity of rs16947 and rs133333/rs5758550
haplotype carriers did not exceed CYP2D6*1 activity, even
though hepatic mRNA levels are elevated (Fig. 4A). This result
may be related to the reduced enzyme activity of the
 altered protein encoded by the non-synonymous SNP rs16947,
as reported (9,10).

We then assigned a CYP2D6 metabolic activity score for each
individual based on genotype using two models. In model 1, the
CYP2D6 activity score was assigned according to guidelines
based on accepted standards (4) (Supplementary Material,
Table S6). In model 2, we removed CYP2D6*2, *41, *29, *35
and *45 while adding rs16947 and enhancer SNPs rs133333/
rs5758550 as markers. As expected, the activity scores using
both model 1 and model 2 were significantly associated with the
log (DM/DX) (∗P < 0.001). In model 2, many individuals
were reassigned to activity score groups predicting higher activity
and thereby improving the linearized relationship with DM/
DX ratios. Specifically, one individual designated as ultra-
rapid metabolizer (activity score = 2.5) using model 1 was re-
classified as extensive metabolizer (activity score = 2) and
seven individuals (all with enhancer SNPs) are now designated
as ultra-rapid metabolizer (score > 2, Fig. 6A and B), bringing
the total number of ultra-rapid metabolizers to 12 subjects.
Moreover, model 1 yielded a significant difference between IM
versus EM phenotypes (∗P < 0.0001) but not EM versus
UM groups (∗P = 0.10). In contrast, model 2 gave significant differ-
ences between all three groups (IM versus EM, ∗P < 0.0001;

Figure 5. Values of log (DM/DX) in individuals with haplotypes containing
CYP2D6 rs16947 alone, reference genotypes or rs16947 in combination with
the enhancer SNPs. Selected individuals had one null allele and one functional
CYP2D6 gene to facilitate the detection of changes imparted by the enhancer
SNPs. The differences between groups were analyzed by t-test. Data are pre-
sented as box plot. The box and horizontal line show the 25th and 75th percentiles
and mean, and whiskers show the minimum and maximum values.

DISCUSSION

While the CYP2D6 gene locus has been extensively studied,
regulatory polymorphisms affecting activity have remained
poorly characterized. We have now identified two distinct mechanisms regulating CYP2D6 expression: rs16947 reduces
CYP2D6 activity by altering exon 6 splicing and downstream enhan-
er SNPs rs5758550/rs133333 enhance transcription. The present study illustrates a robust approach to determine the
effects of a distant enhancer region on its protein-coding target
gene, involving associations with gene expression, analysis of
allelic mRNA expression exclusively testing cis-acting factors,
chromatin-immunoprecipitation and chromatin conformation
capture (3C) assays. Separated by a large distance in the CYP2D6 locus, these two regulatory variants are nevertheless
in high LD with each other, a strong indicator of evolutionary sele-
ction pressures shaping this gene’s architecture. If under posi-
tive selection, such two-allele relationships presented here are
likely also to occur in other gene loci. CYP2D6 activity was
determined by haplotypes containing either rs16947 alone
(decreased activity, 28% minor haplotype frequency), rs16947
plus rs133333/rs5758550 (similar to the wild-type) or the enhan-
cer SNPs alone (3% frequency, with enhanced activity). Each of
these alleles may carry less frequent additional variants that need
to be re-assessed as to their functional activity. Incorporating these findings into biomarker panels will enable predicting
CYP2D6 activity more accurately compared with those current-
lly utilized.

Enhanced exon 6 splicing by rs16947

The CYP2D6*2 allele is considered to be a fully functional allele
(4,5), although a recent pharmacokinetic study suggested that it
is associated with reduced activity, at least with dextromethor-
phan as the probe substrate, possibly due to the R296C substitu-
tion caused by rs16947 (26). Our results, however, demonstrate
that rs16947 decreases CYP2D6 mRNA levels by enhancing
exon 6 skipping, which abrogates CYP2D6 activity. The loca-
tion of rs16947 in exon 6 is consistent with reports demonstrating
that disruption of exonic splicing enhancer sites promotes exon
skipping (27). In other studies, however, rs16947 did not alter, or
was associated with, increased mRNA expression in vivo com-
pared with CYP2D6*1 (8,12,15), suggesting the presence of
yet unidentified additional regulatory variants.

Far-downstream enhancer SNPs rs5758550/rs133333

Data from The Encyclopedia of DNA Elements (ENCODE)
project indicate that distal enhancers can regulate gene
expression via long-range interactions (28). We identified two distal enhancers, one upstream and one downstream of CYP2D6, but only the downstream enhancer physically interacts with the proximal CYP2D6 promoter region in a 3C assay. To investigate the effect of such a distal enhancer site on its target gene, we employed allelic RNA expression analysis, a powerful tool to identify cis-acting regulatory variants (20–23), regardless of the distance between regulatory site and target gene. This approach identified two completely linked SNPs, rs5758550/rs133333, which fully accounted for increased CYP2D6 mRNA expression observed by allelic mRNA analysis in livers. More than half of the subjects carrying rs16947 present in CYP2D6*2 also contained rs5758550/rs133333, accounting for increased mRNA expression in vivo reported previously (12) and in this study. On the other hand, the amino acid substitution R296C caused by rs16947 was found to be associated with 30–50% reduced enzyme activity (9,10,29). This could resolve the discrepancy between enhanced mRNA expression of CYP2D6*2 alleles containing the enhancer SNPs, with unchanged enzyme activity in vivo compared with CYP2D6*1. Enhancer SNP carriers without rs16947 (3% frequency) have increased mRNA expression, and by implication enzyme activity, but because of the relatively small number of subjects with the downstream enhancer SNPs alone in our study, the in vivo effect requires further investigation.

Figure 6. Relationship between log (DM/DX) and activity scores derived from genotype data for model 1 (A) and model 2 (B) (see Supplementary Material, Table S6 for additional details for models 1 and 2). Data are presented as box plot. The box and horizontal lines show the 25th and 75th percentiles and mean, and whiskers show the minimum and maximum values. (C) Log (DM/DX) for individuals with CYP2D6*1/*2 and *1/*41 genotypes. Open circles indicate the mean log (DM/DX). Note: all individuals with CYP2D6*1/*2 genotype also carried enhancer SNPs except two (highlighted by rectangle), while none of the individuals with *1/*41 genotype carried enhancer SNPs except two (oval circle). These outliers exemplify cases where the CYP2D6*2 and *41 classification fails accurately to predict metabolizer status.

Regulatory variants shown not to alter CYP2D6 expression: rs8138080, rs1080985 and CYP2D6*41 SNP rs28371725

Previous reports have revealed a significant association between CYP2D6 mRNA levels and another far downstream SNP, rs8138080 (17,18), which is in high LD (D′ = 1) with rs5758550/rs133333, located ~10 kb further downstream, but present at higher MAF (~40%). Discrepancies with the observed allelic mRNA ratios in livers rule out rs8138080 as a causative variant. Moreover, in agreement with previous studies (12,16), we did not find changes in promoter activity with rs1080985 (−1584 C>G); this SNP was not significantly associated with CYP2D6 mRNA levels even after adjusting for rs16947. The higher CYP2D6 mRNA expression levels previously associated with −1584G (15) can be accounted for by partial LD with the rs5758550/rs133333 enhancer SNPs (Supplementary Material, Fig. S1b), again attesting to the very long haplotypes characteristic of the CYP2D6 locus.

CYP2D6*41 containing rs16947 and another intron 6 SNP, rs28371725, were considered to convey reduced CYP2D6 mRNA/protein activity (8,12). However, we show here that CYP2D6*41 does not appear to have significant independent effects on CYP2D6 mRNA expression other than that exerted by rs16947. Previously, CYP2D6*41 has been associated with an increased level of E6-SV resulting in reduced CYP2D6...
expression and function (12). Our results provide evidence that this effect was largely caused by rs16947 alone rather than that the intron 6 SNP rs28371725 or the combination of rs16947 and rs28371725. Distinct from CYP2D6*2 haplotypes that often contain enhancer SNPs, CYP2D6*41 haplotypes largely exclude the downstream enhancer SNPs (Supplementary Material, Table S5b). Therefore, CYP2D6*41 can serve as a tagging SNP for identifying haplotypes containing only rs16947 without enhancer SNPs, thereby predicting an intermediate metabolizer phenotype. However, the LD between CYP2D6*41 SNP rs28371725 and enhancer SNPs likely differs between populations, rendering this allele unsuitable as a surrogate marker. This contention is supported by our finding that two CYP2D6*1/*41 subjects (both are African Americans) also carrying enhancer SNPs had the highest CYP2D6 enzyme activity in the CYP2D6*1/*41 group, while two CYP2D6*1/*2 subjects without enhancer SNPs had the lowest enzyme activity among all CYP2D6 *1/*2 subjects (Fig. 6C).

Relevance of gene loci with two or more functional variants with high LD in long-range haplotype blocks

The use of robust phenotypes, such as allelic mRNA ratios and drug levels, has enabled the discovery of hidden regulatory variants in CYP2D6 that reveal a unique architecture of this gene locus: both regulatory variants are frequent, more than 100 kb apart, and yet in high LD. These characteristics are a hallmark of evolutionary selection pressures likely to impinge on CYP enzymes as a primary defense against environmental toxins (30). Such genes are likely to accumulate several adaptive mutations, and certainly, CYP2D6 ranks as one of the most polymorphic genes. The example revealed here suggests selection of a haplotype of two regulatory variants with opposing effects in LD with each other but with different allele frequencies, consistent with co-evolution between these distant variants. Could this phenomenon be pervasive among genes under strong selection pressures that are modified by changing environmental factors? We have identified regulatory variants in a number of genes, discovering similar long-range allele interactions in two additional gene loci: CETP encoding cholesterylester transfer protein (31) and CHRNA5 encoding the alpha 5 nicotinic receptor (32). In each of these gene loci, frequent distant enhancer region variants (6–15 kb upstream) occur in strong negative LD with a SNP in the coding region affecting splicing (CETP) or with a non-synonymous SNP altering protein function (CHRNA45), in each case separated by 20–30 kb. Understanding the dynamic interactions between these variants, either on the same or on opposite haplotypes, is critical to revealing the overall effect of a gene locus on relevant phenotypes. Occurrence of these interactions at three gene loci among less than 15 studied in sufficient detail in our group suggests that distal regulatory variants are frequent and may form haplotypes with unique properties, particularly in genes under evolutionary selection. Incorporating such gene architectures into GWAS analyses is critical for correctly estimating the contributions of genetic factors to human phenotypes.

Clinical implications

Guidelines published by the Clinical Pharmacogenomics Implementation Consortium (CPIC) have been developed to predict metabolizer status for CYP2D6 genetic tests for codeine and tricyclic antidepressants therapy (4,5). The guideline includes several rs16947 containing alleles/haplotypes as biomarkers in addition to CYP2D6*2 and *41, namely CYP2D6 *17, *29, *35 and *45. CYP2D6*35 and *45 are currently considered to convey normal enzyme activity with activity score of 1 (4,5), but being in different LD with rs16947 and downstream enhancer SNPs (Supplementary Material, Table S5b), their categorizations without taking the rs16947 and enhancer SNPs into account may lead to over- or under-predicting CYP2D6 activity. Similarly, CYP2D6*29 (V136I and V338M) had little effect on CYP2D6 enzyme activity in cDNA transfected cell culture (33) but is considered a reduced function allele with activity score of 0.5 (4,5,34), possibly confounded by rs16947 and downstream enhancer SNPs. Most (86%) CYP2D6*29 alleles carry rs16947 without the enhancer SNP, while 14% have both (Supplementary Material, Table S5b). On the other hand, CYP2D6 activity conveyed by the CYP2D6*17 allele appears to be mainly determined by its hallmark T107I substitution, because cDNA transfections have revealed over 70% decreased CYP2D6 enzyme activity (9,10,29,35). Most CYP2D6*17 haplotypes (93%) contain both rs16947 and the downstream enhancer SNPs (Supplementary Material, Table S5b). The few CYP2D6*17 alleles (7%) containing only rs16947 may cause even greater reduction in enzyme activity. We conclude that at present, CYP2D6*2, *41, *29, *35 and *45 are inadequate for inferring CYP2D6 activity in vivo, since CYP2D6 activity is affected by haplotypes containing rs16947 and the downstream enhancer SNPs rs5758550/rs133333. Therefore, we suggest that the clinical CYP2D6 biomarker panel be redesigned and simplified to enhance predictive accuracy. We propose to classify CYP2D6*9, *10, *17 and alleles carrying only rs16947 as reduced function alleles, alleles carrying both rs16947 and enhancer SNPs as wild-type function alleles, and alleles carrying only the enhancer SNPs rs5758550/rs133333 as having increased function (Supplementary Material, Table S6). Predictions from this model result in a better fit for the measured metabolic activities (Fig. 6), and importantly, appear more accurately to identify subjects as ultra-rapid metabolizers with an overall activity score >2. Before this revised model is utilized, however, our findings warrant replication in larger cohorts.

In summary, we have identified and functionally characterized two regulatory variants in CYP2D6, the exon 6 SNP rs16947 and two downstream enhancer SNPs rs575550/ rs133333. Haplotypes comprising these two loci determine the enzyme activity of CYP2D6 in vivo, suggesting that their inclusion into genotyping panels will enable predicting CYP2D6 activity with greater accuracy.

MATERIALS AND METHODS

Tissue, DNA and RNA samples

Fifty-eight human liver biopsy samples (55 Caucasians and 3 African Americans) were obtained from the Cooperative Human Network Midwestern and Western Division, under the approval of The Ohio State University Institutional Review Board. Preparation of genomic DNA and RNA from livers or cell cultures was performed as described (20). cDNA was prepared using gene specific primers and oligo(dT) (20).
Genotyping
SNPs were genotyped using multiplex (36) or single primer extension assays using primers shown in Supplementary Material, Table S7. The pediatric study cohort was genotyped as described earlier (3) and included testing for gene duplications (37), the presence of hybrid genes (38,39) and copy number variation analysis (40). Genotype analysis for rs5758550 and rs133333 was carried out using TaqMan assays obtained from Life Technologies (Foster City, CA, USA) using gDNA isolated with a Qiagen Blood Mini Kit.

CYP2D6 total mRNA and SVs quantification
Total CYP2D6 mRNA or SVs were measured using real-time PCR with specific primers (Supplementary Material, Table S7) and SYBR PCR master mix (Life Technologies) using PCR with specific primers (Supplementary Material, Table S7) for total CYP2D6 mRNA and SVs quantification. All constructs driven by a TK promoter were co-transfected with promoter constructs at 1:5 ratio.

Cell culture and transfection
Cells were cultured at 37°C in a humidified incubator at 5% CO₂ in DMEM (HepG2) or DMEM/F12 (HEK293) medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin. The day before transfection, cells were plated into 12-well plates. Transfection of plasmid DNA was performed using lipofectamine 2000 (HEK) (Invitrogen Life Technologies, Carlsbad, CA, USA) or Fugene HD transfection reagent (HepG2). Transfection of UPF1 siRNA was performed using TransIT-TKO transfection reagent (Mirus, Madison, WI, USA). Cells were harvested 48 h after transfection for total RNA preparation. Some cells were treated with puromycin (30 µg/ml) or cycloheximide (10 µg/ml) for 6 h, or actinomycin D (20 µg/ml) for different time periods before harvesting, and RNA was prepared (20). In genomic CYP2D6 DNA expression plasmid transfection experiments, mRNA was purified from total RNA using PolyATtract mRNA isolation systems (Promega) to avoid contamination of total RNA with plasmid DNA.

Quantitative analysis of allelic ratios in genomic DNA and mRNA using a primer extension method (SNaPShot)
SNaPShot assays were performed as described (20). Specific primers shown in Supplementary Material, Table S7 served to amplify nearly full-length CYP2D6 mRNA. The entire CYP2D6 gene (5.1 kb) amplified from genomic DNA served as internal control to normalize allelic RNA ratios. CYP2D6 copy number variations were not detectable in the livers tested by SNaPShot assays and real-time PCR (42). Samples heterozygous for at least one of the three marker SNPs rs1065852 (100C>T, P34S), rs1058164 (1661G>C) and rs16947 (2850C>T, R296C) were selected for measuring allelic RNA expression. Results from each marker gave similar results for the same liver and were combined.

CYP2D6 genomic DNA expression and reporter gene constructs
The entire CYP2D6 gene (5.1 kb, including introns) was PCR amplified with high fidelity Advantage HD polymerase (Clontech, Mountain view, CA, USA) from liver DNA samples with different genotypes and cloned into pcDNA3 vector using Infusion cloning kit (Clontech) according to the manufacturer’s protocol. Fragments of the promoter region (4.3 kb) and downstream enhancer (3.1 kb) regions were also PCR amplified from liver DNA samples and cloned into pGLO3 vectors (see Supplementary Material, Table S7 for PCR primers). All constructs were sequenced to ensure absence of any random mutations during PCR amplification.

Reporter gene assay
Cells were transfected with 1 µg plasmid DNA and luciferase activities measured 48 h post-transfection with Dual-Glo luciferase assay kit (Promega, Madison, WI, USA) on a fluorescence plate reader (PerkinElmer Life and Analytical Science, Waltham, MA, USA). As internal controls, Renilla luciferase

Data analysis
Haplotype structure and LD plots were generated using HelixTree (Golden Helix, Bozeman, MT, USA) and Haploview. Associations between genotype and the level of CYP2D6 total mRNA
expression or SVs were estimated with a linear regression model using Minitab software. Statistical analysis of allelic DNA and RNA ratios or difference in means was performed using Prism (GraphPad Software Inc., San Diego, CA, USA). Data are expressed as mean ± SD.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by National Institute of Health (NIH) grants (U01 GM 092655 to D.W. and W.S., R01 HD 058556 to J.S.L.), and partially by grant from the National Center for Research Resources (UL1RR025755). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institutes of Health.

REFERENCES


