Genetic factors affecting gene transcription and catalytic activity of UDP-glucuronosyltransferases in human liver

Wanqing Liu1,3, Jacqueline Ramírez1, Eric R. Gamazon1, Snezana Mirkov1, Peixian Chen1, Kehua Wu1, Chang Sun2, Nancy J. Cox1,2, Edwin Cook Jr4, Soma Das2 and Mark J. Ratain1,∗

1Department of Medicine and 2Human Genetics, The University of Chicago, Chicago, IL 60637, USA, 3Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, West Lafayette, IN 47907, USA and 4Department of Psychiatry, The University of Illinois at Chicago, Chicago, IL 60612, USA

Received January 10, 2014; Revised May 7, 2014; Accepted May 27, 2014

The aim of this study was to discover cis- and trans-acting factors significantly affecting mRNA expression and catalytic activity of human hepatic UDP-glucuronosyltransferases (UGTs). Transcription levels of five major hepatic UGT1A (UGT1A1, UGT1A3, UGT1A4, UGT1A6 and UGT1A9) and five UGT2B (UGT2B4, UGT2B7, UGT2B10, UGT2B15 and UGT2B17) genes were quantified in human liver tissue samples (n = 125) using real-time PCR. Glucuronidation activities of 14 substrates were measured in 47 livers. We genotyped 167 tagSNPs (single-nucleotide polymorphisms) in UGT1A (n = 43) and UGT2B (n = 124), as well as the known functional UGT1A1∗28 and UGT2B17CNV (copy number variation) polymorphisms. Transcription levels of 15 transcription factors (TFs) known to regulate these UGTs were quantified. We found that UGT expression and activity were highly variable among the livers (median and range of coefficient of variations: 135%, 74–217% and 52%, 39–105%, respectively). CAR, PXR and ESR1 were found to be the most important trans-regulators of UGT transcription (median and range of correlation coefficients: 46%, 6–58%; 47%, 9–58%; and 52%, 24–75%, respectively). Hepatic UGT activities were mainly determined by UGT gene transcription levels. Twenty-one polymorphisms were significantly (FDR-adjusted P < 0.05) associated with mRNA expression and/or activities of UGT1A1, UGT1A3 and UGT2B17. We found novel SNPs in the UGT2B17 CNV region accounting for variability in UGT2B17 gene transcription and testosterone glucuronidation rate, in addition to that attributable to the UGT2B17 CNV. Our study discovered novel pharmacogenetic markers and provided detailed insight into the genetic network regulating hepatic UGTs.

INTRODUCTION

UDP-glucuronosyltransferases (UGTs) are an important group of Phase II (conjugative) metabolizing enzymes that play a critical role in human health and disease. By catalyzing the formation of hydrophilic glucuronides, UGTs are involved in the metabolism and detoxification of numerous endogenous compounds and xenobiotic chemicals including therapeutic agents. Typical UGT substrates include bilirubin, bile acids, sex steroids, thyroid hormones, fatty acids, tobacco smoke carcinogens, dietary components, environmental toxins and pollutants, and a wide variety of prescribed drugs (and metabolites) including morphine, tamoxifen, vorinostat, SN-38, aromatase inhibitors (anastrozole, exemestane and letrozole), ciprofibrate, acetaminophen and mycophenolic acid (1–6). Reduced UGT activity causes or increases the risk of many human disorders, e.g. Gilbert’s syndrome and Crigler–Najjar syndrome (3). Recent genome-wide association studies (GWAS) revealed that the UGT2B17 copy number variation (CNV) is associated with osteoporosis (7), while genetic variants in the UGT1A locus are associated with urinary bladder cancer (8) and gallstone formation (9,10).

*To whom correspondence should be addressed at: The University of Chicago, 5841 S. Maryland Ave., MC 2115, Chicago, IL 60637, USA. Tel: +1 7737024400; Fax: +1 7737023969. Email: mratain@medicine.bsd.uchicago.edu

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
The human UGTs are encoded by a family of 19 protein-coding genes that are classified into two subfamilies, UGT1A and UGT2. The former is located on chromosome 2q37 and consists of nine active genes (UGT1A1, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9 and UGT1A10), while the latter is located on 4q13 and is further subdivided into UGT2A (UGT2A1, UGT2A2 and UGT2A3) and UGT2B (UGT2B4, UGT2B7, UGT2B10, UGT2B11, UGT2B15, UGT2B17 and UGT2B28) (11,12). The genomic organization and transcription mechanism of these UGT genes have been well documented. Briefly, UGT1A genes are organized in a tandem array with unique exons followed by common exons 2–5. In contrast, UGT2A and UGT2B genes are organized and transcribed individually (12). The expression of human UGT genes is tissue-specific. While human UGT2As are predominantly expressed in the olfactory epithelium, UGT1As and UGT2Bs are mainly distributed in aerodigestive tissue (1,13), and play an important role in drug metabolism. As the most important organ and tissue for drug metabolism, the liver significantly expresses five UGT1As (UGT1A1, UGT1A3, UGT1A4, UGT1A6, and UGT1A9) and five UGT2Bs (UGT2B4, UGT2B7, UGT2B10, UGT2B15 and UGT2B17) (1,13).

Glucuronidation is a common drug-metabolizing reaction. Thirty-five percent of the drugs metabolized by Phase II drug-metabolizing enzymes undergo metabolism by UGTs (2). For this reason, UGTs are important subjects of pharmacogenetic studies, among which a well-known example is the UGT1A1 gene where a common promoter (TA)n polymorphism (UGT1A1*28) significantly decreases UGT1A1 gene transcription, leading to reduced glucuronidation and increased toxicity of SN-38, the active metabolite of irinotecan (14,15). A recent GWAS has also found that UGT1A1*28 is associated with unconjugated hyperbilirubinemia in patients receiving tocilizumab (16). As another example, UGT2B17 CNV was found to be a causal variant for graft-versus-host disease after transplantation (17). To date, a large number of DNA polymorphisms have been identified in both UGT1A and 2B genes, many of which have been demonstrated to affect gene transcription and/or catalytic activity. However, the findings thus far only account for limited phenotypic variation in these UGTs, suggesting additional genetic factors affecting their functions remain unidentified (18). In addition to regulation by sequence variations within the loci, UGTs are substantially modulated by many transcription factors (TFs) and possibly by demographic and environmental factors as well (19–21). To date, no study has been conducted towards a systematic evaluation of the influence of these factors in UGT gene function.

Using a collection of human liver tissue samples, our study aimed to identify major cis- and trans-regulating factors conferring variations in hepatic UGT gene transcription and activity. The effect of demographic factors was also investigated. We quantified mRNA levels of 10 hepatic UGTs and 15 TFs known to regulate UGTs. The microsomal activity of 14 substrates for these UGTs was also quantified. Tagged single-nucleotide polymorphisms (tagSNPs) in loci from both gene families were genotyped. Using integrated data analysis, novel UGT single-nucleotide polymorphisms (SNPs) and major TFs significantly associated with UGT gene transcription and/or activity were identified.

RESULTS

UGT and TF mRNA expression

Hepatic UGT and TF mRNA expression was measured using real-time PCR. Given the unique organization of the UGT1A genes (individual exon 1s and shared exons 2–5), their primers were designed based on unique forward primers specific to each individual exon 1 and a universal reverse primer selected from exon 2 (Supplementary Material, Fig. S1). UGT2B genes are organized individually and primers were thereby designed based on the unique sequence of each gene. The specificity of these primers was confirmed by sequencing of PCR products. To normalize the UGT and TF transcript levels, we quantified and compared the variability in four housekeeping genes (Supplementary Material, Table S1). The TRP gene was selected as the best internal control gene as it had the lowest inter-individual variability (CV = 42%) among all housekeeping genes. Normalized UGT expression was highly variable among the population (median and range of the CVs, 135%, 74–217%). Compared with UGTs, TFs had much lower variability (65%, 30–137%). Variability in the relative expression of each UGT and TF genes is summarized in Supplementary Material, Table S1. UGT2B17 and UGT2B4 were the most and least variable UGTs, respectively.

UGT activity

The 14 substrates used in this study are metabolized by one major UGT (e.g. bilirubin and serotonin) or by multiple UGTs (e.g. acetaminophen) according to our experiments and reports from the literature (4,22–40). Hepatic UGTs identified to be involved in glucuronidation of each substrate are listed in Table 1. UGT activities were highly variable among livers (median and range of coefficient of variations: 52%, 39–105%). Among all glucuronidation activities measured, testosterone glucuronidation showed the highest variability (CV = 105%; Supplementary Material, Table S1).

SNP genotyping

TagSNPs were selected based on linkage disequilibrium (LD) (r² ≥ 0.8) using the HapMap data and resequencing data

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Major UGT</th>
<th>Minor UGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN-38</td>
<td>UGT1A1</td>
<td>UGT1A9, UGT1A3, UGT1A6</td>
</tr>
<tr>
<td>Bilirubin</td>
<td>UGT1A1</td>
<td></td>
</tr>
<tr>
<td>Thyroxine</td>
<td>UGT1A3</td>
<td>UGT1A1</td>
</tr>
<tr>
<td>Serotonin</td>
<td>UGT1A6</td>
<td></td>
</tr>
<tr>
<td>Flavopiridol</td>
<td>UGT1A9</td>
<td></td>
</tr>
<tr>
<td>Mycophenolic acid</td>
<td>UGT1A9</td>
<td></td>
</tr>
<tr>
<td>S-oxazepam</td>
<td>UGT2B15</td>
<td>UGT2B7, UGT1A1, UGT1A6</td>
</tr>
<tr>
<td>Testosterone</td>
<td>UGT2B17</td>
<td>UGT2B5</td>
</tr>
<tr>
<td>Epirubicin</td>
<td>UGT2B7</td>
<td></td>
</tr>
<tr>
<td>Morphine</td>
<td>UGT2B7</td>
<td>UGT1A1, UGT1A3, UGT1A6, UGT1A8, UGT1A9, UGT1A10</td>
</tr>
<tr>
<td>Anastrozole</td>
<td>UGT1A4</td>
<td>UGT1A3, UGT2B7</td>
</tr>
<tr>
<td>Imipramine</td>
<td>UGT1A4</td>
<td>UGT2B10</td>
</tr>
<tr>
<td>Acetaminophen</td>
<td>UGT1A9</td>
<td>UGT1A1, UGT2B5, UGT1A6, UGT2B7</td>
</tr>
<tr>
<td>Vorinostat</td>
<td>UGT2B17</td>
<td>UGT2B7, UGT1A9</td>
</tr>
</tbody>
</table>
obtained from HapMap samples (41,42). A total of 44 UGT1A and 125 UGT2B polymorphisms were selected to be genotyped. After genotyping in the liver samples, pair-wise LD ($r^2$) and allele frequency of each SNP were re-calculated based on the liver sample set. Proxy SNPs were selected again from those in high LD ($r^2 \geq 0.8$), and those with allele frequency $<0.05$ were removed from subsequent analyses. As a result, 20 UGT1A and 41 UGT2B SNPs were removed due to their low minor allele frequency ($<5\%$) ($n = 9$ for UGT1A, $n = 5$ for UGT2B), high LD level ($r^2 \geq 0.8$) with other SNPs ($n = 11$ for UGT1A and $n = 33$ for UGT2B) or significant (FDR-adjusted $P < 0.05$) deviation from the Hardy–Weinberg equilibrium (HWE) ($n = 3$ for UGT2B). Therefore, 24 UGT1A and 84 UGT2B SNPs were selected for subsequent analyses (Supplementary Material, Table S2). Note that seven additional SNPs (rs28374627, rs4860985, rs7678636, rs6817882, rs7435827 and rs10028734) located at the UGT2B locus also significantly deviated from HWE (FDR-adjusted $P < 0.05$ for all). However, since the deviation was likely caused by the CNV in females, we avoided inferences from the correlation between normalized expression levels (i.e., relative to the housekeeping gene) to that between the original expression levels, and all reported correlations between genes must be so interpreted (43,44). Similarly, significant correlation patterns were observed between glucuronidation activities (median and range of $r$, 0.32, $-0.34$–0.99). Again, the substrates were clustered into two groups, consistent with the major UGTs involved (Fig. 1B).

The correlation pattern between UGT gene transcription and activity revealed that the variability in UGT activities was attributed to that in UGT gene transcription. In most cases and as shown in Figure 1C, the glucuronidation rate of a substrate could be predicted by the expression of the specific UGTs (listed in Table 1) involved in its metabolism, e.g. bilirubin and SN-38 by UGT1A1 ($r = 0.74$, $P = 4.36 \times 10^{-9}$), serotonin by UGT1A6 ($r = 0.73$, $P = 1.45 \times 10^{-8}$), morphine by UGT2B7 ($r = 0.47$, $P = 1.1 \times 10^{-3}$ for formation of both morphine-3-glucuronide and morphine-6-glucuronide), testosterone by UGT2B17 ($r = 0.41$, $P = 6.25 \times 10^{-9}$) and UGT2B15 ($r = 0.27$, $P = 0.08$), vorinostat by UGT2B17 ($r = 0.52$, $P = 3.31 \times 10^{-4}$) and UGT2B7 ($r = 0.39$, $P = 9.68 \times 10^{-3}$), etc. However, the glucuronidation activities of several other compounds were poorly correlated with the specific UGT expression, e.g. anastrozole and imipramine (Fig. 1C). Notably, UGT2B7 glucuronidation activity represented by epirubicin and morphine was negatively correlated with that of UGT1A6 reflected by serotonin (Fig. 1B). This was consistent with its negative correlation with UGT1A6 gene transcription as well (Fig. 1C). Interestingly, UGT2B7 and UGT1A6 gene transcription were positively correlated ($r = 0.45$, $P = 1.16 \times 10^{-7}$), although not as strongly as observed between many other UGTs (Fig. 1A). The mechanism underlying this observation remains unknown.

We evaluated the degree to which each TF alters the expression of a UGT (see Materials and Methods). Briefly, CAR, PXR and ESR1 expressions were highly significantly associated with those of UGTs (Fig. 1D). The most significant associations showed substantial up-regulation of UGT2B7, UGT2B10 and UGT2B15 ($r = 2.94 \times 10^{-21}$; $P = 1.09 \times 10^{-22}$ and $P = 2.58 \times 10^{-23}$, respectively) with increased ESR1 expression.

### Effect of demographic factors on UGT gene transcription and activity

Associations between demographic data (age, gender and race) and UGT gene transcription/activity were performed. No significant association was found between age or gender and UGT expression/activity. At first sight, UGT2B17 gene transcription was significantly higher in livers from male individuals ($t$-test, $P = 9.35 \times 10^{-4}$). Detailed analysis revealed that this might be due to the slightly higher frequency of the null allele of the UGT2B17 CNV in females. After removing all null-allele homozygotes, there was no significant association ($P = 0.06$, data not shown) between gender and UGT2B17 gene transcription. Significant differences in UGT gene transcription between Caucasian and African-American populations were observed for UGT2B10 ($t$-test, $P = 2.56 \times 10^{-4}$), UGT2B15 ($P = 1 \times 10^{-3}$) and UGT2B17 ($P = 0.015$) (Supplementary Material, Fig. S2). Since the African-American sample set is relatively small ($n = 23$), the following analyses were only focused on Caucasian samples ($n = 125$), without further adjustment for age and gender.

### Relationship between UGT gene transcription and activity

UGT mRNA expression levels were clustered into the UGT1A and UGT2B groups, reflecting the distinct regulation of the two gene families (Fig. 1A). Within each gene family, there was significant evidence for “co-expression” (see Methods) such as between UGT1A1 and UGT1A9 ($P = 1.81 \times 10^{-23}$) in UGT1A, and between UGT2B10 and UGT2B15 ($P = 3.67 \times 10^{-22}$) in UGT2B. Strong correlations were also observed between the two gene families, e.g. UGT1A9 and UGT2B7 ($r = 0.69$, $P = 4.4 \times 10^{-19}$), indicating that certain regulatory factors might play a common role in activating these genes. For any pair of genes,
We identified tagSNPs significantly (FDR < 0.05) associated with UGT gene transcription (Table 2). These SNPs (n = 21) mainly affected the gene transcription of UGT1A1 (n = 5), UGT1A3 (n = 6) and UGT2B17 (n = 10). Three SNPs (UGT1A1*28, rs2741045 and rs6759892) were associated with both UGT1A1 and UGT1A3 expression. The data confirmed the function of the UGT1A1*28 polymorphism and UGT2B17 CNV, as each of them was the most significant polymorphism associated with UGT1A1 and UGT2B17 expression (ANOVA, P = 3.7 x 10^{-4} and 7.04 x 10^{-46}, respectively) (Table 2). Since the tagSNPs were selected based on the LD level, the data suggest that these SNPs may account for additional variability in UGT gene transcription (the interdependence of these SNPs is shown later in this paper). Notably, we found that six SNPs (rs7435827, rs28374627, rs4860305, rs6817882, rs7678636 and rs4860985) located in the UGT2B17 CNV region were significantly associated with UGT2B17 gene transcription (ANOVA, P > 4.84 x 10^{-6} for all tests). Detailed analyses demonstrated that these SNPs accounted for significant additional variability in UGT2B17 expression in both hemizygous individuals with one UGT2B17 copy and homozygous individuals with two copies. As an example, Figure 2A and C show rs6817882, the most significant SNP where the T allele possessed a recessive effect on UGT2B17 expression in individuals bearing either one or two

![Heatmap of correlations between UGT gene transcription (A), UGT activities (B), UGT gene transcription and activity (C) and TFs and UGT gene transcription (D). MPA, mycophenolic acid; M3G, morphine-3-glucuronide; M6G, morphine-6-glucuronide. The color codes indicate correlation coefficients.](human-molecular-genetics.png)
<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Chr</th>
<th>Nucleotide position</th>
<th>Genetic location</th>
<th>Phenotype</th>
<th>ANOVA</th>
<th>FDR-adjusted</th>
<th>Best model</th>
<th>Linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6431558</td>
<td>2</td>
<td>234529643</td>
<td>UGT1A9 5′-flanking</td>
<td>UGT1A1 expression</td>
<td>5.15E-03</td>
<td>3.09E-02</td>
<td>Add</td>
<td>1.26E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bilirubin activity</td>
<td>4.33E-03</td>
<td>2.60E-02</td>
<td>Add</td>
<td>9.04E-04</td>
</tr>
<tr>
<td>rs2741034</td>
<td>2</td>
<td>234548814</td>
<td>UGT1A9 5′-flanking</td>
<td>UGT1A1 expression</td>
<td>4.74E-03</td>
<td>3.09E-02</td>
<td>Rec</td>
<td>1.25E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SN-38 activity</td>
<td>7.23E-04</td>
<td>3.22E-03</td>
<td>Dom</td>
<td>2.91E-04</td>
</tr>
<tr>
<td>rs2741045</td>
<td>2</td>
<td>234580140</td>
<td>UGT1A9 5′-flanking</td>
<td>UGT1A1 expression</td>
<td>2.66E-03</td>
<td>3.09E-02</td>
<td>Add</td>
<td>8.43E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SN-38 activity</td>
<td>1.80E-03</td>
<td>5.41E-03</td>
<td>Rec</td>
<td>4.70E-04</td>
</tr>
<tr>
<td>rs10538910</td>
<td>2</td>
<td>234583295</td>
<td>UGT1A7 5′-flanking</td>
<td>SN-38 activity</td>
<td>4.74E-03</td>
<td>3.09E-02</td>
<td>Rec</td>
<td>1.25E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SN-38 activity</td>
<td>5.21E-03</td>
<td>1.25E-02</td>
<td>Dom</td>
<td>1.67E-03</td>
</tr>
<tr>
<td>rs17868323</td>
<td>2</td>
<td>234590970</td>
<td>UGT1A7 N129K</td>
<td>UGT1A3 expression</td>
<td>4.65E-04</td>
<td>4.56E-03</td>
<td>Rec</td>
<td>1.49E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SN-38 activity</td>
<td>1.64E-03</td>
<td>5.41E-03</td>
<td>Dom</td>
<td>4.80E-04</td>
</tr>
<tr>
<td>rs6759892</td>
<td>2</td>
<td>234601669</td>
<td>UGT1A6 S7A</td>
<td>UGT1A1 expression</td>
<td>9.28E-03</td>
<td>4.46E-02</td>
<td>Rec</td>
<td>2.31E-03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UGT1A3 expression</td>
<td>2.51E-05</td>
<td>6.03E-04</td>
<td>Dom</td>
<td>4.29E-06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bilirubin activity</td>
<td>1.54E-03</td>
<td>1.23E-02</td>
<td>Add</td>
<td>3.84E-04</td>
</tr>
<tr>
<td>rs2011404</td>
<td>2</td>
<td>234627937</td>
<td>UGT1A4 C157C</td>
<td>UGT1A1 expression</td>
<td>1.37E-05</td>
<td>2.04E-04</td>
<td>Rec</td>
<td>4.21E-06</td>
</tr>
<tr>
<td>rs6706232</td>
<td>2</td>
<td>234637853</td>
<td>UGT1A3 E27E</td>
<td>UGT1A3 expression</td>
<td>1.65E-03</td>
<td>7.91E-03</td>
<td>Dom</td>
<td>3.33E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SN-38 activity</td>
<td>1.64E-03</td>
<td>5.41E-03</td>
<td>Dom</td>
<td>4.80E-04</td>
</tr>
<tr>
<td>UGT1A1* 28</td>
<td>2</td>
<td>234668893</td>
<td>UGT1A1 Promoter</td>
<td>UGT1A1 expression</td>
<td>3.70E-04</td>
<td>8.89E-04</td>
<td>Rec</td>
<td>6.59E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UGT1A3 expression</td>
<td>1.05E-03</td>
<td>6.29E-03</td>
<td>Dom</td>
<td>2.06E-04</td>
</tr>
<tr>
<td>rs33979061</td>
<td>2</td>
<td>234680679</td>
<td>UGT1A1 Intron 1</td>
<td>SN-38 activity</td>
<td>2.24E-05</td>
<td>2.04E-04</td>
<td>Rec</td>
<td>8.79E-06</td>
</tr>
<tr>
<td>rs10203853</td>
<td>2</td>
<td>234687418</td>
<td>UGT1A3 3′-flanking</td>
<td>SN-38 activity</td>
<td>8.04E-04</td>
<td>3.22E-03</td>
<td>Add</td>
<td>2.65E-04</td>
</tr>
<tr>
<td>rs2045097</td>
<td>4</td>
<td>69372321</td>
<td>UGT2B17 3′-intergenic</td>
<td>UGT2B17 expression</td>
<td>2.33E-15</td>
<td>9.91E-14</td>
<td>Rec</td>
<td>4.36E-16</td>
</tr>
<tr>
<td>NA</td>
<td>4</td>
<td>69372988</td>
<td>UGT2B17 3′-intergenic</td>
<td>UGT2B17 expression</td>
<td>5.05E-05</td>
<td>4.77E-04</td>
<td>Add</td>
<td>5.05E-05</td>
</tr>
<tr>
<td>UGT2B17 CNV</td>
<td>4</td>
<td>69373758–69491070</td>
<td>UGT2B17 3′-intergenic</td>
<td>UGT2B17 expression</td>
<td>7.04E-04</td>
<td>5.98E-04</td>
<td>Rec</td>
<td>3.57E-04</td>
</tr>
<tr>
<td>rs7455827</td>
<td>4</td>
<td>69414944</td>
<td>UGT2B17 Intron 5</td>
<td>UGT2B17 expression</td>
<td>4.13E-07</td>
<td>5.01E-06</td>
<td>Dom</td>
<td>5.74E-08</td>
</tr>
<tr>
<td>rs28374627</td>
<td>4</td>
<td>69417570</td>
<td>UGT2B17 Y355Y</td>
<td>UGT2B17 expression</td>
<td>1.55E-07</td>
<td>2.20E-06</td>
<td>Dom</td>
<td>2.01E-08</td>
</tr>
<tr>
<td>rs4860305</td>
<td>4</td>
<td>69420232</td>
<td>UGT2B17 Intron 3</td>
<td>UGT2B17 expression</td>
<td>4.84E-06</td>
<td>5.15E-05</td>
<td>Dom</td>
<td>7.32E-07</td>
</tr>
<tr>
<td>rs6817382</td>
<td>4</td>
<td>69436235</td>
<td>UGT2B17 Promoter</td>
<td>UGT2B17 expression</td>
<td>9.33E-14</td>
<td>2.64E-12</td>
<td>Dom</td>
<td>1.10E-14</td>
</tr>
<tr>
<td>rs7678636</td>
<td>4</td>
<td>69436885</td>
<td>UGT2B17 CNV promoter</td>
<td>UGT2B17 expression</td>
<td>9.93E-11</td>
<td>2.11E-09</td>
<td>Dom</td>
<td>1.66E-11</td>
</tr>
<tr>
<td>rs4860985</td>
<td>4</td>
<td>69488510</td>
<td>UGT2B15 5′-flanking</td>
<td>UGT2B17 expression</td>
<td>6.94E-09</td>
<td>1.18E-07</td>
<td>Dom</td>
<td>1.71E-09</td>
</tr>
<tr>
<td>rs12649644</td>
<td>4</td>
<td>70320823</td>
<td>Intergenic UGT2B28-UGT2B4</td>
<td>UGT2B17 expression</td>
<td>8.27E-04</td>
<td>7.03E-03</td>
<td>Rec</td>
<td>2.03E-04</td>
</tr>
</tbody>
</table>

Chr, chromosome; Add, additive; Dom, dominant; Rec, recessive.
copies of the gene. As a result, the metabolism activity of testosterone was affected in a similar manner (Fig. 2B and D).

TagSNPs significantly associated with UGT activities were also identified ($n = 10$). These SNPs affected mainly UGT1A1 activity (bilirubin $n = 4$, SN-38 $n = 10$) (ANOVA, $P < 5.21 \times 10^{-3}$ for all tests). No SNPs significantly associated with UGT2B activities were found after FDR adjustment of $P$ values.

When comparing the results for UGT expression and activity, we found that the majority of SNPs significantly associated with UGT activity were also significantly associated with UGT gene transcription, further demonstrating the strong transcriptional regulation of UGT activities. Given this observation and also our limited sample size, it is highly likely that SNPs nominally ($P < 0.05$) associated with both UGT expression and activity may be tag functional variants, although they did not reach the statistical significance after FDR adjustment. With this criterion, we identified a list of SNPs that could potentially affect UGT activities. An example of this is the aforementioned rs6817882 located in the UGT2B17 CNV region, which was significantly associated with testosterone in hemizygous individuals at the UGT2B17 locus (unadjusted overall ANOVA, $P = 0.001$; Figure 2B and D). These SNPs were thus defined as “suggestive” SNPs for association with UGT activity, subject to further validation (Supplementary Material, Table S3). All these SNPs had over 91% genotyping success rate in our study (Supplementary Material, Table S2).

**Relationship between the identified SNPs and trait-associated SNPs identified in GWAS**

We compared our findings with those annotated in the National Human Genome Research Institute catalog of published GWAS. Eleven SNPs (nine located in the UGT1A locus and two in the...
UGT2B locus) were associated in GWAS with various phenotypes including HIV control, bladder cancer, attention deficit/hyperactivity disorder, serum bilirubin levels, circulating levels of sex hormone-binding globulin and childhood obesity (Supplementary Material, Table S4). We genotyped six of these SNPs (or their LD ($r^2 \geq 0.8$) proxies). Five of these six variants were significantly associated with $UGT1A1$ mRNA expression or bilirubin/SN-38 activities in our study, and with circulating bilirubin levels in GWAS (Supplementary Material, Table S4).

**UGT SNPs and ENCODE regulatory elements**

We investigated the overlap of the UGT SNPs and regulatory elements identified by the ENCODE project (see Methods). We evaluated both the significant and ‘suggestive’ UGT SNPs as well as their LD proxies ($r^2 \geq 0.8$) in the EUR population of the 1000 Genomes Project (45). These annotations include affected regulatory motifs, DNase I hypersensitivity (HS) sites and chromatin states (enhancer and promoter elements) in a variety of cell types, including HepG2 and hepatocytes. We further tested the variants for enrichment for cell-type specific enhancers using the 1000 Genomes SNPs as background.

We found a highly significant enrichment ($P = 0.003$) for overlap with enhancer elements only in hepatocellular carcinoma (HepG2); no such enrichment was observed for the other tissues (Supplementary Material, Table S7). Furthermore, many of the UGT variants and their LD proxies overlap with DNase I HS sites as well as promoter and enhancer histone marks in HepG2 and hepatocytes (Supplementary Material, Table S8). However, we found no significant enrichment (see Methods) for DNase I HS (in HepG2 or any of the other cell types) at the UGT genes relative to other liver-expressed genes. A large proportion (86%) of the UGT SNPs and their LD proxies appear to alter the binding sites of certain TFs (Supplementary Material, Table S9). We compared the effect on regulatory motifs of the variants overlapping these motifs at the UGT genes and at other liver-expressed genes. The change in log-odds (LOD) score between the reference and alternative allele at the UGT genes and the remaining genes had quite distinct distributions ($P = 0.01$). Notably, TF binding sites at the UGT loci showed a significantly greater differential allelic effect on regulatory motifs ($P = 0.048$) than at other liver-expressed genes. Collectively, these results suggest that our eQTL findings are supported by tissue-specific regulatory annotations from ENCODE in closely related cell types.

**Interdependence of significant and “suggestive” SNPs**

Although we selected the tagSNPs based on LD ($r^2 \geq 0.8$), the effect of each tagSNP on UGT transcription or activity might be confounded by the haplotypes containing known functional alleles, e.g. $UGT1A1*28$ and $UGT2B17$ CNV. To explore the dependence of the identified SNPs on each other, we plotted the LD measurements $D'$ and $r^2$. Figure S3 (in Supplementary Material) shows both pair-wise $D'$ and $r^2$ values among the identified significant or “suggestive” SNPs in the $UGT1A$ and $UGT2B$ loci. We found that although the $r^2$ values among all 11 $UGT1A$ SNPs were low (<0.8), the $D'$ values between $UGT1A1*28$ and 7 out of 11 identified SNPs were >0.8, suggesting a confounding effect of $UGT1A1*28$ on these SNPs. The remaining three SNPs (rs2741034, rs10203853 and rs33979061) were more independent of each other as well as $UGT1A1*28$ (all pair-wise $D' \leq 0.8$). However, after conditioning on the $UGT1A1*28$ polymorphism, none of the tagSNPs across the UGT1A locus remained significant ($P > 0.05$), suggesting that $UGT1A1*28$ did confound the associations. Therefore, these SNPs might affect UGT transcription with at most very moderate effect if any. With regard to the UGT2B locus, while most identified SNPs were independent of $UGT2B17$ CNV (all $D' \leq 0.18$ and $r^2 \leq 0.06$, except for rs2045097), the SNPs located within the CNV region were rather dependent on each other ($D' \geq 0.79$). After conditioning on the $UGT2B17$ CNV, we found that the SNP rs28365063 was still associated with mRNA levels of $UGT2B7$ ($P = 0.01$) and $UGT2B15$ ($P = 0.006$), suggesting an independent role of this SNP in regulating $UGT2B$ transcription. As for the six SNPs located in the CNV region, after controlling for the CNV genotype, they were still significantly associated with $UGT2B17$ mRNA expression (rs7435827, $P = 2.82 \times 10^{-7}$; rs28374627, $P = 1.71 \times 10^{-7}$; rs4860305, $P = 2.63 \times 10^{-6}$; rs6817882, $P = 9.63 \times 10^{-11}$, rs7678636, $P = 2.45 \times 10^{-8}$ and rs4860985, $P = 1.89 \times 10^{-6}$), suggesting that they may indeed regulate $UGT2B17$ transcription independently of the CNV. However, the remaining SNPs were found not independently associated with $UGT2B17$ CNV ($P > 0.05$).

**DISCUSSION**

This is the first comprehensive investigation of the regulatory network of the human hepatic UGTs, providing a list of cis- and trans-factors to be considered in clinical pharmacogenetic studies. We confirmed that $UGT1A1*28$ and $UGT2B17$ CNV are associated with gene transcription and glucuronidation activity, implying that our overall approach is valid. These were the polymorphisms with the strongest associations with UGT mRNA expression and/or activity. We also discovered new SNPs responsible for additional functional variability in UGT mRNA expression and activity. Notably, while we observed no $UGT2B17$ expression in $UGT2B17$ nullizygotes and no significant differences in $UGT2B17$ mRNA expression between the carriers of one and two gene copies, six SNPs significantly associated with $UGT2B17$ expression were discovered within the CNV region, one of which also potentially affected testosterone glucuronidation activity (Fig. 2). Based on our discoveries, we assembled a comprehensive list of UGT polymorphisms ($n = 25$) including 21 SNPs with significant associations with UGT mRNA expression and/or activity, and 4 more “suggestive” SNPs with a moderate effect but potentially useful for understanding the full variation in hepatic UGT function. The potential function of these SNPs in regulating UGT gene transcription was verified by the significant enrichment for ENCODE enhancer elements among these SNPs in HepG2 cells. After conditional analysis, these 25 SNPs were further narrowed down to 1 $UGT1A$ SNP ($UGT1A1*28$) and 3 groups of $UGT2B$ SNPs (rs28365063, $UGT2B17$ CNV and the 6 SNPs located in the CNV region) that play an independent role in regulating UGT transcription and/or activity. These four groups of SNPs can serve as ‘very important UGT polymorphisms.
(UGT-VIPs) for future studies of pharmacogenetics or genetic diseases involving UGT genes.

Our newly identified UGT2B17 variants have great potential implications in pharmacogenetics. UGT2B17 CNV has been previously associated with inter-individual variability in drug metabolism, e.g. exemestane (6) and MK-7246 (41). Our finding provides novel alleles that may contribute to additional variability in the metabolism of these drugs. It would be particularly interesting to test clinically whether these variants could help explain inter-patient variability in efficacy and/or adverse reactions related to these drugs.

Our data help understand recent GWAS findings. We found that several SNPs that were significantly associated with circulating bilirubin levels in humans are in high LD with UGT1A1*28. Notably, one SNP rs2361502 that was significantly associated with serum bilirubin levels is in high LD with rs33979061 (Supplementary Material, Fig. S3), an SNP significantly associated with SN-38 activity but in incomplete LD with UGT1A1*28, suggesting that additional functional allele(s) may also contribute to variability in bilirubin levels.

Our data also highlighted the importance of a ‘locus-wide’ tagSNPs genotyping strategy over individual SNP/gene-based design in genetic or pharmacogenetic studies of UGT genes. Since the UGTs are clustered in gene families, the high LD level across the gene locus means that DNA variants altering UGT expression or activity could be far away from the gene (or exon 1) of interest. This is particularly true for UGT1A genes, as high LD level ($r^2 \geq 0.6$) was found in a region of $\sim$90 kb spanning from UGT1A1 to UGT1A9 (46,47). Our study identified a few SNPs located at the UGT1A9 5′-flanking region ($\sim$140 kb upstream of UGT1A9 exon 1) that are strong candidates affecting both UGT1A1/UGT1A3 gene transcription and UGT1A1 activity. Individual-gene-based studies focused on DNA variants within a limited region may therefore miss important information. In addition, many substrates are metabolized by multiple UGTs, further implying that individual-gene/exon 1 locus-based studies cannot completely address genetic variations in drug glucuronidation.

The strong correlations observed between UGT gene transcription and activities demonstrate that hepatic UGTs are controlled mainly by transcriptional regulation. We comprehensively measured mRNA levels of 15 TF genes that are known to regulate UGT genes based on previous studies. The regulation of UGTs by many of these TFs was confirmed by the recently released ENCODE data (genome.ucsc.edu) as multiple binding sites of AHR, GR, HNF1A, HNF4A, STAT3, SP1, PPARα and PPARγ in both UGT1A and UGT2B loci were identified by the ENCODE project. We found that CAR, PXR and ESR1 are the most important TFs modulating UGT gene transcription. While this suggests that UGT polymorphisms affecting the binding of these TFs may significantly change the UGT expression, polymorphisms regulating the activities of these TFs may also indirectly impact UGT gene transcription (48). Our results thus warrant further investigation to understand the causality of the UGT SNPs as well as to identify additional polymorphisms modulating UGT function.

We optimized quantification methods for hepatic UGT genes. Although many genome-wide-based platforms such as microarrays are available and have been widely utilized to analyze hepatic pharmacogenes, most of them place probes in the 3′-UTR region and are thus unable to detect individual UGT1A gene transcription. Moreover, these platforms are hybridization-based techniques, which cannot distinguish each individual gene due to the high sequence similarity between UGTs (data submitted elsewhere). Even in high-throughput sequencing, there is also problematic mapping of the short reads back to these gene loci (data submitted elsewhere). Our study optimized amplification conditions and validated the specificity of the primer set for real-time PCR-based UGT gene quantification. We identified the TBP gene as the best hepatic internal control gene compared with other housekeeping genes. The techniques described here could benefit future studies focusing on hepatic UGTs.

Our study was limited in the power to identify polymorphisms with moderate effect on UGT expression or activity, given the limited sample size. Moreover, although we observed a significant ethnic difference in gene transcription for a few UGT genes, no further analyses were performed due to the small sample size. As the DNA variation and LD structure between the African-American and Caucasian populations are quite different, functional variants specific for the African-American population may exist, which is left as an unaddressed question for future studies. The relative small sample size also precluded evaluation of the function of rare variants (minor allele frequency <5%) among the gene loci. As suggested by previous studies, rare variants may have a much larger effect size on UGT function (49,50). We have previously identified rare UGT variants, many of which are located in exons and are nonsynonymous (46,51).

A recent GWAS also suggested that a rare UGT1A SNP increases the risk of urinary bladder cancer (8,42). However, our study is underpowered and we could not study their functional effects. In addition, our study used a LD-based tagSNP strategy. Although we have identified a few SNPs significantly associated with UGT transcription or activity, these may not be causal SNPs leading to functional changes. Mechanistic studies will be necessary to analyze all SNPs in LD with these SNPs to elucidate possible casual associations. Splicing isoforms and variants have been reported for the UGT1A1 and UGT2B genes (52,53). Additional splicing variants that have not been fully characterized may also exist for other UGT genes (genome.ucsc.edu). As the function of these variants is incompletely understood, they were not included in our study. We also acknowledge that many of the GWAS SNPs were missed by the tag SNP approach. When we designed the study, the tagSNPs (in particular the UGT1A1 tagSNPs) were selected from our previous resequencing data (46) which only covered promoters, exons, exon-intron boundaries and evolutionarily conserved regions. Therefore, many SNPs identified later on in the 1000 Genome Project were not included. We quantified TF genes known to regulate UGT gene transcription. Data recently released from the ENCODE project indicate that many other TFs are involved in UGT gene regulation. How these additional TFs contribute to inter-individual variability of UGT transcription needs further investigation. On the other hand, whether certain SNPs affect mRNA stability would be also worthwhile to clarify. All these questions highlight the importance of performing a genome-wide, multi-ethnic, and large-scale study to reach a more complete understanding of the genetics of human glucuronidation.

In conclusion, our study provides a list of 25 polymorphisms associated with human hepatic UGT gene transcription and activity in vitro. These polymorphisms serve as potential candidates for
future pharmacogenetic studies related to drug glucuronidation or genetic association studies for human diseases.

**MATERIALS AND METHODS**

**Human liver tissue samples**

Normal (non-diseased) donor liver tissues not used for whole organ transplants were collected. Samples were procured through Dr. Mary Relling’s laboratory at St. Jude Children’s Research Hospital, and were provided by the Liver Tissue Cell Distribution System funded by NIH Contract #N01-DK-7-0004/ HHSN267200700004C and by the Cooperative Human Tissue Network. Samples were collected with the approval of institutional review boards. Both the University of Chicago and Purdue University institutional review boards have approved the use for this study. A total of 148 livers were initially included in the study. These livers were collected from unrelated donors of self-reported European (n = 125) and African (n = 23) descent. Demographic information is summarized in Supplementary Material, Table S5.

**DNA, RNA and cDNA preparation**

DNA was isolated from 20 mg of liver tissue using the Blood and Cell culture mini kit (Qiagen, Valencia, CA, USA). Total RNA was extracted using the TRizol® reagent (Invitrogen, Carlsbad, CA, USA) and purified with the RNeasy mini kit (Qiagen). Integrity of total RNAs was examined using agarose gel electrophoresis. Samples with degraded RNA were excluded from the study. Complementary DNA (cDNA) was synthesized with 1 μg total RNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer’s instructions.

**Selection of tagSNPs**

TagSNPs of the UGT1A and UGT2B loci were selected based on a LD threshold (r²) of 0.8, and a minor allele frequency of ≥ 0.05. For UGT1A genes, tagSNPs were chosen from the SNPs (n = 381) we identified by re-sequencing of the exons, promoters, intron–exon boundaries and evolutionary conserved regions of the entire locus in HapMap CEU samples (n = 24) and spanning ∼162 kb (HG B37, Chr2: 234525123–234688882) (41). TagSNPs for UGT2B genes were initially chosen from the HapMap CEU data from the entire UGT2B locus (Chr4: 69372321–70404261) as no re-sequencing SNPs were available when the project started. Re-sequencing data of the HapMap CEU samples (n = 24) were added during project implementation once they became available (51). The selected tagSNPs (n = 43 for UGT1A, n = 124 for UGT2B) are shown in Supplementary Material, Table S2. We also included UGT1A1*28 and UGT2B17 CNV due to their established functional role.

**Genotyping**

TagSNPs were genotyped using approaches combining different platforms, including Sequenom iPLEX, SNaPshot multiplex and PCR sequencing. Samples were analyzed by the Genetics & Informatics Using Statistics Core of the PAAR-Pharmacogenomics of Anticancer Agents Research Group, and the DNA Sequencing & Genotyping Facility of The University of Chicago Comprehensive Cancer Center. Primer sequences and conditions are available upon request. The UGT1A1*28 variant was genotyped as previously reported (18), and UGT2B17 CNV was genotyped using a Taqman-based real-time PCR assay according to the manufacturer’s instructions (Life Technology, CA, USA).

**UGT and TF mRNA expression**

UGT and TF mRNA levels were measured by two-step real-time PCR using the Mx3000P system (Stratagene, Cedar Creek, TX, USA). Total RNA (2 μg) was used to synthesize cDNA in a single experiment using the High Capacity cDNA Reverse Transcription kit (Life Technology) and following the manufacturer’s instructions. The reverse transcription was performed with both poly (dT)n oligo and random hexamer primers (1 : 1 ratio). The thermal profile consisted of 25°C for 10 min, 37°C for 2 h and 85°C for 5 s. Real-time PCRs were performed using IQ™SYBR Green Supermix® (Bio-Rad Laboratories, Hercules, CA, USA). Briefly, cDNA was amplified in 15 μl of the reaction mixture containing IQ™SYBR Green Supermix® and 0.5 μM of specific primers. After preheating (hot start reaction) at 95°C for 10 min, real-time PCR amplifications were performed. The oligonucleotide sequences and annealing conditions of the primers used are shown in Supplementary Material, Table S6. A disassociation curve was used to confirm the specificity of the PCR products. Reactions were performed in triplicate. All PCR products were sequenced for further confirmation of specificity. Initial template quantities were calculated using threshold cycle (Ct) values and a standard curve. To identify the best internal control gene, four housekeeping genes including beta-actin (ACTB), cyclophilin A (PPIA), 18S ribosomal RNA (18S) and TATA-box binding protein gene (TBP) were quantified. The TBP gene was determined to be the best internal control for gene transcription analysis on the basis of the least variability. Data were expressed as the log2-transformed ratios of UGTs or TFs and TBP.

**Measurement of UGT activity**

Microsomal glucuronidation activities of 14 UGT substrates were measured in human liver microsomes (n = 47). These substrates are acetaminophen, anastrozole, bilirubin, epirubicin, flavopiridol, imipramine, morphine, mycophenolic acid, serotonin, SN-38, S-oxazepam, testosterone, thyroxine and vorinostat. To confirm UGT screening results reported in the literature, most compounds (with the exception of bilirubin and morphine) were incubated with recombinant UGTs also.

Glucuronidation activities towards anastrozole (4) epirubicin (30), flavopiridol (36), imipramine (4), morphine (53), mycophenolic acid (54), SN-38 (55), testosterone (32), thyroxine (40) and vorinostat (32) were determined as previously described. Methods for measuring glucuronidation activities towards acetaminophen, bilirubin, serotonin and S-oxazepam were included in the Supplementary Methods.

**Data analysis**

UGT and TF gene transcription as well as the UGT activity data were all log transformed prior to statistical analysis. Correlations
between demographic factors and UGT expression, between UGT expression and activity, between TF expression and activity, and inter-activity correlations were tested using Pearson’s correlation.

We modeled TF regulation of UGT transcription using the following linear model:

\[ Y = \beta_0 + \beta_1 \times X + \epsilon, \]

\[ \epsilon \sim N(0, \sigma^2 I). \]

Here \( X \) denotes the (log-transformed) expression of a given TF, \( Y \) that of a given UGT and \( \epsilon \) is the residual vector with each component normally and independently distributed with variance \( \sigma^2 \).

Suppose \( G_1 \) and \( G_2 \) are the expression levels of two genes (e.g., two UGTs or a UGT and a TF) and \( H \) is the expression level of a housekeeping gene. The correlation between log-transformed ratios (e.g., between two ‘normalized’ UGTs) is, from the properties of log transformation, equal to the correlation between differences of log-transformed values:

\[ r(\log(G_1) - \log(H), \log(G_2) - \log(H)), \]

where \( r \) is the Pearson correlation function. The significance of each correlation was determined by assuming that under the null hypothesis, the test statistic,

\[ t = \frac{r \sqrt{(n-2)}}{\sqrt{1-r^2}}, \]

follows a \( t \)-distribution with \( n-2 \) degrees of freedom with \( n \) equal to the number of samples. As has been previously noted (43), even when the raw expression levels (i.e., prior to normalization) of a pair of genes are not correlated, the normalized expression levels (i.e., expressed as a ratio relative to the expression of the housekeeping gene) may still be correlated—indeed, may be spuriously high or spuriously low (44). Thus, all correlation (and downstream) analyses were performed on the normalized values (the basic unit of analysis), and no inferences were made from these to the original expression values. All reported results from this analysis must therefore be interpreted as correlations between composite measures rather than component measures (11). To obtain correlations that are robust to the underlying distribution of expression levels, we also calculated the Spearman rank correlation.

We performed hierarchical clustering on the UGT genes and on the effect of TF regulation of UGT expression to identify the global patterns of gene regulation or co-expression. Clustering and heatmap plotting were done using Cluster 3.0 program (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm #ctv, last accessed on 6 June 2014) and visualized using Java Treeview (http://jtreeview.sourceforge.net/, last accessed on 6 June 2014).

The HWE was tested using Fisher’s Exact test. Associations between SNPs and UGT gene transcription or activities were screened using one-way ANOVA. A post-hoc linear regression-based test was used to detect the best model (additive, dominant or recessive) for the association. To correct for multiple testing, a FDR < 0.05 was used as cut-off for HWE and ANOVA association tests between SNPs and UGT mRNA expression/activity. No further correction was performed in post hoc tests. A SNP with nominally significant association \( (P < 0.05) \) with both UGT expression and activity was defined as a “suggestive” SNP. Tests for HWE, genotype–phenotype associations and conditional analyses for SNP independence were performed using the PLINK program (http://pngu.mgh.harvard.edu/~purcell/plink/, last accessed on 3 June 2014). Pearson’s correlations were tested using SPSS 19.0 (SPSS, Inc., Chicago, IL, USA). GraphPad Prism 4.0 software was used to calculate mean, median, range, standard deviation (SD) and coefficient of variation (CV), and to plot figures for SNP associations (GraphPad Prism, CA, USA).

Using Haploreg (56), we annotated the UGT SNPs (as well as SNPs in LD with them in the EUR population from the 1000 Genomes Project, \( r^2 \geq 0.8 \)) with a variety of functional ENCODE annotations, including the effect on regulatory motifs and overlap with DNase I HS sites, enhancer and promoter elements in a variety of cell types, including HepG2 and hepatocytes. We also tested the variants for enrichment for cell-type specific enhancers using the 1000 Genomes SNPs as background, as previously described (56). We proceeded to test for enrichment for DNase I HS sites at the UGT genes relative to other liver-expressed genes. We calculated a background frequency using the remaining liver-expressed genes, and a binomial test was then performed. The change in LODs (56) between the reference and alternative allele was used to quantify the effect of a variant on a regulatory motif. A variant may be annotated with multiple TF binding sites. In this case, we used the maximum, over all affected motifs, of the change in LOD score. We compared the distribution of the differential allelic effect, as quantified by the change in LOD score, at the UGT genes and the other liver-expressed genes using the Kolmogorov–Smirnov test. We used the Wilcoxon test on the LOD scores to determine whether the variants at the UGT genes had a significantly greater differential allelic effect on binding motifs.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

Conflict of Interest statement. M.J.R receives royalties from The University of Chicago related to UGT1A1 genotyping. Other authors do not report conflicts of interest.

FUNDING

This work was supported in part by the National Institute of General Medical Sciences (U01GM061393 to PAAR-Pharmacogenomics of Anti-cancer Agents Research Group); the National Institutes of Health (Cancer Center Support Grant P30 CA14599 to the DNA Sequencing and Genotyping Facility of The University of Chicago Comprehensive Cancer Center); the Indiana CTSI Core Pilot Fund to W.L.; and start-up funds from the Department of Medicinal Chemistry and Molecular Pharmacology from Purdue University to W.L.
REFERENCES


atypical glucuronidation kinetics by UGT2B7. Drug Metab. Dispos., 31, 1086–1089.


