UDP-Glucuronosyltransferase 1A1 Gene Polymorphisms and Total Bilirubin Levels in an Ethnically Diverse Cohort of Women

Andrew L. Hong, Dezheng Huo, Hee-Jin Kim, Qun Niu, Donna L. Fackenthal, Shelly A. Cummings, Esther M. John, Dee W. West, Alice S. Whittemore, Soma Das, and Olufunmilayo I. Olopade

Department of Medicine, Section of Hematology/Oncology (A.L.H., H.-J.K., Q.N., S.A.C., O.I.O.), Department of Health Studies (D.H.), and Department of Human Genetics (D.L.F., S.D., O.I.O.), University of Chicago, Chicago, Illinois, Northern California Cancer Center, Fremont, California (E.M.J., D.W.W.); and Stanford University School of Medicine, Stanford, California (D.W.W., A.S.W.)

Received January 3, 2007; accepted April 30, 2007

ABSTRACT:
The objective of this study was to investigate variations in UGT1A1 polymorphisms and haplotypes among African-American and Caucasian women and to assess whether variants other than UGT1A1*28 are associated with total serum bilirubin levels. The (TA)n repeats and 14 single nucleotide polymorphisms (SNPs) in the UGT1A1 gene were genotyped in 335 African Americans and 181 Caucasians. Total serum bilirubin levels were available in a subset of 125 women. Allele frequencies of all SNPs and (TA)n repeats were significantly different between African Americans and Caucasians. In Caucasians, three common haplotypes accounted for 71.8% of chromosomes, whereas five common haplotypes accounted for only 46.6% of chromosomes in African Americans. Mean total serum bilirubin levels were significantly lower (p = 0.005) in African Americans (0.36 mg/dl) than in Caucasians (0.44 mg/dl). The (TA)n repeats explained a significant amount of variation in total bilirubin levels (R^2 = 0.27, p < 0.0001), whereas other SNPs were less correlative. Thus, significant variations in UGT1A1 haplotype structure exist between African Americans and Caucasians in this relatively large cohort of women. The correlation of UGT1A1 with total bilirubin levels was mainly due to (TA)n repeats in Caucasians but a clear correlation was not observed in African Americans because of the high diversity of haplotypes and the small sample size. These data have implications for the design of epidemiologic studies of cancer susceptibility and pharmacogenetic studies for adverse drug reactions in populations of African ancestry.

The UDP-glucuronosyltransferase (UGT) 1A1 gene has been implicated in a number of processes, including conjugation of bilirubin from hemoglobin and hemoprotein turnover, detoxification of potential carcinogens, phase II drug metabolism, and estradiol metabolism (Senafi et al., 1994; Bosma et al., 1995; Malffiti et al., 2005). The nine UGT1A isoforms are expressed to varying degrees in the liver (Strassburg et al., 1997) and mammary tissue (Senafi et al., 1994; Chouinard et al., 2006). The number of TA repeats in the TATA promoter region of UGT1A1 has been shown to be inversely associated with the transcriptional activity of UGT1A1, with five and six repeats (allele *36 and *1, respectively) associated with high UGT1A1 activity and seven and eight repeats (alleles *28 and *37, respectively) associated with low UGT1A1 activity (Beutler et al., 1998). The genotype, (TA)7/(TA)7, has been associated with Gilbert’s syndrome, which presents with mild hyperbilirubinemia (Bosma et al., 1995).

The unifying premise of integrative epidemiology suggested by Spitz et al. (2005) is that the same genes that are implicated in cancer risk may also be involved in a person’s propensity to carcinogenic exposure and/or to modulation of therapeutic outcome. Therefore, constructing genetic profiles that could be used to individualize therapy may also increase our understanding of cancer risk genes and may be applied to cancer development and prediction of outcome. To this end, differences in TA repeats of UGT1A1 have been shown to be responsible for the toxic effects of irinotecan, an anticancer drug (Iyer et al., 2002), and have also been linked with cancer susceptibility (Guillemette et al., 2000; Adegoke et al., 2004). Of interest are studies suggesting an association of UGT1A1 TA repeat polymorphisms with breast cancer among African Americans but not Caucasians (Guillemette et al., 2000, 2001).

ABBREVIATIONS: UGT, UDP-glucuronosyltransferase; PBREM, phenobarbital-responsive enhancer module; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; tSNP, tagging SNP; MAF, minor allele frequency; SBE, single base extension; HWE, Hardy-Weinberg equilibrium; ANOVA, analysis of variance.
Table 1. Characteristics of study sample

<table>
<thead>
<tr>
<th>Study Sample</th>
<th>African Americans</th>
<th>Caucasians</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No.</td>
<td>335</td>
<td>181</td>
<td>516</td>
</tr>
<tr>
<td>No. of related individuals</td>
<td>6</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>No. with breast cancer</td>
<td>282</td>
<td>91</td>
<td>373</td>
</tr>
<tr>
<td>Age (mean ± S.D.)</td>
<td>49.2 ± 10.0</td>
<td>48.0 ± 10.0</td>
<td>48.8 ± 10.0</td>
</tr>
<tr>
<td>Enrollment site</td>
<td>San Francisco Bay area</td>
<td>Chicago</td>
<td>46</td>
</tr>
<tr>
<td>Serum total bilirubin available</td>
<td>36</td>
<td>89</td>
<td>125</td>
</tr>
</tbody>
</table>

**Materials and Methods**

**Study Sample.** The study sample included individuals from two sources: self-reported African-American and Caucasian women who attended the Cancer Risk Clinic at the University of Chicago between 1992 and 2003 and self-reported African-American women from the San Francisco Bay area who enrolled in the Northern California Family Registry for Breast Cancer between 1995 and 2000 (John et al., 2004). DNA samples were available for 227 individuals (46 African Americans and 181 Caucasians) aged 23 to 81 years who attended the Cancer Risk Clinic for a comprehensive cancer risk assessment; of these, 59% had a previous diagnosis of breast cancer and 65% had a family history of breast cancer. DNA samples were also available for 244 African-American patients aged 20 to 64 years with newly diagnosed breast cancer identified through the Greater San Francisco Bay area cancer registry, and 45 African-American women without breast cancer identified through random-digit dialing. Characteristics of the study population are shown in Table 1. All participants provided written consent, and the study was approved by the institutional review boards of the University of Chicago and the Northern California Cancer Center.

**SNP Selection.** Novel SNP identification was based on resequencing of the UGTA1 gene cluster (Maitland et al., 2006). Briefly, optimal sets of mapped based tagging SNPs (tSNPs) were generated from LDSelect (Carlson et al., 2004) for African-American and Caucasian populations (unpublished data). As expected, the number of tSNPs for Caucasians (n = 10) was smaller than that for African Americans (n = 18) because the number of genetic variations is generally smaller in the former. The criteria for selecting these tSNPs were setting the tSNPs with a relative MAF >10% and an r² threshold of 0.8. An additional 6 SNPs were identified as well as (TA)_n. In total, there were 26 unique SNPs based on the resequencing of the UGTA1 genomic region. Of these, 14 SNPs in addition to (TA)_n were selected for the expanded genotyping in the current study, whereas other SNPs were not pursued because of high repetitive sequences within primer sets.

**PCR for (TA)_n.** Variants were genotyped as described previously (Te et al., 2000). Briefly, approximately 5 ng of DNA was amplified by PCR in 10 μl with primers flanking the polymorphic TA locus in the promoter region of the UGTA1 gene. The only protocol modifications to cycling were a final extension at 72°C for 5 min compared with 10 min as described previously. Control DNAs from individuals known to have a 6/6, 6/7, or 7/7 genotype were included in the PCR analysis. The amplified product yielded a 93, 95, 97, or 99-base pair fragment, which corresponded to (TA)_n, (TA)_n, (TA)_n, or (TA)_n, respectively. Samples were diluted 1:4, subjected to sequencing on an ABI 377 3700 Automated DNA Sequencer for (TA)_n, and scored via Gene Mapper 3.7 (Applied Biosystems, Foster City, CA).

**SNP Genotyping.** Genotyping was performed using the SNPstream (Beckman Coulter, Fullerton, CA) and procedures followed the specification for the instrument. Primer design for PCR and single base extension (SBE) was performed with Beckman Coulter Autoprimer software. PCR reactions were organized by SNP type. Reactions were composed of PCR primers at a final concentration of 50 nM, 0.2 U of enzyme Hot Master Taq (Eppendorf, Hamburg, Germany), and 2 to 5 ng of genomic DNA per reaction. Amplification was performed according to the manufacturer’s conditions. All postamplification steps were performed according to SNPstream specifications. Briefly, PCR cleanup was accomplished by treating reactions with 3 μl of Exo/SAP (USB, Cleveland, OH) with incubation for 30 min at 37°C followed by 100°C for 10 min. SBE reactions were performed using reagents and protocols specific to the SNPstream platform. SBE primers were present at a final concentration of 20 nM. Reactions were hybridized to Beckman array plates and scanned by the SNPstream. Genotyping of SNPs was done by DNAPrint Genomics ( Sarasota, FL).

**Bilirubin Levels.** To explore the relationship between UGTA1 genotypes and total bilirubin levels, data on serum total bilirubin levels were obtained from the University of Chicago Hospital medical records. Total bilirubin values for patients diagnosed with cancer were collected before the start of any treatments. If multiple values were available for a patient, the mean total bilirubin level was calculated. The standard deviation for these patients ranged from 0.05 to 0.35 mg/dl, with a median of 0.11 mg/dl. Bilirubin data for individuals who had known liver metastasis or significant hepatic dysfunction were excluded for the analysis.

**Statistical Analysis.** Hardy-Weinberg equilibrium (HWE) for all loci was examined among African Americans and Caucasians separately using a χ² test. Differences in allele frequencies between African Americans and Caucasians were compared using a χ² test or Fisher’s exact test, as appropriate. Pairwise LD was measured by Lewontin’s D’ (Hedrick, 1987), and the LD matrices were plotted using the GOLD software package (Abecasis and Cookson, 2000). Haplotypes were reconstructed using a Bayesian statistical method implemented in phase 2.1 (Stephens et al., 2001; Stephens and Donnelly, 2003). To describe the haplotype diversity, we calculated the effective number

**UGT1A1** is the main isoform that glucuronidates bilirubin (Tukey and Strassburg, 2000), and variants in the TATA promoter and phenobarbital-responsive enhancer module (PBREM) region have been studied extensively, but variation in serum bilirubin levels is not explained solely by TA repeats. Few studies have explored the polymorphisms throughout the UGTA1 gene and assigned functions to these polymorphisms (Sai et al., 2004; Kaniwa et al., 2005). In addition, few studies have examined variants of UGTA1 related to total serum bilirubin within the normal range (≤1.0 mg/dl) (Bosma et al., 1995; Sai et al., 2004; Lin et al., 2006). Nevertheless, it is important to evaluate these to understand whether variants causing hyperbilirubinemia are similar to or different from those involving variation of total bilirubin levels within the normal range. Identifying common alleles of UGTA1 related to bilirubin have important implications in both cancer treatment and prevention. A recent clinical trial has shown that total bilirubin before treatment was a positive predictor of grade 4 neutropenia in cancer patients receiving irinotecan (Innocenti et al., 2004). Additionally, recent works have shown that the risk of cardiovascular disease is lower in those with higher bilirubin levels (Lin et al., 2006). It also has been hypothesized that bilirubin itself may be a potent antioxidant and its slight increase within the normal range may be beneficial in reducing risk of cancer (Grant and Bell, 2000).

Furthermore, population-based studies have shown that serum bilirubin levels vary with gender, race, and smoking status and are 10 to 20% lower in African Americans than in Caucasians (Zucker et al., 2004). Although there are well established differences in UGTA1 TA repeat allele frequencies between Caucasians and African Americans, with (TA)_5 and (TA)_8 being absent in Caucasians (Beutler et al., 1998; Guillemette et al., 2000, 2001), it remains unclear whether there are additional differences in other UGTA1 single nucleotide polymorphisms (SNPs) between the two racial groups, which may explain some of the observed racial difference in serum bilirubin levels.

In this study, we aimed to 1) compare the variations in (TA)_n and SNPs from the PBREM region to the 3’-conserved region of UGTA1 between African-American and Caucasian women; 2) describe the linkage disequilibrium (LD) and construct haplotypes for these two populations; and 3) examine the relationship between UGTA1 polymorphisms and total bilirubin among individuals with total bilirubin levels within the normal range. To our knowledge, this is the first study to examine the entire region of the UGTA1 gene in relation to total bilirubin level in Caucasian and African-American women.
of haplotypes separately for each racial group (Carlson et al., 2004). The effective number of haplotypes, analogous to the effective number of alleles (Hartl and Clark, 1997), was calculated as $N_e = 1/\Sigma p_i^2$, where $p_i$ is the frequency of the $i$th haplotype. To find DNA regions in which there was no apparent historical recombination, we first searched for haplotype blocks using pair-wise D’ implemented in HaploBlockFinder (Zhang and Jin, 2003). Then the minimum blocks were expanded by adding SNPs at the ends using the following criteria. If the calculated D’ value between haplotypes contained in the minimum blocks and the SNP at the end was greater than the prespecified value (e.g., 0.8), we concluded that the added marker belonged to the same block. We proposed this haplotype block finding method because multilocus LD may not be adequately reflected by pairwise LD. Within each block, haplotype-tagging SNP (or microsatellite marker) was identified manually with help from HaploBlockFinder.

Because the distribution of total serum bilirubin was skewed to the right, a natural log transformation was done. For easier interpretation, geometric means and 95% confidence intervals are presented. One-way analysis of variance (ANOVA) was used to test whether the bilirubin level varied significantly by TA repeat or SNP genotypes, without assuming a genetic model. The allele frequencies of TA repeats and all SNPs genotyped were significantly different between Caucasian and African Americans (Table 2). Caucasians had only (TA)$_5$ and (TA)$_8$ repeats, whereas African Americans had broader variability in the number of TA repeats, including (TA)$_5$ and (TA)$_8$. Marker 6 appeared to be a unique SNP in African Americans. The MAFs of markers 13 and 15 in Caucasians and of marker 10 in African Americans were less than 5%.

**Linkage Disequilibrium and Haplotype Structure of the UGT1A1 Gene.** Figure 1 shows pairwise LD measured by Lewontin’s D’ for the two racial groups. In African Americans, there were two LD clusters: markers 1 to 5 (corresponding to the range from promoter to intron 1) and markers 9 to 15 (3 untranslated region of exon 5 and conserved regions). A historical recombination hot spot appears to exist between the two clusters. A similar LD pattern was observed in Caucasians, except that there was a long-range LD between markers 3 and 5 and markers 11 and 15.

Haplotype analysis was performed separately for African Americans and Caucasians. The haplotype diversity was much higher in African Americans than in Caucasians, with 83 haplotypes inferred in African Americans versus 23 haplotypes inferred in Caucasians. The effective number of haplotypes (the number of equally frequent haplotypes that would be required to produce the same homozygosity as observed) was 17.7 and 5.4 in African Americans and Caucasians, respectively. In Caucasians, there were three common haplotypes (frequency >5%), accounting for 71.8% of all chromosomes. In contrast, the five common haplotypes accounted for 46.6% of chromosomes in African Americans (Table 3). The top three haplotypes are the same in the two racial groups but with quite different frequencies.

Two haplotype blocks were observed for both Caucasians and African Americans, with one block consisting of markers 1 to 7 and the other block consisting of markers 9 to 15. Marker 8 did not fall into either block because it was in moderate LD with either block 1 or $2$ ($D’ \sim 0.6$). Although the gene was divided into two haplotype blocks, there was weak but nonignorable LD between the two blocks ($D’ = 0.52$ in African Americans and $D’ = 0.45$ in Caucasians). In Caucasians, the first block was defined by two major and one minor
haplotype, and only two markers (TA repeats and any one of markers 1, 3, 5, and 7) were required to tag them. The second block in Caucasians was defined by four common haplotypes, requiring three haplotype-tagging SNPs. In African Americans, five common haplotypes accounted for 85% of the first block and four markers were required to tag them. The second block in African Americans consisted of five common haplotypes and required four haplotype-tagging SNPs. In both blocks, haplotypes were more diversified in African Americans than in Caucasians.

**Bilirubin Level, UGT1A1 Polymorphism, and Race.** Total serum bilirubin levels were available for 125 patients (89 Caucasians and 36 African Americans). In the one-way ANOVA analysis, bilirubin levels were significantly associated with TA repeats (marker 2) and several other SNPs (markers 1, 4, 5, 9, 12, 13, and 14). Figure 2A shows the percentage of the variability in serum bilirubin levels explained by each marker (\(R^2\)). Specifically, the variability in bilirubin concentration explained by TA repeats was 27% (\(p < 0.0001\)). However, after adjusting for TA repeats, only marker 5 was marginally significant (\(p = 0.05\)). Figure 2B shows that 5% or less of variability in bilirubin level was explained by each SNP beyond TA repeats. The geometric means of bilirubin levels by TA repeat genotypes are presented in Table 4. Homozygous carriers of (TA)\(_7\) had the highest serum bilirubin level, followed by heterozygous carriers of (TA)\(_6/(TA)\(_7\). Homozygous carriers of (TA)\(_8\) and heterozygous carriers of (TA)\(_6/(TA)\(_8\) had low levels of bilirubin. There were only a few carriers of (TA)\(_6/(TA)\(_7\) and (TA)\(_8/(TA)\(_8\), and, thus, no inference was

**Fig. 1.** Linkage disequilibrium (Lewontin’s D') of 15 consecutive markers in UGT1A1 in Caucasians (A) and African Americans (B). These graphical displays were generated by GOLD and are a visual representation of the LD statistics. The scales on the right indicate Lewontin’s D’ values. A color scheme closer to red (D’ = 1.00) represents strong association of the SNPs (linkage disequilibrium) whereas a color scheme closer to blue (D’ = 0.00) represents areas of linkage equilibrium. The SNPs are arranged by the physical order and distance on UGT1A1 along the x and y axes.
made for these categories. TA repeats was still a strong determinant of bilirubin levels (adjusted \( p < 0.0001 \)) after adjusting for age, race, and breast cancer diagnosis. Interestingly, bilirubin levels were significantly lower (adjusted \( p = 0.02 \)) in patients diagnosed with breast cancer (0.40 mg/dl) than in those without breast cancer (0.46 mg/dl).

Serum bilirubin levels were 22% higher (adjusted \( p = 0.005 \)) in Caucasians than in African Americans (0.44 mg/dl) than in African Americans (0.36 mg/dl) (Table 4). This difference in bilirubin level was not explained by the distribution of TA repeats alone: after adjustment for TA repeats, Caucasians still had higher bilirubin level than African Americans (\( p = 0.018 \)). Stratified by race, 29% of the variation in total bilirubin level was explained by TA repeats for Caucasians and 17% for African Americans (Fig. 3).

Of the 89 Caucasian patients with bilirubin data, 3 patients were excluded from the analysis of haplotype and bilirubin because their haplotypes were constructed with a high level of uncertainty (probability \( > 80% \)). Table 5 shows the diplotype configuration (haplotype pairs) and the relationship with total serum bilirubin levels in Caucasians. Bilirubin levels varied significantly across the diplotypes in block 1 (\( R^2 = 0.33, p < 0.0001 \)) and the diplotypes in block 2 (\( R^2 = 0.22, p = 0.01 \)). However, this haplotype-phenotype correlation is mainly due to the effect of TA repeat polymorphisms. After removing the effect of TA repeat polymorphisms using linear regression models, neither block 1 (partial \( R^2 = 0.06, p = 0.17 \)) nor block 2 haplotypes (partial \( R^2 = 0.14, p = 0.14 \)) were significantly correlated with bilirubin. This is because (TA)\(_n\) was one of the tagging markers for block 1 haplotypes, and it was also moderately linked with haplotypes in block 2 (Table 5). In the 36 African Americans with bilirubin data, 20 distinct diplotypes were found in block 1 and 19 diplotypes were found in block 2. Therefore, the study lacked statistical power to examine the relationship between UGT1A1 haplotypes and bilirubin levels in African Americans.
The distribution of the number of TA repeats observed in this study is consistent with previous studies (Bosma et al., 1995; Beutler et al., 1998; Guillemette et al., 2001; Innocenti et al., 2002). (TA)$_4$ and (TA)$_5$ were each found in about 5% of African Americans but in none of the Caucasians. The allele frequency of (TA)$_7$ was slightly higher in African Americans (0.375) than in Caucasians (0.323). UGT1A1 polymorphisms other than the TATA box have not been well studied, and several SNPs (M6, M8 and M10) identified in this study are novel. The allele frequencies and corresponding genotypes of these SNPs were quite different between African Americans and Caucasians as shown in Table 3. In Caucasians, marker 1 (at -1352 upstream of the transcription starting point) and TA repeats were sufficient to tag the common haplotypes in block 1, C7 0.323, A6 0.569, and C6 0.107. Interestingly, these haplotype frequencies were similar to those tagged by markers -3279T>G and TA repeats for Caucasians in previous studies, G7 0.36–0.39, T6 0.45–0.53, and G6 0.09–0.1 (Innocenti et al., 2002; Kaniwa et al., 2005), suggesting that these three haplotypes are the same although different markers were used. For African Americans, the haplotypes tagged by marker 1 and TA repeats in this study were also comparable to that tagged using markers -3279T>G and TA repeats in previous studies (Innocenti et al., 2002; Kaniwa et al., 2005). However, the haplotype diversity within block 1 was greater in African Americans than in Caucasians and thus more tagging markers are required (Table 3). Similarly, the haplotype diversity within block 2 was also greater in African Americans than in Caucasians. In this study, the haplotype block 2 extended from the 3'-conserved region of exon 5 to the conserved region. To our knowledge, no previous TABLE 5

<table>
<thead>
<tr>
<th>Haplotype*</th>
<th>No.</th>
<th>Geometric mean (95% CI)</th>
<th>$R^2$ (p)</th>
<th>Partial $R^2$ (p)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6GCTCA/A6GCTCA</td>
<td>30</td>
<td>0.371 (0.337–0.409)</td>
<td>0.327 ($&lt;0.0001$)</td>
<td>0.060 (0.17)</td>
</tr>
<tr>
<td>A6GCTCA/C6.C.C.C</td>
<td>29</td>
<td>0.445 (0.403–0.491)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6GCTCA/C6.C.C.C</td>
<td>13</td>
<td>0.431 (0.371–0.501)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7GCCCA/C7GCCCA</td>
<td>10</td>
<td>0.697 (0.503–0.966)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6TCCC/G6TCCC</td>
<td>2</td>
<td>0.382 (0.156–0.935)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6TCCC/G7CTCCA</td>
<td>2</td>
<td>0.315 (0.256–0.386)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCACCA/TCTGCAA</td>
<td>26</td>
<td>0.367 (0.332–0.405)</td>
<td>0.220 (0.01)</td>
<td>0.146 (0.14)</td>
</tr>
<tr>
<td>TCTGCAA/TCTGCAA</td>
<td>22</td>
<td>0.494 (0.421–0.579)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTGCCA/C.C.C.C.A</td>
<td>14</td>
<td>0.435 (0.366–0.518)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCACCA/TCCACCA</td>
<td>8</td>
<td>0.408 (0.313–0.532)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTGCCA/CCTGCCA</td>
<td>7</td>
<td>0.524 (0.357–0.769)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCACCA...C/CACACAA</td>
<td>5</td>
<td>0.470 (0.333–0.664)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCACCA/CCTGCAA</td>
<td>2</td>
<td>0.358 (0.221–0.578)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTGCCA/CCTGCAA</td>
<td>1</td>
<td>0.310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTGCCA/CCTGCAA</td>
<td>1</td>
<td>0.920</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Markers 1 and 2 were used to tag haplotypes in block 1. Markers 9, 11, and 12 were used to tag haplotypes in block 2. Dots indicate that alleles can not be fully determined by haplotype-tagging markers.

*Partial $R^2$ and corresponding $p$ values were calculated in linear regressions after controlling for TA repeats.
study had examined this. These ethnic differences in SNP and haplotype distribution suggest that the presence and strength of the association between UGT1A1 genotypes and phenotypes including drug toxicities and cancer susceptibility need confirmation in independent populations.

Consistent with a previous study conducted in Caucasians (Bosma et al., 1995), we showed that TA repeats were strongly correlated with total bilirubin within the normal range, accounting for 29% of variations among Caucasians and 17% among African Americans. It is conceivable that the observed differences between the two groups is due to the diversity in the African Americans and the relatively small sample of African Americans patients with serum total bilirubin levels available for study. Nonetheless, our observation that serum total bilirubin levels were also lower among African American is consistent with results of previous population-based studies (Zucker et al., 2004). A TA repeat polymorphism was also shown to be responsible for the variation in bilirubin in Asian individuals with a similar R² (Ki et al., 2003; Sai et al., 2004). In addition, an Asian-specific variant, 211G>A (G71R) in exon 1 has been associated with variations in bilirubin (Huang et al., 2000; Ki et al., 2003; Sai et al., 2004). A study conducted in Japanese found that 1813C>T in the 3’ untranslated region of exon 5 was weakly but significantly associated with bilirubin level after adjusting for TA repeats (p = 0.03) (Sai et al., 2004). In this study, we showed that the same SNP, M9, was not significant after adjusting for TA repeats (p = 0.11). Among Caucasians, we did not find another variant that was associated with bilirubin conditional on TA repeats, as shown in both the genotype analysis (Fig. 2) and haplotype analysis (Table 5). Our sample size for African Americans was relatively small; thus, we cannot rule out that there may be other UGT1A1 variants accounting for the variation in bilirubin in African Americans. This possibility calls for further investigation of bilirubin levels in a larger study among African Americans as bilirubin can serve as an antioxidant and therefore may be a surrogate biomarker for risk.

A review of the UGT literature suggests that bilirubin is selectively glucuronidated by UGT1A1 (Tukey and Strassburg, 2000). A linkage analysis of pedigree data from the Framingham Study estimated the heritability of serum bilirubin to be 49% ± 6%, and the peak multi-point logarithm of odds score is located 1 centimorgan away from the UGT1A1 gene. The authors concluded that UGT1A1 may be the major gene controlling serum bilirubin (Lin et al., 2003), but almost all subjects in their study were Caucasians. Potentially, there are other factors such as other isoforms of UGT or other genes that moderate the variation in serum bilirubin because bilirubin levels also depend on the rate of bilirubin production, transportation, and elimination. The presence of Gilbert alleles is not necessarily linked to hyperbilirubinemia (Kadakol et al., 2000). In addition, UGT genes in mice are regulated by factors such as other isoforms of UGT or other genes that moderate the variation in serum bilirubin because bilirubin levels also depend on the rate of bilirubin production, transportation, and elimination. The presence of Gilbert alleles is not necessarily linked to hyperbilirubinemia (Kadakol et al., 2000). In addition, UGT genes in mice are regulated by factors such as other isoforms of UGT or other genes that moderate the variation in serum bilirubin because bilirubin levels also depend on the rate of bilirubin production, transportation, and elimination.

References
Innocenti F, Underiva SD, Iyer L, Chen PX, Das S, Kocherginsky M,arrison T, Janisch L,


