Genetic variants and haplotypes of the UGT1A9, 1A7 and 1A1 genes in Chinese Han

Xiaqing Zhang1*, Guokun Ao2*, Yuewen Wang1, Wei Yan1, Min Wang1, Erfei Chen1, Fangfang Yang1 and Jin Yang1

1National Engineering Research Center for Miniaturized Detection System, College of Life Sciences, Northwest University, Xi’an, Shaanxi, China.
2The 309th Hospital of Chinese People’s Liberation Army, HaiDian District, Beijing, China.

Abstract

In this report, we describe combined polymorphisms of the UGT1A9, UGT1A7 and UGT1A1 genes in 100 unrelated, healthy Chinese Han subjects. The functional regions of these genes were sequenced and comprehensively analyzed for genetic polymorphisms. Thirty variants were detected, including five novel forms. Tentative functional predictions indicated that a Cys/c174Arg substitution at position 277 in the UGT1A7 gene could alter the protein conformation and that 12460T > G in the 3’UTR might influence protein translation through specifically expressed miRNAs. UGT1A9*1b was a major functional variant in the subjects examined whereas the*1f allele had a frequency of only 0.5%. A special functional haplotype (GAGAAC) was identified for UGT1A9, 1A7 and 1A1. These findings provide fundamental genetic information that may serve as a basis for larger studies designed to assess the metabolic phenotypes associated with UGT1A polymorphisms. They also provide important data for the implementation of personalized medicine in Chinese Han.

Key words: Chinese Han, genetic polymorphisms, haplotype, UGT1A.

Received: February 1, 2012; Accepted: March 26, 2012.

UDP-glucuronosyltransferases (UGTs) are important enzymes in glucuronidation, the major pathway for the removal of endogenous compounds and potentially toxic chemicals ingested in the diet (King et al., 2001; Chu et al., 2009). UGT1A9 (MIM# 606434), UGT1A7 (MIM# 606432) and UGT1A1 (MIM# 191740) are three important members of the UGT1A subfamily that form a gene cluster on chromosome 2q37 (Miners et al., 2002; Cecchin et al., 2009). Each isoform comprises a unique alternate exon 1 that is regulated by its own promoter and is linked to the same four downstream exons as the other isoforms (Gong et al., 2001). The enzyme activities of UGT1A9, 1A7 and 1A1 show wide interindividual variability that can lead to the failure of drug therapy, severe drug toxicity or unpredictable adverse side effects. Major mutations in the UGT1A1 gene can cause a serious metabolic deficiency in bilirubin that leads to a condition known as Gilbert’s syndrome. The missense mutation 211G > A (UGT1A1*6 allele) markedly attenuates (to <10% of normal) or completely abolishes enzyme activity (Tukey and Strassburg, 2000; Verlaan et al., 2004) and reduces activity towards 7-ethyl-10-hydroxy-camptothecin (SN-38) (Huang et al., 2000). UGT1A7 and UGT1A9 also contribute to SN-38 glucuronidation (Gagné et al., 2002; Villeneuve et al., 2003). The relative Vmax of the UGT1A7*3 allele is 5.8-fold lower than the wild-type and confers a slow glucuronidation phenotype (Guillemette et al., 2000).

Genetic polymorphisms that alter enzyme functions are of clinical value in predicting a predisposition to disease and an individual’s ability to respond to certain therapeutic agents (Evans and Johnson, 2001). Haplotypes that combine variants of more than one gene can be used to investigate genotype-phenotype associations (Clark, 2004; Kitsios and Zintzaras, 2009). The co-occurrence of single nucleotide polymorphisms (SNPs) or segmental haplotypes with functional changes in UGT1A9, 1A7 and 1A1 can result in altered glucuronidation activity (Judson et al., 2000). In this study, we used direct sequencing to screen UGT1A9, 1A7 and 1A1 for genetic polymorphisms in their exons, promoters and surrounding introns and inferred the haplotype structures of each region.

The subjects consisted of 100 Chinese Han (50 males and 50 females, 18-40 years old) from Yulin. All of the participants were in a healthy condition based on physical examination and their medical histories. The aim of the investigation and the procedures involved were explained to...
the volunteers, all of whom provided written informed consent before participating in the study. The project was reviewed and approved by the Ethics Committee of Northwest University.

Blood samples were obtained from all participants and genomic DNA was isolated from peripheral blood leukocytes by a standard procedure (Okuda et al., 2002). Polymerase chain reaction (PCR) amplifications were done using 14 sets of primers. DNA sequencing reactions were done using ABI PRISM Big Dye Terminator V3.1 cycle sequencing chemistry. The sequences of the PCR and sequencing primers are available upon request. Sequencher 4.10.1 software (Gene Codes Corporation, Ann Arbor, MI, USA) was used for base calling, sequence alignment and polymorphism determination based on the DNA sequences. Novel variants were those not listed in the dbSNP database and having no rs record. Haploview 4.2 software (Massachusetts Institute of Technology, Cambridge, MA, USA) was used to estimate minor allele frequencies (MAF), test for Hardy-Weinberg equilibrium (HWE) and analyze linkage disequilibrium (LD) (Barrett et al., 2005).

We detected 30 genetic variants, the allele frequencies of which were in Hardy-Weinberg equilibrium (p > 0.05). Eight polymorphisms were found in the transcription factor binding site in the promoter region were analyzed using TFSEARCH (Transcription Factor Binding Sites Search) software (Heinemeyer et al., 1998) and the MAPPER platform (Marinescu et al., 2005). Two prediction programs, PANTHER (Thomas et al., 2003; Thomas and Kejariwal, 2004) and PolyPhen (Ramensky et al., 2002), were used to estimate the possible impact of novel nonsynonymous mutations on protein function by combining structural information and sequence similarity analysis. Swiss-PdbViewer 3.7 (Peitsch et al., 1997) was used to thread a protein primary sequence onto a three-dimensional template. CASTp was used to identify protein pockets and measure the changes in pocket volume produced by novel amino acid substitutions (Binkowski et al., 2003). MicroRNA (miRNA) target prediction was done using PITA (Kertesz et al., 2007), an algorithm that considers site accessibility and hybridization energy to predict potential miRNAs that target UGT1As and harbor novel SNPs in their binding sites.

We detected 30 genetic variants, the allele frequencies of which were in Hardy-Weinberg equilibrium (p > 0.05). Eight polymorphisms were found in the

<table>
<thead>
<tr>
<th>Location</th>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>Number of subjects</th>
<th>Minor allele frequency</th>
<th>HWE P value</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A9 Promoter</td>
<td>-2189T &gt; C</td>
<td></td>
<td>97</td>
<td>0.015</td>
<td>1.000</td>
<td>novel</td>
</tr>
<tr>
<td></td>
<td>-1888T &gt; G</td>
<td></td>
<td>91</td>
<td>0.045</td>
<td>1.000</td>
<td>rs6731242</td>
</tr>
<tr>
<td></td>
<td>-1819T &gt; C</td>
<td></td>
<td>37</td>
<td>0.415</td>
<td>0.322</td>
<td>rs13418420</td>
</tr>
<tr>
<td>1A7 Promoter</td>
<td>-341C &gt; T</td>
<td>Pro11Pro</td>
<td>75</td>
<td>0.140</td>
<td>0.576</td>
<td>rs28946877</td>
</tr>
<tr>
<td></td>
<td>-577T &gt; G</td>
<td></td>
<td>60</td>
<td>0.225</td>
<td>0.419</td>
<td>rs7586110</td>
</tr>
<tr>
<td>1A1 Promoter</td>
<td>-364C &gt; T</td>
<td>Gly71Arg</td>
<td>82</td>
<td>0.105</td>
<td>0.139</td>
<td>rs887829</td>
</tr>
<tr>
<td></td>
<td>-646C &gt; C</td>
<td></td>
<td>95</td>
<td>0.025</td>
<td>1.000</td>
<td>rs857478</td>
</tr>
<tr>
<td>1A1 Exon 1</td>
<td>-211G &gt; A</td>
<td>Gly71Arg</td>
<td>54</td>
<td>0.265</td>
<td>1.000</td>
<td>rs1483232</td>
</tr>
<tr>
<td></td>
<td>-686C &gt; A</td>
<td>Pro229Gln</td>
<td>97</td>
<td>0.015</td>
<td>1.000</td>
<td>rs35539060</td>
</tr>
<tr>
<td>1A Exon 2</td>
<td>-683G &gt; A</td>
<td>Lys317Lys</td>
<td>98</td>
<td>0.010</td>
<td>1.000</td>
<td>novel</td>
</tr>
<tr>
<td>1A Intron 2</td>
<td>-693T &gt; C</td>
<td></td>
<td>91</td>
<td>0.045</td>
<td>1.000</td>
<td>rs1483227</td>
</tr>
<tr>
<td>1A Exon 4</td>
<td>-793C &gt; T</td>
<td></td>
<td>95</td>
<td>0.025</td>
<td>1.000</td>
<td>rs34996978</td>
</tr>
<tr>
<td>1A Exon 5</td>
<td>-1212T &gt; G</td>
<td>Tyr486Asp</td>
<td>97</td>
<td>0.015</td>
<td>1.000</td>
<td>rs34993780</td>
</tr>
<tr>
<td>1A 3’UTR</td>
<td>-1249T &gt; G</td>
<td></td>
<td>99</td>
<td>0.005</td>
<td>1.000</td>
<td>novel</td>
</tr>
</tbody>
</table>

W – wild-type, M – mutant, HWE – Hardy-Weinberg equilibrium.
mon exons region, including five rare sites: 6833G > A, 6893T > C, 7939C > T (Pro364Leu), 12126T > G (Tyr486Asp) and 12460T > G. The novel nucleic acid replacement of G to A at nucleotide 6833 was a synonymous mutation (Lys317Lys) in exon 2. In the specific promoter and exon 1 regions we detected five well-known SNPs for the UGT1A1 gene, seven promoter variants for the UGT1A9 gene including two novel polymorphisms (-2189T > C, 1.5%, and -40 > G, 0.5%), and nine common variants and one novel rare non-synonymous mutation (829T > C, Cys277Arg) for the UGT1A7 gene. Table 1 summarizes the data for these polymorphisms.

Fourteen alleles were identified across the sequencing regions of the UGT1A9, 1A7 and 1A1 genes. There were six well-known alleles located in the promoters that influenced enzyme activity by affecting pre-transcriptional control. Among these alleles, a common SNP rs67695772 (UGT1A9*1b) that leads to a deletion of T at position 118 occurred at a high frequency (41.5%), with a genotype distribution (47 heterozygous and 18 homozygous subjects) similar to that of UGT1A7*2. In contrast, the allele UGT1A9*If, which combined *1b with three other sites (-1819T > C, -441C > T and -332T > C) in the promoter region of the UGT1A9 gene, showed a very low frequency (0.5%). In the coding regions, four variants (UGT1A7*2, *3, *4 and *11) with a frequency of 25% were found in exon 1 of the UGT1A7 gene while another four variants (UGT1A1*6, *7, *27 and *63) were located in the UGT1A1 gene; the latter variants had relatively low frequencies and none were homozygous. Tables 1 and 2 summarize these findings.

The pairwise LD for polymorphisms in the UGT1A9, 1A7 and 1A1 genes was expressed as r² on a graded red scale. Among the variants listed in Table 1, only common SNPs with frequencies >10% were tested for pairwise LD. The LD map is shown in Figure S1a in which increasing color intensity indicates greater linkage. In view of the potential of rare mutations to alter enzyme function all of the variants were included in the LD analyses. As shown in Figure S1b, two main linkage blocks were observed across the locus. Block 1 located in the promoter and exon 1 of the UGT1A7 gene included three complete LD variants (387T > G, 391C > A and 392G > A) that showed the greatest linkage (densest red color). In addition, three polymorphisms in block 2 also had high LD in the 3'-untranslated region (UTR) of the UGT1A1 gene.

Haplotype analysis using an accelerated EM algorithm identified two major haplotypes, TCCTCGT in block 1 and CCC in block 2 that accounted for 57% and 85.5% of the total haplotype diversity, respectively; four rare haplotypes had a frequency of <5%. The level of recombination between the two blocks was predicted to have a multiallelic D' value of 0.35. The significant LD seen here among several UGT1A SNPs suggests that it would be possible to use only a small number of highly informative SNP markers as

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>N</th>
<th>UGT1A1</th>
<th>UGT1A7</th>
<th>UGT1A9</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Han</td>
<td>100</td>
<td>26.5</td>
<td>1.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>280</td>
<td>15.3</td>
<td>25.2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Dong</td>
<td>273</td>
<td>15.3</td>
<td>10.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>She</td>
<td>264</td>
<td>15.3</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Taiwanese</td>
<td>441</td>
<td>15.3</td>
<td>2.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Singaporean</td>
<td>301</td>
<td>15.3</td>
<td>2.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>50</td>
<td>15.3</td>
<td>10.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>46</td>
<td>15.3</td>
<td>10.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>92</td>
<td>15.3</td>
<td>10.5</td>
<td>10.5</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>0.0</td>
<td>15.3</td>
<td>10.5</td>
<td>10.5</td>
<td></td>
</tr>
</tbody>
</table>

N – number of subjects in each study.
htSNP for genotyping in association studies. Table 3 summarizes the haplotype frequency and htSNP for each LD block.

The TFSEARCH and MAPPER platforms were used to predict the effects that the two novel polymorphisms in the promoter of the UGT1A9 gene would have on the transcription factor binding site. Neither the wild-type nor mutant of the -40C > G variant was predicted to bind to any transcription factor. For the -2189T > C variant, analysis of the E-values revealed some differences between the wild-type and mutant, but neither had a consistent effect on the transcription factor binding site. Further studies are required to determine whether the -2189T > C variant has a role in transcription factor regulation.

Ten miRNAs containing the novel polymorphism 12460T>G in the 3’UTR met the threshold of interaction energy for miRNA target pairing. Of these, has-miR-181a and has-miR-26a lost their important binding site with a G substitution and showed a significant difference (p < 0.005) in expression between colon tumor and normal samples (Cummins et al., 2006). The enhanced or reduced miRNA binding caused by SNP variation may have consequences for the corresponding downregulation or upregulation of mRNA stability and translation (Mishra et al., 2007; Chen et al., 2008). Table S1 shows the miRNA sequence and binding capacity for this variant.

PANTHER subPSEC scores of less than -3.0 (Pdeleterious >0.5) and a PolyPhen PSIC score difference >2.0 indicate that mutations may probably damage protein function. Based on the PANTHER prediction, which yielded a low subPSEC score (-6.08576) and significantly high Pdeleterious value (0.9563), the C277R variant has a good probability of impairing protein function. PolyPhen analysis showed that this variant also tended to enlarge the protein cavity (PSIC score difference = 3.874), yet a further indication of its potential to impair protein function.

CASTp analysis provided a better understanding of the fatal transformation in cavity structure caused by the nonsynonymous SNP (Figure 1A,B). The wild-type cavity (Pocket ID = 40), which has a surface area of 60.9 Å² and a volume of 42 Å³, is composed of atoms from five residues, namely, Cys277, Gly280, Leu357, His360 and Met362 (Figure 1A). Surprisingly, the replacement of Arg277 with Cys increases the surface area by 55.3 Å² and results in a greater than two-fold expansion in volume (Pocket ID = 60, surface area = 116.2 Å², volume = 98.4 Å³). The dramatic change in cavity dimensions in the mutant reflects the recruitment of atoms from an additional seven residues,

---

**Figure 1** - Functional prediction for the novel nonsynonymous mutation 829T>C in the UGT1A7 gene. (A, B) The Cys277Arg substitution enlarges the cavity configuration as shown by red balls. The pocket ID 40 in the wide-type (A) is formed by atoms from five residues while the pocket ID 60 in the mutant (B) is formed by atoms from seven residues. (C, D) The Cys277Arg substitution alters hydrogen bonds with the surrounding residues as shown by the green dotted line. Cys277 forms hydrogen bonds with Gly280 and Lys281 (C) whereas Arg277 regenerates hydrogen bonds with Lys281, Ile272 and Asp356 (D).
namely, Arg277, Gly280, Lys281, Asp356, Leu357, His360 and Met362 (Figure 1B).

The Cys277Arg substitution disrupts a hydrogen bond with Gly280 on the same helix structure but creates two hydrogen bonds with Ile272 and Asp356, and one between them (Figure 1C, D). The three-dimensional structure of the region shown in Figure 1 indicates that the change in hydrogen bonding can explain the expansion of the pocket. These findings suggest that Cys277, together with Ile272 and Asp356, may act as anchors between two helices to regulate the pocket size during interaction with various drugs. The presence of a highly conserved residue in an important domain, together with the striking alteration in cavity configuration caused by this single amino acid substitution, suggests that the new variant could produce a malfunctioning protein.

As shown here, UGT1A7*11 was a common coding variant (41.4%) in Chinese Han, in agreement with previous investigations of Asian populations (35% in 301 Japanese, 39% in 50 Koreans) (Saeki et al., 2006; Yea et al., 2008). In contrast, the frequencies of other variants differ considerably among ethnic populations. In Asians, the frequencies of UGT1A1*6 and *27 were reported to range from 7.1% to 21.3% and 0.9% to 5%, respectively (Chen et al., 2006; Zhou et al., 2009), whereas these variants were not observed in Caucasians (Thomas et al., 2006; Borucki et al., 2009). Among Taiwanese, the frequencies of alleles UGT1A1*2, *27 and *28 were 21.3%, 5.0% and 28.6%, respectively (Tang et al., 2005), compared with the corresponding value of 18.5%, 1.5% and 10.5% observed here. Similar interethnic differences in frequency exist among Dong, She and Han populations (Zhang et al., 2007). These differences may partly reflect the fact that some rare variants are unique to certain ethnic groups, but may also be related to the small sample sizes used. In this study, we used 100 subjects and this number may have been too small for adequate LD analysis, particularly since D' values tend to be higher when small sample sizes or rare alleles are examined. A larger sample size is therefore required to obtain more precise information.

Although we have predicted the potential significance of novel variants in Chinese Han, additional studies in vivo and in vitro are required to determine whether these variants cause functional changes in UGT1A. Overall, the special haplotype patterns and htSNPs described here should facilitate genotype-phenotype studies in larger populations.

Acknowledgments

We thank the volunteers who participated in this study and Changfeng Li for technical assistance. This work was supported by a grant from the National High Technology Research and Development Program of China (863 Program; grant no. 2009AA022710).

References


Verlaan M, te Morsche RHM, Pennington MW, Owens IS, Peters WHM and Drenth JPH (2004) Functional polymorphisms of UDP-glucuronosyltransferases 1A1, 1A6, and 1A8 are not involved in chronic pancreatitis. Pharmacogenetics 14:351-357.


Internet Resources


Supplementary Material

The following online material is available for this article:

Table S1 - Prediction of novel SNP that can influence miRNA targeting.
Figure S1 - Haplovieview analysis of the UGT1A9, IA7 and IA1 polymorphisms in this study.
This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Mara H. Hutz

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
### Table S1 - Prediction of novel SNPs that may influence the targeting of miRNAs.

<table>
<thead>
<tr>
<th>miRNA identification</th>
<th>miRNA sequence</th>
<th>Wild-type score ($\triangle \Delta G$: kcal/mol)</th>
<th>Mutation score ($\triangle \Delta G$: kcal/mol)</th>
<th>Tumor</th>
<th>Normal</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>has-miR-181a</td>
<td>AACAUUCAcgcugucggaguc</td>
<td>-6.06</td>
<td>-</td>
<td>5</td>
<td>62</td>
<td>0.001832</td>
</tr>
<tr>
<td>has-miR-181c</td>
<td>AACAUUCAcgcugucggaguc</td>
<td>-4.16</td>
<td>-</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>has-miR-26a</td>
<td>UUCAAGUAuccaggauagguc</td>
<td>-2.81</td>
<td>-</td>
<td>125</td>
<td>868</td>
<td>0</td>
</tr>
<tr>
<td>has-miR-653</td>
<td>GUGUGAAAaauccauuacaguac</td>
<td>-1.08</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>has-miR-203</td>
<td>GUGAAUGUuuaggaccacauag</td>
<td>-0.21</td>
<td>-</td>
<td>6</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>has-miR-190b</td>
<td>UGAUAUGUUugauauugggguc</td>
<td>-0.18</td>
<td>-</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>has-miR-181b</td>
<td>AACAUCAUcgcugucggaguc</td>
<td>-6.96</td>
<td>-</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>has-miR-181d</td>
<td>AACAUCAUcgcugucggaguc</td>
<td>-3.25</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

miRNA seed sequences are highlighted in uppercase and the complementary SNP sites in the seed sequences are underlined. miRNAs expression data in colon (tumor/normal samples) was cited from Cummins et al (2006) the colorectal microRNAome. Proc Natl Acad Sci USA 103:3687-3692.
**Figure S1** - Haploview analysis of the UGT1A9, IA7 and IA1 polymorphisms in this study. (A) Pairwise LD map for common UGT1A SNPs. A denser color indicates greater linkage. (B) LD block for all UGT1A SNPs identified in Chinese Han subjects. An area corresponding to a haplotype block is displayed in frame.