Characterization of a common deletion polymorphism of the UGT2B17 gene linked to UGT2B15


aCancer Research Program, JLC-Biomedical/Biotechnology Research Institute, North Carolina Central University, 1801 Fayetteville Street, Durham, NC 27707, USA
bDepartment of Genetics, Lineberger Comprehensive Cancer Center and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
cDivision of Molecular Epidemiology, National Center for Toxicological Research, Jefferson, AK, USA
dGenomics Research Program, JLC-Biomedical/Biotechnology Research Institute, North Carolina Central University, 1801 Fayetteville Street, Durham, NC 27707, USA

Received 20 February 2004; accepted 23 June 2004
Available online 3 August 2004

Abstract

Members of the human UDP-glucuronosyltransferase 2B family are located in a cluster on chromosome 4q13 and code for enzymes whose gene products are responsible for the normal catabolism of steroid hormones. Two members of this family, UGT2B15 and UGT2B17, share over 95% sequence identity. However, UGT2B17 exhibits broader substrate specificity due to a single amino acid difference. Using gene-specific primers to explore the genomic organization of these two genes, it was determined that UGT2B17 is absent in some human DNA samples. The gene-specific primers demonstrated the presence or absence of a 150 kb genomic interval spanning the entire UGT2B17 gene, revealing that UGT2B17 is present in the human genome as a deletion polymorphism linked to UGT2B15. Furthermore, it is shown that the UGT2B17 deletion polymorphism shows Mendelian segregation and allele frequencies that differ between African Americans and Caucasians.

© 2004 Elsevier Inc. All rights reserved.

Keywords: UDP Glucuronosyltransferase; Gene deletion; Polymorphisms (genetics); Duplicate genes; Genetic variation; Genetic diversity

Introduction

Based on sequence similarity and evolutionary divergence, the human UDP-glucuronosyltransferases (EC 2.4.1.17) have been subdivided into two families (UGT1 and UGT2) and four subfamilies (UGT1A, UGT1B, UGT2A, and UGT2B) [1]. Seven cDNAs from the UGT2B subfamily have been identified in humans: UGT2B4, UGT2B7, UGT2B10, UGT2B11, UGT2B15, UGT2B17, and UGT2B28 [2–5]. As part of the phase II liver detoxification system, these genes are responsible for maintaining steady-state levels of steroid hormones and other endogenous substances by catalyzing the transfer of glucuronic acid moieties to these molecules and rendering them hydrophilic [6]. This reaction, glucuronidation, facilitates the excretion of steroid compounds into the circulatory system and their elimination from the body through bile and urine [7]. UGT2B17 and UGT2B15 both bind testosterone and 5-α reductase metabolites in liver and extrahepatic tissue such as the prostate, testis, breast, lung, placenta, skin, uterus, kidney, and adrenal gland [8–10]. Their
expression in steroid target tissues such as the breast, uterus, and prostate is suggestive of a role in hormonally influenced diseases.

UGT2B17 cDNA shares over 95% sequence identity with UGT2B15 cDNA and similar substrate specificity [7]. While both proteins are reactive with androgen metabolites at the 17β position, UGT2B17 conjugates at both the 17β and the 3α positions due to a single amino acid change that causes a broader range of substrate specificity [8,11]. The high level of sequence identity between UGT2B15 and UGT2B17 suggests that the genes may have arisen from a duplication event [2,7]. The high level of identity and tight physical clustering of the known UGT2B subfamily genes within 1.0 Mb on chromosome 4q13.2 indicates the presence of multiple gene duplication events.

We used gene-specific markers to characterize the genomic relationship between UGT2B15 and UGT2B17 and to investigate genetic variation at this locus in different ethnic groups. Our results show for the first time that UGT2B17 occurs as a deletion polymorphism that is linked to UGT2B15 and differs in its frequency between two ethnic groups.

Results

Identification of a UGT2B17 polymorphism

To determine the genomic organization of UGT2B17 and UGT2B15, a human DNA panel of 40 samples from the SNP Discovery Repository was screened with UGT2B17- and UGT2B15-specific primers from exon 1 designated markers C and G, respectively. Fig. 1 and Table 1 show the locations and sequence of the primers on the genomic BAC clones that are annotated in the NCBI database and positioned so that gene orientation on both BACs is the same. The PCR amplification of UGT2B15-specific marker G generated the expected 329-bp product from all 40 individual DNA samples (Fig. 2A). Unexpectedly, amplification of the UGT2B17-specific marker C generated the predicted 316-bp product from only 31 samples and failed to amplify from the remaining 9 samples (Fig. 2B). PCR products from markers C and G were sequenced and aligned to confirm specificity for UGT2B17 and UGT2B15, respectively (Fig. 2C). C and G marker specificity was further validated by sequence alignments of PCR products with that of marker I, revealing the same sequence differences (Fig. 2C).

The lack of amplification and the sequence differences suggest the existence of a deletion spanning UGT2B17 in some human samples. To test this hypothesis the same DNA panel was screened with two additional markers, marker D, WIAF-2471-STS, and marker E (exon 6) (Fig. 1). The ability to amplify both markers was concordant with the amplification of marker C. Markers that fail to amplify span 64 kb from WIAF-2471-STS (marker D), located over 30 kb upstream of exon 1, to marker E, located in exon 6 of UGT2B17; markers downstream of

Fig. 1. Gene-specific marker locations. Portions of NCBI annotated BACs RP11-597F18 and RP11-185H6 are shown so that the genes are oriented in the same direction. Gene-specific marker locations are indicated and abbreviated ends are indicated by // marks. A sequence gap is indicated as designated in the University of California at Santa Cruz Genome Database. UGT2B17-specific markers A–E were derived from BAC clone RP11-597F18, while UGT2B15-specific markers F–H were designed from RP11-185H6. Markers C and G map to the exon 1 regions of UGT2B17 and UGT2B15, respectively. Markers D and H were located upstream of the transcriptional start sites of UGT2B17 and UGT2B15, respectively. The locations of markers A and B are within the intron 3 regions of UGT2B17, while marker F is located within the intron 3 region of UGT2B15. Marker E is located within exon 6 of UGT2B17. *Amplification of markers B and F was performed with a single primer pair and distinguished on the basis of fragment size (marker B 640 bp, marker F 706 bp). **Marker I maps to the exon 1 regions of both UGT2B15 and UGT2B17. Marker J amplification was used to assay for the presence or absence of the genomic interval containing UGT2B17.
exon 6 amplify from all individuals. These results suggest that there is a deletion polymorphism spanning UGT2B17.

Segregation analysis of UGT2B15/UGT2B17 markers in CEPH families

To determine the segregation patterns of UGT2B17 and UGT2B15, PCR was performed on genomic DNA samples from two CEPH families (Pedigrees 1344 and 13291) using seven primer pairs described in Table 1 and Fig. 1. In addition to the deletion polymorphism detected by UGT2B17-specific primers, two UGT2B15-specific markers (F and H, Figs. 1 and 3) were polymorphic.

The results for Pedigree 1344 are summarized in Fig. 3A. The mother and three offspring (M, 1, 4, and 8) showed amplification with UGT2B17-specific markers A through E, while the father, his parents, and remaining six offspring (F, GM, GF, 2, 3, 5, 6, 7, and 9) failed to amplify these five markers. As with the previous samples, all individuals amplified with UGT2B15-specific markers F through H. The paternal grandmother was heterozygous for marker H, while the mother was heterozygous for marker F.

The results for Pedigree 13291 are summarized in Fig. 3B. The father and his seven offspring (F, 1, 2, 3, 4, 5, 6, and 7) showed amplification with UGT2B17-specific markers A through E, while the mother and maternal grandparents did not amplify any of these markers. All samples amplified UGT2B15-specific markers F through H. The maternal grandmother was heterozygous for both markers F and H, while the mother and father were heterozygous for marker H. The presence of amplification for all UGT2B17-specific markers (A–E) was concordant in individuals from both pedigrees.

An analysis of marker segregation was performed in both pedigrees (Fig. 4). Only the mother in Pedigree 1344 (Fig. 4A) was informative for the segregation of the UGT2B17 deletion polymorphism; she is heterozygous for a UGT2B15 polymorphism in marker F and for the first SNP identified within marker H (see below). Three offspring inherited UGT2B17 and allele 2 at marker F and a G at all of the alleles for the SNPs from their mother, while six offspring inherited allele 1 at marker F and the A allele at SNP3 but not UGT2B17. The predicted phase of the alleles is shown in Fig. 4. Segregation analysis indicates that UGT2B17 and UGT2B15 are closely linked (LOD = 2.41, θ = 0).
In Pedigree 13291 (Fig. 4B), the segregation of UGT2B17 is not informative in either parent. However, both parents are heterozygous for two SNPs in marker H and transmit the alleles according to Mendelian expectations. Importantly, UGT2B17 is observed in all offspring independent of the paternal UGT2B15 allele inherited, indicating that the father was homozygous for the presence of UGT2B17. We conclude from these analyses that the UGT2B17 deletion polymorphism is closely linked to UGT2B15 and that individuals with two, one, or zero copies of UGT2B17 exist.

Allele frequency of the UGT2B17 deletion polymorphism

Successful amplification with marker C identifies the presence of the UGT2B17 gene. To test reliably the deletion of UGT2B17 in different ethnic groups, we developed a second PCR assay that spans the putative breakpoint of the deletion (marker J). Successful amplification of a 1.2-kb product with this assay identifies chromosomes with the deletion. Chromosomes lacking the deletion fail to amplify marker J due to the presence of an 150-kb interval spanning UGT2B17. The assay for marker J was validated in the two pedigrees. All individuals amplify marker J except the father in Pedigree 13291, which is the only individual homozygous for the presence of the UGT2B17. The combination of both assays allows the unambiguous genotyping of the deletion polymorphism.

We screened 85 African Americans and 103 Caucasians for markers C and J, which specifically recognize the presence or absence of the UGT2B17 gene, respectively. The results in Table 2 show that the distribution of genotypes is significantly different between these two ethnic groups (Ho, test for independence between genotype and ethnic background; \( \chi^2 = 7.41, 2 \text{ df, } p < 0.05 \)). The absence of the UGT2B17 gene was five times more frequent in Caucasians than in African Americans (11 and 2%, respectively). The deletion polymorphism is at Hardy–Weinberg equilibrium in both populations (Table 2).

Discussion

Here we report that DNA spanning the UGT2B17 gene is absent in 9 of 40 human DNA samples, while all samples contain the UGT2B15 gene. Segregation analysis demonstrates that UGT2B15 and UGT2B17 are linked and that the absence of UGT2B17 in some individuals is the result of a deletion polymorphism. Therefore, these genes are distinct and inherited in Mendelian fashion. The deletion polymorphism is common in humans (deletion allele frequency = 0.27) and the allele frequency varies between two ethnic groups, African Americans and Caucasians (deletion allele frequencies are 0.21 and 0.33, respectively; Table 2).

The UGT2B gene cluster on chromosome 4q13 appears to be the site of numerous duplication events that may have led to intergene redundancy and neofunctionalization. Work
by Riedy and co-workers [2] showed that there are numerous pseudogenes, orphan exons, and related sequences surrounding UGT2B4 and UGT2B7. UGT2B15 and UGT2B17 probably arose from a tandem duplication event by unequal recombination. Comparative sequence analysis suggests that this duplication is fairly ancient. Within exons the sequence identity between the genes is only 95%, while average sequence identity genome-wide between humans and chimpanzees is 98.8% [12–14]. Therefore, the duplication should predate the divergence of the human and chimpanzee lineages and is likely to be present in other primates. The different levels of sequence conservation between introns (approximately 91%) and exons indicate that both genes have been under positive selection. The deletion polymorphism described here was analyzed by DNA sequence analysis and the haplotypes of two SNPs (SNP3, G to A, and SNP4, G to A) are indicated.

We have identified two SNPs in the promoter region of UGT2B15 and discovered that most of the UGT2B15 SNPs currently available in public databases are in fact paralogous sequence variants [15]. These SNPs were validated by segregation analysis and resequencing. In addition to these polymorphisms, a SNP in UGT2B15 resulting in a D-Y amino acid change has been reported [16]. The polymorphism leads to proteins with different levels of enzymatic activity and is a possible risk factor for prostate cancer [16,17]. Population studies have shown that this polymorphism has different distributions among various ethnic groups [2,18]. Notably Asians have a higher prevalence for the D85 homozygous genotype compared to Caucasians and about half the proportion of Y85 homozygotes compared to Caucasians [18]. Small studies have been done to determine whether the polymorphism is associated with prostate cancer. Genotyping of DNA samples from a case–control group (64 cases and 64 controls) revealed that prostate cancer patients were significantly more likely to be homozygous for the lower activity D85 UGT2B15 allele than controls [19]. Similar results were obtained in another study with a larger case–control group (155 cases and 155 controls) in which a significant association was found between D85 homozygosity and increased risk of prostate cancer [20]. However, in a third study, the D85 polymorphism was not associated with prostate cancer risk (190 cases and 190 controls) [21]. Unlike the first two studies, in the third study the controls were patients with benign prostatic hyperplasia, which may have confounded the association. Additionally, Murata et al. [22] have shown that the absence of UGT2B17 alters immunogenicity in the enzyme’s role as a human minor histocompatibility antigen.
In addition to the UGT2B15 functional polymorphism, UGT2B17 has been shown to conjugate a broader range of steroid molecules than UGT2B15 as a result of a single amino acid difference in exon 1 [8,11]. The added substrate specificity range of the UGT2B17 gene product may play a significant biological role in maintaining steroid levels in the body.

The presence of two common and functionally relevant polymorphisms in two closely related genes with different allele frequencies between ethnic groups complicates association analyses between polymorphisms within this gene complex and disease phenotypes. Amino acid substitutions, gene copy number, and expression differences either alone or in combination may influence disease susceptibility. Studies combining genotyping with SNP quantitation are being used to facilitate the discovery of cancer-causing genes and cancer diagnosis [23]. Recently, deletion polymorphisms in glutathione S-transferase T1 (GSTT1) were examined in conjunction with SNPs to determine association with risk for lung and brain cancer [24,25]. In both studies the deletion polymorphism was associated with increased risk for cancer in Caucasians. The same deletion, however, was associated with increased risk for breast cancer in Tunisians [26] but not in Chinese [27]. Moreover, an insertion/deletion polymorphism of the angiotensin-converting enzyme (ACE) that results in higher levels of circulating and tissue ACE activity was found to be significantly associated with advanced

![Diagram](image-url)

**Fig. 4.** UGT2B15 and UGT2B17 inheritance. Marker amplifications from Fig. 3 were used to construct the haplotypes for UGT2B15 and UGT2B17 in CEPH families A and B. (A) Family A (CEPH/UTAH Pedigree 1344): paternal grandmother, NA12056; paternal grandfather, NA12057; father, NA10851; mother, NA10850; and offspring, NA12047 (d), NA12048 (d), NA12049 (d), NA12050 (s), NA12051 (s), NA12052 (s), NA12053 (d), NA12054 (d), and NA12055 (s). (B) Family B (CEPH/UTAH Pedigree 13291): maternal grandmother, NA07045; maternal grandfather, NA06986; father, NA06995; mother, NA06997; offspring, NA07018 (s), NA07036 (s), NA06981 (d), NA06980 (d), NA07047 (s), NA07433 (s), and NA07058 (s). Shaded and boxed regions indicate individual haplotypes of UGT2B15 and UGT2B17 markers. *Haplotypes for family B members NA07036 and NA07058 could not be determined.

**Table 2**

<table>
<thead>
<tr>
<th>UGT2B17 genotype distribution</th>
<th>Insertion/insertion</th>
<th>Insertion/deletion</th>
<th>Deletion/deletion</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasians</td>
<td>47</td>
<td>45</td>
<td>11</td>
<td>3.35</td>
</tr>
<tr>
<td>African-Americans</td>
<td>52</td>
<td>31</td>
<td>2</td>
<td>4.05</td>
</tr>
</tbody>
</table>

p < 0.05.
prostate cancer [28]. The fact that \textit{UGT2B17} copy number differs significantly between African Americans and Caucasians along with the prevalence of prostate cancer warrants further research into the role of the \textit{UGT2B} gene family in ethnic-based disease disparities.

\textbf{Materials and methods}

\textbf{DNA samples}

A panel of 40 unrelated, genetically diverse human DNA samples (subset M44PDR) was obtained from the Human SNP Discovery Resource of the Coriell Cell Repositories (Camden, NJ, USA). The panel of DNA represents the following ethnic backgrounds: European, African, Mexican, Native, and Asian-American. Human DNA samples from the members of two Centre d’Etude du Polymorphisme Humain (CEPH) pedigrees were obtained from the NIGMS Human Genetic Cell Repository (CEPH/UTAH Pedigree Subcollection). CEPH/UTAH Pedigree 1344 consists of paternal grandfather, NA12056; paternal grandmother, NA12057; father, NA10851; mother, NA10850; and offspring, NA12047 (d), NA12048 (d), NA12049 (d), NA12050 (s), NA12051 (s), NA12052 (s), NA12053 (d), NA12054 (s), and NA12055. CEPH/UTAH Pedigree 13291 (Family B) consists of maternal grandmother, NA07045; maternal grandfather, NA06986; father, NA06995; mother, NA06997; and offspring, NA07018 (s), NA07036 (s), NA06981 (d), NA06980 (d), NA07047 (s), NA07433 (s), and NA07058 (s).

DNA samples from African Americans and Caucasians were obtained from frozen pancreatic and liver tissues that were purchased from the Southern Division of the Human Tissue Network (Birmingham, AL, USA) and the International Institute for the Advancement of Medicine (Exton, PA, USA). All experiments were conducted under an approved IRB protocol at the National Center for Toxicological Research. Genomic DNA was prepared from tissue samples using Qiagen DNA isolation kits (Qiagen, Valencia, CA, USA).

Primers and product sizes for gene-specific markers A–I are indicated in \textbf{Table 1}. Marker D is a \textit{UGT2B17}-specific sequence-tagged site (WIAF-2471-STS) identified on the University of California at Santa Cruz Genomic Database. Primer sequences for markers A, E, and H were previously described [22,29]. A single pair of primers designed from the NCBI dSNP database (Cluster ID: rs5029389) was used in the amplification of markers B and F.

\textbf{Analysis of UGT2B15 and UGT2B17 markers}

PCR products from markers C, D, E, and G were analyzed by agarose gel electrophoresis. PCR amplification was performed on all human genomic DNA samples using AmpliTaq Gold MasterMix (Applied Biosystems, Foster City, CA, USA) incubated with forward and reverse primers at 100 pmol and 30 ng of template DNA. The incubation was carried out at 94°C for 4 min followed by 33 cycles of 94°C for 30 s, 60°C for 1 min, and 72°C for 3 min in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems). All PCR products were analyzed on 1% agarose/TBE gels and photographed with an Alpha Innotech Imager (Alpha Innotech Corp., San Leandro, CA, USA).

PCR products from \textit{UGT2B15}-specific markers F and H and \textit{UGT2B17}-specific markers A and B were analyzed by SSCP. PCR amplifications were performed on genomic DNA samples using a 10-µl reaction volume that contained \textit{Taq} polymerase buffer and polymerase (Promega, Madison, WI, USA), dNTP’s (Promega), 2 pmol of primers, 2µCi \textalpha;\textsuperscript{32}P-radiolabeled dCTP, and 50 ng genomic DNA. The PCR was performed as described above except 35 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s were used. Amplification of markers B and F was followed by a 1-h digestion with 5–20 units of \textit{Fnu4HI}, \textit{EcoRI}, or \textit{HindIII} restriction enzyme (New England BioLabs). SSCP analysis of PCR products was performed in accordance with the MDE Polycrylamide Gel Solution protocol (BioWhittaker Molecular Applications, Rockland, ME, USA). Electrophoresis of PCR products was for 16–18 h at 2.5 W. The gels were dried prior to an overnight autoradiography exposure on Kodak BioMax X-ray film (Rochester, NY, USA).

\textbf{Analysis of UGT2B17 deletion polymorphism}

PCR amplification of marker J was performed on DNA samples in a 10-µl reaction volume that contained \textit{Taq} polymerase buffer and polymerase (Promega), dNTP’s (Promega), 2 pmol of primers, 2µCi \textalpha;\textsuperscript{32}P-radiolabeled dCTP, and 50 ng genomic DNA. Touchdown PCR amplification was performed with an initial 2 cycles at 95°C for 30 s, followed by 59–56°C for 30 s, 72°C for 30 s; then 20 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. SSCP analysis of PCR products was performed as described above.

\textbf{Sequencing}

PCR products were purified with QIAquick PCR purification kit (Qiagen) and sequenced with a 3100 genetic analyzer (Applied Biosystems). The sequencing reaction was performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA polymerase, FS (Applied Biosystems) and primers from markers C, G, H, and I.

\textbf{Acknowledgments}

The authors thank Beverly Koller (Department of Genetics, University of North Carolina at Chapel Hill),
Ingrid Hall (Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention), and Leroy Worth (DERT, SRB, NIEHS/NIH) for critically reading the manuscript. We appreciate additional technical assistance from Elena Arthur, Keoshia Johnson, and Beverly Word. This work was supported by grants from the NCI/NIH (CA92075-02 UNC, CA992077-02 NCCU), NIGMS/NIH (S06-GM08049), and NSF (MCB-0133526 to F.P.M.V.).

References


学霸图书馆
www.xuebalib.com

本文献由“学霸图书馆-文献云下载”收集自网络，仅供学习交流使用。

学霸图书馆（www.xuebalib.com）是一个“整合众多图书馆数据库资源，提供一站式文献检索和下载服务”的24小时在线不限IP图书馆。

图书馆致力于便利、促进学习与科研，提供最强文献下载服务。

图书馆导航：
图书馆首页 文献云下载 图书馆入口 外文数据库大全 疑难文献辅助工具