Glutathione S-Transferase P1: Gene Sequence Variation and Functional Genomic Studies


1Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Department of Biochemistry and Molecular Biology, and Department of Health Sciences Research, Mayo Medical School-Mayo Clinic, Rochester, Minnesota

Abstract

Glutathione S-transferase P1 (GSTP1) is of importance for cancer research because of its role in detoxifying carcinogens, activating anticancer prodrugs, metabolizing chemotherapeutic agents, and its involvement in cell cycle and apoptosis regulation. Two common GSTP1 genetic polymorphisms have been studied extensively. However, the full range of GSTP1 genetic variation has not been systematically characterized in the absence of disease pathology. We set out to identify common GSTP1 polymorphisms in four ethnic groups, followed by functional genomic studies. All exons, splice junctions, and the 5'-flanking region of GSTP1 were resequenced using 60 DNA samples each from four ethnic groups. The 35 single-nucleotide polymorphisms (SNP) identified included six nonsynonymous SNPs and 17 previously unreported polymorphisms. GSTP1 variant allozymes were then expressed in COS-1 cells, and five displayed significantly altered levels of enzyme activity. One decreased to 22% of the wild-type (WT) activity. Four variant allozymes had Km values that differed significantly from that of the WT, and five showed altered levels of immunoreactive protein compared with WT, with a significant correlation (r = 0.79, P < 0.007) between levels of immunoreactive protein and enzyme activity in these samples. In the Mexican American population, five linked SNPs were significantly associated with GSTP1 mRNA expression, one of which was found by electrophoretic mobility shift assay to alter protein binding. These studies have identified functionally significant genetic variation, in addition to the two frequently studied GSTP1 nonsynonymous SNPs, that may influence GSTP1's contribution to carcinogen and drug metabolism, and possibly disease pathogenesis and/or drug response. [Cancer Res 2008;68(12):4791–801]

Introduction

Glutathione S-transferase P1 (GSTP1) is a member of the cytosolic GST superfamily (1–3). These phase II enzymes catalyze the glutathione conjugation of a variety of electrophilic xenobiotics, including substrates that range from environmental toxins and carcinogens to drugs used in the treatment of cancer (3–6).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprint: Richard Weinshilboum, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905. Phone: 507-284-2246; Fax: 507-284-4455; E-mail: weinshilboum.richard@mayo.edu.

©2008 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-07-6724

4 http://www.hugenavigator.net
sequence variation in GSTP1, followed by the functional characterization of those polymorphisms and determination of mechanisms by which they might alter function. As a first step, all GSTP1 exons, splice junctions, and a portion of the 5′-flanking region (5′-FR) were resequenced using 60 DNA samples each from African American (AA), Caucasian American (CA), Han Chinese American (HCA), and Mexican American (MA) subjects. Functional genomic studies were then performed with all variant allozymes encoded by alleles containing each of the nonsynonymous cSNPs identified during the gene resequencing studies. In addition, transcriptional activity of common 5′-FR SNPs and haplotypes were assessed with mRNA microarray and electrophoretic mobility shift assay (EMSA).

The results of these studies provide comprehensive information with regard to common sequence variation in GSTP1, as well as the functional consequences of that variation. These data also provide a foundation for future genotype-phenotype association studies involving both carcinogenesis risk and inherited variation in antineoplastic drug response.

Materials and Methods

DNA samples. DNA samples from 60 CA, 60 AA, 60 HCA, and 60 MA subjects (Human Variation Panel sample sets HD100CAU, HD100AA, HD100CHL, and HD100MEX) were obtained from the Coriell Cell Repository. The National Institute of General Medical Sciences had anonymized these DNA samples before deposit, and all subjects had provided written consent for the use of their DNA for research purposes. These studies were reviewed and approved by the Mayo Clinic Institutional Review Board.

GSTP1 gene resequencing. The PCR was used to amplify all GSTP1 exons, intron-exon splice junctions, and ~400 bp of the 5′-FR. Amplification conditions and primer sequences are listed in the supplementary material. Amplicons were sequenced on both strands in the Mayo Molecular Biology Core Facility using dye terminator sequencing chemistry. Polymorphisms observed only once, as well as any ambiguous sequences, were confirmed by performing independent amplifications, followed by DNA sequencing.

GSTP1 microarray analysis. Lymphoblastoid cell lines from which the DNA samples used for the gene resequencing had been obtained were acquired from the Coriell Cell Repository. Total RNA was extracted from cell lines corresponding to the DNA samples used in each of the four populations with the RNeasy kit (Qiagen). RNA quality assessment was performed using the Agilent 2100 bioanalyzer before microarray analysis. All RNA samples had an Agilent RNA integrity number of >9.0. The RNA was then reverse-transcribed and biotin labeled for hybridization with Affymetrix U133 Plus 2.0 GeneChips (Affymetrix). The microarray images were then reverse-transcribed and biotin labeled for hybridization with Affymetrix U133 Plus 2.0 GeneChips (Affymetrix). The microarray images were analyzed with quality control techniques established in the Mayo Clinic Microarray Core Facility, and the data were normalized using Fastsl, a version of cyclic loess normalization (35). Data from probe set 200824_at, corresponding to GSTP1, were used in the analyses described here.

GSTP1 EMSA. Biotin-labeled double-stranded oligonucleotides corresponding to the WT sequences and to the GSTP1 (−219) or (−18) variant sequences, together with their corresponding unlabeled oligonucleotides as competitors, were used in these assays. Binding assays were performed, followed by electrophoresis on a 4% nondenaturing gel and transfer to a nylon membrane, with detection according to the manufacturer’s directions using the LightShift Chemiluminescent EMSA kit (Pierce). Nuclear extracts were prepared from a pool of the lymphoblastoid cell lines used to perform the microarray experiments. Specifically, 15 million cells were pelleted at 524 × g for 3 min. Cells were washed in 1 mL of cold PBS and resuspended at 20 mmol/L of cold PBS and resuspended in 200 μL of ice-cold lysis buffer consisting of 10 mmol/L HEPES (pH 7.9), 10 mmol/L KCl, 0.1 mmol/L EDTA, 0.1 mmol/L EGTA, 0.4% NP40, 1 mmol/L DTT, 0.5 mmol/L phenylmethylsulfonyl fluoride (PMSF), and protease inhibitor. After incubation on ice for 15 min, the lysate was centrifuged at 14,000 × g for 30 s. After washing, the pellet was resuspended in a buffer containing 20 mmol/L HEPES (pH 7.9), 400 mmol/L NaCl, 1 mmol/L EDTA, 1 mmol/L EGTA, 1 mmol/L DTT, 0.5 mmol/L PMSF, and protease inhibitor and was vortexed for 15 s at 4°C. The mixture was then centrifuged at 14,000 × g for 10 min and was stored at −80°C. An antibody against acute myelogenous leukemia-1 (AML-1; Santa Cruz Biotechnology) was used in an attempt to perform a super shift assay. Specifically, 2 μL of antibody were added to 20 μL of the binding reaction, followed by incubation at room temperature for 45 min.

GSTP1 expression constructs and transient expression. A WT expression construct was created by amplifying cdNA from a human liver cdNA library and cloning the full-length GSTP1 cdNA open reading frame into the expression vector pcDNA4/HiscMax (Invitrogen). Site-directed mutagenesis was performed using the circular PCR to create variant allozyme constructs. Sequences of all inserts were confirmed by sequencing in both directions. Expression constructs for WT and variant allozyme cdNAs were then transfected into COS-1 cells in serum-free DMEM using the TransFast reagent (Promega) at a charge ratio of 3:1. Expression of variant allozymes was performed in a mammalian cell system to ensure that the mechanism for mammalian posttranslational modification and degradation would be present. Specifically, 7 μg of construct DNA were cotransfected with 7 μg of pSV-β-galactosidase DNA (Promega) as a control to correct for possible variation in transfection efficiency. After 48 h, the cells were washed with PBS, resuspended in 1 mL of homogenization buffer, and lysed with a Polytron homogenizer (Brinkmann Instruments). The homogenates were centrifuged at 100,000 × g for 4°C for 1 h. The resulting cytosol preparations were stored at −80°C.

GSTP1 activity assay and substrate kinetics. Enzyme activity was determined with 1-chloro-2,4-dinitrobenzene (CDNB; Sigma-Aldrich) as the substrate for each recombinant GSTP1 allozyme, using a modification of the spectrophotometric assay described by Habig and colleagues (36). Specifically, 100 mmol/L potassium phosphate buffer (pH 6.5), 1 mmol/L EDTA, 2.5 mmol/L reduced glutathione (Sigma-Aldrich), 1.0 mmol/L CDNB (dissolved in 95% ethanol, final concentration in the assay of 3.2%), and recombinant enzyme were combined in a final volume of 250 μL. This reaction mixture was added to a 96-well plate before the addition of the enzyme and was incubated at room temperature for 2 min. The enzyme source was then added, and the mixture was incubated at room temperature for 7 min, followed by the measurement of absorbance at 340 nm in a Safire2 microplate reader (Tecan). The complete reaction mixture including COS-1 cell high-speed supernatant from cells transfected with empty vectors served as a blank. The signal-to-noise ratio was ≥1 or greater over the range of enzyme concentrations used. Activity present in untransfected COS-1 cells was always subtracted from that of COS-1 cells transfected with GSTP1 allozyme constructs. Activity for the WT enzyme was set at 100%, and all other allozymes are reported as percentages of the WT value. β-Galactosidase activity was measured spectrophotometrically in the same samples using the Promega β-Galactosidase Assay System, and levels of enzyme activity were corrected on the basis of the cotransfected β-galactosidase enzyme activity. The same assay was used to perform substrate kinetic studies with five concentrations of CDNB that ranged from 0.25 to 2.0 mmol/L but only over the course of a 300-s incubation.

Western blot analysis. Levels of immunoreactive protein were determined for each recombinant GSTP1 allozyme by performing quantitative Western blot analysis. A mouse monoclonal anti-His antibody (Sigma-Aldrich) was used to visualize protein bands. Specifically, COS-1 cell cytosol was loaded onto 12% SDS mini-gels (Bio-Rad) on the basis of levels of enzyme activity were corrected on the basis of the cotransfected β-galactosidase enzyme activity. The same assay was used to perform substrate kinetic studies with five concentrations of CDNB that ranged from 0.25 to 2.0 mmol/L but only over the course of a 300-s incubation.


Downloaded from cancerres.aacrjournals.org on July 27, 2017. © 2008 American Association for Cancer Research.
expression values for the genotype-phenotype associations were log2, 0, 1, or 2 on the basis of the number of rare variants present. GSTP1 was evaluated with a 1-degree-of-freedom test, with SNP genotypes coded as phenotype of GSTP1 expression. For these analyses, SNP genotypes were used to assess the association of each SNP with the quantitative parameter. SNP pairs for MA subjects were observed to have both a \( D^* \) value of <0.05. These maximally associated SNP pairs varied among the ethnic groups, as depicted graphically in Fig. 2. The figure shows the presence of a well-defined haplotype block in the 5′-FR of the gene that is most clearly defined in the

Results

**GSTP1 gene resequencing.** We resequenced the exons, splice junctions, and ~400 bp of the 5′-FR of *GSTP1* using 240 DNA samples, 60 each from AA, CA, HCA, and MA subjects. A total of 35 SNPs were observed, including six nonsynonymous cSNPs (Fig. 1 and Table 1). There were striking variations in SNP types and frequencies among the ethnic groups studied. For example, there were two SNPs, 5′-FR (−219) and 5′-untranslated region (UTR) (−18) that had very high MAFs, but only in the MA population. Eight of the SNPs had very high minor allele frequencies (>40%) in one or more of the populations studied. All SNPs were in Hardy-Weinberg equilibrium. Although databases populated with polymorphism locations and frequencies are becoming increasingly complete and useful, there continues to be a need for systematic in-depth gene resequencing studies. When compared with the SNPs in dbSNP and the HapMap, 17 of the 35 SNPs that we observed were novel, and only 10 of our 35 SNPs were present in the HapMap (release 21a). The fact that only half of the SNPs that we identified were represented in publicly available databases underscores the continuing need for gene resequencing. Our gene resequencing data have been deposited in the NIH database PharmGKB (Submission ID PS205605).

\[ D^* \] values were calculated for all pairwise combinations of *GSTP1* SNPs, and haplotype analysis was performed. Values for \( D^* \) can range from 0 when SNPs are randomly associated, to 1.0, when they are maximally associated (39, 40). Forty-two pairs of SNPs for samples from AA, 58 SNP pairs for CA, 44 HCA SNP pairs, and 50 SNP pairs for MA subjects were observed to have both a \( D^* \) value of ≥0.8 and a \( P \) value <0.05. These maximally associated SNP pairs varied among the ethnic groups, as depicted graphically in Fig. 2. The figure shows the presence of a well-defined haplotype block in the 5′-FR of the gene that is most clearly defined in the

![Human GSTP1 Genetic Polymorphisms](image-url)

**Figure 1.** Human GSTP1 genetic polymorphisms. Colored rectangles, coding exons; open rectangles, UTR sequence. Arrows indicate the locations of polymorphisms, with frequencies indicated by the color of the arrow. Polymorphisms altering encoded amino acid sequence are indicated.
Four well-matched haplotypes containing one or more nonsynonymous cSNPs, listed in order of haplotype frequency (Table 2). For example, within this classification, the previously described *B haplotype with the Val¹⁰⁵ variant encompasses a number of haplotypes, designated here as the *2, *6, and *8 haplotype groups. The previously described *C haplotype that included both the Val¹⁰⁵ and Val¹¹⁴ polymorphisms would be the *3 group, and the previously described *D haplotype with Ile¹⁰⁵ and Val¹¹⁴ polymorphisms would be the *4 group, and the previously described *D haplotype with Ile¹⁰⁵ and Val¹¹⁴ is designated *4 (Table 2).

We also calculated nucleotide diversity, a measure of genetic variation, adjusted for the number of alleles studied. Two standard measures of nucleotide diversity are π, average heterozygosity per site, and θ, a population mutation measure that is theoretically equal to the neutral mutation variable. Values for Tajima’s D, a test

**Table 1. Human GSTP1 genetic polymorphisms**

<table>
<thead>
<tr>
<th>SNP Location</th>
<th>Nucleotide sequence change</th>
<th>Amino acid sequence change</th>
<th>Minor allele frequency</th>
<th>Present in databases</th>
<th>rs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 5’-FR (-383)</td>
<td>G to T</td>
<td>0.400 0.317 0.117 0.200 Yes</td>
<td>rs17593068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 5’-FR (-361)</td>
<td>G to T</td>
<td>0.400 0.317 0.117 0.208 Yes</td>
<td>rs36211087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 5’-FR (-360)</td>
<td>C to T</td>
<td>0.400 0.317 0.117 0.208 Yes</td>
<td>rs36211088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 5’-FR (-317)</td>
<td>del T</td>
<td>0.400 0.317 0.117 0.208 Yes</td>
<td>rs36211089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 5’-FR (-316)</td>
<td>C to A</td>
<td>0.400 0.317 0.117 0.208 Yes</td>
<td>rs36211089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 5’-FR (-311)</td>
<td>C to T</td>
<td>0.400 0.317 0.117 0.208 Yes</td>
<td>rs36211089</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 5’-FR (-219)</td>
<td>C to G</td>
<td>0.050 0.017 0.017 0.342 Yes</td>
<td>rs8191438</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 5’-FR (-192)</td>
<td>G to A</td>
<td>0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 5’-FR (-138)</td>
<td>C to T</td>
<td>0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 5’-FR (-103)</td>
<td>C to A</td>
<td>0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 5’-UTR (-20)</td>
<td>C to T</td>
<td>0.008 0.000 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 5’-UTR (-18)</td>
<td>G to A</td>
<td>0.025 0.017 0.017 0.342 Yes</td>
<td>rs8191439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 5’-UTR (-1)</td>
<td>C to T</td>
<td>0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 IVS1 (21)</td>
<td>T to G</td>
<td>0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 IVS1 (-28)</td>
<td>C to G</td>
<td>0.000 0.008 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 IVS1 (-20)</td>
<td>C to G</td>
<td>0.175 0.583 0.758 0.342 Yes</td>
<td>rs4147581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 IVS2 (34)</td>
<td>G to A</td>
<td>0.000 0.042 0.000 0.000 Yes</td>
<td>rs8191445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 IVS2 (103)</td>
<td>T to C</td>
<td>0.008 0.000 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 IVS2 (131)</td>
<td>G to C</td>
<td>0.008 0.000 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 IVS2 (134)</td>
<td>C to T</td>
<td>0.212 0.317 0.121 0.208 No</td>
<td>rs2370143</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 IVS2 (-108)</td>
<td>G to C</td>
<td>0.147 0.000 0.000 0.000 Yes</td>
<td>rs8191446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 Exon 3 (95)</td>
<td>A to T</td>
<td>Gln¹⁰⁵Val 0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 Exon 4 (150)</td>
<td>C to T</td>
<td>0.150 0.000 0.000 0.008 Yes</td>
<td>rs8191448</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 Exon 4 (172)</td>
<td>G to A</td>
<td>0.008 0.000 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 IVS4 (13)</td>
<td>C to A</td>
<td>0.407 0.325 0.117 0.208 Yes</td>
<td>rs762803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 Exon 5 (313)</td>
<td>A to G</td>
<td>Ile¹⁰⁵Val 0.425 0.317 0.108 0.533 Yes</td>
<td>rs1695</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 IVS5 (28)</td>
<td>T to G</td>
<td>0.000 0.000 0.000 0.008 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 IVS5 (112)</td>
<td>G to C</td>
<td>0.008 0.000 0.000 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29 Exon 6 (341)</td>
<td>C to T</td>
<td>Ala¹⁰⁵Val 0.017 0.083 0.000 0.008 Yes</td>
<td>rs1138272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 Exon 6 (372)</td>
<td>C to T</td>
<td>0.000 0.008 0.000 0.008 Yes</td>
<td>rs11553890</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 Exon 6 (439)</td>
<td>G to T</td>
<td>Asp¹⁴⁷Tyr 0.000 0.000 0.000 0.008 Yes</td>
<td>rs4966949</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 IVS6 (-18)</td>
<td>G to A</td>
<td>0.033 0.000 0.000 0.008 No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33 IVS6 (-16)</td>
<td>C to T</td>
<td>0.250 0.300 0.108 0.183 Yes</td>
<td>rs1871042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 Exon 7 (555)</td>
<td>T to C</td>
<td>0.458 0.300 0.117 0.542 Yes</td>
<td>rs4891</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35 Exon 7 (559)</td>
<td>C to T</td>
<td>Arg¹⁴⁷Trp 0.000 0.000 0.008 0.000 No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Polymorphism locations, nucleotide and amino acid sequence changes, variant allele frequencies, and presence in dbSNP are listed for each of the four ethnic groups. Polymorphisms in exons and UTRs are numbered with the “A” in the “ATG” translation initiation codon being designated “1.”

Numbers 5’ to that position are negative, whereas numbers 3’ are positive. Nucleotides located within introns (IVSs) are numbered based on their distance from splice junctions, with distances from 3’ splice junctions assigned positive numbers and distances from 5’ splice junctions assigned negative numbers. Exon nucleotides are “boxed.”
for the neutral mutation hypothesis (42, 43), were also calculated. Under conditions of neutrality, Tajima’s $D$ should equal 0. π and $D$ did not differ significantly among the four populations studied, and Tajima’s $D$ did not differ statistically from 0 in any of the populations (Supplementary Table A).

**GSTP1 microarray analysis.** Expression microarray studies were performed to determine whether a correlation might exist between GSTP1 genotype and basal mRNA expression in the lymphoblastoid cells from which the DNA used for gene resequencing had been obtained. There was wide variation in expression among individual samples (Fig. 3A). It has been shown that sequences that influence transcription can be located 5’ of a gene, in introns, or 3’ of a gene (45). Therefore, to assess polymorphisms that might alter GSTP1 mRNA expression, a SNP association analysis was performed to identify any polymorphisms throughout the length of the gene that might influence expression. That analysis showed that five SNPs (−219, −18, IVS1-20, Exon 5 +313, and Exon 7 +555) were significantly associated with level of GSTP1 expression ($P = 2.8 \times 10^{-4}$, $P = 1.1 \times 10^{-4}$, $P = 0.024$, and $P = 0.013$), and all four were associated with decreased GSTP1 expression for the variant nucleotide. When this analysis was performed separately for each of the four ethnic groups, these same five SNPs were associated with mRNA expression in the MA population, with lower $P$ values than in the overall analysis ($6.8 \times 10^{-9}$, $6.8 \times 10^{-9}$, $2.1 \times 10^{-4}$, $7.3 \times 10^{-7}$, and $3.0 \times 10^{-6}$, respectively), and in this case, all five SNPs remained significant after correction for multiple comparisons. None of the SNPs were significantly associated with GSTP1 mRNA expression when analyzed in the other three populations. That may be true because two of the five SNPs, (−219) and (−18), were found at much higher frequencies in MA subjects than in the other three populations.

Data for the (−18) 5’-FR SNP are displayed graphically in Fig. 3B. In the MA population, all five SNPs were tightly linked, with $D’$ of 1.0, $P < 0.0001$. Based on transcription factor binding prediction (TFSEARCH v.1.3), the (−219) variant nucleotide was predicted

![GSTP1 Haplotype Structure](image)

**Figure 2.** Human GSTP1 haplotype structure. The extent of population-specific linkage disequilibrium is depicted. **Red** combinations where $D’ = 1$ and linkage of disequilibrium (LOD) $> 2$; light red, combinations where $D’ < 1$ and LOD $> 2$; blue squares, combinations where $D’ = 1$ and LOD $< 2$; white squares, combinations with $D’ < 1$ and LOD $< 2$. SNPs are arranged in order from 5’ to 3’ in the gene, as shown in the gene structure above each plot.
to eliminate an E47 transcription factor binding site, whereas the (−18) SNP was predicted to cause the gain of an AML-1a binding site. Presence of the variant nucleotides for both of these SNPs was associated with a decrease in mRNA expression, and each separately explained 41% of the variation in GSTP1 mRNA expression in the MA population. Taken together, these five SNPs explained 51% of the variation in expression in the MA population. None of the other three SNPs were predicted to cause a change in transcription factor binding. For that reason, and because the (−219) and (−18) polymorphisms displayed the lowest P values in all of our analyses, we focused on those two SNPs in subsequent experiments. It should be emphasized that transcription regulation is highly tissue-specific and cell-specific, so the results shown in Fig. 3 apply only to these lymphoblastoid cells.

**GSTP1 EMSA.** To evaluate the (−219) and (−18) SNPs, which were predicted to possibly alter transcription factor binding sites, EMSAs were performed using oligonucleotides corresponding to both WT and variant sequences. A shift was observed when the (−18) variant sequence was exposed to lymphoblastoid cell nuclear extracts (Fig. 3C). This observation was consistent with the TFSEARCH prediction that the variant nucleotide at position (−18) would introduce an AML-1a binding site. However, we were unable to detect a super shift with commercially available AML-1 antibody (data not shown). No shift was observed with either WT or variant sequences at the (−219) position (data not shown).

**GSTP1 variant allozyme activity assay and substrate kinetics.** Functional genomic studies were also performed to explore the possible effects of GSTP1 nonsynonymous SNPs on function. WT GSTP1 and the eight variant allozymes, six with a single variant amino acid, and two with a combination of two amino acid variants each—based on the results of the haplotype analysis (Table 2)—were expressed in COS-1 cells. A mammalian cell line was used to perform these experiments to ensure the presence of mammalian posttranslational modification and protein

### Table 2. GSTP1 haplotypes

<table>
<thead>
<tr>
<th>Previous designation</th>
<th>Allele Frequencies</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
<th>5′-FR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA</td>
<td>CA</td>
<td>HCA</td>
<td>MA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*A</td>
<td>0.131</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.108</td>
<td>0.153</td>
<td>0.092</td>
<td>0.133</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.050</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.033</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.011</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.017</td>
<td>0.008</td>
<td>0.342</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.008</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>*A</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.008</td>
<td>G</td>
<td>G</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

**NOTE:** White type on black highlights variant nucleotides compared with those present in the most common haplotype in the AA population (the WT). Designations for haplotypes were made on the basis of alterations in the encoded amino acid sequence (number) and allele frequency (letters)—starting with the most common haplotype in the AA population. Numbers for variant allozymes were assigned from the NH2-terminus to the COOH-terminus of the protein. Letters refer to haplotypes encoding the same amino acid sequence, from most to least common, beginning with the most common haplotype in the AA population. A previous set of designations based only on nonsynonymous SNPs, excluding all other polymorphisms (previous designations), are also listed (25, 41).

Abbreviation: d, deletion.
Val105/Val114 variant allozymes also differed significantly from WT in apparent $K_m$ values (Fig. 4B). Although these differences in $K_m$ were statistically significant, they may not be biologically significant. Therefore, we also measured the level of immunoreactive protein for WT and each of the eight variant allozymes.

### GSTP1 Western blot analysis

GSTP1 recombinant allozymes were used to perform quantitative Western blot analyses. After correction for transfection efficiency, five of the variant allozymes with one or two changes in encoded amino acids displayed significantly decreased levels of immunoreactive protein when compared with the WT allozyme: Val32 at 71.1 ± 8.8% ($P < 0.05$), Asn58 at 77.3 ± 7.3% ($P < 0.05$), Val105 at 80.2 ± 2.5% ($P < 0.001$), Trp187 at 37.9 ± 4.7% ($P < 0.001$), and Val32/Val105 at 75.6 ± 8.0% ($P < 0.05$; Fig. 4C). These quantitative Western blot results correlated with the observed levels of GSTP1 enzyme activity with an $r_P$ value of 0.79 ($P < 0.007$) or 0.94 ($P < 0.0002$) if the outlier point corresponding to the Val105 allozyme was excluded (Fig. 4D). The next experiment was performed in an attempt to understand why several GSTP1 variant allozymes might have decreases in protein quantity, as well as significant reductions in level of activity. Although there are several possible explanations, accelerated protein degradation has been the most common mechanism responsible for these reductions in protein quantity and, thus, reductions in enzyme activity for a large number of variant allozymes for other enzymes that have been studied in detail (37, 38).

### In vitro translation/degradation

In an attempt to determine whether accelerated degradation might be responsible for decreases in enzyme activity and level of immunoreactive protein for GSTP1 variant allozymes, in vitro translation and degradation experiments for GSTP1 variant allozymes were performed with a rabbit reticulocyte lysate. Although we were able to synthesize...
radioactively labeled protein for all variant allozymes, we did not obtain evidence for accelerated degradation of the variant allozymes, although a positive control, the rapidly degraded TPMT*3A variant allozyme (37), was degraded very rapidly (data not shown).

**Discussion**

GSTP1 catalyzes the conjugation of reactive electrophiles with glutathione (4–6). It also participates in the regulation of MAPK pathways (18, 20, 22). GSTP1 is genetically polymorphic, and two GSTP1 nonsynonymous SNPs have been studied extensively in the epidemiology literature (27–29). For example, a simple PubMed search for GSTP1 polymorphism yielded 552 hits. In the present study, we set out to systematically identify common GSTP1 genetic variation in four ethnic groups, followed by functional genomic studies designed to begin to define the biological significance of that variation. Specifically, we resequenced GSTP1 exons, splice junctions, and a portion of the 5'-FR that contains the core promoter. We also performed functional genomic studies, including enzyme activity assays, substrate kinetics studies, Western blots, and mRNA expression array analysis.

---

**Figure 3.** GSTP1 mRNA expression and EMSA studies. A, GSTP1 mRNA expression in the lymphoblastoid cells from which the resequenced DNA was obtained. Each bar represents an individual sample and the height of the bar represents mRNA expression level as measured by Affymetrix U133 2.0 Plus GeneChip analysis. Data are color-coded by ethnic group. B, quantile boxplot (red) of expression by GSTP1 genotype at 5'-FR nucleotide (−18) in the MA samples. Green lines, group means; dark blue, error bars for each group. C, EMSA experiment showing the shift observed with the (−18) variant nucleotide in the presence of a pooled lymphoblastoid cell nuclear extract.
We identified 35 GSTP1 SNPs, half of which were not represented in databases such as dbSNP and the HapMap. The SNPs observed and their frequencies varied widely among ethnic groups, raising the possibility of variations in cancer risk and differential response to antineoplastic therapy among ethnic groups. In contrast, several novel SNPs were present in multiple ethnic groups with relatively high MAFs. These polymorphisms varied from nonsynonymous SNPs that altered the encoded amino acids to $\gamma$-FR polymorphisms that could potentially influence transcription. Microarray analysis showed large interindividual differences in GSTP1 mRNA expression in lymphoblastoid cells (Fig. 3A). Five SNPs that were significantly associated with level of GSTP1 mRNA in the MA population were identified. However, those SNPs did not seem to explain the full extent of variation in expression and were not significant in the other populations studied, most likely as a result of their lower frequencies in those populations—an observation that was especially striking for the two SNPs in the $\gamma$-FR, (−219) and (−18). Our EMSA studies showed that the variant nucleotide at (−18) resulted in protein binding that was not observed with the WT sequence (Fig. 3C). Because this functional SNP is tightly linked with the heavily studied Ile(105)Val polymorphism, it is possible that a portion of the association of the codon 105 polymorphism with various clinical phenotypes may result from the effect on transcription of the G(−18)A SNP. Although the five linked SNPs in the MA population are associated with $\approx$51% of GSTP1 variation in mRNA expression, there is considerable variation—especially in the populations with low frequency of these SNPs—which is not explained even in this one type of cell. The GSTP1 promoter has been shown to be hypermethylated in a variety of cancers, so one possible additional explanation for differences in mRNA expression could be that genetic polymorphisms result in variation in baseline promoter methylation status and, thus, mRNA expression. However, work by Han and colleagues indicates that, in the absence of disease, GSTP1 promoter methylation is conserved across tissue types and among individual subjects (49). Therefore, other variables, such as differences in transacting factors, posttranscriptional modification, and altered mRNA stability or transcription factor binding due to SNPs outside of the resequenced region, may be responsible for the remaining interindividual variation in mRNA expression.

After expression in COS-1 cells, several GSTP1 variant allozymes showed moderate decreases in enzyme activity, most strikingly for the extensively studied Val105 variant allozyme (Fig. 4A). Our laboratory has shown for a large number of cytosolic proteins that
enzyme function is most often influenced by the change in a single amino acid as a result of changes in protein levels (37, 38, 50–52). The results shown in Fig. 4D show a similar trend for GSTP1. When the mechanism responsible for this relationship has been studied, most often it has been found to result from an alteration in the rate of protein degradation. In this study, levels of GSTP1 enzyme activity for variant allotymes were highly correlated with levels of immunoreactive protein, r = 0.79 or 0.94, excluding Val105. The Val105 variant allotyme was an outlier in this analysis, which was not surprising because it has been reported previously that this amino acid change alters the active site of the enzyme and, as a result, substrate specificity (27, 31, 44, 46). Our substrate kinetic studies confirmed those reports and showed a significantly increased apparent $K_d$ of the Val105 variant allotyme with CDNBr as substrate (Fig. 4B). However, in vitro translation/degradation studies failed to show differences in rates of protein degradation for GSTP1 variant allotymes, including Val105. Of interest was the fact that double variant allotymes—those with both Val105 and an additional amino acid alteration—did not display as dramatic a decrease in enzyme activity as did the Val105 variant allotyme. What is clear is that GSTP1 represents an enzyme for which genetic variation in encoded amino acid sequence is responsible for alteration in level of enzyme activity as a result of at least two mechanisms—changes in the active site, e.g., Val105, and differences in levels of enzyme protein.

Finally, there has been increasing interest in interactions between GSTP1 and cell signaling molecules, such as JNK. Holley and colleagues recently reported differential effects of GSTP1 haplotypes on cell proliferation and apoptosis (34). Those investigators showed that GSTP1*A (WT) reduced cellular proliferation and was antiapoptotic through a JNK-independent mechanism, whereas GSTP1*C (Val105/Val114) did not influence proliferation and was antiapoptotic—through a JNK-mediated mechanism. These observations raise the possibility that other GSTP1 SNPs, such as those reported here, may alter interactions of GSTP1 with JNK, a hypothesis that should be the focus of future studies.

In summary, in the present study, we have applied a comprehensive and systematic genotype-to-phenotype research strategy to characterize common genetic variation in GSTP1, a gene that encodes a protein that plays diverse roles, from phase II drug metabolism to the regulation of apoptosis. Knowledge of common GSTP1 SNPs and haplotypes, as well as understanding of their functional implications, should contribute both to mechanistic and epidemiologic studies of the involvement of GSTP1 in carcinogenesis, as well as individual variation in response to antineoplastic drug therapy.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 12/19/2007; revised 3/14/2008; accepted 4/17/2008.

Grant support: NIH grants T32 GM24774, R01 GM28157, R01 GM35720, and U01 GM61388 (The Pharmacogenetics Research Network).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Luanne Wessouw for her assistance with the preparation of this article and Dr. Krishna (Kuni) Kalari for assistance with the linkage disequilibrium display shown in Fig. 2.

References


doly DNA increases the sensitivity of a colon cancer cell line to adriamycin, cisplatin, melphalan, and etoposide. Cancer Res 1996;56:3577–82.
28. Yang X, Pleifeir RM, Goldstein AM. Influence of glutathione-S-transferase (GSTM1, GSTTI, GSTT1) and cytochrome p50 (CYP1A1, CYP2D6) polymorphisms on numbers of basal cell carcinomas (BCCs) in families with the naevoid basal cell carcinoma syndrome. J Med Genet 2006;43:e16.
41. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002;70:425–34.
Glutathione S-Transferase P1: Gene Sequence Variation and Functional Genomic Studies


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/68/12/4791

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2008/06/06/68.12.4791.DC1

Cited articles
This article cites 50 articles, 21 of which you can access for free at:
http://cancerres.aacrjournals.org/content/68/12/4791.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/68/12/4791.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.