Published in final edited form as:
Cancer Res. 2008 June 15; 68(12): 4791–4801.

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

Ann M. Moyer[1], **Oreste E. Salavaggione**[1], **Tse-Yu Wu**[1], **Irene Moon**[1], **Bruce W. Eckloff**[2], **Michelle A. T. Hildebrandt**[1], **Daniel J. Schaid**[3], **Eric D. Wieben**[2], and **Richard M. Weinshilboum**[1]

1*Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Medical School-Mayo Clinic, Rochester, MN 55905, USA*

2*Department of Biochemistry and Molecular Biology, Mayo Medical School-Mayo Clinic, Rochester, MN 55905, USA*

3*Department of Health Sciences Research, Mayo Medical School-Mayo Clinic, Rochester, MN 55905, USA*

Abstract

Glutathione S-transferase P1 (*GSTP1*) is of importance for cancer research because of its role in detoxifying carcinogens, activating antineoplastic prodrugs, metabolizing chemotherapeutic agents, and because of 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 (SNPs) 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 to 22% of the wild type (WT) activity. Four variant allozymes had K_m values that differed significantly from that of the WT, and five showed altered levels of immunoreactive protein as compared to 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 EMSA 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.

Keywords

Glutathione S-transferases; glutathione transferase; GST; GSTP1; gene resequencing; functional genomics; genetic polymorphisms; SNPs; variant allozymes; expression regulation

To whom correspondence should be addressed: Richard Weinshilboum, M.D., 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.

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). Variation in the expression and activity of GSTP1 has been associated with a variety of human cancers (7-12). GSTP1 can metabolize and inactivate a number of antineoplastic drugs (4,11,13-16), but it can also activate antineoplastic prodrugs such as TLK286/TER286 (17). Finally, GSTP1 has been shown to function not only as a phase II drug-metabolizing enzyme, but also as a regulator of mitogen-activated protein kinases (MAPKs) as a result of non-enzymatic, ligand-binding activity (18-22).

Because of its multiple roles, the genetic regulation of *GSTP1* has been studied extensively, particularly two common nonsynonymous polymorphisms and the methylation state of a CpG island in its promoter. The CpG island in the *GSTP1* promoter has been reported to be hypermethylated in prostate, hepatocellular, breast, renal, lung, and colon cancer, as well as some lymphomas (7,9,10,12,23). Hypermethylation results in reduced GSTP1 expression (24). When the expression of GSTP1 decreases, it has been speculated that cells become more susceptible to mutation and damage as a result of exposure to electrophiles and oxidative stress (25,26).

GSTP1 also has two common nonsynonymous single nucleotide polymorphisms (cSNPs) that result in Ile105Val and Ala114Val alterations in encoded amino acid sequence. These SNPs have been associated with variation in cancer risk and in clinical response to antineoplastic drug therapy (11,27-30). The Val105 substitution results in steric restriction of the H-site due to shifts in the side chains of several amino acids. Thus, the Val105 variant allozyme may be able to accommodate less bulky substrates than the Ile105 allozyme, and, as a result may display substrate specificities that differ from those of the WT allozyme (27,31). In addition, the thermal stability of the codon 105 variant allozyme differs from that of the WT (27,31). These characteristics may be responsible, in part, for the reported association between this allozyme and carcinogenesis or variation in response to antineoplastic drugs (8,32,33). As an example of the interest in the genetics of GSTP1, the CDC HuGEpedia (http://www.hugenavigator.net) lists 462 publications reporting associations of genetic variation in *GSTP1* with health, associations involving 201 human diseases. Finally, GSTP1 appears to have functional roles that extend beyond phase II drug metabolism. Specifically, GSTP1 is an inhibitor of JNK and TRAF2 (18,20,21), resulting in alterations in downstream processes such as cell cycle control and apoptosis (34).

Although *GSTP1* is clearly an important gene, a systematic study of common genetic variation in this gene and characterization of the functional significance of that variation have not been reported. Therefore, we set out to systematically identify common DNA 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 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, HD100CHI, and HD100MEX) were obtained from the Coriell Cell Repository (Camden, NJ). The National Institute of General Medical Sciences had anonymized these DNA samples prior to 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 approximately 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, Valencia, CA). RNA quality assessment was performed using the Agilent 2100 bioanalyzer prior to microarray analysis. All RNA samples had an Agilent RNA Integrity Number (RIN) greater than 9.0. The RNA was then reverse-transcribed and biotin labeled for hybridization with Affymetrix U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, CA). The microarray images were analyzed with quality control techniques established in the Mayo Clinic Microarray Core Facility and the data were normalized using Fastlo, a version of cyclic loess normalization (35). Data from probe set 200824_at, corresponding to GSTP1, were used in the analyses described here.

GSTP1 electrophoretic mobility shift assay (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, Rockford, IL). Nuclear extracts were prepared from a pool of the lymphoblastoid cell lines used to perform the microarray analyses. Specifically, 15 million cells were pelleted at 524 × g for 3 minutes. Cells were washed in 1 mL of cold PBS, and repelleted. The pellet was resuspended in 200 μL of ice-cold lysis buffer consisting of 10 mM HEPES (pH 7.9), 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.4% NP-40, 1 mM DTT, 0.5 mM PMSF, and protease inhibitor. After incubation on ice for 15 minutes, the lysate was centrifuged at 14,000 × g for 30 seconds. After washing, the pellet was resuspended in a buffer containing 20 mM HEPES (pH 7.9), 400 mM NaCl, 1 mM EDTA, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 0.5 mM PMSF, and protease inhibitor, and was vortexed for 15 seconds at 4°C. The mixture was then centrifuged at 14,000 × g for 10 minutes and was stored at -80°C. An antibody against AML-1 (Santa Cruz Biotechnology, Santa Cruz, CA) was used in an attempt

to perform a "super shift" assay. Specifically, 2 μL of antibody was added to 20 μL of the binding reaction, followed by incubation at room temperature for 45 minutes.

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 (ORF) into the expression vector pcDNA4/HisMax (Invitrogen, Carlsbad, CA). 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 Dulbecco's Modified Eagle's Medium using the TransFast reagent (Promega, Madison, WI) 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 post-translational modification and degradation would be present. Specifically, 7 μg of construct DNA was co-transfected with 7 μg of pSV-(β-galactosidase DNA (Promega) as a control to correct for possible variation in transfection efficiency. After 48 hours, the cells were washed with phosphate-buffered saline, resuspended in 1 ml of homogenization buffer, and lysed with a Polytron homogenizer (Brinkmann Instruments, Westbury, NY). The homogenates were centrifuged at 100,000 × g at 4°C for 1 hour. 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, St. Lous, MO) as the substrate for each recombinant GSTP1 allozyme, using a modification of the spectrophotometric assay described by Habig *et al*. (36). Specifically, 100 mM potassium phosphate buffer (pH 6.5), 1 mM ethylenediaminetetraacetic acid, 2.5 mM reduced GSH (Sigma-Aldrich), 1.0 mM CDNB (dissolved in 95% EtOH, 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 prior to the addition of the enzyme, and was incubated at room temperature for 2 minutes. The enzyme source was then added and the mixture was incubated at room temperature for 7 minutes, followed by the measurement of absorbance at 340 nm in a Safire^2 microplate reader (Tecan, Männedorf, Switzerland). The complete reaction mixture including COS-1 cell highspeed supernatant from cells transfected with "empty vectors" served as a blank. The signal-to-noise ratio was 2:1 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 mM but only over the course of a 300 second 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, St. Louis, MO) was used to visualize protein bands. Specifically, COS-1 cell cytosol was loaded onto 12% SDS mini-gels (Bio-Rad, Hercules, CA) on the basis of levels of the cotransfected β-galactosidase activity to correct for possible variation in transfection efficiency. Electrophoresis was performed for 90 minutes at 120 V, followed by the transfer of proteins to nitrocellulose membranes. After blocking for two hours with 5% powdered milk in Tris-buffered saline with Tween 20 (TBST), the membranes were incubated overnight with

primary antibody diluted 1:20,000 with 5% powdered milk in TBST. The next morning, after three washes, goat anti-mouse horseradish peroxidase antibody (Bio-Rad, Hercules, CA) was applied for two hours at a dilution of 1:10,000, followed by three washes. The ECL Western Blotting System (Amersham Biosciences, Piscataway, NJ) was then used to detect bound antibody by enhanced chemiluminescence. The Western blot data were analyzed with the AutoChemi System (UVP BioImaging Systems, Upland, CA). Multiple independent blots were performed for each allozyme, and the results were expressed as a percentage of the intensity of the WT allozyme on the same gel.

In vitro translation/degradation assay

Transcription and translation of GSTP1 allozymes were performed with the TNT® Coupled Rabbit Reticulocyte Lysate (RRL) System (Promega) in the presence of ^{35}S-methionine and cysteine (1000 Ci/mmol, 2.5 mCi total) (Amersham Biosciences). This reaction mixture was incubated at 37°C for 90 minutes, and 5 μL aliquots were used to perform SDS-PAGE. After transcription and translation of the allozymes, protein degradation experiments were performed as described previously (37). Specifically, 10 μL of *in vitro* translated ^{35}S-methionine and cysteine-labeled protein was added to 40 μL of an ATP-generating system and 40 μL of "untreated" RRL. During incubation at 37°C, 10 μL aliquots were removed at 0, 4, 8, and 24 hours, followed by SDS-PAGE and autoradiography. Radioactively labeled protein was quantified using the AutoChemi System (UVP BioImaging Systems). The rapidly degraded protein TPMT*3A (38) was used as a positive control for the degradation studies.

Data analysis

Sequence chromatograms from the gene resequencing studies were analyzed using Mutation Surveyor (Softgenetics, State College, PA). Linkage disequilibrium among *GSTP1* polymorphisms was determined by calculating *D′* values (39,40), and intragene haplotypes were inferred using the method described by Schaid et. al. (41). Graphical representations of population-specific haplotype block patterns across *GSTP1* were generated using Haploview. Values for π, θ, and Tajima's *D* were determined as described by Tajima (42,43). Genotype-phenotype associations for *GSTP1* SNPs were evaluated by race and overall, i.e. for all samples. A linear model was used to assess the association of each SNP with the quantitative phenotype of GSTP1 expression. For these analyses, SNP genotypes were evaluated with a one degree-of-freedom test, with SNP genotypes coded as 0, 1, or 2 on the basis of the number of rare variants present. GSTP1 expression values for the genotype-phenotype association studies were $\log_2$ transformed and adjusted for gender, race, and storage time. Apparent K_m values were calculated using the GraphPad Prizm 3.0 computer program (GraphPad, San Diego, CA). Average levels of recombinant allozyme activity, K_m values, and immunoreactive protein levels were compared with values for the WT allozyme by the use of Student's *t*-test.

Results

GSTP1 gene resequencing

We resequenced the exons, splice junctions, and approximately 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 6 nonsynonymous cSNPs (Figure 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′-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 indepth gene resequencing studies. When compared to 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). 42 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 ≥ 0.8 and a p value < 0.05. These maximally associated SNP pairs varied among the ethnic groups, as depicted graphically in Figure 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 MA samples. Although this block does not include the frequently studied Ile105Val polymorphism, that SNP is tightly linked to this haplotype block (|D′|=1.0, p<0.001).

GSTP1 haplotypes with a frequency of ≥1% in any population, as well as haplotypes containing one or more nonsynonymous cSNPs, are listed in Table 2. Four *GSTP1* haplotypes that involved only nonsynonymous SNPs had previously been designated *A, *B, *C, and *D (27,44). However, because of the large number of SNPs that we had identified, that classification was inadequate to define the full spectrum of common variation in *GSTP1* since those designations often included several haplotypes. Therefore, we have used a system of nomenclature in which we have assigned numbers corresponding to the nonsynonymous cSNPs, numbered in order from the 5′ to the 3′ end of the gene, with letters corresponding to varying combinations of SNPs in addition to the 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 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 parameter. Values for Tajima's *D*, a test for the neutral mutation hypothesis (42,43), were also calculated. Under conditions of neutrality, Tajima's D should equal 0. π and θ did not differ significantly among the 4 populations studied, and Tajima's D did not differ statistically from zero in any of the populations (Supplemental 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 (Figure 3A). It has been demonstrated 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 ($p=7.9\times10^{-6}$, 3.2×10^{-6}, 2.4×10^{-3}, 6.9×10^{-4}, and 3.6×10^{-4}, respectively). Four of these SNPs were still significant after correction for multiple comparisons (-219, $p=2.8\times10^{-4}$, -18, $p=1.1\times10^{-4}$, Exon 5 +313, p=0.024, and Exon 7 +555, 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 4 ethnic groups, these same 5 SNPs were associated with mRNA

expression in the MA population, with lower p values than in the overall analysis (6.8×10^{-9}, 6.8×10^{-9}, 2.1×10^{-4}, 7.3×10^{-7}, and 3.0×10^{-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 Figure 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 to eliminate an E47 transcription factor binding site, while 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 approximately 41% of the variation in GSTP1 mRNA expression in the MA population. Taken together, these 5 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 and cell-specific, so the results shown in Figure 3 apply only to these lymphoblastoid cells.

GSTP1 electrophoretic mobility shift assay (EMSA)

To evaluate the (-219) and (-18) SNPs, which were predicted to possibly alter transcription factor binding sites, EMSA 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 (Figure 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 cSNPs on function. WT GSTP1 and the 8 variant allozymes, 6 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 post-translational modification and protein degradation systems. That was important because it has been demonstrated repeatedly that the alteration of only one or two amino acids as a result of genetic polymorphisms can be associated with drastic changes in the level of protein, often due to rapid protein degradation through a ubiquitin-proteasome-mediated process (37,38). After correction for transfection efficiency, the Val105 and Trp187 variants showed the most striking changes in activity with CDNB as the substrate as compared with the WT allozyme, with 21.8 ± 4.3% ($p<1\times10^{-6}$) and 55.2 ± 5.9% ($p<0.001$) of the activity of the WT allozyme, respectively. The Val114 variant displayed 79.9 ± 5.1% ($p<0.05$), the Val32/Val105 double variant 65.3 ± 5.7% ($p<0.001$), and the Val105/Val114 double variant 74.1 ± 5.2% ($p<0.01$) of the activity of the WT GSTP1 allozyme (Figure 4A).

One way in which changes in the encoded amino acid might alter enzyme activity is through changes in substrate kinetics. Therefore, we also determined apparent K_m values for the WT and variant allozymes after transient expression in COS-1 cells. As reported previously (27, 29,46,47), the Val105 variant had a significantly higher apparent K_m value, 1.51 ± 0.20 mM, than did the WT Ile105 protein, 0.92 ± 0.10 mM (Figure 4B). The Asn58, Val114, and Val105/

Val114 variant allozymes also differed significantly from WT in apparent K_m values (Figure 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 8 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 to 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% ($p<0.05$) (Figure 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 (Figure 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, 48).

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 even though 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.

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 5′-FR polymorphisms that could potentially influence transcription. Microarray analysis demonstrated large interindividual differences in GSTP1 mRNA expression in lymphoblastoid cells (Figure 3A). Five SNPs were identified that were significantly associated with level of GSTP1 mRNA in the MA population. However, those

SNPs did not appear 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 5′-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 (Figure 3C). Since 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 approximately 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 by 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 et. al. 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 trans-acting 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 demonstrated moderate decreases in enzyme activity, most strikingly for the extensively studied Val105 variant allozyme (Figure 4A). Our laboratory has demonstrated 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 Figure 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 (37, 38, 50-52). In this study, levels of GSTP1 enzyme activity for variant allozymes were highly correlated with levels of immunoreactive protein, r=0.79, or 0.94 excluding Val105. The Val105 variant allozyme was an outlier in this analysis, which was not surprising since 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_m of the Val105 variant allozyme with CDNB as substrate (Figure 4B). However, *in vitro* translation/degradation studies failed to show differences in rates of protein degradation for GSTP1 variant allozymes, including Val105. Of interest was the fact that “double variant” allozymes – 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 allozyme. 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 signalling molecules such as JNK. Holley et. al. recently reported differential effects of GSTP1 haplotypes on cell proliferation and apoptosis (34). Those investigators demonstrated that GSTP1*A (WT) reduced cellular proliferation and was anti-apoptotic through a JNK-independent mechanism, while GSTP1*C (Val105/Val114) did not influence proliferation and was anti-apoptotic – once again 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Luanne Wussow for her assistance with the preparation of this article, and Dr. Krishna (Rani) Kalari for assistance with the linkage disequilibrium display shown in Figure 2.

Supported in part by National Institutes of Health (NIH) grants T32 GM72474, R01 GM28157, R01 GM35720, and U01 GM61388 (The Pharmacogenetics Research Network).

References

1. Cowell IG, Dixon KH, Pemble SE, Ketter B, Taylor JB. The structure of the human glutathione S-transferase pi gene. Biochem J 1988;255:79–83. [PubMed: 3196325]
2. Kano T, Sakai M, Muramatsu M. Structure and expression of a human class pi glutathione S-transferase messenger RNA. Cancer Res 1987;47:5626–30. [PubMed: 3664469]
3. Mannervik B, Danielson UH. Glutathione transferases--structure and catalytic activity. CRC Crit Rev Biochem 1988;23:283–337. [PubMed: 3069329]
4. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. Annu Rev Pharmacol Toxicol 2005;45:51–88. [PubMed: 15822171]
5. Keen JH, Jakoby WB. Glutathione transferases. Catalysis of nucleophilic reactions of glutathione. J Biol Chem 1978;253:5654–7. [PubMed: 670218]
6. Ketterer B. A bird's eye view of the glutathione transferase field. Chem Biol Interact 2001;138:27–42. [PubMed: 11640913]
7. Esteller M, Corn PG, Urena JM, Gabrielson E, Baylin SB, Herman JG. Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. Cancer Res 1998;58:4515–8. [PubMed: 9788592]
8. Harries LW, Stubbins MJ, Forman D, HG C, Wolf CR. Identification of genetic polymorphisms at the glutathione S-transferase Pi locus and association with susceptibility to bladder, testicular and prostate cancer. Carcinogenesis 1997;18:641–4. [PubMed: 9111193]
9. Lee JS. GSTP1 promoter hypermethylation is an early event in breast carcinogenesis. Virchows Arch 2007;450:637–42. [PubMed: 17479284]
10. Lee WH, Morton RA, Epstein JI, et al. Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci U S A 1994;91:11733–7. [PubMed: 7972132]
11. McIlwain CC, Townsend DM, Tew KD. Glutathione S-transferase polymorphisms: cancer incidence and therapy. Oncogene 2006;25:1639–48. [PubMed: 16550164]
12. Zhong S, Tang MW, Yeo W, Liu C, Lo YM, Johnson PJ. Silencing of GSTP1 gene by CpG island DNA hypermethylation in HBV-associated hepatocellular carcinomas. Clin Cancer Res 2002;8:1087–92. [PubMed: 11948118]
13. Ban N, Takahashi Y, Takayama T, et al. Transfection of glutathione S-transferase (GST)-pi antisense complementary DNA increases the sensitivity of a colon cancer cell line to adriamycin, cisplatin, melphalan, and etoposide. Cancer Res 1996;56:3577–82. [PubMed: 8758929]
14. Goto S, Iida T, Cho S, Oka M, Kohno S, Kondo T. Overexpression of glutathione S-transferase pi enhances the adduct formation of cisplatin with glutathione in human cancer cells. Free Radic Res 1999;31:549–58. [PubMed: 10630679]

15. Tidefelt U, Elmhorn-Rosenborg A, Paul C, Hao XY, Mannervik B, Eriksson LC. Expression of glutathione transferase pi as a predictor for treatment results at different stages of acute nonlymphoblastic leukemia. Cancer Res 1992;52:3281–5. [PubMed: 1596886]
16. Waxman DJ. Glutathione S-transferases: role in alkylating agent resistance and possible target for modulation chemotherapy--a review. Cancer Res 1990;50:6449–54. [PubMed: 2208103]
17. Morgan AS, Sanderson PE, Borch RF, et al. Tumor efficacy and bone marrow-sparing properties of TER286, a cytotoxin activated by glutathione S-transferase. Cancer Res 1998;58:2568–75. [PubMed: 9635580]
18. Adler V, Yin Z, Fuchs SY, et al. Regulation of JNK signaling by GSTp. Embo J 1999;18:1321–34. [PubMed: 10064598]
19. Ketley JN, Habig WH, Jakoby WB. Binding of nonsubstrate ligands to the glutathione S-transferases. J Biol Chem 1975;250:8670–3. [PubMed: 1184584]
20. Wang T, Arifoglu P, Ronai Z, Tew KD. Glutathione S-transferase P1-1 (GSTP1-1) inhibits c-Jun N-terminal kinase (JNK1) signaling through interaction with the C terminus. J Biol Chem 2001;276:20999–1003. [PubMed: 11279197]
21. Wu Y, Fan Y, Xue B, et al. Human glutathione S-transferase P1-1 interacts with TRAF2 and regulates TRAF2-ASK1 signals. Oncogene 2006;25:5787–800. [PubMed: 16636664]
22. Zhao X, Fan Y, Shen J, Wu Y, Yin Z. Human glutathione S-transferase P1 suppresses MEKK1-mediated apoptosis by regulating MEKK1 kinase activity in HEK293 cells. Mol Cells 2006;21:395–400. [PubMed: 16819303]
23. Millar DS, Ow KK, Paul CL, Russell PJ, Molloy PL, Clark SJ. Detailed methylation analysis of the glutathione S-transferase pi (GSTP1) gene in prostate cancer. Oncogene 1999;18:1313–24. [PubMed: 10022813]
24. Lin X, Tascilar M, Lee WH, et al. GSTP1 CpG island hypermethylation is responsible for the absence of GSTP1 expression in human prostate cancer cells. Am J Pathol 2001;159:1815–26. [PubMed: 11696442]
25. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. Nature 1997;386:761–3. [PubMed: 9126728]
26. Meiers I, Shanks JH, Bostwick DG. Glutathione S-transferase pi (GSTP1) hypermethylation in prostate cancer: review 2007. Pathology 2007;39:299–304. [PubMed: 17558856]
27. Ali-Osman F, Akande O, Antoun G, Mao JX, Buolamwini J. Molecular cloning, characterization, and expression in Escherichia coli of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. J Biol Chem 1997;272:10004–12. [PubMed: 9092542]
28. Lecomte T, Landi B, Beaune P, Laurent-Puig P, Loriot MA. Glutathione S-transferase P1 polymorphism (Ile105Val) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. Clin Cancer Res 2006;12:3050–6. [PubMed: 16707601]
29. Yang XR, Pfeiffer RM, Goldstein AM. Influence of glutathione-S-transferase (GSTM1, GSTP1, GSTT1) and cytochrome p450 (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. [PubMed: 16582078]
30. Tew KD. Glutathione-associated enzymes in anticancer drug resistance. Cancer Res 1994;54:4313–20. [PubMed: 8044778]
31. Johansson AS, Stenberg G, Widersten M, Mannervik B. Structure-activity relationships and thermal stability of human glutathione transferase P1-1 governed by the H-site residue 105. J Mol Biol 1998;278:687–98. [PubMed: 9600848]
32. Hengstler JG, Arand M, Herrero ME, Oesch F. Polymorphisms of N-acetyltransferases, glutathione S-transferases, microsomal epoxide hydrolase and sulfotransferases: influence on cancer susceptibility. Recent Results Cancer Res 1998;154:47–85. [PubMed: 10026993]
33. Stoehlmacher J, Park DJ, Zhang W, et al. Association between glutathione S-transferase P1, T1, and M1 genetic polymorphism and survival of patients with metastatic colorectal cancer. J Natl Cancer Inst 2002;94:936–42. [PubMed: 12072547]

34. Holley SL, Fryer AA, Haycock JW, Grubb SE, Strange RC, Hoban PR. Differential effects of glutathione S-transferase pi (GSTP1) haplotypes on cell proliferation and apoptosis. Carcinogenesis 2007;28:2268–73. [PubMed: 17557902]
35. Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. Bioinformatics 2004;20:2778–86. [PubMed: 15166021]
36. Habig WH, Pabst MJ, Jakoby WB. Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. J Biol Chem 1974;249:7130–9. [PubMed: 4436300]
37. Wang L, Sullivan W, Toft D, Weinshilboum R. Thiopurine S-methyltransferase pharmacogenetics: chaperone protein association and allozyme degradation. Pharmacogenetics 2003;13:555–64. [PubMed: 12972954]
38. Weinshilboum R, Wang L. Pharmacogenetics: inherited variation in amino acid sequence and altered protein quantity. Clin Pharmacol Ther 2004;75:253–8. [PubMed: 15080131]
39. Hartl, D.; Clark, AG. Chapter 3. Organization of genetic variation. Principles of Population Genetics. Sunderland, MA: Sinauer Associates, Inc.; 2000. p. 95-107.
40. Hendrick, P. Genetics of Populations. Sudbury, MA: Jones and Bartlett Publishing; 2000. p. 396-405.
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. [PubMed: 11791212]
42. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989;123:585–95. [PubMed: 2513255]
43. Tajima F. DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. Genetics 1989;123:229–40. [PubMed: 2806885]
44. Watson MA, Stewart RK, Smith GB, Massey TE, Bell DA. Human glutathione S-transferase P1 polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. Carcinogenesis 1998;19:275–80. [PubMed: 9498276]
45. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. Nature 2005;437:1365–9. [PubMed: 16251966]
46. Hu X, Xia H, Srivastava SK, et al. Activity of four allelic forms of glutathione S-transferase hGSTP1-1 for diol epoxides of polycyclic aromatic hydrocarbons. Biochem Biophys Res Commun 1997;238:397–402. [PubMed: 9299520]
47. Zimniak P, Nanduri B, Pikula S, et al. Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties. Eur J Biochem 1994;224:893–9. [PubMed: 7925413]
48. Hildebrandt MA, Carrington DP, Thomae BA, et al. Genetic diversity and function in the human cytosolic sulfotransferases. Pharmacogenomics J 2007;7:133–43. [PubMed: 16801938]
49. Han W, Cauchi S, Herman JG, Spivack SD. DNA methylation mapping by tag-modified bisulfite genomic sequencing. Anal Biochem 2006;355:50–61. [PubMed: 16797472]
50. Ji Y, Salavaggione OE, Wang L, et al. Human phenylethanolamine N-methyltransferase pharmacogenomics: gene re-sequencing and functional genomics. J Neurochem 2005;95:1766–76. [PubMed: 16277617]
51. Salavaggione OE, Wang L, Wiepert M, Yee VC, Weinshilboum RM. Thiopurine S-methyltransferase pharmacogenetics: variant allele functional and comparative genomics. Pharmacogenet Genomics 2005;15:801–15. [PubMed: 16220112]
52. Thomae BA, Rifki OF, Theobald MA, Eckloff BW, Wieben ED, Weinshilboum RM. Human catecholamine sulfotransferase (SULT1A3) pharmacogenetics: functional genetic polymorphism. J Neurochem 2003;87:809–19. [PubMed: 14622112]

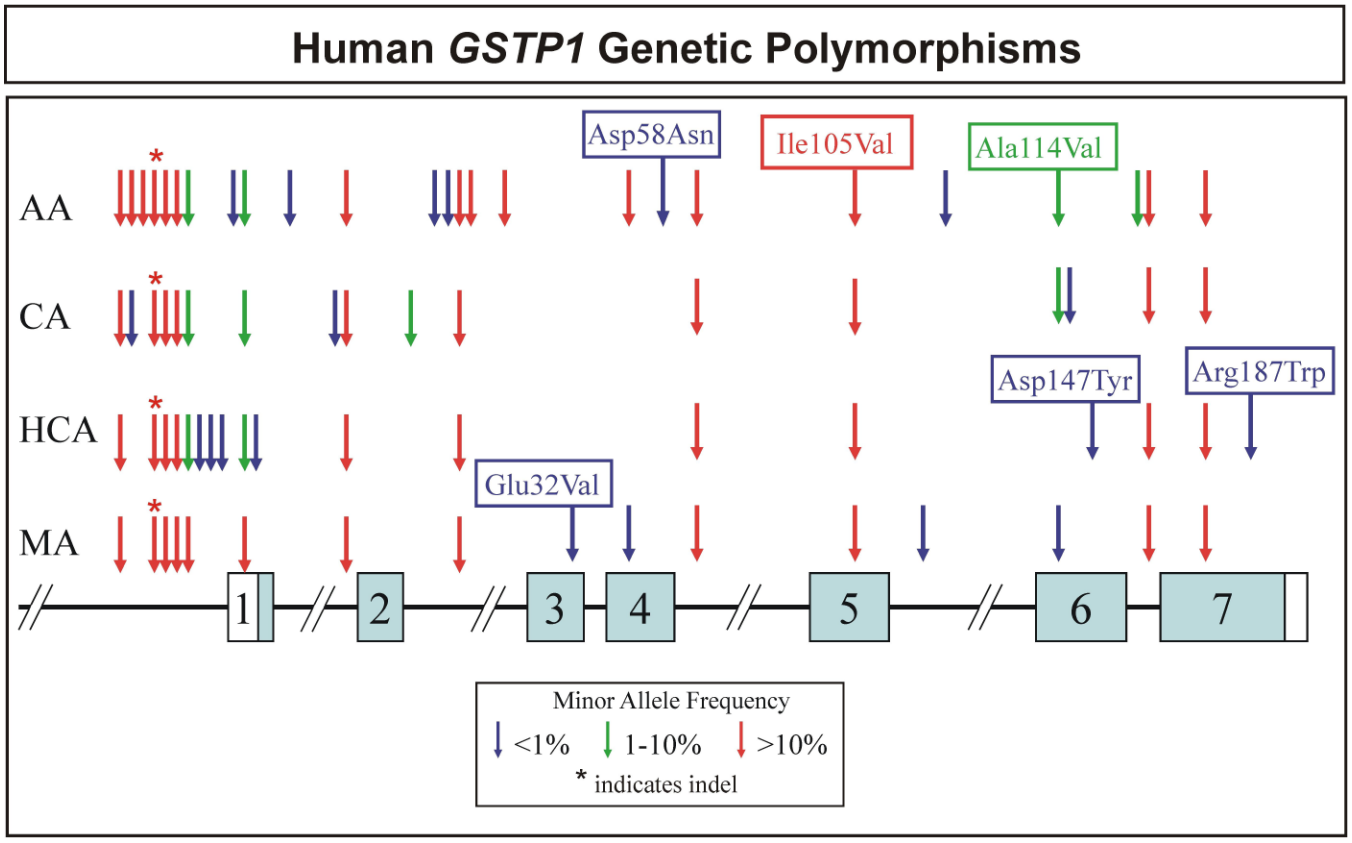

Figure 1. Human *GSTP1* genetic polymorphisms

Colored rectangles represent coding exons, and open rectangles represent 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.

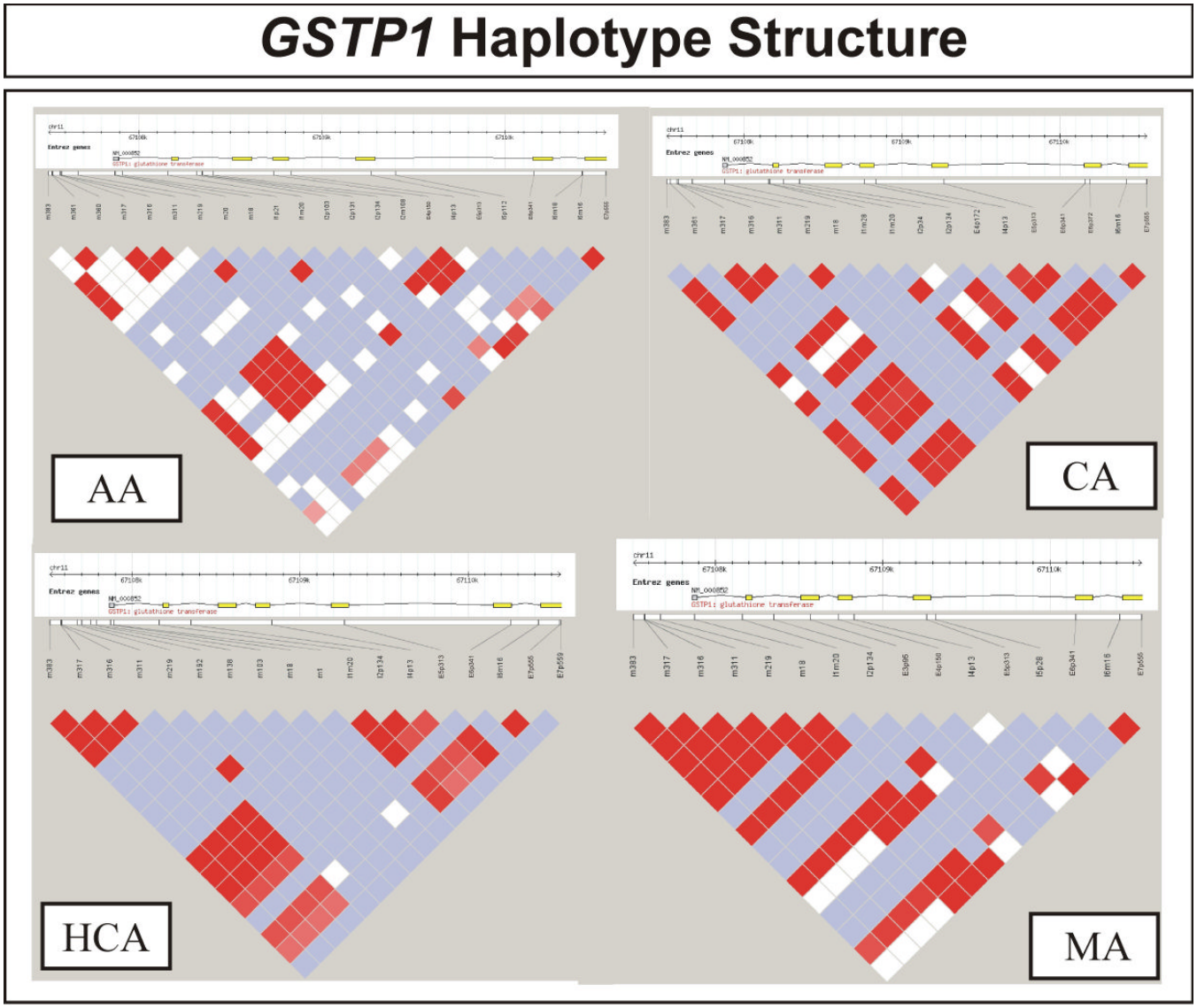

Figure 2. Human *GSTP1* haplotype structure

The extent of population-specific linkage disequilibrium is depicted. Colored in red are combinations where D′=1 and LOD≥2 and in light red, D′<1, LOD≥2. Blue squares represent combinations where D′=1 and LOD<2, while white squares are 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.

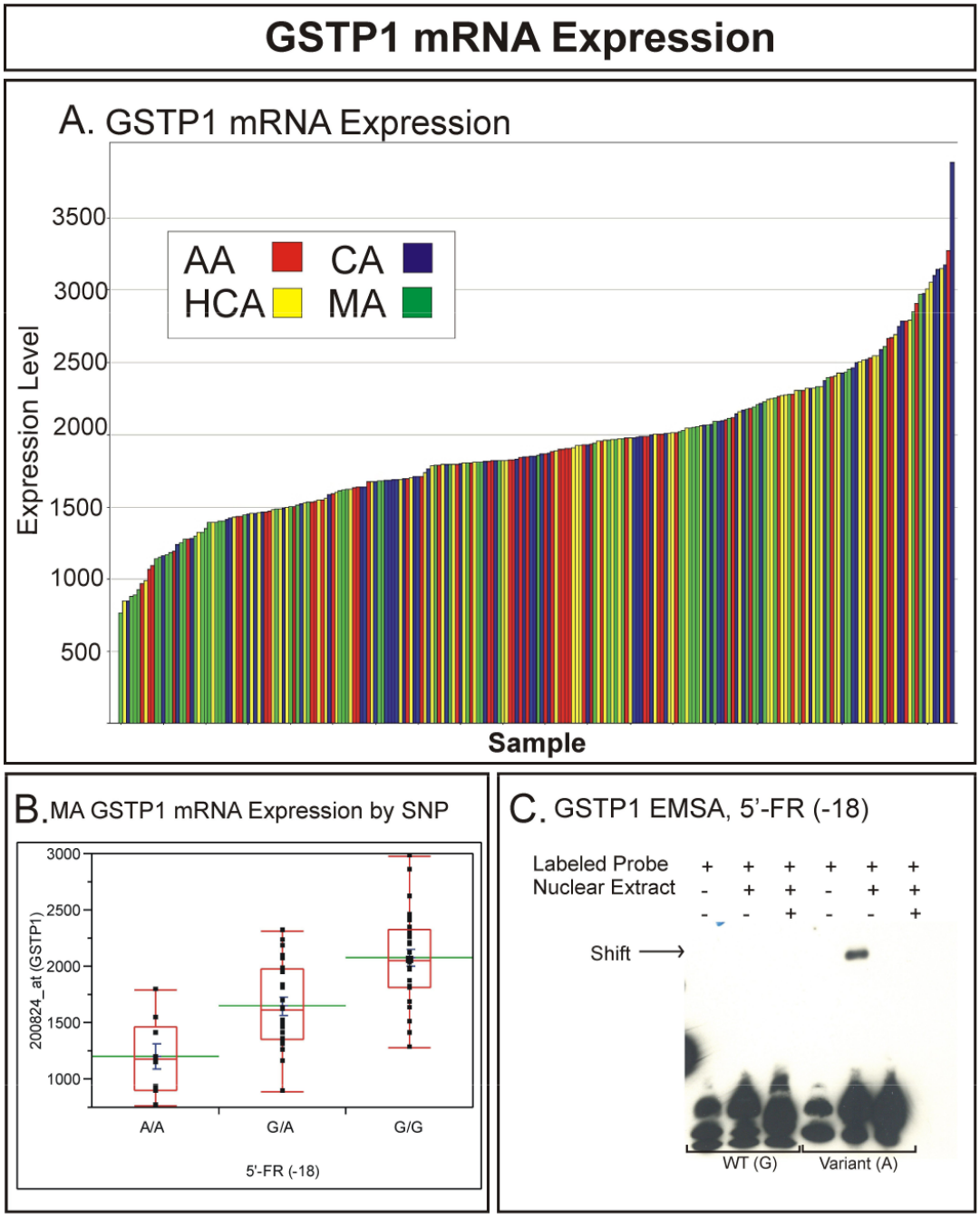

Figure 3. GSTP1 mRNA expression and EMSA studies

(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 (in red) of expression by GSTP1 genotype at 5′-FR nucleotide (-18) in the MA samples. Green lines indicate group means, with error bars for each group in dark blue. (C) EMSA experiment showing the “shift” observed with the (-18) variant nucleotide in the presence of a pooled lymphoblastoid cell nuclear extract.

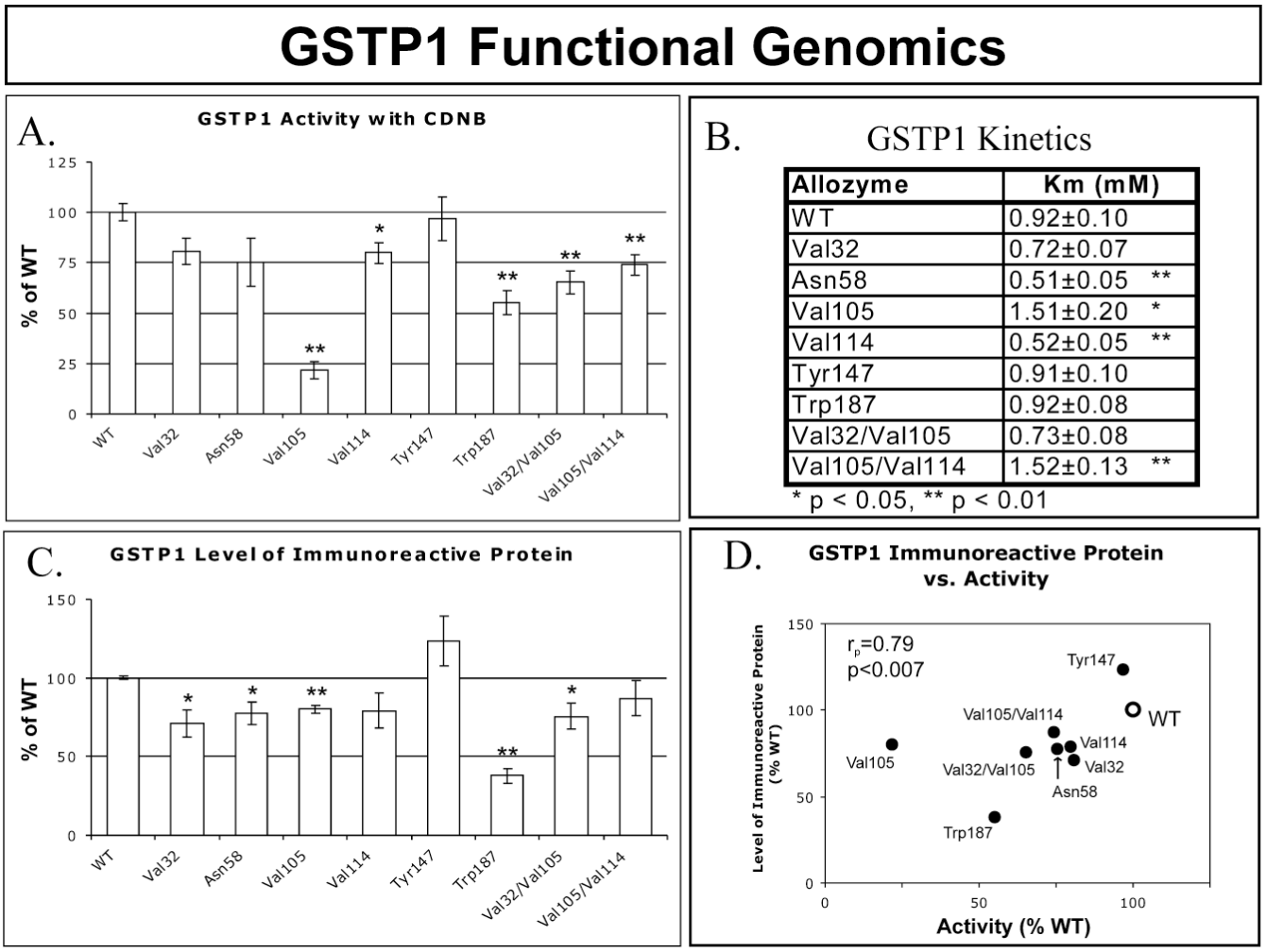

Figure 4. Human GSTP1 recombinant allozyme enzyme activity levels, substrate kinetics and immunoreactive protein levels

Average levels of enzyme activity (A) and immunoreactive protein (C) expressed as a percentage of the respective WT protein after the transfection of COS-1 cells are shown. Each bar represents the average of six independent transfections (mean ± SEM). * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$ when compared to the value for the WT construct. (B) GSTP1 allozyme apparent K_m values. Values are represented as mean ± SEM for three independent determinations. (D) Correlation between human GSTP1 recombinant allozyme enzyme activity and immunoreactive protein after expression in COS-1 cells.

Table 1

Human *GSTP1* genetic polymorphisms. 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, while 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."

SNP	Polymorphism Location	Nucleotide Sequence Change	Amino Acid Sequence Change	Minor Allele Frequency				Present in Databases	rs#
				AA	*CAU*	*HCA*	*MA*		
1	5′-FR (-383)	G to T		*0.400*	*0.317*	*0.117*	*0.200*	Yes	rs17593068
2	5′-FR (-361)	G to T		*0.142*	*0.008*	*0.000*	*0.000*	No	
3	5′-FR (-360)	C to T		*0.108*	*0.000*	*0.000*	*0.000*	No	
4	5′-FR (-317)	del T		*0.400*	*0.317*	*0.117*	*0.208*	Yes	rs36211087
5	5′-FR (-316)	C to A		*0.400*	*0.317*	*0.117*	*0.208*	Yes	rs36211088
6	5′-FR (-311)	C to T		*0.408*	*0.317*	*0.117*	*0.208*	Yes	rs36211089
7	5′-FR (-219)	C to G		*0.050*	*0.017*	*0.017*	*0.342*	Yes	rs8191438
8	5′-FR (-192)	G to A		*0.000*	*0.000*	*0.008*	*0.000*	No	
9	5′-FR (-138)	C to T		*0.000*	*0.000*	*0.008*	*0.000*	No	
10	5′-FR (-103)	C to A		*0.000*	*0.000*	*0.008*	*0.000*	No	
11	5′-UTR (-20)	C to T		*0.008*	*0.000*	*0.000*	*0.000*	No	
12	5′-UTR (-18)	G to A		*0.025*	*0.017*	*0.017*	*0.342*	Yes	rs8191439
13	5′-UTR (-1)	C to T		*0.000*	*0.000*	*0.008*	*0.000*	No	
14	IVS1 (21)	T to G		*0.008*	*0.000*	*0.000*	*0.000*	No	
15	IVS1 (-28)	C to T		*0.000*	*0.008*	*0.000*	*0.000*	No	
16	IVS1 (-20)	C to G		*0.175*	*0.583*	*0.758*	*0.342*	Yes	rs4147581
17	IVS2 (34)	G to A		*0.000*	*0.042*	*0.000*	*0.000*	Yes	rs8191445
18	IVS2 (103)	T to C		*0.008*	*0.000*	*0.000*	*0.000*	No	
19	IVS2 (131)	G to C		*0.008*	*0.000*	*0.000*	*0.000*	No	
20	IVS2 (134)	C to T		*0.212*	*0.317*	*0.121*	*0.208*	Yes	rs2370143
21	IVS2 (-108)	G to C		*0.147*	*0.000*	*0.000*	*0.000*	Yes	rs8191446
22	Exon 3 (95)	A to T	**Glu32Val**	*0.000*	*0.000*	*0.000*	*0.008*	No	
23	Exon 4 (150)	C to T		*0.150*	*0.000*	*0.000*	*0.008*	Yes	rs8191448
24	Exon 4 (172)	G to A	**Asp58Asn**	*0.008*	*0.000*	*0.000*	*0.000*	No	
25	IVS4 (13)	C to A		*0.407*	*0.325*	*0.117*	*0.208*	Yes	rs762803
26	Exon 5 (313)	A to G	**Ile105Val**	*0.425*	*0.317*	*0.108*	*0.533*	Yes	rs1695
27	IVS5 (28)	T to G		*0.000*	*0.000*	*0.000*	*0.008*	No	
28	IVS5 (112)	G to C		*0.008*	*0.000*	*0.000*	*0.000*	No	
29	Exon 6 (341)	C to T	**Ala114Val**	*0.017*	*0.083*	*0.000*	*0.008*	Yes	rs1138272
30	Exon 6 (372)	C to T		*0.000*	*0.008*	*0.000*	*0.000*	Yes	rs11553890
31	Exon 6 (439)	G to T	**Asp147Tyr**	*0.000*	*0.000*	*0.008*	*0.000*	Yes	rs4986949
32	IVS6 (-18)	G to A		*0.033*	*0.000*	*0.000*	*0.000*	No	
33	IVS6 (-16)	C to T		*0.250*	*0.300*	*0.108*	*0.183*	Yes	rs1871042
34	Exon 7 (555)	T to C		*0.458*	*0.300*	*0.117*	*0.542*	Yes	rs4891
35	Exon 7 (559)	C to T	**Arg187Trp**	*0.000*	*0.000*	*0.008*	*0.000*	No	

Table 2

GSTP1 haplotypes. White type on black highlights variant nucleotides as compared to 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 N- to the C-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. "d" = deletion. A previous set of designations based only on nonsynonymous SNPs, excluding all other polymorphisms ("Previous Designations"), are also listed (25,41).

Previous Designation	Allele	AA	CA	HCA	MA	5'-FR -383	5'-FR -361	5'-FR -360	5'-FR -317	5'-FR -316	5'-FR -311	5'-FR -219	5'-UTR -20	5'-UTR -18	IVS1 -20	IVS2 +34	IVS2 +134	IVS2 -108	Exon 3 +95	Exon 4 +150	Exon 4 +172	IVS4 +13	Exon 5 +313	IVS5 +112	Exon 6 +341	Exon 6 +439	IVS6 -16	Exon 7 +555	Exon 7 +559
*A	*1A	0.131	0.000	0.000	0.008	G	G	C	T	C	C	C	C	G	C	G	C	G	A	T	G	C	G	G	C	G	C	C
*A	*1B	0.108	0.153	0.092	0.133	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	G	G	C	G	T	C	C
*A	*1C	0.050	0.000	0.000	0.008	G	G	C	T	C	C	C	C	G	C	G	C	G	A	C	G	C	G	G	C	G	C	C	C
*A	*1D	0.033	0.000	0.000	0.008	T	G	C	d	A	T	C	C	G	C	G	C	G	A	C	G	A	G	G	C	G	T	C	C
*A	*1E	0.017	0.000	0.000	0.000	G	T	T	T	C	C	C	C	G	C	G	C	G	A	C	G	C	G	G	C	G	C	T	C
*A	*1F	0.017	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	C	G	A	C	G	A	G	G	C	G	C	T	C
*A	*1G	0.011	0.000	0.000	0.000	G	G	C	T	C	C	C	C	G	C	G	C	G	A	T	G	C	G	G	C	G	T	C	C
*A	*1H	0.008	0.000	0.000	0.000	G	G	C	T	C	C	C	C	G	C	G	C	G	A	T	G	C	G	G	C	G	C	T	C
*A	*1I	0.008	0.017	0.008	0.342	G	G	C	T	C	C	G	C	A	C	G	C	G	A	C	G	C	G	G	C	G	C	C	C
*A	*1J	0.008	0.008	0.000	0.000	T	T	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	G	G	C	G	T	C	C
*A	*1K	0.008	0.000	0.000	0.000	G	G	C	T	C	C	G	C	G	C	G	C	G	A	C	G	C	G	G	C	G	T	C	C
*A	*1L	0.008	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	G	C	C	G	C	C	C
*A	*1M	0.008	0.000	0.000	0.000	T	T	C	d	A	T	G	C	G	C	G	T	G	A	C	G	A	G	G	C	G	T	C	C
*A	*1N	0.008	0.000	0.000	0.000	T	G	C	d	A	T	C	T	G	C	G	C	G	A	C	G	A	G	G	C	G	C	T	C
*A	*1O	0.000	0.022	0.000	0.000	T	G	C	d	A	T	C	C	G	C	A	T	G	A	C	G	A	G	G	C	G	T	C	C
*B	*2A	0.117	0.555	0.708	0.325	G	G	C	T	C	C	C	C	G	G	G	C	G	A	C	G	C	A	G	C	G	C	T	C
*B	*2B	0.084	0.075	0.100	0.075	G	G	C	T	C	C	C	C	G	C	G	C	G	A	C	G	C	A	G	C	G	C	T	C
*B	*2C	0.072	0.000	0.000	0.000	G	T	T	T	C	C	C	C	G	C	G	C	G	A	C	G	C	A	G	C	G	C	T	C
*B	*2D	0.066	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	C	C	A	C	G	A	A	G	C	G	C	T	C
*B	*2E	0.033	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	G	G	C	C	A	C	G	A	A	G	C	G	C	T	C
*B	*2F	0.027	0.017	0.025	0.042	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	A	G	C	G	C	T	C
*B	*2G	0.024	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	A	G	C	G	T	C	C
*C	*3A	0.008	0.000	0.000	0.000	G	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	A	G	T	G	T	C	C
*C	*3B	0.008	0.000	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	T	C	A	C	G	A	A	G	T	G	C	T	C
*D	*4	0.000	0.083	0.000	0.008	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	G	A	G	G	T	G	T	C	C
*A	*5	0.000	0.008	0.000	0.000	T	G	C	d	A	T	C	C	G	C	G	T	G	A	C	A	A	G	G	C	G	T	C	C
*B	*6	0.000	0.000	0.008	0.000	G	G	C	T	C	C	C	C	G	G	G	C	G	A	C	G	C	A	G	C	T	C	T	C
*A	*7	0.000	0.000	0.000	0.008	T	G	C	d	A	T	C	C	G	C	G	T	G	T	C	G	A	G	G	C	G	T	C	C
*B	*8	0.000	0.000	0.008	0.000	G	G	C	T	C	C	C	C	G	G	G	C	G	A	C	G	C	A	G	C	G	C	T	T