

## Genetic Variation in Genes Associated with Arsenic Metabolism: Glutathione S-Transferase Omega 1-1 and Purine Nucleoside Phosphorylase Polymorphisms in European and Indigenous Americans\*

Lizhi Yu,<sup>1</sup> Kelly Kalla,<sup>1</sup> Erin Guthrie,<sup>1</sup> Amy Vidrine,<sup>1</sup> and Walter T. Klimecki<sup>1,2</sup>

<sup>1</sup>Arizona Respiratory Center, Tucson, Arizona, USA; <sup>2</sup>Southwest Environmental Sciences Center, Tucson, Arizona, USA

Individual variability in human arsenic metabolism has been reported frequently in the literature. This variability could be an underlying determinant of individual susceptibility to arsenic-induced disease in humans. Recent analysis revealing familial aggregation of arsenic metabolic profiles suggests that genetic factors could underlie interindividual variation in arsenic metabolism. We screened two genes responsible for arsenic metabolism, human purine nucleoside phosphorylase (*hNP*), which functions as an arsenate reductase converting arsenate to arsenite, and human glutathione S-transferase omega 1-1 (*hGSTO1-1*), which functions as a monomethylarsonic acid (MMA) reductase, converting MMA(V) to MMA(III), to develop a comprehensive catalog of commonly occurring genetic polymorphisms in these genes. This catalog was generated by DNA sequencing of 22 individuals of European ancestry (EA) and 24 individuals of indigenous American (IA) ancestry. In *hNP*, 48 polymorphic sites were observed, including 6 that occurred in exons, of which 1 was nonsynonymous (G51S). One intronic polymorphism occurred in a known enhancer region. In *hGSTO1-1*, 33 polymorphisms were observed. Six polymorphisms occurred in exons, of which 4 were nonsynonymous. In contrast to *hNP*, in which the IA group was more polymorphic than the EA group, in *hGSTO1-1* the EA group was more polymorphic than the IA group, which had only 1 polymorphism with a frequency > 10%. Populations representing genetic admixture between the EA and IA groups, such as Mexican Hispanics, could vary in the extent of polymorphism in these genes based upon the extent of admixture. These data provide a framework in which to conduct genetic association studies of these two genes in relevant populations, thereby allowing *hNP* and *hGSTO1-1* to be evaluated as potential susceptibility genes in human arsenicism. **Key words:** arsenic, biotransformation, European ancestry, indigenous American ancestry, *hGSTO1-1*, *hNP*, polymorphisms, SNP. *Environ Health Perspect* 111:1421–1427 (2003). doi:10.1289/txg.6420 available via <http://dx.doi.org/> [Online 21 July 2003]

Arsenic-induced human carcinogenesis remains one of the most perplexing mechanistic puzzles in contemporary toxicology. Although current animal models of arsenic-induced carcinogenesis are of equivocal biological relevance, convincing human epidemiologic data have identified the skin, lung, and bladder as targets of carcinogenesis caused by chronic exposure to arsenic (Bates et al. 1992). Several potential carcinogenic mechanisms have been proposed, including inhibition of DNA repair enzymes, tumor promoter-like induction of proliferation, alteration of DNA methylation, and clastogenesis (Kitchin 2001). Notwithstanding their lack of *in vivo* potency in carcinogenesis assays, arsenic compounds are potent toxicants both *in vitro* and *in vivo*. Biological effects of *in vitro* or *in vivo* exposure to arsenic include apoptosis, protein ubiquitination, cellular proliferation, oxidative stress, and enzyme inactivation (Asmuss et al. 2000; Chen et al. 2001; Hunter 2000; Kirkpatrick et al. 2003). Thus, arsenicals are clearly biologically active, but no definitive relationship has been established that links carcinogenesis to specific toxic endpoints.

Toxicologists have been able to assemble the key steps of the arsenic biotransformation pathway by identifying the chemical species of biotransformed arsenic. This pathway includes several oxidation state changes, successive oxidative methylations, and at least four metabolites (Vahter 2002). Notwithstanding the lack of specifically identified carcinogenic mechanisms of action, available evidence suggests that arsenic biotransformation is likely to be key to arsenic-induced disease. First, experimental evidence suggests that arsenic metabolism in humans could produce toxic intermediates and certainly governs the *in vivo* balance between chemical arsenic species with high biological potency and those with low biological potency. Historically, methylation of arsenic has been regarded as a detoxification mechanism (Abernathy et al. 1999). Within that historical context, detoxification typically referred to the lower cytotoxic potency of methylated arsenicals, namely monomethylarsonic acid of valence V [MMA(V)] and dimethylarsonic acid of valence V [DMA(V)], compared with inorganic arsenic species. In addition to the differential cytotoxicity

between inorganic arsenic and organic arsenic in the +V valence state, there is differential cytotoxicity between inorganic arsenic in the +III valence state and inorganic arsenic in the +V valence state. To complicate matters further, more recently, it has been shown that, as opposed to the low potency of MMA(V) and DMA(V), MMA(III) and DMA(III) are potent inducers of apoptosis, cytosolic protein binding, and epithelial hyperplasia (Thomas et al. 2001). In fact these methylated arsenic (III) compounds may be the most potent toxicants in the entire metabolic pathway. Although no direct link exists between arsenic-induced cytotoxicity and arsenic-induced carcinogenesis, the differential biological activity of arsenic metabolites is probably not restricted to cytotoxicity. Thus, understanding arsenic metabolism is likely to be key to a complete understanding of arsenic carcinogenesis.

A second, equally compelling reason to explore arsenic metabolism is the evidence that the metabolism of arsenic may be genetically influenced, with significant interindividual variation in arsenic metabolism among humans. Interspecies comparisons also suggest a genetic component to arsenic metabolism. Comparison of nine different strains of rats in which liver S9 fractions were exposed to inorganic arsenic revealed a 2-fold difference between the strain with the lowest rate of arsenic methylation versus the strain with the highest rate (Thomas et al. 2001). Parallel experiments corroborated the S9 fraction results with intact hepatocytes from the nine strains. In experiments examining the rate of formation of methylated arsenic in hepatocytes taken from different human

Address correspondence to W.T. Klimecki, Arizona Health Sciences Center, BRL-Room C112, 1609 N. Warren Ave., Tucson, AZ 85724 USA. Telephone: (520) 626-7470. Fax: (520) 626-5956. E-mail: walt@resp-sci.arizona.edu

\*The online version of this article (available at <http://www.ehponline.org>) contains Supplemental Material.

This work was supported by National Institute for Environmental Health Sciences grant ES06694. WTK was also supported by National Heart, Lung, and Blood Institute grants HL66801, HL66806, and HL67672.

The authors declare they have no conflict of interest. Received 21 April 2003; accepted 14 July 2003.

donors, nearly a 10-fold difference was observed (Stybło et al. 1999). Although this difference could be due to conditions surrounding donor organ condition or assay conditions, it is interesting to note that a comparison of several large studies measuring the distribution of urinary arsenic metabolites in humans exposed to inorganic arsenic via drinking water demonstrated a nearly 7-fold difference in the fraction of urinary arsenic as MMA (Vahter 2000). It is likely that some of this variability is environmental; however, it is possible that genetic variation may underlie some of the variation in arsenic metabolism. A role for genetic determinants of arsenic biotransformation is supported in recent work by Chung et al. (2002), in which familial aggregation in the pattern of urinary methylated arsenic metabolites was demonstrated in Chilean families.

A prerequisite in the evaluation of potential genetic determinants of arsenic metabolism is the development of a comprehensive catalog of genetic variation in genes whose products are involved in arsenic biotransformation. With such a catalog, one can test individual or combinations of polymorphic sites for association with phenotypes related to arsenic biotransformation, such as patterns of urinary metabolites. This study represents the first systematic polymorphism screening of two genes involved in arsenic biotransformation, human purine nucleoside phosphorylase (*bNP*) and human glutathione *S*-transferase omega 1-1 (*bGSTO1-1*). The products of these genes function as arsenate reductase and MMA, As(V) reductase, respectively (Radabaugh et al. 2002; Zakharyan et al. 2001). Because a major focus of our laboratory is the local arsenic-affected population [European ancestry (EA) individuals and Mexican ancestry individuals], we developed this catalog using two populations of subjects, EA subjects and indigenous American ancestry (IA) subjects. Because Mexican ancestry individuals represent predominantly an admixture between Europeans and indigenous Americans, these two populations make up the primary genetic background of the locally relevant populations (Cerdeira-Flores et al. 2002)

## Materials and Methods

### Subjects

Anonymous DNA samples from healthy individuals of self-reported ancestry were obtained from the Coriell Institute (Camden, NJ). Twenty-two samples from individuals of EA and 24 samples from individuals of IA were studied. EA individuals were selected from unrelated Centre

d'Etude du Polymorphisme Humain samples. The geographical origin of the IA samples consisted of 5 samples from Peru, 9 samples from Mexico, 1 sample from Ecuador, and 9 samples from Brazil. One DNA sample isolated from chimpanzee was also included in the study. These samples are commercially available; individual sample identification and ordering information are shown in the Supplemental Material, Table 1.

### Polymerase Chain Reaction

Genomic sequences for *bNP* and *bGSTO1-1* were accessed from the UCSC Genome Browser (<http://www.genome.ucsc.edu>) November 2002 freeze. Polymerase chain reaction (PCR) amplicons were designed to completely traverse each gene, such that each amplicon was approximately 900 bp long, and consecutive amplicons overlapped each other by approximately 200 bp. PCR reactions contained 20 ng genomic DNA, 1 pmol of each primer, 0.2 U taq polymerase (platinum taq; Invitrogen, Carlsbad, CA) and 0.1 mM deoxynucleotide triphosphates in a total volume of 10  $\mu$ L. Specific reaction conditions, including primer sequences, are available from the authors.

### Direct Polymerase Chain Reaction Sequencing

PCR amplicons were prepared for cycle sequencing by diluting them with water, using a dilution range of 1:3–1:6, depending on the reaction yield determined by agarose gel electrophoresis. Cycle-sequencing reactions were assembled using 0.4  $\mu$ L cycle-sequencing premix (BigDye, version 3.0; Applied Biosystems, Foster, CA), 1 pmol sequencing primer, 1.8  $\mu$ L 5 $\times$  sequencing dilution buffer and 5  $\mu$ L PCR product in a final volume of 10  $\mu$ L. Cycle-sequencing reactions were purified using DNA-affinity magnetic beads (Agencourt Biosciences, Beverly, MA). Purified sequencing reactions were electrophoretically analyzed using a DNA Analyzer 3730 (Applied Biosystems).

### Polymorphism Identification and Analysis

Sequence chromatograms were processed for base calling and assembly using the phred, phrap, and Consed suite of software programs (Ewing et al. 1998; Gordon et al. 1998). Initial polymorphism tagging was performed using Polyphred, with a minimum sequence quality of phred 25 (Nickerson et al. 1997). Potential polymorphic sites initially identified by Polyphred were individually confirmed by visual inspection of sequence traces. A criterion of

this visual inspection-confirmation was that the polymorphism must be observed in multiple chromatograms from singleton polymorphisms (polymorphisms occurring in only one subject) or in multiple subjects. For these confirmed polymorphic sites, each genotype for each subject was also confirmed by visual inspection of chromatograms. Polymorphic sites and associated subject-identified genotypes were automatically output to a relational database for further analysis, which included the automated generation of ethnicity-specific genotype frequencies, allele frequencies, and goodness-of-fit tests for Hardy-Weinberg equilibrium. Haplotypes were inferred using a Gibbs-sampling algorithm as implemented in the Phase software program (Stephens et al. 2001). Because the accuracy of statistically inferred haplotype increases with increasing haplotype frequency, we used polymorphisms with a minimum frequency [minor allele frequency (MAF)] of 0.10 to define relatively common haplotypes (Tishkoff et al. 2000). Pairwise linkage disequilibrium (LD) was calculated as  $r^2$ , a measure of the product-moment correlation coefficient (Devlin and Risch 1995).

### Gene Context of Polymorphisms

Each gene was annotated graphically using the Artemis software program (Rutherford et al. 2000). Annotations included exon location, protein coding exon subset, reading frame, and polymorphism site. Coding region polymorphisms were evaluated for codon changes resulting from polymorphisms and the predicted effect on amino acid sequence.

## Results

### Glutathione *S*-Transferase Omega 1-1, MMA(V) Reductase

Based on mapping data from the November 2002 freeze of the Human Genome Browser, *bGSTO1-1* contains six exons and extends for 12,551 bp of chromosome 10, in banding region q25.1. This is a region of low overall meiotic recombination, with a sex-averaged rate of 0.6 centimorgan (cM)/megabase (Mb) (Kong et al. 2002). The summary of the results of polymorphism screening in *bGSTO1-1* is shown in Figure 1. Overall, 33 polymorphic sites were observed in the approximately 16,000 bases of genomic DNA that we sequenced, including the gene itself, 5' upstream, and 3' downstream genomic regions. The flanking sequences of all polymorphisms are given in Supplemental Material, Table 2, to allow unambiguous identification of polymorphisms. Complete

genotypes for every individual at every site are shown in Supplemental Material, Figure 1 (EA population), and Supplemental Material, Figure 2 (IA population). Six polymorphisms occurred in exons 4 and 6, 2 of which were in the 3' untranslated region. The 4 remaining polymorphisms were nonsynonymous, resulting in two nonconservative amino acid substitutions (A140D, E208K), one amino acid deletion (E155del), and one conservative amino acid substitution (A236V). An interesting feature of these data is the distribution of polymorphisms between the EA and IA groups. Overall, the IA group was minimally polymorphic. The 14 polymorphisms observed in the IA population had a mean heterozygosity of  $0.085 \pm 0.064$  (mean  $\pm$  SD). In contrast,

the mean heterozygosity of the 21 polymorphisms observed in the EA population was  $0.223 \pm 0.195$ . In large part the genetic polymorphisms of each group were exclusive to that group. We observed 19 polymorphisms exclusive to the EA group, 12 polymorphisms exclusive to the IA group, and only 2 shared polymorphisms (668 and 8283). LD among high-frequency (MAF > 0.10) polymorphisms was observed spanning over 11,000 bp in the EA group, with  $r^2$  values approaching 1.0 between sites 2609 and 12707. Of the 8 polymorphisms in the EA group with MAF > 0.10, 7 (890, 1285, 2609, 6398, 8238, 10629, 12707) were in substantial LD with each other, with  $r^2$  values between 0.60 and 1.0. LD in the IA group is difficult to evaluate because of the low

frequency of the polymorphisms. Complete pairwise LD plots for *hGSTO1-1* are shown in Supplemental Material, Figure 3 (EA group), and Supplemental Material, Figure 4 (IA group). The lack of diversity of the IA samples in this genomic region is also reflected in the haplotype analysis, shown in Tables 1 and 2. Considering only relatively common haplotypes, here arbitrarily defined as those composed of polymorphisms with an MAF > 0.10, only two haplotypes defined the entire IA population. These haplotypes, in turn, were defined by polymorphism 668, the only polymorphism that exceeded this MAF threshold. Thus, 84% of the chromosomes studied carried the major allele for polymorphism 668. The remaining 16% of the IA chromosomes carried the minor allele for polymorphism 668. Despite having 8 polymorphisms exceeding the 0.10 MAF threshold, the EA population was divided into only three primary haplotypes, which represented 93% of all EA chromosomes (Table 2). The most common EA haplotype, GGG + AACT, was also the common IA haplotype. The second most common EA haplotype, AAA – GCAC, was entirely absent from the IA population.

**Purine Nucleoside Phosphorylase, Arsenate Reductase**

*hNP* contains six exons and occupies 7,636 bp of chromosome 14 in band region q11.2. Substantially more meiotic recombination occurs in this region, compared with *hGSTO1-1*, with a sex-averaged rate of 3.7 cM/Mb (Kong et al. 2002). The summary of the results of polymorphism discovery in *hNP* is shown in Figure 2. The flanking sequences of all polymorphisms are given in Supplemental Material, Table 3, to allow unambiguous identification of polymorphisms. Complete genotypes for every individual at every site are shown in Supplemental Material, Figure 5 (EA population), and Supplemental Material, Figure 6 (IA population). Forty-eight polymorphisms were observed in the approximately 11,000 bases of DNA sequenced in this genomic region. Six polymorphisms occurred in exons, of which 1 (2544) occurred in the 5' untranslated RNA (UTR) and 1 (9987) in the 3' UTR. Of the remaining 4 polymorphisms, 3 (5483, 5594, 8254) were synonymous, and 1 (5574) resulted in a conservative amino acid substitution (G51S). In contrast to the

ID	ATG Offset	Gene context	Minor allele frequency					
			European ancestry			American ancestry		
			Freq %	0.25	0.50	Freq %	0.25	0.50
668	-1859	5' Genomic	45			16		
890	-1637	5' Genomic	33			0		
1218	-1309	5' Genomic	0			4		
1285	-1242	5' Genomic	36			0		
1724	-803	5' Genomic	2			0		
1817	-710	5' Genomic	0			4		
2075	-452	5' Genomic	0			2		
2207	-320	5' Genomic	9			0		
2307	-220	5' Genomic	2			0		
2609*	82	Intron 1	34			0		
6198	3671	Intron 2	0			2		
6398	3871	Intron 2	34			0		
7062	4535	Intron 2	0			4		
7688	5161	Intron 3	2			0		
7964	5437	Intron 3	0			2		
8238	5711	Intron 3	43			4		
8592	6065	Intron 3	2			0		
8963	6436	Intron 3	0			2		
9173	6646	Intron 3	9			0		
9963	7436	Intron 3	0			2		
9992	7465	Intron 3	0			4		
10031	7504	Intron 3	2			0		
10629	8102	Exon 4: A140D	34			0		
10674*	8147	Exon 4: E155del	5			0		
10703	8176	Intron 5	2			0		
10709	8182	Intron 5	7			0		
11227	8700	Intron 5	0			5		
12707	10180	Intron 5	33			0		
13230	10703	Intron 5	2			0		
14902	12375	Exon 6: E208K	5			0		
14987	12460	Exon 6: A236V	0			4		
15029	12502	Exon 6: 3' UTR	2			0		
15035	12508	Exon 6: 3' UTR	0			8		

**Figure 1.** Summary of frequency and gene context of polymorphisms discovered in *hGSTO1-1* in EA (Europe) and IA (America) subjects. "ID" column indicates the polymorphism identification number relative to the location in the consensus sequence, with the first base of the consensus numbered 1. \*Indicates insertion/deletion. "ATG offset" indicates the polymorphism location relative to the first base "A" of the ATG methionine initiation codon. "Freq %" is the MAF, graphically displayed in the column to the right. Nonsynonymous SNPs are denoted by amino acid substitutions, e.g., A140D.

**Table 1.** *hGSTO1-1* haplotypes: IA population.

Chromosomes (%)	Haplotype ID	668
84.1	1	G
15.9	2	A



**Table 2.** *hGSTO1-1* haplotypes: EA population.

Chromosomes (%)	Haplotype ID	668	890	1285	2609	6398	8238	10629	12707
47.7	3	G	G	G	+	A	A	C	T
34.1	4	A	A	A	-	G	C	A	C
11.4	5	A	G	G	+	A	A	C	T

ID	ATG Offset	Functional context	Minor allele frequency					
			European ancestry			American ancestry		
			Freq %	0.25	0.50	Freq %	0.25	0.50
208	-2432	5' Genomic	2			0		
680	-1960	5' Genomic	2			0		
1014	-1626	5' Genomic	34			48		
1088	-1552	5' Genomic	14			31		
1095	-1545	5' Genomic	16			12		
1671	-969	5' Genomic	7			0		
1748*	-892	5' Genomic	20			15		
2368	-272	5' Genomic	7			0		
2544	-96	Exon1: 5' UTR	2			0		
3207	567	Intron 1	18			15		
3253	613	Intron 1	0			2		
3679*	1039	Intron 1	0			19		
4008	1368	Intron 1	2			0		
4303	1663	Intron 1	20			26		
4557	1917	Intron 1	20			31		
4594	1954	Intron 1	2			0		
5328	2688	Intron 1	0			21		
5483	2843	Exon 2: Syn	14			15		
5574	2934	Exon 2: G51S	14			35		
5594	2954	Exon 2: Syn	14			35		
5703	3063	Intron 2	2			0		
5766	3126	Intron 2	2			0		
5802	3162	Intron 2	14			34		
6037	3397	Intron 2	2			21		
6272*	3632	Intron 2	16			35		
6386	3746	Intron 2	25			35		
6563	3923	Intron 2	2			8		
6586	3946	Intron 2	24			15		
6631	3991	Intron 2	0			2		
6688	4048	Intron 2	0			2		
6938	4298	Intron 2	2			0		
6986	4346	Intron 2	5			15		
7036	4396	Intron 2	21			15		
7424	4784	Intron 2	25			15		
7717	5077	Intron 2	23			15		
8254	5614	Exon 5: Syn	0			2		
8477	5837	Intron 5	23			38		
8506	5866	Intron 5	0			6		
8538	5898	Intron 5	0			6		
8845	6205	Intron 5	0			23		
8883	6243	Intron 5	0			23		
8912	6272	Intron 5	0			15		
9293	6653	Intron 5	7			0		
9400	6760	Intron 5	16			15		
9987	7347	Exon 6: 3' UTR	16			15		
10328	7688	3' Genomic	2			23		
10461	7821	3' Genomic	0			15		
10544	7904	3' Genomic	0			23		

**Figure 2.** Summary of frequency and functional context of polymorphisms discovered in *hNP* in EA and IA subjects. Syn, synonymous SNP. "ID" indicates the polymorphism identification number relative to the location in the consensus sequence, with the first base of the consensus numbered 1. \*Indicates insertion/deletion. "ATG offset" indicates the polymorphism location relative to the first base "A" of the ATG methionine initiation codon. "Freq %" is the MAF, graphically displayed in the column to the right. Nonsynonymous SNP are denoted by amino acid substitutions, e.g., G51S.

situation for *hGSTO1-1*, in the genomic region of *hNP*, the IA group was more polymorphic than was the EA group, with mean heterozygosity values of  $0.292 \pm 0.127$  for the IA group and  $0.200 \pm 0.132$  for the EA group. Unlike *hGSTO1-1*, where 94% of the polymorphisms were exclusive to one group or the other, only 51% of all *hNP* polymorphisms were exclusive to one group, with 11 polymorphisms exclusive to the EA group and 13 polymorphisms exclusive to the IA group. Twenty-four polymorphisms were shared by both groups. The contrast in overall variability in the IA group between *hGSTO1-1* and *hNP* was evident in the number of polymorphisms exceeding the 0.10 MAF threshold for haplotype analysis. In contrast to *hGSTO1-1* (in which the IA group had 1 such polymorphism, whereas the EA group had 8), in *hNP* the IA group had 30 polymorphisms exceeding this threshold, whereas the EA group had only 19. LD in both the IA and EA groups was also more complex in *hNP*; in each group the high-frequency polymorphisms occurred in no less than three clusters, with high LD within a cluster but low LD between clusters (data not shown). Complete pairwise LD plots for *hNP* are shown in Supplemental Material, Figure 7 (EA group) and Supplemental Material, Figure 8 (IA group). The results of haplotype analysis for *hNP* are shown in Tables 3 and 4. Haplotype analysis revealed a greater complexity in the IA population within the *hNP* genomic region than in *hGSTO1-1*. Of the 48 IA chromosomes, 91.7% comprised five haplotypes. Similarly, the EA population chromosomes could largely be described by five haplotypes, which accounted for 81.8% of all EA chromosomes studied.

### Discussion

Genetic determinants of interindividual variability in arsenic metabolism have been speculated upon frequently in the literature, even before specific arsenic-metabolizing genes were identified (Abernathy et al. 1999; Vahter 1999, 2000). This project focused on a thorough screening of the first two characterized, arsenic-metabolic genes, with the objective of producing a catalog of commonly occurring polymorphisms in individuals of EA, IA, and by inference, many individuals of Mexican ancestry.

The genetics and molecular biology of *hNP* have been studied extensively because of the causative role of some nonsynonymous mutations in *hNP* in severe combined immunodeficiency syndrome (SCID) (Markert 1991; Sato and Wakabayashi

1998). Given the severe phenotype of SCID, it is not difficult to imagine that the spectrum of genetic variations one would observe in the general population would be confined to a limited range of functional changes in the protein. In fact for the 92 chromosomes that we studied, we observed only one nonsynonymous change in the predicted amino acid sequence, a conservative change of glycine to serine at residue 51. The 5 remaining polymorphisms that occurred in exons were either synonymous or positioned in untranslated mRNA regions. Although there is clear selection pressure against amino acid polymorphism in *hNP*, it is plausible that genetic polymorphism in gene regulatory regions could occur, as such changes could manifest themselves in a tissue-specific manner. Thus, in the presence of a polymorphism occurring in a gene regulatory region, a lethal target tissue such as T lymphocytes could be spared alterations in *hNP* gene expression, while a tissue critical to arsenic metabolism such as the liver could have altered transcriptional activity. Two polymorphisms, 3207 and 3253, are present within a region of intron 1 that has been defined as an enhancer element; this element may be absolutely necessary for gene transcription (Jonsson et al. 1992, 1994). No polymorphisms were observed in the minimal promoter elements that have been thus far defined for *hNP* (Jonsson et al. 1991).

*hGSTO1-1* is a member of an atypical class of glutathione transferases, with a biochemical activity more closely resembling glutaredoxins than other, more typical, glutathione transferase enzymes (Board et al. 2000). Its involvement in biochemical pathways in the human remains to be fully characterized, but in addition to its involvement in arsenic biotransformation, *hGSTO1-1* may include posttranslational

processing of interleukin-1 $\beta$  (Laliberte et al. 2003). It is strongly expressed in several tissues, including colon, heart, liver, ovary, pancreas, prostate, and spleen (Yin et al. 2001).

Our analysis of *hGSTO1-1* revealed a higher degree of predicted amino acid polymorphism than we observed in *hNP*. As opposed to *hNP*, in which only 17% of polymorphisms occurring in exons were nonsynonymous, 67% (4 of 6) of exon polymorphisms in *hGSTO1-1* were nonsynonymous. Of these 4 nonsynonymous polymorphisms, 3 were predicted to result in nonconservative amino acid substitutions. Only one of these nonconservative substitutions occurred at above 0.10 MAF, 10629 (A140D), which occurred at 0.34 MAF in the EA population only. Because the minimal regulatory region of *hGSTO1-1* has not yet been characterized, we cannot speculate as to which polymorphisms could be occurring in regulatory regions. Notwithstanding the amino acid diversity, perhaps the most notable features of the genetic evaluation of *hGSTO1-1* are the striking lack of polymorphism in the IA group and the low complexity of the EA group. One high-frequency polymorphism, 668, is common to both the IA group and the EA group. Because the IA group probably derives from Asia, the appearance of this polymorphism in both these geographically diverse groups suggests that this may be an ancient polymorphism. Aside from polymorphism 668, the IA group is monomorphic with respect to polymorphisms above a 0.10 MAF. The EA group has 7 additional polymorphisms above this MAF threshold, but they are all in a high degree of LD. This linked cluster of polymorphisms includes 1 nonsynonymous polymorphism, 10629. Thus, for the IA group, only 1 polymorphism, 668, is necessary to define all common haplotypes. In

the EA group, only 2 polymorphisms, 668 and 1 polymorphism from the cluster of 7 linked polymorphisms, are necessary to define all of the common haplotypes. Several factors may explain the difference in overall variation in *hGSTO1-1* between the IA and EA groups, including demographic events or, alternatively, selection pressure on this genomic region. Perhaps of greater significance is that for Mexican individuals that are some admixture between EA and IA, the extent to which an individual might be polymorphic for *hGSTO1-1* is related to the degree of European admixture in his/her ancestry. Thus, genetic association studies of *hGSTO1-1* in Mexican Hispanics should be carefully controlled for potential confounding by admixture.

To date, two studies have evaluated polymorphisms in *hGSTO1-1*. Tanaka-Kagawa et al. (2003), using polymorphism data from the Single Nucleotide Polymorphism database (dbSNP; <http://www.ncbi.nlm.nih.gov/SNP/>) studied 2 polymorphisms that were predicted to result in nonsynonymous changes, A140D and T217N. Substrate-dependent alterations in the resultant polymorphic proteins were observed. No validation of the presence or frequency of these 2 polymorphisms in human subjects was performed. Recently, Whitbread et al (2003), reported functional characterization of *hGSTO1-1*, together with polymorphism screening and validation in human subjects of three ancestries, African, Australian (EA), and Chinese. The authors based their polymorphism screening on existing polymorphisms in several databases. Through successive tiers of validation, only 1 of 26 *hGSTO1-1* polymorphisms originally found in the databases was observed to occur in any of the three ethnic groups. This polymorphism resulted in a predicted

**Table 3.** *hNP* haplotypes: IA population.

Chromosomes (%)	Haplotype ID	1014	1088	1095	1748	3207	3679	4557	5328	5483	5574	5594	5802	6037	6272	6386	6586	6688	6986	7036	7424	7717	8477	8845	8883	8912	9400	9987	10328	10461	10544
33.3	1	A	A	A	-	A	-	G	T	C	G	C	A	A	-	T	A	A	G	T	T	A	T	C	G	C	G	C	C	C	G
16.7	2	G	T	A	-	A	+	T	G	C	A	T	G	G	+	A	A	G	G	T	T	A	A	T	C	C	G	C	T	C	A
14.6	3	G	A	G	+	C	-	T	T	T	A	T	G	A	+	A	G	A	A	C	G	G	A	C	G	T	G	C	C	C	G
14.6	4	A	A	A	-	A	-	G	T	C	G	C	A	A	-	T	A	A	G	T	T	A	T	C	G	C	G	C	C	T	G
12.5	5	G	T	A	-	A	-	G	T	C	G	C	A	A	-	T	A	A	G	T	T	A	T	C	G	C	A	T	C	C	G

**Table 4.** *hNP* haplotypes: EA population.

Chromosomes (%)	Haplotype ID	1014	1088	1095	1748	3207	4557	5483	5574	5594	5802	6272	6386	6586	7036	7424	7717	8477	9400	9987
47.7	6	A	A	A	-	A	G	C	G	C	A	-	T	A	T	T	A	T	G	C
9.1	7	A	A	A	-	A	G	C	G	C	A	-	A	G	C	G	G	A	G	C
9.1	8	G	T	A	-	A	G	C	G	C	A	-	T	A	T	T	A	T	A	T
9.1	9	G	A	G	+	A	T	T	G	C	A	-	T	G	C	G	G	A	G	C
6.8	10	G	T	A	-	C	G	C	A	T	G	+	A	A	T	T	A	T	A	T

nonsynonymous amino acid substitution, A140D. In the course of the study, the authors also discovered an insertion/deletion polymorphism that caused a deletion of residue E155 (E155del). Finally, the authors generated expression constructs for the three observed haplotypes of these two polymorphisms and assayed the resultant *in vitro* expressed protein function in enzymatic assays. Their functional analysis demonstrated altered thiol transferase and glutathione conjugation for the A140, E155del haplotype. Our study also identified A140D (10629) and E155del (10674), and we discovered 1 additional nonsynonymous polymorphism in EA subjects, E208K (14902), and 1 in IA subjects, A236V (14987). The Whitbread study (Whitbread et al. 2003) illustrates two commonly encountered limitations of *in silico*-based polymorphism screening, namely, lack of validation of database polymorphisms and incomplete characterization of the gamut of commonly occurring polymorphisms in the population. In this regard, it is important to note that, of our 22 EA subjects, the same two subjects that were heterozygous for E155del, were also heterozygous for E208K. Thus, it is possible that these 2 polymorphisms are in substantial LD in EA subjects. As a result, it is possible that, by testing the 155 deletion, in the absence of the linked E208K substitution (an A140, E155del protein vs. an A140, E155del, E208K protein) the expression constructs in the Whitbread study do not accurately represent the commonly encountered hGSTO1-1 proteins, at least in their EA Australian group. This is particularly significant in light of the nature of the predicted E208K substitution, in which an acidic amino acid is substituted by a basic amino acid. Further testing is necessary to determine if these polymorphisms occur together in other ethnic groups.

Along with the EA and IA groups, we sequenced both *hNP* and *hGSTO1-1* in one chimpanzee DNA sample (data not shown). Chimpanzees, the closest non-extinct primate relative of humans, do not biotransform arsenic into methylated species but do demonstrate arsenate reductase activity (Vahter et al. 1995; Wildfang et al. 2001). Thus, it was not surprising that we observed a high degree of protein homology between humans and chimpanzees in *hNP* and *hGSTO1-1*, with only one amino acid difference between human and chimpanzee *hGSTO1-1*, a conservative substitution of isoleucine (human) for valine (chimp) at residue 26. Two amino acid differences were observed in *hNP*. At residue 51, a site where we observed a

glycine to serine nonsynonymous polymorphism in the human subjects, the chimp DNA sequence predicted a serine residue. At residue 277, all of our subjects were predicted to code for isoleucine, whereas the chimp coded for valine.

These data provide a functionally annotated catalog of commonly occurring genetic polymorphisms, in both protein coding and potential gene regulatory regions, in two important xenobiotic-metabolizing genes in populations that make up the genetic background of a substantial portion of Europe and the Americas. The combination of functional polymorphisms, clusters of linked polymorphisms, and haplotypes provide an essential framework from which to design genetic association studies to test key phenotypes related to arsenic biotransformation, such as urinary arsenic metabolic profile and susceptibility to arsenic-induced disease.

## REFERENCES

- Abernathy CO, Liu YP, Longfellow D, Aposhian HV, Beck B, Fowler B, et al. 1999. Arsenic: health effects, mechanisms of actions, and research issues. *Environ Health Perspect* 107(7):593–597.
- Asmuss M, Mullenders LH, Hartwig A. 2000. Interference by toxic metal compounds with isolated zinc finger DNA repair proteins. *Toxicol Lett* 112–113:227–231.
- Bates MN, Smith AH, Hopenhayn-Rich C. 1992. Arsenic ingestion and internal cancers: a review. *Am J Epidemiol* 135(5):462–476.
- Board PG, Coggan M, Chelvanayagam G, Eastal S, Jermin LS, Schulte GK, et al. 2000. Identification, characterization, and crystal structure of the Omega class glutathione transferases. *J Biol Chem* 275(32):24798–24806.
- Cerda-Flores RM, Villalobos-Torres MC, Barrera-Saldana HA, Cortes-Prieto LM, Barajas LO, Rivas F, et al. 2002. Genetic admixture in three Mexican Mestizo populations based on D1S80 and HLA-DQA1 loci. *Am J Hum Biol* 14(2):257–263.
- Chen F, Vallyathan V, Castranova V, Shi X. 2001. Cell apoptosis induced by carcinogenic metals. *Mol Cell Biochem* 222(1–2):183–188.
- Chung JS, Kalman DA, Moore LE, Kosnett MJ, Arroyo AP, Beeris M, et al. 2002. Family correlations of arsenic methylation patterns in children and parents exposed to high concentrations of arsenic in drinking water. *Environ Health Perspect* 110(7):729–733.
- Devlin B, Risch N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29(2):311–322.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8(3):175–185.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8(3):195–202.
- Hunter ES III. 2000. Role of oxidative damage in arsenic-induced teratogenesis. *Teratology* 62(4):240.
- Jonsson JJ, Converse A, Mclvor RS. 1994. An enhancer in the first intron of the human purine nucleoside phosphorylase-encoding gene. *Gene* 140(2):187–193.
- Jonsson JJ, Foresman MD, Wilson N, Mclvor RS. 1992. Intron requirement for expression of the human purine nucleoside phosphorylase gene. *Nucleic Acids Res* 20(12):3191–3198.
- Jonsson JJ, Williams SR, Mclvor RS. 1991. Sequence and functional characterization of the human purine nucleoside phosphorylase promoter. *Nucleic Acids Res* 19(18):5015–5020.
- Kirkpatrick DS, Dale KV, Catania JM, Gandolfi AJ. 2003. Low-level arsenite causes accumulation of ubiquitinated proteins in rabbit renal cortical slices and HEK293 cells. *Toxicol Appl Pharmacol* 186(2):101–109.
- Kitchin KT. 2001. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol* 172(3):249–261.
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardson B, et al. 2002. A high-resolution recombination map of the human genome. *Nat Genet* 31(3):241–247.
- Laliberte RE, Perregaux DG, Hoth LR, Rosner PJ, Jordan CK, Peese KM, et al. 2003. Glutathione S-transferase omega 1-1 is a target of cytokine release inhibitory drugs and may be responsible for their effect on interleukin-1beta posttranslational processing. *J Biol Chem* 278(19):16567–16578.
- Markert ML. 1991. Purine nucleoside phosphorylase deficiency. *Immunodef Rev* 3(1):45–81.
- Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25(14):2745–2751.
- Radabaugh TR, Sempayo-Reyes A, Zakharyan RA, Aposhian HV. 2002. Arsenate reductase II. Purine nucleoside phosphorylase in the presence of dihydrolipoic acid is a route for reduction of arsenate to arsenite in mammalian systems. *Chem Res Toxicol* 15(5):692–698.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944–945.
- Sato T, Wakabayashi Y. 1998. [PNP deficiency]. *Ryokibetsu Shokogun Shirizu* (21 pt 2):228–231.
- Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68(4):978–989.
- Styblo M, Del Razo LM, LeCluyse EL, Hamilton GA, Wang C, Cullen WR, et al. 1999. Metabolism of arsenic in primary cultures of human and rat hepatocytes. *Chem Res Toxicol* 12(7):560–565.
- Tanaka-Kagawa T, Jinno H, Hasegawa T, Makino Y, Seko Y, Hanioka N, et al. 2003. Functional characterization of two variant human GSTO 1-1s (Ala140Asp and Thr217Asn). *Biochem Biophys Res Commun* 301(2):516–520.
- Thomas DJ, Styblo M, Lin S. 2001. The cellular metabolism and systemic toxicity of arsenic. *Toxicol Appl Pharmacol* 176(2):127–144.
- Tishkoff SA, Pakstis AJ, Ruano G, Kidd KK. 2000. The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. *Am J Hum Genet* 67(2):518–522.
- Vahter M. 1999. Methylation of inorganic arsenic in different mammalian species and population groups. *Sci Prog* 82 (Pt 1):69–88.

- . 2000. Genetic polymorphism in the biotransformation of inorganic arsenic and its role in toxicity. *Toxicol Lett* 112–113:209–217.
- . 2002. Mechanisms of arsenic biotransformation. *Toxicology* 181–182:211–217.
- Vahter M, Couch R, Nermell B, Nilsson R. 1995. Lack of methylation of inorganic arsenic in the chimpanzee. *Toxicol Appl Pharmacol* 133(2):262–268.
- Whitbread AK, Tetlow N, Eyre HJ, Sutherland GR, Board PG. 2003. Characterization of the human Omega class glutathione transferase genes and associated polymorphisms. *Pharmacogenetics* 13(3):131–144.
- Wildfang E, Radabaugh TR, Vasken Aposhian H. 2001. Enzymatic methylation of arsenic compounds. IX. Liver arsenite methyltransferase and arsenate reductase activities in primates. *Toxicology* 168(3):213–221.
- Yin ZL, Dahlstrom JE, Le Couteur DG, Board PG. 2001. Immunohistochemistry of omega class glutathione S-transferase in human tissues. *J Histochem Cytochem* 49(8):983–987.
- Zakharyan RA, Sampayo-Reyes A, Healy SM, Tsaprailis G, Board PG, Liebler DC, et al. 2001. Human monomethylarsonic acid (MMA(V)) reductase is a member of the glutathione-S-transferase superfamily. *Chem Res Toxicol* 14(8):1051–1057.