Published in final edited form as:

J Appl Toxicol. 2010 April ; 30(3): 260–270. doi:10.1002/jat.1492.

Genetic association between intronic variants in AS3MT and arsenic methylation efficiency is focused on a large linkage disequilibrium cluster in chromosome 10

Paulina Gomez-Rubio¹, Maria M. Meza-Montenegro², Ernesto Cantu-Soto², and Walter T. Klimecki^{1,♦}

¹Department of Pharmacology and Toxicology, University of Arizona

²Department of Environmental Sciences, Instituto Tecnológico de Sonora

Abstract

Differences in arsenic metabolism are known to play a role in individual variability in arsenic-induced disease susceptibility. Genetic variants in genes relevant to arsenic metabolism are considered to be partially responsible for the variation in arsenic metabolism. Specifically, variants in arsenic (3+ oxidation state) methyltransferase (AS3MT), the key gene in the metabolism of arsenic, have been associated with increased arsenic methylation efficiency. Of particular interest is the fact that different studies have reported that several of the AS3MT single nucleotide polymorphisms (SNPs) are in strong linkage-disequilibrium (LD), which also extends to a nearby gene, CYP17A1. In an effort to characterize the extent of the region in LD, we genotyped 46 SNPs in a 347,000 base region of chromosome 10 that included AS3MT in arsenic-exposed subjects from Mexico. Pairwise LD analysis showed strong LD for these polymorphisms, represented by a mean r^2 of 0.82, spanning a region that includes 5 genes. Genetic association analysis with arsenic metabolism confirmed the previously observed association between AS3MT variants, including this large cluster of linked polymorphisms, and arsenic methylation efficiency. The existence of a large genomic region sharing strong LD with polymorphisms associated with arsenic metabolism presents a predicament because the observed phenotype cannot be unequivocally assigned to a single SNP or even a single gene. The results reported here should be carefully considered for future genomic association studies involving AS3MT and arsenic metabolism.

Keywords

Arsenic; AS3MT; Linkage Disequilibrium; polymorphism; SNP

INTRODUCTION

Human environmental exposure to inorganic arsenic is a well-recognized cause of significant morbidity and mortality affecting individuals worldwide. The conundrum of the cause of human disease resulting from environmental arsenic exposure is underscored by the multiplicity of affected organs and organ systems, as well as the dearth of experimental models that mimic arsenic-induced human disease (Waalkes et al. 2004; Rahman et al. 2009). The result of this is our relatively limited understanding of the mechanism by which arsenic induces a diverse set of diseases that include neuropathy, vascular disease, diabetes,

♦To whom correspondence should be addressed: Walt Klimecki, DVM, Ph.D. Dept. of Pharmacology and Toxicology University of Arizona PO 210207 Tucson, AZ 85721-0207 Tel. 520-626-7470 Fax. 520-626-2466 klimecki@pharmacy.arizona.edu .

and different types of cancers (Hafeman et al. 2005; States et al. 2009; Tseng et al. 2002; Morales et al. 2000).

Epidemiological studies, particularly in the absence of experimental models providing mechanistic explanations of arsenic-induced diseases, have played an important role in the characterization of factors that may place certain arsenic-exposed individuals at higher risk of toxicity. As might be expected, the strongest predictor of arsenic-associated disease in human population is the duration-weighted level of exposure to arsenic (Huang et al. 2008). Thus, the most dire public health challenges have occurred in regions where populations are exposed to sustained high levels of arsenic, such as Taiwan and Bangladesh (Singh et al. 2007; Nahar et al. 2008; Chiou et al. 1995). Studies of these populations, as well as others, measuring diverse disease phenotypes have identified different disease risk factors, including sex (males at higher risk), nutritional status (individuals with deficient diets at higher risk), and arsenic metabolism (individuals with lower efficiency in arsenic methylation at higher risk) (Chen et al. 2003; Steinmaus et al. 2005, 2006; Pu et al. 2007). However, these factors are not unrelated, as it has been shown that women and well-nourished individuals tend to have more efficient metabolism of arsenic, measured by indices such as percent urinary arsenic present as methylarsonic acid (%uMMA(V)), or the urinary ratio of dimethylarsinic acid to methylarsonic acid (uDMA/uMMA) (Heck et al. 2007; Hopenhayn-Rich et al. 1996). It is speculated that the explanation for the association observed between arsenic methylation and disease risk lies in the fact that, in the human metabolic scheme of arsenic, among the different arsenic metabolites MMA(V) is proximal to the primary toxic metabolite of arsenic (Lindberg et al. 2007). Consequently, the indices of arsenic metabolism may be surrogates for exposure to the causative metabolite, perhaps monomethylarsonous acid (MMA(III)), which is metabolically created in a two-electron reduction of MMA(V) and has been identified by different in vitro studies as the most toxic form of arsenic in biological systems (Styblo et al. 2000; Petrik et al. 2000; Eblin et al. 2007).

Given the routinely observed individual variability in arsenic methylation efficiency in exposed populations, genetic variation in the gene or genes responsible for arsenic metabolism is a potentially important determinant of this metabolic variability (Ghosh et al. 2008). While the exact metabolic pathways and the identities of the arsenic metabolic proteins are not universally agreed upon, it is likely that the key gene involved in human arsenic metabolism is AS3MT, located on chromosome 10 (Hayakawa et al. 2005; Thomas et al. 2007). Following an initial report in 2005, several reports have identified associations between genetic variants in AS3MT and arsenic metabolic variability in humans (Meza et al. 2005; Schläwicke Engström et al. 2007; Hernández et al. 2008). One genetic association in particular exists between a group of intronic AS3MT polymorphisms and arsenic methylation efficiency (Meza et al. 2005). In 2005, Meza et al., pointed out that, in Mexican subjects, three single nucleotide polymorphisms (SNPs) in AS3MT, dubbed 2393, 7388, and 30585, were in high degree of linkage disequilibrium (LD) and were all associated with high uDMA/uMMA ratio. In this context, linkage disequilibrium refers to the non-random co-inheritance of particular alleles at different polymorphic sites (Wall and Pritchard, 2003). Stated differently, among a group of subjects genotyped for two SNPs on the same chromosome that share high LD, subjects that are variant for one SNP are usually (or always) variant for the other SNP as well. In the study of Meza et al. (2005), subjects that were variant at the AS3MT 2393 SNP were almost always also variant for the 7388 and 30585 positions. The existence of clusters of high LD offers both benefits and pitfalls in genetic association studies. An advantage of LD clusters is that if one SNP in the cluster is positively associated with a phenotype under study, the entire cluster will reflect that association proportional to the degree of LD with the causative SNP, increasing the likelihood of discovering the genetic association in a screening study. In contrast, a

positively associated cluster of polymorphisms with a high degree of LD also presents a dilemma because the polymorphism that is actually creating the altered molecular biology underlying the phenotypic variation, the causative polymorphism, cannot be unambiguously identified through genetic association studies alone. In the case of AS3MT, the situation has become even more complex. Recently, Schläwicke Engström et al. (2008) have pointed out that the linked cluster of SNPs associated with arsenic methylation efficiency extends past AS3MT to a nearby gene, CYP17A1, a cytochrome P450 gene involved in the production of dehydroepiandrosterone, an androgenic hormone precursor (Lieberman et al. 2001). Since this finding clouds the unambiguous assignment of the genetic variation to AS3MT (individuals who were variant at AS3MT 7388 were also variant at a CYP17A1 polymorphic site), we were interested in characterizing more completely the extent of LD in this region of the genome to identify if yet more genes were contained in the cluster of polymorphic sites that have shown significant association with arsenic methylation efficiency. Identification of linked polymorphisms in other genes would further indicate that caution should be exercised in assigning the higher arsenic methylation efficiency seen in AS3MT variants to that gene.

MATERIALS AND METHODS

Study population

405 individuals were recruited from the area surrounding Ciudad Obregon, Sonora, Mexico. Individuals 6 years of age and older were included in the study. Subjects were not excluded based on current health status. No more than one individual from each household was sampled. Participants completed a general health status questionnaire, but were not medically screened for signs of arsenic-induced disease such as skin lesions and neuropathy.

Study subjects donated buccal cells and urine samples for DNA isolation and arsenic determination. Informed consent was obtained from all the participants, as approved by the Human Subjects Committee of the University of Arizona and of the Instituto Tecnológico de Sonora (ITSON). For subjects less than 18 years of age, parental consent was obtained in addition to specific assent from the minor subject.

DNA isolation

Buccal cells were collected from subjects using Omni Swabs (Whatman Scientific). Swab heads were ejected into tubes containing 750 µl of lysis buffer (50 mM Tris pH 8.0, 50 mM EDTA, 25 mM Sucrose, 100 mM NaCl, 1% SDS) and were stored at room temperature until extraction. Immediately prior to extraction the lysis buffer volumes (with the swab head) were normalized to 1 ml, 25 µl of proteinase K (20 mg/ml) was added to all samples, and they were incubated with agitation at 55°C overnight. DNA was isolated from the samples using a protocol modified from the Qiagen BioSprint 96 DNA Blood Kit as follows: The lysate was loaded into two identical 2.2 ml deep well plates. Each plate contained 300 µl lysate, 300 µl Isopropanol, and 300 µl Qiagen buffer AL. The first plate also contained 35 µl of Qiagen MagAttract Suspension G. The BioSprint 96 started in lysis plate 1 with a premix followed by a 4 minute binding step and 1 bead collection. The beads were then transferred into the second lysis plate with a premix followed by a 4 minute binding step and 1 bead collection. The beads were then transferred to a wash plate containing 800 µl of Qiagen buffer AW1 where they had a 4 minute wash step. The beads were then transferred to a second wash plate containing 500 µl of Qiagen buffer AW1 where they had a 3 minute wash step. The beads were then transferred to a third wash plate containing 500 µl of Qiagen buffer AW2 where they had a 2 minute wash step. The beads were then transferred to a fourth wash plate containing 500 µl of Qiagen buffer AW2 where they had a 2 minute wash step. The beads were then transferred to a fifth wash plate containing 500 µl of molecular

grade water. The beads were then transferred to an elution plate containing 250 μ l of Qiagen buffer AE (pre-incubated at 55°C) with a 10 minute elution step. The beads were then collected and transferred into first lysis plate and the above binding, wash, and elution steps were repeated for a second time.

Arsenic analysis

First morning urine samples were collected and stored at -20°C until shipment to the University of Arizona. Samples were shipped at -60°C and then stored at -80°C until the analysis was performed. Quantitative speciation of arsenic metabolites (As(III), As(V), MMA(V), and DMA(V)) as well as arsenobetaine in urine samples was performed using high performance liquid chromatography (Agilent 1100) with inductively coupled plasma mass spectroscopy (Agilent 7500) as described previously (Meza et al. 2005).

Selection of polymorphic sites

The genomic region surrounding AS3MT was evaluated using combined phase 1 and phase 2 project data from the international HapMap project and visualized on the genome browser feature of the Hapmap website (International HapMap Project, 2009). A general block of LD spanning approximately 347 KBases in this region of chromosome 10 was observed in both the Asian HapMap (Japanese in Tokyo, Japan, HapMap designation JPT; and Han Chinese in Beijing, designated CHB) and the European HapMap (CEPH, Utah residents with ancestry from northern and western Europe, HapMap designation CEU) populations. The limits of this block of LD were used as the limits of our search. We were interested in polymorphisms in LD with the three linked polymorphisms in AS3MT reported in 2005 to be associated with arsenic metabolism (Meza et al. 2005). Of those three polymorphisms, one (rs3740393, dubbed “7388” for its AS3MT position) is included in this HapMap data set. We screened all HapMap polymorphism sites within this 347Kbase region in European and Asian populations for polymorphisms that were at LD levels of r^2 greater than 0.5 with rs3740393. This resulted in a group of 53 candidate polymorphisms that were submitted to proprietary design software for the MassARRAY platform (Sequenom, San Diego, California) (Jurinke et al. 2002). Of the initial group, 49 SNPs were designed into panels to test in our arsenic-exposed Mexican subjects; the protocol and reaction conditions followed the manufacturer’s specifications (Oeth et al. 2005). The genotypes for two additional polymorphisms in AS3MT gene, rs12767543 and rs11191453, were obtained by allelic discrimination using the Taqman assay as previously described by Meza et al. (2005). Briefly, 15 ng of genomic DNA or water blank were manually assembled in a total reaction volume of 10 μ L; reaction mixture and thermal cycler specifications as described in Meza et al. (2005). Genotypes were assigned based on fluorescence using the allelic discrimination software (SDS, version 2.3; Applied Biosystems). We also tested one additional polymorphism that was not in LD with the other polymorphisms, namely rs11191439, a non-synonymous AS3MT polymorphism that has been reported to be associated with arsenic metabolic differences in humans (Hernández et al. 2008). The polymorphisms for which we are reporting data are listed in Table 1. Assay oligonucleotide sequence, reaction recipes and conditions for all polymorphisms are available from the corresponding author.

Quality control

All samples included in the analysis produced genotypes for at least 43 of the 47 polymorphisms. All polymorphisms included in the analysis produced genotypes in >90% of the samples. Polymorphisms were evaluated for deviation from the Hardy-Weinberg equilibrium using exact test statistics (Wigginton et al. 2005). Genotyping of all SNPs was performed twice in separate runs for 10% of the samples. 100% concordance (identical genotype assignments in both runs) was obtained from the comparison of both genotype sets.

Pairwise LD calculation

Pairwise LD, calculated as D' and r^2 , and LOD score (the log odds of the likelihood of the observed data from two SNPs occurring due to linkage compared to the likelihood of the data occurring under independent assortment) for each SNP pair, were calculated using Haploview, a publicly available software program for population-level genetic analysis (Barrett et al. 2005).

Genetic association between chromosome 10 variants and arsenic metabolism

The phenotype for arsenic metabolism was determined by calculating the ratio of urinary DMA concentration to MMA concentration (uDMA/uMMA), which was transformed by natural log conversion to approximate a normal distribution. A codominant genetic mode of inheritance was assumed for each polymorphic site and these sites were tested individually in a linear regression model in which uDMA/uMMA ratio was regressed onto allele dosage (0,1, 2, for major homozygous, heterozygous, and minor homozygous individuals, respectively). This was done for each polymorphic site using Plink, a publicly available software program designed for large-scale genetic association testing (Purcell et al. 2007).

RESULTS

Subject characteristics

Both males and females, with ages ranging between 6 to 86 years, were recruited for this study. While we did not exclude subjects with concurrent diseases such as diabetes and cardiovascular disease, the general health of the population based on self-reports was good, with the prevalence of diabetes at 7.4% and cardiovascular disease at 1%. Arsenic concentration in the drinking water source of this population ranged from undetectable to 227.6 ug As/L. Total urinary arsenic mean and range, after subtracting arsenobetaine, were 184.1 ug/L, 2.1 ug/L - 1053.5 ug/L with arsenic metabolite percentages of (mean, range): combined arsenite and arsenate (14.4, 1.7-92.6%), MMA(V) (11.6, 0.8-40.5%), and DMA(V) (73.9, 6.5-91.5%) (Table 2). Because our interest in this study is the characterization of LD across the AS3MT surrounding region in chromosome 10 it was necessary to include only individuals for whom nearly every site had been successfully genotyped. Thus, we chose to exclude individuals that had more than 4 missing genotypes. A total of 327 subjects of our study population met this criterion and were included in the analysis. A t-test analysis of mean urinary total arsenic and mean uDMA/uMMA ratios between included and excluded subjects did not demonstrate significant ($P < 0.05$) difference between groups, suggesting that no obvious bias was introduced through subject exclusion.

Linkage disequilibrium analysis—As an initial screen we examined LD patterns in the International HapMap Project data for the chromosome 10 genomic region around AS3MT, and identified a 347 KBase block of LD present in both Asian and European populations. Using this information as a reference, we then selected 49 SNPs for genotyping within this region that demonstrated LD levels of $r^2 > 0.5$ with rs3740393, an AS3MT SNP we previously found to be associated with arsenic metabolism. Two additional AS3MT polymorphisms, not included in the HapMap data set, that were previously reported in LD with rs3740393 (rs12767543 and rs11191453) were also included in the analysis (Meza et al. 2005). Of these 51 polymorphisms, 5 did not perform to quality control standards ($> 90\%$ genotyping rate) and were dropped, leaving 46 SNPs for the analysis of linked SNPs. Genotype distributions of the SNPs in our subjects were consistent with Hardy-Weinberg equilibrium (data not shown). LD for the polymorphisms in our population followed a pattern similar to published HapMap data. Pairwise analysis demonstrated that LD across this region averaged $r^2 = 0.78$ for the HapMap European population and $r^2 = 0.79$ for the

HapMap Asian population, as shown in figure 1. For our Mexican samples the variants are in strong LD, with a mean r^2 of 0.82 and a range of r^2 from 0.47 to 1.0 (Figure 2). The three AS3MT polymorphisms previously published as associated with arsenic metabolism were also in LD in this set of Mexican samples. Relative to polymorphism AS3MT rs3740393, pairwise LD with AS3MT rs12767543 was 0.68 and with AS3MT rs11191453 was 0.78.

Genetic association analysis—In addition to characterizing LD in this region, we were interested in evaluating whether the previously observed (Meza et al. 2005) genetic association between AS3MT rs12767543, rs3740393, and rs11191453 polymorphisms (2393, 7388 and 30585, respectively) and arsenic methylation efficiency was present in our current study population from Mexico. In addition, and to illustrate the confounding effect of LD across this region of chromosome 10, genetic association for each polymorphism and uDMA/uMMA ratio was carried out. Under the assumption of a codominant mode of inheritance, all linked polymorphisms tested in this region except one (rs11191580) were significantly ($P < 0.05$) associated with uDMA/uMMA ratio, consistent with previous genetic association studies of AS3MT (Meza et al. 2005; Schläwicke Engström et al. 2007). The general pattern of this genetic association is shown in figure 3 for AS3MT polymorphism rs10883795. Common homozygous individuals at this SNP had a mean uDMA/uMMA ratio of 6.7, while individuals heterozygous and homozygous for the variant alleles had uDMA/uMMA ratios of 7.9 and 8.4, respectively. Significance values for all polymorphisms tested in the genetic association study are shown in figure 4. Not surprisingly, in light of the strong LD within this group of polymorphisms, statistical significance is shared among the group of polymorphisms as well. Overall, the strength of the genetic association with uDMA/uMMA values appears to decrease with genomic distance from 5' to 3', relative to AS3MT. It is noteworthy, however, that polymorphisms with the highest levels of statistical significance were present in three distinct genomic structures: two genes, CYP17A1, AS3MT, and an open reading frame, C10orf32 that lies between those two genes.

Although it was not in significant LD with any other polymorphisms in this study, we also tested a non-synonymous AS3MT polymorphism, rs11191439, because it had been previously reported to be associated with human arsenic metabolism (Valenzuela et al. 2009). While our data demonstrated a trend in the same direction as has been reported previously with this polymorphism (Valenzuela et al. 2009), namely that variant individuals have lower uDMA/uMMA ratios, the association did not achieve statistical significance, perhaps in part due to the low frequency of variants for this polymorphism.

DISCUSSION

Epidemiological studies have shown that low arsenic methylation efficiency is associated with increased risk of arsenic-induced disease (Pu et al. 2007; Huang et al. 2007). This has generated interest in determining the factors that contribute to individual differences in arsenic metabolism. Two elements of our study have relevance for the scientific community interested in genetic determinants of variable arsenic metabolism in human populations. First, we have replicated the genetic association between a group of AS3MT polymorphisms and arsenic methylation efficiency as measured by the urinary ratio of DMA to MMA that we first reported in 2005 (Meza et al. 2005). The individuals carrying variant alleles in these AS3MT SNPs have an increased uDMA/uMMA ratio. While we reported this association was significant only for children in the 2005 study, in this study the association was significant for adults as well. It is important to consider that the number of subjects in the first study was relatively small, and variability in ancestry between age groups was also considered a confounding factor as the data suggested that the children had more indigenous American ancestry than adults. Subsequent to that report, a study by Schläwicke Engström

et al. (2007) reported a similar positive correlation between three intronic AS3MT polymorphisms that shared high LD (at rs3740393, rs3740390 and rs10748835) and enhanced arsenic methylation efficiency in a study of adult women in Argentina. Our study includes two of these polymorphisms (rs3740393 and rs3740390), and shows a similar positive association and pairwise LD in this population from Mexico.

The second key element in our study is the determination of the extent to which the AS3MT SNPs are in LD with other chromosome 10 polymorphisms, which proved to be larger than previously suggested, spanning about 347 KBases. In addition to AS3MT, there are 4 other genes present in this region: CYP17A1 (cytochrome P450, family 17, subfamily A, polypeptide 1) a gene that is involved in steroid metabolism, C10orf32 (chromosome 10 open reading frame 32), CNM2 (cyclin M2), involved in the transport of divalent metal cations, and NT5C2 (5'-nucleotidase, cytosolic II), a gene that plays a role in maintaining a constant composition of intracellular purine and pyrimidine nucleotides.

The region of chromosome 10 that includes AS3MT lies in an area of unusually high LD relative to the region's size. On average, LD in European ancestry populations extends for about 60 KBases, while African ancestry populations, owing to their relatively higher population age, exhibit LD that averages about 5 KBases (Goldstein and Weale, 2001). In this genomic region, the extent of LD in European, Asian and Mexican populations is nearly 6 fold higher than the European average. It should be noted that the 46 SNPs analyzed in this study are more likely a subset of the entire group of polymorphisms that form this cluster of LD since HapMap SNPs generally represent only a sampling of genetic variation in a given genomic region. It is also worth noting that we have not established the actual limits of this region of LD, although recently published international HapMap data for Mexican subjects suggest that we may be near the limits of this block of LD in Mexican subjects.

The finding of significant LD among SNPs that are associated with arsenic metabolism in this large region of 347 KBases points out that some caution is necessary in assigning the genetic association to a particular polymorphism or even to a particular gene. LOD score calculations for pairwise LD across this region support the significance of this level of LD. Historically a LOD score of 3 has been used as a threshold for significance (Gerber and Rodolphe, 1994). Our lowest r^2 value of 0.47 was associated with a LOD score of 36.6, suggesting that the probability of the LD we observed being a chance finding is exceedingly low. The practical consequence of this block of LD is that, since most of the subjects that are variant at one of the polymorphisms we studied are also variant at polymorphic sites across the entire region, prudence must be exercised in assuming that a genetic association between polymorphisms in AS3MT and arsenic metabolism is explained by a functional change in the molecular biology of AS3MT that is caused by polymorphisms in this gene. Due to the high LD shared between all the polymorphisms we tested, significant association was found between all of them and arsenic methylation efficiency, with only one exception in a NT5C2 polymorphism that showed borderline significance of $P=0.066$. Consequently, this information makes it difficult to determine which specific SNP or gene is causing the observed effect over uDMA/uMMA ratio. We are not suggesting that investigators should disregard the strength of the evidence pointing to the assumption that polymorphisms in AS3MT underlie a molecular biology effect that alters arsenic metabolism in humans. It is clear that AS3MT has the strongest weight of evidence for functional relevance to arsenic metabolism of any gene in the genome, as it has been proven to be a key enzyme involved in the methylation and reduction steps leading from arsenite (As(III)) to dimethylated arsenic (DMA) in the arsenic metabolic pathway (Thomas et al. 2004; Drobna et al. 2006). The functional relevance of AS3MT may also underlie the fact that statistical significance appears to be strongest near AS3MT, and decreases with increasing genomic distance 3' to

AS3MT. This, in turn relates to the overall lower LD observed between polymorphisms within the immediate region of AS3MT compared to polymorphism pairs outside the immediate AS3MT region (Figure 2), perhaps due to a locally higher recombination rate within AS3MT. Rather, we are suggesting that it is important not to exclude the possibility that all or part of the observed phenotype associated with variant status in this region may functionally involve other polymorphisms in other genes in this cluster of LD. While speculative, this is not an entirely unfounded idea. As observed in figure 4, the lowest *P*-values were obtained in the polymorphisms located in AS3MT as well as C10orf32 and CYP17A1 genes. CYP17A1 lies 5' to AS3MT, in a head-to-head configuration that is sometimes associated with the co-regulation of two proximal genes (Trinklein et al. 2004). Both AS3MT and CYP17A1 are expressed at their highest levels in the body within adrenal tissue, where CYP17A1 plays an important role in sex-steroid biosynthesis (Genomics Institute of the Novartis Research Foundation, BioGPS, 2009; Oskarsson et al. 2006). Furthermore, it has been proposed that sex steroids may play a role in arsenic metabolism. This idea comes from different epidemiological studies that have consistently shown that females have higher methylation efficiency than males (Hopenhayn-Rich et al. 1996; Tseng et al. 2005; Lindberg et al. 2007). In addition, there are reports that show that among women, pregnant women have more efficient arsenic methylation, and this effect increases during pregnancy (Concha et al. 1998; Hopenhayn et al. 2003). It has been speculated that these observations might be associated with differences in steroid production. Thus it is at least worth considering that functional polymorphisms that modulate the expression of other genes such as CYP17A1 could have a 'bystander' effect in modulating AS3MT expression levels by, for example, promoting a more transcription-conducive chromatin configuration.

Taken together with existing genetic association studies, the data that we report here establishes AS3MT as one of the most consistently replicated genes that is associated with arsenic metabolism in humans. In addition, the fact that a large number of polymorphisms in this genomic region share strong LD through at least 347 KBases makes this information relevant for future genetic association studies, and should be carefully considered when associating any of these specific genes or polymorphisms with phenotypes like arsenic methylation efficiency. The replication of this genetic association strongly suggests that there is an underlying molecular biological difference in individuals carrying variant alleles in this genomic region in chromosome 10 that manifests as increased arsenic methylation efficiency. Given the lack of genetic evidence for involvement of amino acid substitutions within variant individuals in this genetic effect, attention turns to the potential for genetic variants in this region to influence gene regulation, perhaps of AS3MT. Unfortunately, our current understanding that cis-acting elements can be located in almost any genomic structure, often at considerable distances from the cis-regulated gene, together with this report of a large block of co-occurring arsenic metabolism-associated polymorphisms that span a large genomic region, create a formidable challenge for researchers who aim to identify the causative polymorphism or polymorphisms that are responsible for this intriguing genetic association.

SHORT ABSTRACT

Several AS3MT single nucleotide polymorphisms (SNPs) that previously were associated with arsenic methylation efficiency have also shown strong linkage-disequilibrium (LD) in the genomic region including AS3MT. To characterize the extent of LD in this region, 46 SNPs were genotyped in chromosome 10. Strong LD was observed spanning a region that includes AS3MT and other 4 genes. Genetic association analysis confirmed the association between this large cluster of linked polymorphisms and arsenic methylation efficiency.

Acknowledgments

The authors wish to acknowledge Michael Kopplin for performing the arsenic speciation analyses. This work was supported by the NIEHS Superfund Basic Research Program (ES 04940) and by a NIEHS Center Grant (ES 006694).

References

- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21:263–265. [PubMed: 15297300]
- Chen YC, Guo YL, Su HJ, Hsueh YM, Smith TJ, Ryan LM, Lee MS, Chao SC, Lee JY, Christiani DC. Arsenic Methylation and Skin Cancer Risk in Southwestern Taiwan. *J Occup Environ Med*. 2003; 45:241–248. [PubMed: 12661181]
- Chiou HY, Hsueh YM, Liaw KF, Horng SF, Chiang MH, Pu YS, Lin JS, Huang CH, Chen CJ. Incidence of internal cancers and ingested inorganic arsenic: a seven-year follow-up study in Taiwan. *Cancer Res*. 1995; 55:1296–1300. [PubMed: 7882325]
- Concha G, Vogler G, Lezcano D, Nermell B, Vahter M. Exposure to inorganic arsenic metabolites during early human development. *Toxicol Sci*. 1998; 44:185–190. [PubMed: 9742656]
- Drobna Z, Xing W, Thomas DJ, Stýblo M. shRNA Silencing of AS3MT Expression Minimizes Arsenic Methylation Capacity of HepG2 Cells. *Chem Res Toxicol*. 2006; 19:894–898. [PubMed: 16841956]
- Eblin KE, Bredfeldt TG, Buffington S, Gandolfi AJ. Mitogenic Signal Transduction Caused by Monomethylarsonous Acid in Human Bladder Cells: Role in Arsenic-Induced Carcinogenesis. *Toxicol Sci*. 2007; 95:321–330. [PubMed: 17093206]
- Genomics Institute of the Novartis Research Foundation, BioGPS. 2009. <http://biogps.gnf.org/?referer=symatlas#goto=genereport&id=57412>
- Gerber S, Rodolphe F. Estimation and test for linkage between markers: a comparison of lod score and χ^2 test in a linkage study of maritime pine (*Pinus pinaster* Ait.). *Theor Appl Genet*. 1994; 88:293–297.
- Ghosh P, Banerjee M, Giri AK, Ray K. Toxicogenomics of arsenic: classical ideas and recent advances. *Mutat Res*. 2008; 659:293–301. [PubMed: 18638567]
- Goldstein DB, Weale ME. Population genomics: linkage disequilibrium holds the key. *Curr Biol*. 2001; 11:576–579.
- Hafeman DM, Ahsan H, Louis ED, Siddique AB, Slavkovich V, Cheng Z, van Geen A, Graziano JH. Association between arsenic exposure and a measure of subclinical sensory neuropathy in Bangladesh. *J Occup Environ Med*. 2005; 47:778–784. [PubMed: 16093927]
- Hayakawa T, Kobayashi Y, Cui X, Hirano S. A new metabolic pathway of arsenite: arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. *Arch Toxicol*. 2005; 79:183–191. [PubMed: 15526190]
- Heck JE, Gamble MV, Chen Y, Graziano JH, Slavkovich V, Parvez F, Baron JA, Howe GR, Ahsan H. Consumption of folate-related nutrients and metabolism of arsenic in Bangladesh. *Am J Clin Nutr*. 2007; 85:1367–1374. [PubMed: 17490975]
- Hernández A, Xamena N, Surrallés J, Sekaran C, Tokunaga H, Quinteros D, Creus A, Marcos R. Role of the Met(287)Thr polymorphism in the AS3MT gene on the metabolic arsenic profile. *Mutat Res*. 2008; 637:80–92. [PubMed: 17850829]
- Hopenhayn C, Huang B, Christian J, Peralta C, Ferreccio C, Atallah R, Kalman D. Profile of Urinary Arsenic Metabolites During Pregnancy. *Environ Health Perspect*. 2003; 111:1888–1891. [PubMed: 14644662]
- Hopenhayn-Rich C, Biggs ML, Smith AH, Kalman DA, Moore LE. Methylation study of a population environmentally exposed to arsenic in drinking water. *Environ Health Perspect*. 1996; 104:620–628. [PubMed: 8793350]
- Huang YK, Huang YL, Hsueh YM, Yang MH, Wu MM, Chen SY, Hsu LI, Chen CJ. Arsenic exposure, urinary arsenic speciation, and the incidence of urothelial carcinoma: a twelve-year follow-up study. *Cancer Causes Control*. 2008; 19:829–839. [PubMed: 18351295]

- Huang YK, Tseng CH, Huang YL, Yang MH, Chen CJ, Hsueh YM. Arsenic methylation capability and hypertension risk in subjects living in arseniasis-hyperendemic areas in southwestern Taiwan. *Toxicol Appl Pharmacol*. 2007; 218:135–142. [PubMed: 17173945]
- International HapMap Project. 2009. www.hapmap.org
- Jurinke C, van den Boom D, Cantor CR, Köster H. The use of MassARRAY technology for high throughput genotyping. *Adv Biochem Eng Biotechnol*. 2002; 77:57–74. [PubMed: 12227737]
- Lieberman S, Warne PA. 17-Hydroxylase: an evaluation of the present view of its catalytic role in steroidogenesis. *J Steroid Biochem Mol Biol*. 2001; 78:299–312. [PubMed: 11717000]
- Lindberg AL, Kumar R, Goessler W, Thirumaran R, Gurzau E, Koppova K, Rudnai P, Leonardi G, Fletcher T, Vahter M. Metabolism of low-dose inorganic arsenic in a central European population: influence of sex and genetic polymorphisms. *Environ Health Perspect*. 2007; 115:1081–1086. [PubMed: 17637926]
- Meza MM, Yu L, Rodriguez YY, Guild M, Thompson D, Gandolfi AJ, Klimecki WT. Developmentally restricted genetic determinants of human arsenic metabolism: association between urinary methylated arsenic and CYT19 polymorphisms in children. *Environ Health Perspect*. 2005; 113:775–781. [PubMed: 15929903]
- Morales KH, Ryan L, Kuo TL, Wu MM, Chen CJ. Risk of internal cancers from arsenic in drinking water. *Environ Health Perspect*. 2000; 108:655–661. [PubMed: 10903620]
- Nahar N, Hossain F, Hossain MD. Health and socioeconomic effects of groundwater arsenic contamination in rural Bangladesh: new evidence from field surveys. *J Environ Health*. 2008; 70:42–47. [PubMed: 18517153]
- Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, Jurinke C. iPLEX assay: Increased plexing efficiency and flexibility for MassArray system through single base primer extension with mass-modified terminators. SEQUENOM Application Note. 2005 <http://www.sequenom.com/getdoc/5a1593a3-08b5-436e-adbe-8d13154b9554/8876-006.aspx>.
- Oskarsson A, Ullerås E, Plant KE, Hinson JP, Goldfarb PS. Steroidogenic gene expression in H295R cells and the human adrenal gland: adrenotoxic effects of lindane in vitro. *J Appl Toxicol*. 2006; 26:484–492. [PubMed: 17080404]
- Petrick JS, Ayala-Fierro F, Cullen WR, Carter DE, Aposhian H Vasken. Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes. *Toxicol Appl Pharmacol*. 2000; 163:203–207. [PubMed: 10698679]
- Pu YS, Yang SM, Huang YK, Chung CJ, Huang SK, Chiu AW, Yang MH, Chen CJ, Hsueh YM. Urinary arsenic profile affects the risk of urothelial carcinoma even at low arsenic exposure. *Toxicol Appl Pharmacol*. 2007; 218:99–106. [PubMed: 17196235]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
- Rahman MM, Ng JC, Naidu R. Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. *Environ Geochem Health*. 2009; 31:189–200. [PubMed: 19190988]
- Schläwicke Engström K, Broberg K, Concha G, Nermell B, Warholm M, Vahter M. Genetic polymorphisms influencing arsenic metabolism: evidence from Argentina. *Environ Health Perspect*. 2007; 115:599–605. [PubMed: 17450230]
- Schläwicke Engström K, Nermell B, Concha G, Strömberg U, Vahter M, Broberg K. Arsenic metabolism is influenced by polymorphisms in genes involved in onecarbon metabolism and reduction reactions. *Mutat Res*. Jul 17.2008 [Epub ahead of print].
- Singh N, Kumar D, Sahu AP. Arsenic in the environment: effects on human health and possible prevention. *J Environ Biol*. 2007; 28:359–365. [PubMed: 17929751]
- States JC, Srivastava S, Chen Y, Barchowsky A. Arsenic and cardiovascular disease. *Toxicol Sci*. 2009; 107:312–323. [PubMed: 19015167]
- Steinmaus C, Bates MN, Yuan Y, Kalman D, Atallah R, Rey OA, Biggs ML, Hopenhayn C, Moore LE, Hoang BK, Smith AH. Arsenic methylation and bladder cancer risk in case-control studies in Argentina and the United States. *J Occup Environ Med*. 2006; 48:478–488. [PubMed: 16688004]

- Steinmaus C, Carrigan K, Kalman D, Atallah R, Yuan Y, Smith AH. Dietary intake and arsenic methylation in a U.S. population. *Environ Health Perspect.* 2005; 113:1153–1159. [PubMed: 16140620]
- Stybło M, Del Razo LM, Vega L, Germolec DR, LeCluyse EL, Hamilton GA, Reed W, Wang C, Cullen WR, Thomas DJ. Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells. *Arch Toxicol.* 2000; 74:289–299. [PubMed: 11005674]
- Thomas DJ, Li J, Waters SB, Xing W, Adair BM, Drobna Z, Devesa V, Stybło M. Arsenic (+3 oxidation state) methyltransferase and the methylation of arsenicals. *Exp Biol Med (Maywood).* 2007; 232:3–13. [PubMed: 17202581]
- Thomas DJ, Waters SB, Stybło M. Elucidating the pathway for arsenic methylation. *Toxicol Appl Pharmacol.* 2004; 198:319–326. [PubMed: 15276411]
- Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res.* 2004; 14:62–66. [PubMed: 14707170]
- Tseng CH, Huang YK, Huang YL, Chung CJ, Yang MH, Chen CJ, Hsueh YM. Arsenic exposure, urinary arsenic speciation, and peripheral vascular disease in blackfoot disease-hyperendemic villages in Taiwan. *Toxicol Appl Pharmacol.* 2005; 206:299–308. [PubMed: 16039941]
- Tseng CH, Tseng CP, Chiou HY, Hsueh YM, Chong CK, Chen CJ. Epidemiologic evidence of diabetogenic effect of arsenic. *Toxicol Lett.* 2002; 133:69–76. [PubMed: 12076511]
- Valenzuela OL, Drobna Z, Hernandez-Castellanos E, Sanchez-Pena LC, Garcia-Vargas GG, Borja-Aburto VH, Stybło M, Del Razo LM. Association of AS3MT polymorphisms and the risk of premalignant arsenic skin lesions. *Toxicol Appl Pharmacol.* 2009; 239:200–207. [PubMed: 19538983]
- Waalkes MP, Liu J, Ward JM, Diwan BA. Animal models for arsenic carcinogenesis: inorganic arsenic is a transplacental carcinogen in mice. *Toxicol Appl Pharmacol.* 2004; 198:377–384. [PubMed: 15276417]
- Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet.* 2003; 4:587–597. [PubMed: 12897771]
- Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. 2005; 76:887–893.

Figure 1. Linkage-disequilibrium profiles for the genotyped SNPs in HapMap a) Asian population, and b) European population. LD as r^2 for SNP pairs is shown inside the squares. Box gray-scale color indicates r^2 value varying from white ($r^2=0$) to black ($r^2=1$). A schematic illustration of the genomic structure of the genomic region studied is above each LD plot, arrowheads on genes indicate the gene orientation.

Figure 2. Linkage-disequilibrium profile for the genotyped SNPs in the Mexican population in this study. LD as r^2 for SNP pairs is shown inside the squares. Box gray-scale color indicates r^2 value varying from white ($r^2=0$) to black ($r^2=1$). A schematic illustration of the genomic structure of the genomic region studied is above the LD plot, arrowheads on genes indicate the gene orientation.

Figure 3. uDMA/uMMA ratio stratified by genotype for AS3MT polymorphism rs10883795. uDMA/uMMA ratio is shown on the Y-axis as the geometric mean, with error bars outlining the 95% confidence interval of the geometric mean values. Genotype groups are shown on the X-axis as homozygous non-variant, heterozygous and homozygous variant (TT, TC and CC respectively). $P = 0.00074$ (codominant inheritance, linear regression).

Figure 4.

Genetic association between the chromosome 10 genomic variants and arsenic metabolism. Negative log-converted P -values for statistical tests of association between genetic variation and uDMA/uMMA ratio are shown (Y-axis) for all the study polymorphisms (X-axis), as well as their location within the genomic region (Lines from bar-graphed polymorphisms leading to genomic map at top of figure). Horizontal line shows $P = 0.05$.

Table 1

Genes and polymorphisms genotyped.

rs No.	Gene	SNP type, Alleles	Chr. 10 position
rs17115100	CYP17A1	intronic G>T	104581383
rs1004467	CYP17A1	intronic T>C	104584497
rs3824755	CYP17A1	intronic G>C	104585839
rs11191416	intergenic	T>G	104594906
rs3824754	C10orf32	intronic G>A	104604340
rs4409766	C10orf32	intronic C>T	104606653
rs7098825	intergenic	T>C	104618224
rs12767543	AS3MT	intronic G>A	104621653
rs3740393	AS3MT	intronic C>G	104626645
rs3740390	AS3MT	intronic G>A	104628470
rs11191439	AS3MT	non-synonymous T>C	104628713
rs10883795	AS3MT	intronic T>C	104644567
rs11191453	AS3MT	intronic T>C	104649842
rs11191454	AS3MT	intronic A>G	104649994
rs1046778	AS3MT	3' UTR T>C	104651474
rs7897654	intergenic	C>T	104652448
rs17115213	CNNM2	intronic A>G	104671133
rs12411886	CNNM2	intronic C>A	104675289
rs10509759	CNNM2	intronic A>T	104679655
rs11191472	CNNM2	intronic A>T	104697006
rs12413409	CNNM2	intronic G>A	104709086
rs10883808	CNNM2	intronic A>T	104711116
rs11191479	CNNM2	intronic T>C	104713610
rs10883815	CNNM2	intronic T>C	104729169
rs11191499	CNNM2	intronic T>C	104754261
rs11191514	CNNM2	intronic C>T	104763354
rs11191515	CNNM2	intronic G>A	104766517
rs11191531	CNNM2	intronic G>C	104796888
rs12219027	CNNM2	intronic T>C	104801193
rs3781285	CNNM2	intronic G>C	104815655
rs1926032	CNNM2	intronic G>A	104819459
rs943037	CNNM2	synonymous G>A	104825909
rs11191548	intergenic	T>C	104836168
rs10786736	NT5C2	3' UTR G>C	104839106
rs12217501	NT5C2	intronic T>C	104841879
rs12412038	NT5C2	intronic G>A	104846152
rs11191555	NT5C2	intronic A>C	104847513
rs11191557	NT5C2	intronic C>G	104854604
rs11191558	NT5C2	intronic G>A	104854668

rs No.	Gene	SNP type, Alleles	Chr. 10 position
rs12413046	NT5C2	intronic A>G	104861194
rs10883832	NT5C2	intronic T>G	104861269
rs9633712	NT5C2	intronic G>C	104863751
rs732998	NT5C2	intronic A>G	104887891
rs12220375	NT5C2	intronic T>C	104891481
rs11191580	NT5C2	intronic T>C	104896201
rs11191582	NT5C2	intronic G>A	104903643
rs11191593	NT5C2	intronic T>C	104929205

Table 2

Population characteristics.

sex	n
female	203
male	124

	mean	range
Age	26	6 - 86
W - As *	72.1	0 - 227.6
U - As *	184.1	2.1 - 1053.53
uDMA/uMMA	7.36	0.79 - 52.04

W-As, total arsenic in water. U-As, total urinary arsenic. uDMA/uMMA, urinary DMA to urinary MMA ratio.

* Arsenic measured as parts per billion.