Arsenic metabolism is influenced by polymorphisms in genes involved in one-carbon metabolism and reduction reactions.

Engström, Karin; Nermell, Barbro; Concha, Gabriela; Strömberg, Ulf; Vahter, Marie; Broberg Palmgren, Karin

Published in:
Mutation Research

DOI:
10.1016/j.mrfmmm.2008.07.003

Published: 2009-01-01

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?
This is an author produced version of a paper published in Journal. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:


Access to the published version may require journal subscription. Published with permission from: Elsevier
Arsenic Metabolism is Influenced by Polymorphisms in Genes
Involved in One-carbon Metabolism and Reduction Reactions

Karin Schläwicke Engström1*, Barbro Nermell2, Gabriela Concha3, Ulf Strömberg1,
Marie Vahter2, Karin Broberg1

1) Department of Laboratory Medicine, Section of Occupational and Environmental
Medicine, Lund University, Lund, Sweden

2) Institute of Environmental Medicine, Section for Metals and Health, Karolinska
Institutet, Stockholm, Sweden

3) National Food Administration, Toxicology Division, Uppsala, Sweden

* Corresponding author
Address correspondence to K. Schläwicke Engström, Section of Occupational and
Environmental Medicine, Lund University Hospital, 22185 Lund, Sweden.
Telephone: +46 (0) 46 173148. Fax: +46 (0) 46 143702. E-mail:
Karin.engstrom@med.lu.se.
Abstract

Objectives: The susceptibility to arsenic (As) -induced diseases differs greatly between individuals, probably to a large extent due to genetic differences in arsenic metabolism. The aim for this study was to identify genetic variants affecting arsenic metabolism.

Methods: We evaluated the association between urinary metabolite pattern and polymorphisms in three gene-groups related to arsenic metabolism: 1) methyltransferases, 2) other genes involved in one-carbon metabolism and 3) genes involved in reduction reactions. Forty-nine polymorphisms were successfully genotyped in indigenous women (N=104) from northern Argentina, exposed to approximately 200 µg/L of arsenic in drinking water, with a unique metabolism with low percent monomethylated arsenic (%MMA) and a high percent dimethylated arsenic (%DMA).

Results: Genetic factors affecting arsenic metabolite pattern included two polymorphisms in arsenic (+III) methyltransferase (AS3MT) (rs3740400, rs7085104), where carriers had lower %MMA and higher %DMA. These single nucleotide polymorphism (SNPs) were in strong linkage disequilibrium (LD) with three intronic AS3MT SNPs, previously reported to be associated with arsenic metabolism, indicating the existence of a strongly methylating, population-specific haplotype. The CYP17A1 rs743572, 27 kilobasepairs (kbs) upstream of AS3MT, was in strong LD with the AS3MT SNPs and thus had similar effects on the metabolite profile. Smaller effects were also seen for one-carbon metabolism genes choline dehydrogenase (CHDH) (rs9001, rs7626693) and 5-methyltetrahydrofolate-homocysteine methyltransferase reductase (MTRR) (rs1801394) and genes involved in reduction
reactions, glutaredoxin ($GLRX$) (rs3822751) and peroxiredoxin 2 ($PRDX2$)
(rs10427027, rs12151144). Genotypes associated with more beneficial arsenic
metabolite profile (low %MMA and/or high %DMA in urine) were more common in
this population, which has been exposed to arsenic in drinking water for thousands of
years.

Conclusions: Polymorphisms in $AS3MT$ and in genes involved in one-carbon
metabolism and reduction reactions affects arsenic metabolism.

Keywords: Arsenic, $AS3MT$, metabolism, methylation, one-carbon metabolism,
polymorphisms
1. Introduction

Many millions of people around the world are exposed to high levels of arsenic (As) in drinking water. As is associated with skin, lung and bladder cancers [1,2], as well as vascular diseases, hepatotoxicity, neurotoxicity, diabetes, chronic cough and impaired fetal and child development [2-4]. The susceptibility to arsenic-induced health effects differs greatly between individuals, partly due to a large individual variability in arsenic metabolism, affecting retention and distribution of toxic metabolites. As is metabolized by reduction and methylation reactions via one-carbon metabolism. The most toxic metabolite is monomethylated As (MMA), while dimethylated As (DMA) has the lowest body retention [5]. The association between the fraction of MMA in urine, probably reflecting MMA\textsuperscript{III} in the tissues, and the risk of various health effects is well documented [6-8]. The inter-individual variation in As metabolism is partly due to environmental factors, but hereditary differences are very likely to contribute [9]. So far, knowledge of these genetic factors and their way of action is limited.

The main methyltransferase in As metabolism is As (+III) methyltransferase (AS3MT) (see Table 1 for accession numbers for all genes included in this study), which can catalyze all methylation steps when a reductant is present. We have previously found three intronic polymorphisms in AS3MT, the carriers of which had lower %MMA and higher %DMA [10]. The importance of AS3MT polymorphisms for the metabolism of arsenic has also been seen in other studies [11-14]. Functional
data for more common polymorphisms is scarce, apart from a few studies for the
p.M287T variant [14,15]. The methylation of As may also partly be due to an
AS3MT-independent process, since AS3MT-silenced hepatic cells can methylate As,
although less efficiently than cells with AS3MT expressed [16]. As-methylating
capacity of other methyltransferases may explain these results. Candidates could be
the DNA methyltransferases (DNMT1, DNMT3b), which methylate DNA at the
cytosine residues, as several parallel characteristics of As methylation and DNA
methylation have been reported [17]. As exposure may affect DNA methylation,
possibly due to that DNMTs, as well as AS3MT, are dependent of S-adenosyl-
methionine (SAM) [18,19], or hypothetically, because DNMTs may also methylate
As.

Many different enzymes take part in the metabolism of SAM, i.e. one-carbon
metabolism (Figure 1) and several display genetic variation. Known polymorphisms
affecting As metabolism have been found in 5,10-methylenetetrahydrofolate
reductase (MTHFR)[10,12,20] and 5-methyltetrahydrofolate-homocysteine
methyltransferase (MTR) [10] but several others may have similar effects. Little is
known about polymorphisms in genes involved in AsV reduction (Figure 2), which is
thiol-dependent [21]. Effects on metabolite pattern have been suggested for some
 glutathione (GSH) transferases (GSTs) GSTM1 [10,20,22,23], GSTO1 [24], GSTP1
[23] and GSTT1 [10,22,25]. The rate-limiting step in the synthesis of GSH is the
formation of γ-glutamyl-cysteine by glutathione cysteine ligase, encoded by two
genes (GCLC and GCLM). Gamma-glutamyl-transferase (GGT1) degrades GSH. A
coupled system involving glutaredoxin (Glrx)/GSH/glutathione reductase
(Gsr)/NADPH has As-reducing power in bacteria, yeast and recombinant rat [26, 27].
The coupled system thioredoxin (Txn)/thioredoxin reductase (TXNRD)/NADPH has
also As-reducing power in the recombinant rat studies for As3mt. Peroxiredoxin 2 (PRDX2) is capable of oxidizing TXN to TXNS2. Our aim was to elucidate how variations in the above-discussed genes affect the metabolism of As. Therefore, we genotyped and phenotyped inhabitants from San Antonio de los Cobres (SAC), Argentina, exposed to As in drinking water. This population shows an unusual As metabolism, with a low %MMA and a high %DMA [28]. Polymorphisms were selected in genes (including nearby gene regions) grouped into: 1) methyltransferases, 2) other genes involved in one-carbon metabolism and 3) genes involved in reduction reactions.

2. Materials and methods

2.1 Study areas and populations

Participants were women living in SAC, a village in northern Argentinean Andes, and had drinking water from the same source with about 200 µg As/L [10] with a fairly small variation over time [29]. Men were excluded from the studies, because they often were in other geographical locations for extended periods of time for work. Urine and blood samples were collected from a total of 111 women in 2004 and 2005. For individuals with enough DNA and successful genotype data (N=104), information (mean, range, number of individuals) was available for age (35, 15-76 years, N =104), weight (57, 40-88 kg, N =87), body mass index (BMI) (25, 17-38 kg/m², N =87), coca usage (46% users, N =104) and urinary As (U-As) (290, 94-720 µg/L, N =104). The metabolite values were (mean, range) (N =103): iAs (13, 1.0-48%), MMA (7.8, 1.2-18%) and DMA (79, 47-93%). The subjects were recruited via the local radio and hospital registers and interviewed about history of residence,
residences of parents and grandparents, parity and water consumption. Interviews revealed that individuals from SAC were mainly of indigenous (Atacameño) origin, with varying Hispanic origin. No individuals were first-degree relatives.

The Health Ministry of Salta, Argentina, and the Ethics Committee of the Karolinska Institutet, Sweden, have approved this study. Before sampling, the responsible community health worker and the interviewer (one of the coauthors) informed the women of all details of the research project, and informed consent was given in writing.

2.2 As analysis
Speciation of As metabolites (iAs, MMA and DMA) in urine for assessment of the metabolite pattern was performed using HPLC-HG-ICPMS (Agilent 1100 series system; Agilent 7500ce; Agilent Technologies, Japan and Germany) employing adequate quality control [10,30].

2.3 Genotyping
DNA was isolated from either whole blood or buccal cells. The isolation and extraction procedures have been described previously [10]. Twenty candidate genes were chosen, according to the literature involved in 1) methyltransferases, 2) other genes involved in one-carbon metabolism and 3) genes involved in reduction reactions. We have chosen the single nucleotide polymorphisms (SNPs) according to two different approaches.

Our first approach was to select SNPs that had a higher possibility to have a functional impact (according to dbSNP; website: http://www.ncbi.nlm.nih.gov/SNP/), like non-synonymous SNPs that may affect the protein structure/ enzyme activity or 5’
SNPs at putative promoter sites that may affect gene expression. However, a number of genes did not demonstrate these types of SNPs, or the SNPs had no known/very low allele frequencies for the CEU HapMap population (CEPH, Utah residents with ancestry from northern and western Europe; N=60, website: www.hapmap.org), which we used as reference, rendering it difficult to know if they might be very rare SNPs. In addition, the function of a given variant is not always easily predicted. Thus, as a second approach, we also used the indirect tagging approach, where no prior hypothesis is required in relation to the function of the SNP, and the objective instead was to capture as much genetic variation as possible. We then used HapMap data (with the CEU population as reference) in order to get potentially tagging SNPs in order to raise the possibility of finding a functional variant due to linkage disequilibrium (LD). TagSNPs were selected in Haplovie [31]. Thus, fifty-two SNPs were selected either by function (non-synonymous coding or putative promoter SNPs) or from HapMap data (website: www.hapmap.org), using the CEU population data (CEPH, Utah residents with ancestry from northern and western Europe; N=60). Genotyping for most SNPs was performed using Sequenom™ (San Diego, CA, USA) technology; performed by Swegene, Malmö University Hospital, Malmö, Sweden, while four SNPs were genotyped by Taqman allelic discrimination assay (GCLC, GCLM, GSTA1 and GSTP1) on an ABI 7000 instrument (Applied Biosystems, Foster City, CA, USA). The primers, probes and reaction conditions for the Taqman assays have been described earlier [32].

2.4 Statistical analyses

Deviations from Hardy-Weinberg equilibrium were tested using chi-square analysis.
Linkage disequilibrium (LD) analysis was performed using Haploview. Haplotypes were inferred by PHASE [33].

Associations between genotypes/haplotypes (independent variables) and urinary As metabolites (dependent variables) were analyzed with ANOVA and multivariate regression. The dependent variables used were %iAs, %MMA and %DMA (all natural log transformed due to consideration of normally distributed residuals).

ANOVA was performed with each SNP grouped into three genotypes (reference homozygotes, heterozygotes and variant homozygotes). ANOVA was then performed with genotypes combined into a dichotomous variable, based on the results from the analysis with three genotypes. Genotypes with effect estimates that had the same direction were combined. However, when the frequency of variant homozygotes was very low with just a few individuals, these individuals were pooled into the group with heterozygotes. For genotypes that did not meet the requirements of normal distribution and/or equal variances, Mann-Whitney/Kruskal-Wallis tests were performed.

In the multivariate regression analyses, all SNPs/haplotypes were employed as dichotomous variables. Total U-As (natural log transformed) was used as an exposure marker. Each potentially influential independent variable (U-As, weight, BMI, age and coca usage) was tested; an independent variable was included provided a p-value below 0.2. The final model was performed with an interaction term between genotype and U-As in order to account for multiplicative effect modification. If no significant interaction was present, a model without interaction term was employed in order to explore main effects.

To assess false positives, the false positive report probability (FPRP) [34] was calculated, based on observed association data, according to the formula; $\alpha \times (1-\pi)/(\alpha$ $+ \pi)$.
\[
\alpha \times (1-\pi) + (1-\beta) \times \pi.
\]
\(\alpha\) denotes the p-values from the multivariate regression analyses, \(1-\beta\) denotes the statistical power for the tests (based on obtained standard error for the observed p-value as significance threshold) and \(\pi\) denotes the prior probability of a true association of the tested genetic variant and outcome. The prior probabilities employed were 0.25 (high probability, for \(AS3MT\)) and 0.1 (moderate probability, for the other SNPs). Statistically significant SNPs with a FPRP above 0.5 were in this study not considered reliable to classify as true positives.

All statistical analyses were performed using SPSS (Version 15; SPSS, Chicago, IL, USA).

2.5 Bioinformatics

SNPs with a statistically significant association with arsenic metabolite pattern were further evaluated \textit{in silico} for potential function. PupaSNP (website: www.pupasnp.org) \cite{35,36} was employed in order to detect SNPs potentially affecting transcription factor binding sites, exonic splicing enhancers/silencers, triplices, splice sites and microRNA target sites. Emboss CpGPlot (website: http://www.ebi.ac.uk/emboss/cpgplot/) was employed to detect CpG-rich areas.

3. Results

3.1 Genotyping data

A list of all genotyped SNPs and their allele frequencies in SAC and the CEU population is shown in Table 1.

Of the 52 SNPs selected, assays for 48 SNPs in 16 genes were run with the Sequenom™ technology and 4 SNPs in 4 genes were analyzed by the Taqman allelic discrimination assay. Three of the Sequenom™ assays failed having too much
missing data. One individual was missing metabolite data, while two were missing
data for all genotypes, while the other individuals all had less than 10% missing data.
All SNPs were in Hardy-Weinberg equilibrium.

3.2 Metabolite pattern and genotype

The statistical analyses for %iAs were in several cases largely dependent on one
individual with an outlying metabolite value (1.0% iAs). The statistical analyses for
%DMA were affected by another outlier, with 47% DMA, compared with the
population mean of 79%. These extreme values were not excluded in the analyses,
but p-values for multivariate analyses without the outliers are presented within
brackets in Table 2 to facilitate the interpretation of the results.

3.2.1 Interaction: all groups

The multivariate model with an interaction term was evaluated first in order to
account for multiplicative effect modification. The model was as follows:

\[ \%\text{Metabolite} (\%i\text{As or } \%\text{MMA or } \%\text{DMA}) = \text{Intercept} + \beta_1 \times \text{Genotype} + \beta_2 \times \text{U-As} + \beta_3 \times \text{Age} + \beta_4 (\text{Genotype} \times \text{U-As}) \]

Few SNPs displayed significant interactions; only
the ones affecting %DMA (two SNPs) had a FPRP below 0.5 and could thus be
considered reliable to classify as true positives: steeper regression slopes with
increasing U-As were seen for \textit{DNMT1} rs8111085 T-allele carriers compared with
CC carriers (p-value = 0.003, $\beta = 0.026$; with the %DMA outlier removed the p-value
was 0.028 and $\beta = 0.018$) and \textit{GGT1} rs2236626 TT individuals compared with
individuals with at least one C-allele (p-value = 0.042, $\beta = 0.012$; with the %DMA
outlier removed the p-value was 0.048 and $\beta = 0.010$).
3.2.2 Main effects

3.2.2.1 Methyltransferases

If no significant interaction was present, a model without interaction term was employed in order to explore main effects. Results from this model are shown in Table 2. The model was as follows: %Metabolite (%iAs or %MMA or %DMA) = Intercept + β1×Genotype + β2×U-As + β3×Age. The polymorphisms in AS3MT (rs7085104, rs3740400) were in strong LD with each other ($R^2 = 0.97$) and also with the polymorphisms in our previous study [10](Fig. 3), resulting in a strong influence on %MMA and %DMA (Table 2, Figure 4). The effects were strongest in the ANOVA analyses, revealing a clear allele dose effect (geometric mean %MMA for different rs3740400 genotypes: AA: 13.6%, CA: 7.9% and CC: 6.5%). The rs3740400 is a putative triplex disrupting SNP and was also situated in a CpG island. Additionally, rs743572 located 27 kbs upstream of AS3MT, near the gene CYP17A1, also displayed a strong linkage to the two AS3MT polymorphisms ($R^2 = 0.97 - 1$, Figure 3) and showed a similar profile for %MMA and %DMA. Due to the strong LD, the haplotype analysis gave very similar results, even if the three SNPs from our earlier study were included.

Individuals with one or two variant alleles for DNMT1 rs16999593 (p.A97G) had significantly lower %DMA. However, this result was outlier-dependent and the variant allele frequency was low (0.07).

3.2.2.2 Genes involved in one-carbon metabolism
The rs1801394 (p.I22M) in MTRR affected %iAs, although this result was outlier-dependent. Individuals with at least one variant Met-allele had lower %iAs than individuals with the IleIle genotype (Table 2).

Two linked polymorphisms of CHDH (rs9001, rs7626693) affected both %MMA (Figure 5) and %DMA (Table 2). CC individuals for rs9001 had lower %MMA and higher %DMA. The rs9001 may act by disrupting exonic splicing enhancers (ESEs), and was also situated in a CpG island. CA individuals were in some cases (n=13) hard to distinguish from the AA individuals and pooled into CA+AA. Thus no ANOVA was performed for all three genotypes. However, for the CA and AA with clear genotype, individuals had similar metabolite pattern (data not shown).

3.2.2.3 Genes involved in reduction reactions

The GLRX intronic rs3822751 exerted the strongest effect in this group. Individuals with the variant CC genotype had significantly lower %MMA and higher %DMA (Figure 6) compared with individuals with at least one G-allele (GG+GC) (Table 2).

Two intronic SNPs in PRDX2 showed a tendency towards lower %iAs (Table 2, Figure 7), although the results were non-significant in the multivariate models. However, the p-values reached significance when the %iAs outlier was removed. These SNPs were in strong LD (R^2 = 0.94) and the heterozygotes had lower %iAs. No variant homozygotes were found.

A comparison of genotype frequencies for significant SNPs between SAC and the other HapMap populations can be seen in Figure 8.

The haplotype analyses did not reveal additional information about the metabolite pattern for any gene. We also performed dual gene combinations for the
two genes showing the strongest effect on As metabolite pattern, *AS3MT* and *GLRX*. Geometric mean %MMA for different combinations were (genotypes denoted as high or low depending on %MMA in the single SNP analyses): *AS3MT* AA+AG (high) and *GLRX* GG+GC (high): 8.9% (N = 35), *AS3MT* AA+AG (high) and *GLRX* CC (low): 6.9% (N = 9), *AS3MT* GG (low) and *GLRX* GG+GC (high): 6.9% (N = 42) and *AS3MT* GG (low) and *GLRX* CC (low): 5.5% (N = 16) (ANOVA, p=0.001).

4. Discussion

This study demonstrates that, so far, *AS3MT* genotype is the strongest effect marker for As metabolism, to a part explaining the unusually low %MMA in urine of Andean women exposed to As via drinking water. However, a number of other influential genetic markers with smaller effects were identified, and SNPs in the genes *MTRR* and *CHDH* (involved in one-carbon metabolism) as well as *GLRX* and *PRDX2* (involved in reduction reactions) are for the first time shown to possibly affect As metabolism. These genetic markers might have a functional impact on the As metabolism by themselves, or be in LD with other influential markers.

The SNPs in *AS3MT* and *CYP17A1* were in strong LD with the polymorphisms described in our earlier study [10] and thus part of a large haplotype block (at least 63 kbs) that had a marked effect, resulting in low %MMA and high %DMA. The SNPs were located upstream of the SNPs identified and described in our previous study. None of the HapMap populations had a haplotype block of a similar strength encompassing all SNPs. Due to the strong LD, it is not possible to say which of the SNPs, if any, that has a functional effect. However, we have some hypothetical candidates among the SNPs in this study: The functional impact of the rs3740400 SNP in *AS3MT* on As methylation is not yet known, but it may potentially act as a
putative triplex disrupting SNP. DNA triplexes are sequences larger than 10 polypurines or polypyrimidines, which inhibits mRNA synthesis, especially when the triplex is situated in a regulatory region [37,38]. SNPs located in the middle of those sequences can possibly affect to the triplex formation and thus potentially increase the mRNA synthesis. The rs3740400 SNP was situated in a CpG rich island in intron 1, about 200 basepairs from the transcription start site. The SNP is an A>C substitution, where the C is situated next to a G and thus an extra CpG is introduced with the SNP. The addition of methyl groups to cytosine in CpG nucleotides provides a way for differential regulation of gene expression. CpG regions tend to be associated with gene promoter regions and unmethylated CpGs allow gene expression [39]. The potential CpG site created by the variant allele and the location within a triplex suggest that this SNP is involved in regulation of AS3MT expression.

The rs743572 was situated 27 kbs upstream of AS3MT and may be indirectly associated with As metabolism profile due to LD. However, this SNP may also be of interest in itself, since it is located in the promoter of CYP17A1, a key enzyme in the steroidogenic pathway, which produces progestins, mineralocorticoids, glucocorticoids, androgens and estrogens. As methylation is more efficient in women of childbearing age than in men [7], indicating that sex steroids may play a role in As metabolism. Possibly, this is related to the fact that estrogen up-regulates one-carbon metabolism via phosphatidylethanolamine N-methyltransferase (PEMT) [17,40], catalyzing endogenous synthesis of choline, which via betaine can remethylate homocysteine to methionine in premenopausal women [41]. The variant allele of CYP17A1-34 T>C could create an additional Sp-1-type (CCACC box) promoter site and carriers of the variant allele may then have a higher gene expression [42,43]
resulting in an increased steroid production. However this has not been verified in vitro [44]. Individuals with one or two variant alleles for DNMT1 rs16999593 (His→Arg) had significantly lower %DMA. This result was outlier-dependent, and is thus questionable, and the MAF is low (0.07). However, this SNP has potential functional interest due to the replacement of the chemically unique histidine to arginine. Besides AS3MT and CYP17A1, GLRX showed the strongest effect on As metabolism. The variant CC genotype of the intronic SNP had significantly lower %MMA and higher %DMA compared with individuals with at least one G-allele. Possibly, GLRX may catalyze the reduction of disulfide bonds in As metabolism related proteins, thereby activating proteins and possibly influencing As metabolism. In studies with recombinant rat (rrat) As3mt, Waters et al. [27] showed that Glrx increased the methylation rate of As, but only when GSH was present. To our knowledge, there are no functional data for this SNP in GLRX. PRDX2 converts TXN to TXNS2 via H2O2. If TXN functions as a reducing agent in As metabolism, directly on As or via conversion of protein disulfides in As metabolizing genes, decreased function of PRDX2 would result in higher levels of TXN and thus a more efficient reduction of As. Indeed, individuals with a variant allele for any of the two intronic PRDX2 SNPs probably had a more efficient reduction of As, since they had a lower %iAs. The associations increased considerably in strength after removal of the %iAs outlier. The finding of influential genetic markers in the GLRX and TXN reduction complexes are supported by the role of these systems in animal models for As methylation [26,27]. Mass spectral evidence also suggests that the pentavalent arsenicals could be reduced by TXN [45].
The MTRR polymorphism p.I22M affected %iAs. However, this association lost in strength after removal of the %iAs outlier, and the significance of this result is thus somewhat unclear. MTRR regenerates a functional methionine synthase (MTR), which catalyzes the synthesis of methionine from homocysteine.

CHDH catalyzes the formation of betaine from choline, which is the only alternative to the folate-dependent conversion of homocysteine to methionine, particularly effective in women [41,51]. We discovered that two strongly linked CHDH SNPs (rs7626693, rs9001) affected %MMA. The strongest effect was seen for rs9001 (p.E40A). Individuals with the CC (AlaAla) genotype had lower %MMA and higher %DMA. The CC genotype was at all not present in the European HapMap population and rare in the other Hapmap populations. The effect of the rs9001 could be due to the amino acid substitution itself or by disruption of ESEs [52,53]. The C-allele has earlier been shown to have a protective effect on susceptibility to choline deficiency [54]. This is in line with our findings of a more efficient As methylation, possibly due to a better conversion of choline to betaine by CHDH.

The frequencies of the polymorphisms influencing As metabolism often deviated markedly from the frequencies in the European HapMap population, indicating a low degree of European ancestry. It also deviated from the other HapMap populations, in line with the unique arsenic metabolism in the indigenous Andean population [28]. Except for MTRR, the genotypes with the largest methylation capacity were more common in the SAC population than in most other populations. The population from SAC has been exposed to As for thousands of years [55,56] and one may speculate that this has resulted in a positive selection of genotypes which metabolize As in a less toxic way. Nevertheless, there is one study from the region that demonstrates genetic drift as an operating factor in populations in this area. The
endogamy factor was high due to isolated population and the small population size [57]. This might explain the large LD blocks, rather than selection for arsenic-beneficial genotypes.

We found few SNPs that modified the relationship between exposure and metabolite pattern, so-called interaction. This might be due to insufficient power for detection of interactions.

The absence of positive results for some genes may partly be due to low variant allele frequencies, where in many cases no variant homozygotes were found. Thus, in order to detect smaller effects, a larger population size is needed. For example, all DNMTs except for one (rs8111085, p.I311V) had a MAF<10%.

The fact that several genotypes have been analyzed on a number of outcome variables increases the possibility of false positive findings, and we have conducted FPRP analyses in order to assess the risk of false positives. A FPRP value of 0.5 is frequently used in FPRP analyses for smaller initial studies. According to Wacholder et al [34], large studies or pooled analyses that attempt to be more definitive evaluations of a hypothesis should use a more stringent FPRP value, perhaps below 0.2. If 0.2 is used, CHDH rs9001, the two PRDX2 SNPs and GGT1 rs2236626 would be classified as not reliable true positives. However, we have included the FPRP values so the reader can evaluate the results.

To conclude, two SNPs in AS3MT and one in CYP17A1 were in a strong LD with the polymorphisms reported in our earlier study [10] and thus had a very strong lowering effect on %MMA and increasing effect on %DMA. The strong LD in the indigenous SAC population stretches throughout the whole AS3MT gene and it also includes at least one SNP in CYP17A1, which gives an LD-block of at least 63 kbs. Other SNPs that affected As metabolite pattern were found in the one-carbon-
metabolizing genes \textit{MTRR} and \textit{CHDH}, as well as in the reducing reaction genes \textit{GLRX} and \textit{PRDX2}. The frequencies of the SNPs studied deviated in most cases from the other HapMap populations, and the genotype with the largest methylation capacity was more common in the SAC population.

Acknowledgements

This research was supported by grants from the Swedish Council for Working Life and Social Research, The EU within the Sixth Framework Programme for RTD (“PHIME” contract no FOOD-CT-2006-016253. The work reflects only the author’s views. The Community is not liable for any use that may be made of the information contained therein), The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, and the Karolinska Institutet. The authors declare they have no competing financial interest.

References


Smith Bronchiectasis in persons with skin lesions resulting from arsenic in 

drinking water, Epidemiology 16 (2005) 760-765.

M.A. Wahed, M. Yunus and L.Å. Persson Association of arsenic exposure 
during pregnancy with fetal loss and infant death: a cohort study in 


Chen, Y.H. Wang, Y.M. Hsueh, H.Y. Chiou and M.M. Wu Biomarkers of 
exposure, effect, and susceptibility of arsenic-induced health hazards in 

Persson and M. Vahter Gender and age differences in the metabolism of 
inorganic arsenic in a highly exposed population in Bangladesh, Environ Res 
106 (2008) 110-120.

diseases: current perspective, J Environ Sci Health C Environ Carcinog 

D.N. Mazumder, A.L. Hernandez and A.H. Smith Family correlations of 
arсенic methylation patterns in children and parents exposed to high 
concentrations of arsenic in drinking water, Environ Health Perspect 110 


Figure legends

Figure 1. Candidate genes in the group one-carbon metabolism genes. Genes are illustrated in ellipses. All genes in the picture were included in this study except for MTR. In this picture, the second methylation step is illustrated. However, the candidate genes are the same for the first methylation step (iAs$^{III}$ to MMA$^{V}$). Abbreviations: BHMT: betaine-homocysteine methyltransferase, CBS: cystathionine-beta-synthase, CHDH: choline dehydrogenase, MTHFR: 5,10-methylenetetrahydrofolate reductase, MTR: 5-methyltetrahydrofolate-homocysteine methyltransferase, MTRR: 5-methyltetrahydrofolate-homocysteine methyltransferase reductase, TCN2: transcobalamin II.

Figure 2. Candidate genes in the group reduction reactions. Genes are illustrated in ellipses. The reduction for iAs$^{III}$ is illustrated. The candidate genes are the same for the reduction of MMA$^{III}$ and DMA$^{III}$. The figure is divided into two parts, where the upper shows reduction of As$^{V}$ by GSH and the lower shows reduction via TXN. All genes in the picture were included in this study. Abbreviations: GCLC: glutathione cysteine ligase catalytic subunit, GCLM: glutathione cysteine ligase modifier subunit; GGT1: gamma-glutamyl-transferase 1, GLRX: glutaredoxin, GSR: glutathione reductase, GSTA1: glutathione S-Transferase A1, GSTP1: glutathione S-transferase P1, PRDX2: peroxiredoxin 2, TXV: thioredoxin, TXNRD1,2: thioredoxin reductase 1,2.

Figure 3. LD-values ($R^2$) for AS3MT SNPs. (A) depicts SAC while (B)-(E) depicts the different HapMap populations; (B) CEU: CEPH, Utah residents with ancestry from northern and western Europe, (C) CHB: Han Chinese in Beijing, China, (D) JPT: Japanese in Tokyo, Japan and (E) YRI; Yoruba in Ibadan, Nigeria. Data for rs3740400 is missing in the Hapmap
populations. Squares in black with no digits depicts full LD (R²=1). Rs3740393 (SNP12390),
r3740390 (SNP14215) and rs10748835 (SNP35991) were included in our previous study [10].

Figure 4. Scatterplot of %DMA for individuals with different AS3MT rs3740400 genotypes. The %DMA outlier is labeled in the figure.

Figure 5. Scatterplot of %MMA for individuals with different CHDH rs9001 genotypes.

Figure 6. Scatterplot of %DMA for individuals with different GLRX rs3822751 genotypes (genotype data is missing for the %DMA outlier).

Figure 7. Scatterplot of %iAs for individuals with different PRDX2 rs10427027 genotypes. The %iAs outlier is labeled in the figure.

Figure 8. Genotype frequencies in SAC compared with the HapMap populationsa for statistically significant SNPs. Efficient methylation genotypes (low %iAs, low %MMA and high %DMA) are depicted in black.

aFor PRXD2, reference populations are from Coriell Cell Repository data (NCBI SNP Database)
Table 1. Genes and polymorphisms genotyped.

<table>
<thead>
<tr>
<th>Group of genes</th>
<th>Gene</th>
<th>Accession no.</th>
<th>Rs nr</th>
<th>SNP type</th>
<th>SNP Position</th>
<th>Allele frequencies</th>
<th>Allele frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAC (%)</td>
<td>European population</td>
</tr>
<tr>
<td>Methyltransferases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS3MT</td>
<td>Hs.34492</td>
<td>rs7085104</td>
<td>5' near gene, g.A&gt;G</td>
<td>23377399, NT_030059</td>
<td>25/75</td>
<td>68/32</td>
</tr>
<tr>
<td></td>
<td>AS3MT</td>
<td>rs3740400</td>
<td></td>
<td>Intron, g.A&gt;C</td>
<td>23377991, NT_030059</td>
<td>25/75</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>CYP17A1</td>
<td>Hs.438016</td>
<td>rs743572</td>
<td>5'UTR, g.T&gt;C</td>
<td>23345678, NT_030059</td>
<td>25/75</td>
<td>40/60</td>
</tr>
<tr>
<td></td>
<td>DNMT1</td>
<td>Hs.202672</td>
<td>rs16999593</td>
<td>p.H97R, g.T&gt;C</td>
<td>1553983, NT_011295</td>
<td>93/7</td>
<td>100/0</td>
</tr>
<tr>
<td></td>
<td>DNMT1</td>
<td>rs8111085</td>
<td>p.I311V, g.T&gt;C</td>
<td>1536174, NT_011295</td>
<td>63/37</td>
<td>96/4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNMT1</td>
<td>rs7253062</td>
<td>Intron, g.G&gt;A</td>
<td>1557926, NT_011295</td>
<td>92/8</td>
<td>56/44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNMT1</td>
<td>rs11880388</td>
<td>Intron, g.G&gt;A</td>
<td>1516375, NT_011295</td>
<td>n/a</td>
<td>49/51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNMT3b</td>
<td>Hs.655708</td>
<td>rs2424913</td>
<td>Intron, g.C&gt;T</td>
<td>1570351, NT_028392</td>
<td>6/94</td>
<td>66/34</td>
</tr>
<tr>
<td></td>
<td>DNMT3b</td>
<td>rs6087990</td>
<td>5' near gene, g.T&gt;C</td>
<td>1546000, NT_028392</td>
<td>6/94</td>
<td>71/29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNMT3b</td>
<td>rs2424932</td>
<td>3'UTR, g.G&gt;A</td>
<td>1592628, NT_028392</td>
<td>6/94</td>
<td>52/48</td>
<td></td>
</tr>
<tr>
<td>One-carbon metabolism</td>
<td>BHMT</td>
<td>Hs.80756</td>
<td>rs585800</td>
<td>3' UTR, g.A&gt;T</td>
<td>29021566, NT_006713</td>
<td>97/3</td>
<td>69/31</td>
</tr>
<tr>
<td></td>
<td>BHMT</td>
<td>rs3733890</td>
<td>p.R239Q, g.G&gt;A</td>
<td>29016317, NT_006713</td>
<td>52/48</td>
<td>72/28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CBS</td>
<td>Hs.533013</td>
<td>rs234715</td>
<td>Intron, g.G&gt;T</td>
<td>1482836, NT_030188</td>
<td>99/1</td>
<td>78/22</td>
</tr>
<tr>
<td>Gene</td>
<td>rsID</td>
<td>Position</td>
<td>Reference Sequence</td>
<td>Genotype</td>
<td>Allele 1</td>
<td>Allele 2</td>
<td>MAF 1</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>----------</td>
<td>--------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>CBS</td>
<td>rs2124485</td>
<td>g.C&gt;T</td>
<td>30325934, NT_010966</td>
<td>99/1</td>
<td>100/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBS</td>
<td>rs234705</td>
<td>Intron, g.C&gt;T</td>
<td>1478213, NT_030188</td>
<td>98/2</td>
<td>66/34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBS</td>
<td>rs706209</td>
<td>3' UTR, g.C&gt;T</td>
<td>1467866, NT_030188</td>
<td>98/2</td>
<td>58/42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHDH</td>
<td>rs9001</td>
<td>p.E40A, g.A&gt;C</td>
<td>53797957, NT_022517</td>
<td>n/a</td>
<td>93/7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHDH</td>
<td>rs7626693</td>
<td>Intron, g.C&gt;T</td>
<td>53798581, NT_022517</td>
<td>72/28</td>
<td>45/55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHDH</td>
<td>rs9836592</td>
<td>Intron, g.T&gt;C</td>
<td>53795123, NT_022517</td>
<td>98/2</td>
<td>67/32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHDH</td>
<td>rs11718497</td>
<td>Intron, g.G&gt;C</td>
<td>53812057, NT_022517</td>
<td>n/a</td>
<td>64/36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTHFR</td>
<td>rs1537516</td>
<td>Intron/3' UTR, g.C&gt;T</td>
<td>6385228, NT_021937</td>
<td>97/3</td>
<td>90/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTHFR</td>
<td>rs1801131</td>
<td>p.E429A, g.A&gt;C</td>
<td>6391843, NT_021937</td>
<td>97/3</td>
<td>64/36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTHFR</td>
<td>rs1703796</td>
<td>Intron, g.C&gt;T</td>
<td>6399414, NT_021937</td>
<td>97/3</td>
<td>87/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTRR</td>
<td>rs10380</td>
<td>p.H595Y, g.C&gt;T</td>
<td>7887191, NT_006576</td>
<td>48/52</td>
<td>80/20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTRR</td>
<td>rs162036</td>
<td>p.K377R, g.A&gt;G</td>
<td>7875959, NT_006576</td>
<td>49/51</td>
<td>77/23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTRR</td>
<td>rs2287779</td>
<td>Synonymous, g.G&gt;A</td>
<td>7879216, NT_006576</td>
<td>75/25</td>
<td>96/4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTRR</td>
<td>rs1801394</td>
<td>122M, g.A&gt;G</td>
<td>7860973, NT_006576</td>
<td>83/17</td>
<td>55/45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCN2</td>
<td>rs2240433</td>
<td>5' near gene, g.A&gt;G</td>
<td>10393707, NT_011520</td>
<td>55/44</td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCN2</td>
<td>rs1801198</td>
<td>Arg259Pro, g.G&gt;C</td>
<td>10402179, NT_011520</td>
<td>25/75</td>
<td>45/55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCN2</td>
<td>rs5749134</td>
<td>Intron, g.C&gt;G</td>
<td>10400811, NT_011520</td>
<td>25/75</td>
<td>42/58</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Reduction reactions**

<table>
<thead>
<tr>
<th>Gene</th>
<th>rsID</th>
<th>Position</th>
<th>Reference Sequence</th>
<th>Genotype</th>
<th>Allele 1</th>
<th>Allele 2</th>
<th>MAF 1</th>
<th>MAF 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCLC</td>
<td>rs17883901</td>
<td>Promotor, g.C&gt;T</td>
<td>44268268, NT_007592</td>
<td>99/1</td>
<td>94/6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCLM</td>
<td>rs41303970</td>
<td>Promotor, g.C&gt;T</td>
<td>64347228, NT_029777</td>
<td>48/52</td>
<td>84/16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGT1</td>
<td>rs2236626</td>
<td>5' near gene, g.C&gt;T</td>
<td>4370016, NT_011520</td>
<td>32/68</td>
<td>21/79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Unigene Accession</td>
<td>SNP Accession</td>
<td>Effect</td>
<td>Reference Sequence</td>
<td>Minor Allele Frequency</td>
<td>Major Allele Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------------</td>
<td>---------------</td>
<td>--------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLRX</td>
<td>rs3822751</td>
<td>NT_023148</td>
<td>Intron, g.G&gt;C</td>
<td>3468082</td>
<td>51/49</td>
<td>30/70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLRX</td>
<td>rs871775</td>
<td>NT_023148</td>
<td>5' near gene, g.C&gt;T</td>
<td>3472640</td>
<td>99/1</td>
<td>91/9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLRX</td>
<td>rs4561</td>
<td>NT_023148</td>
<td>Synonymous, g.C&gt;T</td>
<td>3466185</td>
<td>7/93</td>
<td>41/59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSR</td>
<td>rs8190955</td>
<td>NT_007995</td>
<td>p.R153C, g.C&gt;T</td>
<td>886398</td>
<td>99/1</td>
<td>100/0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSR</td>
<td>rs2253409</td>
<td>NT_007995</td>
<td>Intron, g.G&gt;C</td>
<td>867740</td>
<td>86/14</td>
<td>74/26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSR</td>
<td>rs2978296</td>
<td>NT_007995</td>
<td>Intron, g.G&gt;C</td>
<td>895714</td>
<td>76/24</td>
<td>83/17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSTA1</td>
<td>rs3957356</td>
<td>NT_007592</td>
<td>Promotor, g.C&gt;T</td>
<td>43526901</td>
<td>7/93</td>
<td>61/39^k</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSTP1</td>
<td>rs1695</td>
<td>NT_033903</td>
<td>p.I105V, g.A&gt;G</td>
<td>12658484</td>
<td>27/73</td>
<td>61/39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRDX2</td>
<td>rs10427027</td>
<td>NT_011295</td>
<td>3' near gene, g.T&gt;C</td>
<td>12772285</td>
<td>88/12</td>
<td>95/5^t</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRDX2</td>
<td>rs12151144</td>
<td>NT_011295</td>
<td>Intron, g.A&gt;C</td>
<td>4175198</td>
<td>95/5^t</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXN</td>
<td>rs1049927</td>
<td>NT_008470</td>
<td>5'UTR, g.A&gt;G</td>
<td>20339960</td>
<td>99/1</td>
<td>92/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXN</td>
<td>rs2301242</td>
<td>NT_008470</td>
<td>5' near gene, g.T&gt;A</td>
<td>20340207</td>
<td>99/1</td>
<td>71/29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD1</td>
<td>rs11111979</td>
<td>NT_019546</td>
<td>5'UTR, g.C&gt;G</td>
<td>28162972</td>
<td>26/74</td>
<td>49/51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD1</td>
<td>rs6539137</td>
<td>NT_019546</td>
<td>Intron, g.T&gt;A</td>
<td>28189370</td>
<td>93/7</td>
<td>86/14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD2</td>
<td>rs5748647</td>
<td>NT_011519</td>
<td>Intron, g.C&gt;T</td>
<td>3073153</td>
<td>45/55</td>
<td>57/43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD2</td>
<td>rs1139793</td>
<td>NT_011519</td>
<td>p.L370T, g.A&gt;G</td>
<td>3020368</td>
<td>n/a</td>
<td>32/68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD2</td>
<td>rs5992495</td>
<td>NT_011519</td>
<td>p.S229R, g.T&gt;G</td>
<td>3035134</td>
<td>92/8</td>
<td>85/15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXNRD2</td>
<td>rs5748649</td>
<td>NT_011519</td>
<td>p.A66S, g.C&gt;A</td>
<td>3059249</td>
<td>91/9</td>
<td>65/35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


When applicable, amino acid position/gene region is denoted. Ancestral allele, according to NCBI SNP Database is denoted first when known.


Allele frequencies from HapMap CEU population (website: www.hapmap.org).

n/a = not available.

° denotes SNPs not included in the statistical analysis due to a failure rate of over 10%

No frequencies are presented due to difficulties to distinguish between CA and AA genotypes.

Allele frequencies from Coriell Cell Repository at NCBI SNP Database.

Allele frequencies from a Danish/Swiss population (Submitter handle CNPSCZ at NCBI SNP Database).

Allele frequencies from the Caucasian population in the NCI SNP500Cancer Database (website: http://snp500cancer.nci.nih.gov).