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 - Fundamental and Molecular Mechanisms of Mutagenesis

DOI:
10.1016/j.mrfmmm.2008.07.003

2009

Arsenic Metabolism is Influenced by Polymorphisms in Genes Involved in One-carbon Metabolism and Reduction Reactions

Karin Schläwicke Engström[1*], Barbro Nermell[2], Gabriela Concha[3], Ulf Strömberg[1], Marie Vahter[2], Karin Broberg[1]

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^{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 As^V 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 $TXNS_2$.

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/m^2, 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 Haploview [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; $\alpha \times (1-\pi)/(\alpha$

Formaterat: Engelska (USA)

× (1-π) + (1-β) × π). α Denotes the p-values from the multivariate regression analyses, 1-β denotes the statistical power for the tests (based on obtained standard error for the observed p-value as significance threshold) and π 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 *in silico* for potential function. PupaSNP (website: www.pupasnp.org) [35,36] was employed in order to detect SNPs potentially affecting transcription factor binding sites, exonic splicing enhancers/silencers, triplexes, 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: %Metabolite (%iAs or %MMA or %DMA) = Intercept + $\beta_1 \times$ Genotype + $\beta_2 \times$ U-As + $\beta_3 \times$ Age + β_4 (Genotype $\times$ 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 *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 *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 $\beta = 0.010$).

Formaterat: Engelska (USA)
Formaterat: Engelska (USA)
Formaterat: Engelska (USA)
Formaterat: Engelska (USA)

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 + β_1×Genotype + β_2×U-As + β_3×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

Formaterat: Engelska (USA)

Formaterat: Engelska (USA)

Formaterat: Engelska (USA)

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 $TXNS_2$ via H_2O_2. 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 *DNMT*s 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 *MTRR* and *CHDH,* as well as in the reducing reaction genes *GLRX* and *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

[1] IARC Some drinking-water disinfectants and Contaminants, including arsenic: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Lyon, International Agency for Research on Cancer; 84 (2004).

[2] NRC Arsenic in drinking water: 2001 update., Washington D.C., National Academy Press; (2001).

[3] D.N. Mazumder, C. Steinmaus, P. Bhattacharya, O.S. von Ehrenstein, N. Ghosh, M. Gotway, A. Sil, J.R. Balmes, R. Haque, M.M. Hira-Smith and A.H.

Smith Bronchiectasis in persons with skin lesions resulting from arsenic in drinking water, Epidemiology 16 (2005) 760-765.

[4] A. Rahman, M. Vahter, E.C. Ekström, M. Rahman, A.H. Golam Mustafa, 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 Bangladesh, Am J Epidemiol 165 (2007) 1389-1396.

[5] M. Vahter Mechanisms of arsenic biotransformation, Toxicology 181-182 (2002) 211-217.

[6] C.J. Chen, L.I. Hsu, C.H. Wang, W.L. Shih, Y.H. Hsu, M.P. Tseng, Y.C. Lin, W.L. Chou, C.Y. Chen, C.Y. Lee, L.H. Wang, Y.C. Cheng, C.L. Chen, S.Y. 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 Taiwan, Toxicol Appl Pharmacol 206 (2005) 198-206.

[7] A.L. Lindberg, E.C. Ekström, B. Nermell, M. Rahman, B. Lönnerdal, L.A. 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.

[8] C.H. Tseng Arsenic methylation, urinary arsenic metabolites and human diseases: current perspective, J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 25 (2007) 1-22.

[9] J.S. Chung, D.A. Kalman, L.E. Moore, M.J. Kosnett, A.P. Arroyo, M. Beeris, D.N. Mazumder, A.L. Hernandez and A.H. Smith Family correlations of arsenic methylation patterns in children and parents exposed to high concentrations of arsenic in drinking water, Environ Health Perspect 110 (2002) 729-733.

[10] K. Schlawicke Engstrom, K. Broberg, G. Concha, B. Nermell, M. Warholm and M. Vahter Genetic polymorphisms influencing arsenic metabolism: evidence from Argentina, Environ Health Perspect 115 (2007) 599-605.

[11] A. Hernandez, N. Xamena, J. Surralles, C. Sekaran, H. Tokunaga, D. Quinteros, A. Creus and R. Marcos Role of the Met(287)Thr polymorphism in the AS3MT gene on the metabolic arsenic profile, Mutat Res 637 (2008) 80-92.

[12] A.L. Lindberg, R. Kumar, W. Goessler, R. Thirumaran, E. Gurzau, K. Koppova, P. Rudnai, G. Leonardi, T. Fletcher and M. Vahter Metabolism of low-dose inorganic arsenic in a central European population: influence of sex and genetic polymorphisms, Environ Health Perspect 115 (2007) 1081-1086.

[13] M.M. Meza, L. Yu, Y.Y. Rodriguez, M. Guild, D. Thompson, A.J. Gandolfi and W.T. Klimecki Developmentally restricted genetic determinants of human arsenic metabolism: association between urinary methylated arsenic and CYT19 polymorphisms in children, Environ Health Perspect 113 (2005) 775-781.

[14] T.C. Wood, O.E. Salavagionne, B. Mukherjee, L. Wang, A.F. Klumpp, B.A. Thomae, B.W. Eckloff, D.J. Schaid, E.D. Wieben and R.M. Weinshilboum Human arsenic methyltransferase (AS3MT) pharmacogenetics: gene resequencing and functional genomics studies, J Biol Chem 281 (2006) 7364-7373.

[15] Z. Drobna, S.B. Waters, F.S. Walton, E.L. LeCluyse, D.J. Thomas and M. Styblo. Interindividual variation in the metabolism of arsenic in cultured primary human hepatocytes, Toxicol Appl Pharmacol 201 (2004) 166-177.

[16] Z. Drobna, W. Xing, D.J. Thomas and M. Styblo shRNA silencing of AS3MT expression minimizes arsenic methylation capacity of HepG2 cells, Chem Res Toxicol 19 (2006) 894-898.

[17] M.E. Vahter Interactions between arsenic-induced toxicity and nutrition in early life, J Nutr 137 (2007) 2798-2804.

[18] E. Marafante, M. Vahter and J. Envall The role of the methylation in the detoxication of arsenate in the rabbit, Chem Biol Interact 56 (1985) 225-238.

[19] J.L. Martin and F.M. McMillan SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold, Curr Opin Struct Biol 12 (2002) 783-793.

[20] C. Steinmaus, L.E. Moore, M. Shipp, D. Kalman, O.A. Rey, M.L. Biggs, C. Hopenhayn, M.N. Bates, S. Zheng, J.K. Wiencke and A.H. Smith Genetic polymorphisms in MTHFR 677 and 1298, GSTM1 and T1, and metabolism of arsenic, J Toxicol Environ Health A 70 (2007) 159-170.

[21] T.R. Radabaugh and H.V. Aposhian Enzymatic reduction of arsenic compounds in mammalian systems: reduction of arsenate to arsenite by human liver arsenate reductase, Chem Res Toxicol 13 (2000) 26-30.

[22] H.Y. Chiou, Y.M. Hsueh, L.L. Hsieh, L.I. Hsu, Y.H. Hsu, F.I. Hsieh, M.L. Wei, H.C. Chen, H.T. Yang, L.C. Leu, T.H. Chu, C. Chen-Wu, M.H. Yang and C.J. Chen Arsenic methylation capacity, body retention, and null genotypes of glutathione S-transferase M1 and T1 among current arsenic-exposed residents in Taiwan, Mutat Res 386 (1997) 197-207.

[23] R. Marcos, V. Martinez, A. Hernandez, A. Creus, C. Sekaran, H. Tokunaga and D. Quinteros Metabolic profile in workers occupationally exposed to

arsenic: role of GST polymorphisms, J Occup Environ Med 48 (2006) 334-341.

[24] L.L. Marnell, G.G. Garcia-Vargas, U.K. Chowdhury, R.A. Zakharyan, B. Walsh, M.D. Avram, M.J. Kopplin, M.E. Cebrian, E.K. Silbergeld and H.V. Aposhian Polymorphisms in the human monomethylarsonic acid (MMA V) reductase/hGSTO1 gene and changes in urinary arsenic profiles, Chem Res Toxicol 16 (2003) 1507-1513.

[25] K.M. McCarty, Y.C. Chen, Q. Quamruzzaman, M. Rahman, G. Mahiuddin, Y.M. Hsueh, L. Su, T. Smith, L. Ryan and D.C. Christiani Arsenic methylation, GSTT1, GSTM1, GSTP1 polymorphisms, and skin lesions, Environ Health Perspect 115 (2007) 341-345.

[26] R. Mukhopadhyay and B.P. Rosen Arsenate reductases in prokaryotes and eukaryotes, Environ Health Perspect 110 Suppl 5 (2002) 745-748.

[27] S.B. Waters, V. Devesa, L.M. Del Razo, M. Styblo and D.J. Thomas Endogenous reductants support the catalytic function of recombinant rat cyt19, an arsenic methyltransferase, Chem Res Toxicol 17 (2004) 404-409.

[28] M. Vahter, G. Concha, B. Nermell, R. Nilsson, F. Dulout and A.T. Natarajan A unique metabolism of inorganic arsenic in native Andean women, Eur J Pharmacol 293 (1995) 455-462.

[29] G. Concha, B. Nermell and M. Vahter Spatial and temporal variations in arsenic exposure via drinking-water in northern Argentina, J Health Popul Nutr 24 (2006) 317-326.

[30] A.L. Lindberg, W. Goessler, M. Grander, B. Nermell and M. Vahter Evaluation of the three most commonly used analytical methods for

determination of inorganic arsenic and its metabolites in urine, Toxicol Lett 168 (2007) 310-318.

[31] J.C. Barrett, B. Fry, J. Maller and M.J. Daly Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2005) 263-265.

[32] H.M. Custodio, R. Harari, L. Gerhardsson, S. Skerfving and K. Broberg Genetic influences on the retention of inorganic mercury, Arch Environ Occup Health 60 (2005) 17-23.

[33] M. Stephens and P. Donnelly A comparison of bayesian methods for haplotype reconstruction from population genotype data, Am J Hum Genet 73 (2003) 1162-1169.

[34] S. Wacholder, S. Chanock, M. Garcia-Closas, L. El Ghormli and N. Rothman Assessing the probability that a positive report is false: an approach for molecular epidemiology studies, J Natl Cancer Inst 96 (2004) 434-442.

[35] L. Conde, J.M. Vaquerizas, H. Dopazo, L. Arbiza, J. Reumers, F. Rousseau, J. Schymkowitz and J. Dopazo PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes, Nucleic Acids Res 34 (2006) W621-625.

[36] J. Reumers, L. Conde, I. Medina, S. Maurer-Stroh, J.V. Durme, J. Dopazo, F. Rousseau and J. Schymkowitz Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNPeffect and PupaSuite databases, Nucleic Acids Res (2007).

[37] J.R. Goni, X. de la Cruz and M. Orozco Triplex-forming oligonucleotide target sequences in the human genome, Nucleic Acids Res 32 (2004) 354-360.

[38] J.R. Goni, J.M. Vaquerizas, J. Dopazo and M. Orozco Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions, BMC Genomics 7 (2006) 63.

[39] M.J. Fazzari and J.M. Greally Epigenomics: beyond CpG islands, Nat Rev Genet 5 (2004) 446-455.

[40] M. Resseguie, J. Song, M.D. Niculescu, K.A. da Costa, T.A. Randall and S.H. Zeisel Phosphatidylethanolamine N-methyltransferase (PEMT) gene expression is induced by estrogen in human and mouse primary hepatocytes, Faseb J 21 (2007) 2622-2632.

[41] S.H. Zeisel Gene response elements, genetic polymorphisms and epigenetics influence the human dietary requirement for choline, IUBMB Life 59 (2007) 380-387.

[42] A.H. Carey, D. Waterworth, K. Patel, D. White, J. Little, P. Novelli, S. Franks and R. Williamson Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17, Hum Mol Genet 3 (1994) 1873-1876.

[43] H.S. Feigelson, G.A. Coetzee, L.N. Kolonel, R.K. Ross and B.E. Henderson A polymorphism in the CYP17 gene increases the risk of breast cancer, Cancer Res 57 (1997) 1063-1065.

[44] E.K. Kristensen VN, Andersson KB, Lønning PE, Erikstein B, Kåresen R, et al CYP17 and breast cancer. The polymorphism in the 5-flanking area of CYP17 does not influence binding to SP-1, Cancer Res 59 (1999) 2825–2828.

[45] Z. Wang, H. Zhang, X.F. Li and X.C. Le Study of interactions between arsenicals and thioredoxins (human and E. coli) using mass spectrometry, Rapid Commun Mass Spectrom 21 (2007) 3658-3666.

[46] A. Fredriksen, K. Meyer, P.M. Ueland, S.E. Vollset, T. Grotmol and J. Schneede Large-scale population-based metabolic phenotyping of thirteen genetic polymorphisms related to one-carbon metabolism, Hum Mutat 28 (2007) 856-865.

[47] D.J. Gaughan, L.A. Kluijtmans, S. Barbaux, D. McMaster, I.S. Young, J.W. Yarnell, A. Evans and A.S. Whitehead The methionine synthase reductase (MTRR) A66G polymorphism is a novel genetic determinant of plasma homocysteine concentrations, Atherosclerosis 157 (2001) 451-456.

[48] R.M. Gueant-Rodriguez, Y. Juilliere, M. Candito, C.E. Adjalla, P. Gibelin, B. Herbeth, E. Van Obberghen and J.L. Gueant Association of MTRRA66G polymorphism (but not of MTHFR C677T and A1298C, MTRA2756G, TCN C776G) with homocysteine and coronary artery disease in the French population, Thromb Haemost 94 (2005) 510-515.

[49] A. Laraqui, A. Allami, A. Carrie, A.S. Coiffard, F. Benkouka, A. Benjouad, A. Bendriss, N. Kadiri, N. Bennouar, A. Benomar, A. Guedira, A. Raisonnier, S. Fellati, J.E. Srairi and M. Benomar Influence of methionine synthase (A2756G) and methionine synthase reductase (A66G) polymorphisms on plasma homocysteine levels and relation to risk of coronary artery disease, Acta Cardiol 61 (2006) 51-61.

[50] J.D. Vaughn, L.B. Bailey, K.P. Shelnutt, K.M. Dunwoody, D.R. Maneval, S.R. Davis, E.P. Quinlivan, J.F. Gregory, 3rd, D.W. Theriaque and G.P. Kauwell Methionine synthase reductase 66A->G polymorphism is associated with increased plasma homocysteine concentration when combined with the homozygous methylenetetrahydrofolate reductase 677C->T variant, J Nutr 134 (2004) 2985-2990.

[51] S.E. Chiuve, E.L. Giovannucci, S.E. Hankinson, S.H. Zeisel, L.W. Dougherty, W.C. Willett and E.B. Rimm The association between betaine and choline intakes and the plasma concentrations of homocysteine in women, Am J Clin Nutr 86 (2007) 1073-1081.

[52] L. Cartegni, S.L. Chew and A.R. Krainer Listening to silence and understanding nonsense: exonic mutations that affect splicing, Nat Rev Genet 3 (2002) 285-298.

[53] L. Cartegni, J. Wang, Z. Zhu, M.Q. Zhang and A.R. Krainer ESEfinder: A web resource to identify exonic splicing enhancers, Nucleic Acids Res 31 (2003) 3568-3571.

[54] K.A. da Costa, O.G. Kozyreva, J. Song, J.A. Galanko, L.M. Fischer and S.H. Zeisel Common genetic polymorphisms affect the human requirement for the nutrient choline, Faseb J 20 (2006) 1336-1344.

[55] Figueroa LT, Razmilic BB and González MU Corporal distribution of arsenic in mummied bodies owned to an arsenical habitat., In: International seminar proceedings arsenic in the environment and its incidence on health (Sancha AM, ed); (1992).

[56] Núñez ALN, Hill HG and Martinez AL Guía Museo Arqueológico, Universidad Catoli´ca del Norte Chile. San Pedro de Atacama (II Regíon): Instituto de Investigaciones Arqueológicas y Museo RP. Gustavo Le Paige SJ., (1991).

[57] Albeza MV, Acreche NE, Caruso GB. Biodemografía en poblaciones de la Puna (Chañarcito, Santa Rosa de los Pastos Grandes y Olacapato), Salta, Argentina. Revista de Antropología Chilena, (2002) 119-126.

1

2

3
4

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, *TXN:* 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 (R^2=1). Rs3740393 (SNP12390), rs3740390 (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 populations[a] for statistically significant SNPs. Efficient methylation genotypes (low %iAs, low %MMA and high %DMA) are depicted in black.

[a]For PRXD2, reference populations are from Coriell Cell Repository data (NCBI SNP Database)

Table 1. Genes and polymorphisms genotyped.

Group of genes						
Gene	Accession no. [a]	Rs nr[b]	SNP type[c]	SNP Position[d]	Allele frequencies SAC (%)	Allele frequencies European population[e]
Methyltransferases						
AS3MT	Hs.34492	rs7085104	5' near gene, g.A>G	23377399, NT_030059	25/75	68/32
AS3MT		rs3740400	Intron, g.A>C	23377991, NT_030059	25/75	n/a[f]
CYP17A1	Hs.438016	rs743572	5'UTR, g.T>C	23345678, NT_030059	25/75	40/60
DNMT1	Hs.202672	rs16999593	p.H97R, g.T>C	1553983, NT_011295	93/7	100/0
DNMT1		rs8111085	p.I311V, g.T>C	1536174, NT_011295	63/37	96/4
DNMT1		rs7253062	Intron, g.G>A	1557926, NT_011295	92/8	56/44
°[g] *DNMT1*		rs11880388	Intron, g.G>A	1516375, NT_011295	n/a	49/51
DNMT3b	Hs.655708	rs2424913	Intron, g.C>T	1570351, NT_028392	6/94	66/34
DNMT3b		rs6087990	5' near gene, g.T>C	1546000, NT_028392	6/94	71/29
DNMT3b		rs2424932	3'UTR, g.G>A	1592628, NT_028392	6/94	52/48
One-carbon metabolism						
BHMT	Hs.80756	rs585800	3' UTR, g.A>T	29021566, NT_006713	97/3	69/31
BHMT		rs3733890	p.R239Q, g.G>A	29016317, NT_006713	52/48	72/28
CBS	Hs.533013	rs234715	Intron, g.G>T	1482836, NT_030188	99/1	78/22

CBS		rs2124485	g.C>T	30325934, NT_010966	99/1	100/0
CBS		rs234705	Intron, g.C>T	1478213, NT_030188	98/2	66/34
CBS		rs706209	3' UTR, g.C>T	1467866, NT_030188	98/2	58/42
CHDH	Hs.126688	rs9001	p.E40A, g.A>C	53797957, NT_022517	n/a[h]	93/7
CHDH		rs7626693	Intron, g.C>T	53798581, NT_022517	72/28	45/55
CHDH		rs9836592	Intron, g.T>C	53795123, NT_022517	98/2	67/32
°*CHDH*		rs11718497	Intron, g.G>C	53812057, NT_022517	n/a	64/36
MTHFR	Hs.214142	rs 1537516	Intron/3' UTR, g.C>T	6385228, NT_021937	97/3	90/10
MTHFR		rs1801131	p.E429A, g.A>C	6391843, NT_021937	97/3	64/36
MTHFR		rs17037396	Intron, g.C>T	6399414, NT_021937	97/3	87/13
MTRR	Hs.481551	rs10380	p.H595Y, g.C>T	7887191, NT_006576	48/52	80/20
MTRR		rs162036	p.K377R, g.A>G	7875959, NT_006576	49/51	77/23
MTRR		rs2287779	Synonymous, g.G>A	7879216, NT_006576	75/25	96/4
MTRR		rs1801394	I22M, g.A>G	7860973, NT_006576	83/17	55/45
TCN2	Hs.417948	rs2240433	5' near gene, g.A>G	10393707, NT_011520	55/44	n/a
TCN2		rs1801198	Arg259Pro, g.G>C	10402179, NT_011520	25/75	45/55
TCN2		rs5749134	Intron, g.C>G	10400811, NT_011520	25/75	42/58
Reduction reactions						
GCLC	Hs.654465	rs17883901	Promotor, g.C>T	44268268, NT_007592	99/1	94/6[i]
GCLM	Hs.315562	rs41303970	Promotor, g.C>T	64347228, NT_032977	48/52	84/16[j]
GGT1	Hs.645535	rs2236626	5' near gene, g.C>T	4370016, NT_011520	32/68	21/79[i]

GLRX	Hs.28988	rs3822751	Intron, g.G>C	3468082, NT_023148	51/49	30/70
GLRX		rs871775	5' near gene, g.C>T	3472640, NT_023148	99/1	91/9
GLRX		rs4561	Synonymous, g.C>T	3466185, NT_023148	7/93	41/59
GSR	Hs.271510	rs8190955	p.R153C, g.C>T	886398, NT_007995	99/1	100/0
GSR		rs2253409	Intron, g.G>C	867740, NT_007995	86/14	74/26
GSR		rs2978296	Intron, g.G>C	895714, NT_007995	76/24	83/17
GSR		rs2280820	g.T>C	1955535, NT_023736	100/0	36/65
GSTA1	Hs.446309	rs3957356	Promotor, g.C>T	43526901, NT_007592	7/93	61/39[k]
GSTP1	Hs.523836	rs1695	p.I105V, g.A>G	12658484, NT_033903	27/73	61/39
PRDX2	Hs.695971	rs10427027	3' near gene, g.T>C	12772285, NT_011295	88/12	95/5[i]
PRDX2		rs12151144	Intron, g.A>C	4175198, NT_011295	88/12	95/5[i]
TXN	Hs.435136	rs1049927	5'UTR, g.A>G	20339960, NT_008470	99/1	92/8
TXN		rs2301242	5' near gene, g.T>A	20340207, NT_008470	99/1	71/29
TXNRD1	Hs.696144	rs11111979	5'UTR, g.C>G	28162972, NT_019546	26/74	49/51
TXNRD1		rs6539137	Intron, g.T>A	28189370, NT_019546	93/7	86/14
TXNRD2	Hs.443430	rs5746847	Intron, g.C>T	3073153, NT_011519	45/55	57/43
°*TXNRD2*		rs1139793	p.I370T, g.A>G	3020368, NT_011519	n/a	32/68
TXNRD2		rs5992495	p.S229R, g.T>G	3035134, NT_011519	92/8	85/15
TXNRD2		rs5748469	p.A66S, g.C>A	3059249, NT_011519	91/9	65/35

[a] Unigene Accession number from NCBI (National Center for Biotechnology Information) Unigene Database (website: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene).

[b] Rs numbers for the SNPs from NCBI SNP Database (website: http://www.ncbi.nlm.nih.gov/SNP).

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

[d] Contig accession numbers (NT_) and SNP positions from NCBI Nucleotide database (website: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide).

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

[f] n/a = not available.

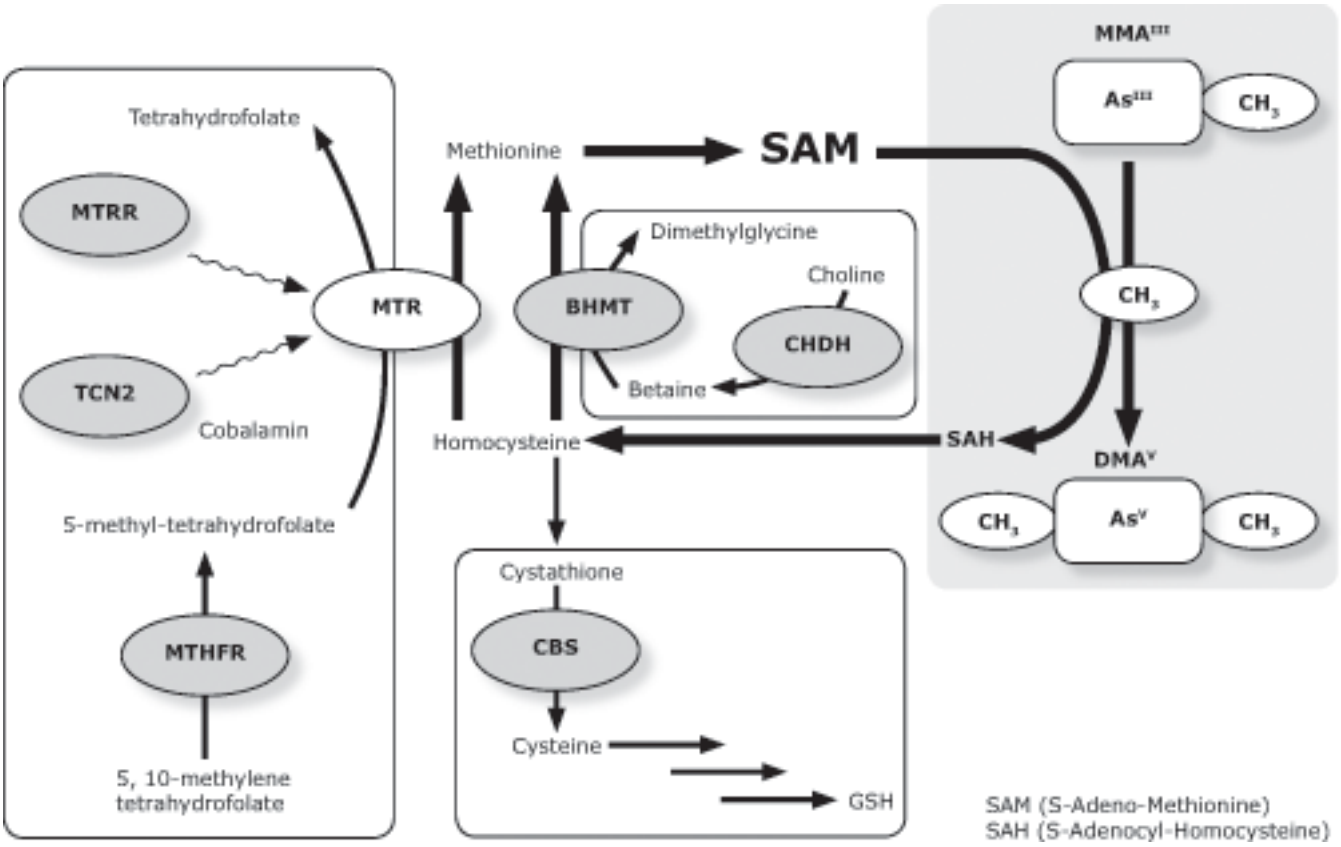
[g] ° denotes SNPs not included in the statistical analysis due to a failure rate of over 10%

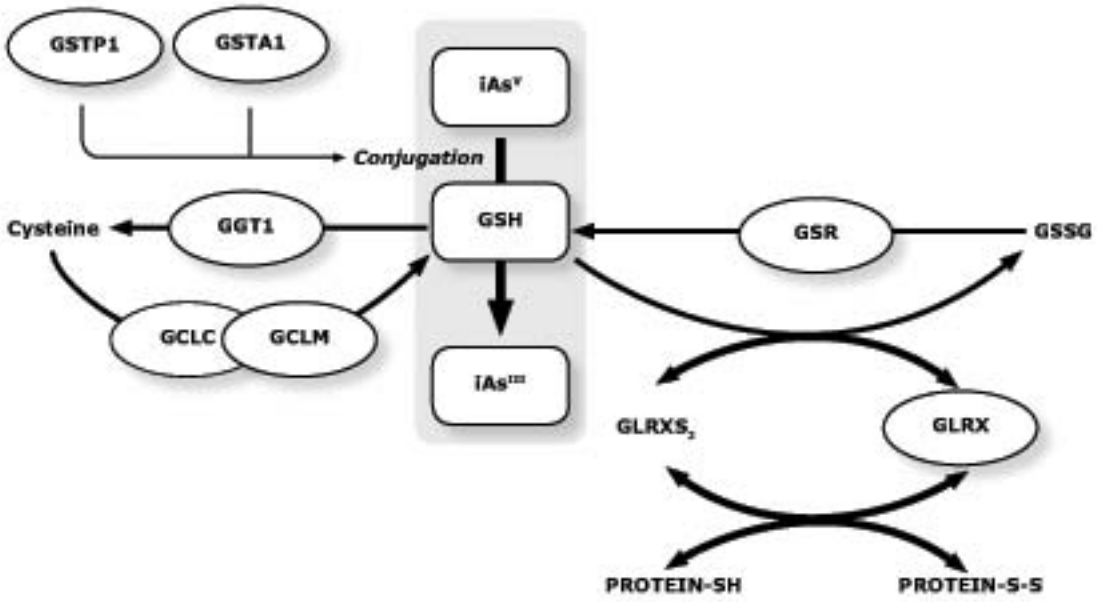
[h] No frequencies are presented due to difficulties to distinguish between CA and AA genotypes.

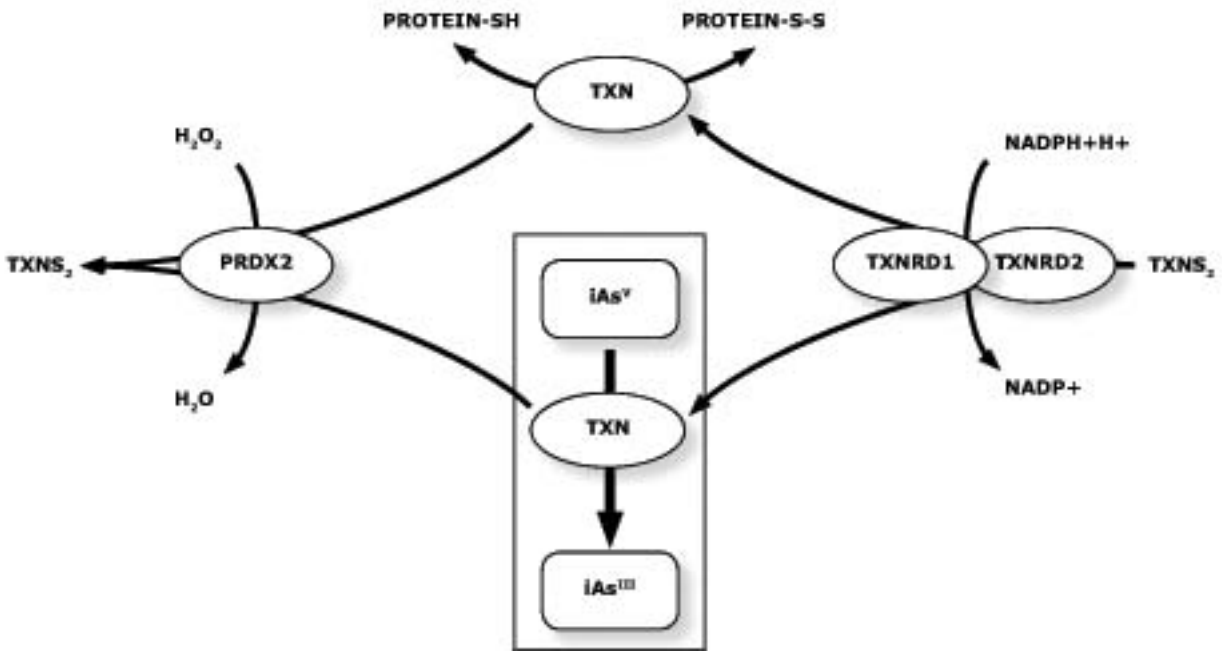
[i] Allele frequencies from Coriell Cell Repository at NCBI SNP Database.

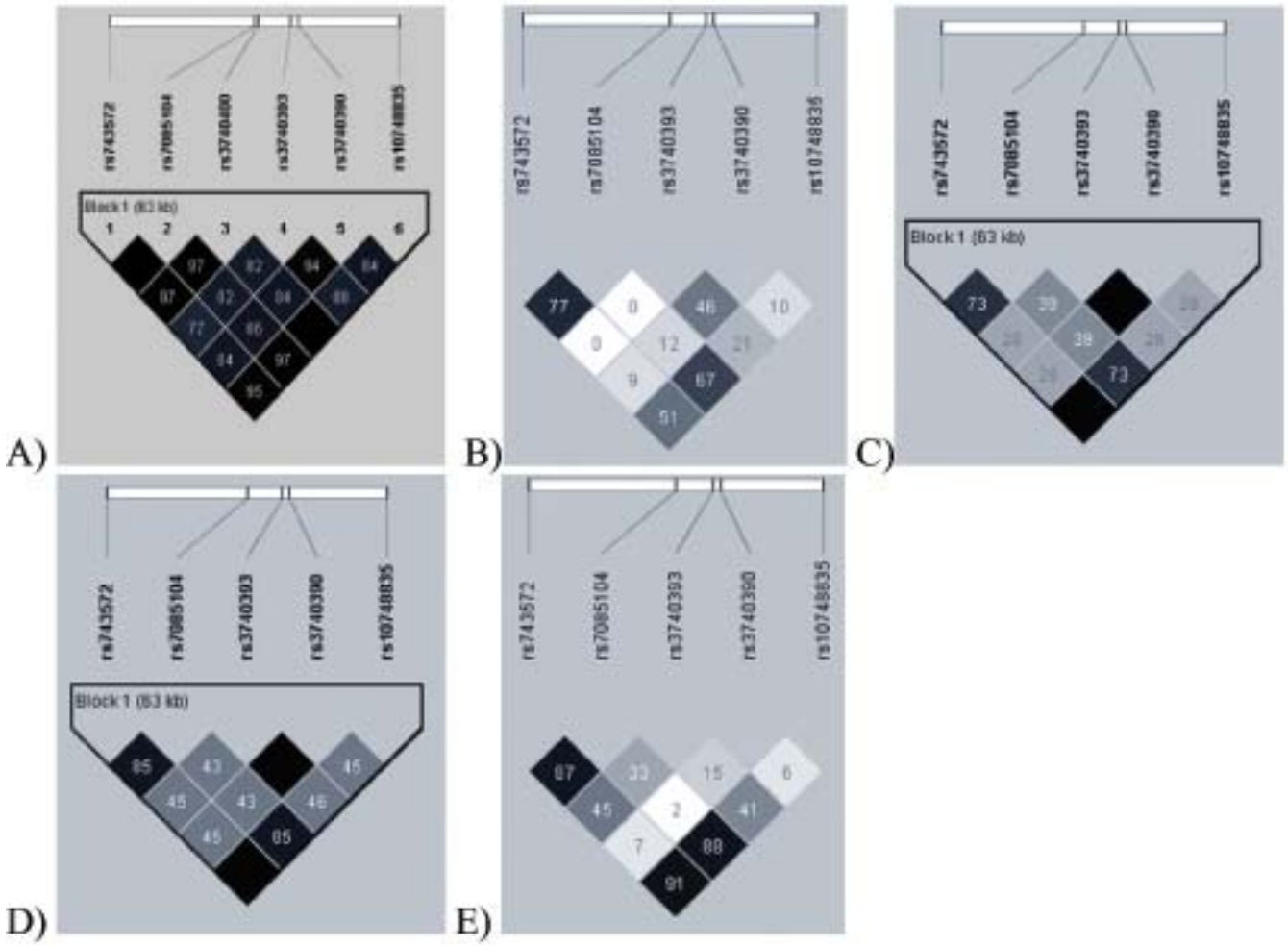
[j] Allele frequencies from a Danish/Swiss population (Submitter handle CNPSCZ at NCBI SNP Database).

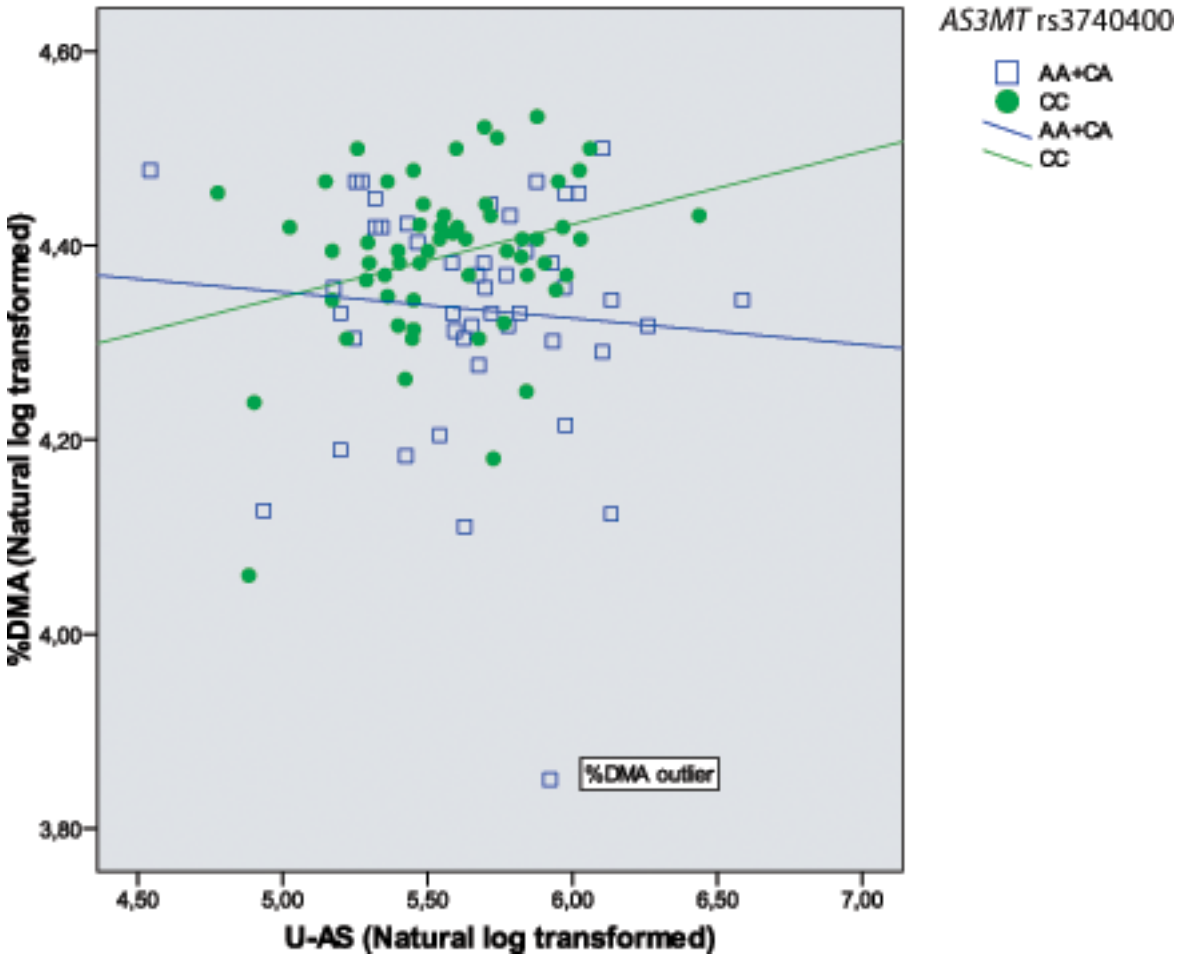
[k] Allele frequencies from the Caucasian population in the NCI SNP500Cancer Database (website: http://snp500cancer.nci.nih.gov).

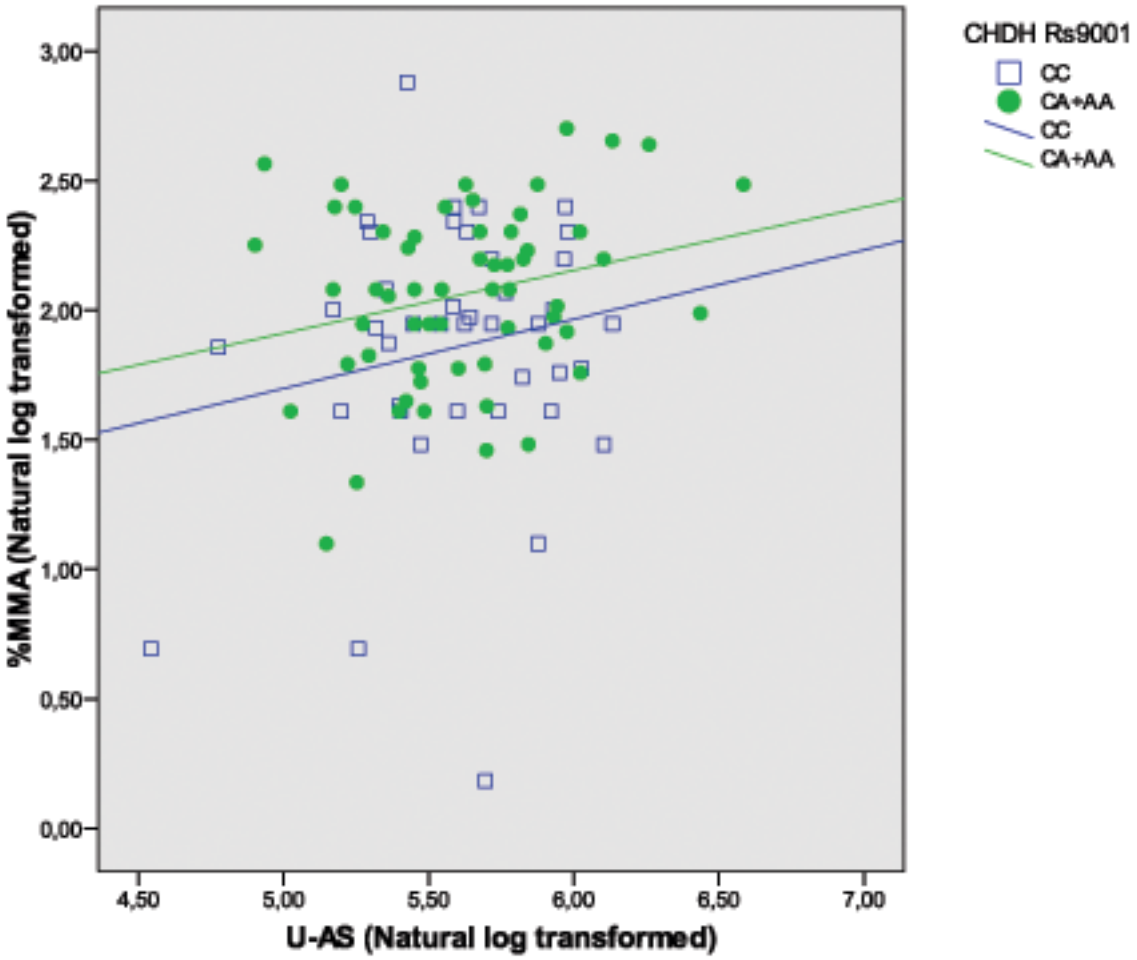

MMA^{III}
As^{III}
CH_3
Tetrahydrofolate
Methionine
SAM
MTRR
Dimethylglycine
Choline
MTR
BHMT
CHDH
CH_3
TCN2
Betaine
Cobalamin
Homocysteine
SAH
DMA^{V}
5-methyl-tetrahydrofolate
CH_3
As^{V}
CH_3
Cystathione
MTHFR
CBS
Cysteine
GSH
5, 10-methylene
tetrahydrofolate
SAM (S-Adeno-Methionine)
SAH (S-Adenocyl-Homocysteine)

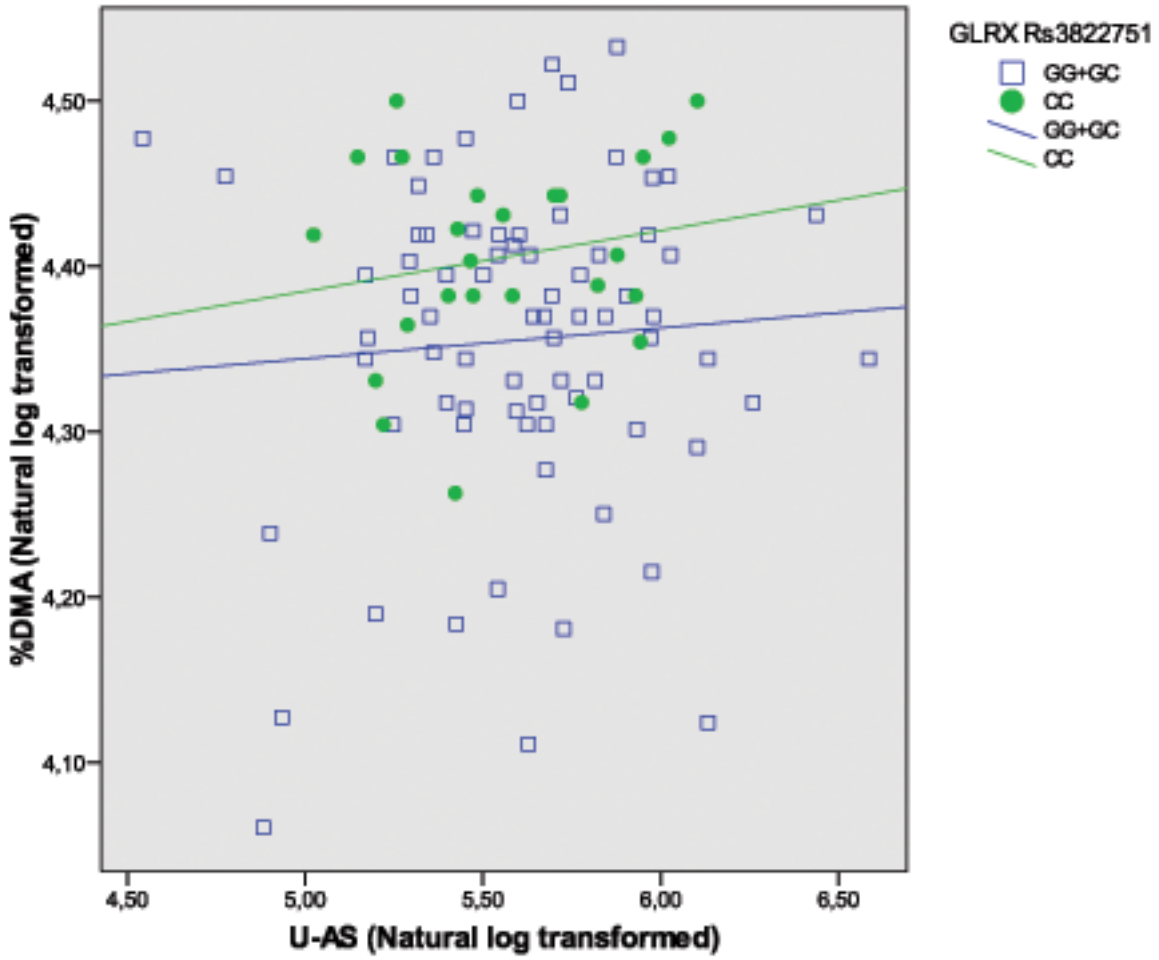

GSTP1
GSTA1
iAs^{V}
Conjugation
Cysteine
GGT1
GSH
GSR
GSSG
GCLC
GCLM
iAs^{III}
GLRXS$_2$
GLRX
PROTEIN-SH
PROTEIN-S-S

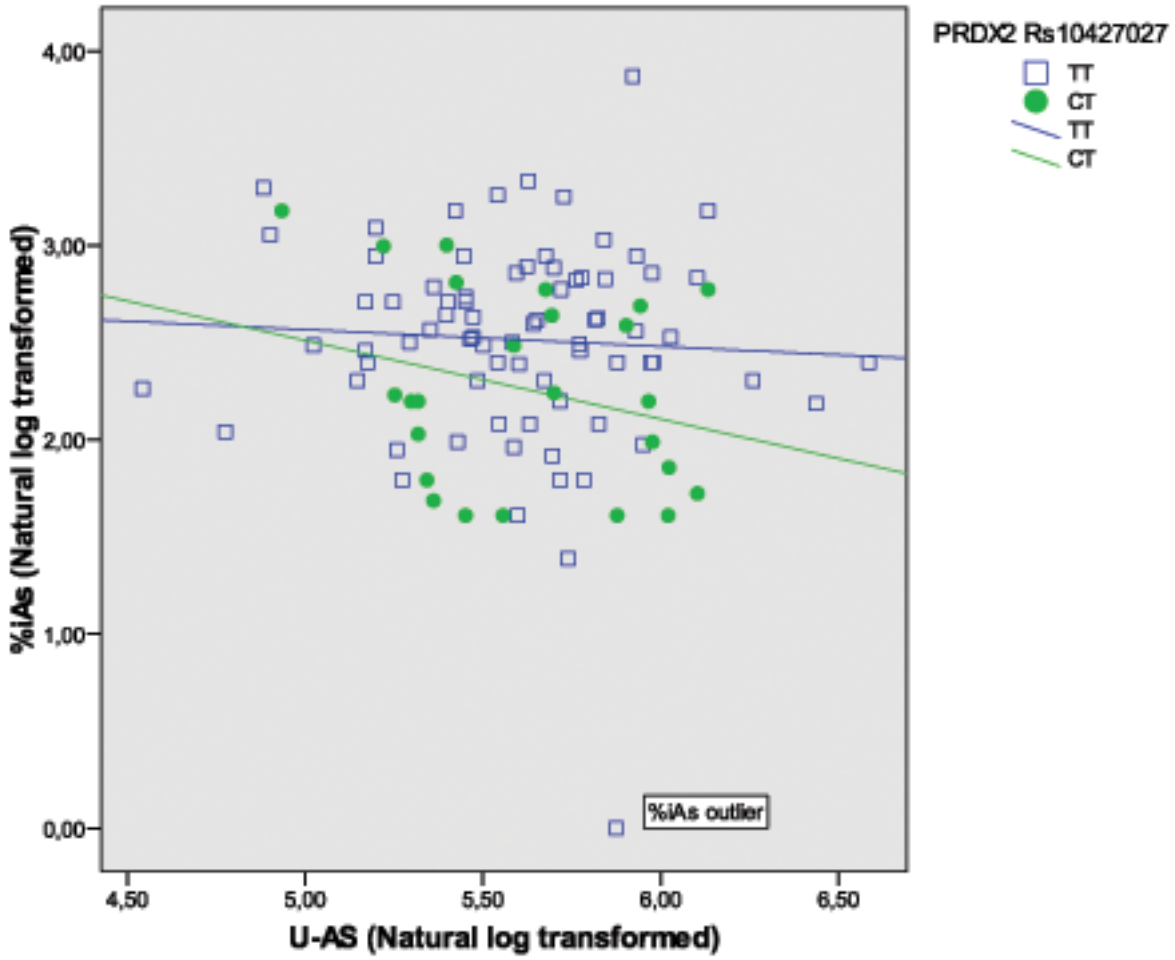

PROTEIN-SH
PROTEIN-S-S
TXN
H$_2$O$_2$
NADPH+H+
TXNS$_2$
PRDX2
TXNRD1
TXNRD2
TXNS$_2$
iAs^{V}
H$_2$O
NADP+
TXN
iAs^{III}

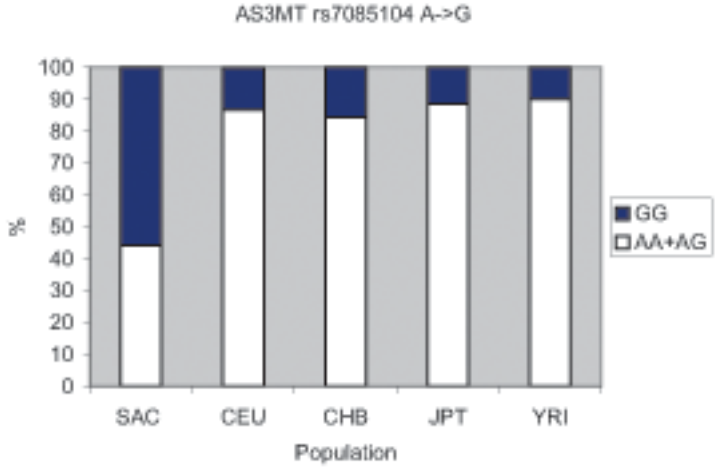

A)
rs743572
rs7085104
rs3740393
rs3740390
rs10748835
Block 1 (63 kb)
1 2 3 4 5 6
97 82 84 84
87 92 84 88
77 86
84 97
95
B)
rs743572
rs7085104
rs3740393
rs3740390
rs10748835
77 0 46 10
0 12 21
9 67
51
C)
rs743572
rs7085104
rs3740393
rs3740390
rs10748835
Block 1 (63 kb)
73 29 26
28 38 28
29 73
D)
rs743572
rs7085104
rs3740393
rs3740390
rs10748835
Block 1 (63 kb)
85 43 45
45 43 46
45 85
E)
rs743572
rs7085104
rs3740393
rs3740390
rs10748835
87 33 15 6
45 2 41
7 88
91

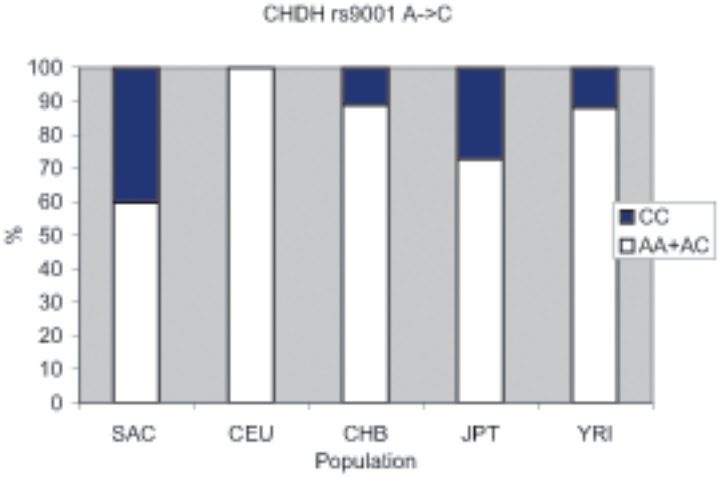

AS3MT rs3740400
AA+CA
CC
AA+CA
CC
%DMA (Natural log transformed)
4,60
4,40
4,20
4,00
3,80
%DMA outlier
4,50
5,00
5,50
6,00
6,50
7,00
U-AS (Natural log transformed)

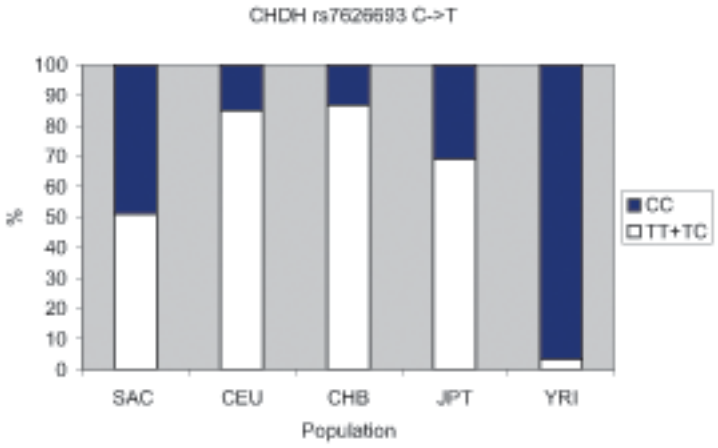

CHDH Rs9001
CC
CA+AA
CC
CA+AA
%MMA (Natural log transformed)
3,00
2,50
2,00
1,50
1,00
0,50
0,00
4,50
5,00
5,50
6,00
6,50
7,00
U-AS (Natural log transformed)

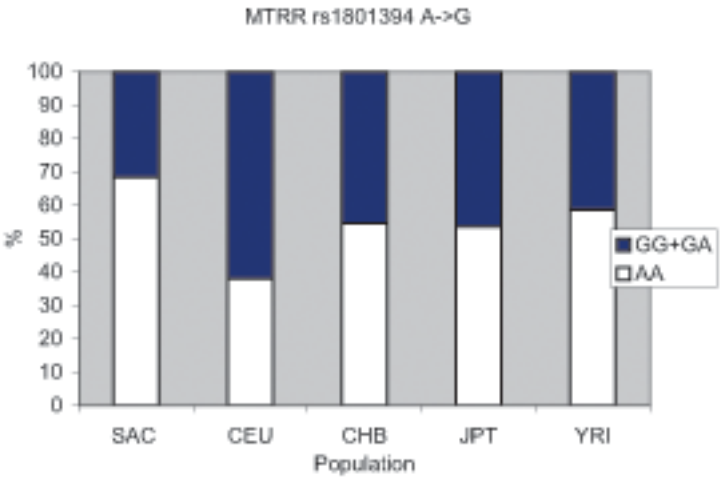

GLRX Rs3822751
GG+GC
CC
GG+GC
CC
%DMA (Natural log transformed)
4,50
4,40
4,30
4,20
4,10
4,50
5,00
5,50
6,00
6,50
U-AS (Natural log transformed)

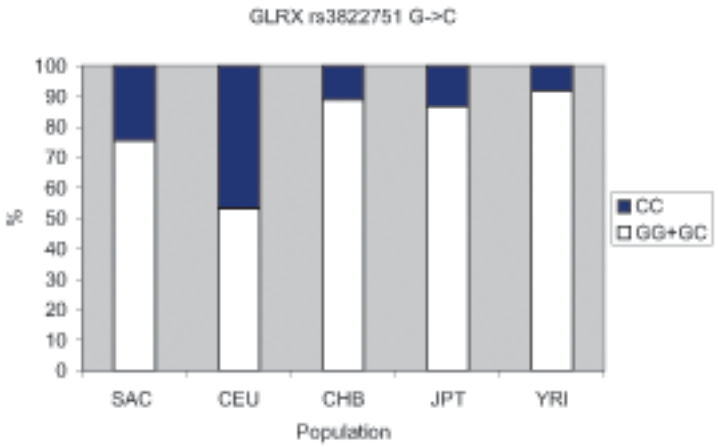

PRDX2 Rs10427027
TT
CT
TT
CT
%iAs (Natural log transformed)
4,00
3,00
2,00
1,00
0,00
%iAs outlier
4,50
5,00
5,50
6,00
6,50
U-AS (Natural log transformed)

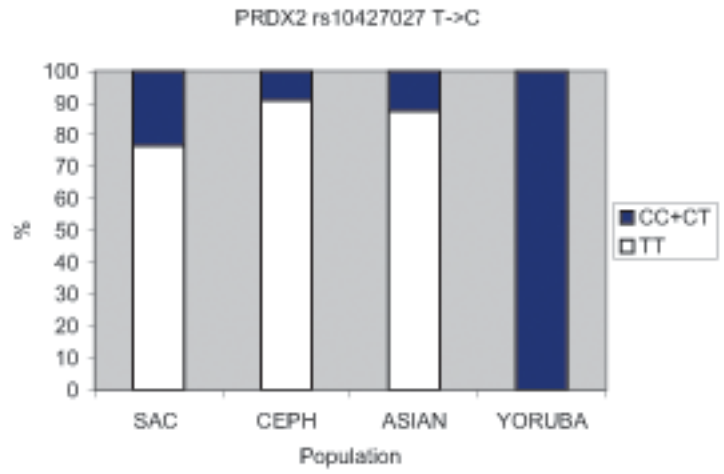

AS3MT rs7085104 A->G
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEU
CHB
JPT
YRI
Population
GG
AA+AG


CHDH rs9001 A->C
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEU
CHB
JPT
YRI
Population
CC
AA+AC


CHDH rs7626693 C->T
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEU
CHB
JPT
YRI
Population
CC
TT+TC


MTRR rs1801394 A->G
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEU
CHB
JPT
YRI
Population
GG+GA
AA


GLRX rs3822751 G->C
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEU
CHB
JPT
YRI
Population
CC
GG+GC


PRDX2 rs10427027 T->C
%
100
90
80
70
60
50
40
30
20
10
0
SAC
CEPH
ASIAN
YORUBA
Population
CC+CT
TT