OXFORD

SOT | Society of Toxicology
academic.oup.com/toxsci

TOXICOLOGICAL SCIENCES, 2020, 382–395

doi: 10.1093/toxsci/kfaa075
Advance Access Publication Date: 20 May 2020
Research article

Genetic Determinants of Reduced Arsenic Metabolism Efficiency in the 10q24.32 Region Are Associated With Reduced *AS3MT* Expression in Multiple Human Tissue Types

Meytal Chernoff,[*,†] Lin Tong,[*] Kathryn Demanelis,[*] Donald Vander Griend,[‡] Habib Ahsan,[*] and Brandon L. Pierce[*,§,1]

[*]The Department of Public Health Sciences, The University of Chicago, Chicago, Illinois 60637 - 1447; [†]The Interdisciplinary Scientist Training Program, The University of Chicago, Chicago, Illinois 60637; [‡]The Department of Pathology, The University of Illinois at Chicago, Chicago, Illinois 60612; and [§]The Department of Human Genetics, The University of Chicago, Chicago, Illinois 60637

The authors certify that all research involving human subjects was done under full compliance with all government policies and the Helsinki Declaration.

[1]To whom correspondence should be addressed at the Department of Public Health Sciences, The University of Chicago Medical Center, 5841 South Maryland Avenue, M2000, Chicago, IL 60637. Fax: 773-834-0139. Email: brandonpierce@uchicago.edu.

ABSTRACT

Approximately 140 million people worldwide are exposed to inorganic arsenic through contaminated drinking water. Chronic exposure increases risk for cancers as well as cardiovascular, respiratory, and neurologic diseases. Arsenic metabolism involves the *AS3MT* (arsenic methyltransferase) gene, and arsenic metabolism efficiency (AME, measured as relative concentrations of arsenic metabolites in urine) varies among individuals. Inherited genetic variation in the 10q24.32 region, containing *AS3MT*, influences AME, but the mechanisms remain unclear. To better understand these mechanisms, we use tissue-specific expression data from GTEx (Genotype-tissue Expression project) to identify *cis*-eQTLs (expression quantitative trait loci) for *AS3MT* and other nearby genes. We combined these data with results from a genome-wide association study of AME using "colocalization analysis," to determine if 10q24.32 SNPs (single nucleotide polymorphisms) that affect AME also affect expression of *AS3MT* or nearby genes. These analyses identified *cis*-eQTLs for *AS3MT* in 38 tissue types. Colocalization results suggest that the casual variant represented by AME lead SNP rs4919690 impacts expression of *AS3MT* in 13 tissue types (> 80% probability). Our results suggest this causal SNP also regulates/coregulates expression of nearby genes: *BORCS7* (43 tissues), *NT5C2* (2 tissues), *CYP17A1-AS1* (1 tissue), and *RP11-724N1.1* (1 tissue). The rs4919690 allele associated with decreased AME is associated with decreased expression of *AS3MT* (and other coregulated genes). Our study provides a potential biological mechanism for the association between 10q24.32 variation and AME and suggests that the causal variant, represented by rs4919690, may impact AME (as measured in urine) through its effects on arsenic metabolism occurring in multiple tissue types.

Key words: arsenic metabolism efficiency; colocalization; *AS3MT*; multitissue; expression quantitative trait locus (eQTL); single nucleotide polymorphism (SNP); Genotype-tissue Expression (GTEx) project.

The World Health Organization (WHO) reports that ≥ 140 million people across 50 countries consume water containing arsenic ≥ 10 μg/L (Ravenscroft *et al.*, 2009; WHO, 2018) (often naturally contaminated groundwater). Chronic exposure to arsenic has been linked to many health conditions including arsenical skin lesions, cardiovascular disease, diabetes, respiratory and neurologic diseases, and obstetric problems (Hong *et al.*, 2014; Rahman *et al.*, 2009). The International Agency for Research on Cancer classifies arsenic as a human carcinogen (Group 1), and chronic arsenic exposure has been linked to risk of bladder, liver, kidney, lung, prostate, and skin cancers (Hong *et al.*, 2014; Hopenhayn-Rich *et al.*, 1998; WHO, 2018).

Inorganic arsenic (iAs) (arsenite, As^{III}, and arsenate, As^{V}) is the primary arsenic species ingested via contaminated drinking water, and is believed to be primarily metabolized in the liver. Based on the Challenger model (Challenger, 1945; Gao *et al.*, 2015; Rehman and Naranmandura, 2012) of arsenic metabolism, iAs in the pentavalent state (As^{V}) is reduced to the trivalent state (As^{III}); As^{III} then undergoes oxidative methylation by arsenic methyltransferase (encoded by *AS3MT*) using S-adenosylmethionine (SAM) as the methyl donor. This produces monomethylarsonic acid (MMA^{V}), which can be reduced to monomethylarsonous acid (MMA^{III}). A second methylation reaction produces dimethylarsinic acid (DMA^{V}) that can then be reduced to dimethylarsinous acid (DMA^{III}), the end product of arsenic metabolism. These arsenic species can be excreted in urine, where DMA accounts for the largest proportion of arsenic species excreted (Ameer *et al.*, 2016; De Loma *et al.*, 2019; Vahter, 1999, 2002). The relative percentage of urinary arsenic metabolites (eg, DMA%) is often used as a measure of arsenic metabolism efficiency (AME) (Ahsan *et al.*, 2007; Lindberg *et al.*, 2008). Therefore, high DMA% represents an enhanced AME, which is believed to facilitate the clearance of arsenic from the body and decrease the risk of arsenic-related toxicities (Pierce *et al.*, 2013).

A previous genome-wide association study (GWAS) of Bangladeshi individuals, who were exposed to a wide range of arsenic levels in drinking water, showed that genetic variation in the 10q24.32 region influences AME (Pierce *et al.*, 2012). This region contains the *AS3MT* gene, which resides within a large block of linkage disequilibrium (LD) covering approximately 1 Mb. Genetic polymorphisms in and around this gene have been reported as associated with AME, as measured by urinary arsenic metabolites, in multiple exposed populations (Agusa *et al.*, 2009; Balakrishnan *et al.*, 2017; Engström *et al.*, 2007; Meza *et al.*, 2005; Wood *et al.*, 2006). The importance of *AS3MT* for arsenic metabolism is well-established. *AS3MT* knockout mice exposed to iAs have markedly reduced capacities to eliminate arsenic compared with wild-type mice (Drobna *et al.*, 2009; Marafante and Vahter, 1984). Knockout mice retain higher levels of arsenic, particularly unmethylated arsenic, in tissues (Chen *et al.*, 2011; Drobna *et al.*, 2009; Hughes *et al.*, 2010), and are at much higher risk for multiple toxicity-related phenotypes including regenerative hyperplasia and cytotoxicity of the uroepithelium (Yokohira *et al.*, 2010), toxicity-associated mortality (Yokohira *et al.*, 2010), inflammatory cell infiltration of the liver (Yokohira *et al.*, 2010), and obesity and insulin resistance (Douillet *et al.*, 2017).

In our previous GWAS, we reported rs9527 as the single nucleotide polymorphism (SNP) most significantly associated with DMA% in the *AS3MT* region (after imputation rs4919690 was identified as the lead SNP). Conditioning on this primary association signal revealed a secondary signal, reported initially as rs11191527 (rs75075426 was identified as the lead SNP after imputation). For this study, we utilized existing array-based SNP data for the Bangladeshi individuals included in this prior analysis. This array measured only a subset of SNPs in the 10q24.32 region, so we used imputation to generate information on SNPs not measured on the array and performed conditional association analyses to identify the lead SNPs for each independent association signal. For each signal, the lead SNP may not be the causal SNP; however, colocalization analysis does not require specifying a causal or lead SNP. The nonimputed results were shown in our previous paper (Pierce *et al.*, 2012); we show the imputed results in Figure 1. After adjusting for these 2 independent association signals, no further evidence of association with DMA% was observed (Figure 1) (Pierce *et al.*, 2012).

Given the global impact of arsenic exposure on health, there remains a need to understand the biological mechanisms by which these variants impact arsenic-related phenotypes. One possibility is that SNPs alter the amino acid sequence of the *AS3MT* protein product in a way that impacts the enzyme's function; however, as none of our previously identified top SNPs reside in exons, this mechanism is unlikely. It is also possible that these variants impact AME through regulation of *AS3MT* expression (or expression of other genes). Using results from our previous GWAS of AME and tissue-specific gene expression data from human tissue donors from the Genotype-tissue Expression (GTEx) project, we sought to determine if SNPs affecting DMA% also have local (ie, *cis*) effects on the expression of *AS3MT* or other nearby genes. We quantify the probability of a common causal variant (CCV) driving the association with DMA% and gene expression using Bayesian colocalization analysis. Through this work, we aim to provide a potential mechanism for the observed association between 10q24.32 variation and AME and identify tissue types in which this mechanism likely operates.

MATERIALS AND METHODS

Study Descriptions

Our previous GWAS of urinary arsenic species was conducted using data on 2056 individuals randomly selected from the Health Effects of Arsenic Longitudinal Study (HEALS). Data on tissue-specific gene expression and *cis*-eQTLs (expression quantitative trait loci) were obtained from the GTEx project (v7).

The Health Effects of Arsenic Longitudinal Study (HEALS). HEALS is a prospective cohort study of a population from Araihazar, Bangladesh exposed to iAs through naturally contaminated drinking water (Ahsan *et al.*, 2006). Established in 2000, the study seeks to understand a range of health outcomes associated with arsenic exposure, including arsenic-induced skin lesions and skin cancers. Recruitment for the original HEALS cohort occurred from October 2000 through May 2002; during this time 11 746 men and women (age 18–75 years) were enrolled. At baseline, study physicians, blinded to arsenic measurements, used structured protocols to conduct in-person interviews and clinical evaluations, and to collect spot urine and blood samples. In-person follow-up interviews were conducted biennially. Information on arsenic exposure was obtained by testing all 5966 wells in the study area for arsenic and from individual reporting of the primary well from which they obtained their drinking water. The protocol for this study was approved by the Institutional Review Boards of the University of Chicago, Columbia University, and the Bangladesh Medical Research Council. Informed consent was obtained from all participants.

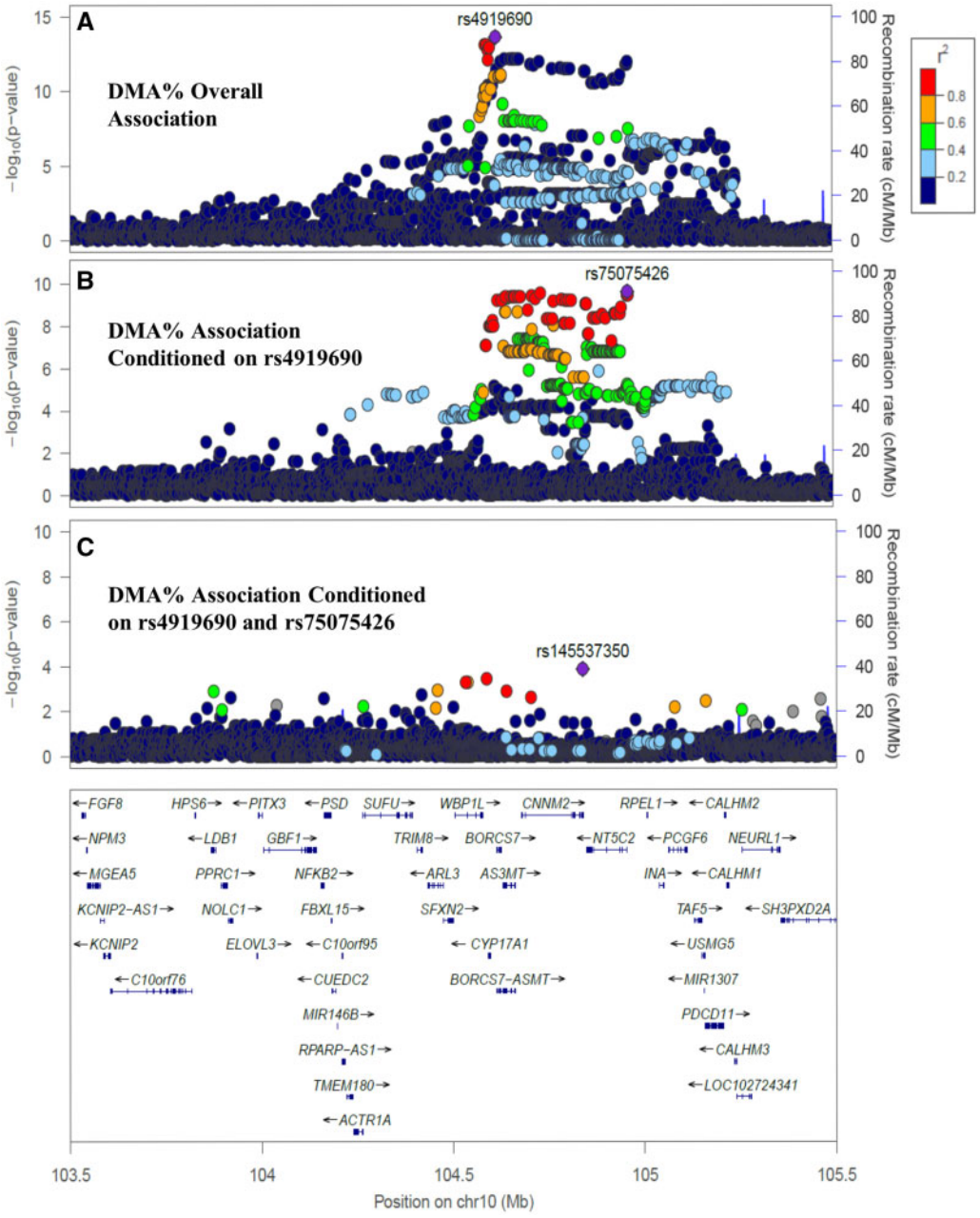

Figure 1. Two variants in the 10q24.32 region show independent association with DMA% in the Health Effects of Arsenic Longitudinal Study (HEALS). Results from a prior genome-wide association study of arsenic metabolism efficiency (DMA%) using genotyping and arsenic metabolite data from the HEALS cohort, plot generated using LocusZoom (Pruim *et al.*, 2010). The single nucleotide polymorphism (SNP) with the strongest association with DMA% is labeled in each panel. Panel A shows the overall association results. Panel B shows *p* values from models adjusted for rs4919690. Panel C shows the *p* values from models adjusted for both rs4919690 and rs75075426, which here yielded no significant SNPs.

The Columbia University Trace Metals Core Laboratory used graphite furnace atomic absorption spectrometry (Sigma, St Louis, Missouri) (Nixon *et al.*, 1991) to measure urinary arsenic. Urinary arsenic species were distinguished using high-performance liquid chromatography for metabolite separation and plasma mass spectrometry with a dynamic reaction cell for detection (as previously described by Ahsan *et al.* [2007]). Total concentration of arsenic metabolites (iAs, MMA, and DMA) were represented as a percentage of total arsenic after subtracting organic arsenic from dietary sources (arsenobetaine and arsenocholine).

The Genotype-tissue Expression (GTEx) project data version 7. Established in 2010 by the NIH, this project established a centralized data source for information on inherited genetic variation and gene expression in multiple human tissue types (Carithers *et al.*, 2015; GTEx Consortium, 2017). Numerous tissues are collected from each participating donor, enabling the study of tissue-specific gene expression and its relationship with genetic variation. Expression data for the project was obtained through RNA sequencing using Illumina's TruSeq library construction protocol (GTEx Consortium, 2017; GTEx Portal, 2019). Genotype data was generated from whole genome

sequencing obtained using Illumina's HiSeq X (with the first batch run on HiSeq 2000). Genotype data on 620 individuals was used in this study.

Genotyping and quality control for HEALS data. For 3454 HEALS participants, DNA was extracted from clot blood using the Flexigene DNA kit (Cat No. 51204) from Qiagen, and sample concentration and purity was assessed by NanoDrop 1000. Genome Reference Consortium Human Build 37 was used for sequencing and analysis. From each sample, 250 ng of DNA was then genotyped using Illumina HumanCytoSNP-12 v2.1 chips, which included 299 140 SNP markers, and read on the BeadArray reader. Genotype calls were generated by processing image data using the BeadStudio software.

All Quality Control (QC) was performed using PLINK (Purcell *et al.*, 2007). We excluded DNA samples with poor call rates (< 90%) ($n = 6$), as well as monomorphic ($n = 38\,753$) or poorly called (< 90%) SNPs. Individuals with mismatched sex ($n = 36$) based on genetics and self-reporting were also removed. Once we examined the distributions of sample and SNP call rates, any sample with call rate < 97% ($n = 4$) or SNP with call rate < 95% ($n = 1045$) was also excluded. Finally, we excluded SNPs with Hardy-Weinberg equilibrium p values $< 10^{-10}$ ($n = 634$). From our initial pool of 3454 individuals, this QC resulted in 3364 individuals with high-quality genotype data for 257 747 SNPs. Of these individuals, a subset of 2056 individuals with arsenic metabolite data was included in the GWA analysis. The Michigan Imputation Server (Das *et al.*, 2016) was used to conduct genotype imputation using 1000 Genomes reference haplotypes (1KG phase3 v5, which includes overall populations). Only high-quality imputed biallelic SNPs (imputation $r^2 > 0.3$) and SNPs with minor allele frequency >0.005 were included in the analysis.

The initial identification of arsenic-metabolism associated SNPs in the 10q24.32 region using HEALS data has been previously reported by Pierce *et al.* (2012). In short, pairwise kinship coefficients were estimated using GEMMA (Zhou and Stephens, 2012). All SNP association tests were performed in GEMMA (Zhou and Stephens, 2012) using a mixed model that accounted for cryptic relatedness and adjusted for sex, age, and genotyping batch. Regional association plots were then generated using LocusZoom (Pruim *et al.*, 2010).

Identifying eQTLs analyzed using GTEx data. The *cis*-eQTLs used in this paper were identified using expression data from the 48 tissue types in GTEx release v7. This GTEx release included 620 donors whose data was used for eQTL analysis, and the sample sizes for each tissue used in this analysis ranged from 88 to 491. Donor enrollment and the consent process as well as specimen procurement, sample fixation, and histological review procedures for the GTEx project have been previously described (Carithers *et al.*, 2015; GTEx Consortium, 2017).

Expression quantification for GTEx samples has been described in detail previously by the GTEx Consortium (2017). Gene expression data for a particular tissue were normalized between samples using TMM as implemented in edgeR (Robinson *et al.*, 2010; Robinson and Oshlack, 2010). Expression values for each gene were normalized across samples using an inverse normal transformation. RNA-seq expression outliers were identified and excluded along with samples that produced < 10 million mapped reads. The expression data was downloaded from data version 7 on the GTEx portal. (https://gtexportal.org/home/datasets; last accessed May 27, 2020)

We performed *cis*-eQTL mapping for *AS3MT* (and genes in the 10q24.32 region within 500 kb of SNP rs4919690) using linear regression, implemented with Matrix eQTL (Shabalin, 2012). The eQTL analyses used a 1 Mb window (500 kb on each side of rs4919690) to designate SNPs to be tested for association with a gene. The restriction of *cis*-eQTL mapping to a 1 Mb distance is common in studies of *cis*-eQTLs, including those performed by GTEx (GTEx Consortium, 2017). However, to ensure our lead DMA% SNPs did not have long-range *cis* effects on genes beyond 500 kb, we assessed association between these SNPs and all genes within 5 Mb (on each side), but we observed no additional genes showing association ($p < 5\times10^{-8}$); our DMA%-associated SNPs were only associated with the expression of genes within 500 kb of the SNPs. Regressions were adjusted for all suggested covariates provided by GTEx within the covariate file for each tissue. This included adjustment for 3 genotyping principle components, sex, and a set of covariates identified with the Probabilistic Estimation of Expression Residuals (PEER) method. The PEER method is designed to account for nongenetic factors impacting observed gene expression (Stegle *et al.*, 2010). This enables researchers to understand the factors contributing to variation in gene expression by jointly modeling the underlying (nongenetic) causes of variability and the genetic effects to determine the contributions from different sources of variation. The PEER factors were calculated for the normalized expression matrices and provided by GTEx for each tissue. The number of PEER factors for each tissue was determined as a function of the sample size available for each tissue, resulting in a different number of PEER covariates for each tissue. In studies from the GTEx consortium, these PEER factors were strongly associated with known technical confounders and captured 59%–78% of the total variation in gene expression levels (GTEx Consortium, 2017).

For each variant-gene pair, nominal p values were estimated with a 2-tailed t test. We used a p value threshold of 5×10^{-8} to identify *cis*-eQTLs. This threshold is loosely based on a Bonferroni correction for the total number of tests performed, which included testing approximately 2655 SNPs for each of 19 genes within our region of interest in a total of 48 tissue types ($2655 \times 19 \times 48 = 2\,421\,360$ total tests; $0.05/2\,421\,360 = 2.1\times10^{-8}$ Bonferroni corrected threshold). Using the standard GWAS significance threshold of $p = 5\times10^{-8}$ is slightly less stringent, but reasonable (and likely overly conservative) in light of the correlation among test statistics due to LD in the region.

For each *cis*-eQTL identified, we performed a conditional analysis of the region to identify secondary *cis*-eQTLs for the gene being analyzed by adjusting for the lead SNP of the primary eQTL. For identified secondary eQTLs ($p < 5\times10^{-8}$), we conducted additional analyses to identify additional eQTLs, adjusting for the top SNPs of both the primary and secondary eQTLs. Continuing this way, we were able to identify all *cis*-eQTLs for each gene in a given tissue ($p < 5\times10^{-8}$) and "isolate" each eQTL's association signal using adjustments for the lead SNPs of all other *cis*-eQTLs for a given gene.

To identify lead-SNP eQTL pairs that potentially share a CCV with 1 of our 10q24.32 association signals for DMA%, we first made a list of the lead SNPs for all identified eQTLs. Then we identified lead eQTL SNPs that were in strong LD ($r^2 > 0.7$) with either of the lead SNPs from our 2 association signals for DMA% (based on 1000 Genomes data for European and Bengali in Bangladesh populations).

Colocalization analyses. To estimate the probability that the associations we observe for DMA% (in HEALS) and the tissue-

specific *cis*-eQTLs in the 10q24.32 region (from GTEx) are due to the same causal variant(s), we used a Bayesian test for colocalization (Giambartolomei *et al.*, 2014) implemented in the coloc package in R (Wallace, 2013, Wallace *et al.*, 2012). We applied this test to the DMA% association signals paired with tissue-specific eQTL association signals (applying this test to all pairs for which the lead SNPs were in LD with $r^2 > 0.7$) to estimate the probability that each pair shares a CCV. Because our Bangladeshi participants are not well-matched to the GTEx donors on ancestry (ie, GTEx is primarily comprised individuals of European ancestry), we provide LD estimates based on both South Asian and European populations from the 1000 Genomes Project. For this method, we used 2 sets of summary statistics: association results from our GWAS and results from the GTEx *cis*-eQTL analysis. For each DMA% association signal, we included results for all SNPs (including imputed SNPs) within 500 kb of the lead SNP. For each of the 2 DMA% association signals, our results are adjusted for the lead SNP representing the other DMA% association signal. For each corresponding eQTL, we used the same set of SNPs selected from our eQTL results. The summary statistics for each eQTL signal were adjusted for the lead SNP(s) representing additional nearby eQTLs for the same gene (based on our conditional eQTL analyses). The R coloc package only utilized information on SNPs present in both sets of summary statistics (Wallace, 2013; Wallace *et al.*, 2012).

Bayesian colocalization requires specifying a prior probability for a SNP being associated only with trait 1 (p_1), only with trait 2 (p_2), and with both traits (p_{12}). For our analysis, p_1 represents the probability of being associated only with DMA%, p_2 represents the probability of being only an eQTL, and p_{12} represents the probability of being associated with DMA% and an eQTL. For our study, we set the overall probability of being associated with DMA% ($p_1 + p_{12}$) as 10^{-5} and the overall probability of being an eQTL as 10^{-4} ($p_2 + p_{12}$), both are which are typical in studies of complex traits and eQTLs. The value of p_{12} was varied in our analysis to correspond to a 5%, 20%, and 50% probability that a causal variant for DMA% is also an eQTL. This led us to examine 3 values of p_{12}: 5×10^{-7}, 2×10^{-6}, and 5×10^{-6}. Thus, we covered a range of potential assumptions regarding the likelihood that SNPs associated with DMA% are also eQTLs. Through our colocalization analyses, we evaluated the posterior probability (PP) of colocalization (H4) as well as the PP of distinct causal variants underlying the association with DMA% and the identified eQTL (H3). In addition to these posteriors, we evaluated the PP that there is an identifiable causal variant underlying the association with DMA%, but no detectable eQTL signal (H2) as well as the PP that there is no detectable DMA% signal and a detectable eQTL (H1). Both H2 and H1 reflect the possibility that we do not have sufficient power for a robust colocalization analysis within a given tissue.

RESULTS

Identifying Cis-*eQTLs for Colocalization Analysis in GTEx Tissues*

Our prior GWAS of AME using data from HEALS yielded 2 independent association signals in the 10q24.32 region for DMA%. These signals are represented by lead SNPs rs4919690 and rs75075426. These 2 associations are in close proximity to the *AS3MT* (arsenic methyltransferase) gene (Figure 1).

Using GTEx data, we examined *AS3MT* expression across all available tissues and found that expression ranged from 411.5 transcripts per million (TPM) in the adrenal gland to 3.6 TPM in the esophageal mucosa (GTEx Portal, 2019). Median *AS3MT* expression for all GTEx tissues is presented in Supplementary Table 1 and Supplementary Figures 1 and 2. Excluding the adrenal gland, most tissues showed expression levels between 28.7 and 3.6 TPM with the liver showing relatively high expression at 25.1 TPM (GTEx Portal, 2019). We identified *cis*-eQTLs for the *AS3MT* gene across all available tissues. At a p value threshold of 5×10^{-8}, *cis*-eQTLs were identified in 38 tissue types. Multiple eQTLs were observed in 11 tissues with a maximum of 3 signals in a single tissue. Among these *cis*-eQTLs, we identified *cis*-eQTLs in 13 tissues with a lead eSNP in high LD ($r^2 > 0.7$) with rs4919690 (the lead SNPs in the 10q24.32 region identified in our GWAS of DMA%). We found rs4919690 to be the lead SNP for the *cis*-eQTL in 7 of these tissue types (Table 1). In contrast, none of the 48 observed eQTLs had a lead SNP in strong LD with the lead SNP of our secondary association signal for DMA% (rs75075426). This led us to focus exclusively on rs4919690 for the remainder of our analyses.

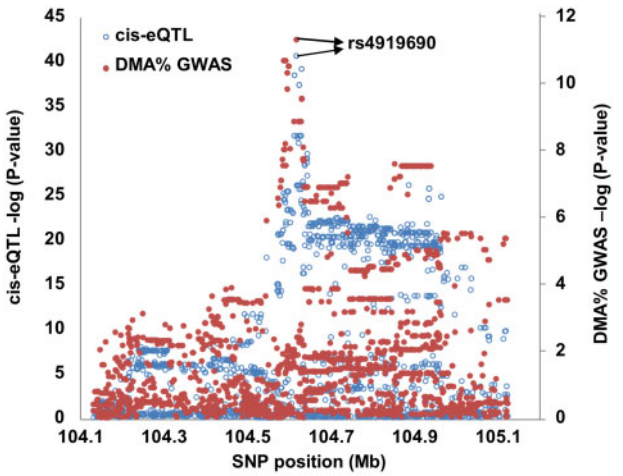

Figure 2. Colocalization of arsenic metabolism efficiency association signal (lead single nucleotide polymorphism [SNP] rs4919690) with a *cis*-eQTL (expression quantitative trait locus) for AS3MT in the aortic artery. Filled circles representing DMA% GWAS SNPs are plotted based on −log10 (p values), and empty circles representing *cis*-eQTLs are also plotted by −log10(p values). The posterior probability of colocalization for the *cis*-eQTL and GWAS association signal ranged from 0.997 to 1 depending on the prior probability (Table 1) revealing strong evidence for signal colocalization and suggesting that the causal SNP that impacts DMA% likely regulates AS3MT expression in the aorta.

Due to the presence of multiple genes in the10q24.32 region near our lead SNPs for DMA% (rs4919690) we extended our eQTL analyses beyond *AS3MT* to all genes within 500 kb of *AS3MT*. In total, across all available GTEx tissue types, we identified 18 genes with an eQTL ($p < 5\times10^{-8}$) in at least 1 tissue type. Among these genes, 4 had an eQTL with a lead SNP that was in high LD ($r^2 > 0.7$) with the DMA% lead SNP, rs4919690: *NT5C2*, *CIP17A1-AS1*, *RP11-724N1.1*, and *BORCS7*. These genes, their associated eQTLs, and their observed LD with our lead SNP are listed in Table 2.

Colocalization of Cis*-eQTLs for AS3MT and DMA% Association Signals*

For each eQTL in the 10q24.32 region identified above, we conducted colocalization analysis to assess the evidence that the eQTL shares a CCV with one of the association signals observed for DMA%. Under the assumption that 50% of DMA% SNPs are eQTLs ($p_{12} = 5\times10^{-6}$), we found evidence for colocalization between *cis*-eQTLs for *AS3MT* and rs4919690 in 13 tissue types at a PP of having a CCV >80% (Table 1). These included: subcutaneous and visceral omental adipose, the aorta and left ventricle of

Table 1. Colocalization of *AS3MT* Expression Quantitative Trait Loci (eQTLs) Identified in GTEx Tissues With the Primary Association Signal for DMA% (Lead Single Nucleotide Polymorphism [SNP] rs4919690) Identified in a Bangladeshi Population Across a Range of Prior Probabilities

Tissue	rsIDs of eQTLs	PP of Colocalization Assuming 5% of GWAS SNPs Are eQTLs	PP of Colocalization Assuming 20% of GWAS SNPs Are eQTLs	PP of Colocalization Assuming 50% of GWAS SNPs Are eQTLs	LD Between eQTL and Lead SNP in BEB	LD Between eQTL and Lead SNP in EUR
Subcutaneous adipose	rs4919690	0.219	0.575	0.848	1	1
Visceral omental adipose	rs12416687	0.989	0.998	0.999	0.60	0.96
Aorta	rs4919690	0.997	0.999	1	1	1
Coronary artery	rs12775431	0.988	0.997	0.999	0.60	0.96
Tibial artery	rs4919690	0.990	0.998	0.999	1	1
Brain: amygdala	rs12775431	0.466	0.965	0.991	0.60	0.96
Brain: hippocampus	rs4919690	0.887	0.979	0.993	1	1
Sigmoid colon	rs12775431	0.970	0.994	0.998	0.60	0.96
Transverse colon	rs9527	0.310	0.682	0.896	0.60	1
Heart: left ventricle	rs4919690	0.985	0.997	0.999	1	1
Lung	rs4919690	0.995	0.999	1	1	1
Tibial nerve	rs4919690	0.582	0.870	0.965	1	1
Vagina	rs11191421	0.970	0.994	0.998	0.60	0.94

Abbreviations: LD, linkage disequilibrium; PP, posterior probability.

the heart, the coronary and tibial arteries, the amygdala and hippocampus of the brain, the sigmoid and transverse colon, the lung, the tibial nerve, and the vagina. Across all tissues, *cis*-eQTLs for *AS3MT* in the aortic tissue showed the strongest PP of colocalization (colocalized association signals shown in Figure 2). In contrast, no evidence of colocalization was observed for *AS3MT* eQTLs in the liver, the primary tissue in which arsenic is believed to be metabolized.

Varying the prior probability (p_{12}) concerning the percentage of DMA% SNPs that are also eQTLs from 50% to 5% decreased the number of tissues in which colocalization was observed (PP of CCV > 80%) from 13 to 9. For subcutaneous adipose and the tibial nerve, decreasing p_{12} decreased the PP of CCV and primarily increased the PP of 2 distinct causal variants (Supplementary Table 2). In contrast, for brain (amygdala) and colon (transverse), decreasing p_{12} decreased the PP of CCV and primarily increased the PP of H2 (association for DMA% only, no eQTL), suggesting that the eQTL associations are too weak to conduct well-powered colocalization analyses (Supplementary Table 2).

The minor allele of the lead SNP identified in our previous GWA study of AME, rs4919690, is associated with lower metabolism efficiency (lower DMA%). This minor allele is also associated with low *AS3MT* expression across all tissues in which colocalization was identified (Figure 3). Thus, we observe consistency in the directionality of association as the allele associated with low *AS3MT* expression is the same allele associated with low AME.

Colocalization of Cis*-eQTLs for Genes Near* AS3MT *and DMA% Association Signals*

We further examined colocalization of genes near *AS3MT* (Table 2). *Cis*-eQTLs, for *BORCS7* in 43 tissue types (lead SNP rs11191421 in 29 tissues, rs4919690 in 12 tissues, rs9527 in 1 tissue, and rs3740392 in 1 tissue) were found to colocalize (PP >80%) with our primary DMA% association signal (lead SNP rs4919690). Of note, 1 of the 43 tissue types containing a colocalizing *cis*-eQTL for *BORCS7* was the liver (Figure 4). Evidence of colocalization (PP >80%) with the primary DMA% association signal was also observed for tissue-specific eQTLs impacting the following genes: *CYP17A1-AS1* (thyroid) and *RP11-724N1.1* (thyroid). Colocalization with *cis*-eQTLs for *NT5C2* was observed in the cerebellum (lead SNP rs12779263) and pancreas (lead SNP rs10883790). However, evidence of colocalization was generally not as strong for eQTLs for *CYP17A1-AS1*, *RP11-724N1.1*, and *NT5C2*, as compared with *AS3MT* and *BORCS7* (in terms of robustness of the PP to each of the 3 choices of prior, see Table 2).

Overall, the results of our colocalization analyses were fairly consistent across values of p_{12}, which can be observed in Table 2. In general, we observed lower probabilities of colocalization with lower LD between the lead DMA% SNP and lead *cis*-eQTL SNP in a given tissue, as expected. Varying the prior assumption (p_{12}) concerning the percentage of DMA% SNPs that are also eQTLs, also altered the observed colocalization (Supplementary Table 3). Like our analysis of *AS3MT*, we observed evidence for distinct causal variants (an increased PP of H3) as well as evidence that some eQTL signals were too weak to conduct well-powered colocalization analyses (increased PP of H2).

Correlation Between AS3MT and BORCS7 Expression

There is no prior evidence suggesting a role for *BORCS7* in arsenic metabolism; however, colocalization between the DMA% association signal (rs4919690) and *cis*-eQTLs for *BORCS7* was observed in 43 tissues including the liver. One potential explanation for this lies in the coregulation of *AS3MT* and *BORCS7*. Correlation between *BORCS7* expression and *AS3MT* expression was observed in nearly all tissues in which colocalization was found for both genes (Figure 5), as well as in the liver. This finding suggests that a CCV regulates expression of both genes (or 2 distinct causal regulatory variants in strong LD) and contributes to the observed coexpression of these genes. To determine whether this coexpression was entirely driven by genotype, we used a linear regression model to examine the association after adjusting for rs4919690. We found that for some tissues the coexpression of *BORCS7* and *AS3MT* appears entirely due to the SNP. However, we observe a correlation between the expression of these 2 genes after adjustment for this SNP in the aorta, amygdala, left ventricle of the heart, and the liver, suggesting

Table 2. Colocalization of Expression Quantitative Trait Loci (eQTLs) for Genes in the 10q24.32 Region (Excluding AS3MT) With the Primary Association Signal for DMA% (Lead Single Nucleotide Polymorphism [SNP] rs4919690) With Posterior Probability of Colocalization > 80% (for $p_{12} = 5 \times 10^{-6}$)

Gene	Top SNP (eQTL)	Tissue	PP if 5% of GWAS SNPs Are eQTLs ($p_{12} = 5 \times 10^{-7}$)	PP if 20% of GWAS SNPs Are eQTLs ($p_{12} = 2 \times 10^{-6}$)	PP if 50% of GWAS SNPs Are eQTLs ($p_{12} = 5 \times 10^{-6}$)	LD Between eQTL and rs4919690 in BEB	LD Between eQTL and rs4919690 in EUR
NT5C2	rs12779263	Cerebellum	0.778	0.944	0.986	0.298	0.707
	rs10883790	Pancreas	0.491	0.823	0.950	0.276	0.873
CYP17A1-AS1	rs11191401	Thyroid	0.891	0.975	0.994	0.496	0.708
RP11-724N1.1	rs11191401	Thyroid	0.324	0.697	0.903	0.496	0.708
BORCS7	rs11191421	Adrenal gland	0.981	0.996	0.999	0.604	0.974
		Coronary artery	0.886	0.974	0.994	0.604	0.974
		Tibial artery	0.992	0.998	1.000	0.604	0.974
		Brain amygdala	0.186	0.521	0.814	0.604	0.974
		Brain anterior cingulate cortex	0.881	0.973	0.993	0.604	0.974
		Brain caudate	0.984	0.997	0.999	0.604	0.974
		Brain cortex	0.972	0.994	0.999	0.604	0.974
		Cerebellar hemisphere	0.984	0.997	0.999	0.604	0.974
		Hypothalamus	0.933	0.985	0.996	0.604	0.974
		Hippocampus	0.621	0.886	0.969	0.604	0.974
		Brain putamen	0.962	0.992	0.998	0.604	0.974
		Brain spinal cord	0.886	0.974	0.993	0.604	0.974
		Brain substantia nigra	0.581	0.869	0.964	0.604	0.974
		Mammary	0.976	0.995	0.999	0.604	0.974
		Sigmoid colon	0.968	0.993	0.998	0.604	0.974
		Esophageal mucosa	0.934	0.985	0.996	0.604	0.974
		Esophageal muscularis	0.804	0.952	0.988	0.604	0.974
		Atrial appendage	0.979	0.996	0.999	0.604	0.974
		Left ventricle	0.994	0.999	1.000	0.604	0.974
		Liver	0.983	0.996	0.999	0.604	0.974
		Lung	0.978	0.995	0.999	0.604	0.974
		Minor salivary	0.870	0.970	0.992	0.604	0.974
		Tibial nerve	0.996	0.999	1.000	0.604	0.974
		Ovary	0.363	0.731	0.917	0.604	0.974
		Pituitary	0.966	0.993	0.998	0.604	0.974
		Nonsun-exposed skin	0.969	0.993	0.998	0.604	0.974
		Sun-exposed skin	0.710	0.922	0.980	0.604	0.974
		Small intestine	0.980	0.996	0.999	0.604	0.974
		Stomach	0.951	0.989	0.997	0.604	0.974
	rs4919690	Subcutaneous adipose	0.995	0.999	1.000	1	1
		Visceral omentum adipose	0.993	0.999	1.000	1	1
		Aorta	0.996	0.999	1.000	1	1
		Cerebellum	0.989	0.998	0.999	1	1
		Skeletal muscle	0.957	0.991	0.998		
		Transverse colon	0.996	0.999	1.000	1	1
		Gastroesophageal Junction	0.996	0.999	1.000	1	1
		Prostate	0.967	0.993	0.998	1	1
		Pancreas	0.908	0.979	0.995		
		Spleen	0.759	0.938	0.984	1	1
		Thyroid	0.997	0.999	1.000	1	1
		Uterus	0.274	0.643	0.879	1	1
	rs9527	Brain frontal cortex	0.983	0.997	0.999	0.6	1
	rs3740392	Brain nucleus accumbens	0.175	0.504	0.805	0.3	0.8

Abbreviations: LD, linkage disequilibrium; PP, posterior probability.

that factors other than genotype are contributing to the coregulation of these genes in these tissue types (adjusted regression results shown in Supplementary Figure 3). The factors that may contribute to the observed coexpression include the conformation of the local chromatin structure, and the sharing of local regulatory elements including transcription factors, enhancers, and promoters (Michalak, 2008). In considering these additional factors, we used an additional linear regression

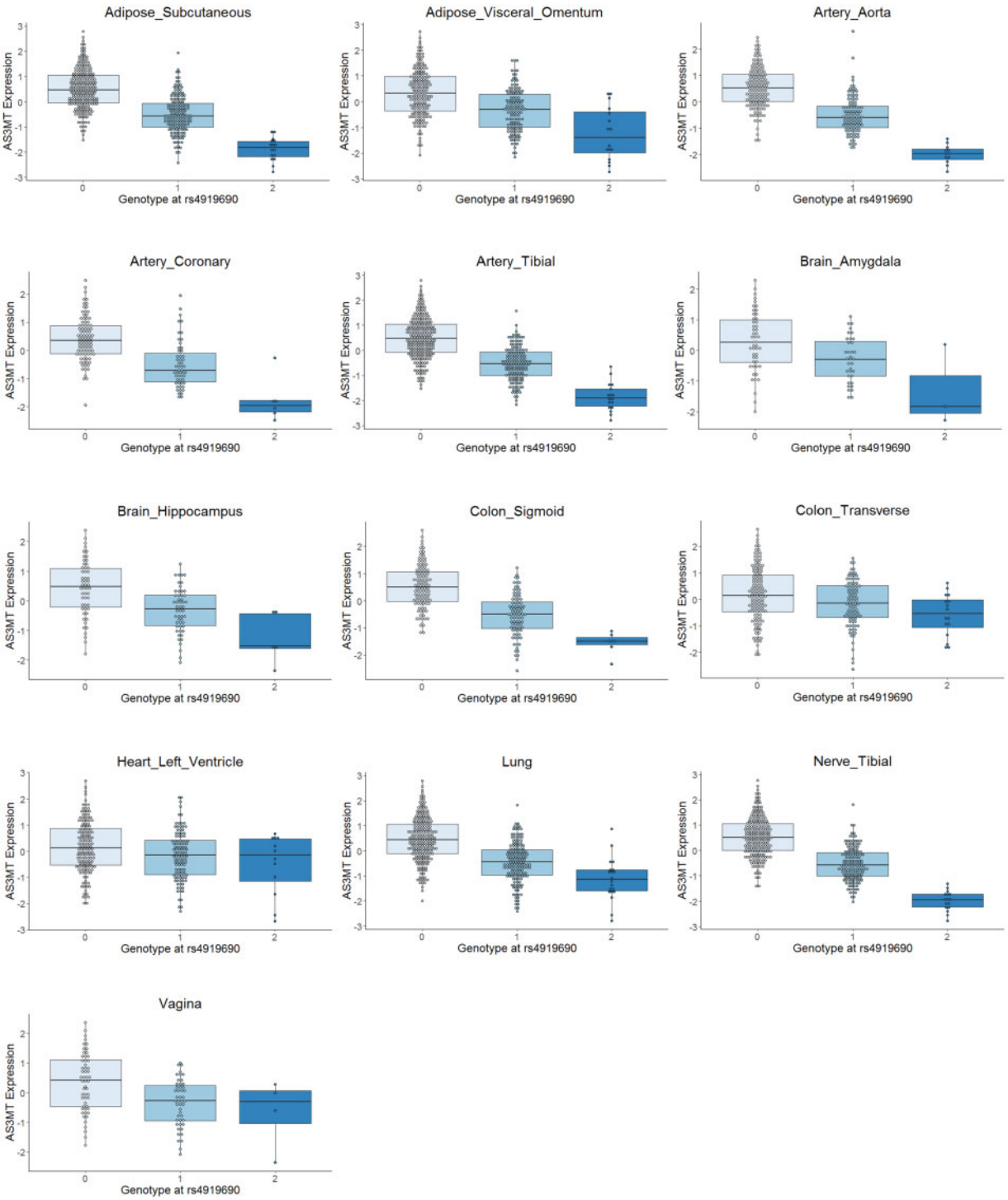

Figure 3. Expression of AS3MT across GTEx tissue types by genotype at lead single nucleotide polymorphism, rs4919690. The lightest shading represents AS3MT expression levels among individuals homozygous for the high efficiency allele (C). The darkest shading represents AS3MT expression levels among those individuals homozygous for the low efficiency allele (T). Here, we observe a clear pattern of lower AS3MT expression among those who have at least 1 copy of the low efficiency allele (T) across all tissues in which colocalization was observed.

model that adjusted for rs4919690 as well as sex, genotyping platform, genotyping principle components, and all tissue-specific PEER factors. The inclusion of PEER factors allowed us to control for confounding due to differences in cell-type composition that may impact our results. Here, we observe correlation between *AS3MT* and *BORCS7* expression after adjustment

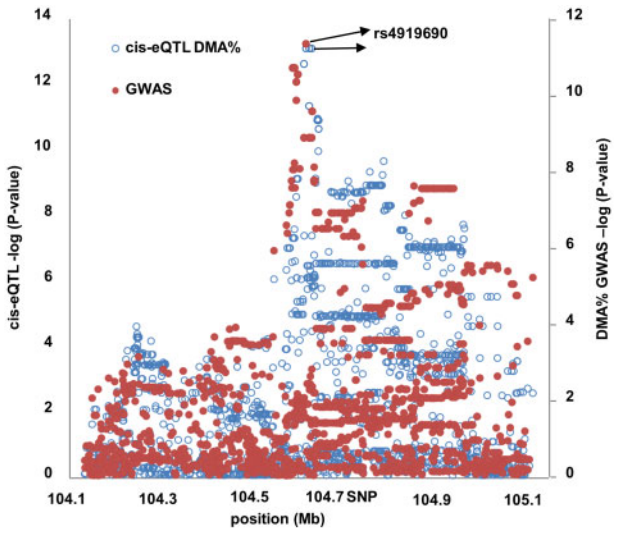

Figure 4. Colocalization of arsenic metabolism efficiency association signal (lead single nucleotide polymorphism [SNP] rs4919690) with a *cis*-eQTL (expression quantitative trait locus) for BORCS7 in the liver. Filled circles representing DMA% GWAS SNPs are plotted based on −log10(*p* values), and empty circles representing *cis*-eQTLs are also plotted by −log10(*p* values). The posterior probability of colocalization for the *cis*-eQTL and GWAS association signal ranged from 0.983 to 0.999 depending on the prior probability (Table 2) revealing strong evidence for signal colocalization and suggesting that the causal SNP that impacts DMA% regulates BORCS7 expression in the liver.

in the aorta, tibial artery, liver, and tibial nerve (Supplementary Figure 4).

DISCUSSION

In this study, our goal was to assess the evidence that genetic variants in the 10q24.32 region known to impact AME (as measured by DMA% in urine) exert their effects through regulation of *AS3MT* gene expression. Using tissue-specific eQTL results from GTEx, we conducted colocalization analyses focused on *AS3MT* and its surrounding genes to estimate the probability that a CCV affects both AME and expression of nearby genes. Our results provide evidence that the genetic variant impacting AME (represented by rs4919690) also affects expression of *AS3MT* (and other nearby genes) in 13 distinct tissue types, including several tissues relevant to arsenic toxicity: heart tissues (coronary artery, aorta, and left ventricle) and lung. In light of *AS3MT*'s central role in arsenic metabolism (Lin *et al.*, 2002), these results suggest that the allele that decreases AME does so by altering *AS3MT* activity (potentially in multiple tissues), thereby increasing the internal dose of arsenic in tissues and increasing susceptibility for arsenic-related diseases in individuals with low AME genotypes (Antonelli *et al.*, 2014; de la Rosa *et al.*, 2017). Epidemiologic studies of *AS3MT* SNPs have reported associations with risk of multiple arsenic-associated diseases (Antonelli *et al.*, 2014), including arsenic-induced skin lesions (Valenzuela *et al.*, 2009), carotid atherosclerosis (Hsieh *et al.*, 2011), coronary heart disease (Gong and O'Bryant, 2012), lung cancer (de la Rosa *et al.*, 2017), and bladder cancer (Beebe-Dimmer, 2012).

Arsenic was not measured in GTEx (and we expect most individuals to have very low exposure, with exposure likely depending on factors like smoking, source of drinking water, and occupational history), so we were unable to assess the impact of arsenic on *AS3MT* expression in GTEx. Studies examining *AS3MT* expression in human lymphocytes (Pierce *et al.*, 2013) and peripheral blood (Engström *et al.*, 2013), and in multiple brain regions of mice (Sánchez-Peña *et al.*, 2010) found no association between *AS3MT* expression and arsenic dose. One possible explanation for this lack of association may be a role for *AS3MT* outside arsenic metabolism, resulting in ubiquitous expression (as observed in GTEx) regardless of the presence of arsenic. Although additional roles for *AS3MT* have not been well described, strong associations between *AS3MT* variants and schizophrenia risk suggest a potential neurological role (D'Ambrosio *et al.*, 2019). In addition, *AS3MT* is coregulated with other genes in the 10q24.32 region, and these genes may serve important roles in a variety of tissues resulting in the expression of *AS3MT* in these tissue types.

Previous studies in arsenic-dosed mice found arsenic metabolites in blood, liver, lung, bladder, and kidney as well as in urine and that the levels of arsenicals were both tissue-specific and dose-dependent (Kenyon *et al.*, 2005, 2008) Arsenite methyltransferase activity, as measured by incubating mouse tissues with sodium arsenite and S-[methyl-3*H*]adenosyl-L-methionine and measuring MMA formation, has also been observed in multiple mouse tissues including liver, testis, kidney, and lung (Healy *et al.*, 1998). This observation combined with our results suggests that SNPs associated with AME may impact this metabolism in multiple tissues, and that this multi-tissue activity may collectively impact DMA% in urine. We note that despite the liver's central role in arsenic metabolism, we did not observe colocalization between our DMA% association signal and *cis*-eQTLs for *AS3MT* in the liver. One reason for this may be the generally low level *AS3MT* expression observed across GTEx tissues combined with relatively small sample sizes (226 in liver). This may result in low power and prevent colocalization from being detectable in some tissues.

The directionalities of associations in our GWA and *cis*-eQTL analyses provides further support for *AS3MT* regulation as a mechanism by which 10q24.32 SNPs impact AME. The allele associated with low *AS3MT* expression identified in our *cis*-eQTL analysis is the same allele associated with low DMA% and increased arsenic toxicity risk.

The association signal for DMA%, represented by rs4919690, was also found to colocalize with *cis*-eQTL for 4 additional genes, including *BORCS7* (Stelzer *et al.*, 2016; BORCS7-ASMT, 2020). Across these genes, we observed colocalization for *cis*-eQTLs in a variety of tissue types. Some observed *cis*-eQTLs appeared to be tissue specific, with colocalization observed in only a single tissue type, whereas other eQTLs influenced expression in many tissues. Colocalization between the DMA% association signal (rs4919690) and the *cis*-eQTLs for *BORCS7* was observed in 43 tissues including the liver. *BORCS7* (previously known as *c10orf32*) encodes a subunit of BORC, a multisubunit complex that recruits Arl8 to lysosomes and initiates a process that promotes microtubule-guided transport of lysosomes toward the cell periphery (Pu *et al.*, 2015). Studies of a truncated *BORCS7* transcript show that it is important for lysosomal transport in neurons and truncation results in impaired motor function (Snouwaert *et al.*, 2018). The lack of a functional BORCS7 in mice is perinatally lethal, revealing its importance for normal function; however, this lethality prevents studying *AS3MT* function in *BORCS7* knockout mice (Snouwaert *et al.*, 2018). A study exploring the mechanism underlying the association between arsenic exposure and increased diabetes risk found that arsenic may induce pancreatic beta cell apoptosis through activation of the lysosome-mitochondrial pathway (Pan *et al.*, 2016). Beyond

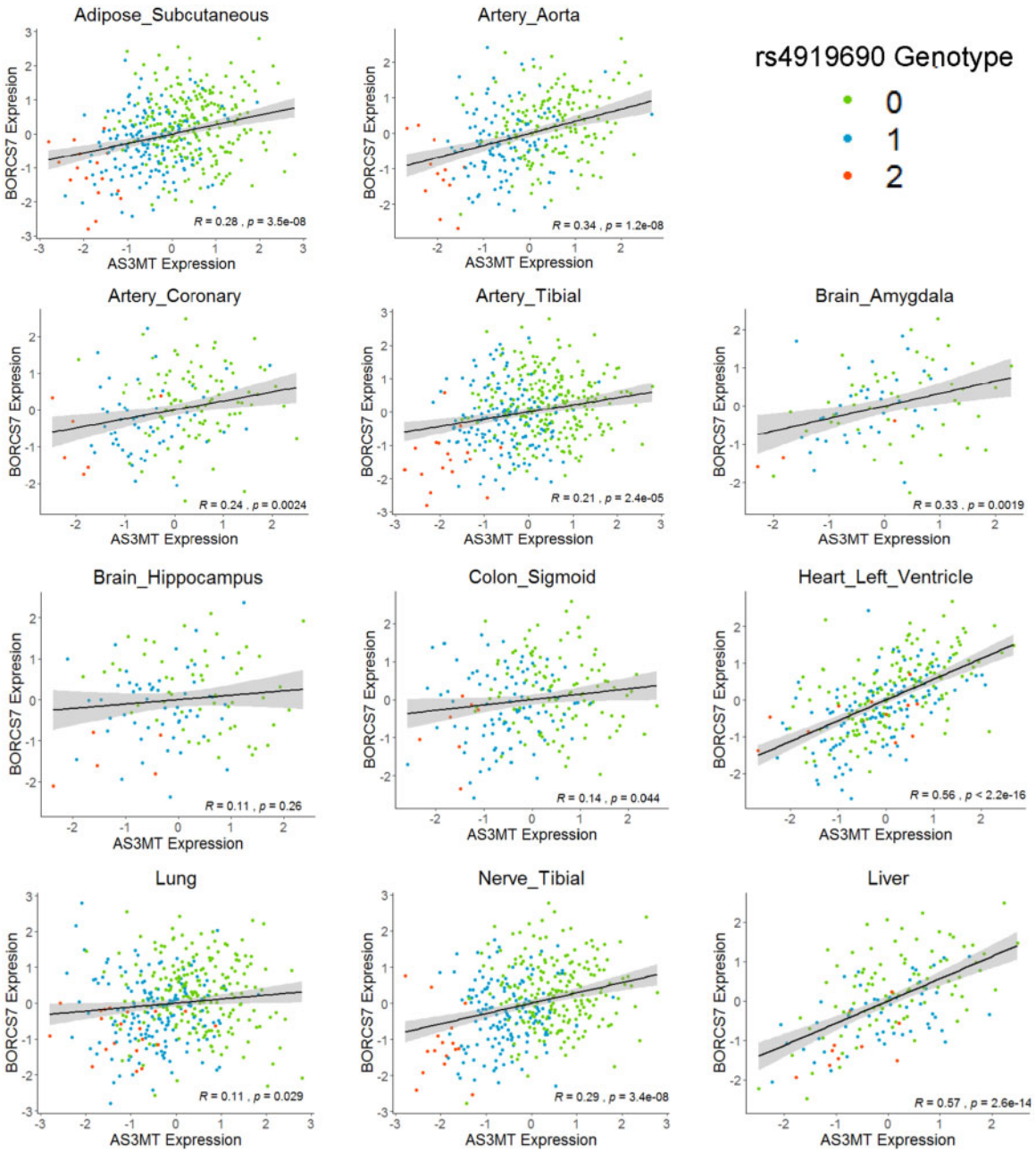

Figure 5. Correlation between expression of AS3MT and BORCS7 in tissues for which colocalization was observed for both genes (as well as liver). These plots show correlation between AS3MT and BORCS7 expression in nearly all tissues in which colocalization between lead single nucleotide polymorphism rs4919690 and expression of these genes was observed (with the exception of the Amygdala). In addition, we observe correlation between AS3MT and BORCS7 expression in the liver, despite the lack of colocalization for AS3MT in this tissue. The colors of the points reflect the genotype of each evaluated individual at rs4919690 and demonstrate that the low-efficiency allele corresponds to lower expression in both of the examined genes.

the potential that coregulation of *BORCS7* and *AS3MT* produces the observed colocalization, the connections among *BORCS7*, lysosomal function, and arsenic-induced activation of the lysosomal-mitochondrial pathway provides a potential alternate explanation for the colocalization. In addition, our previous work found an association between arsenic exposure (measured by total urinary arsenic) and *BORCS7* expression in lymphocytes (Pierce *et al.*, 2013).

The observed colocalization of a *BORCS7* eQTL and an AME association signal may also be due to the coregulation of *AS3MT* and *BORCS7* (Figure 5), genes separated by approximately 4 kb. Coexpression of *AS3MT* and *BORCS7* has been reported previously. A study of schizophrenia associated variants using GTEx tissues found as association between a risk SNP (rs7085104) and expression of a particular *AS3MT* isoform in multiple neuronal tissues; this association was independent of *BORCS7* expression, but the adjusted expression of the genes were weakly correlated (Li *et al.*, 2016). Similarly, another schizophrenia study using human brain tissue from the London Neurodegenerative Diseases Brain Bank found that heterozygosity at a schizophrenia risk variant (rs11191419) was associated with increased expression of *BORCS7* and *AS3MT* in

both fetal and adult brains (Duarte *et al.*, 2016). This evidence of the 2 genes being regulated by the same eQTL is similar to our own observations of *BORCS7* and *AS3MT* coregulation across GTEx tissue types. Finally, the 2 genes have been reported to be involved in a read-through transcription that produces *BORCS7-AS3MT* (Kent *et al.*, 2002).

The higher expression of *BORCS7* than *AS3MT* across GTEx tissues may have increased power to detect colocalization in *BORCS7* compared with *AS3MT*, despite the impact of the same *cis*-eQTLs on both genes. Providing further evidence of coregulation is the fact that the lead SNP for the colocalizing eQTL for *BORCS7* in the liver was also the lead SNP for the colocalizing eQTL for *AS3MT* in multiple tissues. We observed correlation between *BORCS7* and *AS3MT* expression in nearly all tissues in which colocalization was found in both genes. When we examined this correlation further, we found that coexpression of *BORCS7* and *AS3MT* appears entirely due to the SNP in some tissues, whereas in other tissues we found residual correlation suggesting that factors other than genotype contribute to coregulation in some tissue types.

Our secondary association signal for DMA% was not found to colocalize with *cis*-eQTLs for *AS3MT* or nearby genes, suggesting an alternate (ie, non-eQTL) mechanism for its impact on AME. The causal SNP underlying this secondary association signal may impact arsenic metabolism in multiple tissues, but not through regulation of local gene expression levels. One way in which this might occur is through the SNP's impact on messenger RNA conformation, stability, or translational efficiency (Battle *et al.*, 2015).

We observed that AS3MT expression was substantially higher in adrenal glands compared with other GTEx tissues, a finding consistent with prior studies (Gomez-Rubio *et al.*, 2009). However, GTEx samples consist of an entire adrenal gland cross section, so we cannot differentiate between expression in different zones of the adrenal gland. Protein expression studies however, have observed high levels of *AS3MT* in both the adrenal cortex and glomerular cells (Uhlen *et al.*, 2015; The Human Protein Atlas; Tissue Expression of AS3MT 2020). The high expression of *AS3MT* in the adrenal gland may be due to coregulation with the highly expressed nearby gene *CYP17A1*, which plays an important role in the formation of steroid hormones in the adrenal gland (Gilep *et al.*, 2011; Shu *et al.*, 2020; van den Akker *et al.*, 2002). It is also possible that *AS3MT* expression is high in the adrenal gland for protection against arsenic. Within the adrenal gland, arsenic can disrupt endocrine function, including repression of steroid-hormone regulated gene transcription and interference with hormone receptor binding (Sun *et al.*, 2016). These effects may impact cognitive development, learning, and memory by modifying the activity of the glucocorticoid receptor and blocking steroid binding (Sun *et al.*, 2016).

There are several limitations to this work. First, our DMA% association signals were identified in a population of Bangladeshi individuals, whereas the majority of GTEx donors are of European ancestry. This ancestry mismatch is likely to weaken evidence for colocalization in cases where a shared causal variant affects both DMA% and gene expression. Arsenic exposure is quite different within these 2 studies, with high levels of exposure in Bangladeshi individuals and very low to no arsenic exposure among GTEx participants. In addition, our analysis focused on gene expression and did not consider other cellular phenotypes that reflect gene regulation such as epigenetics, protein levels, and RNA isoforms, which may limit our ability to identify regulatory mechanisms underlying our observed associations. Finally, the sample sizes available for our eQTL analyses were not especially large, limiting our ability to detect weak eQTL effects. This would be particularly true if the effect of a 10q24.32 SNP on gene expression in the liver was weak.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

Genotype-Tissue Expression Project Acknowledgement

The Genotype Tissue Expression (GTEx) Project was supported by the Common Fund of the Director of the National Institutes of Health as well as by NCI, NHGRI, NIDA, NHLBI, NINDS, and NIMH. The data used for the analysis described in this publication were downloaded from the GTEx Portal on 10/11/2017.

FUNDING

Research reported in this publication was supported by National Institute of Environmental Health Sciences of the National Institutes of Health under award numbers R01ES023834 (to B.L.P), R35ES028379 (to B.L.P. and K.D.), R21ES024834 (to B.L.P. and M.A.), P42ES010349 (to J.H.G.), and P30 ES027792 (to H.A. and G.P.). Support for this research was also provided by the National Human Genome Research Institute of the National Institutes of Health under award number U01HG007601 (to B.L.P.), and by the National Cancer Institute of the National Institutes of Health under award number R01CA107431 (to H.A.).

Additional support for the research reported in this publication came from the National Institute of General Medical Sciences of the National Institutes of Health under award number T32M007281 and the Susan G. Komen Research Training Grant under award number GTDR16376189 (training support to M.C.). Fellowship support for Dr. Demanelis was provided by the National Institute on Aging (NIA) Specialized Demography and Economics of Aging Training Program under award number 2T32AG000243 at the University of Chicago.

This work used existing datasets, and most of the work associated with this publication was federally funded. Specifically, 80% ($40,000) was federally funded. Additionally, 20% ($10,000) was financed through nongovernmental sources.

The content of this work is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Agusa, T., Iwata, H., Fujihara, J., Kunito, T., Takeshita, H., Minh, T. B., Trang, P. T. K., Viet, P. H., and Tanabe, S. (2009). Genetic polymorphisms in AS3MT and arsenic metabolism in residents of the Red River Delta, Vietnam. *Toxicol. Appl. Pharmacol.* **236**, 131–141.

Ahsan, H., Chen, Y., Kibriya, M. G., Slavkovich, V., Parvez, F., Jasmine, F., Gamble, M. V., and Graziano, J. H. (2007). Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol. Biomarkers Prev.* **16**, 1270–1278.

Ahsan, H., Chen, Y., Parvez, F., Argos, M., Hussain, A. I., Momotaj, H., Levy, D., van Geen, A., Howe, G., Graziano, J., *et al.* (2006). Health Effects of Arsenic Longitudinal Study (HEALS): Description of a multidisciplinary epidemiologic investigation. *J. Expo. Sci. Environ. Epidemiol.* **16**, 191–205.

Ameer, S. S., Xu, Y. Y., Engström, K., Li, H., Tallving, P., Nermell, B., Boemo, A., Parada, L. A., Peñaloza, L. G., Concha, G., *et al.* (2016). Exposure to inorganic arsenic is associated with increased mitochondrial DNA copy number and longer telomere length in peripheral blood. *Front. Cell Dev. Biol.* **4**, 87.

Antonelli, R., Shao, K., Thomas, D. J., Sams, R., and Cowden, J. (2014). AS3MT, GSTO, and PNP polymorphisms: Impact on arsenic methylation and implications for disease susceptibility. *Environ. Res.* **132**, 156–167.

Balakrishnan, P., Vaidya, D., Franceschini, N., Voruganti, V. S., Gribble, M. O., Haack, K., Laston, S., Umans, J. G., Francesconi, K. A., Goessler, W., *et al.* (2017). Association of cardiometabolic genes with arsenic metabolism biomarkers in American Indian communities: The Strong Heart Family Study (SHFS). *Environ. Health Perspect.* **125**, 15–22.

Battle, A., Khan, Z., Wang, S. H., Mitrano, A., Ford, M. J., Pritchard, J. K., and Gilad, Y. (2015). Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667.

Beebe-Dimmer, J. L. (2012). Genetic variation in glutathione S-transferase omega-1, arsenic methyltransferase and methylene-tetrahydrofolate reductase, arsenic exposure and bladder cancer: A case-control study. *Environ. Health* **11**, 43.

BORCS7-ASMT Gene. (2020). Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=BORCS7-ASMT. Accessed June 1, 2020.

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., Compton, C. C., DeLuca, D. S., Peter-Demchok, J., Gelfand, E. T., *et al.* (2015). A Novel approach to high-quality postmortem tissue procurement: The GTEx project. *Biopreserv. Biobank.* **13**, 311–319.

Challenger, F. (1945). Biological methylation. *Chem. Rev.* **36**, 315–361.

Chen, B., Arnold, L. L., Cohen, S. M., Thomas, D. J., and Le, X. C. (2011). Mouse arsenic (+3 oxidation state) methyltransferase genotype affects metabolism and tissue dosimetry of arsenicals after arsenite administration in drinking water. *Toxicol. Sci.* **124**, 320–326.

D'Ambrosio, E., Dahoun, T., Pardiñas, A. F., Veronese, M., Bloomfield, M. A., Jauhar, S., Bonoldi, I., Rogdaki, M., Froudist-Walsh, S., Walters, J. T., .. (2019). The effect of a genetic variant at the schizophrenia associated AS3MT/BORCS7 locus on striatal dopamine function: A PET imaging study. *Psychiatry Res. Neuroimaging* **291**, 34–41.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., *et al.* (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287.

de la Rosa, R., Steinmaus, C., Akers, N. K., Conde, L., Ferreccio, C., Kalman, D., Zhang, K. R., Skibola, C. F., Smith, A. H., Zhang, L., .. (2017). Associations between arsenic (+3 oxidation state) methyltransferase (AS3MT) and N-6 adenine-specific DNA methyltransferase 1 (N6AMT1) polymorphisms, arsenic metabolism, and cancer risk in a chilean population. *Environ. Mol. Mutagen.* **58**, 411–422.

De Loma, J., Tirado, N., Ascui, F., Levi, M., Vahter, M., Broberg, K., and Gardon, J. (2019). Elevated arsenic exposure and efficient arsenic metabolism in indigenous women around Lake Poopó, Bolivia. *Sci. Total Environ.* **657**, 179–186.

Douillet, C., Huang, M. C., Saunders, R. J., Dover, E. N., Zhang, C., and Stýblo, M. (2017). Knockout of arsenic (+3 oxidation state) methyltransferase is associated with adverse metabolic phenotype in mice: The role of sex and arsenic exposure. *Arch. Toxicol.* **91**, 2617–2627.

Drobna, Z., Naranmandura, H., Kubachka, K. M., Edwards, B. C., Herbin-Davis, K., Styblo, M., Le, X. C., Creed, J. T., Maeda, N., Hughes, M. F., *et al.* (2009). Disruption of the arsenic (+3 oxidation state) methyltransferase gene in the mouse alters the phenotype for methylation of arsenic and affects distribution and retention of orally administered arsenate. *Chem. Res. Toxicol.* **22**, 1713–1720.

Duarte, R. R. R., Troakes, C., Nolan, M., Srivastava, D. P., Murray, R. M., and Bray, N. J. (2016). Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered *cis*-regulation of BORCS7, AS3MT, and NT5C2 in the human brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **171**, 806–814.

Engström, K. S., Broberg, K., Concha, G., Nermell, B., Warholm, M., and Vahter, M. (2007). Genetic polymorphisms influencing arsenic metabolism: Evidence from Argentina. *Environ. Health Perspect.* **115**, 599–605.

Engström, K. S., Hossain, M. B., Lauss, M., Ahmed, S., Raqib, R., Vahter, M., and Broberg, K. (2013). Efficient arsenic metabolism—The AS3MT haplotype is associated with DNA methylation and expression of multiple genes around AS3MT. *PLoS One* **8**, e53732.

Gao, J., Tong, L., Argos, M., Bryan, M. S., Ahmed, A., Rakibuz-Zaman, M., Kibriya, M. G., Jasmine, F., Slavkovich, V., Graziano, J. H., *et al.* (2015). The genetic architecture of arsenic metabolism efficiency: A SNP-based heritability study of Bangladeshi adults. *Environ. Health Perspect.* **123**, 985–992.

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383.

Gilep, A. A., Sushko, T. A., and Usanov, S. A. (2011). At the crossroads of steroid hormone biosynthesis: The role, substrate specificity and evolutionary development of CYP17. *Biochim. Biophys. Acta* **1814**, 200–209.

Gomez-Rubio, P., Meza-Montenegro, M. M., Cantu-Soto, E., and Klimecki, W. T. (2009). Genetic association between intronic variants in AS3MT and arsenic methylation efficiency is focused on a large linkage disequilibrium cluster in chromosome 10. *J. Appl. Toxicol.* **30**, 260–270.

Gong, G., and O'Bryant, S. E. (2012). Low-level arsenic exposure, AS3MT gene polymorphism and cardiovascular diseases in rural Texas counties. *Environ. Res.* **113**, 52–57.

GTEx Consortium. (2017). Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213.

GTEx Portal. (2019). Available at: https://gtexportal.org/home/.

Healy, S. M., Casarez, E. A., Ayala-Fierro, F., and Aposhian, H. V. (1998). Enzymatic methylation of arsenic compounds: V. Arsenite methyltransferase activity in tissues of mice. *Toxicol. Appl. Pharmacol.* **148**, 65–70.

Hong, Y.-S., Song, K.-H., and Chung, J.-Y. (2014). Health effects of chronic arsenic exposure. *J. Prev. Med. Pub. Health* **47**, 245–252.

Hopenhayn-Rich, C., Biggs, M. L., and Smith, A. H. (1998). Lung and kidney cancer mortality associated with arsenic in

drinking water in Córdoba, Argentina. *Int. J. Epidemiol.* **27**, 561–569.

Hsieh, Y.-C., Lien, L.-M., Chung, W.-T., Hsieh, F.-I., Hsieh, P.-F., Wu, M.-M., Tseng, H.-P., Chiou, H.-Y., and Chen, C.-J. (2011). Significantly increased risk of carotid atherosclerosis with arsenic exposure and polymorphisms in arsenic metabolism genes. *Environ. Res.* **111**, 804–810.

Hughes, M. F., Edwards, B. C., Herbin-Davis, K. M., Saunders, J., Styblo, M., and Thomas, D. J. (2010). Arsenic (+3 oxidation state) methyltransferase genotype affects steady-state distribution and clearance of arsenic in arsenate-treated mice. *Toxicol. Appl. Pharmacol.* **249**, 217–223.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, A. D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.

Kenyon, E. M., Del Razo, L. M., and Hughes, M. F. (2005). Tissue distribution and urinary excretion of inorganic arsenic and its methylated metabolites in mice following acute oral administration of arsenate. *Toxicol. Sci.* **85**, 468–475.

Kenyon, E. M., Hughes, M. F., Adair, B. M., Highfill, J. H., Crecelius, E. A., Clewell, H. J., and Yager, J. W. (2008). Tissue distribution and urinary excretion of inorganic arsenic and its methylated metabolites in C57BL6 mice following subchronic exposure to arsenate in drinking water. *Toxicol. Appl. Pharmacol.* **232**, 448–455.

Li, M., Jaffe, A. E., Straub, R. E., Tao, R., Shin, J. H., Wang, Y., Chen, Q., Li, C., Jia, Y., Ohi, K., *et al.* (2016). A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nat. Med.* **22**, 649–656.

Lin, S., Shi, Q., Nix, F. B., Styblo, M., Beck, M. A., Herbin-Davis, K. M., Hall, L. L., Simeonsson, J. B., and Thomas, D. J. (2002). A NovelS-adenosyl-L-methionine: arsenic(III) methyltransferase from rat liver cytosol. *J. Biol. Chem.* **277**, 10795–10803.

Lindberg, A.-L., Rahman, M., Persson, L.-Å., and Vahter, M. (2008). The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. *Toxicol. Appl. Pharmacol.* **230**, 9–16.

Marafante, E., and Vahter, M. (1984). The effect of methyltransferase inhibition on the metabolism of [74As]arsenite in mice and rabbits. *Chem. Biol. Interact.* **50**, 49–57.

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

Michalak, P. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* **91**, 243–248.

Nixon, D. E., Mussmann, G. V., Eckdahl, S. J., and Moyer, T. P. (1991). Total arsenic in urine: Palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin. Chem.* **37**, 1575–1579.

Pan, X., Jiang, L., Zhong, L., Geng, C., Jia, L., Liu, S., Guan, H., Yang, G., Yao, X., Piao, F., *et al.* (2016). Arsenic induces apoptosis by the lysosomal-mitochondrial pathway in INS-1 cells. *Environ. Toxicol.* **31**, 133–141.

Pierce, B. L., Kibriya, M. G., Tong, L., Jasmine, F., Argos, M., Roy, S., Paul-Brutus, R., Rahaman, R., Rakibuz-Zaman, M., Parvez, F., *et al.* (2012). Genome-wide association study identifies chromosome 10q24.32 variants associated with arsenic metabolism and toxicity phenotypes in Bangladesh. *PLoS Genet.* **8**, e1002522.

Pierce, B. L., Tong, L., Argos, M., Gao, J., Jasmine, F., Roy, S., Paul-Brutus, R., Rahaman, R., Rakibuz-Zaman, M., Parvez, F., *et al.* (2013). Arsenic metabolism efficiency has a causal role in arsenic toxicity: Mendelian randomization and gene-environment interaction. *Int. J. Epidemiol.* **42**, 1862–1872.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., and Willer, C. J. (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337.

Pu, J., Schindler, C., Jia, R., Jarnik, M., Backlund, P., and Bonifacino, J. (2015). BORC, a Multisubunit Complex that Regulates Lysosome Positioning. *Dev. Cell* **33**, 176–188.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., *et al.* (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Rahman, M. M., Ng, J. C., and Naidu, R. (2009). Chronic exposure of arsenic via drinking water and its adverse health impacts on humans. *Environ. Geochem. Health* **31**, 189–200.

Ravenscroft, P., Brammer, H., and Richards, K. (2009). *Arsenic Pollution: A global synthesis*. *Wiley-Blackwell*, Chichester, UK; Malden, MA.

Rehman, K., and Naranmandura, H. (2012). Arsenic metabolism and thioarsenicals. *Metallomics* **4**, 881–892.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.

Sánchez-Peña, L. C., Petrosyan, P., Morales, M., González, N. B., Gutiérrez-Ospina, G., Del Razo, L. M., and Gonsebatt, M. E. (2010). Arsenic species, AS3MT amount, and AS3MT gene expression in different brain regions of mouse exposed to arsenite. *Environ. Res.* **110**, 428–434.

Shabalin, A. A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358.

Shu, T., Zhai, G., Pradhan, A., Olsson, P.-E., and Yin, Z. (2020). Zebrafish cyp17a1 knockout reveals that androgen-mediated signaling is important for male brain sex differentiation. *Gen. Comp. Endocrinol.* **295**, 113490.

Snouwaert, J. N., Church, R. J., Jania, L., Nguyen, M., Wheeler, M. L., Saintsing, A., Mieczkowski, P., Manuel de Villena, F. P., Armao, D., Moy, S. S., *et al.* (2018). A mutation in the Borcs7 subunit of the lysosome regulatory BORC complex results in motor deficits and dystrophic axonopathy in mice. *Cell Rep.* **24**, 1254–1265.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput. Biol* **6**, e1000770.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T. I., Nudel, R., Lieder, I., Mazor, Y., .. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* **54**, 54:1.30.1–1.30.33.

Sun, H.-J., Xiang, P., Luo, J., Hong, H., Lin, H., Li, H.-B., and Ma, L. Q. (2016). Mechanisms of arsenic disruption on gonadal, adrenal and thyroid endocrine systems in humans: A review. *Environ. Int.* **95**, 61–68.

The Human Protein Atlas; Tissue Expression of AS3MT. (version 19.1; 2019). *Staining in Adrenal Gland—The Human Protein Atlas*. Available at: https://www.proteinatlas.org/ENSG00000214435-AS3MT/tissue/adrenal+gland. Accessed June 1, 2020.

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., .. (2015). Tissue-based map of the human proteome. *Science* **347**, 1260419–1260419.

Vahter, M. (1999). Variation in human metabolism of arsenic. In *Arsenic Exposure and Health Effects III: Proceedings of the Third International Conference on Arsenic Exposure and Health Effects, July 12-15, 1998, San Diego, California*, pp. 267–279. Elsevier Science, Amsterdam, Holland.

Vahter, M. (2002). Mechanisms of arsenic biotransformation. *Toxicology* **181–182**, 211–217.

Valenzuela, O. L., Drobná, Z., Hernández-Castellanos, E., Sánchez-Peña, L. C., García-Vargas, G. G., Borja-Aburto, V. H., Stýblo, M., and Del Razo, L. M. (2009). Association of AS3MT polymorphisms and the risk of premalignant arsenic skin lesions. *Toxicol. Appl. Pharmacol.* **239**, 200–207.

van den Akker, E. L. T., Koper, J. W., Boehmer, A. L. M., Themmen, A. P. N., Verhoef-Post, M., Timmerman, M. A., Otten, B. J., Drop, S. L. S., and De Jong, F. H. (2002). Differential inhibition of 17α-hydroxylase and 17,20-lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. *J. Clin. Endocrinol. Metab.* **87**, 5714–5721.

Wallace, C. (2013). Statistical Testing of Shared Genetic Control for Potentially Related Traits. *Genet. Epidemiol.* **37**, 802–813.

Wallace, C., Rotival, M., Cooper, J. D., Rice, C. M., Yang, J. H. M., McNeill, M., Smyth, D. J., Niblett, D., Cambien, F., Tiret, L., *et al.* (2012). Statistical colocalization of monocyte gene expression and genetic risk variants for type 1 diabetes. *Hum. Mol. Genet.* **21**, 2815–2824.

WHO. (2018). *Arsenic*. Available at: https://www.who.int/news-room/fact-sheets/detail/arsenic. Accessed June 1, 2020.

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

Yokohira, M., Arnold, L. L., Pennington, K. L., Suzuki, S., Kakiuchi-Kiyota, S., Herbin-Davis, K., Thomas, D. J., and Cohen, S. M. (2010). Severe systemic toxicity and urinary bladder cytotoxicity and regenerative hyperplasia induced by arsenite in arsenic (+3 oxidation state) methyltransferase knockout mice. A preliminary report. *Toxicol. Appl. Pharmacol.* **246**, 1–7.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824.