Alternative Splicing of the Human AS3MT Gene Based on Number of Variable Nucleotide Tandem Repeats

By
Pragnya Dontu

Senior Honors Thesis
Department of Chemistry
University of North Carolina at Chapel Hill

April 26, 2019

Approved:

Abstract

Arsenic is an environmental toxin that is metabolized in humans by reduction and methylation of inorganic arsenic. Arsenite methyltransferase, encoded by AS3MT, methylates trivalent arsenic so that it can be excreted in urine. Genetic studies indicate that variants at the AS3MT locus confer risk for the development of schizophrenia. Alternate splicing of this gene leads to the production of an alternate isoform, AS3MTd2d3, predominately in the brain, and this alternative isoform is present at higher levels in tissue from individuals with schizophrenia. AS3MTd2d3, lacks arsenite methyltransferase activity, and its formation is linked to variable nucleotide tandem repeats (VNTRs) in exon 1 of AS3MT. It is hypothesized that VNTR copy number determines the relative production of the two transcripts by altering DNA methylation at the locus, and this in turn alters risk for schizophrenia. This hypothesis is tested using embryonic stem (ES) cell lines humanized with either the common AS3MT allele, an allele with three VNTRs in exon 1, or humanized with an allele we generated that carries four VNTRs. Expression of the two AS3MT isoforms is compared in ES cell lines that carry either the triple or quadruple VNTR human DNA. While no correlation between VNTR number and ratio of the AS3MTd2d3 isoform exists in ES cells, the ES cells will be differentiated into neuronal cells for further analysis to confirm or deny the original hypothesis.

Introduction

Arsenic is pervasive in the environment and exists in many different forms and oxidation states – - III (arsine), 0 (arsenic), + III (arsenite), or + V (arsenate) (Figure 1).[1] Organic arsenic, which contains carbon, is found in fish and shellfish, while inorganic arsenic is in soil, sediment and groundwater. Inorganic arsenic can be naturally occurring or may enter the environment as a

result of industries such as mining, fracking, and ore smelting. It has been historically used in pesticides, paint, and wood preservatives, but due to its toxicity, use of inorganic arsenic today is restricted. Exposure to this environmental toxin in large doses may lead to nausea, vomiting, diarrhea, or shock, or more chronic health effects such as diabetes, high blood pressure, or hepatotoxicity. Inorganic arsenic has also been found to be a carcinogen in humans.[2]

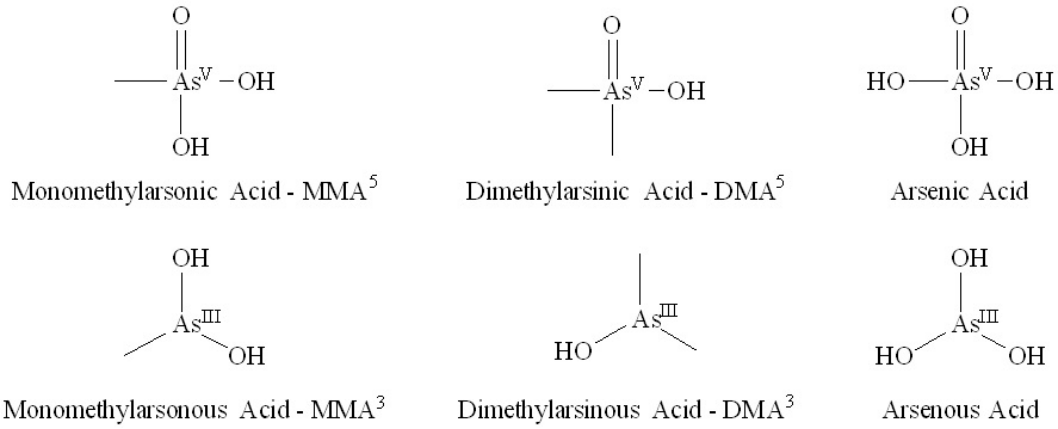

Figure 1. Structures of various forms of arsenic that are relevant in arsenic metabolism.

The mechanism of metabolism of arsenic in the body is a reduction of pentavalent arsenic to trivalent arsenic, then an oxidative methylation to pentavalent organic arsenic,[1] as shown by the following equation:

$$As^{V}O(OH)_3 + 2e^- \longrightarrow As^{III}(OH)_3 + CH_3^+ \longrightarrow CH_3As^{V}O(OH)_2 + 2e^- \longrightarrow CH_3As^{III}(OH)_3 + CH_3^+$$
$$\longrightarrow (CH_3)_2As^{V}O(OH)_2 + 2e^- \longrightarrow (CH_3)_2As^{III}(OH)_3 + CH_3^+ \longrightarrow (CH_3)_3As^{V}O$$

Monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA) are the methylated forms of inorganic arsenic that are excreted in urine.[1]

Trivalent forms of arsenic have shown a more potent toxicity than the pentavalent state.[1] The exact mechanisms of action are not known, but it is thought that the pentavalent state may

replace phosphate in reactions,[3] as arsenic is a chemical analog of phosphorous, and the trivalent state may react with thiols in functional groups.[1] Arsenate can react with glucose and gluconate and replace phosphate in processes such as the red blood cell anion exchange transport system.[4] During enzymatic reactions that create ATP, there has been diminished ATP formation in vitro due to the substitution of arsenate for phosphate.[1] Trivalent arsenite also plays a role in decreased production of ATP, as it binds and inhibits pyruvate dehydrogenase (PDH).[5] PDH normally oxidizes pyruvate to form acetyl CoA, which is a precursor for the citric acid cycle, but with less of the citric acid cycle intermediates, there is less ATP production and lower levels of gluconeogenesis.[1] Methylated trivalent arsenic interestingly has been shown to have a greater acute toxicity than arsenic; human cells are more sensitive to MMA^{III} than arsenite,[6] which may indicate another function for methylated arsenic apart from a detoxification mechanism.

While arsenic can be toxic, it has paradoxically been used as a medical treatment option in the past. In more recent years, this use of arsenic has become prevalent again, as it used for patients with acute promyelocytic leukemia (APL). This acute bone marrow cancer involves the accumulation of immature white blood cells, promyelocytes, in bone marrow. Cells are arrested in maturation at the promyelocyte state of myeloid cell development but arsenic trioxide (ATO) (Figure 2) can induce nonterminal cytodifferentiation of these cells.[7]

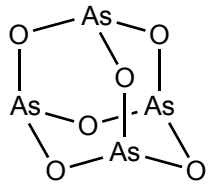

Figure 2. Structure of arsenic trioxide.

APL patients have a chromosomal translocation t(15;17)(q22;q21) that causes the retinoic acid receptor (RAR) gene on chromosome 17 to fuse to the promyelocytic leukemia (PML) gene on chromosome 15. This abnormal fusion yields PML-RARα fusion proteins. These proteins can

form a complex with PML and disrupts nuclear bodies that play a key role in apoptosis, epigenetic control of chromatin, transcriptional expression, and modulation of certain nuclear proteins.[8] ATO shows beneficial results in APL but not for other acute leukemia subtypes because it specifically binds to the cysteine residues in PML-RARα fusion proteins, promoting degradation (Figure 3).[7] Combined with all-*trans* retinoic acid (ATRA) as a treatment option, there is terminal differentiation of APL cells and clearance of the PML-RARα transcripts, bringing the five-year overall survival rate to 91.7%.[8] Figure 4 depicts the efficiency of the different treatment options for APL, including ATO.

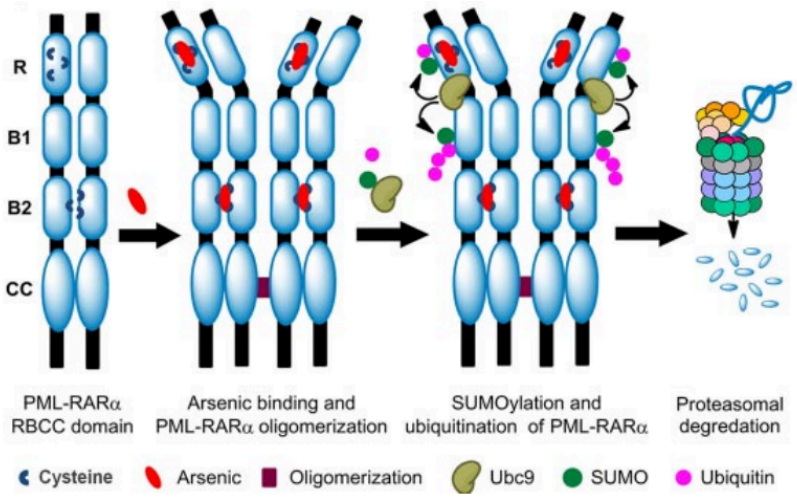

Figure 3. From Chen et al.[8] Representation of arsenic's interaction with PML-RARα fusion proteins.

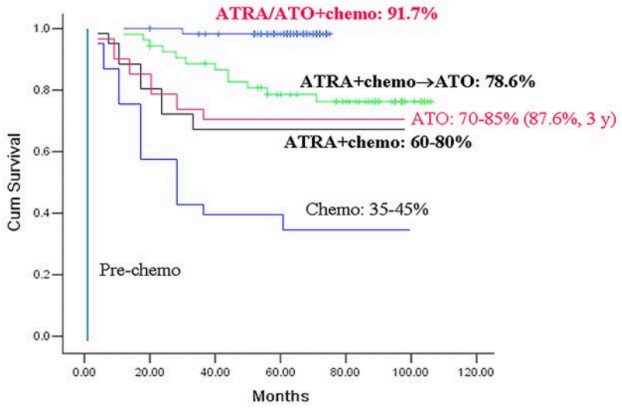

Figure 4. From Chen et al.[8] Treatment survival outcomes over a five-year period for patients with APL. The most efficient form of treatment is a combination of ATRA, ATO and chemo.

Apart from its effect in APL, ATO's ability to recognize PML proteins has been further studied in different pathways and seems to have more interactions than previously thought. For example, it inhibits the Notch pathway in gliomas[9] and the Hedgehog pathway[10], and may also help eliminate leukemia-initiating cells (Figure 5).[8]

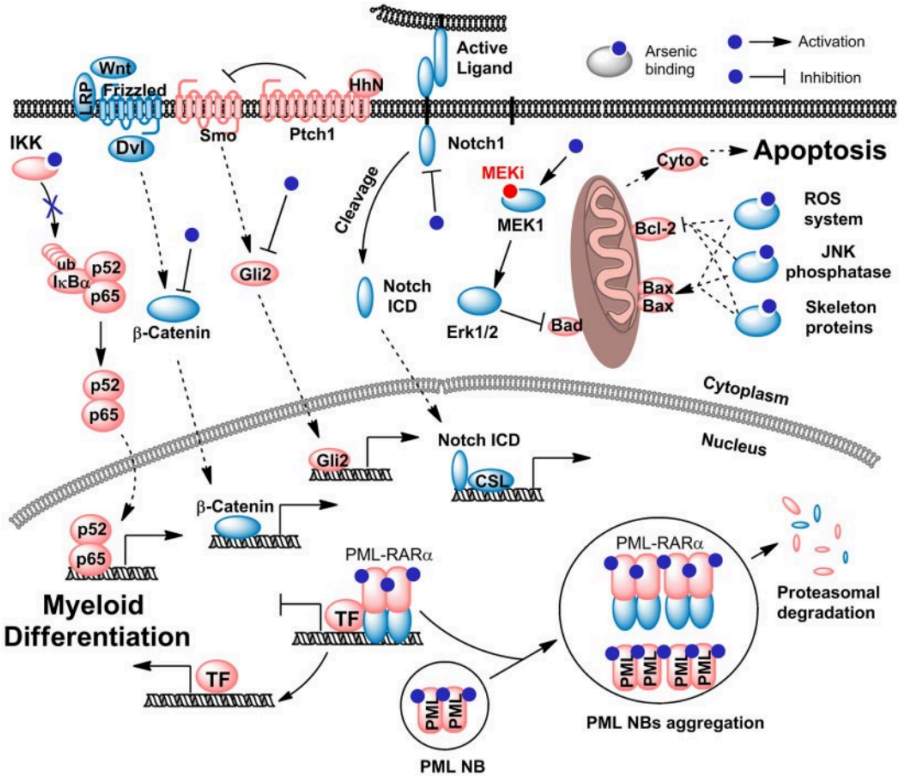

Figure 5. From Chen et al.[8] Targeting of arsenic in critical pathways in leukemia-initiating cells.

With the multiple, contradictory roles arsenic can have in the human body, it is informative to focus on its metabolism and the AS3MT gene. The human AS3MT gene confers expression of the arsenite methyltransferase protein. Arsenite methyltransferase catalyzes the methylation of trivalent arsenicals in the body to MMA^{V} by transferring methyl groups from s-adenosyl methionine to $As^{III}(OH)_3$ (Figure 6).[11]

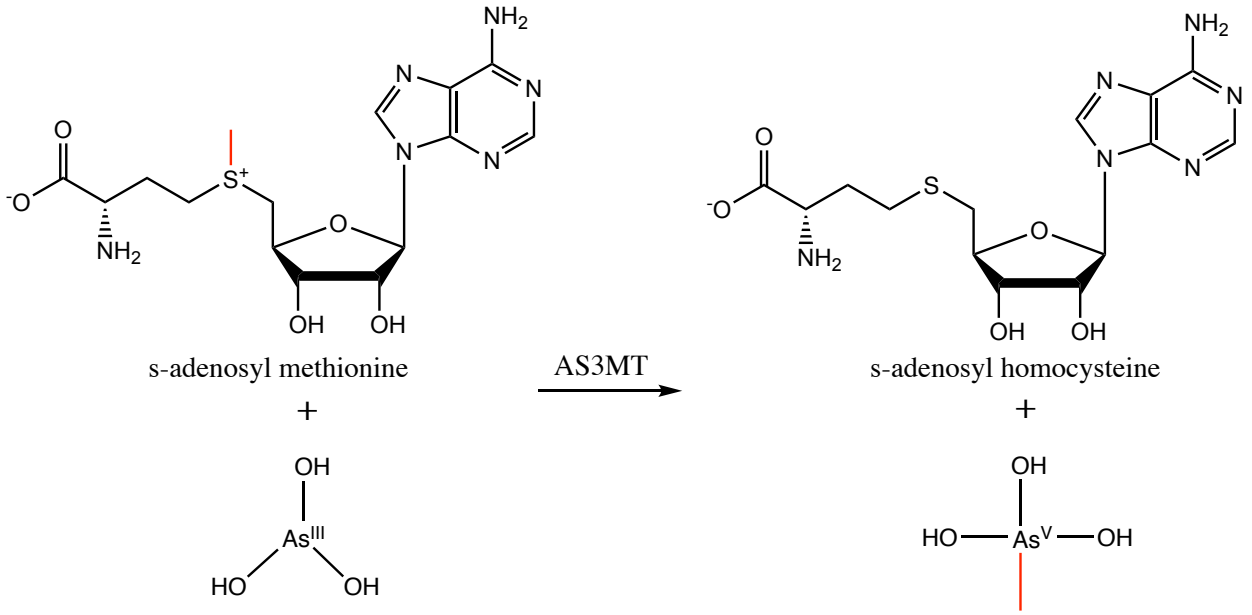

Figure 6. Reaction driven by AS3MT involving the transfer of a methyl group from s-adenosyl methionine to trivalent arsenous acid.

During treatment of APL, varying genotypes of AS3MT small nucleotide polymorphisms (SNPs) confer different side effects to the ATO. AS3MT genotypes with a higher efficiency to methylate arsenic have lower arsenic toxicity related side effects, and alterations in gene expression are based on the location of SNPs in the gene.[12] The non-mutated forms of this protein are on average found in equal amounts in the brain and somatic tissues, but are more highly expressed in the adrenal glands, cerebrospinal fluid, liver, heart, and testes.[13] Alternate splicing leads to the production of an alternate isoform of AS3MT predominantly in the brain. Genetic studies[13] indicate that the variant isoform AS3MTd2d3 confers risk for the development of schizophrenia. AS3MTd2d3 is present at higher levels in tissue from individuals with schizophrenia and lacks arsenite methyltransferase activity, which may imply a link of schizophrenia with arsenite accumulation. The lack of arsenite methyltransferase activity but the prevalence of this isoform may also suggest an alternate function of the protein. AS3MTd2d3 formation is linked to variable nucleotide tandem repeats (VNTRs) in the 5' untranslated region in the first exon of AS3MT.[13]

The number of VNTRs causes a variation in transcription of the gene,[14] and in this particular case, an increased number of repeats increases transcriptional activity of the AS3MT promoter. Each repeat is 36 base pairs and there can be two to four repeats (Figure 7). There are 11 exons in the full AS3MT gene, and the isoform AS3MTd2d3 is missing exons 2 and 3 (Figure 8). The presence of VNTRs has been appreciably associated with d2d3 mRNA expression, and the copy number shows an additive effect on d2d3 expression with an increased number of repeats. The d2d3 expression is significantly higher in individuals with schizophrenia and/or major depression compared to healthy subjects.[13] It is hypothesized that the relative production of d2d3 to full transcripts is altered by DNA methylation at the locus.

Figure 7. From Li et al.[13] Sequence of the VNTR repeats in exon 1 of AS3MT.

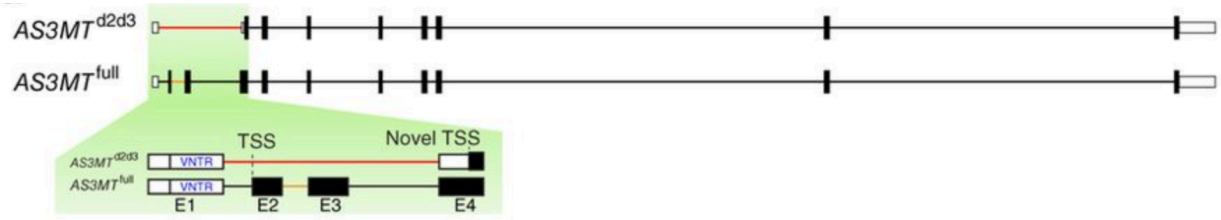

Figure 8. From Li et al.[13] Schematic depicting the variation of exons in the AS3MTfull and AS3MTd2d3 transcripts.

We test the effects of increased copy number of VNTRs in the AS3MT gene using mouse embryonic stem (ES) cell lines humanized with either the common AS3MT allele, which has 3 VNTRs in exon 1, or humanized with an allele we generated that carries 4 VNTRs. ES cells are used because of their high homologous recombineering ability and their potential to be differentiated into different cell types. Using ES cell lines also provides us with the ability to inject the cells into mouse embryos and develop chimera mice. We take advantage of

homologous recombination, which occurs when similar sequences are exchanged between DNA, in order to create the cell lines. Cells use homologous recombination as a repair mechanism for double-stranded breaks, so by introducing homologous DNA sequences into ES cells, we can rely on the cell's endogenous machinery to incorporate the new, modified DNA into the original chromosomes. Homologous recombination ensures that the only difference between the cell lines is the changed locus, and it also assures that the altered DNA sequence inserts where we want it to (Figure 9). This limits the number of other variables between the two ES cell lines, replaces the mouse endogenous *As3mt* sequence with the inserted human sequence, and reduces the number of off target cut sites that you may experience with CRISPR/cas9 cloning methods. The ratio of AS3MTd2d3 to AS3MTfull transcripts in the two types of cell lines are examined.

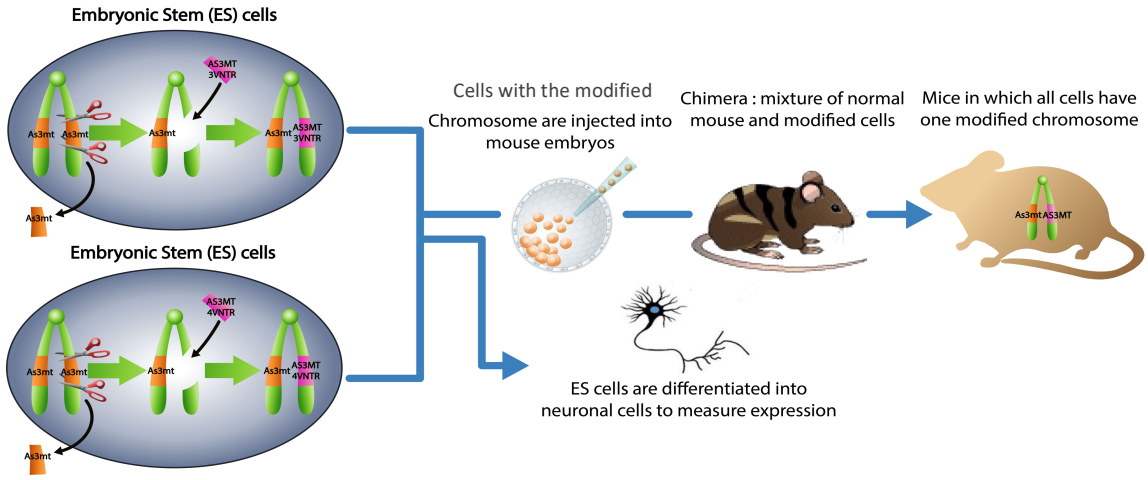

Figure 9. Schematic of the overall strategy. ES cells are humanized with either the AS3MT allele with 3 VNTRs (top) or 4 VNTRs (bottom) by homologous recombineering techniques. The ES cells can be further differentiated into neuronal cells or used to create mice with the modified AS3MT gene.

Results

AS3MT locus

The human BORCS7 gene and AS3MT gene are located adjacent to each other in the same orientation on chromosome 10. Given their proximity, a BORCS7-AS3MT read-through transcript is often formed, but is subject to nonsense-mediated mRNA decay. Bacterial recombineering techniques were used in *E.coli* carrying human bacterial artificial chromosomes to create the humanized BORCS7-AS3MT locus. The human DNA is introduced into mouse ES cells by homologous recombination, replacing the endogenous mouse As3mt locus. The inserted human DNA is introduced with a Neo sequence into the mouse genome (Figure 10).

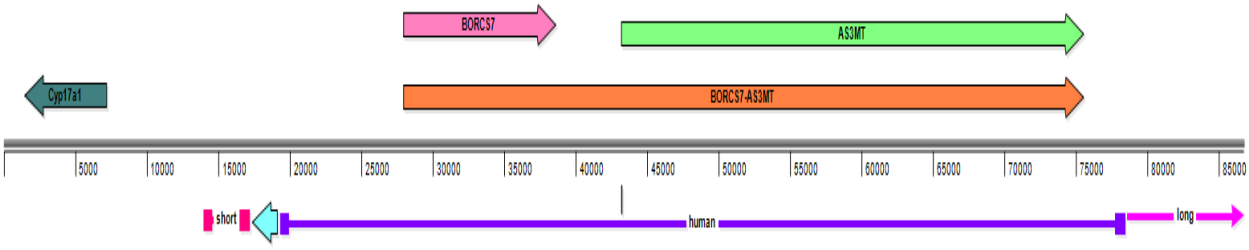

Figure 10. Model of the genome of the ES cells after homologous recombination. Mouse endogenous Cyp17a1 gene (dark green), and the inserted human sequence (purple) including the short and long arms and Neo sequence (blue) inserted with it. The human sequence leads to BORCS7 (pink), AS3MT (light green), and BORCS7-AS3MT (orange) transcript formation.

Mouse phenotypes with triple VNTR AS3MT gene

Cell lines 1488-5G12 and 1489-4D3 had previously been created from mouse ES cells. These cell lines contain the BORCS7-AS3MT locus with the triple VNTR sequence in exon 1 of AS3MT. By injecting ES cell line 1489-4D3 into mice blastomas, mice that are heterozygous for the huAS3MT gene where created and bred.

These mice weighed more than the control B6J mice and had larger white adipose tissue fat pads. The weights of the adrenals, testes, brown adipose tissue (BAT), and white adipose

tissue (WAT) for the control (+/+), heterozygous (+/hu), and homozygous (hu/hu) mice were determined relative to the total body weight (Tables 1 and 2). While AS3MT is known to be more highly expressed in adrenals and testes, there was no evident trend in weight differences between the three genotypes.

Table 1. Tissue weights from wild type (+/+), heterozygous (+/hu) for the humanized AS3MT gene, and homozygous (hu/hu) for the humanized AS3MT gene mice.

huAS3MT mice		**sex**	**Birth Date**	**Weight (g)**	**Adrenals (mg)**	**Testes (mg)**	**BAT (mg)**	**WAT/epididymal (mg)**
B6J.1	+/+	M	10/10/18	32.3	8.21	206.99	124.05	351.45
B6J.2	+/+	M	10/10/18	30.9	2	219.72	100.94	293.37
1968.3	+/hu	M	10/6/18	35.9	3.28	244.56	136.86	1073.44
1976.98	+/hu	M	10/16/18	34.4	2.56	205.37	160.73	1068.59
1968.7	hu/hu	M	10/6/18	34.5	2.2	222.72	138.21	1078.62
1968.4	hu/hu	M	10/6/18	37.9	2.58	163.72	106.71	1840.26
1968.8	hu/hu	M	10/6/18	34.5	9.06	207.09	84.3	753.08
1976.83	hu/hu	M	10/16/18	37.1	4.65	179.91	121.12	1142.79
1976.84	hu/hu	M	10/16/18	34.5	5.34	148.33	71.97	509.4
1867.91	hu/hu	M	9/2/18	33.1	4.18	160.64	82.41	824.96

Table 2. The average of the relative adrenal and testes weight for each genotype. There is no apparent correlation of these numbers with the genotype.

	Adrenals/total weight (mg/g)	Testes/total weight (mg/g)
+/+	0.159	6.76
+/hu	0.0829	6.39
hu/hu	0.133	5.13

After some of the heterozygous and homozygous mice that were a few days old were allowed to breed, we noticed that only female mice that were heterozygous for huAS3MT got pregnant while the homozygous females did not. This potential trend was found among four heterozygous and three homozygous mice.

Generating the quadruple VNTR construct

The quadruple VNTR AS3MT Displacer vector was constructed using recombineering techniques to modify an existing triple VNTR AS3MT Displacer vector in *E. coli* carrying human bacterial artificial chromosomes. The vector was modified by Red/ET recombination, which is mediated by the phage protein pairs Redα and Redβ. Red/ET recombination keeps unspecific events at a minimum, works independently of restriction sites, and the method is not limited by DNA size. The correct sequences are selected for based on antibiotic resistance. The modified vector was transformed into bacteria and the colonies were selected by growth on streptomycin. These manipulations result in a "seamless" alteration to the DNA such that the correct VNTR and AS3MT sequence is inserted into the chromosome. The presence of the correctly inserted VNTR cassette was confirmed by PCR amplification of the region followed by Sanger sequencing.

The AS3MT Displacer construct carrying the quadruple VNTR was introduced into embryonic stem (ES) cells by electroporation. Homologous recombination occurs and the modified AS3MT gene replaces the mouse endogenous *As3mt* locus. Cells carrying at least one copy of the construct were identified by their ability to grow in medium containing the antibiotic G418. Lysate was prepared from individual colonies of G418 resistant ES cell clones and screened using a PCR designed to specifically amplify the locus in which the Displacer construct had undergone homologous recombination with the endogenous mouse As3mt locus.

Verifying construct

The ES cells lines were verified as containing the correct quadruple VNTR AS3MT DNA sequence through a PCR screen of the cell lysate. Genomic DNA from the positive cell lines was

further verified with a Taqman quantitative PCR (qPCR) with As3mt and Borcs7 primer probes. The As3mt and Borcs7 probes recognize and specifically amplify the endogenous mouse As3mt and Borcs7 DNA. The ratio of As3mt and Borcs7 concentration were determined with respect to the Hprt gene, which is found on the mouse X chromosome. Cell lines with a fold change of 0.5 of both As3mt and Borcs7 compared to the control mouse DNA (Phnx43) were considered to be positives, as they only have the mouse Borcs7-As3mt sequence on only one chromosome rather than two. ES cell lines 1506-B9 and 1509-C9 were verified as having the quadruple VNTR locus by qPCR (Figure 11).

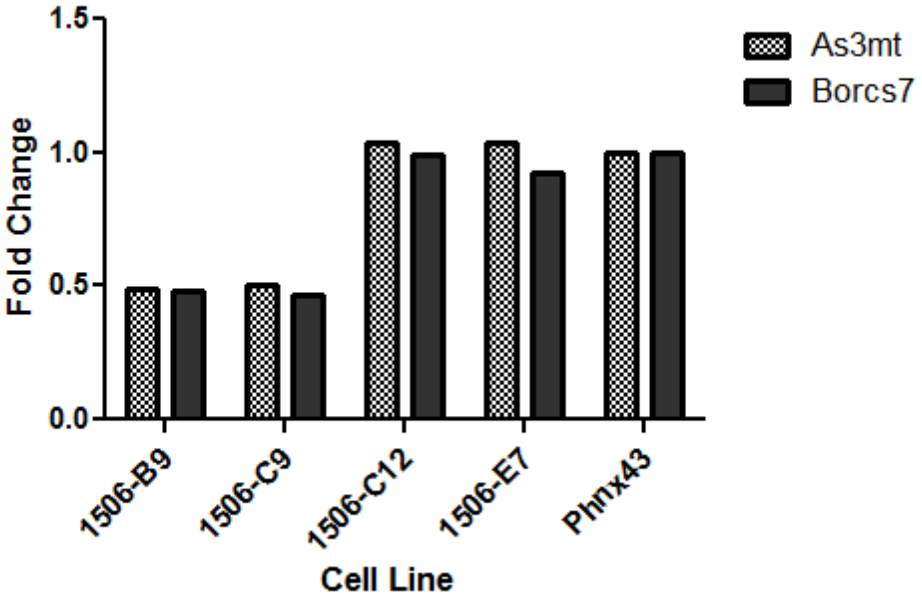

Figure 11. qPCR analysis of the endogenous mouse As3mt and Borcs7 genes. 1506-B9, 1506-C9, 1506-C12, and 1506-C12 cell lines were potential humanized candidates, but only 1506-B9 and 1506-C9 had deleted mouse As3mt and Borcs7 genes. Phnx43 is the male mouse control.

Primer probes that recognize the human BORCS7-AS3MT sequence were also used to confirm that the full sequence was in place. Mouse DNA (Phnx43), human DNA (Hct116), and the triple VNTR AS3MT locus cell line 1488-3A6 were used as controls for the six sets of

primers. PCR reactions were performed to amplify the regions of DNA that the probes recognize and screen for PCR product (Figure 12).

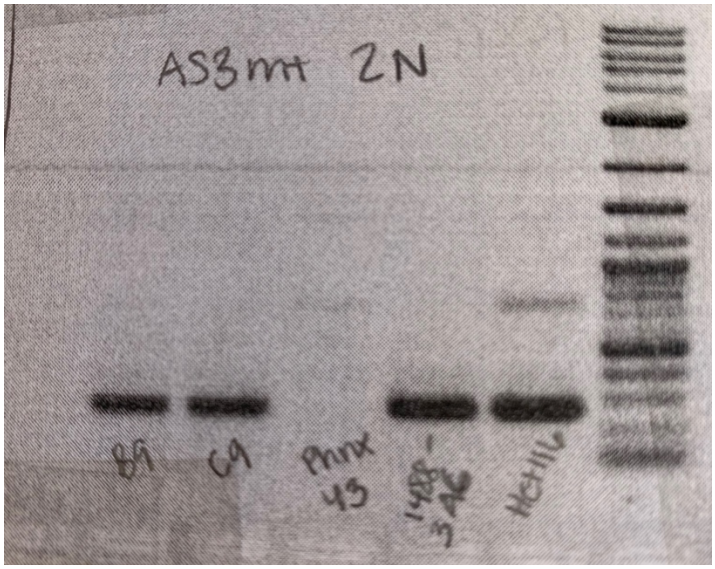

Figure 12. An example of a primer probe PCR reaction. This PCR checks for a sequence in the AS3MT gene that we have called AS3MT hu 2N. Quadruple VNTR cell lines 1506-B9 and 1506-C9 were tested using Phnx 43 (mouse), 1488-3A6 (triple VNTR) and Hct116 (human) DNA as controls. A 2-log ladder is in the furthest right lane of the gel.

ES cell line cDNA analysis

RNA is extracted from the two verified cell lines, 1506-B9 and 1506-C9, and the triple VNTR cell lines 1488-5G12 and 1489-4D3. cDNA is reverse transcribed from these four cell lines so that the ratios of full to d2d3 AS3MT transcripts can be determined using SYBR Green qPCR methods. Primers were designed to amplify the full and d2d3 AS3MT sequences. The resulting PCR product fluoresces, and the amount of relative fluorescence can be used to determine relative transcript concentration.

In order to verify that the mice can be used as a model for the alternative splicing on the AS3MT gene and the subsequent protein isoforms, the ES cells with the previously generated triple VNTR locus were used to create mice humanized with the AS3MT gene. The ration of the d2d3 to full AS3MT transcripts were tested in tissues from these mice to determine whether the

VNTRs cause alternative splicing in the mice (Figure 13 a, b). Based on these results, it was confirmed that mice are capable of producing the d2d3 transcript.

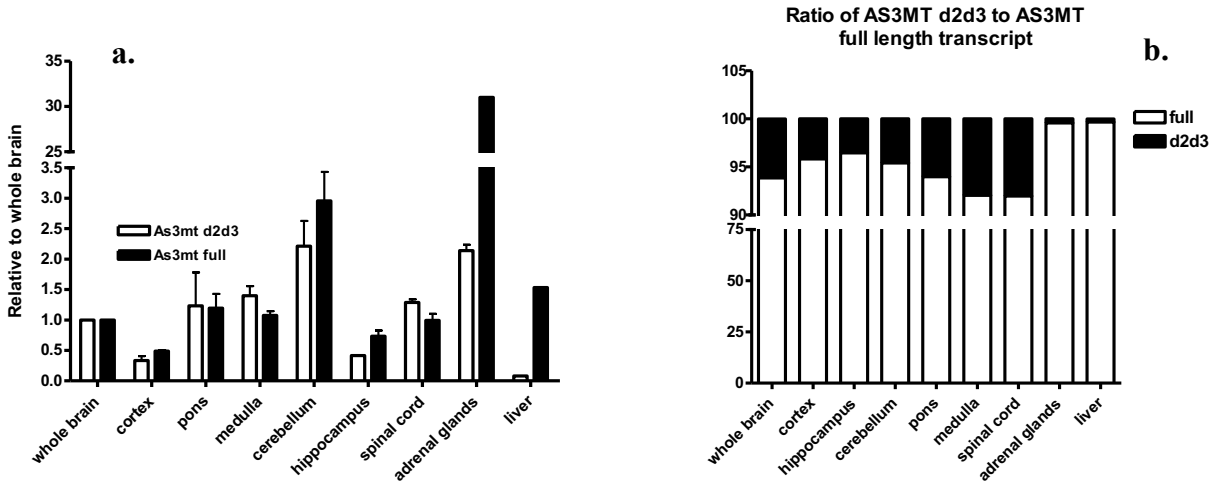

Figure 13. a) Fold change in relative concentration of AS3MT d2d3 transcripts (white) and AS3MT full transcripts (black) in tissues of the mice humanized with the triple VNTR AS3MT construct compared to the concentration in the mouse whole brain. b) Percentage of the d2d3 transcripts (black) and of the full transcripts (white) in these tissues of the total AS3MT transcript number.

The same experimental setup is repeated to determine the ratio of the transcripts in the ES cell lines 1506-B9, 1506-C9, and 1488-5G12, along with cDNA from the humanized mouse spinal cord and liver. In the ES cells, there is no apparent trend in the ratios of the full to d2d3 transcripts (Figure 14 and 15).

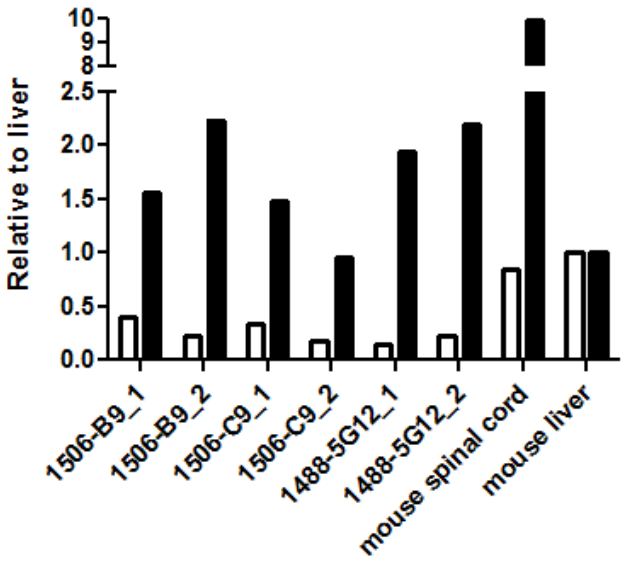

Figure 14. Fold change analysis of the concentrations of the AS3MTfull (white) and AS3MTd2d3 (black) transcripts. Two cDNA samples of each 1506-B9, 1506-C9, and 1488-5G12 cell lines were analyzed along with cDNA from mouse spinal cord and mouse liver. The fold change was calculated relative to the concentrations in the mouse liver cDNA. There is no apparent trend or consistency in the ES cell lines, likely due to a high background and low AS3MTd2d3 transcript number.

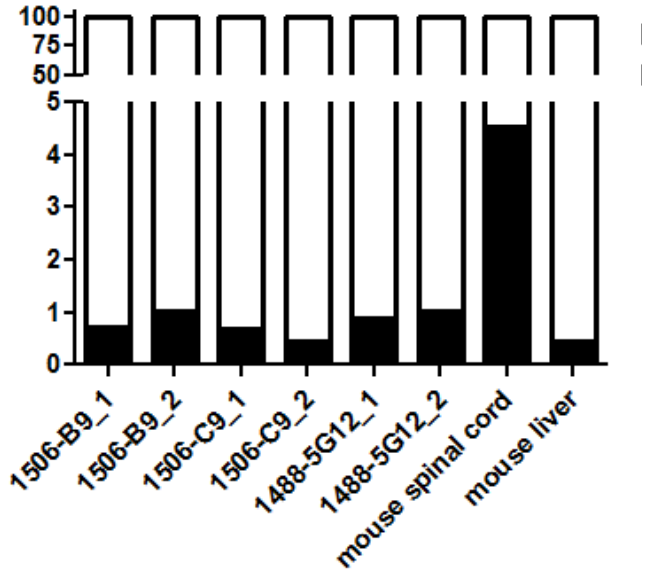

Figure 15. Percentage of AS3MTd2d3 (black) and of AS3MTfull (white) transcripts of the total recognized AS3MT cDNA. Two cDNA samples of each 1506-B9, 1506-C9, and 1488-5G12 cell lines were analyzed along with cDNA from mouse spinal cord and mouse liver. There is no apparent trend or consistency in the ES cell lines, likely due to a high background and low AS3MT transcript amount.

Droplet digital PCR (ddPCR) methods with EvaGreen Supermix is also used to verify these results. The SYBR Green qPCR relies on determining the concentrations relative to a chosen sample, but ddPCR allows for the calculation of absolute concentrations of the transcripts. ddPCR creates droplets of the samples and measures the relative number of droplets with the chosen PCR product. AS3MT full and the AS3MT d2d3 primers are used with the cDNA of the 1506-B9, 1506-C9, 1488-5G12, and 1489-4D3 cells lines, along with cDNA made from humanized mouse spinal cord and liver (Figure 16). The percent of d2d3 transcript in the cell lines was at such a low level that the margin of error of the samples gave results that were not consistent. As with the qPCR methods, there was no apparent change in the ratio of the transcripts between the cells with the quadruple VNTR locus and those with the triple VNTR locus due to the inconsistency between runs and high background.

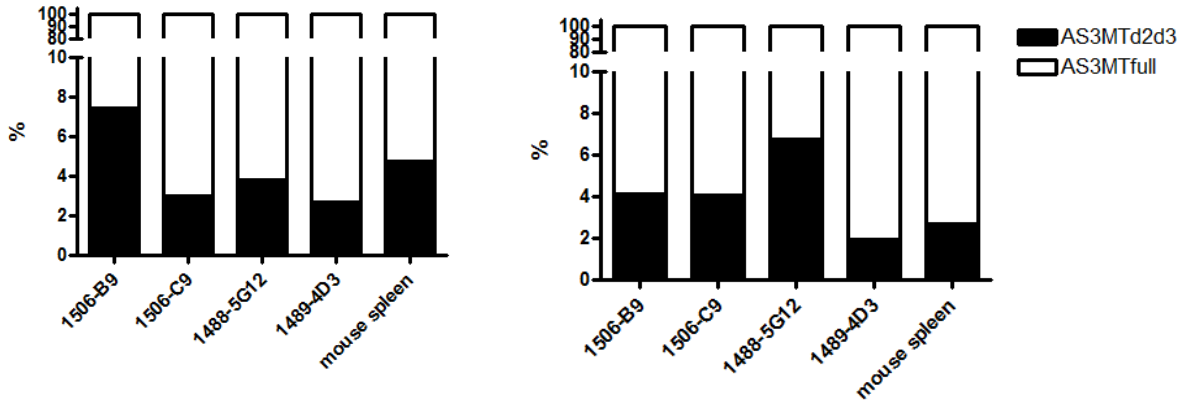

Figure 16. Two trials of the ddPCR runs testing the ratio of AS3MT d2d3 to AS3MT full transcripts. The percentage of d2d3 transcripts of the total AS3MT transcript number is depicted by the black bars for each sample. cDNA samples of 1506-B9 and 1506-C9 (quadruple VNTR ES cell lines), and of 1488-5G12 and 1489-4D3 (triple VNTR cell lines) were analyzed along with cDNA from humanized mouse spleen. There is no evident consistency between the two trials and the transcript percentage is at lower numbers for each of the samples.

Discussion

The role of AS3MT in the human body may not be as clear as simply methylation of trivalent arsenic for metabolism and detoxification. The presence of the gene in vivo seemed to cause the mice to develop more fat and may have a role in fertility. While more experiments and trials need to be done on these subjects, our findings hint at the fact that AS3MT may have a more widespread role than previously thought. We cannot definitively say that it plays a role in fertility, but the inserted locus has a Neo site that may interfere with the upstream Cypt17a1 gene (Figure 10). Cypt17a1 is involved in the formation of sex hormones such as estrogen and testosterone,[15] so the interference could have been affecting the sexual maturation of the AS3MT hu/hu female mice. More studies have to be done to confirm any relationship between AS3MT and the reproductive system and rule out the chance that the pregnancy trend may have been a coincidence.

In regards to the copy number of VNTRs in the AS3MT locus, we have not yet confirmed or denied that the quadruple VNTR sequence increases the amount of d2d3 transcript produced, which would be related to the schizophrenia risk factor. However, the correlation of AS3MT to schizophrenia is related to an increased ratio of the d2d3 mRNA transcript in the brain.[13] Based on Li et al's study,[13] it is known that there is a higher amount of AS3MT transcripts in the brain, adrenal glands, and testes compared to other tissues. In the brain specifically, AS3MT transcripts are seen in both astrocytes and neurons, but they tend to be more highly concentrated in the astrocytes. Given the low levels of both of the transcript types in our ES cells, we are assuming that there is too much background noise to determine a significant difference between the cell lines in their current state.

The ES cells will be differentiated into neuronal cell fates in order to increase the overall expression of AS3MT and further tested to determine if any correlation exists between the VNTR number and the amount of d2d3 transcript. These experiments will be conducted with qPCR and ddPCR methods across the two quadruple VNTR cell lines and the two triple VNTR cell lines used in the ES cell studies. The differentiated cells will hopefully give us more insight into the role of AS3MT and whether the VNTR copy number impacts that splicing of the gene. If the neuronal cell types do not provide conclusive data that the quadruple VNTR leads to an increase in d2d3 transcripts, we will use the ES cell lines to create humanized mice with chromosomes containing a couple of the quadruple VNTR AS3MT sequence to further test the ratio of the transcripts and protein isoform. If we are not able to prove a direct relationship between VNTR repeat number and an increase in the AS3MT d2d3 isoform, this could point to some holes in Li *et. al's* study. The correlation that they saw may not necessarily indicate causation. It is important to verify their finding with a biochemical analysis in order to find specific genes that can be targeted in treatment options of schizophrenia. If the VNTR repeat number does shows an increase of d2d3 transcripts in the differentiated cells and mice, we can further look to see if the mice humanized with the 4 VNTR locus versus the 3 VNTR locus display schizophrenic characteristics. The existence of such characteristics would promote the notion that d2d3 isoforms are linked to schizophrenia risk factors.

Arsenic is very prevalent in the environment and is toxic to humans, so while AS3MT plays an important role in arsenic metabolism, it is possible that it has other functions. This is an idea that warrants further investigation, given that cells may be more sensitive to the methylated forms of arsenic that come from the reaction driven by AS3MT. If shown that the number of VNTR repeats in the AS3MT gene leads to alternative splicing, our data has the potential to offer

the first biochemical evidence to support the genetic studies linking VNTRs to the d2d3 isoform, in turn conferring risk to schizophrenia.

Methods

Bacterial Recombination

The quadruple VNTR AS3MT Displacer vector was constructed using recombineering techniques to modify an existing triple VNTR AS3MT Displacer vector. An existing counterselection cassette carrying an aminoglycoside phosphotransferase gene, which confers resistance to kanamycin, and a ribosomal S12 protein gene, which confers sensitivity to streptomycin, was modified by using standard cloning techniques to replace the kanamycin resistance gene with a beta-lactamase gene, which confers resistance to ampicillin. Red/ET recombination was then used to replace the triple VNTR region of the starting construct with the counterselection cassette. The Red/ET plasmid is added with the displacer vector and counterselection cassette and induced with arabinose. The temperature is then increased for an hour to get transient recombineering ability. The resulting phage protein pairs Redα and Redβ have exonuclease activity and are single stranded bringing proteins, respectively, and work to replace the target DNA sequence with the counterselection cassette. After selection on ampicillin, the correct insertion of the counterselection cassette in the modified vector was confirmed by a polymerase chain reaction (PCR). A second Red/ET recombination step was then used to replace the counterselection cassette in the vector with a cassette carrying the AS3MT quadruple VNTR. Bacteria carrying a vector in which the desired modification had occurred were selected by growth on streptomycin, and the presence of the correctly inserted VNTR cassette was confirmed by PCR. The correct sequence of the VNTR region of the modified

vector was confirmed by PCR amplification of the region followed by Sanger sequencing. The AS3MT Displacer construct carrying the quadruple VNTR was introduced into embryonic stem (ES) cells by electroporation, and cells carrying at least one copy of the construct were identified by their ability to grow in medium containing the antibiotic G418.

ES cell line screening

Colonies expressing G418 resistance were plated at low dilutions and colonies arising from individual cells were sub-cultured and cell lysates prepared from ~ 300 to ~1,000 cells. Cell lysate was made from the colonies by adding 20 µL lysate buffer (Protease K, triton, TE) to 13 µL of cells and incubating at 55°C for 30 min. The reaction was ended by heating for 10 min at 95°C. Lysates were screened by polymerase chain reaction (PCR) to screen for the presence of the BORCS7 and AS3MT genes. The screen PCR was designed to specifically amplify the locus in which the Displacer construct had undergone homologous recombination with the endogenous mouse As3mt locus. Colonies expressing two copies of the wildtype As3mt locus were eliminated.

RNA extraction

After lysing the cells, RNA is isolated using the RNA-Bee protocol. The lysate is separated into organic and aqueous phases using chloroform and the RNA is recovered from the aqueous phase. The isolated RNA is used as a template for a PCR reserve transcription reaction. cDNA was obtained at a concentration of about 100 ng/uL.

RT-PCR to determine the ratio of AS3MTd2d3 to AS3MTfull transcripts

The cDNA is used for a qPCR analysis targeting the AS3MT full transcript and the AS3MT d2d3 transcript, which is missing exons 2 and 3. The reaction uses iTaq Universal SYBR Green Supermix, which causes the cDNA transcripts of interest to fluoresce. Analysis of the data is done by comparing the cycle number it takes for each sample to reach the threshold fluorescence value. The analysis is conducted with 18S cDNA as an internal normalizing control. The results are compared using delta-Ct, fold change analysis.

ddPCR analysis to determine the ratio of AS3MTd2d3 to AS3MTfull transcripts.

Droplet digital PCR technology targeting the AS3MT full transcript and the AS3MT d2d3 transcript of the cDNA is used with an EvaGreen Supermix. Water-oil emulsion droplets are created, and PCR amplification of either the full or d2d3 transcript occurs in each droplet. The proportion of droplets with PCR products are related to the concentration of each transcript.

References

1. Hughes, M. F. Arsenic Toxicity and Potential Mechanisms of Action. *Toxicology Letters* 2002, *133* (1), 1–16.
2. "National Biomonitoring Program." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 7 Apr. 2017.
3. Dixon, H.B.F., 1997. The biochemical action of arsonic acids especially as phosphate analogues. Adv. Inorg. Chem. 44, 191–227.
4. Kenney, L.J., Kaplan, J.H., 1988. Arsenate substitutes for phosphate in the human red cell sodium pump and anion exchanger. J. Biol. Chem. 263, 7954–7960.
5. Peters, R.A., 1955. Biochemistry of some toxic agents. I. Present state of knowledge of biochemical lesions induced by trivalent arsenical poisoning. Bull. John Hopkins Hospital 97, 1–20.
6. Petrick, J.S., Ayala-Fierro, F., Cullen, W.R., Carter, D.E., Aposhian, H.V., 2000. Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes. Toxicol. Appl. Pharmacol. 163, 203–207.
7. Soignet, S. L.; Maslak, P.; Wang, Z.-G.; Jhanwar, S.; Calleja, E.; Dardashti, L. J.; Corso, D.; DeBlasio, A.; Gabrilove, J.; Scheinberg, D. A.; et al. Complete Remission after Treatment of Acute Promyelocytic Leukemia with Arsenic Trioxide. *The New England Journal of Medicine; Boston* 1998, *339* (19), 1341–1348.
8. Chen, S.-J.; Zhou, G.-B.; Zhang, X.-W.; Mao, J.-H.; de The, H.; Chen, Z. From an Old Remedy to a Magic Bullet: Molecular Mechanisms Underlying the Therapeutic Effects of Arsenic in Fighting Leukemia. *Blood* 2011, *117* (24), 6425–6437.
9. Zhen Y, Zhao S, Li Q, Li Y, Kawamoto K. Arsenic trioxide-mediated Notch pathway inhibition depletes the cancer stem-like cell population in gliomas. Cancer Lett. 2010;292(1):64-72.
10. Kim J, Lee JJ, Kim J, Gardner D, Beachy PA. Arsenic antagonizes the Hedgehog pathway by preventing ciliary accumulation and reducing stability of the Gli2 transcriptional effector. Proc Natl Acad Sci U S A. 2010;107(30):13432-13437.
11. AS3MT Gene - GeneCards | AS3MT Protein | AS3MT Antibody (accessed Feb 27, 2019).
12. Lu, J.; Hu, S.; Wang, W.; Li, J.; Dong, Z.; Zhou, J.; Hai, X. AS3MT Polymorphisms, Arsenic Metabolism, and the Hematological and Biochemical Values in APL Patients Treated with Arsenic Trioxide. *Toxicol Sci* 2018, *166* (1), 219–227.
13. 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. A Human-Specific AS3MT Isoform and BORCS7 Are Molecular Risk Factors in the 10q24.32 Schizophrenia-Associated Locus. *Nature Medicine* 2016, *22* (6), 649–656.
14. 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. Human Arsenic Methyltransferase (AS3MT) Pharmacogenetics: GENE RESEQUENCING AND FUNCTIONAL GENOMICS STUDIES. *Journal of Biological Chemistry* 2006, *281* (11), 7364–7373.
15. CYP17A1 Gene - GeneCards | CP17A Protein | CP17A Antibody (accessed Mar 25, 2019).