

Analysis of the transcriptional regulation of cancer-related genes by aberrant DNA methylation of the cis-regulation sites in the promoter region during hepatocyte carcinogenesis caused by arsenic

Zhuang Miao¹, Lin Wu¹, Ming Lu², Xianzhi Meng¹, Bo Gao¹, Xin Qiao², Weihui Zhang¹, Dongbo Xue¹

¹Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, PR China

²Department of Surgery, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA

Correspondence to:

Dongbo Xue, e-mail: dongbo_xue@126.com

Keywords: carcinogenesis, DNA methylation, arsenic exposure

Received: March 21, 2015

Accepted: May 11, 2015

Published: May 25, 2015

ABSTRACT

Liver is the major organ for arsenic methylation metabolism and may be the potential target of arsenic-induced cancer. In this study, normal human liver cell was treated with arsenic trioxide, and detected using DNA methylation microarray. Some oncogenes, tumor suppressor genes, transcription factors (TF), and tumor-associated genes (TAG) that have aberrant DNA methylation have been identified. However, simple functional studies of genes adjacent to aberrant methylation sites cannot well reflect the regulatory relationship between DNA methylation and gene transcription during the pathogenesis of arsenic-induced liver cancer, whereas a further analysis of the cis-regulatory elements and their trans-acting factors adjacent to DNA methylation can more precisely reflect the relationship between them. *MYC* and *MAX* (*MYC* associated factor X) were found to participating cell cycle through a bioinformatics analysis. Additionally, it was found that the hypomethylation of cis-regulatory sites in the *MYC* promoter region and the hypermethylation of cis-regulatory sites in the *MAX* promoter region result in the up-regulation of *MYC* mRNA expression and the down-regulation of *MAX* mRNA, which increased the hepatocyte carcinogenesis tendency.

INTRODUCTION

As a cytotoxic substance, arsenic is widely existed in natural environment, particularly in Asian. Long-term exposure to a certain level of arsenic that can result in poisoning has been established as an issue that seriously harms individuals' health [1]. Arsenic not only causes nerve and skin lesions but is also centrally involved in liver damage, cirrhosis and the pathogenesis of liver cancer [2]. Studies have shown that the liver is an important target organ of the carcinogenic effects of arsenic. A study by Li J [3] proposed that the liver was the potential target of arsenic-induced cancer, Lin HJ [4] reported that chronic arsenic poisoning through drinking water could promote the pathogenesis of liver cancer, Chen H [5] reported that long-term exposure to arsenic

could induce abnormal gene expression and develop liver cancer. Previous studies have explored the mechanism of arsenic-induced hepatocarcinogenesis. For example, studies have confirmed that arsenic can directly or indirectly injure hepatocytes through DNA damage, lipid peroxidation, the inhibition of apoptosis and by promoting excessive cell proliferation [6]. Recent reports found that during its hepatocarcinogenesis process, arsenic stimulated the hypomethylation (HypoM) of *DNMT3* and *HA-RAS* genes [7] and the hypermethylation (HyperM) of *P16* and *P53* genes [3], indicating that the effect of arsenic on hepatocyte gene methylation levels may be an important mechanism during its hepatocarcinogenesis process.

Inorganic arsenic toxicity is decreased inside the body through methylation, and the liver is the major location for methylation metabolism. The methylation

metabolism process of inorganic arsenic in the body competes for methyl donors with the DNA methylation modification process, which affects the DNA methylation-demethylation modification. With the increase of the accumulation and uptake of inorganic arsenic, the human body, particularly the liver, has a larger burden for arsenic methylation. Therefore, we speculate that under the conditions of long-term arsenic exposure, the methylation level of hepatocytes may change, which in turn affects the expression of proto-oncogenes and tumor suppressor genes, thereby increasing the tendency of hepatocarcinogenesis. To verify this inference, we performed DNA methylation detection and analysis using a methylation microarray on normal human liver cells that have been long-term exposed to arsenic to explore the possible mechanism for the pathogenesis of arsenic-induced liver cancer.

RESULTS

Analysis of the aberrant DNA methylation sites and adjacent gene annotation

Based on the analysis results of the aberrant DNA methylation sites, we detected that the DNA methylation signal in the arsenic-exposure group was up-regulated in 1148 DNA methylation sites, which represents the HyperM phenomenon. In combination with the annotation of genes adjacent to the DNA methylation sites, 637 gene promoter regions contained these HyperM sites. While in the neighborhood of the 1159 HypoM sites in the arsenic-exposure group, we isolated a total of 683 genes that met the criteria (Table 1).

This result indicates that the transcription status of these genes may be affected by arsenic treatment. It is noteworthy that in all of the annotation results, we found 41 genes with promoter regions that contained both HyperM sites and HypoM sites simultaneously.

Pathway enrichment analysis of the genes adjacent to the aberrant DNA methylation sites

Based on the above analysis results, we used the types of aberrant DNA methylation sites in the gene promoter region as a standard and categorized the genes adjacent to the aberrant DNA methylation sites as genes adjacent to hypermethylated sites (HyperM adjacent genes), genes adjacent to hypomethylated sites

(HypoM adjacent genes) and mixed types of genes (Mixed genes, i.e., the promoter region contained both HyperM sites and HypoM sites at the same time). We performed a pathway enrichment analysis for each category. The detailed results for the pathway enrichment of each type of gene were shown in Figure 1.

Based on the KEGG pathway enrichment analysis results, we found that some HyperM adjacent genes in the arsenic-exposure group were enriched in the Wnt signaling pathway (including *MYC*, *APC*, *CACYBP*, *CAMK2G*, *FZD7*, *LEF1*, *LRP6*, *NFATC3*, *ROCK2*, *WNT2B* and *WNT7A*) and the TGF-beta signaling pathway (including *MYC*, *ACVR1C*, *FST*, *ROCK2*, *TGFBR2* and *THBS1*). In addition, genes related to the cellular metabolic pathways, such as phenylalanine, tyrosine and tryptophan biosynthesis, glycosaminoglycan biosynthesis, valine, leucine and isoleucine biosynthesis, may also be affected by arsenic-induced DNA HyperM. It is noteworthy that in the arsenic-exposure group, we found 4 nucleotide excision repair pathway genes, including *ERCC2*, *POLD3*, *POLE2* and *RPA1*. In addition, we found that HyperM adjacent genes were enriched in the ribosome pathway.

HypoM adjacent genes were enriched in the pathways of lysosome, endocytosis, protein processing in the endoplasmic reticulum, RNA transport, and fatty acid elongation in the mitochondria, whereas Mixed genes were mainly enriched in the pathways of focal adhesion, pyrimidine metabolism, oocyte meiosis, carbohydrate digestion and absorption.

In addition, in combination with the information from the Disease Ontology database, we found that 12 known primary hepatic cancer (PHC) related genes, including *MYC*, *CDC16*, *DDRI*, *EBAG9*, *IGF2*, *KDR*, *MAD2L2*, *MEF2C*, *NDRG2*, *SIX1*, *TNFRSF10B* and *TSGA10*, were HyperM in the arsenic-exposure group.

Screening of transcription factors (TF) and tumor-associated genes (TAG) from genes adjacent to aberrant DNA methylation sites

In combination with known gene function information, we performed screening on HyperM adjacent genes, HypoM adjacent genes and Mixed genes. Additionally, we analyzed TF, oncogenes, tumor suppressors and other tumor-related genes without known functions from each type of gene. The details of the analysis results were shown in Table 2.

Table 1: Analysis results of the aberrant DNA methylation sites

	CpG Site	Adjacent Genes
HyperM	1148	637
HypoM	1159	683
Total	2307	1279

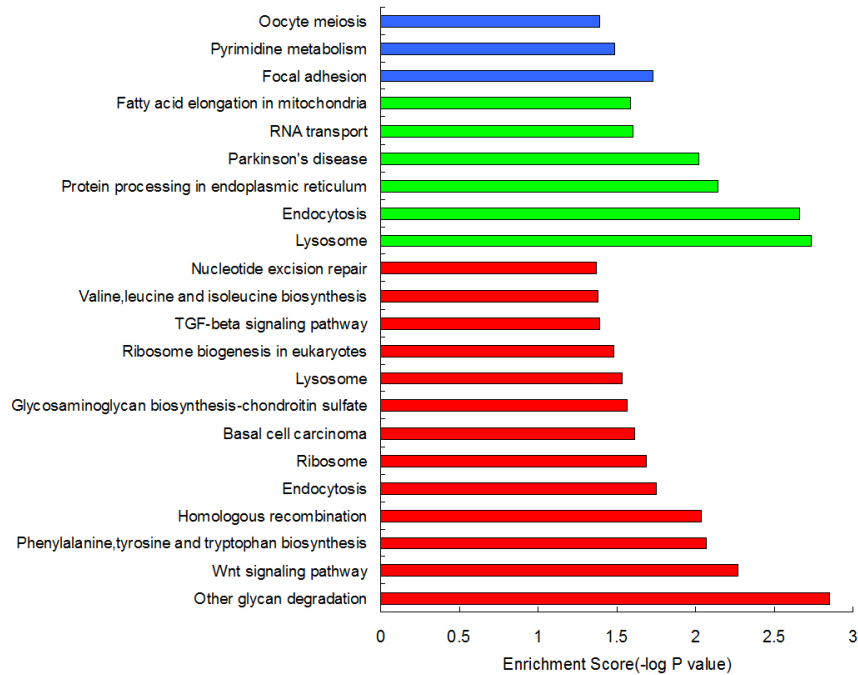


Figure 1: The pathway enrichment analysis results of aberrantly methylated genes in the promoter region. The red columns represents HyperM adjacent genes, the green columns represent HypoM adjacent genes, and the blue ones represent Mixed genes.

Construction of the PPI (protein-protein interaction) sub-network of the genes adjacent to the aberrant DNA methylation sites

Based on the String PPI database search results, we constructed a PPI sub-network of the genes adjacent to the aberrant DNA methylation sites (Figure 2). The results showed that in the relevant PPI sub-network, the key Hub genes with a node degree ≥ 10 included *UBB* (degree = 20), *MYC* (degree = 18), *EIF4E* (degree = 12), *RPL6* (degree = 12), *RPS14* (degree = 11), *NCBP1* (degree = 4), *RPL18A* (degree = 11), *NCBP1* (degree = 11), *RPL37* (degree = 11), *CDK4* (degree = 11), *RPS29* (degree = 11), *BARD1* (degree = 10), *BUB3* (degree = 10), and *TGFB1* (degree = 10).

Annotation and enrichment analysis of the aberrant DNA methylation region related cis-regulatory sites

The detailed annotation results of the aberrant DNA methylation region related cis-regulatory sites are shown in Table 3. We demonstrated that one or more cis-regulatory sites existed in the neighboring regions of the 804 HyperM sites and the 834 HypoM sites. Considering that DNA methylation in the promoter region of the gene has the closest relationship to the transcriptional regulation of this gene, we performed further screening and collected all of the annotation results of the aberrant DNA methylation sites in the

promoter region of the genes and their cis-regulatory elements. We used an enrichment analysis to detect the enrichment significance of each cis-regulatory element. The results showed that we identified 550 HyperM sites and 600 HypoM sites that met the abovementioned screening criteria.

Based on the enrichment analysis of the annotation results of aberrant DNA methylation sites in the promoter region, we showed that the enrichment around the aberrant DNA methylation sites of a total of 32 TF was significant. Among these factors, there were significant enrichment phenomena around both HyperM and HypoM sites in the binding sites of 24 trans-acting factors, including *MYC*, *MAX*, *MAFK*, *CTCF*, *TAF1*, *IRF4*, *RAD21*, *E2F6*, *GABP*, *ELF1*, *SETDB1*, *HEY1*, *SIN3A*, *GATA2*, *ER1*, *TBP*, *STAT3*, *E2F1*, *CHD2*, *SP1*, *YY1*, *FOXA1*, *NRF1*, and *EGR1*. The cognate binding sites of 4 factors, *HDAC2*, *USF1*, *USF2* and *SP11*, only had significant enrichment in the HyperM sites, whereas the cognate binding sites of 4 factors, *ATF3*, *CEBPB*, *ELK4* and *ZBTB33*, only had significant enrichment in the HypoM sites in the arsenic-exposure group (see Figure 3). The results of the classification of these 32 trans-acting factors according to their functions showed that they included 21 TF, 3 known oncogenes, 4 known tumor suppressor genes and 3 TAG. The details of each type of gene are listed in Table 4. Meanwhile, the results of the analysis based on the TF motifs around the aberrant DNA methylation sites showed that the DNA binding sites of a total of 6 TF, including *MYC*,

Table 2: Results of the functional screening of the genes adjacent to the aberrant DNA methylation sites

Class	TF	Oncogene	Tumor Suppressor	Other TAG
HyperM adjacent gene				
	<i>ARNTL2, BRCA2, ELF3, ERCC2, ETV7, FOXP1, GTF3A, HEYL, HNF4A, HOXA11, HOXC13, HOXD11, KCNIP3, LEF1, MEF2B, MEF2C, MYC, NFATC3, NPAS1, PBX1, SIX1, TAF10, TAF5, TBP, TBX22, TCEB1, TEAD4, TFAM, ZNF76</i>	<i>CEP55, COPS5, LF3, GNAS, MYC, PBX1, TAF15</i>	<i>APC, BECN1, BLCAP, BRCA2, CLU, COL18A1, CYLD, DLG1, EEF1A1, FOXP1, HI9, L3MBTL4, LIMD1, MIR124-2, MIR17, MOB1A, NDRG2, PACRG, PTPN13, RND3, RPL10, SFN, TGFBR2, THBS1, TNFRSF10B, TRIM3, ZC3H10</i>	<i>DDR1, ERBB4, EVI2B, FZD7, NCOA4, WNT2B</i>
HypoM adjacent gene				
	<i>AHR, FOXD1, FOXI1, FOXJ1, FOXL2, HDAC5, HES1, HLF, HMGA2, HOXA4, HOXC10, HOXC9, IRF7, LDB1, MEIS2, MXI1, NFE2L2, NKX2-1, RELB, RXRB, SMAD7, SMARCA4, SNAPC1, SOX14, STAT4, TEF, THRB, TMF1, TP73, ZNF143</i>	<i>CDK4, CTTN, ELL, FUS, HMGA2, NET1, NKX2-1, RALA, ZNF217</i>	<i>BARD1, BRMS1, DEFB1, FBXO31, HLF, KLF6, MXI1, NDUFA13, PAF1, PHF6, PRR5, PTEN, PTENP1, RASSF5, SMARCA4, SUFU, TGFBR3, TP53BP2, TP73</i>	<i>ATP5B, BAMBI, BUB1, CTSZ, DHX16, DNMT3B, LAMP3, MBLAC1, NTN1, PCDH1, TFAP2A, TGFBI, THRB</i>
Mixed gene				
	<i>ETV6, SIX5, ZNF24</i>	<i>GFII</i>	<i>ETV6</i>	

TF: transcription factors, TAG: tumor associated gene.

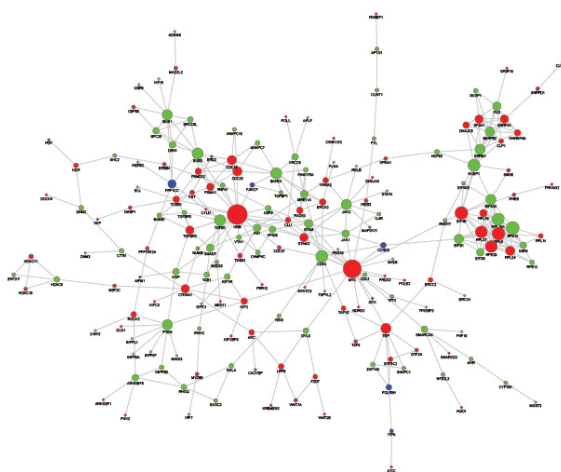


Figure 2: PPI sub-network of genes related to the aberrant DNA methylation sites. The red nodes represent HyperM adjacent genes in the arsenic-exposure group, the green nodes represent HypoM adjacent gene, and the blue ones represent Mixed genes. The diameter of the node is positively proportional to the degree of that node.

Table 3: Compilation of annotation results of the cis-regulatory sites that are related to aberrant DNA methylation regions

	All CpG site covered one or more by cis-element	Related cis-element count	Promoter CpG Site covered one or more by cis-element	Related cis-element count
HyperM	804	71	550	66
HypoM	834	78	600	72

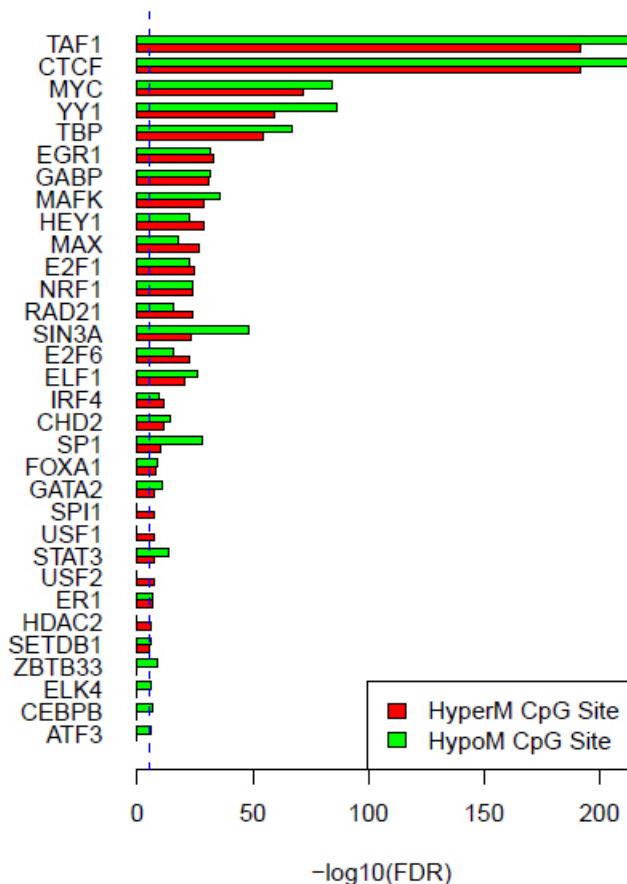


Figure 3: Enrichment analysis results of cis-regulatory elements of the aberrant methylation sites in the promoter region. The red columns represent HyperM sites, and the green columns represent HypoM site. The blue dotted line is the enrichment significance threshold level, anything above this level indicates a significant enrichment result of the cis-regulatory element.

Table 4: Functional classification of trans-acting factors corresponding to significantly enriched cis-regulatory elements

	Gene Count	Genes
TF	21	<i>ATF3, ELK4, USF2, MAFK, CTCF, TAF1, IRF4, HDAC2, MAX, ELF1, HEY1, GATA2, TBP, SPI1, USF1, STAT3, SP1, YY1, FOXA1, NRF1, EGR1</i>
Oncogene	3	<i>MYC, IRF4, SPI1</i>
Tumor Suppressor	4	<i>CTCF, SIN3A, E2F1, EGR1</i>
Other TAG	3	<i>MAX, SETDB1, STAT3</i>

TF: transcription factors, TAG: tumor associated gene.

MAX, *USF1*, *TAF1*, *USF2* and *E2F1*, were enriched in the surrounding region of these DNA methylation sites (see Table 5). This result is consistent with the enrichment analysis results of the cis-regulatory sites.

Construction of aberrant DNA methylation site related transcriptional regulation network

Based on the annotation results of the genes adjacent to the aberrant DNA methylation sites and the enrichment information of TF binding sites from the enrichment results of the cis-regulatory elements of the aberrant DNA methylation sites, we further constructed an aberrant DNA methylation site related transcriptional regulation network (Figure 4). This transcriptional regulation network contained a total of 2290 TF downstream target gene pairs, of which the cis-regulatory element enrichment results included 21 TF, corresponding to 961 target genes. In this network, the network core genes with the 10 highest node degrees in order were *TAF1* (degree = 803), *YY1* (degree = 338), *TBP* (degree = 184), *EGR1* (degree = 159), *IRF4* (degree = 148), *MAX* (degree = 126), *ELF1* (degree = 96), *NRF1* (degree = 87), *HEY1* (degree = 83), and *SP1* (degree = 64).

Detection results of *MYC* and *MAX* mRNA expression

The mRNA expression of *MYC* and *MAX* in the liver cells in the arsenic exposure group and the control group was examined using a real-time RT-PCR method. The results were shown in Figure 5. *MYC* mRNA expression was significantly increased in the arsenic-exposure group (1.76 ± 0.21 vs. 1.0 ± 0.03 , $P < 0.05$), whereas *MAX* mRNA expression was significantly decreased in the arsenic-exposure group (0.45 ± 0.09 vs. 1.0 ± 0.05 , $P < 0.05$).

DISCUSSION

In the present study, we found that arsenic treated liver cells had significantly increased aberrant DNA methylation. DNA methylation is a typical epigenetic marker. Under the catalysis of methyltransferase, the cytosine of the CG dinucleotides in DNA is selectively modified by a methyl group, forming 5-methyl-cytosine. DNA methylation can induce changes in DNA structure and stability and can affect gene transcription and expression. Many studies have shown that DNA methylation can promote or inhibit the pathogenesis of tumors [8]. A study by Nishida N [9] reported that methylation in the promoter region of a gene may inactivate the corresponding tumor suppressor gene and that the genome instability likely caused by the HypoM of the entire gene also activates the pathogenesis of cancer.

Omura K [10] noted that the methylation and expression of normal liver genes and the immune response are involved in early carcinogenesis. Therefore, aberrant DNA methylation is an important mechanism for the pathogenesis of cancer.

Inorganic arsenic can interfere with DNA methylation in its *in vivo* metabolic processes [11]. As an enzyme inhibitor, inorganic arsenic selectively inhibits S-adenosylmethionine-dependent methyltransferase, thereby reducing the utilization of S-adenosylmethionine and causing its concentration to increase, whereas the uninhibited methyltransferase HyperM cytosine [12]. Likewise, some reports have shown that inorganic arsenic can also lead to the HypoM of genes. Its mechanism could be that inorganic arsenic consumes the methyl group of S-adenosylmethionine, causing insufficient intracellular methyl group status, thereby leading to the instability of the methylation mode and the demethylation of genes [13]. Additionally, HyperM and HypoM gene can co-exist to regulate gene functions.

In the present study, through the KEGG pathway enrichment analysis of genes adjacent to aberrant methylation sites, we found that the DNA methylation levels in the promoter regions of 4 nucleotide excision repair pathway genes (*ERCC2*, *POLD3*, *POLE2* and *RPA1*) were higher than in the control group. The DNA repair ability binds to maintaining the human genome integrity and stability and cellular functions. Some reports noted that when DNA repair related genes were inhibited, the sensitivity to cancer causing factors would increase [14]. Studies by Abbasi R [15] showed that the down-regulation of nucleotide excision repair genes would significantly increase the risk of laryngeal cancer. Long-term exposure to arsenic may lead to the transcriptional inhibition of genes related to chromosome base repair inside the cell, making the cells fail to maintain their ability to correct and repair base errors. For example, some studies reported that the *ERCC2* promoter was HyperM in arsenic treated hepatocytes, which inhibited the CAK (cyclin-dependent protein kinase activating kinase)-complex and eventually suppressed DNA repair. This could be one of the important factors for long-term arsenic exposure induced liver cancer [16].

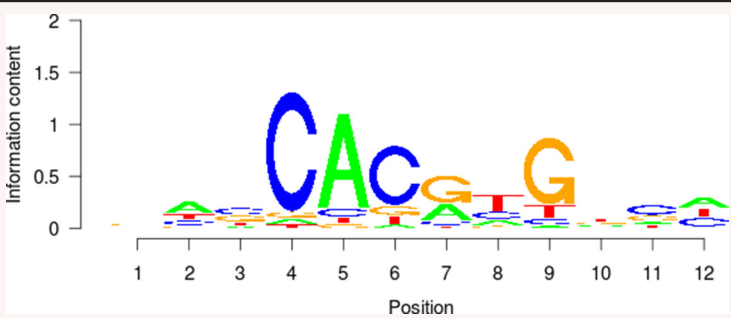
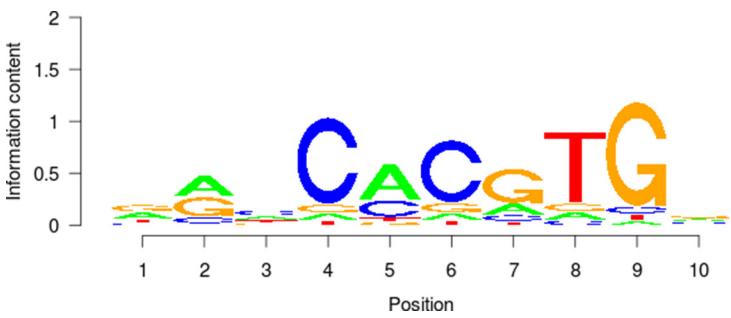
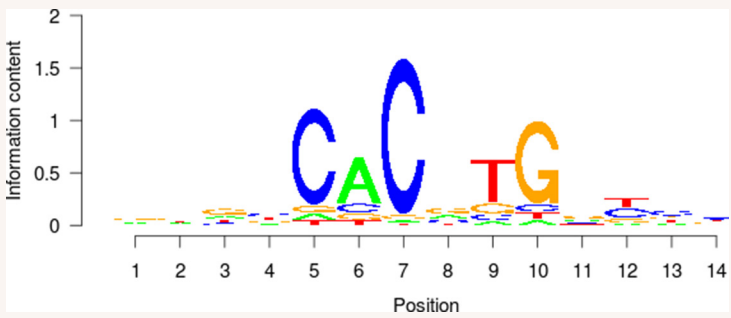
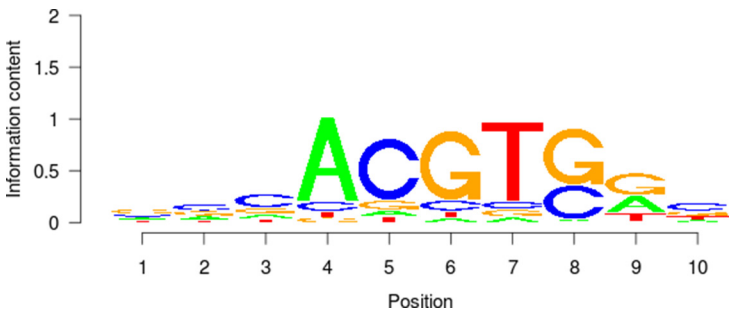
In the present study, we found that HyperM adjacent genes in the arsenic exposure group were mainly enriched in the Wnt signaling pathway and the TGF-beta signaling pathway. Reports showed that these signaling transduction pathways participating the pathogenesis and progression of liver cancer [17, 18]. As shown in Figure 6, the Wnt signaling pathway and the TGF-beta signaling pathway pass extracellular molecular signals through the cell membrane to the inside of the cell to exert their effects. Both pathways are involved in activating cell cycle, and

MYC is located downstream of both pathways, which is a key factor in regulating the cell cycle. Therefore, *MYC* should be a key node of the present study.

Tumorigenesis is inextricably linked to the changes of oncogenes, particularly proto-oncogenes and

tumor suppressor genes. Many studies have confirmed that the activation and excessive expression of proto-oncogenes can promote tumor formation. Meanwhile, the loss or inactivation of tumor suppressor genes may also lead to tumorigenesis. Therefore, we screened TF,

Table 5: Results of the enrichment analysis of transcription factor motifs in the neighborhood of aberrant DNA methylation sites

TF	Motif	MDSseqpos enrichment score ($-\log P$ value)
<i>MYC</i>		142.211
<i>MAX</i>		124.265
<i>USF1</i>		122.075
<i>TAF1</i>		171.884

(Continued)

TF	Motif	MDSseqpos enrichment score ($-\log P$ value)
<i>USF2</i>		130.845
<i>E2F1</i>		122.043

TF: transcription factors.

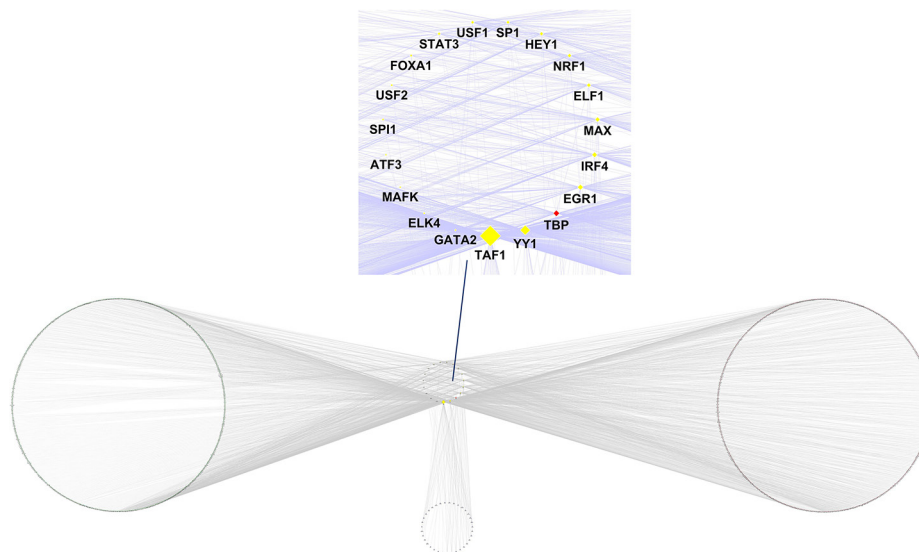


Figure 4: Transcriptional regulatory network that is related to the aberrant DNA methylation sites.

proto-oncogenes, tumor suppressor genes, and other TAG from the genes adjacent to aberrant DNA methylation sites, of which multiple known cancer related genes, including *MYC*, *DDRI*, *MEF2C*, *NDRG2*, *SIX1*, *TNFRSF10B* and *TSGA10*, had HyperM phenomena. In addition, from the PPI sub-network of the genes adjacent to the aberrant DNA methylation sites, we found that among the key genes with a gene node degree greater or equal to 10,

MYC [19], *EIF4E* [20], *CDK4* [21] and *TGFBI* [22] could stimulate the pathogenesis of liver cancer.

However, according to the functional analysis of the abovementioned genes adjacent to the aberrant methylation sites, we found that the adjacent site of the proto-oncogene *MYC* was in a HyperM state. Based on this observation, we inferred that the transcription of *MYC* should be inhibited and *MYC* mRNA expression

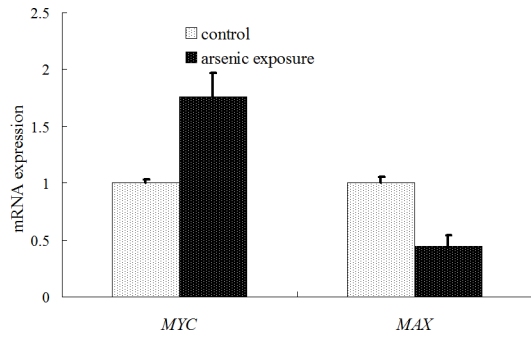


Figure 5: Diagram of the detection results of MYC and MAX mRNA expression using real-time PT-PCR.

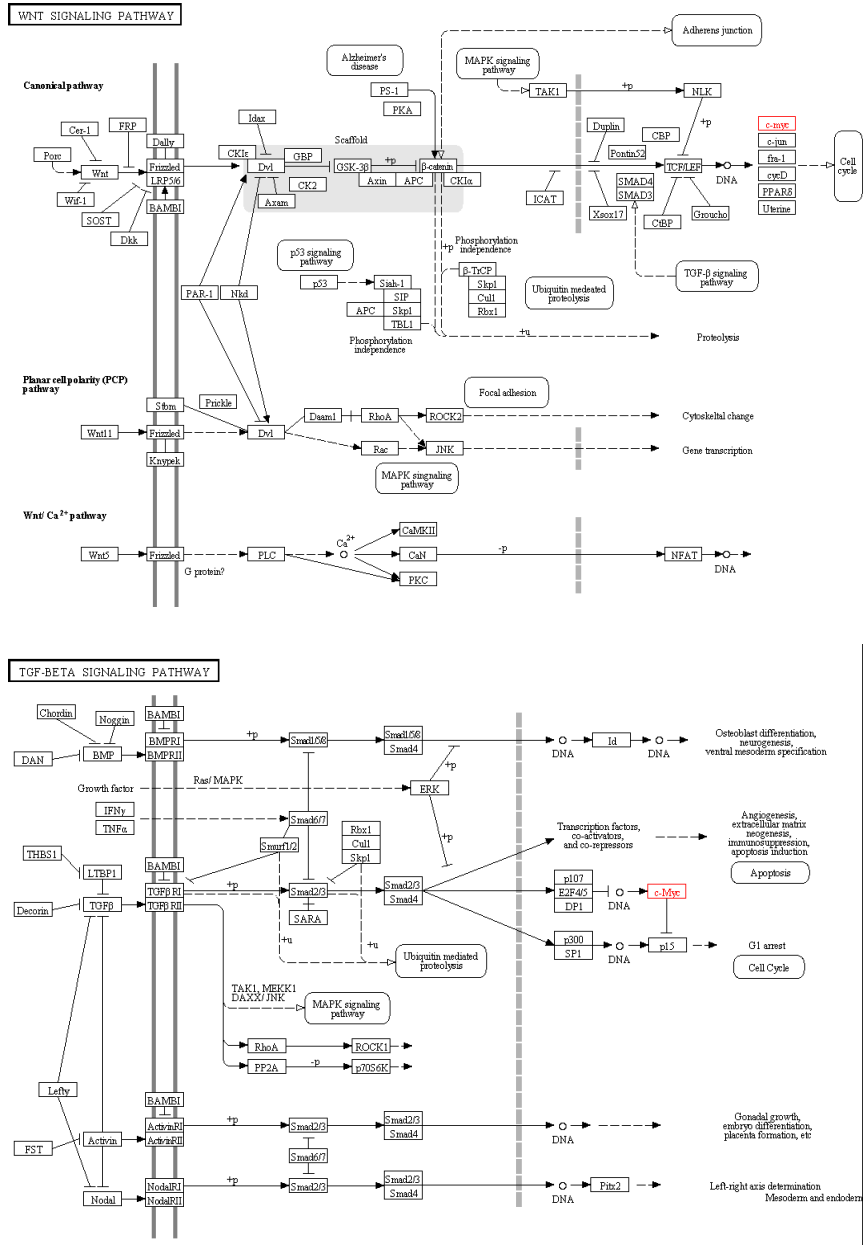


Figure 6: Map of Wnt signaling pathway and TGF-beta signaling pathway.

should be reduced. However, as a proto-oncogene, *MYC* should be increased in its carcinogenic process [23]. Jang KY [19] reported that the high expression of *MYC* could promote the pathogenesis of liver cancer and that *MYC* was considered an indicator of the prognosis of liver cancer. Our real-time RT-PCR detection results also showed that the expression of *MYC* mRNA increased after long-term exposure to arsenic (Figure 5). Obviously, a simple functional analysis of genes adjacent to aberrant methylation sites cannot well explain the effects of DNA methylation on gene transcription during arsenic induced carcinogenic processes.

Therefore, we further performed the annotation and enrichment analysis of cis-regulatory sites that are related to aberrant DNA methylation regions. DNA methylation can result in conformational changes of DNA in the same region on the chromosome, thereby affecting the interaction between the protein and DNA and inhibiting the binding efficiency of the trans-acting factor to DNA. Therefore, we combined the information of the binding sites on the chromosome of all known trans-acting factors, and we searched and analyzed the cis-regulatory sites (i.e., binding sites) of the known trans-acting factors, including TF binding proteins, enhancer binding proteins and insulator proteins, in the proximal regions of both HyperM sites and HypoM sites. We combined enrichment analysis methods to further screen the regulatory elements that were significantly affected by aberrant DNA methylation in the arsenic exposure group.

It is noteworthy that the functional classification of trans-acting factors, whose corresponding cis-regulatory elements are significantly enriched, showed that *MYC* is an important oncogene (Table 4). The results of the enrichment analysis of the TF motifs surrounding the aberrant DNA methylation sites indicated that the MDSeqpos enrichment score of *MYC* was 142.211 (the default enrichment significance threshold was 0.00001, see Table 5). In combination with the known information recorded in the Disease Ontology database, we found in Figure 3 that the cis-regulatory element of *MYC* had both HyperM sites and HypoM sites simultaneously, and the HypoM was predominant. Based on this result, we can infer that the expression of *MYC* should be up-regulated. This result is exactly consistent with our RT-PCR detection results (Figure 5) and literature reports.

Recent studies showed that *MYC* gene family proteins and their associated proteins MAX (MYC-associated factor X) form a network to regulate cell cycle. When these regulatory factors are abnormal, it will lead to uncontrolled cell proliferation and differentiation and the formation of tumors [24, 25]. Ecevit O [26] and Adhikary S [27] reported that known *MYC* functions required the formation of a dimer with MAX to work. *MYC* can form a stable heterodimer with MAX and activate the transcription of the correlated gene. During the entire cell cycle, with the change of *MYC* and MAX

protein levels, either *MYC*-MAX activates transcription or MAX-MAX inhibits transcription and replication [28]. Studies have shown that many malignant tumors have over-expressed *MYC* and under-expressed MAX [29]. Cascon A [30] reported that the lack of MAX in PC12 cells (derived from rat adrenal pheochromocytoma) make it easier to experience tumorigenesis, Albanus RD [31] reported that the expression level of MAX is positively correlated to the survival rate of breast cancer and lung cancer patients, Comino-Méndez I [32] reported that the loss of function of MAX would increase the invasiveness of pheochromocytoma, Yang L [33] reported that the loss of expression of MAX may activates the pathogenesis of pancreatic cancer.

In the present study, the functional classification of trans-acting factors, whose corresponding cis-regulatory elements are significantly enriched, showed that both TF and TAG contain *MAX* (Table 4). The results of the enrichment analysis of the TF motifs surrounding the aberrant DNA methylation sites indicated that the MDSeqpos enrichment score of *MAX* was 124.265 (the default enrichment significance threshold was 0.00001, see Table 5). After the construction of the transcriptional regulation network, which was related to the aberrant DNA methylation sites, we found that *MAX* (degree = 126) was one of the most important network core genes (Figure 4). In the analysis in Figure 3, the *MAX* cis-regulatory element contains both HyperM sites and HypoM sites, and the HyperM was dominant. We speculated that *MAX* transcription would be inhibited and *MAX* mRNA would be expressed at a low level. Our real-time RT-PCR also detected the low expression of *MAX* mRNA (Figure 5). This indicates that during the long-term arsenic exposure process, the HypoM of *MYC* up-regulates *MYC* expression and the HyperM of *MAX* down-regulates *MAX* expression, thereby increasing the carcinogenesis tendency of hepatocytes.

In this study, based on the DNA methylation microarray analysis results under long-term arsenic exposure conditions and the results of the database search and literature search, we found that some genes adjacent to DNA methylation were related to the pathogenesis of liver cancer. The analysis of cis-regulatory elements and their trans-acting factors that are adjacent to DNA methylation sites can more accurately reflect the correlation between methylation and gene transcription and expression. We deduced that the HypoM of *MYC* and the HyperM of *MAX* may activate the pathogenesis of arsenic induced liver cancer.

MATERIALS AND METHODS

Cell culture

The normal human liver cell line L02 was purchased from the China Center for Type Culture Collection (CCTCC, Wuhan, China). The cells were cultured in

DMEM (high glucose) supplemented with 10% FBS at 37°C and 5% CO₂ in an incubator.

The medium for the arsenic exposure group was supplemented with a low concentration of arsenic trioxide, which was purchased from Yida Pharmaceutical Co. Ltd., Harbin Medical University, Harbin, China. The final concentration was 0.2 μM. The cells were split every 48 h, and a fresh arsenic-containing medium was used. After 3 months of culture, the cells were harvested for detection. The control group cells were cultured in DMEM medium only, and the cells were harvested for detection at the same time point. All experiments were performed in triplicate.

Approval for this study was obtained from the Committee for Medical Research Ethics of the First Affiliated Hospital of Harbin Medical University, which are the authority of research ethics in China.

Infinium® HumanMethylation450 BeadChip microarray

DNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany, Cat No. 69504), the DNA samples were quantified, and their quality was examined. The DNA samples from each group were mixed together for further detection. An Infinium® HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA, USA) was used for the transformation, amplification and labeling hybridization experiments. The procedures were conducted according to the Illumina Infinium® HumanMethylation450 BeadChip operating manual. Then, bioinformatics analyses were performed followed the flowchart as shown in Figure 7.

DNA methylation microarray data preprocessing

The MethyLumi package [34] was used to read the raw microarray data in .idat format. The Lumi package [34] was

used to complete the pretreatment operation for the Illumina Infinium DNA methylation microarray, which included dual-channel microarray fluorescence signal value imbalance adjustment (color imbalance adjustment, quantile based), background signal correction (background correction), and data normalization (quantile based). Afterwards, the beta value of each DNA methylation site was calculated to reflect the DNA methylation level at each probe site. The formula for the calculation of the beta value was as follows:

$$Beta_i = \frac{\max(y_{i, methyl}, 0)}{\max(y_{i, unmethyl}, 0) + \max(y_{i, methyl}, 0) + \alpha}$$

In which *i* represents the serial number of each probe on the chip, and α is the compensation value. Here, the default value of α was set to 0.

Detection of aberrant DNA methylation sites

Based on the pretreatment results of the microarray, the differences of the beta value of each DNA methylation site in the arsenic exposure group and in the control group were compared one by one. We ensured that the methylation sites, whose fold coefficient differences of DNA methylation levels between samples were not smaller than 1.5 (i.e., $|\log_2(\text{arsenic exposure group vs. control group})| \geq 0.58$), were significantly aberrant sites.

Annotation of genes adjacent to the aberrant DNA methylation sites

The region between 2 kb upstream of TSS (transcription start site) and 0.5 kb downstream of TSS of each gene was defined as the promoter region of the gene. Thus, combined with the coordinate information of the probe, we determined whether each probe was within the promoter region of the gene. Aberrant DNA methylation sites that met the criteria were kept, and the relevant gene

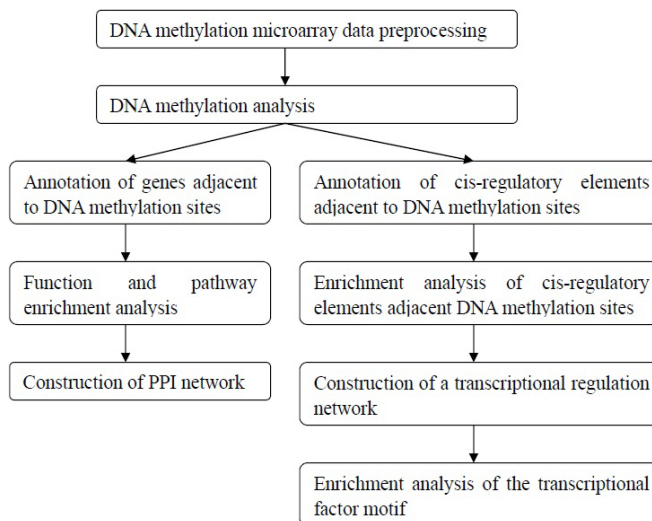


Figure 7: The bioinformatics analysis flowchart of the DNA methylation microarray detection results.

was defined as the adjacent gene of the aberrant DNA methylation sites.

Pathway enrichment analysis of the genes adjacent to the aberrant DNA methylation sites

The enrichment of pathways of the genes adjacent to the aberrant DNA methylation sites were identified based on the information from the KEGG pathway database [35]. During the analysis, the default significant threshold value of the hypergeometric test enrichment was set at 0.05, and the number of genes included in each significantly enriched KEGG pathway was not less than 2.

Screening of genes with specific functions

In combination with the TF data information recorded by TRANSFAC [36], further screening and annotation were performed on the genes adjacent to the aberrant DNA methylation sites to determine whether these genes had transcriptional regulation functions. Second, in combination with the tumor suppressor genes [37] and the TAG database [38], we further screened for all known oncogenes and tumor suppressor genes.

Construction of a sub-network of genes adjacent to aberrant DNA methylation sites based on PPI network

Based on the information provided by the String database [39], we collected other protein-encoding genes that can interact with these genes and established a PPI sub-network. Only those interacting protein pairs that had been verified by experiments, had been reported before, had previously undergone a co-expression analysis or had been recorded in a relevant database were selected as the input data for the sub-network construction. The PPI value of each interacting protein pair was not lower than 0.9 (PPI score was between 0 and 1; the PPI value was positively correlated with the PPI interaction pair credibility).

Annotation and enrichment analysis of cis-regulatory elements adjacent to aberrant DNA methylation sites and the construction of a transcriptional regulation network

First, based on the ENCODE database [40], information regarding the cognate cis-regulatory binding sites of all known human trans-acting factors was extracted. Second, to enhance the credibility of these cis-regulatory binding sites, we statistically analyzed the reproducibility of each binding site, and the cis-binding sites that showed up in at least two independent samples were selected and used for subsequent analysis. Then, in combination with relevant DNA methylation site data, we selected all the

cis-regulatory regions that overlapped with DNA methylation sites and defined these regions as the cis-regulatory elements adjacent to DNA methylation sites.

Based on the results from the above treatment, we annotated the known cis-regulatory sites that were located around the aberrant DNA methylation sites. In combination with Fisher's exact test (FET) enrichment analysis, we performed further screening to predict the transcriptional regulatory elements that were greatly affected by DNA methylation. Based on this information, we built a transcriptional regulatory network. The FET enrichment significance threshold was set to 0.00001.

Enrichment analysis of the TF motif of the aberrant DNA methylation sites

First, the aberrant DNA methylation site was used as the center and the region consisting of 50 bp upstream and downstream of the site was used as the detection range for the motif analysis. Then, in combination with each TF DNA binding motif properties that were recorded in the TRANSFAC [36] and JASPAR databases [41], we used the MDSeqpos tool to perform a TF motif enrichment analysis in the detection region. The default enrichment significance threshold was set at 0.00001.

Detection of *MYC*, *MAX* mRNA using real-time RT-PCR

Total RNA was extracted using the RNeasy Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNA quantity and quality were measured by NanoDrop ND-1000 and RNA integrity was assessed by standard denaturing agarose gel electrophoresis. Freshly prepared RNA (1mg) was reverse transcribed using the HiFi-MMLV cDNA reverse transcription kit (CWbiotech, Beijing, China, Cat No. CW0744) and RealSuper Mixture (with Rox) (CWbiotech, Beijing, China, Cat No. CW0767) according to the manufacturer's instructions. All reactions were performed on an ABI Prism 7500 Real-Time PCR system (Applied Biosystems, Foster City, USA). *MYC*, primers: forward (TCAAGAGGTGCCACGTCTCC), reverse (TCTTGGCAGCAGGATAGTCCTT), and probe (CAGCACAACACTACGCAGCGCCTCC). *MAX*, primers: forward (AGGTGGAGAGCGACGAAGAG), reverse (GTGCATTATGATGAGCCCGTTT), and probe (CCGAGGTTTCAATCTGCGGCTGAC). Relative gene expression was normalized to 1.0 of controls. All experiments were performed in triplicate.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (81370566). Thanks for the help of bioinformatics analysis from FengHe (Shanghai) Information Technology Co. Ltd, China.

REFERENCES

1. Chang Q, Chen B, Thakur C, Lu Y, Chen F. Arsenic-induced sub-lethal stress reprograms human bronchial epithelial cells to CD61⁻ cancer stem cells. *Oncotarget*. 2014; 5:1290–1303.
2. Yu MC, Yuan JM. Environmental factors and risk for hepatocellular carcinoma. *Gastroenterology*. 2004; 127:S72–S78.
3. Liu J, Waalkes MP. Liver is a target of arsenic carcinogenesis. *Toxicol Sci*. 2008; 105:24–32.
4. Lin HJ, Sung TI, Chen CY, Guo HR. Arsenic levels in drinking water and mortality of liver cancer in Taiwan. *J Hazard Mater*. 2013; 262:1132–1138.
5. Chen H, Li S, Liu J, Diwan BA, Barrett JC, Waalkes MP. Chronic inorganic arsenic exposure induces hepatic global and individual gene hypomethylation: implications for arsenic hepatocarcinogenesis. *Carcinogenesis*. 2004; 25:1779–1786.
6. Bashir S, Sharma Y, Irshad M, Nag TC, Tiwari M, Kabra M, Dogra TD. Arsenic induced apoptosis in rat liver following repeated 60 days exposure. *Toxicology*. 2006; 217:63–70.
7. Reichard JF, Schnekenburger M, Puga A. Long term low-dose arsenic exposure induces loss of DNA methylation. *Biochem Biophys Res Commun*. 2007; 352:188–192.
8. Deng J, Liang H, Ying G, Zhang R, Wang B, Yu J, Fan D, Hao X. Methylation of CpG sites in RNF180 DNA promoter prediction poor survival of gastric cancer. *Oncotarget*. 2014; 5:3173–3183.
9. Nishida N, Nishimura T, Nakai T, Chishina H, Arizumi T, Takita M, Kitai S, Yada N, Hagiwara S, Inoue T, Minami Y, Ueshima K, Sakurai T, et al. Genome-wide profiling of DNA methylation and tumor progression in human hepatocellular carcinoma. *Dig Dis*. 2014; 32:658–663.
10. Omura K, Uehara T, Morikawa Y, Hayashi H, Mitsumori K, Minami K, Kanki M, Yamada H, Ono A, Urushidani T. Comprehensive analysis of DNA methylation and gene expression of rat liver in a 2-stage hepatocarcinogenesis model. *J Toxicol Sci*. 2014; 39:837–848.
11. Brocato J, Costa M. Basic mechanics of DNA methylation and the unique landscape of the DNA methylome in metal-induced carcinogenesis. *Crit Rev Toxicol*. 2013; 43:493–514.
12. Majumdar S, Chanda S, Ganguli B, Mazumder DN, Lahiri S, Dasgupta UB. Arsenic exposure induces genomic hypermethylation. *Environ Toxicol*. 2010; 25:315–318.
13. Reichard JF, Puga A. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. *Epigenomics*. 2010; 2:87–104.
14. Tell G, Demple B. Base excision DNA repair and cancer. *Oncotarget*. 2015; 6:584–585.
15. Abbasi R, Ramroth H, Becher H, Dietz A, Schmezer P, Popanda O. Laryngeal cancer risk associated with smoking and alcohol consumption is modified by genetic polymorphisms in ERCC5, ERCC6 and RAD23B but not by polymorphisms in five other nucleotide excision repair genes. *Int J Cancer*. 2009; 125:1431–1439.
16. Paul S, Banerjee N, Chatterjee A, Sau TJ, Das JK, Mishra PK, Chakrabarti P, Bandyopadhyay A, Giri AK. Arsenic-induced promoter hypomethylation and over-expression of ERCC2 reduces DNA repair capacity in humans by non-disjunction of the ERCC2-Cdk7 complex. *Metalomics*. 2014; 6:864–873.
17. Mazzocca A, Antonaci S, Giannelli G. The TGF- β signaling pathway as a pharmacological target in a hepatocellular carcinoma. *Curr Pharm Des*. 2012; 18:4148–4154.
18. Takigawa Y, Brown AM. Wnt signaling in liver cancer. *Curr Drug Targets*. 2008; 9:1013–1024.
19. Jang KY, Noh SJ, Lehwald N, Tao GZ, Bellovin DI, Park HS, Moon WS, Felsner DW, Sylvester KG. SIRT1 and c-Myc promote liver tumor cell survival and predict poor survival of human hepatocellular carcinomas. *PLoS One*. 2012; 7:e45119.
20. Jiang Y, Zhang SH, Han GQ, Qin CY. Interaction of Pcd4 with eIF4E inhibits the metastatic potential of hepatocellular carcinoma. *Biomed Pharmacother*. 2010; 64:424–429.
21. LU JW, Lin YM, Chang JG, Yeh KT, Chen RM, Tsai JJ, Su WW, Hu RM. Clinical implications of deregulated CDK4 and Cyclin D1 expression in patients with human hepatocellular carcinoma. *Med Oncol*. 2013; 30:379.
22. Wang Y, Deng B, Tang W, Liu T, Shen X. TGF- β 1 secreted by hepatocellular carcinoma induces the expression of the Foxp3 gene and suppresses antitumor immunity in the tumor microenvironment. *Dig Dis Sci*. 2013; 58:1644–1652.
23. Sato M, Rodriguez-Barrueco R, Yu J, Do C, Silva JM, Gautier J. MYC is a critical target of FBXW7. *Oncotarget*. 2015; 6:3292–3305.
24. Bindra RS, Glazer PM. Co-repression of mismatch repair gene expression by hypoxia in cancer cells: role of the Myc/Max network. *Cancer Lett*. 2007; 252:93–103.
25. Walker W, Zhou ZQ, Ota S, Wynshaw-Boris A, Hurlin PJ. Mnt-Max to Myc-Max complex switching regulates cell cycle entry. *J Cell Biol*. 2005; 169:405–413.
26. Ecevit O, Khan MA, Goss DJ. Kinetic analysis of the interaction of b/HLH/Z transcription factors Myc, Max, and Mad with cognate DNA. *Biochemistry*. 2010; 49:2627–2635.
27. Adhikary S, Eilers M. Transcriptional regulation and transformation by Myc proteins. *Nat Rev Mol Cell Bio*. 2005; 6:635–645.
28. Raffener P, Röck R, Schraffl A, Hartl M, Hartl JR, DJanda K, Vogt PK, Stefan E, Bister K. *In vivo* quantification and perturbation of Myc-Max interactions and the impact on oncogenic potential. *Oncotarget*. 2014; 5:8869–8878.
29. Frenzel A, Zirath H, Vita M, Albiñ A, Henriksson MA. Identification of cytotoxic drugs that selectively target

- tumor cells with MYC overexpression. *PLoS One*. 2011; 6:e27988.
30. Cascon A, Robledo M. MAX and MYC: a heritable breakup. *Cancer Res*. 2012; 72:3119–3124.
 31. Albanus RD, Juliani Siqueira Dalmolin R, Alves Castro MA, Augusto de Bittencourt Pasquali M, de Miranda Ramos V, Pens Gelain D, Fonseca Moreira JC. Reverse engineering the neuroblastoma regulatory network uncovers MAX as one of the master regulators of tumor progression. *PLoS One*. 2013; 8:e82457.
 32. Iñaki Comino-Méndez, Francisco J Gracia-Aznárez, Francesca Schiavi, Iñigo Landa, Luis J Leandro-García, Rocío Letón, Emiliano Honrado, Rocío Ramos-Medina, Daniela Caronia, Guillermo Pita, Alvaro Gómez-Graña, Aguirre A de Cubas, Lucía Inglada-Pérez, Agnieszka Maliszewska, et al. Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet*. 2011; 43:663–667.
 33. Yang L, Tan X, Yang Z, Li Q, Miao X. Expression of RAD51 and MAX in pancreatic cancer rats. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*. 2010; 35:146–151.
 34. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010; 11:587–596.
 35. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000; 28:27–30.
 36. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003; 31:374–378.
 37. Zhao M, Sun J, Zhao Z. TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res*. 2013; 41:D970–D976.
 38. Chen JS, Hung WS, Chan HH, Tsai SJ, Sun HS. In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. *Bioinformatics*. 2013; 29:420–427.
 39. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41:D808–D815.
 40. Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, Cline MS, Karolchik D, Barber GP, Clawson H, Diekhans M, Fujita PA, Goldman M, et al. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res*. 2012; 40:D912–D917.
 41. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2014; 42:D142–D147.