

## RESEARCH ARTICLE

# Computational prediction and functional analysis of arsenic-binding proteins in human cells

Shichao Pang<sup>1</sup>, Junchen Yang<sup>2</sup>, Yilei Zhao<sup>2</sup>, Yixue Li<sup>2,3</sup>, Jingfang Wang<sup>2,4,\*</sup>

<sup>1</sup> Department of Statistics, School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>2</sup> Department of Bioinformatics and Biostatistics, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>3</sup> Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>4</sup> Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

\* Correspondence: jfwang8113@sjtu.edu.cn

Received August 31, 2018; Revised November 13, 2018; Accepted January 4, 2019

**Background:** Arsenic has a broad anti-cancer ability against hematologic malignancies and solid tumors. To systematically understand the biological functions of arsenic, we need to identify arsenic-binding proteins in human cells. However, due to lack of effective theoretical tools and experimental methods, only a few arsenic-binding proteins have been identified.

**Methods:** Based on the crystal structure of ArsM, we generated a single mutation free energy profile for arsenic binding using free energy perturbation methods. Multiple validations provide an indication that our computational model has the ability to predict arsenic-binding proteins with desirable accuracy. We subsequently apply this computational model to scan the entire human genome to identify all the potential arsenic-binding proteins.

**Results:** The computationally predicted arsenic-binding proteins show a wide range of biological functions, especially in the signaling transduction pathways. In the signaling transduction pathways, arsenic directly binds to the key factors (e.g., Notch receptors, Notch ligands, Wnt family proteins, TGF-beta, and their interacting proteins) and results in significant inhibitions on their enzymatic activities, further having a crucial impact on the related signaling pathways.

**Conclusions:** Arsenic has a significant impact on signaling transduction in cells. Arsenic binding to proteins can lead to dysfunctions of the target proteins, having crucial impacts on both signaling pathway and gene transcription. We hope that the computationally predicted arsenic-binding proteins and the functional analysis can provide a novel insight into the biological functions of arsenic, revealing a mechanism for the broad anti-cancer of arsenic.

**Keywords:** arsenic-binding proteins; free energy profile; position-specific score matrix; signaling transduction

**Author summary:** In the current study, we reported a computational approach to predict arsenic-binding sites in proteins using a single mutation free energy profile. Using MALDI-MS and proteome microarray data as a testing set, this computational approach shows an over 90% accuracy. Thus, we used this approach to scan the entire human proteomics to identify all the potential arsenic-binding proteins. The predicted arsenic-binding proteins in human cells show a wide range of biological functions, and have significant impacts on both signaling transduction and gene transcription.

## INTRODUCTION

Arsenic and its derivatives exhibit two sides for human health. Chronic exposure to arsenic (especially inorganic arsenic compounds) can cause a number of diseases, *e.g.*, cancers, diabetes, as well as hepatic and digestive diseases [1]. Thus, the International Agency for Research on Cancer (IARC) and the US Environmental Protection Agency (EPA) have defined arsenic as a human carcinogen. Additionally, the US Agency for Toxic Substances and Disease Registry (ATSDR) has ranked arsenic as No. 1 on the Substance Priority List for many years. Arsenic and its derivatives also have significant medical applications, especially in treating hematologic diseases [2]. For instance, the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have approved arsenic trioxide combined with all-trans retinoic acid for the treatment of acute promyelocytic leukemia (APL). Because the combination use of arsenic trioxide and all-trans retinoic acid for APL patients shows a superior five-year survival rate [3]. In addition to the excellent clinical performance on APL, arsenic also exhibits a broad-spectrum anti-cancer ability against both hematologic malignancies and solid tumors [4].

To understand the mechanism for the anti-cancer ability of arsenic and its derivatives, we start from the arsenic-binding proteins in human cells. By now, only several arsenic-binding proteins are experimentally confirmed in human cells. Promyelocytic leukemia protein (PML) is a tumor suppressor protein in APL. Trivalent arsenic directly binds the RING-finger domain of PML and PML-retinoic acid receptor  $\alpha$  (RAR $\alpha$ ) fusion protein in APL cells to induce PML oligomerization, enhancing the interaction between PML and UBC9, further accelerating the ubiquitination and degradation of PML and PML-RAR $\alpha$  fusion protein [5]. The ubiquitin E3 ligases c-CBL (Casitas B-lineage lymphoma) is another experimentally confirmed arsenic-binding protein. Similar to PML, arsenic directly binds the RING-finger domain of c-CBL to inhibit its self-ubiquitination and degradation, up-regulating the protein expression of c-CBL [6]. In addition to PML and c-CBL, arsenic has a significant impact on the thioredoxin system by directly binding to the thioredoxin reductase (TrxR) [7].

The identification of arsenic-binding proteins is limited by lacking effective experimental methods and theoretical tools. Position specific score matrix (PSSM) and machine learning algorithms have been successfully used to predict the binding sites of the co-factors of proteins and enzymes, *e.g.*, metal ions and acid radical ions [8]. These computational methods need to utilize a large amount of experimental data as training sets to generate accurate computational models. For instance, the predic-

tion of calcium-binding proteins needs 179 proteins with 329 experimentally validated calcium-binding sites as a training set to reach an accuracy of more than 90% with PSSM and machine learning algorithms [9]. However, not enough experimentally validated arsenic-binding proteins and arsenic-binding sites are available for prediction. Only four human proteins with arsenic-binding sites have been experimentally validated (*e.g.*, mass spectrum data), which is not enough for a training set for machine learning.

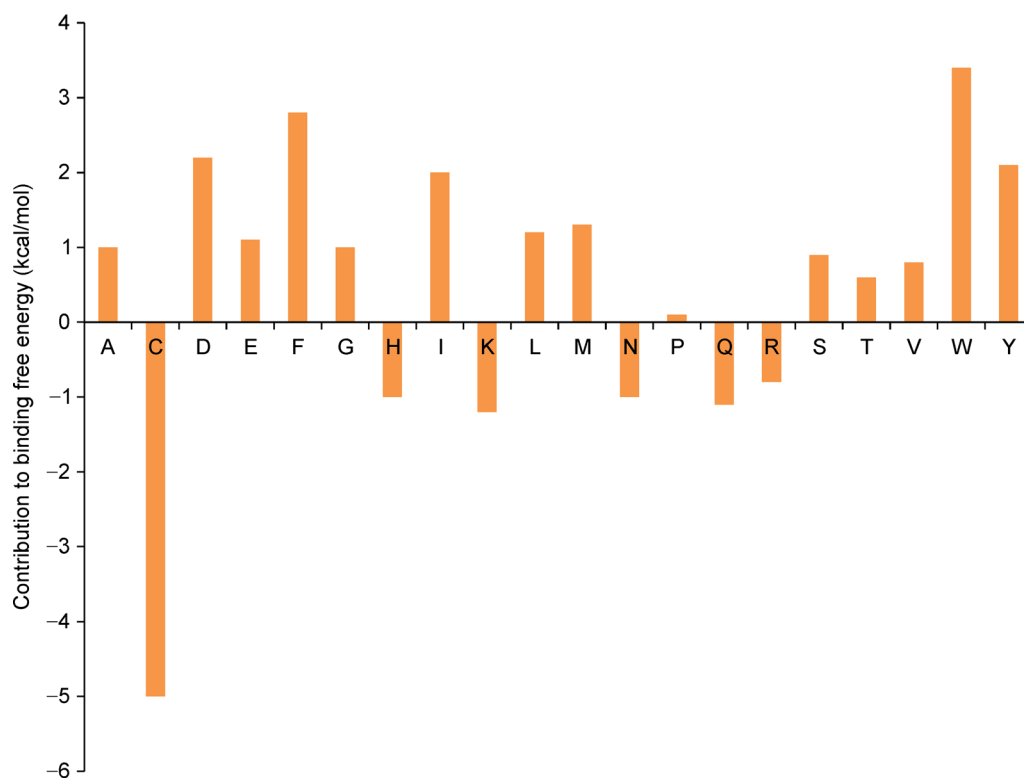
To address this issue, we used binding free energies for the single mutation to replace the multiple sequence alignment in the PSSM generation. Firstly, we constructed a structural model for arsenic binding based on the crystal structure of the arsenic methyltransferase ArsM and generated a number of single mutations based on the structural model for arsenic binding. Subsequently, we performed molecular dynamics simulations on these mutations and calculated the free energy profile for arsenic binding using the free energy perturbation approach. The analysis of molecular dynamics simulations and binding free energies for arsenic binding can provide useful information to understand the molecular mechanism for the binding specificity of arsenic. Based on these calculations, we predicted all the potential arsenic-binding sites and arsenic-binding proteins in human cells, further performing systematical analysis on these arsenic-binding proteins and providing a novel insight into the biological functions of arsenic in human cells.

## RESULTS AND DISCUSSION

### Molecular basis for the arsenic-binding specificity

As trivalent arsenic shows a high affinity to sulfhydryl groups, arsenic is inclined to bind cysteine residues in proteins. As shown in Figure 1, the residue contribution to the free energy for arsenic binding indicates that the cysteine residue shows the most favorable binding free energy for arsenic ( $-5.0$  kcal/mol). In addition to cysteine residues, histidine ( $-1.0$  kcal/mol), lysine ( $-1.2$  kcal/mol), asparagine ( $-1.0$  kcal/mol), glutamine ( $-1.1$  kcal/mol), as well as arginine ( $-0.8$  kcal/mol) residues are preferred for arsenic binding. All the residues favored for arsenic binding are polar amino acids and can be classified into two types: non-ionized amino acids (cysteine, asparagine, and glutamine) and alkaline (histidine, lysine, and arginine) amino acids. These polar amino acids can provide an additional lone pair electrons for arsenic binding.

Both experimental and computational evidence indicate that arsenic specifically binds to reduced cysteine

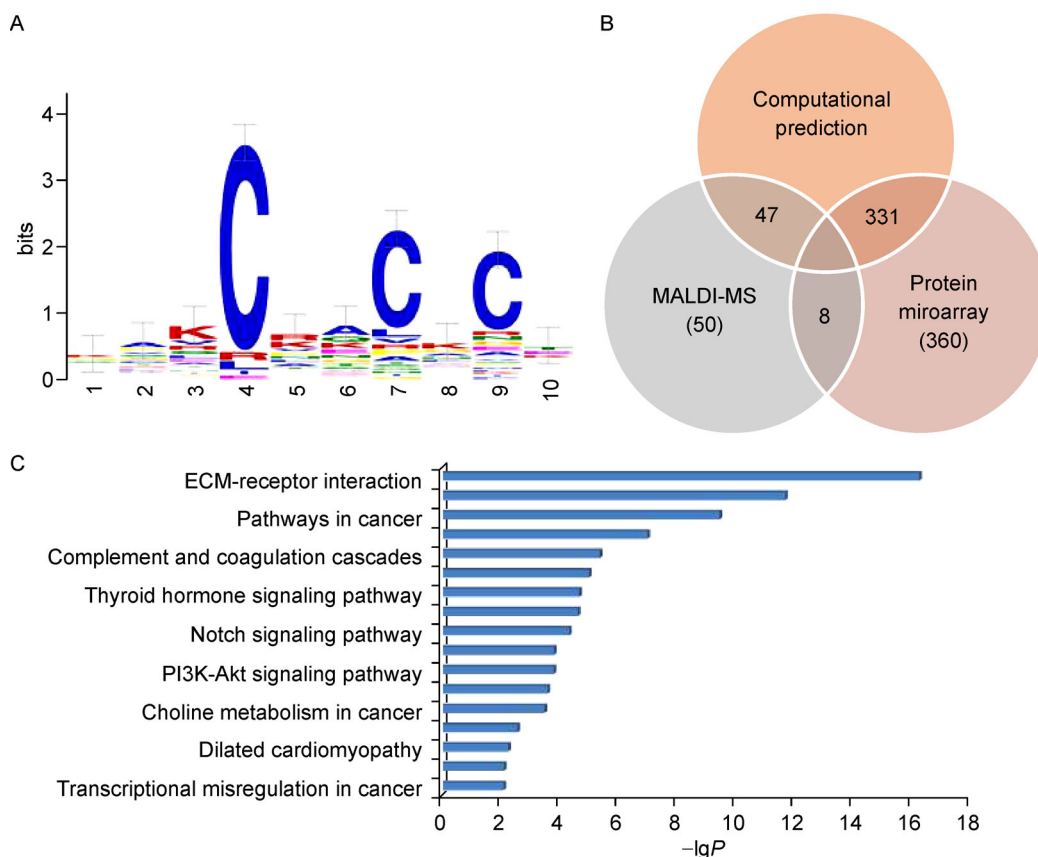


**Figure 1. Residue contribution to arsenic-binding free energy.** Residues are represented in single letter abbreviations: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine. Among all the residues, cysteine, histidine, lysine, asparagine, glutamine, and arginine are preferred for arsenic-binding in proteins with a binding free energy contribution of  $-5.0$  kcal/mol,  $-1.0$  kcal/mol,  $-1.2$  kcal/mol,  $-1.0$  kcal/mol,  $-1.1$  kcal/mol, and  $-0.8$  kcal/mol, respectively.

residues. How does arsenic choose the specific cysteine to bind? To answer this question, we generated ten single-cysteine motifs and four double-cysteine motifs based on the template structure (VISNCVCNLS), and calculated their free energies for arsenic binding. As a result, the double-cysteine motifs exhibit more favorable free energies for arsenic binding than the single-cysteine motif (Supplementary Table S1). This observation provides an indication that adjacent cysteine residues are preferred for arsenic binding. To figure out the favorable binding motif for arsenic, we constructed a single mutation free energy profile for arsenic binding using the free energy perturbation approach. Based on the single mutation free energy profile for arsenic binding, we generated a PSSM for arsenic binding (as shown in Table 1), which can be further used to predict arsenic-binding sites and arsenic-binding proteins. After analyzing the single mutation free energy profile, we obtained a consensus sequence motif xxxCxxCx for arsenic binding (Figure 2A).

### Computational validation for the arsenic-binding PSSM

To examine the predicting ability and accuracy of the arsenic-binding PSSM, we collect nine proteins with experimentally verified arsenic-binding sites for the computational validation. For detailed information, please see the Supplementary Table S2. A slidable window with a width of 10 AA is applied to slide along the local sequence of the collected proteins, to generate a number of 10-AA segments. The 10-AA segments that contain the experimentally verified binding sites for arsenic are considered as the arsenic-binding motifs. Conversely, those without experimentally verified binding sites for arsenic are considered as non-binding motifs. The independent test set is composed of the arsenic-binding motifs and non-binding motifs mentioned above, and the single mutation free energy profile based arsenic-binding PSSM is subsequently applied to the test set for assessing



**Figure 2. Functional analysis for the computationally predicted arsenic-binding proteins.** (A) Consensus sequence motif for arsenic binding based on the single mutation free energy profile PSSM. (B) Validation results using MALDI-MS and protein microarray datasets. (C) Top pathways enriching arsenic-binding proteins.

its predicting ability and accuracy. As a result, the single mutation free energy profile based arsenic-binding PSSM shows 88.7% sensitivity, 90.6% specificity and 92.3% accuracy.

To further check the predicting ability of the single mutation free energy profile based arsenic-binding PSSM, we also used two independent datasets (named as MALDI-MS dataset [10] and protein microarray dataset [11] respectively) for validation. For detailed information, please see Supplementary Table S3 and Table S4. Only eight MALDI-MS proteins are overlapped with protein microarray datasets. The arsenic-binding PSSM is subsequently carried out on the proteins in MALDI-MS and protein microarray datasets. As a result, 94.0% proteins in the MALDI-MS dataset and 91.9% in the protein microarray dataset are successfully predicted as the arsenic-binding proteins (Figure 2B). These findings suggest that our computational model can repeat the majority of previous experimental results. Fairly reliable prediction of identifying arsenic-binding proteins is possible by usage of the tool.

### Identifying arsenic-binding proteins in the human genome

Using the single mutation free energy profile based arsenic-binding PSSM, we scan the entire human genome to search for arsenic-binding proteins in human cells. In total, 11,315 arsenic-binding sites are identified with a threshold of q-value less than 0.05. Together with the chromosome location information, the identified arsenic-binding sites are mapped into genes using the BioMart community portal [12]. After mapping these computational predictions into the human genome, nearly 2,800 proteins are finally discovered as arsenic-binding proteins. Additional function analysis indicates that these arsenic-binding proteins in the human genome are mainly enriched in signaling transduction and related pathways, especially in Notch signaling, Wnt, and TGF-beta signaling pathways (Figure 2C).

In the Notch signaling pathway, 16 arsenic-binding proteins are enriched with a significant P-value of  $8.6 \times 10^{-35}$  (Figure 3 and Supplementary Table S5). The Notch

**Table 1. Single mutation free energy profile for arsenic binding**

Res.	Position									
	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	P <sub>8</sub>	P <sub>9</sub>	P <sub>10</sub>
A	0.7	-1.0	0.3	3.9	-0.1	-1.3	0.8	-0.4	0.8	2.6
C	-0.4	-1.4	0.5	-5.2	4.8	4.8	-4.5	4.8	-4.4	4.8
D	-0.1	2.7	1.1	5.4	1.1	0.3	5.0	2.7	3.7	0.3
E	0.6	0.6	0.2	4.2	0.6	-0.4	4.6	0.6	0.8	-0.7
F	-0.6	4.1	4.1	4.7	4.1	4.1	2.9	4.1	0.8	-0.6
G	-0.6	-0.1	0.8	5.6	1.6	-0.1	1.8	-0.4	0.9	0.3
H	-0.8	-0.8	2.3	4.7	2.3	2.3	3.6	-1.3	3.0	-1.7
I	1.3	3.9	0.4	3.3	0.4	3.9	0.7	0.4	0.9	3.9
K	-1.0	0.0	-1.9	5.7	-1.2	-0.8	4.2	-1.4	1.4	-0.6
L	2.0	0.7	0.3	3.6	0.3	2.0	0.4	0.3	1.7	0.7
M	-0.6	3.2	0.3	4.1	0.3	3.2	2.2	0.3	0.2	-0.6
N	-0.7	0.5	-0.2	4.7	-1.1	-1.1	0.8	-1.1	-0.1	-0.2
P	0.5	-0.8	3.8	5.6	-0.5	1.3	0.5	0.5	1.4	-0.1
Q	0.0	0.0	0.0	5.3	0.0	-1.4	1.1	0.7	2.9	-0.5
R	-0.6	-0.3	-0.9	3.0	-1.6	0.2	0.4	-0.3	0.3	0.9
S	-0.6	-0.3	-0.3	4.7	0.5	0.0	1.5	0.5	1.4	1.2
T	0.9	-0.3	0.2	3.8	-0.3	-0.3	1.0	0.2	2.6	-1.4
V	0.8	-0.1	-0.4	4.8	-0.1	0.8	0.4	0.2	1.2	0.2
W	3.8	3.8	3.8	5.0	3.8	-0.9	3.3	3.8	3.4	3.8
Y	3.3	-0.4	3.3	5.2	3.3	3.3	3.2	-0.4	0.4	-0.4

Residues are represented in single letter abbreviations: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

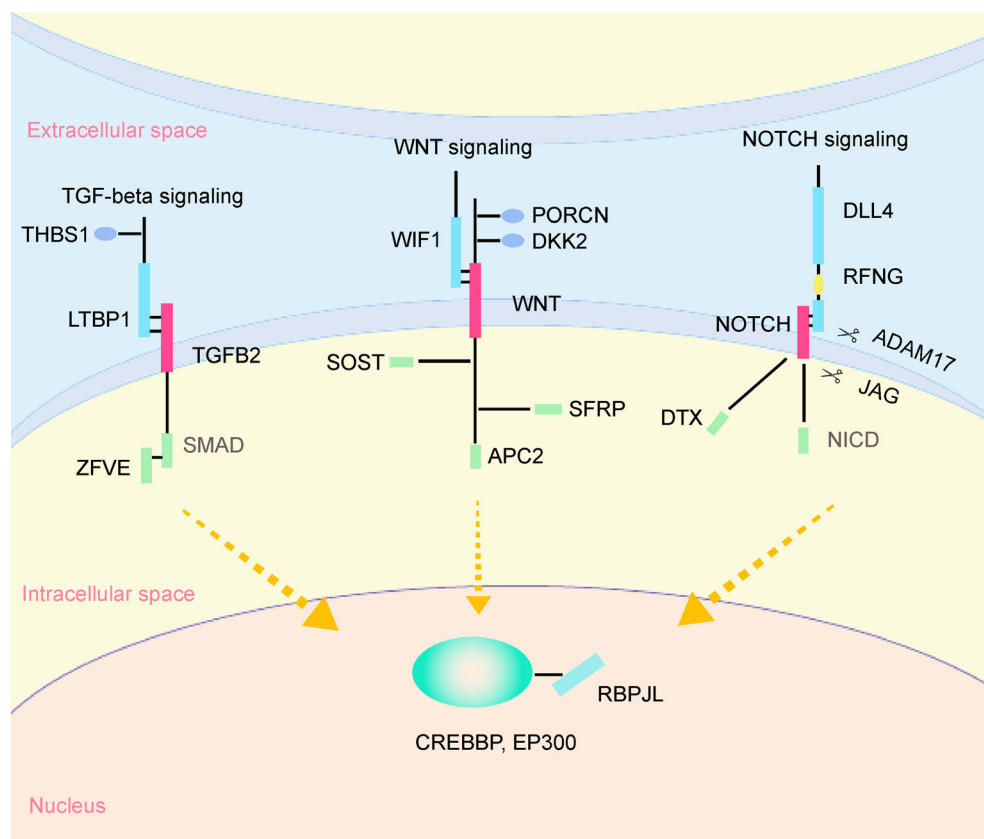
signaling is essential for many fundamental cellular processes, such as cell proliferation and apoptosis, and its dysfunction can result in many diseases, *e.g.*, degenerative diseases, fibrosis, and cancers. Our computational prediction indicates that arsenic directly binds to Notch receptors (Notch1–4) and Notch ligands (Delta-like proteins and Jagged families). Additionally, ADAM proteases, and transcription factor CSL, which is regulated by the released intracellular domain of Notch receptors, are also detected to have arsenic-binding sites. Thus, arsenic binding to these proteins will lead to the dysfunction of Notch signaling, further affecting cell proliferation and apoptosis. This opinion is also supported by the experimental results in human keratinocyte cell [13] and glioma cell lines [14].

In addition to Notch signaling pathway, arsenic-binding proteins are found to enrich in the Wnt and TGF-beta signaling pathways (Figure 3 and Supplementary Table S5). In these signaling pathways, arsenic directly binds to Wnt family proteins, TGF-beta, as well as their regulators, such as porcupine homolog (PORCN), secreted frizzled related proteins (SFRPs), Wnt inhibitory factors (WIFs), latent transforming growth factor beta binding proteins (LTBPs), and thrombospondins (THBSs). Besides, sclerostin (SOST) and dickkopf Wnt signaling pathway inhibitors (DKKs) are also detected to have arsenic-

binding sites. These proteins are regarded to have significant impacts on the Wnt signaling pathway by regulating the expression and biological functions of Wnt receptors (frizzled class receptor FZD and LDL receptors). Thus, arsenic binding to these proteins is able to induce significant inhibitions on their biological activities, further regulating the corresponding signaling pathway, which is partly supported by mouse model studies [15].

Transcriptional regulation is an essential process in all living organisms. The dysfunction of transcriptional regulation can lead to a variety of diseases including cancers. Many biological studies confirm that arsenic is able to bind the RING finger domain of transcriptional factors [16]. In the current study, functional analysis shows that a number of arsenic-binding proteins are transcription factors, enriched in the transcriptional misregulation pathway. For instance, a series of zinc finger and BTB domain-containing proteins (*e.g.*, ZBTB16, ZBTB17, and ZBTB27) are identified as arsenic-binding proteins. These proteins contain C<sub>2</sub>H<sub>2</sub> zinc finger domain and BTB domain (also known as the POZ domain), functioning as transcriptional factors in many diseases, *e.g.*, leukemia and neuroblastoma [17].

In addition to these transcription factors, some significant fusion proteins, which are considered to play crucial roles in cancers, are also identified to have arsenic



**Figure 3. Arsenic-binding proteins in the Notch, Wnt, and TGF-beta signaling pathways.** Notch receptors, Notch ligands, Wnt family proteins, TGF-beta, as well as their interacting proteins are detected to have arsenic-binding sites. Arsenic binding to these proteins is able to induce significant inhibitions on their biological activities, regulating the corresponding signaling transduction.

binding sites. Among these fusion proteins, PML-RAR $\alpha$  and ZBTB16-RAR $\alpha$  fusion proteins are essential in APL, both of which are predicted to have arsenic binding sites by the arsenic PSSM. PML-RAR $\alpha$  fusion protein has been demonstrated to be a direct target for arsenic binding in APL cells, having a significant impact on its ubiquitination and degradation, to achieve an effective strategy against APL. ZBTB16-RAR $\alpha$  fusion protein is a transcriptional repressor to recruit the SMRT/N-CoR deacetylase transcriptional repression complex [18]. The binding of arsenic causes dysfunctions of ZBTB16-RAR $\alpha$  fusion protein, going against APL. Another interesting fusion protein having arsenic binding sites is Tmprss2-Kcnnh2 fusion protein. This fusion protein can induce the activation of the androgen receptor pathway and prostate cancer [19]. If Arsenic binding could accelerate the degradation of Tmprss2-Kcnnh2 fusion protein, it would be a new strategy to treat prostate cancer.

## CONCLUSION

In summary, we developed a single mutation free energy profile for arsenic binding. Based on this binding energy

profile, we generated a position-specific score matrix for identifying arsenic-binding proteins. Both computational validation and experimental evidence show that our computational model is able to predict arsenic-binding proteins with satisfying accuracy. Thus, we scanned the human genome to identify all the potential arsenic-binding proteins. Additional functional analysis indicates that the arsenic-binding proteins in the human genome show a wide range of biological functions, especially in signaling conduction and related pathways. Hopefully, these findings can provide a genomic insight for in-depth understanding of the effective ability of arsenic against malignancies and solid tumors.

## MATERIALS AND METHODS

### The structural model for arsenic-binding

The structural model for arsenic-binding was constructed based on the crystal structure of arsenic methyltransferase (ArsM) from *Cyanidioschyzon* sp. 5508 (RCSB PDB entry 4 fsd) [20]. ArsM is a homolog of human AS3MT that catalyzes transfer to the methyl group of S-

adenosylmethionine (SAM) to trivalent arsenic. It is also an important methyltransferase for arsenic biomethylation. The crystal water molecules and non-polar hydrogen atoms in the crystal structure 4 fsd were removed [21]. pKa values for each residue were calculated based on a Poisson-Boltzmann solver with a dielectric constant of 4.0 [22]. Hydrogens were subsequently added based on the computational pKa values using H++ web server [23]. The structural template was then solvated in a rectangular simulation box with explicit TIP3P water molecules, and subjected to steepest descent energy minimization for 5,000 steps, followed by conjugate gradient for the next 5,000 steps.

### Estimating arsenic-binding free energy

Amber force field parameters were used to parameterize the structural models of arsenic-binding [24]. The force field parameters of arsenic were obtained from the previous density function calculation of arsenic compounds [25]. The free energy perturbation (FEP) approach [26] was used to estimate arsenic-binding free energies. In the FEP calculations, the free energy distinction between the arsenic binding state and the free state is expressed by the equation:

$$\Delta G = -k_B T \ln \langle \exp\left(-\frac{H_c(r,p) - H_f(r,p)}{k_B T}\right) \rangle_f$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature, and  $H_c(r,p)$  and  $H_f(r,p)$  are the Hamiltonians characteristic of the arsenic binding state and free state, respectively.  $\langle \dots \rangle_f$  denotes an ensemble average over configurations representative of the arsenic free state. Molecular dynamics simulations with the periodic boundary condition and NPT ensemble using NAMD package [27] were used to generate the configurations for arsenic binding as well as free states. The weak-coupling algorithm [28] and Berendsen thermostat [29] were applied to maintain the simulation system at a constant temperature and pressure. The SHAKE algorithm [30] was used to constrain all the chemical bonds, and atom velocities for start-up runs were obtained according to the Maxwell distribution at 310 K. The electrostatic interaction was treated by the PME algorithm [31] with an interpolation order of 4.0 and a grid spacing of 0.12 nm. The van der Waals interaction was computed using a cut-off of 12 angstroms. All the molecular dynamics simulations were performed with a time step of 2 fs, and the simulation coordinates were saved every 1 ps.

### Generating binding free energy profiles for single mutations

Based on the template model of arsenic-binding mentioned above, we abstracted a 10-amino-acid (10-AA)

segment (VISNCVCNLS) as the template structure for arsenic-binding. Subsequently, we mutated each residue of the template structure into the other 19 residues to generate a number of single mutation segments. Arsenic-binding free energies for each single mutation segment were then calculated using the free energy perturbation approach mentioned above. The binding free energy profile for single mutations was subsequently converted into a  $10 \times 20$  matrix, representing the free energy distinction for arsenic-binding from the 10-AA segment (VISNCVCNLS) in the template structure. This matrix can be also used as a position-specific scoring function  $\sum_i^{10} M_{S,i} \delta(S, S_i)$ .  $M_{S,i}$  is the score of the residue  $S$  at the  $i^{\text{th}}$  position in the binding free energy profile for single mutations and  $S_i$  is the residue at the  $i^{\text{th}}$  position of the 10-AA segment.  $\delta$  is represented for the binding free energy distinction for arsenic-binding from the 10-AA segment (VISNCVCNLS) in the template structure.

### SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.1007/s40484-019-0169-6>.

### ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (Nos. 2016YFC0901704 and 2017YFA0505500), National High-Tech R&D Program (863 Program, No. 2015AA020105), the National Natural Science Foundation of China (Nos. 21377085 and 31770070), MOE New Century Excellent Talents in University (No. NCET-12-0354), and SJTU Med-Eng Joint Program (No. YG2016MS33) for financial supports.

### COMPLIANCE WITH ETHICS GUIDELINES

The authors Shichao Pang, Junchen Yang, Yilei Zhao, Yixue Li and Jingfang Wang declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

### REFERENCES

- Hong, Y. S., Song, K. H. and Chung, J. Y. (2014) Health effects of chronic arsenic exposure. *J. Prev. Med. Public Health*, 47, 245–252
- List, A., Beran, M., DiPersio, J., Slack, J., Vey, N., Rosenfeld, C. S. and Greenberg, P. (2003) Opportunities for Trisenox (arsenic trioxide) in the treatment of myelodysplastic syndromes. *Leukemia*, 17, 1499–1507
- Burnett, A. K., Russell, N. H., Hills, R. K., Bowen, D., Kell, J., Knapper, S., Morgan, Y. G., Lok, J., Grech, A., Jones, G., *et al.* (2015) Arsenic trioxide and all-trans retinoic acid treatment for acute promyelocytic leukaemia in all risk groups (AML17): results of a randomised, controlled, phase 3 trial. *Lancet Oncol.*, 16, 1295–1305
- Hoonjan, M., Jadhav, V. and Bhatt, P. (2018) Arsenic trioxide:

- insights into its evolution to an anticancer agent. *J. Biol. Inorg. Chem.*, 23, 313–329
- Zhang, X. W., Yan, X. J., Zhou, Z. R., Yang, F. F., Wu, Z. Y., Sun, H. B., Liang, W. X., Song, A. X., Lallemand-Breitenbach, V., Jeanne, M., *et al.* (2010) Arsenic trioxide controls the fate of the PML-RAR $\alpha$  oncoprotein by directly binding PML. *Science*, 328, 240–243
  - Mao, J. H., Sun, X. Y., Liu, J. X., Zhang, Q. Y., Liu, P., Huang, Q. H., Li, K. K., Chen, Q., Chen, Z. and Chen, S. J. (2010) As4S4 targets RING-type E3 ligase c-CBL to induce degradation of BCR-ABL in chronic myelogenous leukemia. *Proc. Natl. Acad. Sci. USA*, 107, 21683–21688
  - Lu, J., Chew, E. H. and Holmgren, A. (2007) Targeting thioredoxin reductase is a basis for cancer therapy by arsenic trioxide. *Proc. Natl. Acad. Sci. USA*, 104, 12288–12293
  - Hu, X., Dong, Q., Yang, J. and Zhang, Y. (2016) Recognizing metal and acid radical ion-binding sites by integrating *ab initio* modeling with template-based transferals. *Bioinformatics*, 32, 3260–3269
  - Yang, J., Roy, A. and Zhang, Y. (2013) BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.*, 41, D1096–D1103
  - Ge, F., Lu, X. P., Zeng, H. L., He, Q. Y., Xiong, S., Jin, L. and He, Q. Y. (2009) Proteomic and functional analyses reveal a dual molecular mechanism underlying arsenic-induced apoptosis in human multiple myeloma cells. *J. Proteome Res.*, 8, 3006–3019
  - Zhang, H. N., Yang, L., Ling, J. Y., Czajkowsky, D. M., Wang, J. F., Zhang, X. W., Zhou, Y. M., Ge, F., Yang, M. K., Xiong, Q., *et al.* (2015) Systematic identification of arsenic-binding proteins reveals that hexokinase-2 is inhibited by arsenic. *Proc. Natl. Acad. Sci. USA*, 112, 15084–15089
  - Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M. H., Baldock, R., Barbiera, G., *et al.* (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.*, 43, W589–W598
  - Zhen, Y., Zhao, S., Li, Q., Li, Y. and Kawamoto, K. (2010) Arsenic trioxide-mediated Notch pathway inhibition depletes the cancer stem-like cell population in gliomas. *Cancer Lett.*, 292, 64–72
  - Cialfi, S., Palermo, R., Manca, S., De Blasio, C., Vargas Romero, P., Checquolo, S., Bellavia, D., Uccelletti, D., Saliola, M., D'Alessandro, A., *et al.* (2014) Loss of Notch1-dependent p21 (Waf1/Cip1) expression influences the Notch1 outcome in tumorigenesis. *Cell Cycle*, 13, 2046–2055
  - Neumann, J. E., Wefers, A. K., Lambo, S., Bianchi, E., Bockstaller, M., Dorostkar, M. M., Meister, V., Schindler, P., Korshunov, A., von Hoff, K., *et al.* (2017) A mouse model for embryonal tumors with multilayered rosettes uncovers the therapeutic potential of Sonic-hedgehog inhibitors. *Nat. Med.*, 23, 1191–1202
  - Zhang, F., Paramasivam, M., Cai, Q., Dai, X., Wang, P., Lin, K., Song, J., Seidman, M. M. and Wang, Y. (2014) Arsenite binds to the RING finger domains of RNF20-RNF40 histone E3 ubiquitin ligase and inhibits DNA double-strand break repair. *J. Am. Chem. Soc.*, 136, 12884–12887
  - Furukawa, M., He, Y. J., Borchers, C. and Xiong, Y. (2003) Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat. Cell Biol.*, 5, 1001–1007
  - Girard, N., Tremblay, M., Humbert, M., Grondin, B., Haman, A., Labrecque, J., Chen, B., Chen, Z., Chen, S. J. and Hoang, T. (2013) RAR $\alpha$ -PLZF oncogene inhibits C/EBP $\alpha$  function in myeloid cells. *Proc. Natl. Acad. Sci. USA*, 110, 13522–13527
  - Bose, R., Karthaus, W. R., Armenia, J., Abida, W., Iaquinta, P. J., Zhang, Z., Wongvipat, J., Wasmuth, E. V., Shah, N., Sullivan, P. S., *et al.* (2017) ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. *Nature*, 546, 671–675
  - Ajees, A. A., Marapakala, K., Packianathan, C., Sankaran, B. and Rosen, B. P. (2012) Structure of an As(III) S-adenosylmethionine methyltransferase: insights into the mechanism of arsenic biotransformation. *Biochemistry*, 51, 5476–5485
  - Gunner, M. R., Mao, J., Song, Y. and Kim, J. (2006) Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations. *Biochim. Biophys. Acta*, 1757, 942–968
  - Antosiewicz, J. M. and Shugar, D. (2011) Poisson-Boltzmann continuum-solvation models: applications to pH-dependent properties of biomolecules. *Mol. Biosyst.*, 7, 2923–2949
  - Anandkrishnan, R., Aguilar, B. and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, 40, W537–W541
  - Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, 11, 3696–3713
  - Xu, X. and Truhlar, D. G. (2011) Accuracy of effective core potentials and basis sets for density functional calculations, including relativistic effects, as illustrated by calculations on arsenic compounds. *J. Chem. Theory Comput.*, 7, 2766–2779
  - Jorgensen, W. L. and Thomas, L. L. (2008) Perspective on free-energy perturbation calculations for chemical equilibria. *J. Chem. Theory Comput.*, 4, 869–876
  - Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26, 1781–1802
  - Wagner, G. J. and Liu, W. K. (2003) Coupling of atomistic and continuum simulations using a bridging scale decomposition. *J. Comput. Phys.*, 190, 249–274
  - Hünenberger, P. (2005) Thermostat algorithms for molecular dynamics simulations. *Adv. Polym. Sci.*, 173, 105–149
  - Krautler, V., Van Gunsteren, W. F. and Hunenberger, P. H. (2001) A fast SHAKE: algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.*, 22, 501–508
  - Cerutti, D. S., Duke, R. E., Darden, T. A. and Lybrand, T. P. (2009) Staggered Mesh Ewald: an extension of the Smooth Particle-Mesh Ewald method adding great versatility. *J. Chem. Theory Comput.*, 5, 2322–2338