EFFECTS OF SINGLE NUCLEOTIDE POLYMORPHISMS ON ARSENIC-CYSTEINE INTERACTIONS

Raphael D. Isokpehi\textsuperscript{1,2}, Hari H.P. Cohly\textsuperscript{1,2}, Matthew N. Anyanwu\textsuperscript{2,3}, Rajendram V. Rajnarayanan\textsuperscript{4}, Paul B. Tchounwou\textsuperscript{1}, Udensi K. Udensi\textsuperscript{1,2} and Barbara E. Graham-Evans\textsuperscript{1}

\textsuperscript{1}RCMI-Center for Environmental Health, College of Science, Engineering & Technology, Jackson State University, Jackson MS 39217; \textsuperscript{2}Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, PO Box 18540, Jackson MS 39217; \textsuperscript{3}Department of Computer Science, University of Memphis, Memphis TN; \textsuperscript{4}Department of Pharmacology and Toxicology, State University of New York at Buffalo, Buffalo, New York, USA.

Abstract: Arsenic is a toxic metalloid that causes skin cancer and binds to cysteine residues—a property that could be used to infer arsenic responsiveness of a target protein. Non-synonymous Single Nucleotide Polymorphisms (nsSNPs) result in amino acid substitutions and may alter arsenic binding with cysteine residues. Thus, the objective of this investigation was to identify and analyze nsSNPs that lead to substitutions to or from cysteine residues as an indication of increased or decreased arsenic responsiveness. We hypothesize that integration of data on molecular impacts of nsSNPs and arsenic-gene relationships will identify nsSNPs that could serve as arsenic responsiveness markers. We have analyzed functional and structural impacts data for 5,811 nsSNPs linked to 1,224 arsenic-annotated genes. In addition to the identified candidate nsSNPs for increased or reduced arsenic responsiveness, we observed (i) a nsSNP that results in the breakage of a disulfide bond, as candidate marker for reduced arsenic responsiveness in KLK7, a secreted serine protease participate in normal shedding of the skin; and (ii) 6 pairs of vicinal cysteines in KLK7 protein that could be binding sites for arsenic. In order to facilitate additional investigations on prioritized SNPs, protein targets and molecular mechanisms of arsenic action, we have constructed a collection of over 100,000 sentences from over 16,000 PubMed abstracts on arsenic. Finally, a web resource Arsenic Sentence Database (http://compbio.jsums.edu/arsenic_pubmed/) was developed to enable web-based searches of the sentences by keywords and PubMed identifiers.

Keywords: Arsenic, cysteines, information extraction, KLK7, SNPs

Acknowledgments: Pittsburgh Supercomputing Center's National Resource for Biomedical Supercomputing (T36 GM008789); U.S. Department of Homeland Security Science & Technology Directorate (2007-ST-104-000007; 2009-ST-062-000014; 2009-ST-104-000021); and Research Centers in Minority Institutions (RCMI) – Center for Environmental Health (NIH-NCRR G12RR13459). Disclaimer: “The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.