RECENT ADVANCES IN ARSENICOGENOMICS: FUNCTIONAL IMPACT AND POPULATION DIVERSITY OF SINGLE NUCLEOTIDE POLYMORPHISM IN ARSENIC RESPONSIVE GENES

Raphael D. Isokpehi, Paul B. Tchounwou, Barbara Graham-Evans and Hari H.P. Cohly

Department of Biology and RCMI Center for Environmental Health, Jackson State University, Jackson, Mississippi USA

Abstract: Arsenic is recognized as an environmental toxicant of global public health concern. Long-term exposure to arsenic principally through drinking water has been correlated with increased risk to diseases such as skin cancer, diabetes, blackfoot disease, spontaneous abortions, and arteriosclerosis. The avalanche of genome sequences combined with genome-enabled datasets from high-throughput gene expression, genotyping, haplotyping and protein assays is making it possible to gain biological insights into previously unknown gene-toxicant interactions. Arsenicogenomics, an aspect of toxicogenomics, therefore, provides a means to (i) understand how various genes respond to arsenic and (ii) how arsenic modifies the function and expression of specific genes in the genome. It has been established from population-based studies in arsenic-endemic regions that not all individuals exposed to arsenic develop disease. The diversity of single nucleotide polymorphisms (SNPs) derived from arsenic responsive genes in different populations could provide a prediction of an individual’s susceptibility to arsenic-induced diseases. There are over 1,000 human genes curated in the Comparative Toxicogenomics Database with evidence for interaction with arsenicals (Medical Subject Heading Identifier [MeSH ID]: D001152). The objective of this study was to determine the impact of Reference SNPs (RefSNP) location on gene function as well as their heterozygosity in different populations. We took advantage of availability of the central public SNP database (dbSNP) to identify reference SNPs. We also developed computational pipelines to extract functional impact of SNP genomic context and its population diversity information. A total of 13,406 SNPs associated with 1,355 human genes were retrieved and classed into the following functional impacts: frame shift (770), missense (6,693), nonsense (185) and synonymous (5,758). For the 5,882 SNPs in which the population diversity measure of heterozygosity could be obtained, the highest average heterozygosity of 0.731 was observed in a SNP mapped to the NEK4 gene. We have organized the synthesized gene-arsenic SNP dataset into subsets based on susceptibility genes extracted from PubMed abstracts. Subsets include SNPs from genes in the base excision DNA repair pathway such as X-ray repair cross-complementing group 1 (XRCC1), human 8-oxoguanine DNA glycosylase (hOGG1) and apurinic/apyrimidinic endonuclease (APEX1), that may influence individual susceptibility to arsenic-induced skin lesions. We have also developed a collection of over 80,000 PubMed citations on genes annotated to interact with arsenic. The citations can be searched using an ontology-based Textpresso Text Mining Engine that is available at http://compbio.jsums.edu/textpresso/. In conclusion, our arsenicogenomics study has yielded integrated evidence-based datasets for evaluation as biomarkers for susceptibility to arsenic-induced diseases.

Acknowledgements: Research Centers in Minority Institutions (RCMI) – Center for Environmental Health at Jackson State University (NIH-NCRR G12RR13459-09).