VISUAL ANALYTICS FOR ENVIRONMENTAL HEALTH: BIOLOGICAL CONCEPTS ENRICHED IN ARSENIC RESPONSIVE GENES

Raphael D. Isokpehi\textsuperscript{1,2}, Andreas N. Mbah\textsuperscript{1}, Udensi K. Udensi\textsuperscript{1,2}, and Matthew N. Anyanwu\textsuperscript{1}

\textsuperscript{1}Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson Mississippi 39217, USA
\textsuperscript{2}RCMI Center for Environmental Health, College of Science, Engineering and Technology, Jackson State University, Jackson MS 39217, USA

Abstract: Visual Analytics is a multidisciplinary field that is defined as the science of analytical reasoning facilitated by interactive visual interfaces. There are increasing efforts to develop visual analytics software that facilitates discussion between user and disease surveillance data for environmental and public health policy formulation, resource allocation, and decision making. Public and environmental health data are often collected in categories requiring statistical methods to model and find associations between categorical variables. Interactive exploratory analysis via visual interfaces to identify partitions in a dataset can prevent missing information on noteworthy associations between variables. As a case study we have used visual analytics tools to integrated enriched biological information on 2058 genes that have been investigation for interaction with arsenicals. A total of 339,543 enriched biological concepts for the 2,058 genes were obtained from ConceptGen database (http://conceptgen.ncibi.org/). Visual analytics tools were then used process the datasets to gain insights into shared and unique biological concepts such as Gene Ontology (GO) annotation, differential gene expression and microRNA association. At least 20 genes had enriched biological concepts containing “arsenic”. In particular, AS3MT [arsenic (+3 oxidation state) methyltransferase], ASNA1 [arsA arsenite transporter, ATP-binding, homolog 1] and GCLC (glutamate-cysteine ligase, catalytic subunit) were annotated with the GO Biological Process term “response to arsenic”. These prioritized genes have vicinal cysteines, which are known to be arsenic binding sites in protein sequences. We are determining the protein structure and population diversity of these genes to predict susceptibility to arsenic-induced diseases.

Acknowledgements: RCMI-Center for Environmental Health, Jackson State University; National Institutes of Health (NIH-NCRR G12RR13459; NIH-NIGMS T36GM095335; NIH-NIMHD 1P20MD002725-01; NIH-NCRR-P20RR016476; NIH-NCRR-P20RR016460); National Science Foundation (EPS-0903787; NSF-DBI-0958179; DBI-1062057); and US Department of Homeland Security Science and Technology Directorate (2007-ST-104-000007; 2009- ST-062-000014; 2009-ST-104-000021). 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.