

BIOINFORMATICS ANALYSIS FOR THE ENVIRONMENTAL AND GENETIC RISK FACTORS OF BREAST CANCER

Antoinasha L. Hollman¹, Paul B. Tchounwou^{1,2} and Hung-Chung Huang^{1,2}

¹*NIH/NIMHD RCMI Center for Environmental Health, College of Science, Engineering, and Technology (CSET), Jackson State University, Jackson, MS 39217, USA*

²*Department of Biology, CSET, Jackson State University, Jackson, MS 39217, USA*

Abstract: Breast cancer has been identified as one of the most common cancers among women in the United States. Genetics is always taken into consideration with this cancer. However, certain chemicals have been proposed as a potential environmental factor. According to previous studies, polychlorinated biphenyls (PCBs) are estrogen-mimicking endocrine disruptors (EEDs) that may be linked to the development of breast cancer. It was proposed that CYP19A1, EGFR, ESR2, FOS, and IGF1 are the genes influenced by PCBs, Bisphenol A (BPA), and phthalates. Polycyclic aromatic hydrocarbons (PAHs) from indoor air pollution are suggested to be strongly associated with this particular cancer. Environmental tobacco smoke, fireplace use, and grilled meat intake are a few indoor pollution sources. The consumption of alcohol seems to possibly play a role in the influence of breast cancer in women. Studies have also suggested a gene-environment interaction between SLC4A7-rs4973768 and smoking, as well as CASP8 and alcohol consumption. We hypothesized that gene-environment interaction induces a change in the genetic material, such as single nucleotide polymorphisms (SNPs) that will affect some genes' expression levels in the presence of breast cancer tumors. Our objectives are to (1) obtain datasets that pertain to breast cancer, (2) evaluate the gene expression levels associated with breast cancer to identify the over- and under-representation of classes of genes or proteins, (3) determine disease-causing SNPs or loci that may be associated with breast cancer based on eQTL analysis on gene expression and genotype data altogether, and (4) evaluate the role, if any, of environmental factors may play in the affected genes. Gene Expression Omnibus (GEO) is a repository used to obtain next-generation sequencing (NGS) datasets for breast cancer on tumor samples with different treatments. Linux-based environment and R software were used to analyze the gene expression levels. Functional enrichment analysis of the differentially expressed genes is performed by the WebGestalt web program for the geneset enrichment studies. dbSNP database provides a broad collection of SNPs that associated with this particular cancer which will be used as reference to compare with the SNPs we found related to breast cancer. Comparative Toxicological Database (CTD) that revealed toxins that could induce cancerous tumors in the breast would be used for this study too.

Keywords: Next-generation sequencing, single nucleotide polymorphism (SNP), gene expression, gene-environment interaction, breast cancer, polychlorinated biphenyls (PCBs), estrogen-mimicking endocrine disruptors (EEDs), Bisphenol A (BPA), phthalates, polycyclic aromatic hydrocarbons (PAHs)

Acknowledgements: This research supported by the National Institutes of Health/National Institute on Minority Health and Health Disparities Grant # G12MD007581, through the RCMI Center for Environmental Health at Jackson State University.