INTEGRATIVE TEXT MINING AND GENE INFORMATION ANALYSES TO UNCOVER GENES PREDISPOSING CO-MORBIDITY OF BIPOLAR DISORDER AND SUBSTANCE ABUSE IN AFRICAN-AMERICAN WOMEN

Sharon A. Lewis1, Tonya M. Gerald2, Wellington K. Ayensu3 and Raphael D. Isokpehi3

1Department of Chemistry, Langston University, Langston, OK 73050, USA
2Department of Chemistry, North Carolina Central University, Durham, NC 27707, USA
3Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson, MS 39217, USA

Abstract: Bipolar disorder (BPD) is a highly heritable, severe and chronic mental illness characterized by episodes of elation and high activity; alternating with periods of low mood and low energy. When compared to major depressive disorder (MDD), bipolar disorder is less prevalent but more persistent and more impairing. Bipolar disorder poses a major challenge to the United States and the global healthcare system. This burden is complicated by the co-morbidity of bipolar disorder with narcotics and alcohol abuse. In the United States, there is consensus from published research that bipolar disorder is more chronic and severe in African-Americans compared to other populations. Furthermore, African-Americans are prescribed anti-psychotics for treatment of bipolar disorder though this type of medication is primarily prescribed to treat patients with acute mania. In the entire population, there are some evidence for gender disparity disproportionately affecting women in hospitalization during manic episodes, rapidity of cycling, and predominance of depressive features. We assume that the PubMed literature database and linked databases (such as Gene Expression Omnibus (GEO), HomoloGene, the Single Nucleotide Database [dbSNP] and The database of Genotype and Phenotype [dbGAP]), contain genes that can be prioritized for the study of the co-morbidity of bipolar disorder and substance abuse in African-American women. We hypothesize that high-throughput data integration will enable us to prioritize genes that can be further studied for the co-morbidity of bipolar disorder and substance abuse in African-American women. The objective of this study was to prioritize a list of genes that are mentioned in PubMed citations annotated with the Medical Subject Heading (MeSH) term: Bipolar Disorder. We constructed a collection of PubMed citations and used a Textpresso Text Mining Engine as the search interface for the citations. Abstracts and or title that contain the word “polymorphism” were retrieved and integrated with a dataset of 12.5 million Entrez Gene to PubMed mappings. The collection consisted of 23066 articles and 126391 sentences can be searched at http://compbio.jsums.edu/textpresso/. A total of 1159 sentences from 547 abstracts were identified as containing mentions of the term polymorphism. Furthermore, 214 human genes were linked to these abstracts. The functional roles gleaned of these genes included receptor and enzyme activities. Additional research will integrate information from dbSNP as well as studies from chromosomal coordinates, linkage analysis and genome-wide association, to further filter genes for a role in co-morbidity of bipolar disorder and substance abuse in African-American women.

Acknowledgements: Langston University; Oklahoma IdeA Network of Biomedical Research Excellence (OK-INBRE); OK-INBRE- National Institutes of Health Grant # P20RR016478-07; the Research Centers in Minority Institutions (RCMI) – Center for Environmental Health at Jackson State University (NIH-NCCRR G12RR13459-09); Carnegie Mellon University, Pittsburgh Supercomputing Center (PSC) -National Institutes of Health Grant #5 T36 GM008789