Factomics: A Cloud-Enabled Application Incorporating Integrative Genomics and GWAS Facilitating Disease Causation Analysis

Revanur, Swetha

Researchers and medical professionals are inundated with tremendous amounts of genomic ‘big data’. However, there is a lack of tools that are capable of offering insightful interpretations of this raw data. To address this void, I developed Factomics, a cloud application, which provides a novel built-in workflow for performing disease correlation/causation analysis based on integrative genomics data. The Factomics workflow is divided into 4 phases. These phases take the user from an initial set of diseases to their candidate causal SNPs, genes, and pathways based on standard analytical algorithms, configurable gene expression data providing phenotypic context, and genome-wide association studies (GWAS). Factomics follows a 3-tier architecture. Google Apps Script (server-side JavaScript) is used for the UI. The middle-tier is integrated with external analytical modules and public data repositories including NCBI databases. Analytical findings are stored on Google Drive. I demonstrated a use-case of Factomics with Alzheimer’s Disease, Type 2 Diabetes Mellitus, Ovarian Cancer, and Pancreatic Cancer. The built-in workflow distilled thousands of SNPs, genes, and pathways per disease to about 25 disease causation hypotheses. Analytical modules identified missing gene to disease and gene to pathway associations, and also highlighted 9 non-synonymous (deleterious) and 13 regulatory candidate causal SNPs. A multi-dimensional view of these diseases showed several overlapping up-regulated genes and pathways. Some of the findings were corroborated by literature, and others were novel. This information can be used to drive new hypotheses leading to optimized drug development, drug repositioning, and diagnostic tools.