KnowEnG: Cloud-Based Environment for Scalable Analyses of Genomic Signatures

***Charles Blatti, University of Illinois at Urbana-Champaign**; Matthew Berry; Lisa Gatzke; Amin Emad; Nahil Sobh; Colleen Bushell; Saurabh Sinha*

Medical and biological researchers require advanced and convenient tools to cluster, characterize, and uncover important features from the genomic datasets they produce (e.g. transcriptomic, epigenetic, or genotypic profiling experiments). A body of recent research has shown these types of analysis are often improved by incorporating prior biological knowledge, such as known relationships and interactions between entities like genes, proteins, functional roles, disease phenotypes, etc. KnowEnG (the "Knowledge Engine for Genomics" strives to make the results of these advanced machine learning and graph processing algorithms, as well as dozens of collections of standardized biological knowledge datasets available to researchers through an intuitive user interface and interpretable visualizations. Our platform is designed to be deployed on heterogeneous, commercial cloud services and relies on technologies such as Docker and Apache Mesos to decompose large, complex analysis workflows and datasets into independent components to maximize parallel computation. In KnowEnG, multiple classes of analysis workflows for genomic datasets are available in both standard and knowledge-informed implementations including 1) gene prioritization methods for tasks like ranking genes for their role in drug response and 2) gene set characterization methods for finding relevant biological annotations. Through analysis available in our platform, we highlight novel insights gained from using random walks on prior knowledge networks over conventional approaches.