

Integration of genetic studies for multiple diseases with biomedical big data has recently been considered to be a powerful approach to improve identification of risk genetic variants. However, it still remains challenging to integrate genome-wide association studies (GWAS) datasets for multiple diseases and effectively utilize information in biomedical big data for GWAS data analyses. In this presentation, I will discuss our novel DDNet-graph-GPA framework which addresses these challenges. Specifically, we developed graph-GPA, a novel Bayesian model that integrates multiple GWAS datasets using a latent Markov random field architecture and allows to incorporate external prior biological knowledge (<https://dongjunchung.github.io/GGPA/>). In addition, we also generated a data resource to infer disease-gene relationships by implementing a comprehensive text mining of biomedical literature utilizing gene ontology knowledge. We further developed DDNet, a public database and web interface that allow researchers to mine genetic relationships among diseases based on the biomedical literature (<http://dnet.chunglab.io/ddnet/>). We applied the proposed approach to simulation studies and real GWAS datasets, where the disease-disease graph obtained from DDNet was used as prior knowledge for graph-GPA. The results show that the proposed approach does not only improve identification of risk genetic variants, but also facilitates understanding of genetic relationships among complex diseases. Finally, I will briefly discuss our ongoing research projects for more effective utilization of biomedical literature mining data, including GAIL (<http://chunglab.io/GAIL/>) and bayesGO (<https://dongjunchung.github.io/bayesGO/>).