[BIOINFO 2023]

Hierarchical Structural Component Analysis of Biological Pathways using Graph Attention Network (HisCoM-GAT)

Teawan Goo[1] and Taesung Park[2*]

[1] *Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea*

[2] *Department of Statistics, Seoul National University, Seoul, Republic of Korea*

* *Corresponding author: tspark@stats.snu.ac.kr*

Recent advances in pathway analysis have significantly enhanced our understanding of disease mechanisms. However, while many analyses have been developed for identifying disease-associated pathways using omics data, a significant majority rely on individual pathway analysis with linear assumption and subsequently overlook correlations among multiple pathways. DeepHisCoM has addressed many of the previous limitations, but it still have been limited short in considering the biological relationships, like protein-protein interaction (PPI), co-expression and co-mutated, among genes within a pathway. To address these gaps and to foster a more holistic understanding, especially, we introduce the Hierarchical Structural Component Analysis using Graph Attention Network (HisCoM-GAT), which encapsulates the hierarchical structure between genes and pathways while accounting for inter-pathway correlations. Using the well-known PPI database, we built a pathway graph. This setup allowed the GAT model to learn the complex relationships between genes. After GAT, a fully connected layer summarized gene attention scores into a pathway score. Association between the pathway scores and phenotype were tested using permutation tests. We evaluated our model through simulation studies in terms of statistical power and false positive rate, comparing it to existing pathway analysis methods. When we applied it to PDAC datasets from Australia, Canada, America, and Korea, it revealed intricate gene relationships within pathways. In conclusion, the HisCoM-GAT offers a comprehensive and advanced approach to pathway analysis in omics data. It not only addresses the limitations of existing methods like DeepHisCoM but also provides a more detailed and interconnected view of gene-pathway structures. The model's ability to allow interactivity and visualization, especially when applied to real-world datasets, sets it apart as a valuable tool for researchers. As omics data continues to play a pivotal role in understanding disease mechanisms, HisCoM-GAT will be crucial in unlocking deeper insights and enhancing our overall comprehension of intricate biological processes.