[BIOINFO 2024]

PathCLAST: a Pathway-augmented Contrastive Learning with Attention mechanism for Enhanced Spatial Domain Identification in Spatial Transcriptomics Data

Minho Noh[1], Sungkyung Lee[1], and Sangsoo Lim[2,*]

[1]*Department of Computer Science and Artificial Intelligence, Dongguk University-Seoul, Korea*
[2]*Division of AI Software Convergence, Dongguk University-Seoul, Korea*
**Corresponding author:*
Sangsoo Lim, *sslim@dgu.ac.kr*

Spatial transcriptomics integrates the spatial organization of cells with their gene expression profiles, providing new opportunities to study cellular function in its native tissue context. Integrating these types of data creates vast potential to interpret the biology of cell types in their native morphological context. Several computational methods have been developed to elucidate the spatial context within tissue by combining histopathological images with transcriptomics data. However, recent methods on dimensionality reduction of gene expression profiles do not consider gene interactions or innate functional potential, making it challenging to decipher biological functions such as cell-cell interaction, in a spatial framework. To address this, we introduce PathCLAST, which transforms the gene expression profiles of each spot into pathway graphs — gene interactions that represent biological functions — allowing us to consider gene-level functional relationships in a spatial context. By leveraging multi-modal contrastive learning, PathCLAST integrates both histopathological images and pathway-based gene expression profiles to generate low-dimensional, interpretable latent embeddings. We also incorporate data-dependent augmentation techniques to enhance model robustness and prevent overfitting. An attention mechanism is employed to derive pathway activation profiles, enabling the identification of specific pathways active in distinct tissue regions. Additionally, in-depth analysis of boundary and interior tissue regions reveals PathCLAST's ability to distinguish spatial domains at a pathway level. Our method not only improves the identification of spatial domains within benchmark datasets but also offers novel insights into the biological functions of tissue regions by elucidating pathway activations within a spatial context. PathCLAST provides a powerful framework for understanding the spatial organization of biological functions at the pathway level, facilitating more precise interpretations of tissue architecture and function.