[BIOINFO 2025]

SPAM : Spatial transcriptomics Prediction with self-supervised Alignment of Multi-modalities

Jaeyun Park[1], Dongsin Kim[2], and Minsik Oh[2,*]

[1] *Department of Data Technology. Myongji University*

[2] *Department of Converged Software. Myongji University*

* *Corresponding author: msoh@mju.ac.kr*

Spatial transcriptomics (ST) enables spatial localization of gene expression within tissue sections, providing insights into disease pathogenesis and spatially organized cellular heterogeneity. Despite its promise, ST remains constrained by high cost, limited resolution, sparse capture rates, and batch variability. Existing computational approaches often rely on paired supervision and lack efficient self-supervised strategies for robust cross-modal representation. Moreover, many prediction models treat each gene independently, neglecting condition-specificity and the contextual dependencies of gene activity. In our framework, incorporating a global prior based on the mean expression profile stabilizes predictions and embeds condition-specific biological context.

We propose a two-step framework for predicting spatially resolved gene expression from H&E histology by integrating histology images, spatial coordinates, and transcriptomic profiles. In the first step, pretraining employ multi-branch contrastive learning that aligns modality pairs, encouraging histology-derived features, spatial neighborhoods, and transcriptional profiles from the same location to share a representation. In the second step, fine-tuning fuses modality-specific encoders with cross-attention. In our framework, a global prior derived from the gene expression profile is incorporated during inference, stabilizing predictions while embedding condition-specific biological context. This design produces biologically consistent and spatially resolved maps directly from routine histology.

We evaluate our method on in-house colorectal cancer and public Xenium datasets, including human tonsil, breast cancer, skin, and kidney. It outperforms baseline models (ST-Net, BLEEP, STMCL) in predicting spatially resolved expression, measured by Pearson correlation. Qualitative spatial plots confirmed the biological validity of predictions, showing concordance with histology. Ablation studies demonstrate that contrastive alignment, modality-specific encoders, and mean expression priors each improve performance, while cross-dataset analysis highlights robustness to heterogeneity.