[BIOINFO 2025]

stLENS identifies unbiased complex spatial patterns in spatial transcriptomics data

Yuhui Jeong[1,#], Hyeon-Min Kim[1,#], Hyun Kim[2,#], Jae Kyoung Kim[2,3,4,*], Jeongbin Park[1,*]

[1]Department of Biomedical Convergence Engineering, Pusan National University
[2]Biomedical Mathematics Group, Pioneer Research Center for Mathematical and Computational Sciences, Institute for Basic Science
[3]Department of Mathematical Sciences, KAIST
[4]Department of Medicine, College of Medicine, Korea University
[#]Co-first Authors
[*]Corresponding author: J.K.K.(jaekkim@kaist.ac.kr) or J.P. (jeongbin.park@pusan.ac.kr)

We recently reported scLENS (single-cell Low-dimensional embedding using the effective Noise Subtract), a tool that computes optimized dimensionality reduction by extracting biologically meaningful signals from noisy single-cell RNA sequencing datasets using random matrix theory. A major advantage of scLENS lies in its exceptional performance on datasets with high sparsity and skewed data distribution—features also prominent in array-based spatially resolved transcriptomics (SRT) data. However, as the number of spatial spots per chip continues to grow rapidly due to technological advances, the original scLENS implementation struggles to scale for such datasets.

To this end, we introduce stLENS, a Python package designed for large-scale SRT datasets. By leveraging parallel and lazy computation frameworks such as CuPy, Dask, Zarr, and OpenMP, stLENS significantly enhances computational scalability, enabling the analysis of datasets containing up to ~1 million cells or spatial spots — a scale previously infeasible with scLENS. Together, these features make stLENS ideally suited for large-scale spatial transcriptomics analysis. Furthermore, the stLENS package seamlessly integrates within the existing large Python ecosystem, such as Scverse. This enables smooth interoperability with widely used spatial transcriptomics analysis packages such as Scanpy, Squidpy, or SpatialData.

We assessed the accuracy of the downstream analysis using the optimized dimensionality-reduced matrix computed by stLENS using public SRT datasets, by comparing the Leiden clustering results with the conventional principal component analysis (PCA) with default parameters using Scanpy. By comparing the differences in the best Jaccard Index of each cell type, we found that stLENS outperformed Scanpy in both STOmics and VisiumHD datasets.