[BIOINFO 2024]

Enhancing cell type classification in scRNA-seq data using Transformer based model and singular value decomposition

Sanghun Sel[1], Subin Lee[1], Hanul Park[1], and Giltae Song [1,2,3,*]

[1] *Division of Artificial Intelligence, Department of Information Convergence Engineering, Pusan National University*

[2] *Center for Artificial Intelligence Research, Pusan National University*

[3] *School of Computer Science and Engineering, Pusan National University*

* *Corresponding author: gsong@pusan.ac.kr*

Single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of cellular heterogeneity and function within complex biological systems. However, analyzing scRNA-seq data remains challenging, especially in cell type classification. Current methods often rely on limited prior knowledge or use opaque neural networks, which compromise interpretability. To address these challenges, we propose a novel approach that enhances both classification accuracy and biological interpretability by utilizing transformer models combined with singular value decomposition (SVD). This method captures key patterns across cells and genes, optimizing dimensionality reduction while preserving critical information. The transformed data allows for comprehensive exploration of the cellular landscape, facilitating robust classification and annotation. Additionally, the approach identifies marker genes associated with individual cells and broader cell types, offering deeper insights into cellular identity and function. This method represents a significant advancement in scRNA-seq analysis, providing a versatile and interpretable framework to unravel cellular diversity and function. In essence, the combination of SVD with transformer-based modeling extracts meaningful insights from scRNA-seq data, offering a concise yet thorough solution to the challenges of scRNA-seq analysis with implications for a wide range of biological research areas.