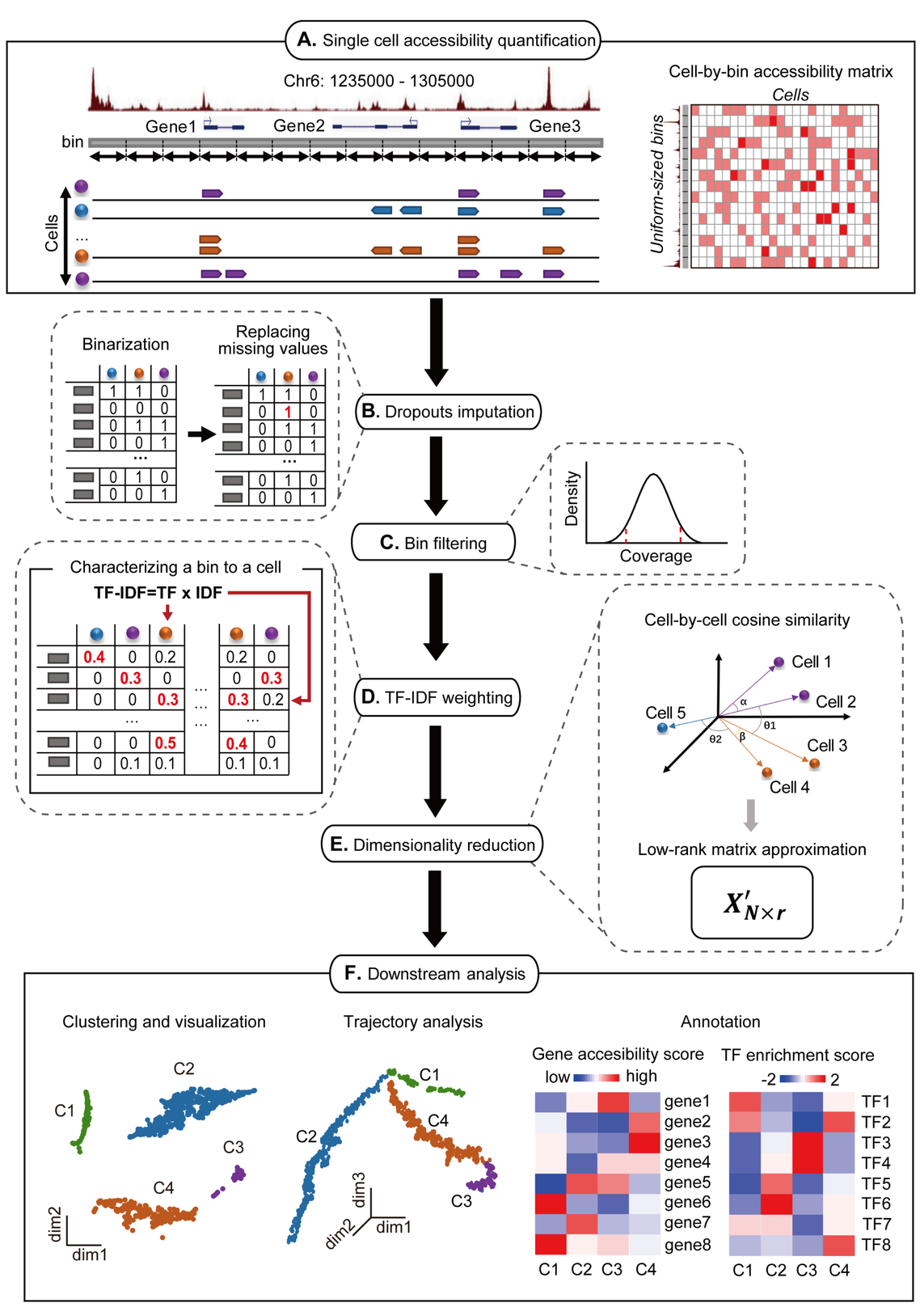
scART: recognizing cell clusters and constructing trajectory from single-cell epigenomic data

Jingxin Guo, Jingyu Li, Fei Huang, Jiadong Chen and Li Shen
Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang 310058, China

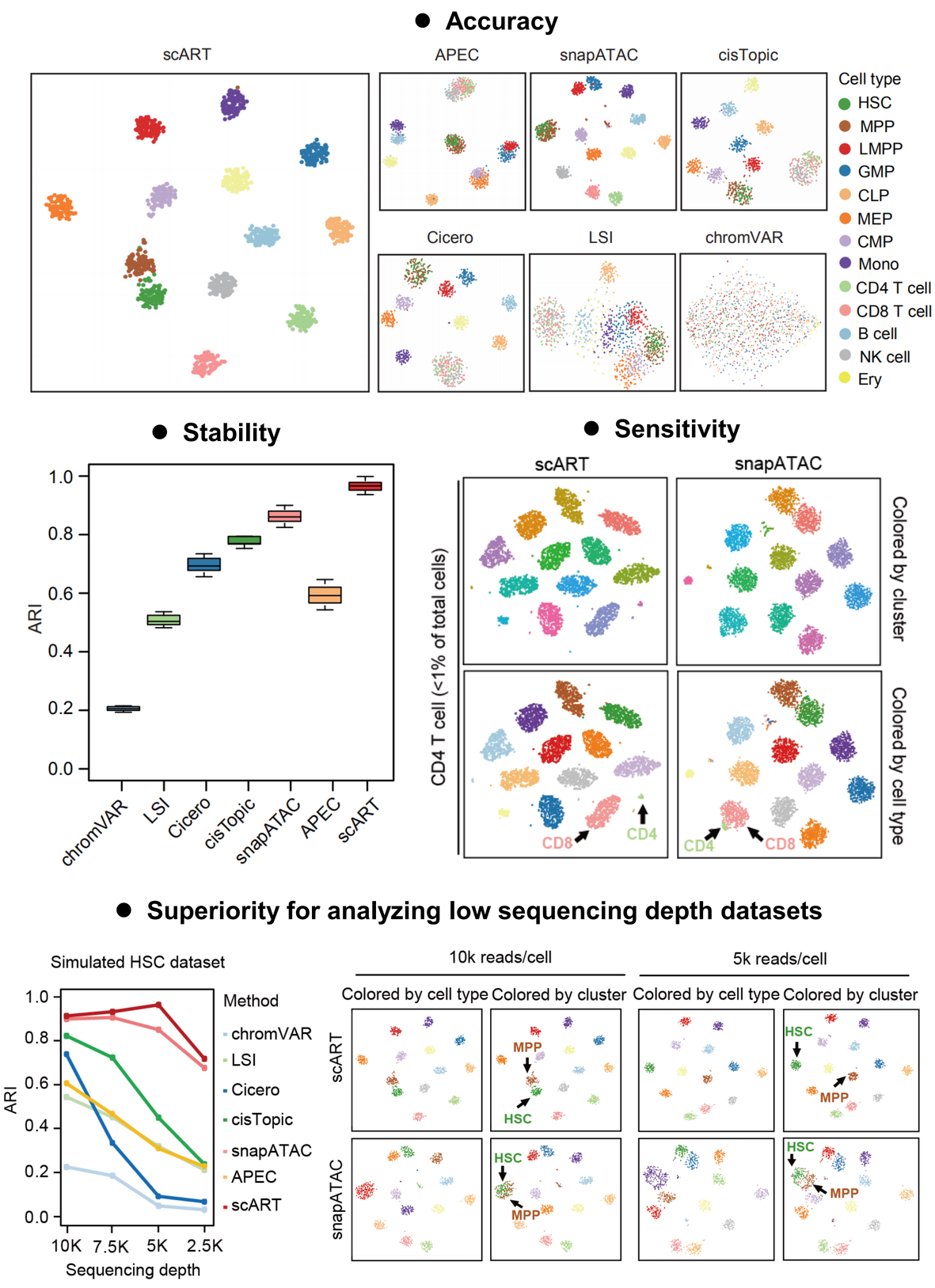
Introduction

Gene expression is a multifaceted process controlled by the combinatorial activity of regulatory elements. Expression heterogeneity could arise from cell-to-cell differences in gene activation and repression regulatory mechanisms such as the action of TFs, chromatin modifiers, and other regulatory factors. The scATAC-seq method is a powerful tool to reveal the regulatory logic of gene expression programs in single cells. However, the high signal sparsity and high levels of noise in scATAC-seq data raised new computational challenges for data analysis. Here, we developed a bioinformatics tool for scATAC-seq data analysis, which is termed Single-cell Chromatin Accessibility-based cluster Recognition and Trajectory reconstruction (scART). In order to reduce noise in the sparse data, scART combined KNN imputation and TF-IDF weighting scheme and inferred cell-to-cell similarities by cosine distance. Compared with published methods, scART exhibited its superiority in the accuracy and sensitivity of cell clustering, especially when used for datasets with low sequencing depth. Besides, scART was capable to construct developmental trajectories from scATAC-seq data. By reanalysis of a published dataset, scART revealed the dynamic changes of cortical layer neurogenesis during mouse embryo forebrain development. Furthermore, scART can be used for motif enrichment analysis and gene accessibility analysis when processing scATAC-seq data.

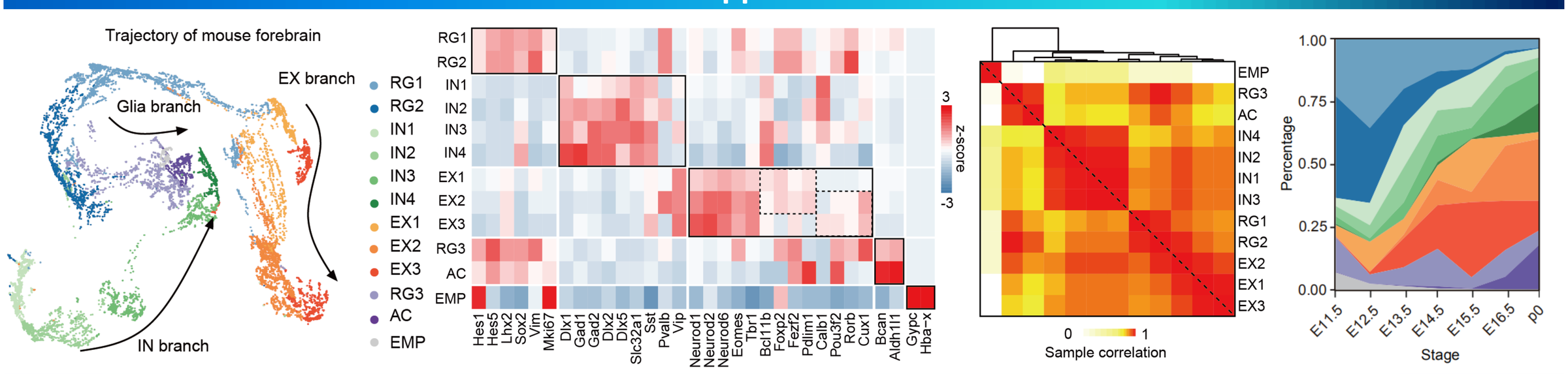
Workflow



Performance



Application



Summary

1. scART combined de-noising algorithms including KNN imputation, TF-IDF weighting scheme and cell-to-cell cosine similarities to reduce the high-level noise in the scATAC-seq data.
2. scART outperforms current methods in accuracy, sensitivity, and stability of identifying cellular heterogeneity in simulated and real datasets.
3. scART is more robust to noise for sparse, high dimensional data analysis and is sensitive to recognize cell types for low coverage datasets.
4. scART reconstructs the development trajectory of mouse forebrain and dissected excitatory neuron cells into different layer neuron clusters.