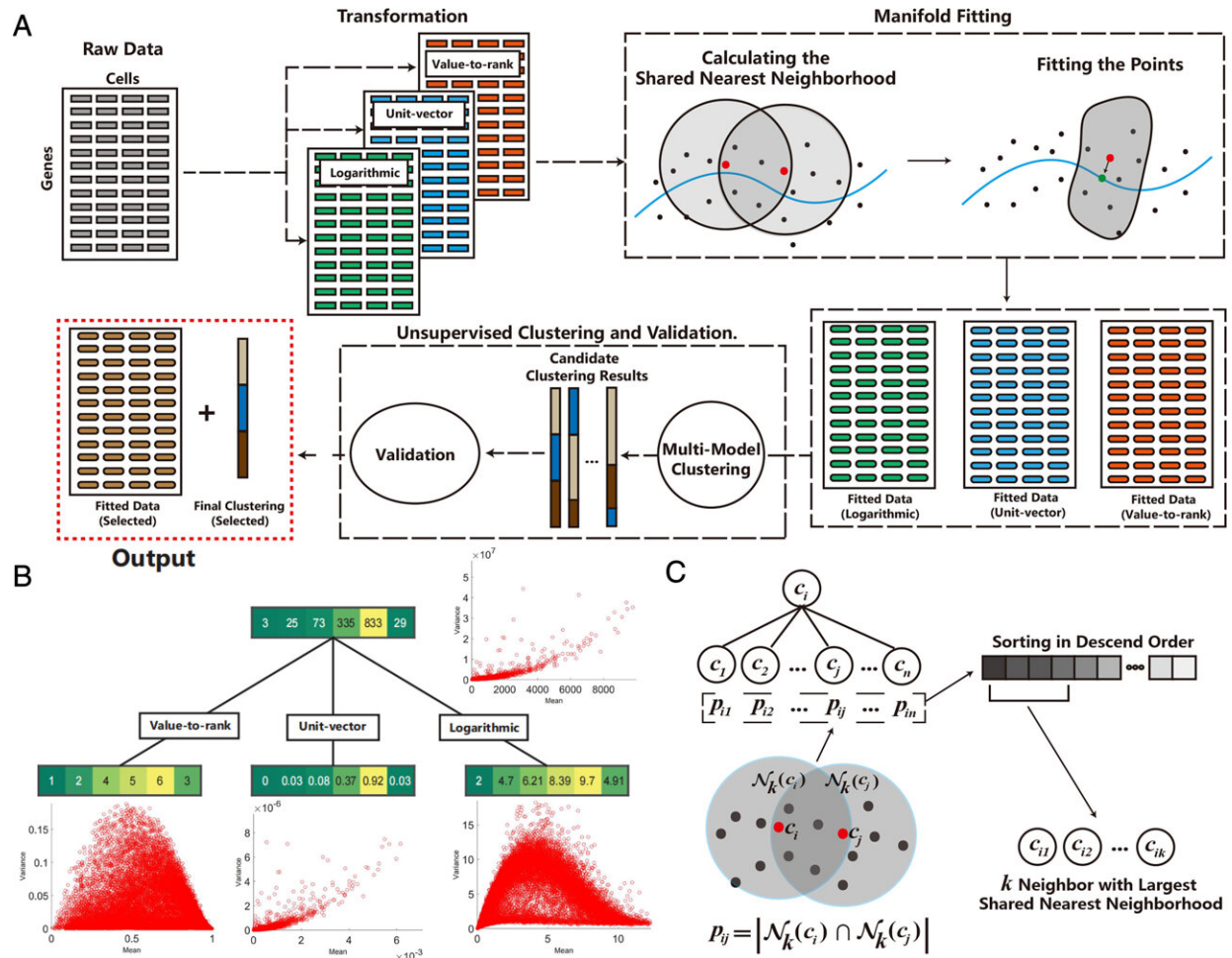


Elevating analysis of genomic data with advanced mathematical techniques

September 6 2024



The scAMF framework for the analysis of scRNA-seq data. Credit: *Proceedings of the National Academy of Sciences* (2024). DOI: 10.1073/pnas.2400002121

A novel approach to analyzing single-cell RNA sequencing (scRNA-seq) data has been unveiled by NUS researchers. This method promises to enhance both the precision and speed of data interpretation, potentially accelerating progress in numerous areas of biomedical investigation, including studies on cancer and Alzheimer's disease.

The innovative framework, dubbed scAMF (Single-cell Analysis via Manifold Fitting), was developed by a team of scientists led by Associate Professor Zhigang Yao from the Department of Statistics and Data Science at the NUS Faculty of Science.

The framework employs advanced mathematical techniques to fit a low-dimensional manifold within the high-dimensional space where the gene expression data are measured. By doing so, scAMF effectively reduces noise while preserving crucial biological information. This allows for more accurate characterization of cell types and states.

This research was done in collaboration with Professor Yau Shing-Tung at Tsinghua University. Their findings have been [published](#) in the *Proceedings of the National Academy of Sciences* on 3 September 2024.

Harnessing manifold fitting techniques to overcome hurdles in data analysis

Single-cell RNA sequencing has become a crucial tool in [genomic research](#), offering unprecedented insights into cellular diversity and disease mechanisms. However, the inherent noise in scRNA-seq data, arising from both biological variability and technical errors, has long posed challenges for accurate analysis.

Traditional scRNA-seq analysis methods, including genomic imputation approaches, graph-based methods, and deep learning-based algorithms,

often struggle to accurately characterize cell relationships due to inherent noise.

The scAMF framework represents a significant step forward in overcoming these limitations. It operates on the principle of fitting a low-dimensional manifold within the ambient space of gene expression data, effectively reducing noise while preserving crucial information.

At the heart of scAMF lies the manifold fitting module which effectively denoises scRNA-seq data by unfolding their distribution in the ambient space. This technique aims to reconstruct a smooth manifold within the original space where the data is measured, capturing the low-dimensional structure of the data in a manner that minimizes information loss and effectively eliminates noise.

The key innovation of scAMF lies in its ability to improve the spatial distribution of the data, bringing gene expression vectors of cells from the same type closer together while maintaining clear separation between different cell types. This enhancement leads to more precise and reliable clustering in subsequent analyses.

"Our approach effectively denoises scRNA-seq data by fitting a low-dimensional manifold in the high-dimensional space," explained Assoc Prof Yao. "This method significantly improves the accuracy of cell type classification and the clarity of data visualization."

The scAMF method employs a unique combination of data transformation, manifold fitting using shared nearest neighbor metrics, and unsupervised clustering validation.

When compared to other methods, scAMF demonstrates superior performance in several key areas, including more effective noise reduction, improved clustering accuracy, better preservation of

biological information, competitive computational efficiency, clearer visualization, and robust performance across diverse datasets.

These improvements position scAMF as a powerful new tool in single-cell analysis, potentially enabling researchers to uncover previously hidden cellular heterogeneity and rare cell populations.

Future work—driving greater understanding of cellular diversity and function

Building on the success of scAMF, the research team is now developing a novel framework for constructing high-resolution, multiscale cell atlases. This new approach aims to overcome current methodological limitations in cell atlas construction, such as challenges in identifying small cell populations and outdated unsupervised learning techniques.

A key focus is the development of a multi-resolution cell analysis framework based on scAMF. This advanced framework aims to identify rare cell populations and contribute to the construction of comprehensive cell atlases.

The multi-resolution approach will allow researchers to analyze cellular heterogeneity at various levels of granularity, from broad cell types to subtle subpopulations. This is particularly crucial for identifying rare cell types that may be overlooked by conventional analysis methods.

"Our ongoing work has already shown promising results across numerous benchmark datasets, revealing novel biological insights," Assoc Prof Yao noted. "We've applied it to the Human Brain Cell Atlas and identified new subtypes and marker genes for various cell types."

This ongoing research promises to push the boundaries of single-cell

analysis even further, potentially revolutionizing our understanding of cellular diversity and function across various biological systems.

More information: Zhigang Yao et al, Single-cell analysis via manifold fitting: A framework for RNA clustering and beyond, *Proceedings of the National Academy of Sciences* (2024). [DOI: 10.1073/pnas.2400002121](https://doi.org/10.1073/pnas.2400002121)

Provided by National University of Singapore

Citation: Elevating analysis of genomic data with advanced mathematical techniques (2024, September 6) retrieved 6 September 2024 from <https://phys.org/news/2024-09-elevating-analysis-genomic-advanced-mathematical.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--