



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Xing, Y., Owomero, S., Tsoka, S., Borgo, R., & Abdul-Rahman, A. (in press). *scFlowVis: Streamlining scRNA-seq Analysis through Visual Design*. Poster session presented at IEEE VIS: Visualization & Visual Analytics 2024.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

scFlowVis: Streamlining scRNA-seq Analysis through Visual Design

Yiwen Xing*

Stanley Owomero*

Sophia Tsoka*

Rita Borgo*

Alfie Abdul-Rahman*

King's College London



Figure 1: The interface of *scFlowVis*, highlighting its key components.

ABSTRACT

The rise of single-cell RNA sequencing (scRNA-seq) technologies has increased the need for visual analytical tools. Current visualization tools often cater to specific analysis stages rather than the entire pipeline, and many analyses rely on computational notebooks requiring significant programming skills. To address these challenges, we developed *scFlowVis*, a web-based visualization application designed in collaboration with bioinformatics and biological experts. *scFlowVis* supports the entire analysis pipeline for scRNA-seq data, providing visualization and interactive assistance throughout the process without programming. It dynamically links panel functionalities, allowing users to explore data attributes, compare parameters, and document workflows visually. The tool can also recommend appropriate visualizations based on data characteristics. Our experiences highlight the challenges and lessons in developing visualization tools for similar research contexts. *scFlowVis* has been released on <https://github.com/EavanXing0416/scFlowVis>.

Index Terms: Design study, biological visualization.

1 INTRODUCTION

As the demand for single-cell RNA sequencing (scRNA-seq) analysis grows, a standardized workflow has been established, encompassing preprocessing, dimensionality reduction, clustering, and

advanced analyses like differential expression and trajectory inference. Numerous data processing and visualization toolkits have been developed to support this workflow, primarily focusing on specific steps within the pipeline. Integrated platforms such as Galaxy [4] offer tools for various analysis purposes within a single platform. However, they lack seamless integration, making it difficult for users to experiment iteratively between different analysis stages. Python and R notebooks, while powerful and capable of supporting the entire workflow, rely heavily on the user's programming skills.

Despite these advancements, there is still a notable lack of comprehensive visualization tools that span the entire analysis pipeline without requiring extensive programming knowledge. Additionally, tools are significantly needed to support workflow documentation and analysis operation recording [2]. These gaps in the domain motivated us to collaborate closely with domain experts to develop *scFlowVis*, a web-based visualization tool designed to streamline the scRNA-seq analysis pipeline.

Guided by the Nine-Stage Framework [6] and focusing on a user-centered, iterative design approach, *scFlowVis* was developed. It offers a comprehensive data processing and visualization suite, allowing users to complete the entire analysis pipeline within a single interface. It minimizes the need to switch between different tools and reduces programming effort. The design cleverly links panel functionalities, facilitating dynamic observation of data structures, exploration of parameters and methods, and visual documentation of the workflow. Additionally, it guides the selection of appropriate data processing and visualization methods.

Our contributions include collaborating with domain experts, developing *scFlowVis*, and documenting the design study process, sharing insights and lessons learned from this collaboration.

*e-mail: {yiwen.xing | stanley.owomero | sophia.tsoka | rita.borgo | alfie.abdulrahman}@kcl.ac.uk

2 RELATED WORK

scRNA-seq Analysis Pipeline: The field of computational biology has developed sophisticated methods for scRNA-seq analysis, including quality control, normalization, dimensionality reduction, clustering, trajectory inference, and differential expression analysis, forming the core of a typical scRNA-seq data analysis pipeline [7].

Visualization Techniques: Provenance tracking is crucial across domains for documenting and sharing analytical workflows [3], yet it is underutilized in scRNA-seq analysis tools. Visual programming can simplify the coding process through graphical elements, making it accessible for beginners [5]. It encompasses various graphical presentation formats such as flowcharts and node-based interfaces [1]. However, it is challenging scalability.

3 REQUIREMENT ANALYSIS

In collaboration with domain experts, we established three primary goals for the tool: 1) *support seamless conduction of the entire data analysis process*, 2) *support decision-making alongside the entire analysis process*, and 3) *support workflow and result presenting and sharing*. Derived from these goals, more specific requirements were summarized:

R1 Integrating sufficient data processing and visualization methods to accommodate the entire analysis process.

R2 Implementing dynamic data structure tracking mechanisms and enabling checking at every analytical phase.

R3 Documenting the workflow & supporting visual comparisons.

R4 Recommending suitable methods at different analytical phases.

R5 Enabling workflow presenting & sharing.

R6 Reducing coding workload.

4 scFlowVis

scFlowVis is the output of sustained iterative design and research conducted in collaboration with researchers and practitioners in the scRNA-seq analysis domain. It is a comprehensive visualization tool that assists researchers throughout the entire scRNA-seq analysis pipeline. Integrating a wide range of mainstream data processing and visualization methods, it categorizes them based on the structured workflow informed by domain experts' insights. *scFlowVis* allows for real-time querying and exploration of data and metadata and features a visualization recommendation engine that suggests suitable methods based on data attributes. The tool offers four distinct visualization display layouts, providing versatile data presentation options. All functionalities are seamlessly integrated, enabling users to import data, conduct analyses, and obtain ready-to-use visual results without leaving the interface. Fig. 1 illustrates the tool interface.

The **Data Structure Viewing Panel (DSV)** is designed with requirement R2 and the complexities of scRNA-seq data in mind. Using the AnnData [8] structure as a prototype, it provides a thumbnail overview (Fig. 1 a) that highlights key data components and their relationships. A detailed data information list (Fig. 1 b) below displays variable information, such as key names, dimensions, and data types. The DSV panel integrates with other panels, allowing users to dynamically inspect data structure changes and access visualization recommendations via trigger buttons.

The **Flowchart Visual Programming Panel (FVP)** serves as the central operational interface in *scFlowVis*, combining flowchart design with visual programming. Users can construct and document their analysis pipeline by connecting blocks for different operations, reducing the programming workload and enhancing workflow presentation (R1, R3, and R6). The panel features a library of scRNA-seq data processing methods (Fig. 1 c) and a user operation board (Fig. 1 d) for adding and linking blocks. It integrates with other panels to reflect updates and maintain consistency across the tool.

The **Visualization Methods Recommendation Panel (VMR)** supports data visualization needs with a button group (Fig. 1 i) for

managing the workflow and an export function for sharing workflows (R5). The core feature is the visualization recommendation engine (Fig. 1 j), which suggests appropriate methods based on data attributes using a rule-based algorithm (R4). Users can incorporate recommended visualization blocks into their flowcharts, ensuring a seamless transition from data exploration to visualization. An organized list of visualization methods (Fig. 1 k) groups techniques by analysis stage for easy access (R1).

The **Visualization Results Display Panel (VRD)** presents all visualization outcomes, designed with four main functional areas. Users can customize the display area (Fig. 1 h) with various layout options to compare multiple visualizations simultaneously (R3). The visualization list (Fig. 1 f) stores results chronologically, allowing users to activate and display them as needed. Interactive charts enable data engagement, with selected elements displaying information in a textbox (Fig. 1 g), which also serves as a note-taking area to integrate data interaction with user insights.

5 DISCUSSION AND FUTURE WORK

Discussions with domain experts highlighted some limitations of *scFlowVis*: 1) its methodological scope, while comprehensive, lacks the latest and deep learning-related methods, and 2) its local-only installation limits sharing and exporting workflows due to the absence of public server deployment. To address data privacy concerns and backend API accessibility, we maintained local installation while releasing the code to an open-source repository. This approach allows users to deploy the tool locally or on their servers, balancing accessibility and privacy and fostering continuous enhancement and community contribution.

For future work, we aim to expand *scFlowVis* in two key areas: 1) updating and enriching data processing methods, particularly integrating the latest techniques, and 2) extending the tool's applicability across various bioinformatics domains beyond scRNA-seq analysis, based on user feedback. By making *scFlowVis* open-source, we encourage community contributions to incorporate more methods and enhance functionalities.

REFERENCES

- [1] M. M. Burnett and D. W. McIntyre. Visual programming. *Computer*, 28(3):14–16, mar 1995. doi: 10.5555/618995.620167 2
- [2] B. Cakir, M. Prete, N. Huang, S. Van Dongen, P. Pir, and V. Y. Kiselev. Comparison of visualization tools for single-cell rnaseq data. *NAR Genomics and Bioinformatics*, 2(3):lqaa052, 2020. doi: 10.1093/nargab/lqaa052 1
- [3] M. Chakhchoukh, N. Boukhefifa, and A. Bezerianos. Understanding how in-visualization provenance can support trade-off analysis. *IEEE Transactions on Visualization and Computer Graphics*, 29(9):3758–3774, 2023. doi: 10.1109/TVCG.2022.3171074 2
- [4] V. Jalili, E. Afgan, Q. Gu, D. Clements, D. Blankenberg, J. Goecks, J. Taylor, and A. Nekrutenko. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, 48(W1):W395–W402, 06 2020. doi: 10.1093/nar/gkaa434 1
- [5] B. A. Myers. Visual programming, programming by example, and program visualization: a taxonomy. *SIGCHI Bull.*, 17(4):59–66, apr 1986. doi: 10.1145/22339.22349 2
- [6] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213 1
- [7] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902.e21, June 2019. doi: 10.1016/j.cell.2019.05.031 2
- [8] I. Virshup, S. Rybakov, F. J. Theis, P. Angerer, and F. A. Wolf. anndata: Annotated data. *bioRxiv*, Dec 2021. doi: 10.1101/2021.12.16.473007 2