Bioinformatics

https://doi.org/10.1038/s41592-026-03018-9

Pushing the boundaries of autonomous biological discovery

Juexiao Zhou, Xiaonan He, Kai Kang & Xin Gao

A new AI framework autonomously explores single-cell RNA sequencing datasets and produces creative, biologically relevant findings.

Modern single-cell biology is routinely generating datasets of unprecedented scale and complexity, particularly those from single-cell RNA sequencing (scRNA-seq)[1]. These datasets allow gene expression to be resolved at cellular resolution across diverse tissues, developmental stages and disease contexts. However, the richness of the data brings steep analytical challenges. Their dimensionality and structural diversity often confine analyses to the scope of a researcher's own expertise, computational skill set and creative imagination. As a result, a vast landscape of possible hypotheses remains unexplored and important biological signals can be overlooked. Current AI agents for biology, although promising, typically act as responsive executors that perform user-specified analyses faithfully[2,3]. They rarely initiate investigations or build on what has already been attempted, which means that the 'unknown unknowns' in complex datasets often remain hidden.

Addressing this limitation, CellVoyager[4], developed by Alber et al., is an autonomous, large language model–driven computational biology agent specifically designed for scRNA-seq analysis. CellVoyager takes in the dataset and a detailed record of prior analyses, summarizes the biological background and generates stepwise, hypothesis-driven workflows (Fig. 1a). Each workflow begins from a fresh analytical idea rather than revisiting existing ones. Operating inside a controlled, reproducible Jupyter environment equipped with leading single-cell analysis packages such as Scanpy[5] and scVI-tools[6], the agent executes analyses in real time, critiques intermediate outputs and iteratively refines its approach. This architecture enables CellVoyager to retain domain awareness, avoid redundant work and pivot toward promising directions revealed during the analysis.

To systematically measure its ability to propose relevant analyses, the team created CellBench, a benchmarking suite drawn from 76 scRNA-seq publications comprising 659 reported analyses (Fig. 1b). Given only the background sections of these papers and without access to methods or results sections, CellVoyager was tasked with predicting analyses the authors had actually performed. The agent outperformed strong large language model baselines (o3-mini and GPT-4o) by up to 23.5 percent. This result highlights the impact of explicit planning, iterative reflection and guideline-informed design in producing biologically appropriate workflows typical of single-cell studies.

The true test, however, came in three real-world case studies, each using published datasets already extensively analyzed by human experts. In a COVID-19 peripheral blood mononuclear cell atlas, CellVoyager introduced a completely new line of inquiry: assessing pyroptosis, a pro-inflammatory form of cell death[7]. While the original paper focused on antigen presentation and interferon signaling, the agent's analyses revealed increased pyroptosis scores in $CD8^+$ T cells from patients with COVID-19, validated across two independent datasets. In a human endometrium atlas, CellVoyager expanded receptor–ligand correlation analyses between stromal fibroblasts and other cell types, identifying menstrual cycle–dependent FGF2–FGFR1 and TGF-β signaling patterns not explored originally[8]. In a mouse brain aging dataset, the agent quantified transcriptional noise across neurogenic niche cell types, finding significant age-associated increases in oligodendrocytes, microglia and mural cells[9]. These observations were reproduced with an independent dataset. Across all case studies, evaluations by original paper authors and independent reviewers rated CellVoyager's outputs highly for creativity, with a mean score of 3.03 out of 4, and biological relevance, with 80% of hypotheses judged compelling enough to merit further exploration.

Furthermore, CellVoyager's modular architecture enables human-in-the-loop iteration. Expert feedback can be integrated into subsequent runs to refine strategies, correct methodological weaknesses and increase statistical rigor without sacrificing novelty. Because the framework maintains a detailed record of previous analyses and results, researchers can guide the AI toward more promising investigative threads or away from already well-explored terrain. This collaborative mode hints at a future where autonomous agents and human scientists work in a genuinely symbiotic manner, during which AI systems are running persistently in the background to explore datasets, flagging unexpected but plausible findings, and proposing high-confidence hypotheses ready for biomolecular or clinical validation. Over time, such systems could streamline the research cycle, continually mining newly published datasets, integrating public and in-house data resources, and even adjusting their search strategies in real time on the basis of evolving priorities in a lab or the broader scientific field. In this vision, the human role shifts toward strategic oversight, experimental design and nuanced interpretation while the AI agent undertakes the exhaustive, hypothesis-generating exploration, transforming both the pace and scope of discovery.

Where might such agents go next? The authors anticipate extending CellVoyager beyond scRNA-seq to other high-dimensional domains including spatial transcriptomics, proteomics and multimodal omics, supported by modality-specific tools and guidelines. They also emphasize the untapped opportunity for systematic reanalysis at scale. With more than 100,000 single-cell studies already published, exhaustive re-examination by human researchers is unrealistic. Autonomous agents could continually mine this corpus, uncover overlooked biological signals and generate new insights without further experimental cost.

Ultimately, CellVoyager signals an important shift in computational biology from reactive assistants toward proactive collaborators. By combining LLM reasoning with domain-specific execution and awareness of prior work, it shows how AI can extend the life cycle of

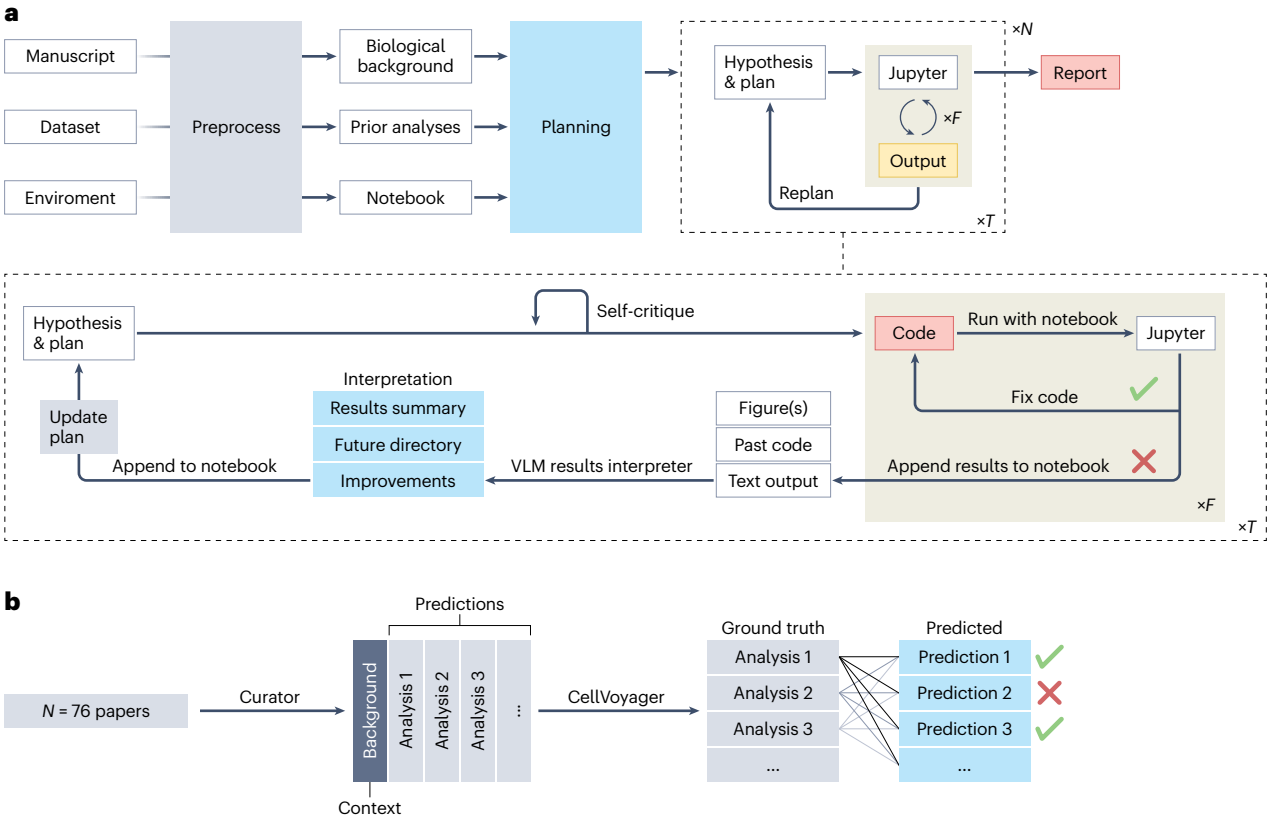

Fig. 1 | Overview of CellVoyager's design, evaluation, and biological applications. a, Workflow schematic showing how CellVoyager ingests a processed scRNA-seq dataset together with a record of prior analyses, summarizes the biological background and generates stepwise, hypothesis-driven workflows. These workflows are executed inside a reproducible Jupyter environment equipped with domain-specific packages, with iterative refinement based on self-critique. **b**, Benchmarking results from the CellBench suite, comprising 76 published scRNA-seq studies and 659 reported analyses. Given only the background sections of each paper, CellVoyager more accurately predicted analyses performed by the authors than did baseline large language models, achieving up to 23.5% better accuracy than GPT-4o.

biological datasets, maximize the return on existing data and accelerate the emergence of new biological understanding.

Juexiao Zhou [1,2,3,4,7], **Xiaonan He** [5,7], **Kai Kang** [6] ✉ & **Xin Gao** [1,2,3] ✉
[1]Computer Science Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia. [2]Center of Excellence for Smart Health, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia. [3]Center of Excellence on Generative AI, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia. [4]School of Data Science, Chinese University of Hong Kong, Shenzhen (CUHK–Shenzhen), Guangdong, China. [5]Emergency Critical Care Center, Beijing AnZhen Hospital, Capital Medical University, Beijing, China. [6]Department of Critical Care Medicine, First Affiliated Hospital of Harbin Medical University, Harbin, China. [7]These authors contributed equally: Juexiao Zhou, Xiaonan He.
✉e-mail: janekk79@126.com; xin.gao@kaust.edu.sa

Published online: 17 March 2026

References

1. Tirosh, I. & Suva, M. L. *Cancer Cell* **42**, 1497–1506 (2024).
2. Zhou, J. et al. *Adv. Sci.* **11**, 2407094 (2024).
3. Zhou, J., Jiang, J., Han, Z., Wang, Z. & Gao, X. *Brief. Bioinform.* **26**, bbaf505 (2025).
4. Alber, S. & Chen, B. *Nat. Methods* https://doi.org/10.1038/s41592-026-03029-6 (2026).
5. Wolf, F. A., Angerer, P. & Theis, F. J. *Genome Biol.* **19**, 15 (2018).
6. Gayoso, A. et al. *Nat. Biotechnol.* **40**, 163–166 (2022).
7. Wilk, A. J. et al. *Nat. Med.* **26**, 1070–1076 (2020).
8. Wang, W. et al. *Nat. Med.* **26**, 1644–1653 (2020).
9. Buckley, M. T. et al. *Nat. Aging* **3**, 121–137 (2023).

Competing interests

The authors declare no competing interests.