https://doi.org/10.1038/s41588-026-02584-0

The Single Cell Notebooks for inclusive and accessible training in single-cell and spatial omics

Adolfo Rojas-Hidalgo, Raúl Arias-Carrasco, Joyce Karoline Silva, Erick Armingol, Sebastián Urquiza-Zurich, Bruno Vinagre, Cesar A. Prada-Medina, Sergio Triana, Daniela D. Russo, Diego Pérez-Stuardo, Emiliano Vicencio, Gerardo Muñoz, Cristóvão Antunes de Lanna, Leandro Santos, Gabriela Rapozo, Natalia Tavares, Andrés Moreno-Estrada, John Randell, Patricia Severino, Ricardo Khouri, Orr Ashenberg, Alex K. Shalek, Mariana Boroni, Yesid Cuesta-Astroz, Benilton S. Carvalho & Vinicius Maracaja-Coutinho

Check for updates

The Single Cell Notebooks provide multilingual, open-access training materials for single-cell and spatial transcriptomics analysis through reusable notebooks. They have evolved into a community-driven platform for education and capacity building, lowering language and computational barriers to support equitable participation in omics research.

Single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics have emerged as transformative technologies in contemporary biology, enabling gene expression analysis at the level of individual cells and with spatial resolution[1–3]. Effective implementation of these technologies demands not only experimental precision, but also interdisciplinary training in statistics, bioinformatics and data science.

Despite these advances, a considerable gap remains between data generation and the skills required for effective analysis, particularly given the volume and complexity of modern scRNA-seq and spatial datasets[4–6]. These requirements are more difficult to meet in under-resourced settings, in which limited funding, limited internet connectivity and scarce computing resources constrain access to state-of-the-art workflows. The predominance of English-only training materials further limits the access of researchers in the global south to cutting-edge tools. Capacity-building and training initiatives should therefore be tailored to local contexts, ensuring accessibility under constrained computational infrastructure and fostering equitable participation in the global research ecosystem[7,8].

To overcome these challenges, we developed 'The Single Cell Notebooks' (scNotebooks), an open-access training suite for single-cell and spatial transcriptomics data analysis, comprising 13 modules (Table 1), which are available in English, Portuguese and Spanish. Hosted on GitHub (https://github.com/integrativebioinformatics/scNotebooks), the notebooks run in a web browser via cloud services (for example, Google Colab; no local installation required) or locally via Jupyter and Docker (local setup required; installation instructions provided in the GitHub repository). The materials are also mirrored on a dedicated website (https://integrativebioinformatics.github.io/scNotebooks/).

scNotebooks combine hands-on tutorials, end-to-end workflows and real-world datasets to guide learners through core concepts and tools for scRNA-seq and spatial transcriptomics data analysis. Since 2021, they have been adopted in multiple training activities in Africa, the Americas and Asia, and their impact is now emerging within local research communities[7,8] (Box 1).

A competency framework for single-cell and spatial omics data analysis

For inclusive and effective scRNA-seq and spatial transcriptomics training, a clear set of interdisciplinary competencies is essential. Accordingly, we curated a practical competency framework to guide the scope and progression of scNotebooks. Although expertise in multiple areas, such as molecular and cellular biology, computational workflows, statistics, data management and visualization, is valuable, no single researcher is expected to master them all. Instead, the notebooks were designed to establish a shared set of core foundations for all learners, while enabling individuals to selectively engage with specialized modules aligned with their interests.

A cornerstone of scRNA-seq analysis is computational expertise, enabling the transformation of raw sequencing data into biologically meaningful insights. Familiarity with R or Python underpins the use of libraries for normalization, clustering, dimensionality reduction and downstream analysis, and this is integrated into the notebooks through hands-on exercises. Additional skills, such as automating pipelines with UNIX or Linux shell scripting, encapsulating environments (for example, Nextflow) to safeguard reproducibility, and applying version control for transparent code sharing, management and collaborative development, can be further strengthened through specialized courses, online forums and collaborative projects beyond the notebooks.

In parallel, grounding in molecular and cellular biology provides essential context for meaningful analysis. Understanding gene regulation, transcriptional dynamics and cell-fate decisions enables researchers to interpret the heterogeneity revealed by single-cell data and combine more advanced analysis to extract meaningful biological insights[3,9,10]. These biological concepts are reinforced across modules

Table 1 | General overview of scNotebooks curriculum

Module	Brief description
Module 01: Introduction to notebooks and databases	In this module, learners explore the structure of Jupyter Notebooks and Google Colab as interactive environments for learning and data analysis, focusing on the use of code and text cells. They also become familiar with major public repositories and portals for single-cell gene expression data across humans and other organisms. Through hands-on exercises, users practice accessing, exploring and working with datasets and metadata.
Module 02: Introduction to R and data visualization with ggplot2	Learners are introduced to the R programming language, a widely used environment for statistical analysis and data manipulation in bioinformatics. The notebooks cover basic syntax and core data structures (for example, vectors and data frames) and introduce the package ggplot2 for data visualization. Through hands-on exercises, users practice writing R codes, generating their own plots and interpreting biological datasets, establishing a solid foundation for subsequent analytical modules.
Module 03: Processing raw scRNA-seq data	This module introduces essential Linux command-line operations, focusing on core commands that are broadly transferable across computational workflows and programming languages. The notebook then walks through the main steps of processing raw single-cell sequencing reads into gene-cell count matrices with Cell Ranger, highlighting key outputs, quality-control summaries and how these products feed into downstream analyses.
Module 04: Quality control, clustering, annotation, and functional analysis	In this module, learners will import raw count matrices, perform filtering and quality-control visualizations and assess standard quality-control metrics. The notebooks then guide users through normalization to mitigate technical variability, dimensionality reduction and clustering to resolve cellular heterogeneity. Building on this foundation, participants carry out gene marker discovery, differential expression and abundance analyses, cell-type annotation and functional enrichment to connect transcriptional programs to development and disease. The module also introduces practical approaches to diagnose and mitigate ambient RNA contamination.
Module 05: Integrating single-cell transcriptomes from multiple samples	This notebook focuses on data integration and batch correction, key steps for ensuring that biological signals are not confounded by technical variation or donor-specific effects. The notebooks examine how differences in protocols, platforms and sample origin introduce batch effects, and how the choice of batch covariates shapes integration results. Through hands-on exercises with Seurat and Harmony, users apply correction methods, tune key parameters and assess integration quality using standard diagnostics. Benchmarking activities compare alternative strategies, illustrating practical trade-offs between removing unwanted variation and preserving meaningful biological structure.
Module 06: Trajectory inference and pseudotemporal ordering	The module explores how scRNA-seq data can be used to reconstruct cell-state trajectories and model dynamic changes during development or differentiation. The notebooks introduce key concepts behind trajectory inference, including how gradual transcriptional programs can be captured and ordered computationally. Through hands-on activities with Monocle3, users perform pseudotime analysis and trajectory reconstruction, enabling them to map putative lineage paths, identify genes that vary along transitions and interpret functional shifts across cellular states.
Module 07: Deciphering cell–cell communication in single-cell transcriptomics data	In this notebook, learners explore how signaling interactions can be inferred from scRNA-seq data, with an emphasis on curated ligand–receptor resources and the assumptions underlying these approaches. Through hands-on activities, users apply multiple state-of-the-art methods to compare interaction predictions, assess consistency across tools and interpret biological relevance in context.
Module 08: Multimodal single-cell analysis	In this module, learners analyze umbilical cord blood mononuclear cells to examine the relationship between RNA expression and surface protein abundance. Working with paired RNA and antibody-derived tag count matrices, users explore concordance and divergence between modalities, interpret marker patterns and connect these profiles to immune cell identity and function.
Module 09: TCR profiling in single-cell analysis	In this module, learners examine how T cell receptor (TCR) sequencing captures clonal diversity and antigen-driven expansion, and how CITE-seq integrates transcriptomic and surface-protein measurements to refine cell-state and cell-type resolution. Building on these core concepts, hands-on activities guide participants through practical analysis workflows to characterize immune repertoires, link clonotypes to transcriptional phenotypes and interpret multimodal datasets.
Module 10: An introduction to spatial transcriptomics approaches	Spatial transcriptomics provides a way to map gene expression directly within tissues, revealing cell types, their distribution and interactions in native contexts. In this module, learners will examine how spatially resolved profiles advance our understanding of development and disease. Practical activities guide users to recover gene expression across regions, and apply cell-type deconvolution approaches to interpret spatial organization.
Module 11: An introduction to single cell assay for transposase-accessible chromatin sequencing (scATAC-seq)	In this module, hands-on activities guide users through core steps including for scATAC-seq data analysis, including quality control, dimensionality reduction, peak/feature definition, motif and transcription factor activity analysis, and linkage to putative target genes. The module also covers integration with matched scRNA-seq profiles and includes benchmarking exercises to compare integration strategies and support robust interpretation of gene regulatory programs.
Module 12: Understanding alternative polyadenylation (APA)	In this module, learners will explore scRNA-seq data to detect, quantify and interpret APA events. The workflow introduces input formats, preprocessing steps, and strategies for visualizing APA dynamics across cell types.
Module 13: Fair data and sharing data	In this notebook, learners practice structuring and documenting metadata using consistent standards and assemble submission-ready packages for public repositories and data portals. The notebooks emphasize common file formats, required metadata fields and practical submission workflows, helping ensure that datasets remain discoverable, interoperable and readily reusable across platforms and studies.

A brief description of the 13-module program.

BOX 1

Early uptake and educational outcomes of scNotebooks

scNotebooks is a multilingual training initiative designed to lower barriers to single-cell and spatial transcriptomics analysis. Since 2021, the project has expanded to 13 modules (39 notebooks across 3 languages; 1 notebook per module per language) and has been adopted in multiple international courses and workshops, reaching 314 participants.

Quantifying longer-term outcomes (for example, publications attributable to the training) is currently challenging because most scNotebooks-based courses were delivered by the core team in collaboration with partner institutions that manage enrollment and evaluation locally, and participant-level follow-up is not centrally collected across jurisdictions with different data-protection frameworks. In addition, the canonical scNotebooks resources (repository and dedicated website) are being released in their current public-facing form alongside this publication, so web-based usage analytics are not yet informative.

To provide evidence of early educational and research impact, we highlight here user-reported outcomes that illustrate how scNotebooks support both research training and local capacity building.

"I adapted scNotebooks for *Caenorhabditis elegans* scRNA-seq during my undergraduate research at the University of São Paulo. The notebooks were essential for producing the results in my dissertation and significantly accelerated the analysis while improving my understanding of the full scRNA-seq workflow." — Felipe Sánchez (undergraduate student in applied mathematics, University of São Paulo, Brazil).

"I used scNotebooks in our Center's journal clubs to demonstrate what single-cell approaches can enable, which helped catalyze a collaboration to implement single-cell methods locally. I am now adapting scNotebooks for a doctoral elective course on omics technologies." — Juan Francisco Calderón (associate professor, Universidad del Desarrollo, Chile).

Together, these examples illustrate how scNotebooks can be reused and adapted beyond short-format workshops, supporting thesis work, local dissemination and incorporation into formal graduate-level teaching and research.

through datasets that base computational steps on interpretable biological questions.

Competency in public data repositories and reuse is also fundamental. The notebooks introduce guided steps to retrieve and handle public datasets and metadata from major repositories and portals, including the Gene Expression Omnibus (GEO)[11], the Human Cell Atlas Data Portal (https://data.humancellatlas.org/), CELLxGENE[12] and the Single-Cell Expression Atlas[13], thereby allowing the user to progressively develop the capacity to expand analyses or initiate new projects by integrating and reusing public data.

Taken together, this competency framework provides a common reference of the analytical workflow, clarifying what is needed, what is produced and the assumptions made at each stage. In this way, individuals and teams can build complementary expertise without requiring every member to learn every component at once.

Development and design of scNotebooks

The scNotebooks initiative began in 2021 with an inaugural online workshop linked to the Human Cell Atlas focused on Latin America, where early versions were first deployed for hands-on training. In 2023, the initiative shifted to in-person workshops in Salvador, Rio de Janeiro (Brazil) and Santiago (Chile), deepening regional capacity in South America and building networks of local instructors. By 2024, the same workshop-based training approach extended across continents, including activities in Bangkok (Thailand), Bogotá (Colombia), Dakar (Senegal), Nairobi (Kenya) and San José (Costa Rica) (Fig. 1). These courses were delivered by the core scNotebooks team together with local co-instructors at host institutions and supported by research grants, philanthropic funding and institutional contributions.

The educational resources were developed as open-access, multilingual materials to support hands-on training in single-cell and spatial transcriptomics. Hosted on GitHub and via a dedicated website, they were designed for use in both Google Colab and Jupyter platforms, integrating executable codes, interactive visualizations and explanatory text in a single document, connecting theory with practice. Additionally, a Docker environment was developed to encapsulate dependencies and configurations for the Jupyter-based version, ensuring reproducibility and seamless local execution across systems. Content is maintained by the core scNotebooks team through a version-controlled GitHub workflow (issues, reviewed pull requests and tagged releases), enabling regular updates and community contributions.

By leveraging cloud-based execution, scNotebooks ensure accessibility without local installation, a key advantage in diverse educational settings. Google Colab is especially valuable in this scenario, as it offers real-time collaboration and free-tier graphics processing unit (GPU) and tensor processing unit (TPU) resources, supporting the use of computationally demanding techniques. To accommodate Colab's resource limitations, datasets were manually reduced, allowing users to explore clustering, trajectory inference and other techniques without the need for a high-performance infrastructure.

The notebooks are structured into 13 modules, progressing from foundational concepts to advanced approaches. Modules 1–2 introduce public databases and R programming. Modules 3–5 cover the foundations of scRNA-seq analysis, moving through essential steps, such as processing raw sequencing data, quality control, clustering, cell type annotation (both computational and manual), functional analysis and sample integration with batch correction. Modules 6–8 introduce more advanced techniques, including trajectory inference and pseudotime analysis, cell–cell communication and multimodal single-cell analysis. Modules 9–12 provide specialized workflows, including T cell receptor profiling, spatial transcriptomics, single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) and alternative polyadenylation characterization. Finally, module

13 introduces the FAIR principles (findable, accessible, interoperable and reusable), emphasizing good data practices and strategies for sharing datasets in public repositories (Table 1).

By combining technical training with principles of reproducibility and openness, scNotebooks provide a framework that supports immediate analytical needs, while building sustainable capacity for future research across diverse contexts. Although initially developed for human biomedical contexts, the notebooks can be easily adapted to broader applications. For non-human datasets (for example, plants or non-model organisms), users should tailor reference annotations and gene marker sets. Likewise, for cell–cell communication analyses, organism-appropriate ligand–receptor resources (including ortholog mappings where needed) are essential to ensure biologically meaningful results.

Strengths, limitations and practical considerations of scNotebooks

For researchers and students entering the world of single-cell analysis, scNotebooks offer a structured and accessible entry point into a complex and rapidly evolving field. Their modular design and curated workflows can be empowering, helping users build foundational skills and explore advanced topics with guided support. However, some techniques are presented in a simplified or introductory format, which can restrict their immediate applicability to highly specialized or large-scale research questions.

To offer practical experience with widely adopted tools in the scientific community, scNotebooks incorporate commonly used packages. However, many of these tools are sensitive to parameter choices, which can introduce bias, for example by excluding rare cell types. Similarly, cell–cell interaction predictions depend on curated ligand–receptor databases that may be incomplete or context-specific[14]. Spatial transcriptomics may provide spatial context, but its methods can suffer from constraints regarding resolution or gene coverage, limiting the biological insights they can support. Additionally, some of the notebooks lack comprehensive benchmarking across tools, making it difficult to guide methodological selection.

The pedagogical structure of the notebooks supports progressive learning, allowing modules to be used independently or sequentially. The multilingual availability expands regional accessibility, and deployment in both in-person and virtual workshops has enabled iterative refinement of materials to better fit diverse teaching contexts. However, some chapters may benefit from additional theory to reinforce statistical and biological principles underlying computational analyses. For example, module 4 could be supported by deeper knowledge of dimensionality-reduction methods (for example, principal component analysis (PCA) and uniform manifold approximation and projection (UMAP)) and pathway enrichment approaches. Similarly, the cell–cell communication content can be complemented with background on ligand–receptor biology, and the spatial transcriptomics module with studies of tissue architecture and strategies for integrating spatial and single-cell data.

Although Google Colab improves accessibility in low-infrastructure settings, limits in session time and memory, as well as dependency on internet connectivity, can constrain compute-intensive steps such as the use of Cell Ranger, the 10x Genomics pipeline for processing raw data into gene expression matrices. To address this, we provide instructional subsets, including a small selection of samples from the original datasets and truncated FASTQ files (first ~1,000 reads) for the raw-data processing to enable rapid execution. This allows learners to run workflows within Colab's typical resource limits while still following the complete pipelines. For environments with stronger computational resources, local execution in Jupyter (optionally containerized with Docker) provides flexibility, uninterrupted workflows (no cloud session timeouts), deeper customization of software versions and parameters, and the use of full datasets rather than instructional subsets. Local execution also facilitates integration with institutional storage and internal pipelines, enabling straightforward reuse and adaptation of the notebooks within local projects.

Overall, the strengths of scNotebooks lie in their accessibility, relevance and flexibility, whereas limitations largely reflect the inherent challenges of teaching complex, rapidly evolving methodologies and resource-intensive analysis. By combining an interactive format with open and reusable infrastructure, and iteratively refining the materials through community feedback from workshops, we provide a robust foundation for high-resolution transcriptomics training and a framework for continuous improvement. A GitHub-hosted forum further supports adoption, peer learning and collaboration.

Global reach and capacity building

Across formal, hands-on courses and workshops, our core team directly trained 314 participants from 33 countries with strong representation from the global south (Fig. 1). Here, participants refer to trainees enrolled in instructor-led activities, rather than passive online users. Latin America and the Caribbean accounts for 60.8% of trainees (191 individuals from 13 countries), followed by Asia (19.4%, with 61 participants from 13 countries), Africa (18.8%, with 59 participants from 6 countries) and Oceania (1%, with 3 individuals from 1 country). Evidence of downstream uptake through teaching and research reuse is illustrated in Box 1.

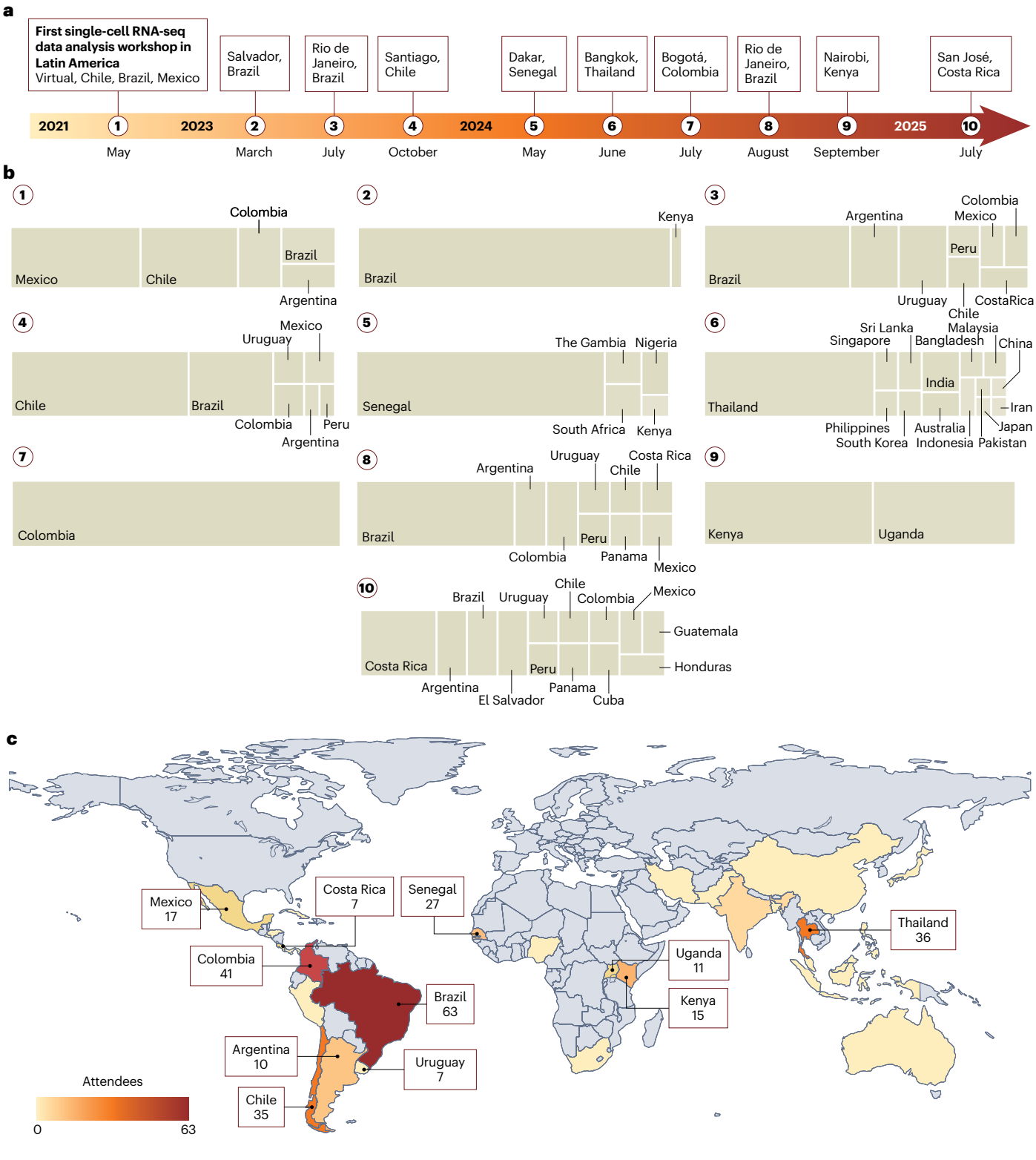
Beyond these formal training events, the open-access materials have enabled broader adoption, with researchers and students using the multilingual notebooks for self-directed learning and local institutional training, often adapted to specific curricula and research contexts (Box 1). Aggregate usage metrics for independent online users are not available, because the public repository and project website are being released alongside this publication. Looking ahead, we plan to expand language support beyond the current languages, including potential versions in Arabic, French, Hindi and Mandarin Chinese, in

Fig. 1 | Global reach and geographic distribution of scNotebooks training programs. a, Timeline of the ten workshops and courses organized between 2021 and 2025, showing the chronological expansion of training activities. The initiative began in May 2021 with a virtual meeting connecting participants from Chile, Brazil and Mexico, coordinated by the Human Cell Atlas Latin America. In 2023, the first in-person course took place in Salvador, followed by Rio de Janeiro, both in Brazil, and Santiago, Chile. These courses expanded the scNotebooks content and introduced the modular structure. In 2024, training reached Dakar (Senegal), Bangkok (Thailand), Bogotá (Colombia), Rio de Janeiro again and Nairobi (Kenya). In 2025, the program extended to San José (Costa Rica). **b**, Graphic representing the frequency of trainees according to participant countries across all training events, with box size proportional to the number of individuals from each country. **c**, Geographic heatmap displaying the distribution of 314 participants across 33 countries, with the top 10 countries (including ties) annotated. Brazil was the largest hub with 63 participants, followed by Colombia (41), Thailand (36), Chile (35) and Senegal (27). The color gradient indicates the number of participants. Publ. note: Springer Nature is neutral about jurisdictional claims in maps.

collaboration with regional partners. Together, hands-on workshops, cloud-based implementation and continuous community-driven refinement supported this sustainable model for democratizing single-cell training in resource-limited settings.

Conclusions and perspectives

scNotebooks provide a practical, multilingual platform that lowers barriers to learning and applying single-cell and spatial transcriptomics. Through repeated deployment in courses and workshops, the materials



have been iteratively refined based on learner feedback, with worked examples adapted to local contexts. Addressing regionally relevant questions in training activities, such as locally prevalent diseases, environmental exposures or biodiversity, can foster deeper engagement and contextual understanding.

Future efforts will include benchmarking-oriented modules to help users compare tools and make best-practice choices by using transparent, reproducible criteria. We also plan to expand language support and new data modalities, broadening the scope of training. Sustained community engagement, through train-the-trainer programs and responsiveness to evolving methods via GitHub, will be key to maintaining scNotebooks as an open, multilingual resource adaptable to diverse data-intensive research areas.

Adolfo Rojas-Hidalgo[1,22], Raúl Arias-Carrasco[2,22], Joyce Karoline Silva[1,3,4,22], Erick Armingol[5], Sebastián Urquiza-Zurich[1], Bruno Vinagre[3], Cesar A. Prada-Medina[6], Sergio Triana[7,8,9], Daniela D. Russo[7,8,9,10], Diego Pérez-Stuardo[1], Emiliano Vicencio[1], Gerardo Muñoz[11], Cristóvão Antunes de Lanna[12], Leandro Santos[12], Gabriela Rapozo[12], Natalia Tavares[3,13,14], Andrés Moreno-Estrada[15], John Randell[16], Patricia Severino[17], Ricardo Khouri[3,4,18,19], Orr Ashenberg[7], Alex K. Shalek[7,8,9], Mariana Boroni[12], Yesid Cuesta-Astroz[20], Benilton S. Carvalho[21] & Vinicius Maracaja-Coutinho[1,3,19]

[1]Unidad de Genómica Avanzada - UGA, Advanced Center for Chronic Diseases - ACCDiS, Departamento de Bioquímicas y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, Santiago, Chile. [2]Instituto Universitario de Investigación y Desarrollo Tecnológico - IDT, Universidad Tecnológica Metropolitana - UTEM, Santiago, Chile. [3]Laboratório de Medicina e Saúde Pública de Precisão - MESP2, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz (Fiocruz), Salvador, Brazil. [4]Programa de Pós-graduação em Ciências da Saúde - PPGCS, Faculdade de Medicina da Bahia, Universidade Federal da Bahia, Salvador, Brazil. [5]Wellcome Sanger Institute, Hinxton, UK. [6]Novo Nordisk, Oxford, UK. [7]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [8]Ragon Institute of Massachusetts General Brigham, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [9]Institute for Medical Engineering and Science (IMES), Department of Chemistry, and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [10]Department of Immunology, Harvard Medical School, Boston, MA, USA. [11]Ingeniería Civil en Bioinformática, Universidad de Talca, Talca, Chile. [12]Laboratorio de Bioinformatica e Biologia Computacional, Brazilian National Cancer Institute - INCA, Rio de Janeiro, Brazil. [13]Instituto Nacional de Ciência e Tecnologia, Instituto de Investigação em Imunologia (INCT - iii), São Paulo, Brazil. [14]Instituto Nacional de Ciência e Tecnologia em Doenças Crônicas das Mucosas e Pele (INCT - Mucosa e Pele), Belo Horizonte, Brazil. [15]Human Evolutionary and Population Genomics Laboratory, Centro de Investigación sobre el Envejecimiento, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), Mexico City, Mexico. [16]Human Cell Atlas, Cambridge, MA, USA. [17]Instituto Israelita de Ensino e Pesquisa Albert Einstein, Hospital Israelita Albert Einstein, São Paulo, Brazil. [18]Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Leuven, Belgium. [19]Instituto Nacional de Ciência e Tecnologia em Saúde Digital (INCT-DigiSaúde), Salvador, Brazil. [20]Escuela de Microbiología, Universidad de Antioquia, Medellín, Colombia. [21]Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, São Paulo, Brazil. [22]These authors contributed equally: Adolfo Rojas-Hidalgo, Raúl Arias-Carrasco, Joyce Karoline Silva.
✉e-mail: vinicius.maracaja@fiocruz.br

Published online: 05 May 2026

References

1. Picelli, S. *RNA Biol.* **14**, 637–650 (2017).
2. Maracaja-Coutinho, V. & Severino, P. in *Transcriptomics in Health and Disease* (ed. Passos, G. A.) 127–141 (Springer, 2022).
3. Wang, S. et al. *Int. J. Mol. Sci.* **24**, 2943 (2023).
4. Atta, L. & Fan, J. *Nat. Commun.* **12**, 5283 (2021).
5. Lähnemann, D. et al. *Genome Biol.* **21**, 31 (2020).
6. Liu, X. et al. *J. Hematol. Oncol.* **17**, 72 (2024).
7. Possik, P. A. et al. *Cell* **188**, 5790–5796 (2025).
8. Maracaja-Coutinho, V. & Nakaya, H. I. *Nat. Rev. Mol. Cell Biol.* **27**, 1–3 (2026).
9. Didier, J. et al. *Endocr. Relat. Cancer* **32**, e250024 (2025).
10. Hemberg, M. et al. *Genome Biol.* **26**, 358 (2025).
11. Clough, E. et al. *Nucleic Acids Res.* **52**, D138–D144 (2024).
12. CZI Cell Science Program et al. *Nucleic Acids Res.* **53**, D886–D900 (2025).
13. George, N. et al. *Nucleic Acids Res.* **52**, D107–D114 (2024).
14. Armingol, E., Baghdassarian, H. M. & Lewis, N. E. *Nat. Rev. Genet.* **25**, 381–400 (2024).

Acknowledgements

This work was supported by the Agencia Nacional de Investigación y Desarrollo (ANID, Chile): FONDECYT (1211731 to V.M.C. and 3250452 to D.P.S.), FONDAP (15130011 and 1523A0008 to V.M.C.), Anillo (ATE220016 to V.M.C.) and BECAS CHILE (2018–72190270 to E.A.); by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil), through the PrInt program (to V.M.C.); by the Chan Zuckerberg Initiative (CZI, USA) (to A.M.E., J.R., P.S., M.B., Y.C.A., B.S.C. and V.M.C.); by the Gates Foundation (to A.K.S.); by the European Union through a Marie Skłodowska-Curie Actions Postdoctoral Fellowship (grant agreement no. 101208051 to E.A.); and by Pew Latin American Fellowship Program in Biomedical Sciences (to S.T.). We are also grateful to the Wellcome Connecting Sciences for the continuous support in several practical workshops and symposiums organized over the years in the region.

Competing interests

The authors declare no competing interests.