

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/159750/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Czarnewski, Paulo, Mahfouz, Ahmed, Calogero, Raffaele A., Palagi, Patricia M., Portell-Silva, Laura, Gonzalez-Uriarte, Asier, Sonesson, Charlotte, Burdett, Tony, Szomolay, Barbara, Videm, Pavankumar, Hotz, Hans-Rudolf, Papatheodorou, Irene, Hancock, John M., Grüning, Björn, Haerty, Wilfried, Krause, Roland, Capella-Gutierrez, Salvador, Leskosek, Brane, Alessandri, Luca, Arigoni, Maddalena, Rezen, Tadeja, Botzki, Alexander, Ferk, Polonca, Lindvall, Jessica, Heil, Katharina F, Ishaque, Naveed and Korpelainen, Eija 2022. Community-driven ELIXIR activities in single-cell omics. *F1000Research* 11, 869. 10.12688/f1000research.122312.1 file

Publishers page: <http://dx.doi.org/10.12688/f1000research.122312.1>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





OPINION ARTICLE

Community-driven ELIXIR activities in single-cell omics

[version 1; peer review: awaiting peer review]

Paulo Czarnewski ¹, Ahmed Mahfouz ², Raffaele A. Calogero ³,
Patricia M. Palagi ⁴, Laura Portell-Silva⁵, Asier Gonzalez-Uriarte⁵,
Charlotte Soneson ^{4,6}, Tony Burdett⁷, Barbara Szomolay⁸, Pavankumar Videm⁹,
Hans-Rudolf Hotz ⁶, Irene Papatheodorou ⁷, John M. Hancock¹⁰,
Björn Grüning ⁹, Wilfried Haerty¹¹, Roland Krause ¹²,
Salvador Capella-Gutierrez ⁵, Brane Leskošek¹⁰, Luca Alessandri³,
Maddalena Arigoni³, Tadeja Rezen¹⁰, Alexander Botzki ¹³, Polonca Ferk¹⁰,
Jessica Lindvall¹, Katharina F. Heil ¹⁴, Naveed Ishaque ¹⁵, Eija Korpelainen ¹⁶

¹Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg, Luxembourg

²VIB Bioinformatics Core, Ghent, Belgium

³ELIXIR Hub, Wellcome Genome Campus, Hinxton, UK

⁴Digital Health Center, Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Berlin, Germany

⁵CSC – IT center for science, Espoo, Finland

⁶Science for Life Laboratory, Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Stockholm University, Solna, Sweden

⁷Department of Human Genetics, The Netherlands & Delft Bioinformatics Lab, Leiden University Medical Center, Delft University of Technology, Leiden, Delft, The Netherlands

⁸Bioinformatics and Genomics unit, Dept. Molecular Biotechnology and Health Science, University of Torino, Torino, Italy

⁹SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

¹⁰Barcelona Supercomputing Center, Barcelona, Spain

¹¹Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

¹²European Bioinformatics Institute, Hinxton, UK

¹³Division of Infection and Immunity, School of Medicine, Cardiff University, Wales, UK

¹⁴Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany

¹⁵Institute of Biochemistry and Molecular Genetics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

¹⁶Earlham Institute, Norwich, UK

V1 First published: 29 Jul 2022, 11(ELIXIR):869
<https://doi.org/10.12688/f1000research.122312.1>

Latest published: 29 Jul 2022, 11(ELIXIR):869
<https://doi.org/10.12688/f1000research.122312.1>

Open Peer Review

Approval Status AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Single-cell omics (SCO) has revolutionized the way and the level of resolution by which life science research is conducted, not only impacting our understanding of fundamental cell biology but also providing novel solutions in cutting-edge medical research. The rapid development of single-cell technologies has been accompanied by the active development of data analysis methods, resulting in a plethora of new analysis tools and strategies every year. Such a rapid

development of SCO methods and tools poses several challenges in standardization, benchmarking, computational resources and training. These challenges are in line with the activities of ELIXIR, the European coordinated infrastructure for life science data. Here, we describe the current landscape of and the main challenges in SCO data, and propose the creation of the ELIXIR SCO Community, to coordinate the efforts in order to best serve SCO researchers in Europe and beyond. The Community will build on top of national experiences and pave the way towards integrated long-term solutions for SCO research.

Keywords

Single cell, multi-omics, spatial transcriptomics, FAIR, data analysis, data standards, training, computing infrastructure



This article is included in the [ELIXIR](#) gateway.



This article is included in the [Bioinformatics](#) gateway.

Corresponding authors: Paulo Czarnewski (paulo.czarnewski@scilifelab.se), Naveed Ishaque (naveed.ishaque@bih-charite.de), Eija Korpelainen (eija.korpelainen@csc.fi)

Author roles: **Czarnewski P:** Data Curation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mahfouz A:** Writing – Original Draft Preparation, Writing – Review & Editing; **Calogero RA:** Writing – Original Draft Preparation; **Palagi PM:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Portell-Silva L:** Writing – Original Draft Preparation, Writing – Review & Editing; **Gonzalez-Uriarte A:** Writing – Original Draft Preparation, Writing – Review & Editing; **Sonesson C:** Writing – Review & Editing; **Burdett T:** Conceptualization; **Szomolay B:** Writing – Original Draft Preparation, Writing – Review & Editing; **Videm P:** Writing – Review & Editing; **Hotz HR:** Writing – Review & Editing; **Papatheodorou I:** Conceptualization; **Hancock JM:** Writing – Review & Editing; **Grüning B:** Writing – Review & Editing; **Haerty W:** Writing – Review & Editing; **Krause R:** Writing – Review & Editing; **Capella-Gutierrez S:** Writing – Original Draft Preparation; **Leskošek B:** Writing – Review & Editing; **Alessandri L:** Writing – Original Draft Preparation; **Arigoni M:** Writing – Original Draft Preparation; **Rezen T:** Writing – Review & Editing; **Botzki A:** Writing – Review & Editing; **Ferk P:** Writing – Review & Editing; **Lindvall J:** Conceptualization; **Heil KF:** Conceptualization, Writing – Review & Editing; **Ishaque N:** Writing – Original Draft Preparation, Writing – Review & Editing; **Korpelainen E:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by ELIXIR, the research infrastructure for life-science data. P.C. is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at ScilifeLab. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2022 Czarnewski P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Czarnewski P, Mahfouz A, Calogero RA *et al.* **Community-driven ELIXIR activities in single-cell omics [version 1; peer review: awaiting peer review]** F1000Research 2022, 11(ELIXIR):869 <https://doi.org/10.12688/f1000research.122312.1>

First published: 29 Jul 2022, 11(ELIXIR):869 <https://doi.org/10.12688/f1000research.122312.1>

Introduction

Single-cell omics (SCO) is an umbrella term that encompasses multiple technologies that are able to profile various omic modalities at the single-cell level. These high-throughput single-cell approaches have rapidly become the method of choice over traditional bulk methods which average data across a population of cells. Single-cell approaches are better suited for characterizing many biological phenomena and exploring cellular heterogeneity such as characterisation of rare cell types and diverse cell states. Besides single-cell/nucleus RNA sequencing,¹⁻⁴ SCO approaches include nuclear epigenetic profiling such as chromatin accessibility,⁵ histone profiling,⁶ DNA methylation,⁷ chromatin conformation⁸ as well as high throughput single-cell proteomics.⁹ Recent developments allow simultaneous profiling of two or more of the aforementioned modalities,¹⁰ opening up unprecedented opportunities to study diverse processes such as development, gene expression dynamics, tissue heterogeneity and disease pathogenesis. More recently, several approaches have been developed to deliver spatial resolution of single cell expression within tissues, adding another layer of complexity. While several grand challenges in exploratory data analysis remain,¹¹ a parallel issue is the provision of infrastructure to support such analysis in the rapidly developing field.

New SCO profiling technologies are mushrooming, and new analysis methods are published weekly (Figure 1a). As the scale and modality of data sets grow, new computational methods are required. The data also need to be stored and annotated in a standardized manner in order to enable their reuse. This in turn adds an extra challenge and makes it hard for most institutes to handle alone, and calls for international collaboration on training, tools, compute, data, interoperability and standardization in SCO.

ELIXIR, the European infrastructure for life sciences data, brings Europe’s national centers and core bioinformatics resources into a single, coordinated infrastructure.¹² This intergovernmental organization currently has 23 Nodes, and facilitates collaboration between its member institutes and researchers with two intersecting organizational groupings: Platforms and Communities. ELIXIR Platforms (Data, Interoperability, Tools, Compute and Training) provide services,

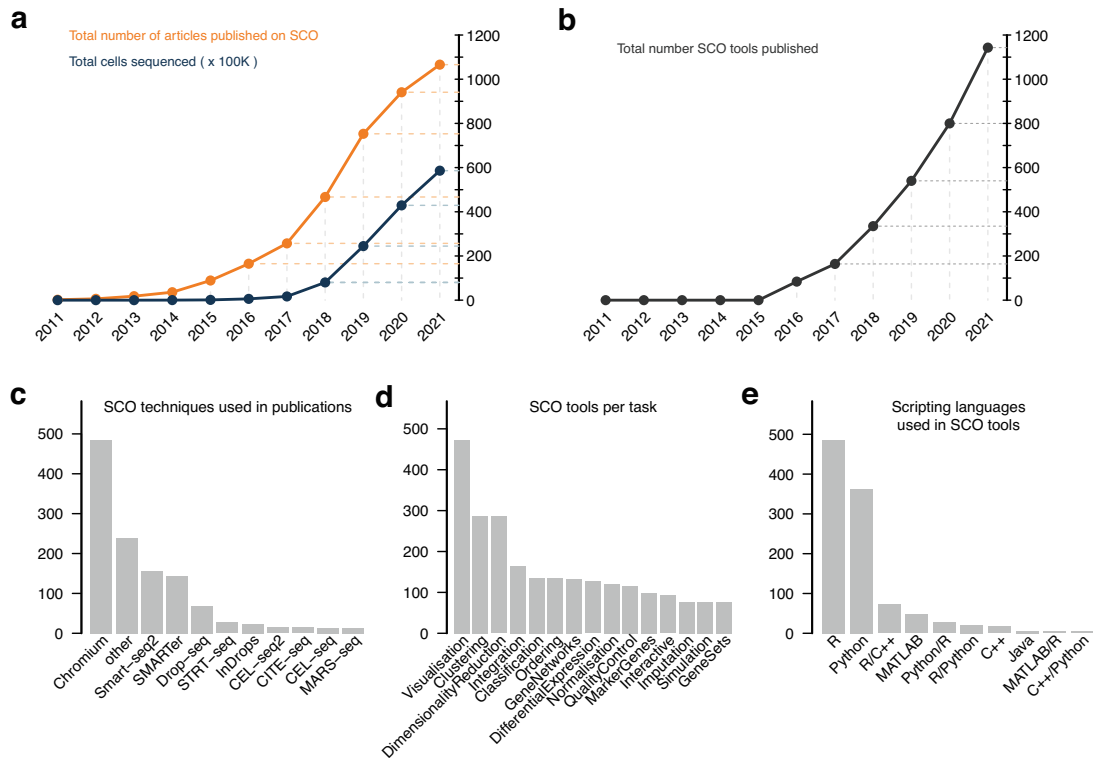


Figure 1. The current landscape in SCO surveyed up until January 2022. (a) Current count of articles using SCO technologies and cumulative number of cells sequenced and deposited in public databases. (b) Number of tools developed specifically to work with SCO. (c) Most common SCO molecular profiling technologies mentioned in publications. (d) Top 15 most targeted categories for software development in SCO. (e) Number of tools developed for SCO, split by which scripting languages are used. Data were taken from public databases.^{34,74} See the Data and Software Availability section for details.

Table 1. Overview of past single-cell course collaborations between ELIXIR Nodes. Tutorial training style refers to teaching bioinformatics by following pre-made sequential analysis steps with code. PBL training style refers to teaching using project-based learning, where students develop their own analysis code to solve analysis tasks. All courses listed below were taught in English.

Course name, link and year	ELIXIR node	Training style	Computing Language	Recorded lectures	FAIR Tool	Length	Technologies
Advanced topics in single cell omics (2021)	SE, CH	PBL	R and python	Yes	Docker	5 days	scRNAseq, scATACseq, ST, Deep Learning
Single Cell School (2019)	SE, CH	Tutorial	R	No	-		scRNAseq, CyTOF, scProteomics
Galaxy single-cell omics training materials (ongoing)	DE, UK, CH	Tutorial	-	Yes	Galaxy	3-5 days	scRNAseq
Single-cell RNAseq analysis using R (2021)	UK EMBL-EBI	Tutorial	R	No	-	5 days	scRNAseq
Gene Expression at Spatial Resolution (2021)	EMBL-EBI SE	Tutorial	-	Yes	-	4 Days	Spatial transcriptomics (10X Visium)
Single-cell RNAseq data analysis with R (2019)	FI, SE, NL, DE, NO, FR	Tutorial	R	Yes	Conda	3 days	scRNAseq
Single-cell RNAseq data analysis with Chipster (2020)	FI, LU	Tutorial	-	Yes	Chipster	3 days	scRNAseq
Single-cell RNAseq data analysis with Chipster (ongoing)	FI, NL, SE, FR	Tutorial, eLearning	-	Yes	Chipster	3 days	scRNAseq
Single cell RNA-seq analysis workshop (2022)	SE, NL	Tutorial	R and Python	Yes	Conda	5 days	scRNAseq, ST

and ELIXIR Communities identify the needs of domain- or technology-specific research around a theme. There are currently 13 Communities ranging from Metabolomics and Proteomics to Federated Human Data and Galaxy.

Many ELIXIR Nodes already have single-cell facilities, and others are setting them up. The Nodes are facing a huge demand for single-cell data analysis and training, and some knowledge transfer between the Nodes already exists. Examples of past data analysis courses co-organized by ELIXIR Nodes are listed in [Table 1](#). The Nodes have also co-organized workshops to discuss FAIR data management and best training practices for SCO. Together, 17 ELIXIR Nodes proposed to create the ELIXIR SCO Community to connect these grass-roots efforts and strengthen European and international cohesion in SCO.

Landscape of SCOs

Technologies

Single-cell omics technologies have seen widespread adoption since its announcement as *Nature* Method of the Year in 2013.¹³ The most widely used SCO technologies are single-cell RNA-seq, single-nucleus RNA-seq, and single-cell ATAC-seq. The early days of these technologies were dominated by heterogeneous implementations of handling, preparation and sequencing protocols,¹⁴ which left its mark in the large number of software tools that had been developed, which in part led to lack of standardization of data and metadata. In recent years, we have seen a number of technology providers prevail for instrumentation (*e.g.* Fluidigm, 10x Genomics), reagents (*e.g.* ThermoFisher, QiaGen, Roche), and sequencing (*e.g.* Illumina, MGI, ONT and PacBio). This has led to the community converging around a few workflows based around the popular 10x Genomics Chromium and SmartSeq chemistries ([Figure 1c](#)), which have been exploited for large scale sequencing efforts such as the Human Cell Atlas (HCA)¹⁵ and Human BioMolecular Atlas (HuBMAP),¹⁶ that, in turn, have resulted in large investments into solving the sample handling, data integration and data management problems underpinning the vast array of data being generated.

While it has been a decade since the SCO technologies have taken the center stage in unraveling bio-molecular heterogeneity, technology developers are far from stagnant and we are seeing rapid evolution of these technologies. We witness the adoption of single-cell genomics (*e.g.* MissionBio Tapestry) and some progress in the field of single-cell proteomics.¹⁷ Single-cell multimodal omics was announced as *Nature* Method of the Year 2019.¹⁰ These assays provide multiple readouts that can be used to define cells. For example, antibody profiling (*e.g.* CITE-seq¹⁸) allows scientists to contextualize novel cell types and states in the context of well-established cell biology markers; immune repertoire profiling can also link how transcriptional profiles of immune cells differ based on the receptor specificity¹⁹; true multimodal omics such as simultaneous profiling of chromatin accessibility, DNA methylation and transcriptomics (*e.g.* scNMT-seq²⁰) allows us to also decipher the regulatory changes that underpin the transcriptional landscape of cells.

In 2020, spatially resolved transcriptomics (SRT) was announced as *Nature* Method of the Year.²¹ These technologies fall under three broad areas of laser capture microscopy combined with single-cell sequencing (*e.g.* GeoMX DSP, Tomoseq), *in situ* capture arrays (*e.g.* ST, Visium, Slide-seq, HDST), and image-based single-molecule expression quantifications (*e.g.* *in situ* sequencing,⁸ seqFISH+,²² Molecular Cartography,²³ MERFISH²⁴). Despite its youth, there are over 20 SRT profiling technologies,²⁵ which are also expanding into other omics modalities, such as proteomics (*e.g.* CODEX²⁶) and metabolomics.²⁷ However, the community quickly realized the potential of this technology and early efforts were pushed by the Chan Zuckerberg Initiative to harmonize the field by funding efforts such as StarFish,²⁸ a platform to uniformly process raw single molecule SRT data, and the [SpaceTX consortium](#), which aim to benchmark and harmonize data from various SRT platforms and analytics methods. Efforts to make comprehensive cell atlases available to the community now facilitate more single-cell study designs to include perturbation and lineage tracing experiments (*e.g.* *via* CRISPR^{29–33}).

While the standards for scRNA-seq have by-and-large converged, the extension of single-cell technologies to new modalities and experimental setups places even more emphasis on establishing adaptable and extensible standards.

Analysis tools

There is a large number of single-cell analysis methods and tools available that cover a wide range of analysis steps. In January 2022, the scRNA-tools database recorded nearly 1,200 tools divided over more than 30 categories ([Figure 1b and d](#)).³⁴ Computational and analytical challenges in single-cell genomics have been discussed extensively.^{11,35} The analysis steps vary depending on the modality of single-cell data. For the most widely used modality, transcriptomic data, the community has converged on a consensus regarding the analysis steps.³⁶ Yet, even some foundational steps remain active areas for research, such as how to normalize scRNA-seq data^{37–39} or how best to perform differential expression analysis.^{40,41} Also, annotation of cell types varies drastically between studies, with many resorting to *ad hoc* decisions. A complete atlas of all cell types would be required to improve the standardization of cell-type nomenclature and

ontologies (*e.g.* Cell Ontology, UBERON).⁴² Scientists analyzing SCO data need to have sufficient information on the strengths and limitations of the available analysis tools in order to select the most suitable ones for their data and purpose. However, systematic comparison of these tools is challenging, especially given the ever-increasing number of methods and their parameter combinations.

Studies have identified essential guidelines for benchmarking computational methods^{43–49} and reviewed published benchmarking studies of computational tools for omics data, highlighting the advantages and limitations of benchmarking across various domains of the life sciences.⁵⁰ Systematic benchmarking frameworks can enable crowdsourcing and community challenges, which have been a successful means for fostering community creativity and expertise to address open problems.⁵¹ In a concerted effort to address “grand challenges” for the SCO community,¹¹ the Open Problem in Single-Cell Analysis group⁵² is devising competitions to address those challenges, *e.g.* the Multimodal Single-Cell Data Integration competition (NeurIPS 2021).

Despite community efforts, the major challenges facing SCO benchmarking studies are the lack of appropriate experimental data and/or realistic simulated data that can be used for benchmarking, as well as the lack of agreed-upon measures to evaluate different methods. There is also a need for a common platform to conduct benchmark studies. The Open Problems NeurIPS challenge provides a leading example for evaluating methods using common datasets, performance metrics, as well as providing a compute infrastructure to run these methods. However, there is still a need for platforms that allow for continuous update of results as new tools and/or metrics become available and to dynamically respond to the needs of individual communities within the life sciences.

Currently available tools and pipelines differ in their usability. While the majority require programming knowledge, several pipelines provide GUIs for users without programming experience (*e.g.* Galaxy, Chipster).^{53,54} Most tools are available as R and Python packages or as a collection of scripts on GitHub (Figure 1e). To keep up with the technology developments, these methods and tools are continuously updated. Yet, maintaining tools and providing support is often challenging for research groups. Interoperability between methods and tools is limited despite efforts by popular packages such as Seurat⁵⁵ to provide wrappers around other tools. However, frequent updates to tools to keep up with technology developments (*e.g.* updating single-cell objects to cater for multi-modal data) limits interoperability, emphasizing the importance of a concerted effort to address robust data and metadata standards.

Standards and research data management

The major factor in realizing interoperability is the definition and adoption of robust data format standards. While more than 1,000 SCO tools exist, there is broad acceptance of widely adopted raw data standards (*e.g.* FASTQ, FAST5, BAM, CRAM) and convergence to a few processed data formats (*e.g.* tab-separated files, AnnData, HDF5, loom, SingleCellExperiment, Seurat). The data formats and structures employed by some of the most popular tools for SCO data analysis^{55,56} have had to change to adapt to new technologies that rendered previous formats inadequate. While these changes in data formats are frustrating for maintaining data analysis workflows, they are necessary for keeping up to date with the rapid technological developments in this field. This places a strong emphasis on planning to adapt to changes by employing extensible structures that do not break the chain of backwards compatibility.

Furthermore, metadata standards and minimal reporting guidelines enable the appropriate archiving and subsequent reuse of SCO data. For some specific library construction or sequencing technologies, provision of platform specific metadata is routine and standardized (*e.g.* for the 10x Chromium), however additional care is required for in-house solutions and for reporting metadata for other parts of the experimental design. Establishing the Minimum Information about a SEQuencing Experiment (MINSEQE) guideline was an important achievement for reporting metadata for sequencing data.⁵⁷ Recently the Minimum Information about a Single-Cell Experiment (minSCe) guidelines were established,⁵⁸ which defines 48 attributes that describe the biosource, isolation method, protocols, library construction, sequencing assay, raw data files and sequences, and cell- and sample-associated information derived from data analysis. However, as we move towards atlasing entire organisms and the rapid emergence of spatially resolved SCO technologies, further refinement of these minimal reporting standards is required to allow for describing common landmarks to facilitate integration of reference maps at differing scales into a single common framework, *e.g.* through the adoption of the Common Coordinate Framework⁵⁹ that aims to uniquely and reproducibly define any location in the human body. A major consequence of establishing robust metadata standards is that they facilitate the establishment of SCO portals that provide access to uniformly processed data from a wide variety of SCO studies (*e.g.* the [EBI Single Cell Expression Atlas](#), the [Broad Single Cell Portal](#), and the [HCA Data Portal](#)). Such portals rely on accurate and sufficient metadata to enable appropriate processing of SCO data from a wide variety of studies in a uniform way.

Given the fast pace of technological developments in the SCO field, the community has identified that both adaptability and extensibility are key considerations in defining sustainable standards. This has been achieved in the field of medical

imaging with the Digital Imaging and Communications in Medicine (DICOM) format,⁶⁰ which has been constantly extended and updated without breaking backwards compatibility for nearly 30 years. However, this level of flexibility was only achieved by the third version of the DICOM standard, 10 years after its initial inception, and it was a concerted effort between medical and trade associations. Part of the successful adoption of the DICOM format is that despite all major medical imaging players having their own proprietary formats, they provide an interface to the DICOM format. In order for the SCO community to reach a similar level of interoperability as has been achieved in medical imaging, technology providers and tool developers should also either adopt the most common standards in the SCO community, or provide interfaces to them. While it is not clear how the current landscape of SCO data and metadata standards will stand the test of time, some aspects that will determine their success will be their ability to adapt to change (*e.g.* through using extensible formats such as JSON), use of controlled nomenclature (*e.g.* utilising ontologies for defining attributes), and adopting versioning (*e.g.* semantic versioning).

Training

Upskilling life scientists to analyze SCO data is a moving target, given the fast development of the field. The cutting edge analysis methods for SCO data tend to be rather computationally complex, making them harder to grasp for life scientists who typically lack a solid background in mathematics, statistics and machine learning and often R/Python skills too.

Trainers, on the other hand, find themselves updating training materials constantly and, in general, struggle to keep up with the fast development of new analysis methods in order to choose what to teach. To make things worse, often only a small fraction of their working time is dedicated to training, or training is offered on a voluntary basis on top of their workload. It is therefore not surprising that even though single-cell courses are offered by several ELIXIR Nodes, many flavors of SCO are not yet covered. For example, courses on single-cell epigenetic, multi-omics as well as image-based spatially resolved SCOs are still rare. The demand for training continues to grow, but the lack of competent experts with enough training experience and time available is a major bottleneck in scaling up training provision. While pedagogical train-the-trainer (TtT) courses^{61,62} can empower experts to feel more comfortable to teach, the constant evolution of the SCO field can intimidate newcomers.

There are also more practical hurdles: the analysis of single-cell data requires a sophisticated computational environment with many tools and their dependencies, often requiring high-end computational resources. These environments have to be ready-to-go or at least easy to set up, and reproducible across heterogeneous hardware infrastructure, allowing the participants to re-run the practical and to analyze their (probably much larger) own data in their own setting. It is also challenging to find good training datasets that are small enough to be run in a class but meaningful enough to prove the concepts.

Alignment with ELIXIR Platforms and Communities

The ELIXIR SCO Community will bring together current efforts and produce guidelines and training. It creates a communication channel to exchange experiences, collect user requests and feedback and push for standards. Given its needs for training, tools, compute, data and interoperability, the SCO Community aligns well with all the ELIXIR Platforms. It also has synergies with the ELIXIR Human Data Communities and the Galaxy Community, as well as some ELIXIR Focus Groups like Cancer Data and FAIR Training.

Training platform

Upskilling scientists in SCO data analysis and standards lies at the heart of the ELIXIR SCO Community, and particular efforts will be made to make the training scalable and FAIR in coordination with the ELIXIR Training platform (Table 2).

The SCO Community will ensure that training materials and expertise are shared efficiently and following FAIR and open research principles.^{63,64} We will collaborate with ELIXIR's Training Portal TeSS⁶⁵ to establish a well-curated SCO training portal, listing national and international training providers, web resources and upcoming training events. To help the current trainers and encourage new ones, we will annotate training materials with appropriate metadata, curate training datasets, provide detailed explanation on how to run courses, and share best practices and best ways to teach the more advanced concepts. In order to identify SCO areas which lack sufficient training, we will participate in designing the annual training gap survey by the Training Platform, and also perform more detailed SCO training surveys if needed. We will regularly host trainer workshops targeting the areas identified as lacking sufficient training to exchange experiences and discuss materials.

Anyone should be able to learn about SCO data analysis independently of time and place. To make the training scalable, lectures and video tutorials will be recorded for asynchronous learning, and combined into modular eLearning courses. Resources will be gathered and annotated on a single site for easy discovery. In addition to organizing training in SCO

Table 2. Goals of the ELIXIR single-cell omics community.

Timeframe	Goals
Short-term goals (~2 years)	<p>Training:</p> <ul style="list-style-type: none"> • Provide training in data analysis and standards to complement ELIXIR Nodes activities. • Create an ELIXIR SCO website, with a dedicated training section for easy discovery. • Create a catalog of SCO video lectures and tutorials for self-study and asynchronous learning with links to training resources which enable anyone to learn SCO data analysis independently of time and place. • Organize workshops for SCO data analysis trainers for exchanging ideas about best practices, methods and datasets. • Collaborate with the ELIXIR Train-the-Trainer programme to provide pedagogical techniques for trainers. <p>Tools:</p> <ul style="list-style-type: none"> • Perform periodic reviews of methods for registration in bio.tools. • Provide a public Slack channel to exchange information about software benchmarks and datasets for benchmarking. • Explore OpenEBench for benchmarking SCO data analysis methods. • Collect and curate datasets for benchmarking. • Provide ready-made containers, Conda recipes and Notebooks with popular SCO software environments. <p>Compute:</p> <ul style="list-style-type: none"> • Keep the Compute Platform up to date with the computing needs of SCO data analysis. <p>Interoperability:</p> <ul style="list-style-type: none"> • Define requirements of a framework for an efficient, effective and flexible single cell omics FAIR data and metadata standards. • Disseminate knowledge of and promote standards in preparation for creation of ELIXIR core data resources.
Long-term goals (~5 years)	<p>Training:</p> <ul style="list-style-type: none"> • Collaborate with TeSS, ELIXIR's Training Portal, to establish a well-curated ELIXIR SCO training portal, listing national and international bodies, web resources and upcoming training events. • Keep training resources up to date. <p>Tools:</p> <ul style="list-style-type: none"> • Benchmarking and reproducibility: develop several software benchmarks within the OpenEBench infrastructure. • Develop cloud-deployable analysis pipelines for SCO data and make them available also through Galaxy and Chipster for non-programming scientists. • Provide long-term cloud-based solutions for making tools open and FAIR. <p>Compute:</p> <ul style="list-style-type: none"> • Benchmark, update and optimize tools to run as efficiently as possible across different ELIXIR computing nodes. <p>Interoperability:</p> <ul style="list-style-type: none"> • Support existing efforts for aggregating and disseminating related metadata standards, e.g. from ArrayExpress, the HCA and further efforts. • Establish a user forum to restructure and unify data structure of sequencing, spatial and image data across SCO. • Encourage all data generators to ensure their data is available from an ELIXIR core data resource. • Leverage EMBL-EBI connections to the HCA Data Coordination Platform to broker the HCA data to ELIXIR core data resources and deposition resources.

data analysis and standards to complement ELIXIR Nodes activities, we will provide training in best practices for trainers (TrT) to increase the number of expert trainers.

The course software installation challenge will be addressed together with the ELIXIR Tools Platform, as described below, using Conda environments, containers and Notebooks. Both Galaxy⁶⁶ and Chipster⁵⁴ offer specific training access and a comprehensive collection of training materials. There is no setup required, and the same environment is available when analyzing one's own data after the course.

Tools platform

SCO data analysis typically requires a large number of tools and their dependencies. The installation challenge can be eased by providing Conda environments⁶⁷ and containers⁶⁸ for SCO, in alignment with the work developed in the Tools

Platform's Packaging, containerisation and deployment activity. Also, RStudio or Jupyter Lab based SCO Notebooks can be made to support courses and self-study. The Community will develop cloud-deployable analysis pipelines for SCO data and make them available also through the web-based [Galaxy Single Cell Omics](#), [Galaxy Human Cell Atlas project](#), and [Chipster](#) analysis platforms for researchers lacking programming skills. The analysis pipelines will be deposited in [WorkflowHub](#)⁶⁹ for easy discovery, re-use and assessment.

The SCO Community will take several actions to address the aforementioned challenges in benchmarking. Liaising with data analysis experts, we will carefully curate data collections suitable for addressing specific tasks within the SCO data analysis workflow (*e.g.* multi-modal data integration, deconvolution of bulk data). For this, we will survey the landscape of existing benchmarking studies and identify the datasets they used and how they were evaluated. Whenever possible, our focus will be on real datasets rather than simulated ones, given the bias introduced by simulated data towards methods using the same underlying model. In order to address the lack of agreed-upon performance metrics to evaluate different types of methods, we will collect and curate existing metrics, and develop/suggest new measures when necessary ([Table 2](#)).

Regarding the need for a common platform to conduct benchmark studies, we will explore using [OpenEBench](#).⁷⁰ This ELIXIR benchmarking platform offers a flexible computational framework that allows individual communities to design and perform their benchmarking experiments. Communities are responsible for defining the reference datasets and the evaluation metrics and designing and developing evaluation workflows. Software developers are then able to use these workflows to evaluate their tools against the reference datasets, and the computed metrics are compiled, analyzed and publicly exposed in tables and visualizations. The results of the evaluation are then used by the community or any other OpenEBench user to decide which is the most suitable tool to do their analysis. The SCO Community will provide guidelines for the setup of single-cell benchmarking experiments. The guidelines have to cover three topics: 1) the scope of the benchmark, 2) the evaluation metrics that will be used to measure the performance of the tools and 3) the reference or gold standard datasets. The SCO Community will establish a benchmarking environment for SCO data analysis tools within the OpenEBench infrastructure, to facilitate a variety of community-driven challenges to address the diversity of the SCO applications.

The SCO community will perform periodic reviews of highly performant and rapidly adopted methods for registration in the [bio.tools catalogue](#). To this end we will work closely also with the [EDAM ontology](#) to define single-cell specific keywords, which will help us not only to annotate the tools but also tag courses in TeSS.

Compute platform

The computing resource requirements of SCO data analysis increase constantly as the scale and modality of the data sets grow. The discussion between the SCO Community and the ELIXIR Compute Platform is therefore vital to ensure sufficient resources. The Community will also benefit from the Compute Platform's Container Orchestration task, which will allow execution of containerised software tools and workflow workloads supporting public and sensitive data across ELIXIR Nodes. The ELIXIR Authentication and Authorisation Infrastructure (AAI) will be supported in the context of sensitive SCO data and whenever controlled access will be needed, we count on learning from the HCA's experience on this matter.

Interoperability platform

The ELIXIR SCO Community will promote the development and usage of standards of metadata and file formats to ensure reproducibility of analyses and data reuse across biological and bioinformatics research communities. We will support existing efforts for aggregating and disseminating related metadata standards, *e.g.* from ArrayExpress, the HCA and further efforts. This is particularly important for emerging spatially resolved data, for which we aim to investigate efficient and scalable reporting structures, in line with current efforts in imaging and omics databases ([Table 2](#)).

Data platform

An important consideration for sensitive human data is the General Data Protection Regulation (GDPR). To comply with the GDPR, raw human sequencing data deposited in EGA is protected and requires approval of the Data Access Committee Officer (DACO) as well as Data Transfer Agreements (DTA) outlining the conditions for allowing access to sensitive data. However, there is heterogeneous interpretation of the GDPR across Europe, and to facilitate this there have been a number of nationally Federated EGAs being established. Other non-human raw sequencing data would be deposited in ENA and would not be subject to these restrictions.

The ELIXIR SCO Community will encourage all data generators to ensure their data is available from an ELIXIR core data resource, or is deposited with a suitable ELIXIR core deposition resource, wherever possible, to ensure maximum

data reuse and long term sustainability of all SCO data across the broader community. *Via* connections to key ELIXIR resources at EMBL-EBI (ENA, EGA, ArrayExpress and BioSamples database), we will promote discussions with these data resources to encourage the adoption and development of standards, where needed, to support the rapid pace of technology change in the single-cell field. We will leverage EMBL-EBI connections to the HCA Data Coordination Platform to broker the HCA data to ELIXIR core data resources and deposition resources.

Alignment with other European and global SCO initiatives

The SCO Community will bring together data standardization efforts across Europe and combine them with global collaborations. The EMBL-EBI Node is a member of the global HCA community, whose mission is to create comprehensive reference maps of all human cells as a basis for both understanding human health and diagnosing, monitoring, and treating disease.¹⁵ It is also involved in the NIH-supported HuBMAP consortium,¹⁶ which develops tools to create an open, global atlas of the human body at the cellular level. The EMBL-EBI has already led an international effort to define the first guidelines for metadata standards of scRNA-seq experiments,⁵⁸ involving members of the HCA and HuBMAP data platforms. As SCO techniques develop, we expect these guidelines to evolve to enable reproducible analysis of other methods, such as scATAC-seq, CITE-seq, single-cell HiC, to name a few.

Importantly, the ELIXIR SCO Community will align its activities with the LifeTime FET initiative,⁷¹ which combines single-cell multi-omics technologies with artificial intelligence and machine learning in order to revolutionize healthcare by tracking, understanding, and treating human cells during diseases. The LifeTime consortium includes over 90 research institutes and 70 supporting companies across Europe. Scientists from some of ELIXIR Nodes belong to both the LifeTime initiative and the SCO Community, thereby providing a direct link between them.

While HCA, HuBMAP and LifeTime focus on human cells, it is important to note that SCO technologies are used for different organisms, and thereby the ELIXIR SCO Community is not limited to human research. For example, the EMBL-EBI is also involved in the Fly Cell Atlas consortium.⁷²

The training activities of the SCO Community will be enriched by collaboration with the Global Organization for Bioinformatics Learning, Education and Training (GOBLET).⁷³ GOBLET's mission is to cultivate the global bioinformatics trainer community, set standards and provide high-quality resources to support learning, education and training. The emerging SCO Community and GOBLET co-organized a [global workshop](#) for single-cell RNA-seq data analysis trainers in 2021. Sharing information about different training approaches, materials and datasets was considered very useful by the participants, and follow-up workshops are planned.

Finally, the SCO Community is discussing with the emerging [SCO Community of Australian BioCommons](#), which is currently collecting user needs and finding solutions to the challenges identified, similar to us.

Conclusions

The SCO paradigm represents a revolution in the life sciences that pushes the boundaries of what can be explored, creating both new opportunities and challenges. We are witnessing increasing numbers of individual- and multi-omics modalities, and spatio-temporally resolved read outs. Both the rapid pace of advancement and adoption indicate that SCO will become the new normal in the life sciences. In the past five years, many ELIXIR Nodes have been working to assemble resources with the goal of developing future-proof guidelines and infrastructure as well as delivering training to SCO scientists. Here, we defined key goals at different infrastructural areas in order to create the ELIXIR SCO Community ([Table 2](#)) to ultimately strengthen current and foster new collaborations, and establish sustainable European and global frameworks for SCO research.

Data and Software Availability

Data on scientific publications on SCO and number of cells sequenced was obtained from the Single-cell studies database.⁷⁴ Data on SCO tools was taken from the [publicly available repository](#) of the scRNA-tools database.³⁴

Author contributions

PC, PMP, IP, TB, JL, KH and EK conceptualized the study. PC performed data curation and visualization. KH performed project administration. PC, AM, RAC, PMP, LPS, AGU, LA, SCG, BS, MA, NI and EK wrote the original draft of the manuscript. PC, AM, PMP, LPS, AGU, CS, BS, PV, H-RH, BL, JMH, BG, WH, RK, TR, AB, PF, KH, NI, and EK reviewed and edited the manuscript.

Acknowledgments

We would like to thank several members of ELIXIR Nodes for useful discussions and input: Ana Melo (PT), Andrei Zinovyev (FR), Åsa Björklund (SE), Celia van Gelder (NL), Ernesto Picardi (IT), Philip Lijnzaad (NL), Jan Korbel (DE), Joaquin Dopazo (ES) Loredana Le Pera (IT), Priit Adler (EE), Ricardo Leite (PT), Silvie Fexova (EMBL-EBI), Ståle Nygård (NO), Victoria Dominguez del Angel (FR).

References

- Lacar B, Linker SB, Jaeger BN, *et al.*: **Nuclear RNA-seq of single neurons reveals molecular signatures of activation.** *Nat. Commun.* 2016 Apr; **7**: 11022.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tang F, Barbacioru C, Wang Y, *et al.*: **mRNA-Seq whole-transcriptome analysis of a single cell.** *Nat. Methods.* 2009 May; **6**(5): 377–382.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Picelli S, Björklund ÅK, Faridani OR, *et al.*: **Smart-seq2 for sensitive full-length transcriptome profiling in single cells.** *Nat. Methods.* 2013 Nov; **10**(11): 1096–1098.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zheng GXY, Terry JM, Belgrader P, *et al.*: **Massively parallel digital transcriptional profiling of single cells.** *Nat. Commun.* 2017 Jan 16; **8**(1): 14049.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Buenrostro JD, Wu B, Litzenburger UM, *et al.*: **Single-cell chromatin accessibility reveals principles of regulatory variation.** *Nature.* 2015 Jul; **523**(7561): 486–490.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Angermueller C, Clark SJ, Lee HJ, *et al.*: **Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity.** *Nat. Methods.* 2016 Mar; **13**(3): 229–232.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Luo C, Keown CL, Kurihara L, *et al.*: **Single Cell Methylomes Identify Neuronal Subtypes and Regulatory Elements in Mammalian Cortex.** *Science.* 2017 Aug; **357**(6351): 600–604.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nagano T, Lubling Y, Stevens TJ, *et al.*: **Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.** *Nature.* 2013 Oct; **502**(7469): 59–64.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bandura DR, Baranov VI, Ornatsky OI, *et al.*: **Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry.** *Anal. Chem.* 2009 Aug; **81**(16): 6813–6822.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zhu C, Preissl S, Ren B: **Single-cell multimodal omics: the power of many.** *Nat. Methods.* 2020 Jan; **17**(1): 11–14.
[Publisher Full Text](#)
- Lähnemann D, Köster J, Szczurek E, *et al.*: **Eleven grand challenges in single-cell data science.** *Genome Biol.* 2020 Feb; **21**(1): 31.
[Publisher Full Text](#)
- Harrow J, Hancock JELIXIR-EXCELERATE Community, *et al.*: **ELIXIR-EXCELERATE: establishing Europe's data infrastructure for the life science research of the future.** *EMBO J.* 2021 Mar 15 [cited 2022 Feb 2]; **40**(6).
[Publisher Full Text](#)
- Method of the Year 2013.** *Nat. Methods.* 2014 Jan; **11**(1): 1–1.
[Publisher Full Text](#)
- Svensson V, Vento-Tormo R, Teichmann SA: **Exponential scaling of single-cell RNA-seq in the past decade.** *Nat. Protoc.* 2018 Apr; **13**(4): 599–604.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Regev A, Teichmann SA, Lander ES, *et al.*: **The Human Cell Atlas.** *Elife.* 2017 Dec; **6**: e27041.
[PubMed Abstract](#) | [Publisher Full Text](#)
- HuBMAP Consortium: **The human body at cellular resolution: the NIH Human Biomolecular Atlas Program.** *Nature.* 2019 Oct; **574**(7777): 187–192.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vistain LF, Tay S: **Single-Cell Proteomics.** *Trends Biochem. Sci.* 2021 Aug 1; **46**(8): 661–672.
[Publisher Full Text](#)
- Stoeckius M, Hafemeister C, Stephenson W, *et al.*: **Simultaneous epitope and transcriptome measurement in single cells.** *Nat. Methods.* 2017 Sep; **14**(9): 865–868.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Redmond D, Poran A, Elemento O: **Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq.** *Genome Med.* 2016 Dec; **8**(1): 80.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Clark SJ, Argelaguet R, Kapourani C-A, *et al.*: **scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells.** *Nat. Commun.* 2018 Dec; **9**(1): 781.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Method of the Year 2020: spatially resolved transcriptomics.** *Nat. Methods.* 2021 Jan; **18**(1): 1–1.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Eng C-HL, Lawson M, Zhu Q, *et al.*: **Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+.** *Nature.* 2019 Apr; **568**(7751): 235–239.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lee JH, Daugharthy ER, Scheiman J, *et al.*: **Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues.** *Nat. Protoc.* 2015 Mar; **10**(3): 442–458.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Xia C, Fan J, Emanuel G, *et al.*: **Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression.** *Proc. Natl. Acad. Sci. U. S. A.* 2019 Sep; **116**(39): 19490–19499.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moses L, Pachter L: **Museum of Spatial Transcriptomics.** *Bioinformatics.* 2021 May [cited 2022 Jan 25].
[Publisher Full Text](#)
- Goltsev Y, Samusik N, Kennedy-Darling J, *et al.*: **Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging.** *Cell.* 2018 Aug; **174**(4): 968–981.e15.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rappez L, Stadler M, Triana S, *et al.*: **SpaceM reveals metabolic states of single cells.** *Nat. Methods.* 2021 Jul; **18**(7): 799–805.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Perkel JM: **Starfish enterprise: finding RNA patterns in single cells.** *Nature.* 2019 Aug; **572**(7770): 549–551.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smith RH, Chen Y-C, Seifuddin F, *et al.*: **Genome-Wide Analysis of Off-Target CRISPR/Cas9 Activity in Single-Cell-Derived Human Hematopoietic Stem and Progenitor Cell Clones.** *Genes.* 2020 Dec; **11**(12): 1501.
[PubMed Abstract](#) | [Publisher Full Text](#)
- ten Hacken E, Clement K, Li S, *et al.*: **High throughput single-cell detection of multiplex CRISPR-edited gene modifications.** *Genome Biol.* 2020 Dec; **21**(1): 266.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Yang L, Chan AKN, Miyashita K, *et al.*: **High-resolution characterization of gene function using single-cell CRISPR tiling screen.** *Nat. Commun.* 2021 Dec; **12**(1): 4063.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zafar H, Lin C, Bar-Joseph Z: **Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data.** *Nat. Commun.* 2020 Dec; **11**(1): 3055.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dixit A, Parnas O, Li B, *et al.*: **Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens.** *Cell.* 2016 Dec 15; **167**(7): 1853–1866.e17.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zappia L, Phipson B, Oshlack A: **Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database.** *Schneidman D, editor. PLoS Comput. Biol.* 2018 Jun; **14**(6): e1006245.
[PubMed Abstract](#) | [Publisher Full Text](#)

35. Stegle O, Teichmann SA, Marioni JC: **Computational and analytical challenges in single-cell transcriptomics.** *Nat. Rev. Genet.* 2015 Mar; **16**(3): 133–145.
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Luecken MD, Theis FJ: **Current best practices in single-cell RNA-seq analysis: a tutorial.** *Mol. Syst. Biol.* 2019 Jun [cited 2022 Feb 3]; **15**(6): e8746.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Lytal N, Ran D, An L: **Normalization Methods on Single-Cell RNA-seq Data: An Empirical Survey.** *Front. Genet.* 2020 Feb; **11**: 41.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Hafemeister C, Satija R: **Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.** *Genome Biol.* 2019 Dec 23; **20**(1): 296.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Lause J, Berens P, Kobak D: **Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data.** *Genome Biol.* 2021 Sep 6; **22**(1): 258.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Sonesson C, Robinson MD: **Bias, robustness and scalability in single-cell differential expression analysis.** *Nat. Methods.* 2018 Apr; **15**(4): 255–261.
[Publisher Full Text](#)
41. Squair JW, Gautier M, Kathe C, et al.: **Confronting false discoveries in single-cell differential expression.** *Nat. Commun.* 2021 Sep 28; **12**(1): 5692.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Osumi-Sutherland D, Xu C, Keays M, et al.: **Cell type ontologies of the Human Cell Atlas.** *Nat. Cell Biol.* 2021 Nov; **23**(11): 1129–1135.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Tran HTN, Ang KS, Chevrier M, et al.: **A benchmark of batch-effect correction methods for single-cell RNA sequencing data.** *Genome Biol.* 2020 Dec; **21**(1): 12.
[PubMed Abstract](#) | [Publisher Full Text](#)
44. Saelens W, Cannoodt R, Todorov H, et al.: **A comparison of single-cell trajectory inference methods.** *Nat. Biotechnol.* 2019 May; **37**(5): 547–554.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Tian L, Dong X, Freytag S, et al.: **Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments.** *Nat. Methods.* 2019 Jun; **16**(6): 479–487.
[Publisher Full Text](#)
46. Abdelaal T, Michielsen L, Cats D, et al.: **A comparison of automatic cell identification methods for single-cell RNA sequencing data.** *Genome Biol.* 2019 Dec; **20**(1): 194.
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Hou W, Ji Z, Ji H, et al.: **A systematic evaluation of single-cell RNA-sequencing imputation methods.** *Genome Biol.* 2020 Dec; **21**(1): 218.
[PubMed Abstract](#) | [Publisher Full Text](#)
48. Luecken MD, Büttner M, Chaichoompu K, et al.: **Benchmarking atlas-level data integration in single-cell genomics.** *Nat. Methods.* 2022 Jan; **19**(1): 41–50.
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Weber LM, Saelens W, Cannoodt R, et al.: **Essential guidelines for computational method benchmarking.** *Genome Biol.* 2019 Jun; **20**(1): 125.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Mangul S, Martin LS, Hill BL, et al.: **Systematic benchmarking of omics computational tools.** *Nat. Commun.* 2019 Mar; **10**(1): 1393.
[PubMed Abstract](#) | [Publisher Full Text](#)
51. Meyer P, Saez-Rodriguez J: **Advances in systems biology modeling: 10 years of crowdsourcing DREAM challenges.** *Cell Syst.* 2021 Jun 16; **12**(6): 636–653.
[Publisher Full Text](#)
52. CZI: **Open Problems in Single Cell Analysis.** *Open Problems in Single Cell Analysis.* [cited 2021 Dec 2].
[Reference Source](#)
53. Afgan E, Baker D, Batut B, et al.: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.** *Nucleic Acids Res.* 2018 Jul; **46**(W1): W537–W544.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Kallio MA, Tuimala JT, Hupponen T, et al.: **Chipster: user-friendly analysis software for microarray and other high-throughput data.** *BMC Genomics.* 2011 Dec; **12**(1): 507.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Hao Y, Hao S, Andersen-Nissen E, et al.: **Integrated analysis of multimodal single-cell data.** *Cell.* 2021 Jun; **184**(13): 3573–3587.e29.
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Wolf FA, Angerer P, Theis FJ: **SCANPY: large-scale single-cell gene expression data analysis.** *Genome Biol.* 2018 Feb 6; **19**(1): 15.
[PubMed Abstract](#) | [Publisher Full Text](#)
57. Brazma A, Ball C, Bumgarner R, et al.: **INSEQE: Minimum Information about a high-throughput Nucleotide Sequencing Experiment - a proposal for standards in functional genomic data reporting.** 2012 Jun 1 [cited 2022 Feb 6].
[Reference Source](#)
58. Füllgrabe A, George N, Green M, et al.: **Guidelines for reporting single-cell RNA-seq experiments.** *Nat. Biotechnol.* 2020 Dec; **38**(12): 1384–1386.
[PubMed Abstract](#) | [Publisher Full Text](#)
59. Rood JE, Stuart T, Ghazanfar S, et al.: **Toward a Common Coordinate Framework for the Human Body.** *Cell.* 2019 Dec 12; **179**(7): 1455–1467.
[PubMed Abstract](#) | [Publisher Full Text](#)
60. Best DE, SCH Md, Bennett WC, et al.: **Update of the ACR-NEMA digital imaging and communications in medicine standard.** *Medical Imaging VI: PACS Design and Evaluation.* SPIE; 1992 [cited 2022 Feb 5]. p. 356–61.
[Publisher Full Text](#)
61. Via A, Attwood TK, Fernandes PL, et al.: **A new pan-European Train-the-Trainer programme for bioinformatics: pilot results on feasibility, utility and sustainability of learning.** *Brief. Bioinform.* 2019 Mar; **20**(2): 405–415.
[PubMed Abstract](#) | [Publisher Full Text](#)
62. Morgan SL, Palagi PM, Fernandes PL, et al.: **The ELIXIR-EXCELERATE Train-the-Trainer pilot programme: empower researchers to deliver high-quality training.** *F1000Res.* 2017 Aug; **6**: 1557.
[Publisher Full Text](#)
63. Wilkinson MD, Dumontier M, Aalbersberg J, et al.: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data.* 2016 Mar; **3**(1): 160018.
[PubMed Abstract](#) | [Publisher Full Text](#)
64. Garcia L, Batut B, Burke ML, et al.: **Ten simple rules for making training materials FAIR.** *PLoS Comput. Biol.* 2020 May 21; **16**(5): e1007854.
[PubMed Abstract](#) | [Publisher Full Text](#)
65. Beard N, Bacall F, Nenadic A, et al.: **TeSS: a platform for discovering life-science training opportunities.** *Bioinformatics.* 2020 May 1; **36**(10): 3290–3291.
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Batut B, Hiltmann S, Bagnacani A, et al.: **Community-Driven Data Analysis Training for Biology.** *Cell Syst.* 2018 Jun 27; **6**(6): 752–758.e1.
[Publisher Full Text](#)
67. Grüning B, Dale R, Sjödin A, et al.: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat. Methods.* 2018 Jul; **15**(7): 475–476.
[Publisher Full Text](#)
68. da Veiga LF, Grüning BA, Alves Afifos S, et al.: **BioContainers: an open-source and community-driven framework for software standardization.** *Bioinformatics.* 2017 Aug 15; **33**(16): 2580–2582.
[Publisher Full Text](#)
69. Goble C, Soiland-Reyes S, Bacall F, et al.: **Implementing FAIR Digital Objects in the EOOSC-Life Workflow Collaboratory.** *Zenodo.* 2021 Mar [cited 2022 Feb 9].
[Reference Source](#)
70. Capella-Gutierrez S, de la Iglesia D, Haas J, et al.: **Lessons Learned: Recommendations for Establishing Critical Periodic Scientific Benchmarking.** *bioRxiv.* 2017 [cited 2022 Feb 4]; p. 181677.
[Publisher Full Text](#)
71. Rajewsky N, Almouzni G, Gorski SA, et al.: **LifeTime and improving European healthcare through cell-based interceptive medicine.** *Nature.* 2020 Nov; **587**(7834): 377–386.
[PubMed Abstract](#) | [Publisher Full Text](#)
72. Li H, Janssens J, De Waegeneer M, et al.: **Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly.** *Genomics.* 2021 Jul [cited 2022 Jan 13].
[Publisher Full Text](#)
73. Attwood TK, Atwood TK, Bongcam-Rudloff E, et al.: **GOBLET: the Global Organisation for Bioinformatics Learning, Education and Training.** *PLoS Comput. Biol.* 2015 Apr; **11**(4): e1004143.
[PubMed Abstract](#) | [Publisher Full Text](#)
74. Svensson V, da Veiga BE, Pachter L: **A curated database reveals trends in single-cell transcriptomics.** *Database.* 2020 Jan 1; **2020**: baaa073.
[PubMed Abstract](#) | [Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research