

Challenges and solutions for teaching single-cell omics data analysis

11.10.2022

Eija Korpelainen, CSC, Finland (eija.korpelainen@csc.fi)



CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

Outline

- Challenges
- International collaboration
 - GOBLET-ELIXIR workshop for single-cell RNA-seq data analysis trainers
 - ELIXIR Single-cell Omics Community - joining forces in Europe and beyond
- National training efforts in Finland
 - Flipped classroom approach
 - eLearning

Single-cell omics



- Multiple technologies able to profile cells for different modalities at a single-cell level
 - Spatially resolved transcriptomics provides location information too!
- Great for studying development, tissue heterogeneity, disease pathogenesis, etc

nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [editorials](#) > article

Published: 30 December 2013

Method of the Year 2013

[Nature Methods](#) 11, 1 (2014) | [Cite this article](#)

32k Accesses | 49 Citations | 133 Altmetric | [Metrics](#)

Methods to sequence the DNA and RNA of single cells are poised to transform m

nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [editorials](#) > article

Editorial | [Published: 06 January 2020](#)

Method of the Year 2019: Single-cell multimodal omics

[Nature Methods](#) 17, 1 (2020) | [Cite this article](#)

39k Accesses | 51 Citations | 134 Altmetric | [Metrics](#)

Multimodal omics measurement offers opportunities for gaining holistic views of cells one by one.

nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [editorials](#) > article

Editorial | [Published: 06 January 2021](#)

Method of the Year 2020: spatially resolved transcriptomics

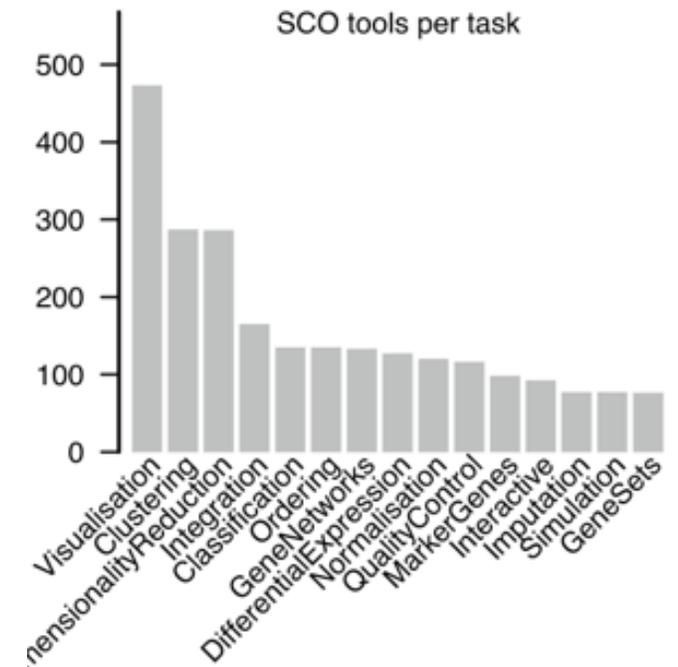
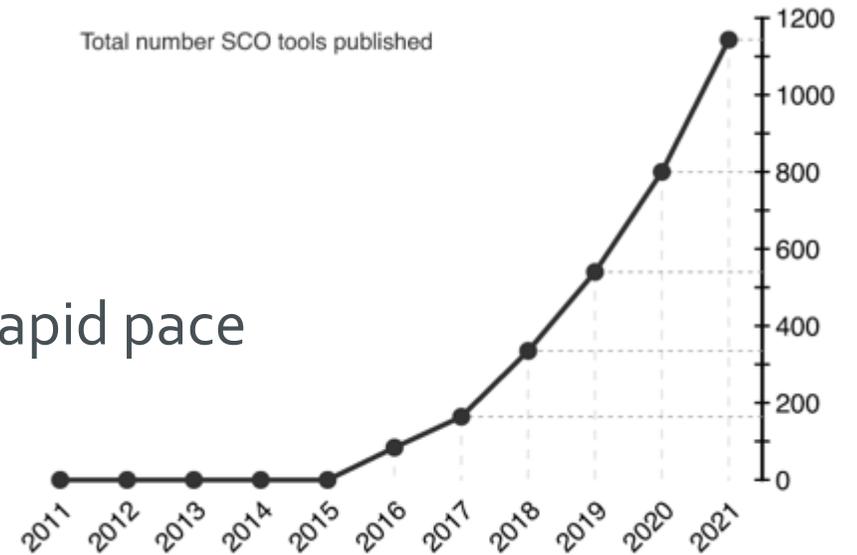
[Nature Methods](#) 18, 1 (2021) | [Cite this article](#)

32k Accesses | 29 Citations | 229 Altmetric | [Metrics](#)

Spatially resolved transcriptomics methods are changing the way we understand complex tissues.

Challenge

- New single-cell omics technologies are developed at a rapid pace
 - They produce different types of data
 - New analysis methods and data standards are needed
- Data analysis methods are mushrooming
 - Multi-step analysis, every step has several methods available
 - How to keep up and choose the best?
- Huge demand for **training**



Teaching single-cell omics data analysis is a headache for trainers

- Fast moving field: what should you teach?
- Complex theory: how do you teach it to biologists?
- Several frameworks and complex objects (Scanpy, Seurat, Bioconductor)
- Environment: everybody should have the same packages and enough memory
- Exercise data: needs to show many things (e.g. batch effect removal), yet fast to run

...and for participants

Pearson residuals from regularized negative binomial regression

Uniform Manifold Approximation and Projection (UMAP)

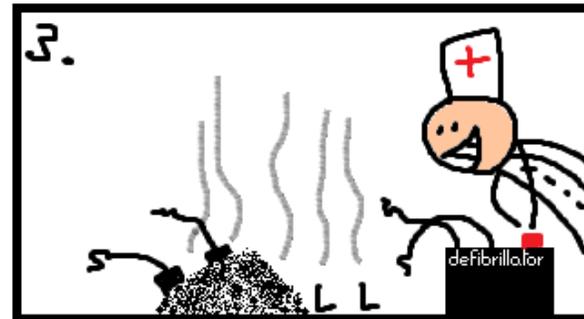
t-distributed stochastic neighbor embedding (t-SNE)

Graph-based clustering using SNN, optimize modularity with Louvain algorithm

Canonical correlation analysis (CCA), L2 normalization of CCVs

Identify mutual nearest neighbours (MMNs), score anchors in PC space

Dynamic time warping



Drawing by Stefan Rua

Workshop for single-cell RNA-seq data analysis trainers

- Co-organized by GOBLET and ELIXIR in 30.11.2021
- Brought together over 30 trainers from four continents
- Presentations and discussions on experiences and best practices
- Slides available at <https://ssl.eventilla.com/event/xP1QM>
- Recorded presentations available as a playlist on GOBLET's YouTube channel
 - https://www.youtube.com/playlist?list=PLN-3-7onnm9WxUY_7yniEv3hWV1fnBnFm

Workshop for scRNA-seq data analysis trainers

🕒 30.11.2021 13:00 +02:00 EET - 30.11.2021 16:00 +02:00 EET 📍 Online

Single cell RNA sequencing (scRNA-seq) offers great potential, and new experimental technologies and data analysis methods have been developed at rapid pace. However, these fast developments put high pressure on life scientists to keep up with the challenging demands of data analysis, and the need for training is enormous.

This **GOBLET-ELIXIR** workshop brought together scRNA-seq trainers to share experiences, challenges and best practices. The **presentation videos** are available in YouTube.

- *Welcome, Eija Korpelainen, CSC*
- **The Swedish efforts in teaching single cell bioinformatics**, *Paulo Czarnewski (ELIXIR-SE, NBIS, Sweden)*
- **Flipped classroom approach and Chipster in scRNA-seq training**, *Eija Korpelainen (ELIXIR-FI, CSC, Finland)*
- **eLearning in scRNA-seq training**, *Maria Lehtivaara (ELIXIR-FI, CSC, Finland)*
- **Galaxy and scRNA-Seq training**, *Pavankumar Videm (ELIXIR-DE, Galaxy, Germany)*
- **scRNAseq hands on (Italy, EMBL, Singapore)**, *Raffaele Calogero (ELIXIR-IT, University of Torino, Italy)*
- **Single-cell RNA seq training in the Netherlands**, *Ahmed Mahfouz (ELIXIR-NL, LUMC, Netherlands)*
- **Organizing advanced single-cell courses**, *Charlotte Soneson (ELIXIR-CH, SIB, Switzerland)*
- **Single-cell RNA-seq training at Cambridge**, *Hugo Tavares (ELIXIR-UK, University of Cambridge, UK)*

🕒 Event time

Starts: 30.11.2021 13:00 +02:00

Ends: 30.11.2021 16:00 +02:00

📍 Event location

Online



Workshop for single cell RNA-seq data analysis trainers 2021

8 videos • 63 views • Last updated on Dec 2, 2021



These presentations were recorded during the GOBLET-ELIXIR workshop for single cell RNA-seq data analysis trainers 30.11.2021. Please see the workshop page <https://ssl.eventilla.com/event/xP1QM> for the program and slides.



GOBLET PR

SUBSCRIBE



1 The Swedish efforts in teaching single cell bioinformatics, Paulo Czarnewski

GOBLET PR



2 Flipped classroom approach and Chipster in scRNA-seq training, Eija Korpelainen

GOBLET PR



3 eLearning in scRNA-seq training, Maria Lehtivaara

GOBLET PR



4 Galaxy and scRNA-Seq training, Pavankumar Videm

GOBLET PR



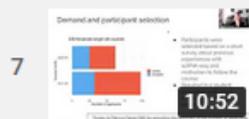
5 scRNAseq hands on (Italy, EMBL, Singapore), Raffaele Calogero

GOBLET PR



6 Single-cell RNA seq training in the Netherlands, Ahmed Mahfouz

GOBLET PR



7 Organizing advanced single-cell courses, Charlotte Soneson

GOBLET PR



8 Single-cell RNA-seq training at Cambridge, Hugo Tavares

GOBLET PR

Ideas from the presentations

- Paulo Czarnewski, NBIS: Problem-based learning
 - Biological research question given to groups of four people, 5-day course
 - Instructions and glossary with essential information
 - Questions (e.g. why do we normalize data)
 - Groups produce a project report and presentation
- Charlotte Sonesson, SIB: Advanced courses, container-based cloud environments
 - Saves course time
 - Access to GPUs too
- Hugo Tavares, University of Cambridge: Online course spread over 3 weeks
 - Google doc for questions
 - Good training data set!
- Maria Lehtivaara, CSC: eLearning

ELIXIR Single-Cell Omics Community

- People from 19 ELIXIR countries
- F1000 paper describing the goals
 - Training
 - Tool benchmarking
 - Data standards
- SCONE project 2023-2024
 - Single-cell Omics Network for ELIXIR
 - Training portal inc. lecture videos, datasets, etc

Home » Browse » Community-driven ELIXIR activities in single-cell omics

Check for updates

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

 Author details



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



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



ALL METRICS

352

VIEWS

28

DOWNLOADS

Get PDF

Get XML

Cite

Export

Track

Email

Share

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.

eLearning course Single-cell RNA-seq data analysis with Chipster



- Focus on data analysis methods
- Lecture videos (beginner and advanced), exercises, quizzes
- <https://e-learn.csc.fi/course/view.php?id=17>

Single cell RNA-seq data analysis with Chipster

Home / Courses / Bioinformatics / scRNAseqChipster

About the course

Course contents

In this course, you will learn how to analyse single-cell RNA-seq data using the Seurat single cell tools integrated in the easy-to-use Chipster software. The exercises and course data are based on the Seurat guided analyses "Guided tutorial - 2700 PBMCs" and "Stimulated vs Control PBMCs".

You will learn the following topics, and how to perform these steps in the Chipster software:

- perform quality control and filter out low quality cells
- normalize gene expression values
- scale data and remove unwanted sources of variation
- select highly variable genes
- perform dimensionality reduction (PCA, tSNE, UMAP, CCA)
- cluster cells
- find marker genes for a cluster
- integrate two samples
- find conserved cluster marker genes for two samples
- find genes which are differentially expressed between two samples in a cell type specific manner
- visualize genes with cell type specific responses in two samples

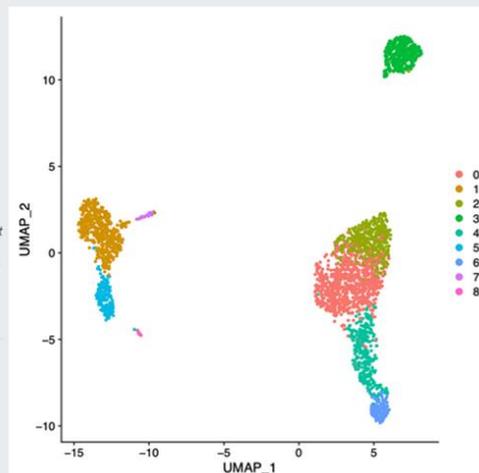
"It is so nice to be able to do the whole workflow in Chipster, compared to the old model, where I had to transfer the tsv file to R-studio and run Seurat there. ... I learned how to use the Seurat tools in Chipster and what all the steps really mean. I learned to check the results after every step to adjust the next steps parameters and to test different PCA plotting tools. I also learned how to find different genes in the clusters and how to visualize them. I never got this far using the R-pipeline." -Pinja, course participant & PhD student from University of Helsinki

Learning objectives

After this course you should be able to:

- use the Seurat tools available in Chipster to undertake basic analysis of single cell RNA-seq data
- name and discuss the different steps of single cell RNA-seq data analysis
- understand the advantages and limitations of single cell RNA-seq data analysis in general and in Chipster

Keywords: Chipster, Seurat, single cell sequencing, RNA-seq, clustering, alignment, cells, cluster markers



Open all Close all

- > 1. Introduction to Chipster
- > 2. Introduction to scRNA-seq
- > 3. Set up a Seurat object and check the quality of cells
- > 4. Normalize gene expression values
- > 5. Identify highly variable genes
- > 6. Scale data and regress out unwanted sources of variation
- > 7. Dimension reduction
- > 8. Cluster cells and detect marker genes for clusters
- > 9. Comparing two samples
- > 10. Test your skills
- > Course certificate and feedback

Flipped classroom: allows to teach the theory in small doses



- Make short lecture videos on each analysis step
- Give lecture videos to course participants 2 weeks *before* the course
- Provide a Google doc where people can write questions
- During the course (online or F2F): recap the theory on topic X, exercises on topic X, check the results and discuss, move to topic Y...
- Our experience:
 - People come to the course well prepared, ask good questions and get more out of the course
 - Course time can be used more efficiently for discussions
- Note: you can use lecture videos by others too!

Summary

- Flipped classroom approach seems to work well
- International collaboration is useful
 - Let me know if you would like to be involved!
 - Topic for the next workshop: how to teach spatial transcriptomics?