

Single cell RNA-seq module

Objectives & Scheme

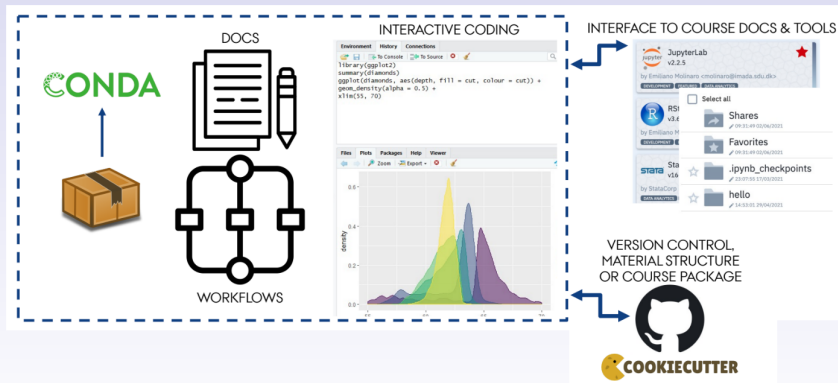
January 2022: "standard" structure for courses and an example scRNA-seq course

- Template for course structure
- Tools and packages organization
- Data collection
- Alignment and analysis of dataset
- Workflows/pipelines
- Course material and test
- Work on a version with Cerebro (?)
- Set on the sandbox

Single cell RNA-seq module

Objectives & Scheme

January 2022: "standard" structure for courses and an example scRNA-seq course



Single cell RNA-seq module

Getting data and preprocessing

16-22.September 2021

- Dataset
 - 15K cells
 - Available from Di Persio et al
 - No sensitive data
- Tools
 - Python (on notebooks)
 - Just began preprocessing the dataset
 - Everything organized in the "cookiecutter" standard structure
 - Includes course documentation in a webpage format
- Next step
 - Do a first coding of the necessary parts of the module
 - Keep an eye on reproducibility/package versions

Single cell RNA-seq module

Getting data and preprocessing

29.September 2021

- Work on the data and course repository
 - 15K cellsData clean and preprocessed, starting to structure some analysis
 - I am collecting a few alignment methods for the long version of the course and for illustration in the short one
 - In parallel I am organizing the skeleton of the webpage/course documentation and the cookiecutter structure
- Tools
 - Python (on notebooks) with integration of some necessary R code (rpy2 package)
 - Mkdocs for the webpage/documentation - hosted on github and easy to write
- Next step
 - Get alignment done and generate also RNA velocity files
 - Resolve the package conflicts
 - Try to finalize the webpage skeleton

Single cell RNA-seq module

Organize course structure and packages

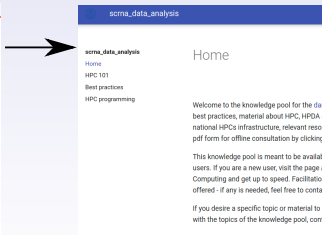
6.October 2021

- Packages organized in conda package manager
- Course structure and documentation with `cookiecutter` and `mkdocs`

```
(SandboxProject) samuele@samuele-ThinkPad-T450s:~$ cookiecutter https://github.com/SamueleSoraggi/cookiecutter-teaching.git
course_name [Amazing-course]: scRNA course
author_name [Health Data Science Team - Aarhus University]: Samuele S
description [Template for structuring an online course at the Health Data Science Sandbox project]: scRNA analysis
Select open_source_license:
1 - No license file
Choose from 1 [1]:
Select python_interpreter:
1 - python3
Choose from 1 [1]:
```



Standard Folder Structure



Webpage/Documentation
coursename.github.io

Single cell RNA-seq module

Organize course structure and packages

6.October 2021

- Packages organized in conda package manager
- Course structure and documentation with `cookiecutter` and `mkdocs`
- two pipelines choices for raw data → gene counts and dynamics.
- Both `snakemake` and `gwf` to write pipelines
- Code, notes and assignments in `jupyter notebooks` - also integrated in webpage

Course repository and material soon available as a repository on GitHub.

22. February 2022

- **scRNA-seq module in Python and R** → Python version ready, R in development → feedback
 - extension into 1 week summer course (AU/KU)
 - recycling material for semester 2 weeks session (AU)

22. February 2022

- **scRNA-seq module in Python and R** → Python version ready, R in development → feedback
 - extension into 1 week summer course (AU/KU)
 - recycling material for semester 2 weeks session (AU)
- **Research single cell data and analysis**, 2 datasets
 - integrated data from lean vs obese patients tutorial → data and code ready, waiting for publication (AU)
 - Spatial data w/Alex → waiting news from phd student (AU/KU)

22. February 2022

- **scRNA-seq module in Python and R** → Python version ready, R in development → feedback
 - extension into 1 week summer course (AU/KU)
 - recycling material for semester 2 weeks session (AU)
- **Research single cell data and analysis**, 2 datasets
 - integrated data from lean vs obese patients tutorial → data and code ready, waiting for publication (AU)
 - Spatial data w/Alex → waiting news from phd student (AU/KU)
- **Environmental DNA** → Detection of leptospirosis and possible vectors in Okinawa → data ready (AU)

22. February 2022

- **scRNA-seq module in Python and R** → Python version ready, R in development → feedback
 - extension into 1 week summer course (AU/KU)
 - recycling material for semester 2 weeks session (AU)
- **Research single cell data and analysis**, 2 datasets
 - integrated data from lean vs obese patients tutorial → data and code ready, waiting for publication (AU)
 - Spatial data w/Alex → waiting news from phd student (AU/KU)
- **Environmental DNA** → Detection of leptospirosis and possible vectors in Okinawa → data ready (AU)
- **GWAS** study and/or course → only in the planning phase (AU/KU)

22.February 2022

Testing material is ready.

- `conda` **environment** auto-loading
- `python` **kernel** auto-loading and shown in `jupyterlab`
- **documentation** webpage builds correctly
- material on **github** (not the data) as per cookiecutter template
- running cost **<5 DKK** (8-32 vCPUs). No GPUs.

22.February 2022

Testing material is ready.

- `conda` **environment** auto-loading
- `python` **kernel** auto-loading and shown in `jupyterlab`
- **documentation** webpage builds correctly
- material on **github** (not the data) as per cookiecutter template
- running cost **<5 DKK** (8-32 vCPUs). No GPUs.

But!!!

22. February 2022

Testing material is ready.

- `conda` **environment** auto-loading
- `python` **kernel** auto-loading and shown in `jupyterlab`
- **documentation** webpage builds correctly
- material on **github** (not the data) as per cookiecutter template
- running cost **<5 DKK** (8-32 vCPUs). No GPUs.

But!!!

- multiple users would **overwrite** each other cache and savefile/logs
- each user must have **its own** notebooks to work with

22.February 2022

Implementing final changes

- only environment and data are at a **fixed location** (ideally read-only by sharing settings)
- auto-loading includes **cloning github** repo in a personal folder → few MBs taken
- eventually a last command **removes excess savefiles**

22.February 2022

Implementing final changes

- only environment and data are at a **fixed location** (ideally read-only by sharing settings)
- auto-loading includes **cloning github** repo in a personal folder → few MBs taken
- eventually a last command **removes excess savefiles**

I will soon send around the information for testing and feedback.