Single cell RNA sequencing data analysis
Bring Your Own Data information
10 April, 2024

Åsa Björklund, Jennifer Fransson & Susanne Reinsbach

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

SciLifeLab

April 10th

9:00 - Short introduction

9:15 - Data managers present

You work on your own and we will be around to help.

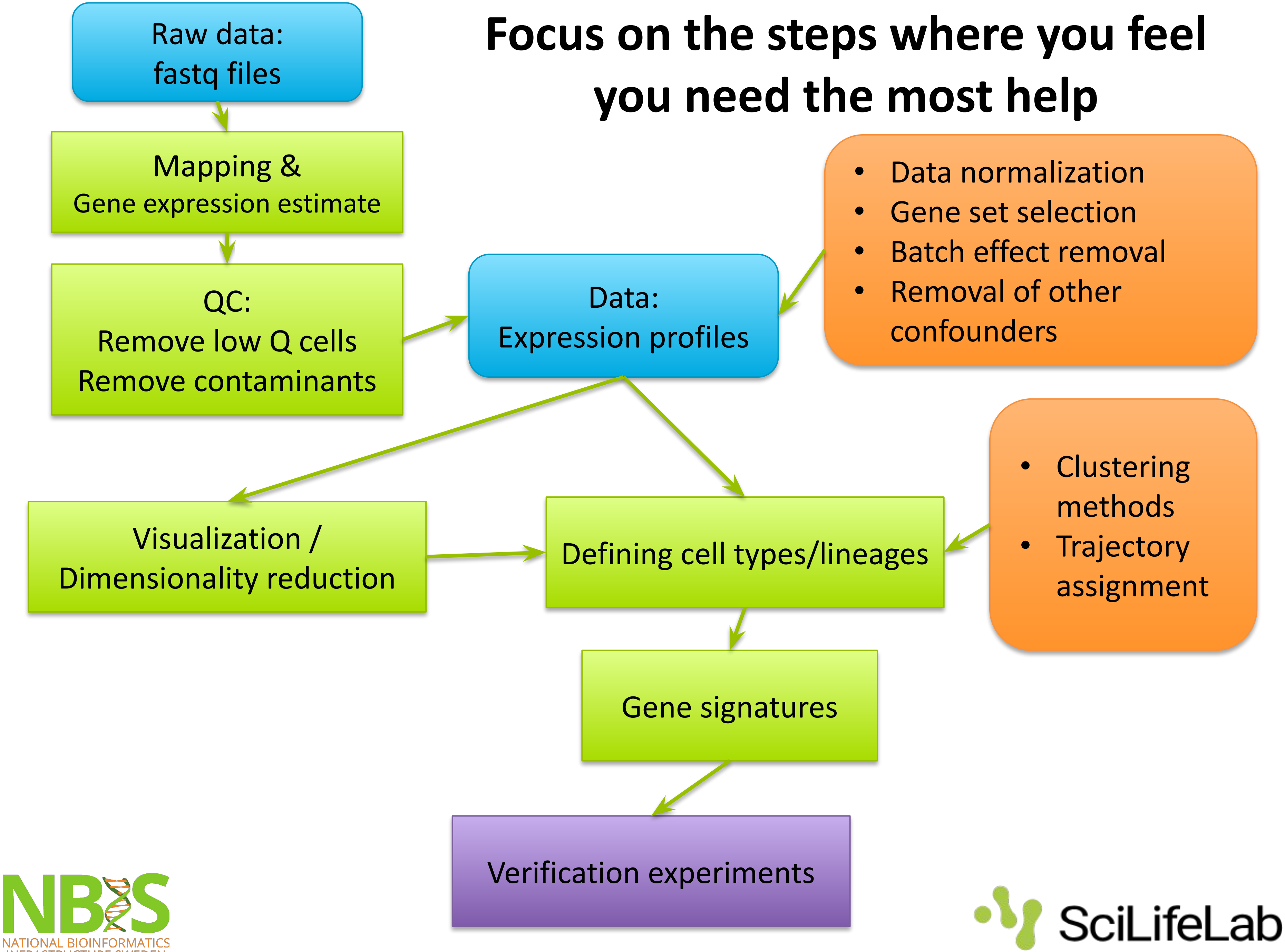
14:30 - Presentations of all your projects.

16:00 - Summary of the day.

Datasets

- Do you have your own data - use that!
- If not - select a dataset from a tissue/celltype that you are interested in.
- Do not select a very large dataset, will be too much computational time during one day!
  - Subsample the cells or select another dataset.
- **If you have sensitive human data, make sure you have permission to work on count matrices locally or on the server.**

Focus on the steps where you feel you need the most help



Raw data:
fastq files

Mapping &
Gene expression estimate

QC:
Remove low Q cells
Remove contaminants

Data:
Expression profiles

- Data normalization
- Gene set selection
- Batch effect removal
- Removal of other confounders

Visualization /
Dimensionality reduction

Defining cell types/lineages

- Clustering methods
- Trajectory assignment

Gene signatures

Verification experiments

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

Breakout groups

Please answer our short survey (link also in slack)

https://docs.google.com/forms/d/e/1FAIpQLSeVtYQmluN0EnCNEpw0D2CLQn-P-1x0skAQ7aZQ6wonSvWG6g/viewform

We will try to make groups that makes sense.

NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

SciLifeLab

Compute environments

- If you have a small dataset - less than 50K cells, probably okay to work on the Serve system. Upload it using the Serve File Manager
- For large datasets, use one of the following options:
  - Downsample the dataset and upload to Serve
  - Use the course containers and Docker locally
  - Use a local setup (at your own risk)
    - Conda-files are available on GitHub but not tested outside of the containers
    - Packages can be installed as you go, but different versions can cause issues with dependencies
    - If you haven't already - consider reproducibility! https://uppsala.instructure.com/courses/107067
- Make sure it runs before April 10th
- Ask in the slack channel **#precourse** if you have problems.

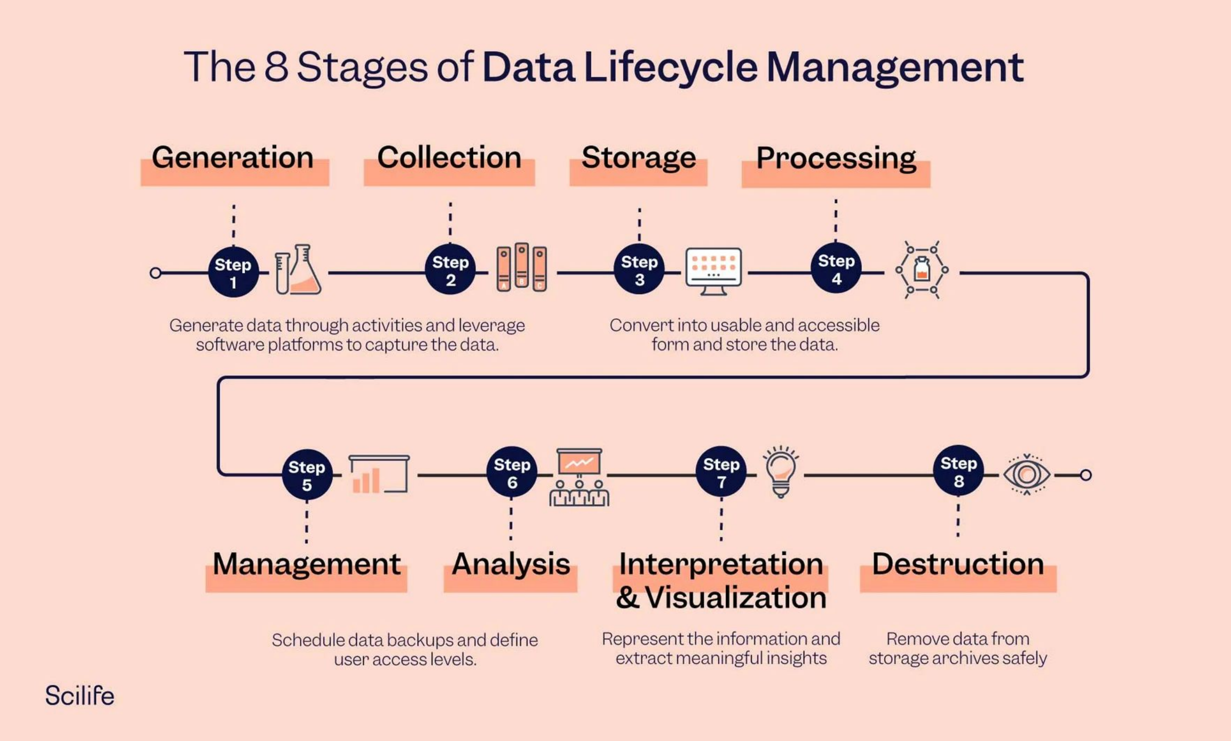
More information

https://nbisweden.github.io/workshop-scRNAseq/other/containers.html

- Option A - running labs on Serve.
  - Use the File Manager at the bottom of the project dashboard to upload your data.
- Option B - running labs locally.

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

Data Management



Generate data through activities and leverage software platforms to capture the data.

Convert into usable and accessible form and store the data.

Schedule data backups and define user access levels.

Represent the information and extract meaningful insights

Remove data from storage archives safely

NBIS Data Managers - our guides through the data life cycle will have a short presentation and be present for consultations during the day.

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

Presentation

We want a **very** short - max 5 minutes presentation of what you have done.

- 1 slide
- What dataset
- Which analysis steps did you perform?
- Which were your main results?
- What were the biggest problems you encountered?

QUESTIONS?

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

Docker tutorial

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN