Single cell RNA sequencing data analysis
Practical exercises
4-6 February 2019

Åsa Björklund

asa.bjorklund@scilifelab.se

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Practical exercises

- https://nbisweden.github.io/workshop-scRNAseq/exercises

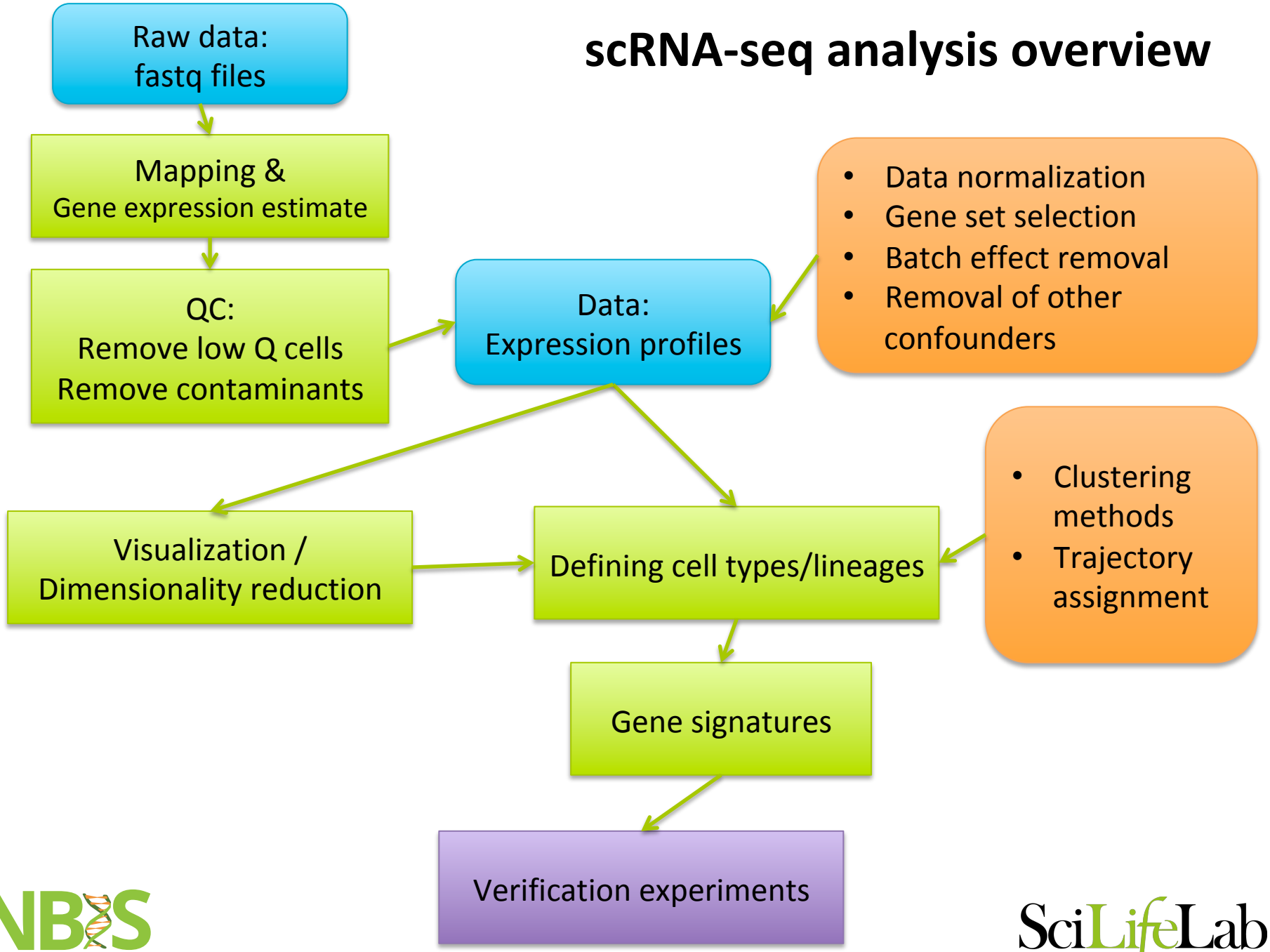
Smörgåsbord of exercises:

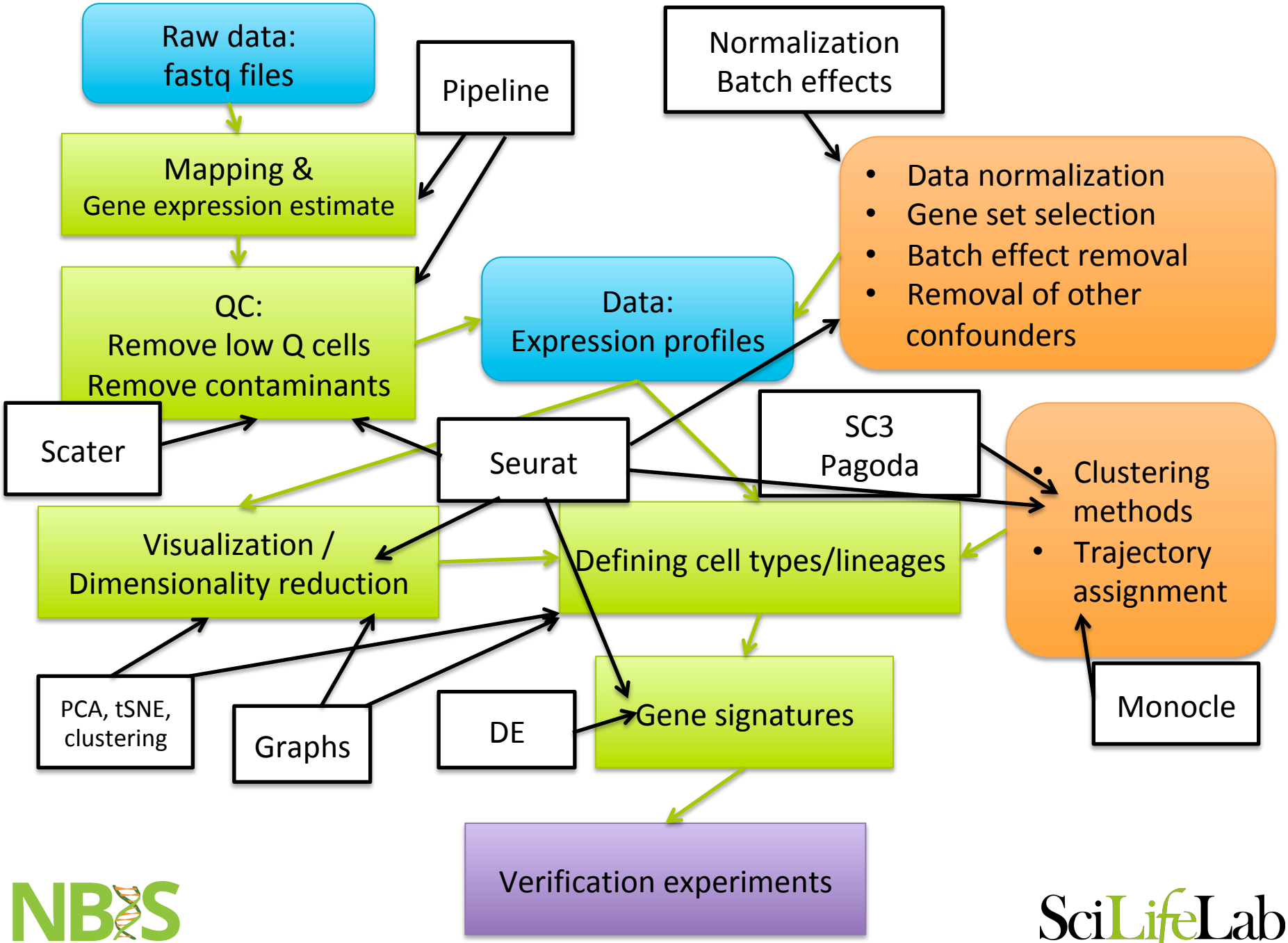
- Pipeline – SmartSeq2
- PCA, tSNE, clustering – own code
- KNN-graphs – igraph package
- QC with Scater package
- Batch effects
- Normalization methods
- SC3 clustering
- Pagoda

- Seurat package
- Differential expression
- Monocle pseudtime

- Biomart
- Example Sbatch script

scRNA-seq analysis overview





NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Practical exercises

- We have several different exercises and we do not expect all of you to have time to go through all of them during the few hours that we have.
- Focus on the ones you feel is most relevant for your research and start there.

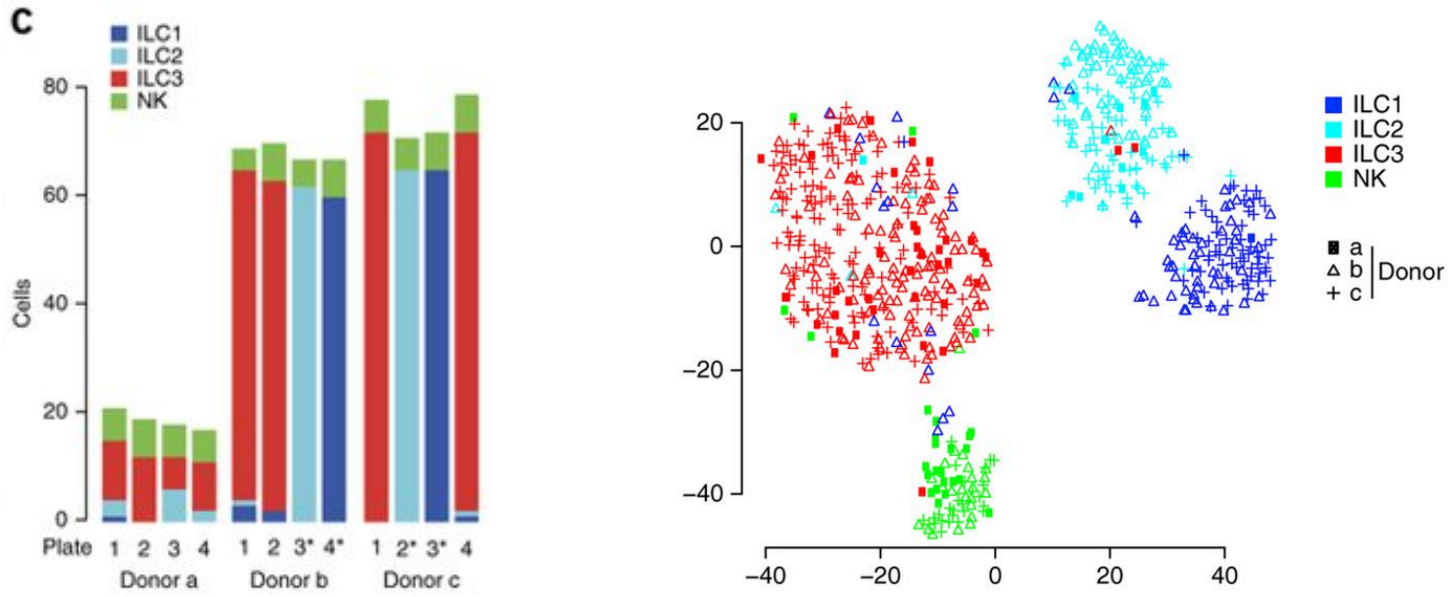
Troubleshooting

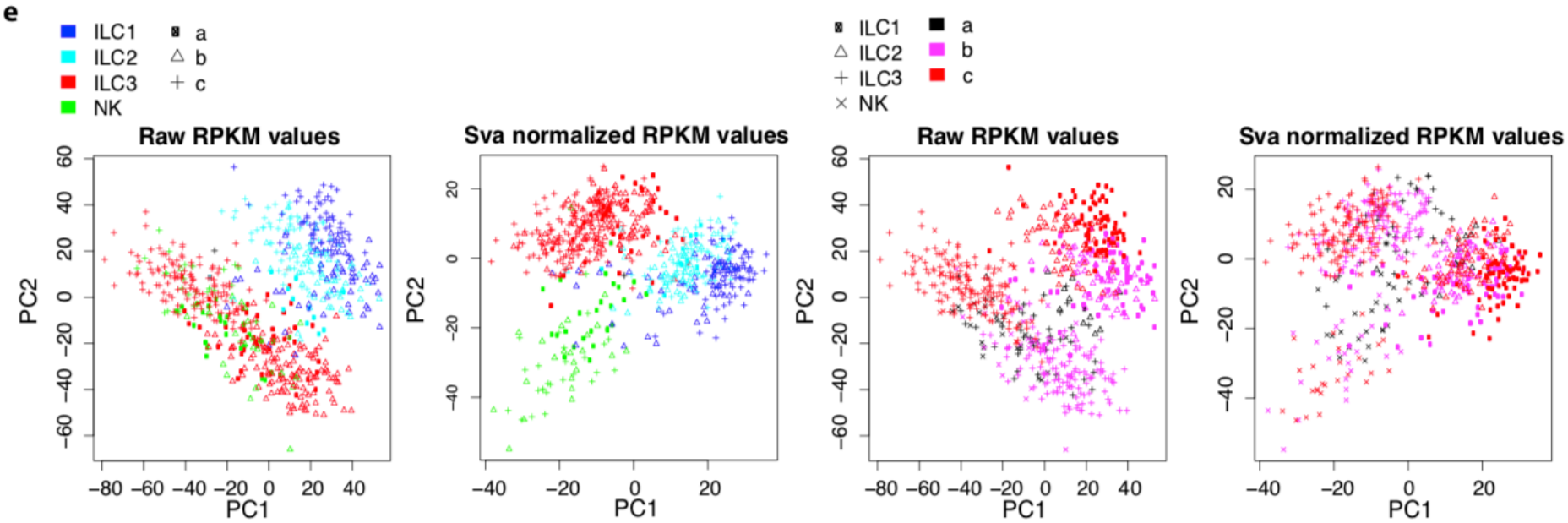
- We have put up a FAQ page where we will try to put up common problems & questions
- It is important that you learn how to troubleshoot yourselves.
  - Look at your error messages, perhaps the answer is there?
  - If not – Google is your best friend! Forums like Seqanswers, Stackexchange, Bioconductor support forum and some specific forums for each package may have the answer.
- TAs are there to answer any questions and give suggestions, but we may not always have the answer.

The datasets

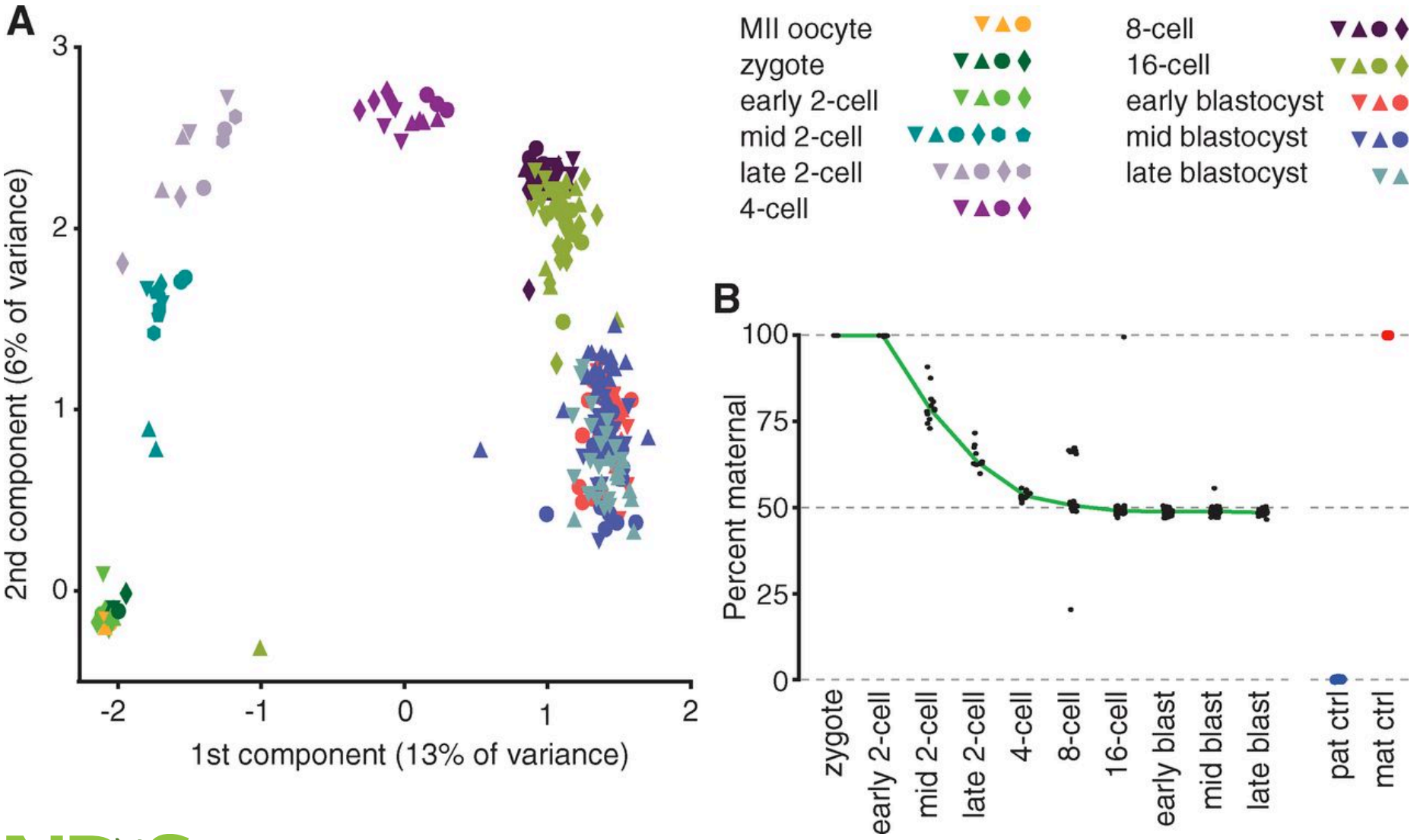
- Human tonsil innate lymphoid cells (ILCs)
  - SmartSeq2 data from 3 donors
- Mouse embryonic development
  - SmartSeq2 data from oocyte to blastocyst stage

ILCs from human tonsils





Mouse embryonic development



NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Using the Uppmax allocations

- Project space /proj/g2019002
- Allocations each day: g2019002_04 / g2019002_05

- **OBS!** Wednesday 6th of February is Maintenance window for Uppmax – if you want to run the pipeline exercise, do it Monday/Tuesday!

Working with your own data

- Keep in mind that some steps of the tutorials have smaller datasets or pre-computed files for compute-intensive steps. Work with parallel R-sessions and run other exercises while you wait or submit to Uppmax as sbatch jobs.
- If you need to convert between gene names, ensembl ids etc, have a look at the BiomaRt tutorial
- Our main focus will be to get everyone through their tutorials, but if time permits we can help you with conclusions from your data.