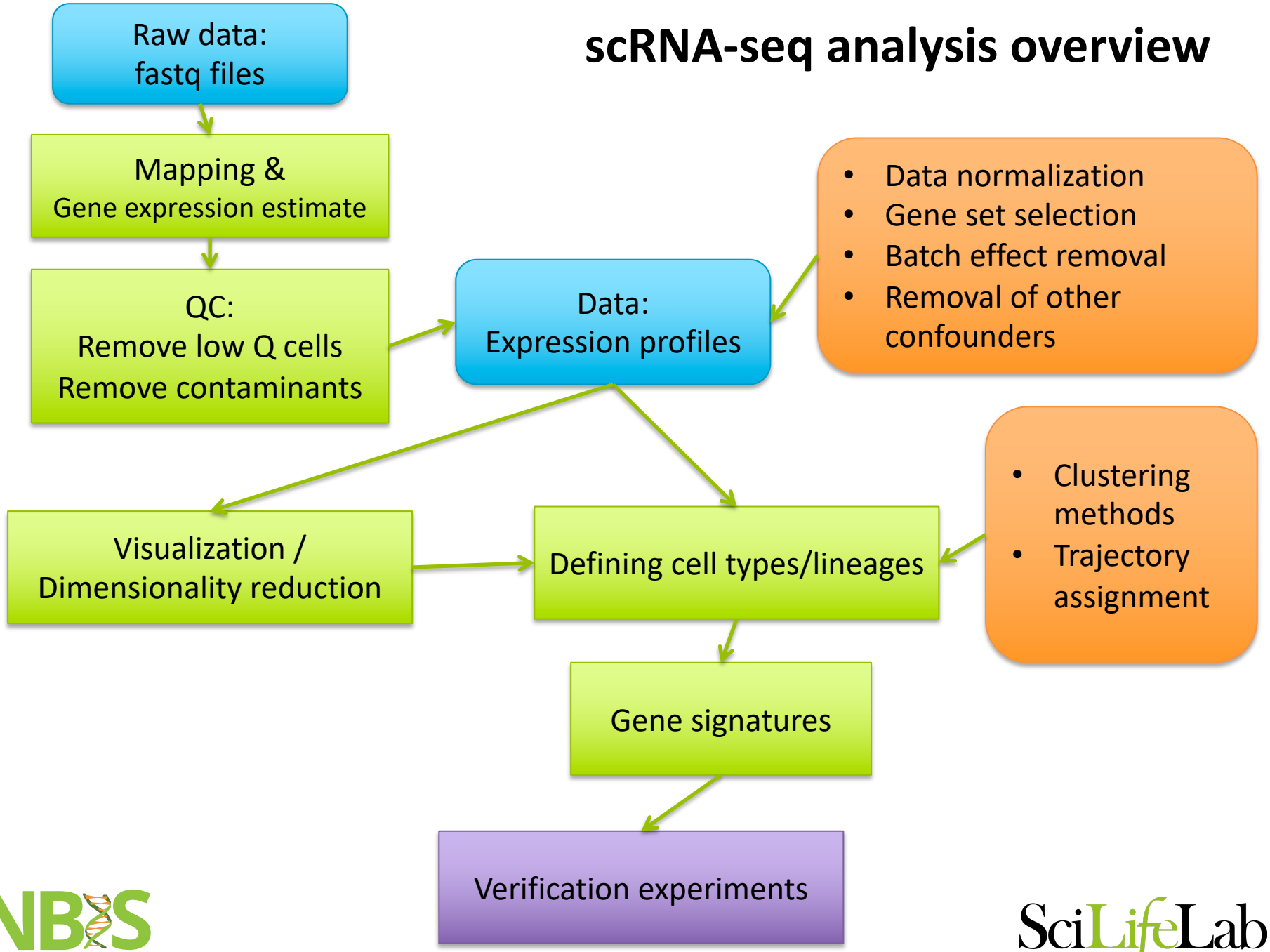
Single cell RNA sequencing data analysis, 2023

Åsa Björklund & Paulo Czarnewski

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

scRNA-seq analysis overview



Some take-home messages

- Data analysis is very seldom a straight line – one pipeline fits all.
  - Often requires several iterations of filtering data, exploring data, refiltering, exploring again, discovering technical artifacts, normalization, exploring again, etc. etc.

- Get to know your data – what types of variation do you have?
  - PCA is a good tool for exploring data
- Apply appropriate methods to control for problems that you see.

- Always check for:
  - Batch effects – think of all possible batches.
  - Cell cycle effects if appropriate
  - Separation due to nUMI / nGene / percent mito
- Both at the start of a project and at the end for your final clustering.

- Variable gene selection is a very critical step
  - Filter to much and you may lose populations
  - Keep to much and you may have too much noise

- Similar for choice of PCs

- Clustering – try out a few different approaches
  - Consensus of different methods gives confidence
  - If they do not agree – figure out why!

- Use your biological knowledge to evaluate the results
- Warning! Do not overfit your data to fit your initial hypotheses. Keep an open mind ;-)

- Remember that bioinformatics tools are giving predictions not the truth – always keep a critical mind!
  - Clustering
  - Differential expression
  - GSEA
  - Celltype prediction
  - Deconvolution

- In this course we point out many of the problems that can occur..
- Do not worry too much, in most cases, a standard workflow works well!

- scRNAseq analysis is a fast evolving field with new methods being published all the time.
  - Try to keep up with development
  - **BUT!** You cannot test every new method out there!

Reproducible research in R

- R / Rstudio in Docker containers
  - https://www.andrewheiss.com/blog/2017/04/27/super-basic-practical-guide-to-docker-and-rstudio/
  - https://github.com/rocker-org/rocker
- OBS! On Uppmax – only Singularity containers are allowed. Most Docker images can be converted.
- Learn more on containers etc:
  - http://nbis-reproducible-research.readthedocs.io/en/latest/
- Rstudio package management – Packrat
  - https://rstudio.github.io/packrat/
- Conda installations of packages – can use conda on both bianca and rackham – `module load conda`

NBIS course in reproducible research:
https://uppsala.instructure.com/courses/73110

Compute resources

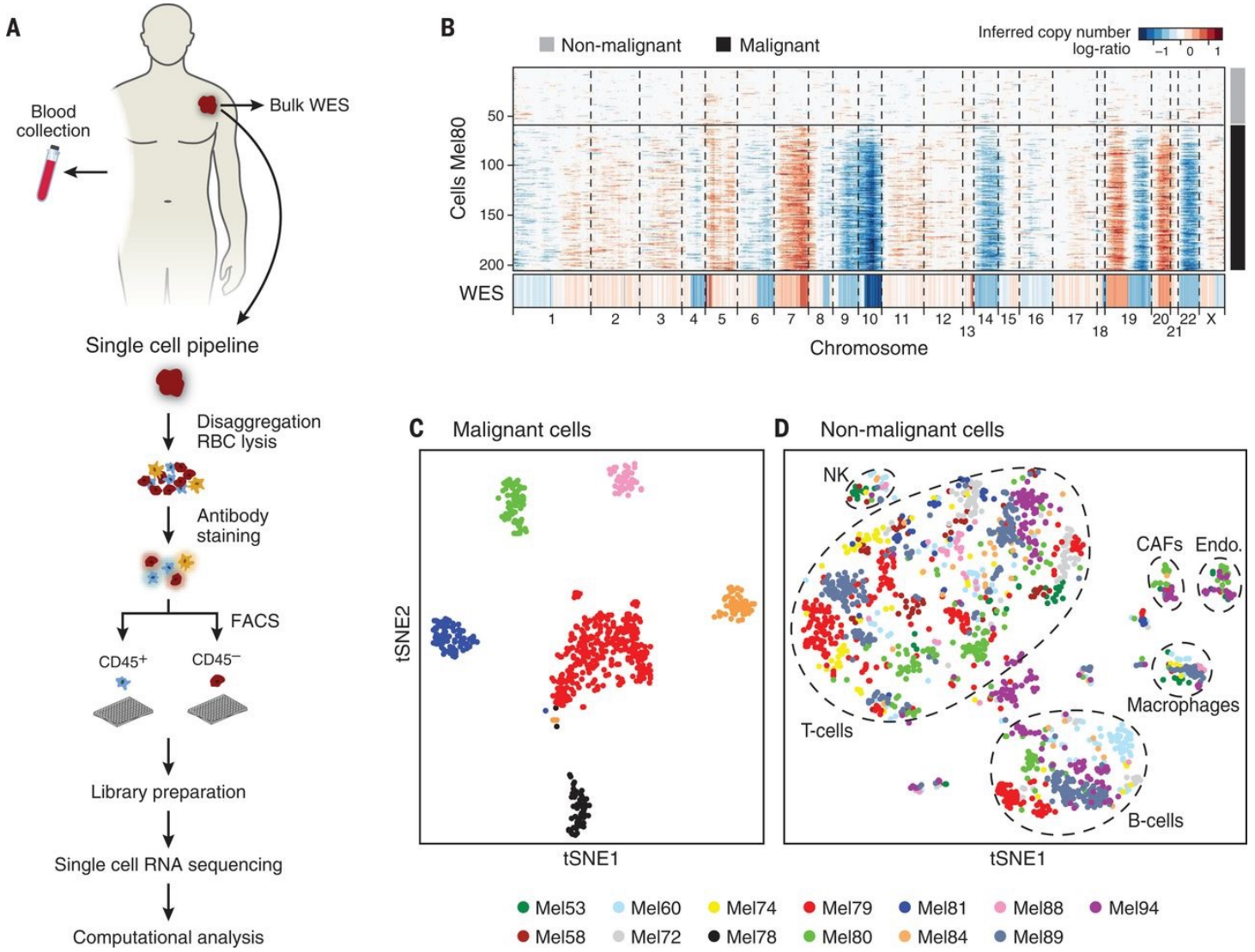
- In these exercises the datasets were small, but you may have many more cells/samples.
- Structure your code to avoid duplication of matrices and expansion of sparse matrices
  - `rm() & gc()`
- Plan ahead for compute resourses, local computer, uppmax or other HPC clusters.
- Human data – raw reads only on encrypted servers like Bianca. Count matrices is fine to use in other places.

Some conda comments

- Conda disc space usage
- `conda env remove –n myenv`
  - Will remove an environment
- `conda clean –all`
  - Will remove all tarballs and packages that are not used.

- We have covered the basic processing, but there is much more you can do…

Copy-number variation (CNV) profiling with RNAseq

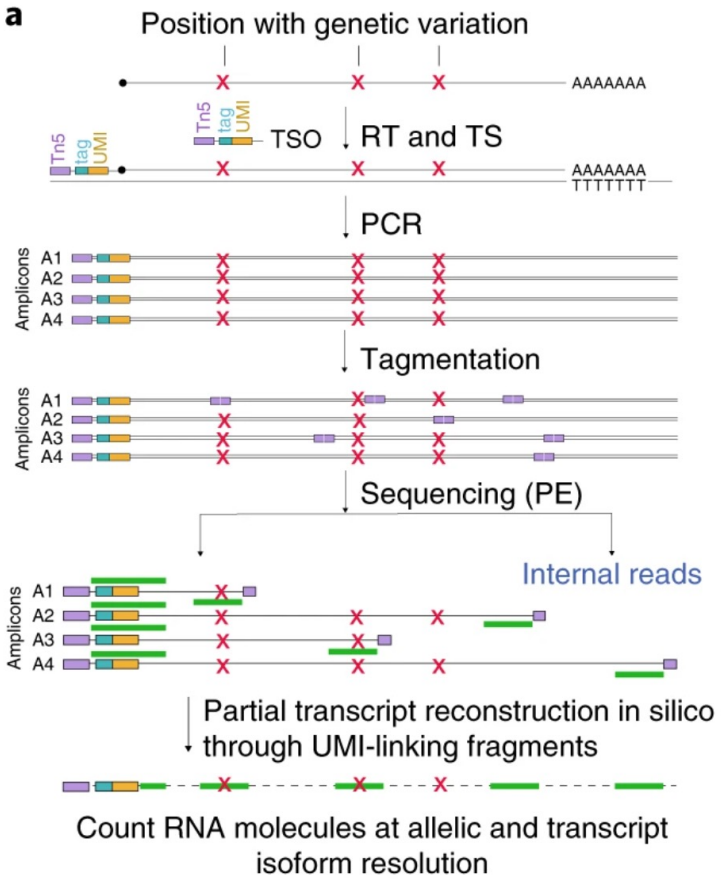


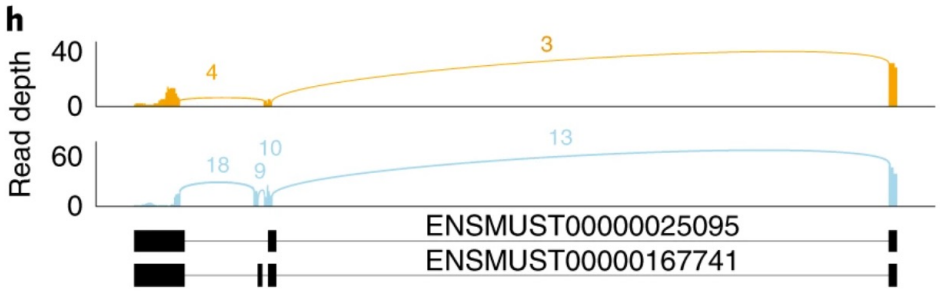
(Tirosh et al. *Science* 2016)

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

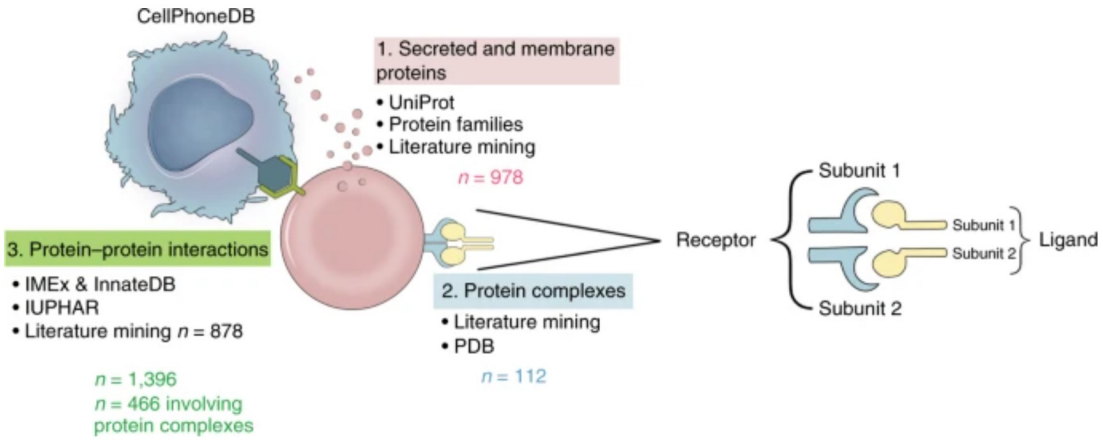
Allele and isoform information with SmartSeq3

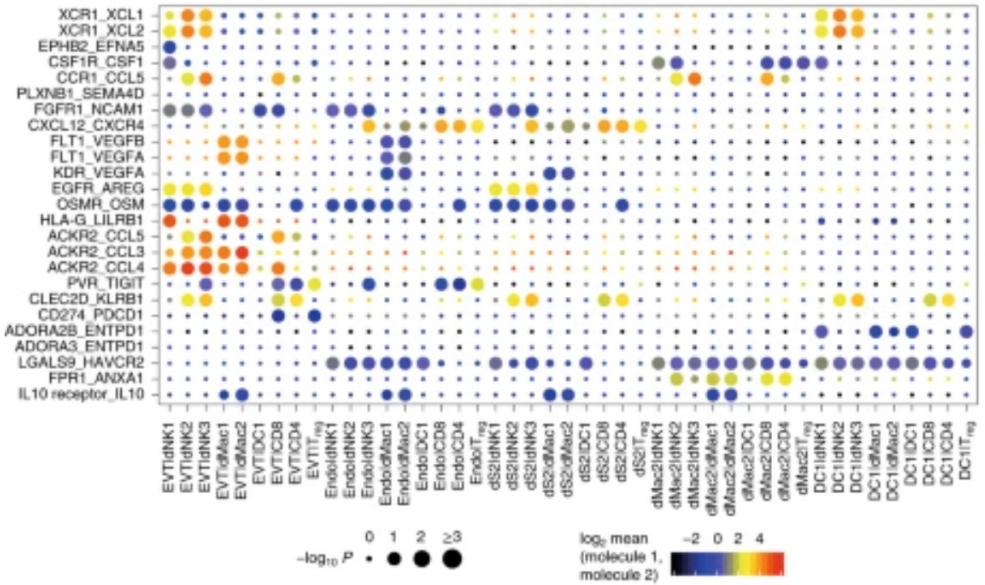




(Hagemann-Jensen et al. Nat. Biotech 2020)

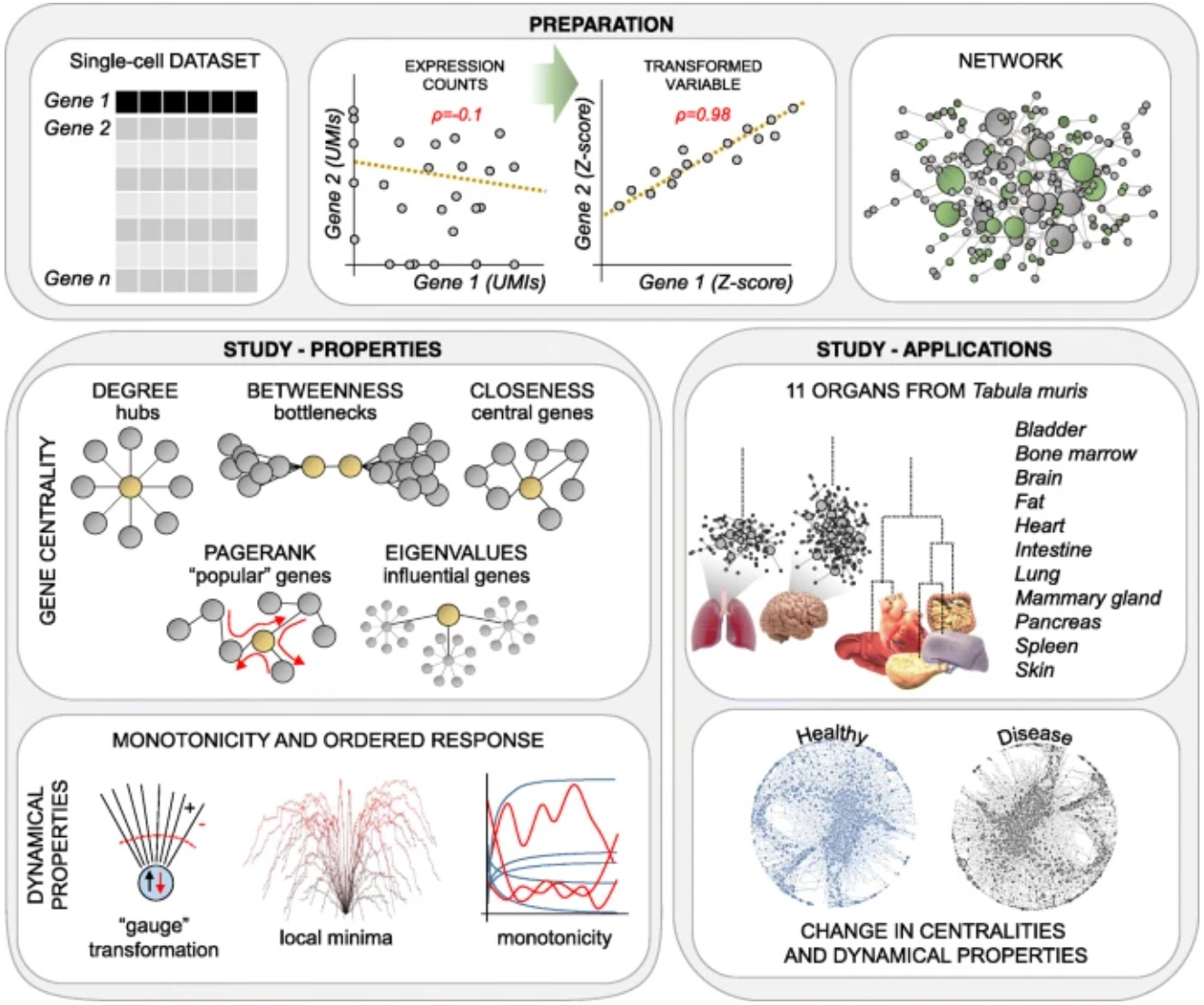
Receptor ligand interaction





(Efremova et. Al. Nat. Protocols 2020)

Gene regulatory networks

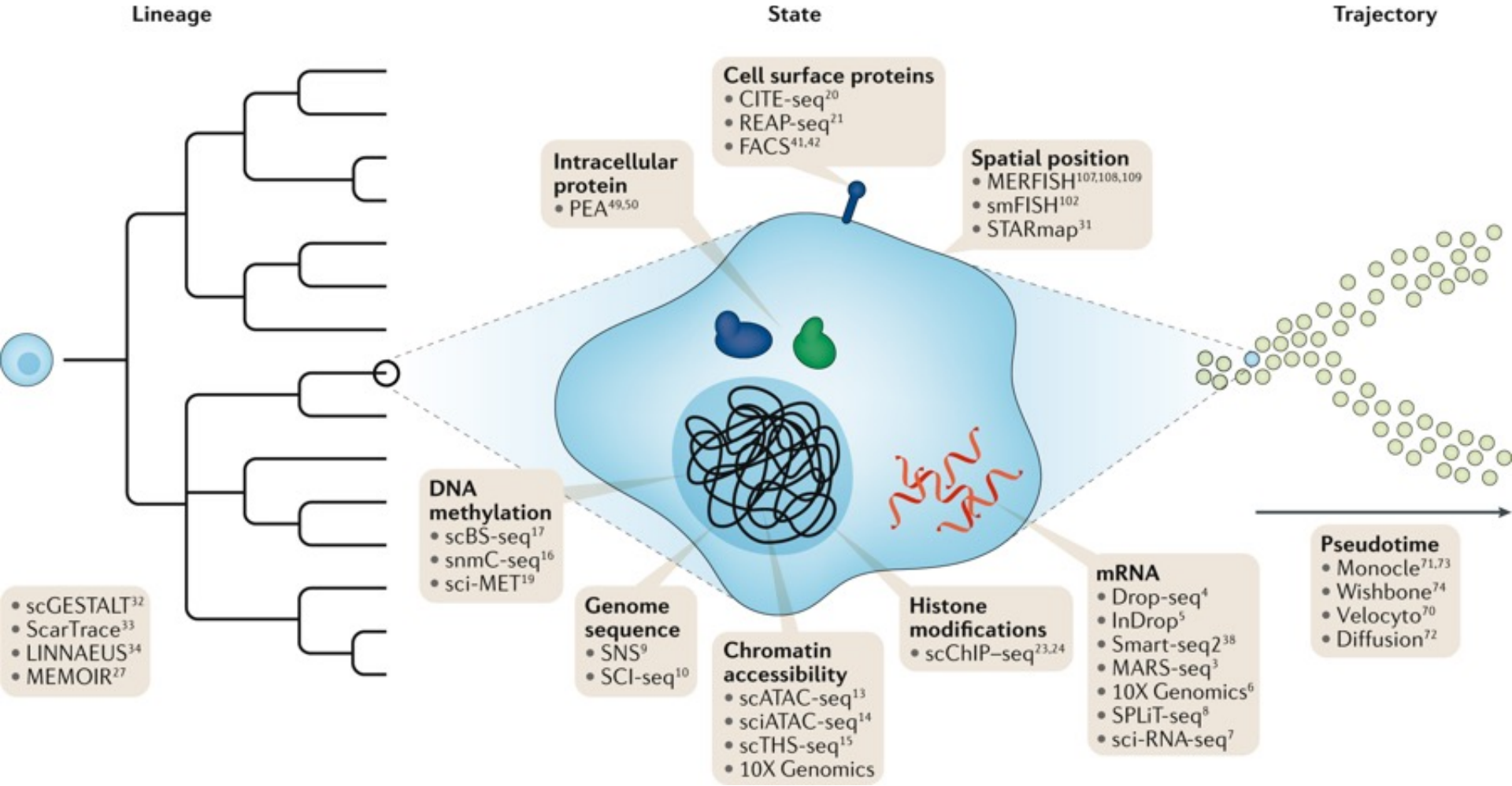


(Iacono et al. Genome Biology 2019)

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

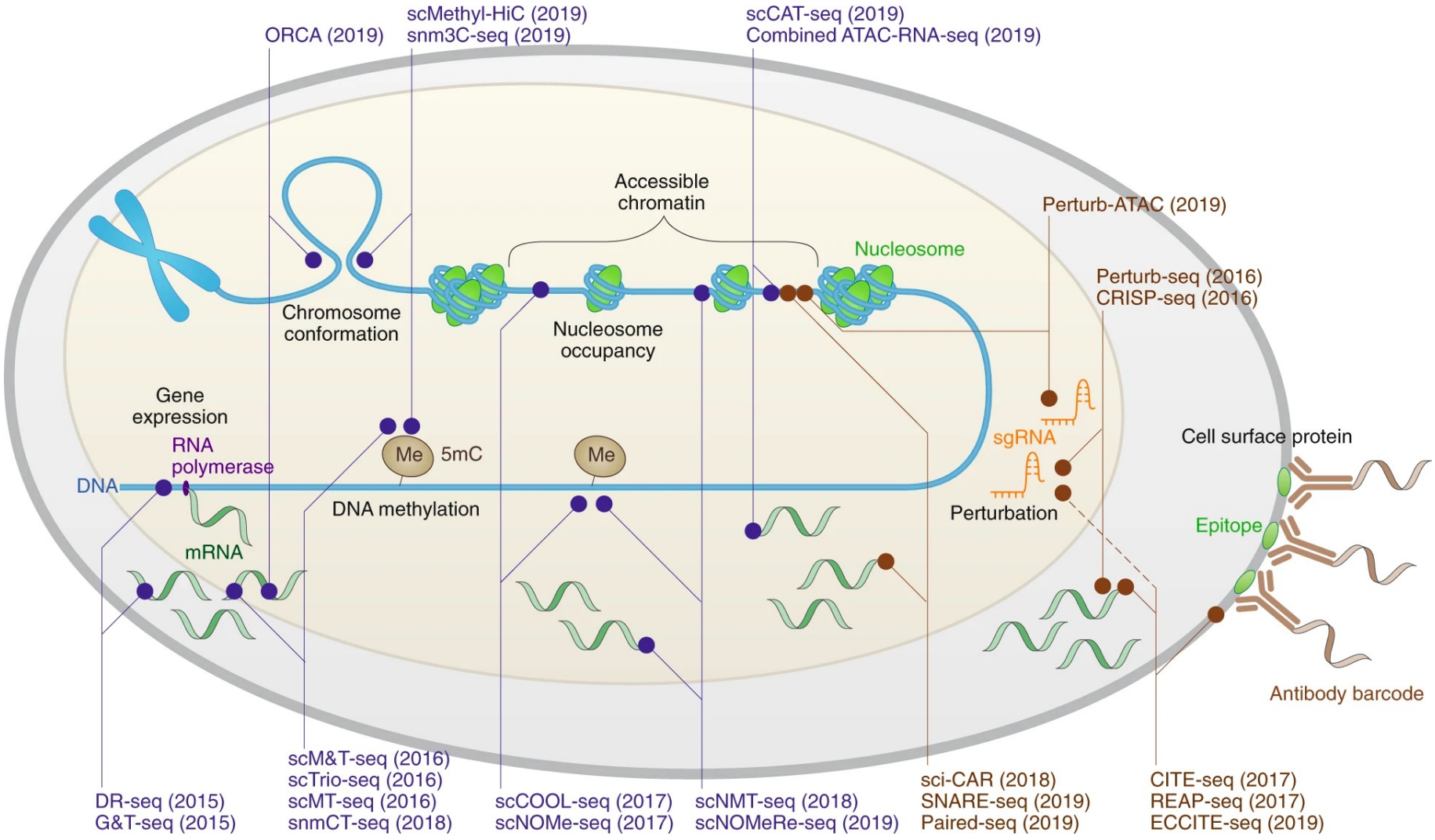
SciLifeLab

Single cell omics



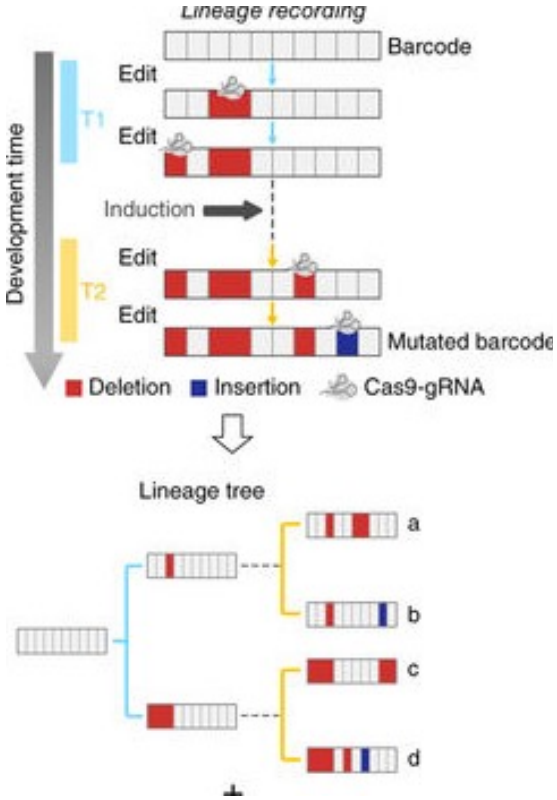
(Stuart & Satija, Nature Rev. Genetics 2019)

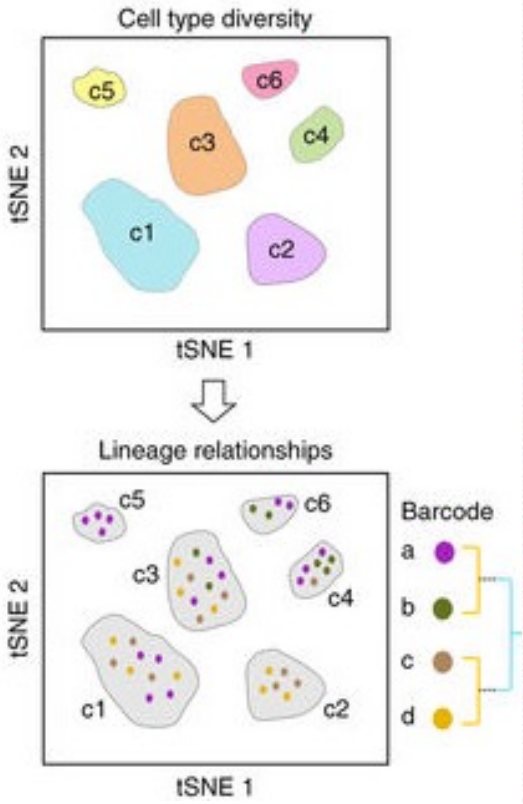
SC Multimodal omics

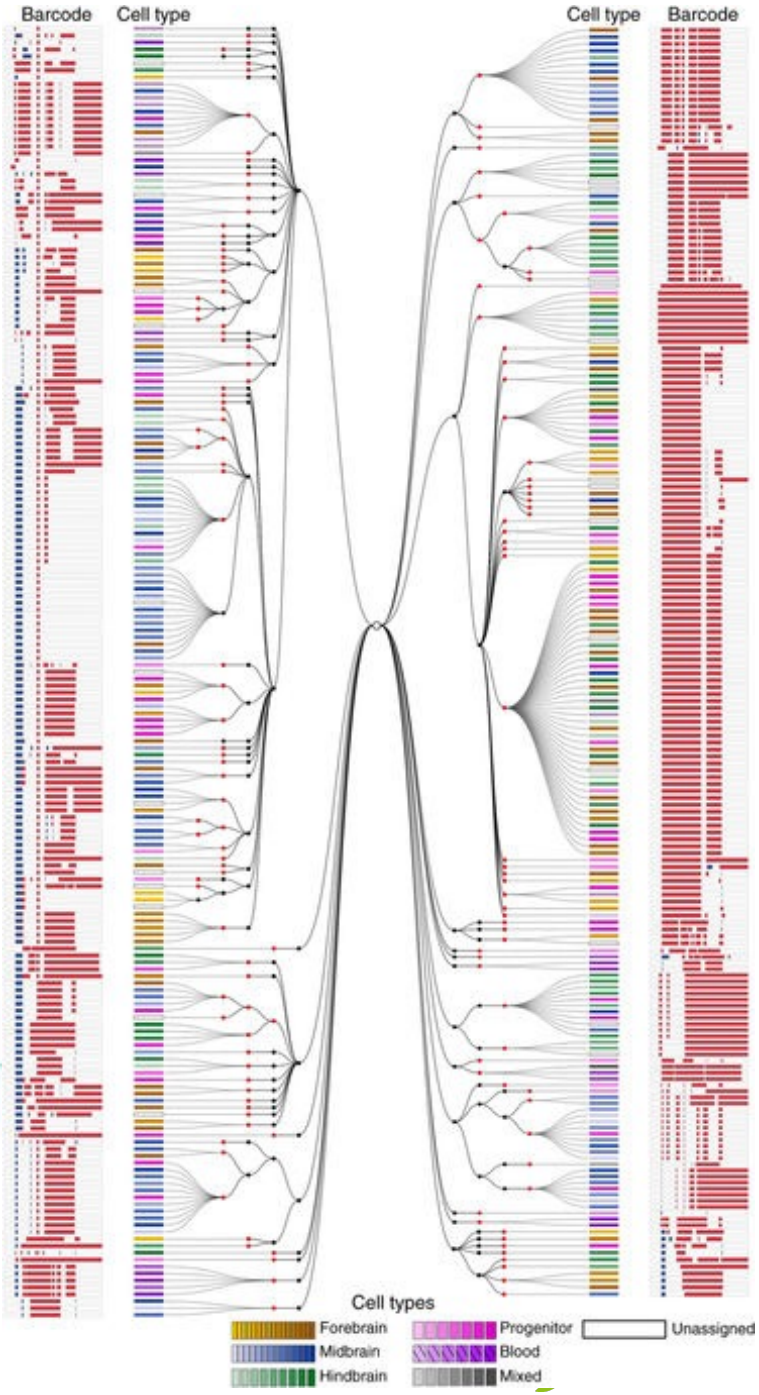


(Zhu et al, Comment in Nature Methods, 2020)

scGESTALT – lineage tracing and cell profiling with CRISPR-Cas9 editing of barcodes



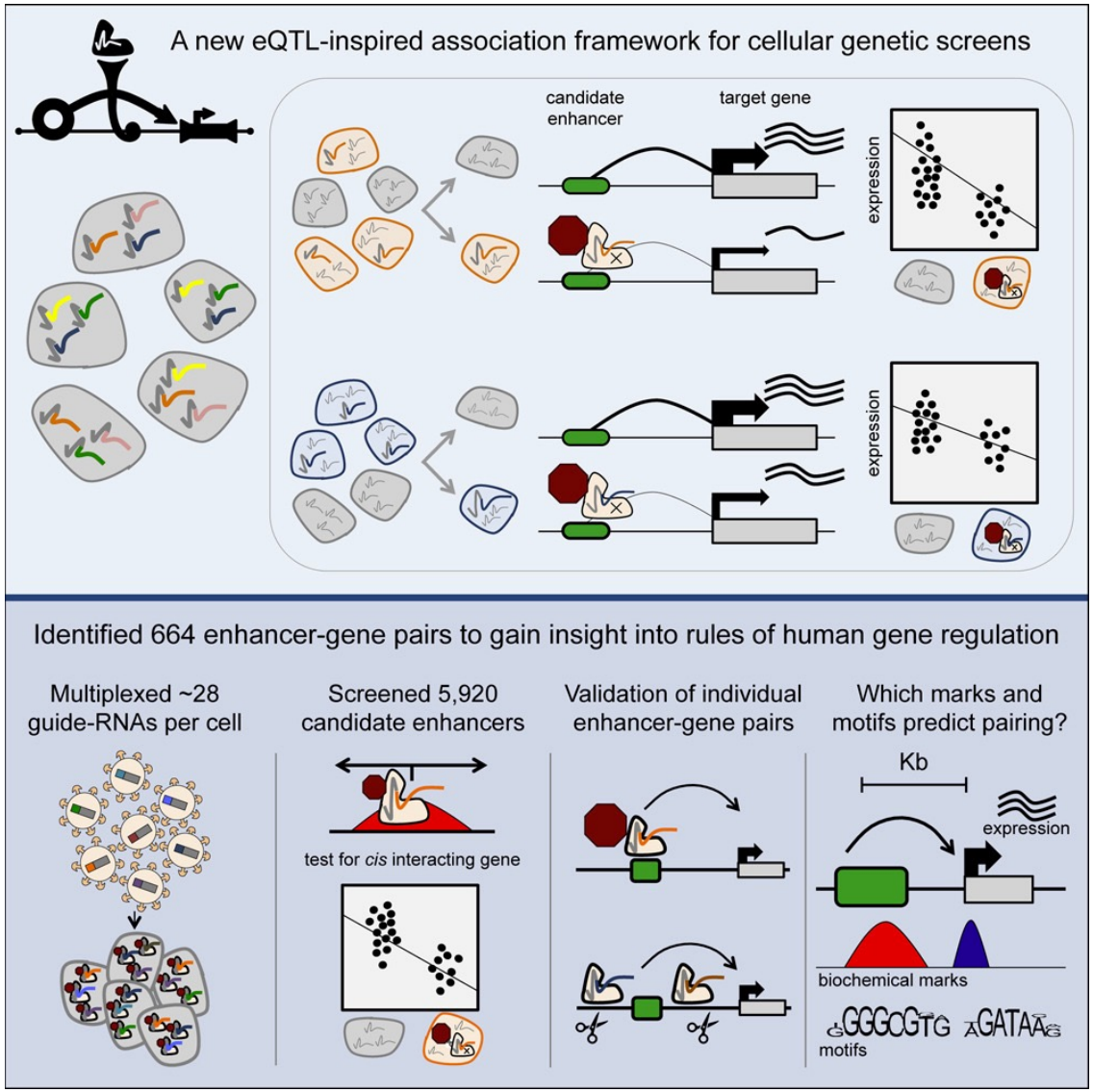




(Raj et al. Nature Biotech 2018)

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

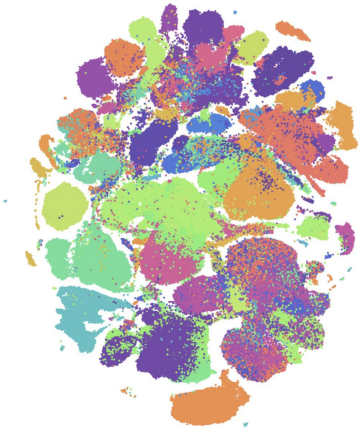
crisprQTL mapping for enhancer-gene pairs



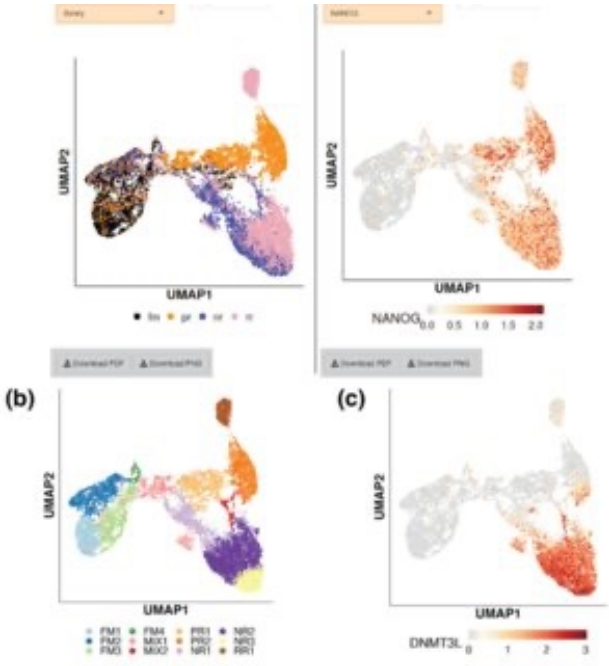
(Gasperini et. al. Cell 2019)

Interactive visualization

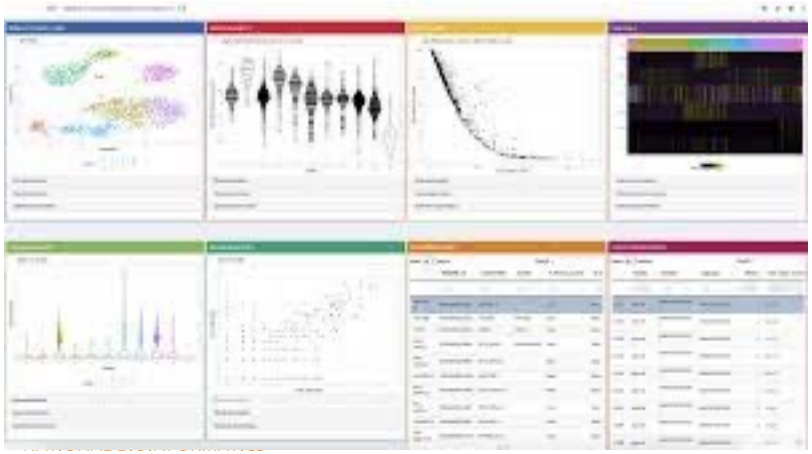
Cellxgene



Shinycell



iSEE



TissUUmaps

Resources

- Course at: https://hemberg-lab.github.io/scRNA.seq.course/
- Scanpy course: https://www.sc-best-practices.org/
- Orchestrating Single-Cell Analysis with Bioconductor http://bioconductor.org/books/3.13/OSCA/
- Many of the packages have good tutorials on their websites
- Repo with scRNA-seq tools: https://github.com/seandavi/awesome-single-cell
- Single cell assay objects for many datasets: https://hemberg-lab.github.io/scRNA.seq.datasets/
- EBI Single cell expression atlas: https://www.ebi.ac.uk/gxa/sc

Need help?

- NBIS project support
- Courses in programming and other types of analyses.
- Drop-in sessions every Tuesday 14.00
- More info at: http://nbis.se/

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

Please fill in the Evaluation Form

Your feedback is important so that we can help improve the course.

Good luck with your analyses!

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab