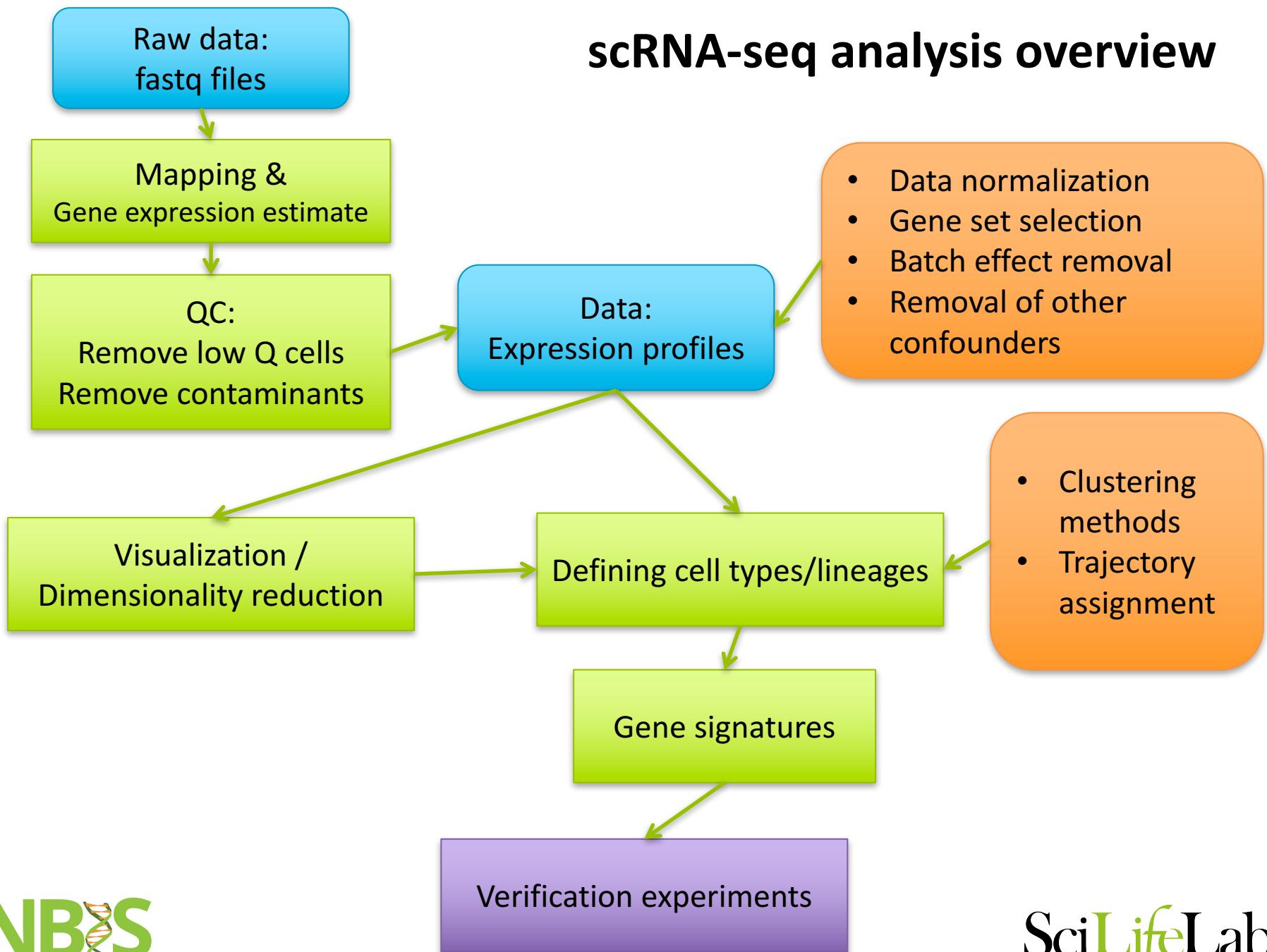




Single cell RNA sequencing data analysis, 25-29 January 2021

Åsa Björklund & Paulo Czarnewski

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
- Both at the start of a project and at the end for your final clustering.

- Variable gene selection is a very critical step
 - Filter too much and you may lose populations
 - Keep too 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 ;-)

- 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

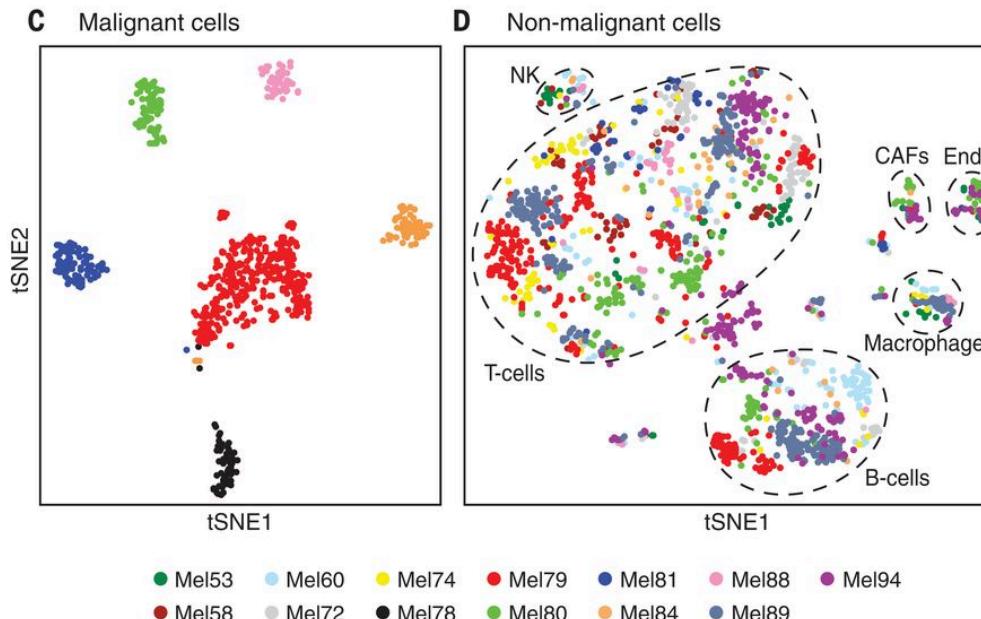
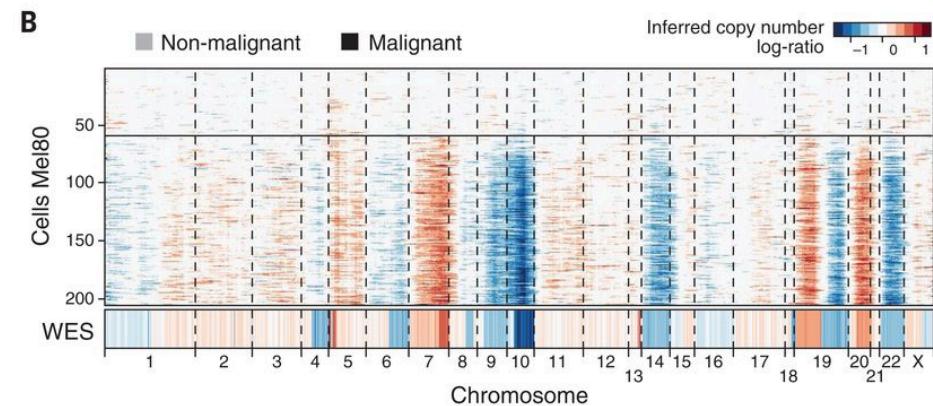
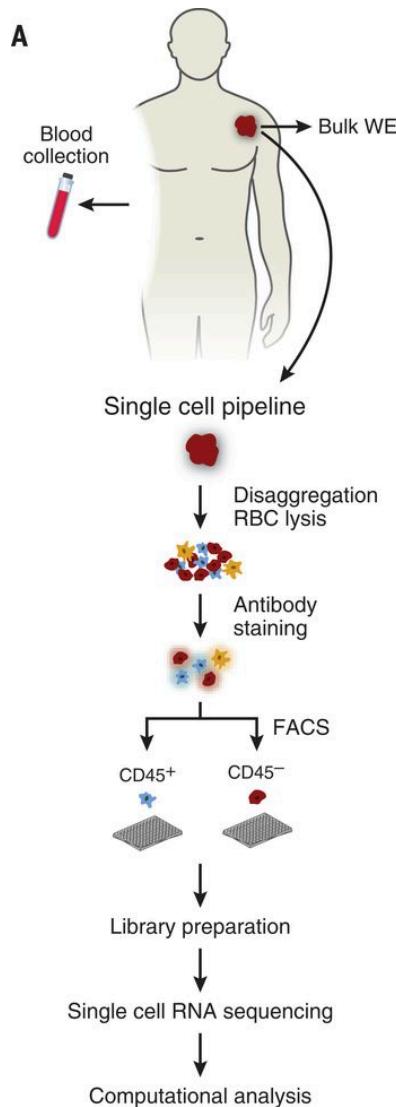
- R / Rstudio in Docker containers
 - <https://www.andrewheiss.com/blog/2017/04/27/super-basic-practical-guide-to-docker-and-rstudio/>
 - <https://github.com/rockr-org/rockr>
- OBS! On Uppmax – only Singularity containers are allowed. Most Docker images can be converted.
- Learn more on containers etc:
 - <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

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.

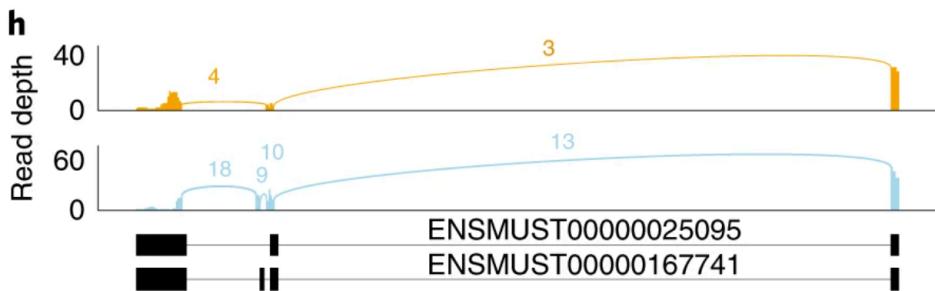
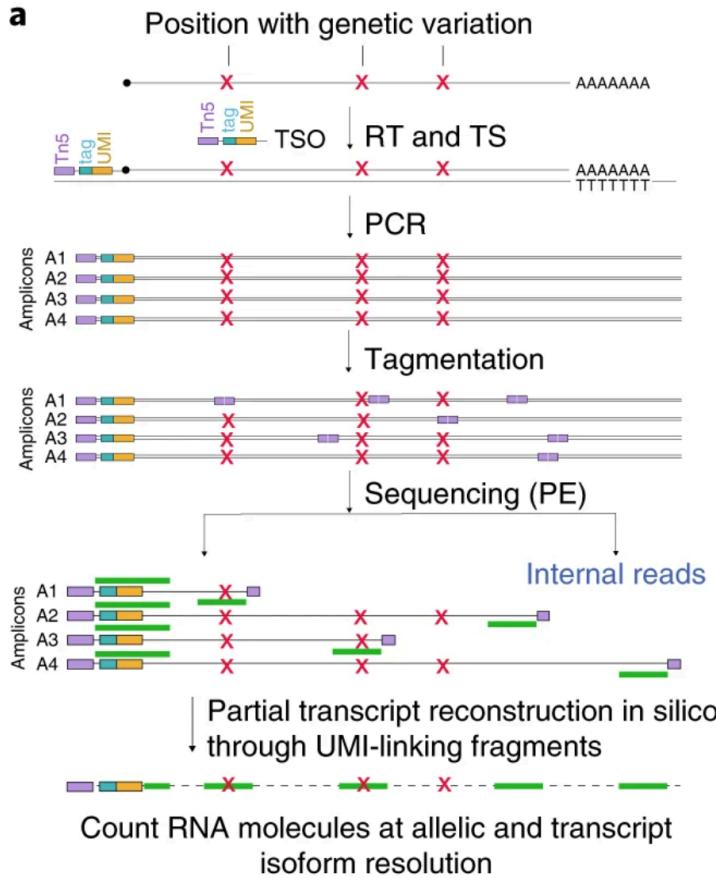
- We have covered the basic processing, but there is much more you can do...

Copy-number variation (CNV) profiling with RNAseq

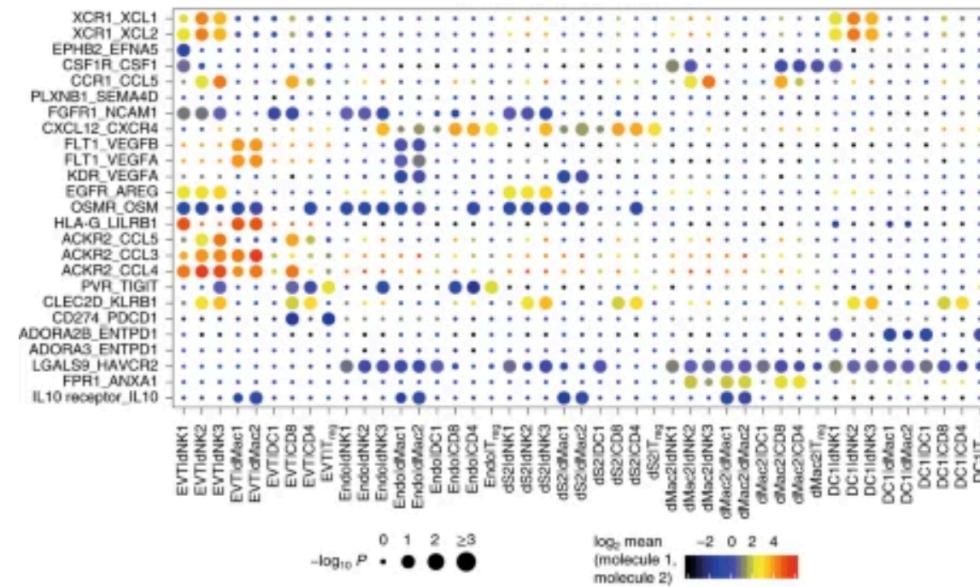
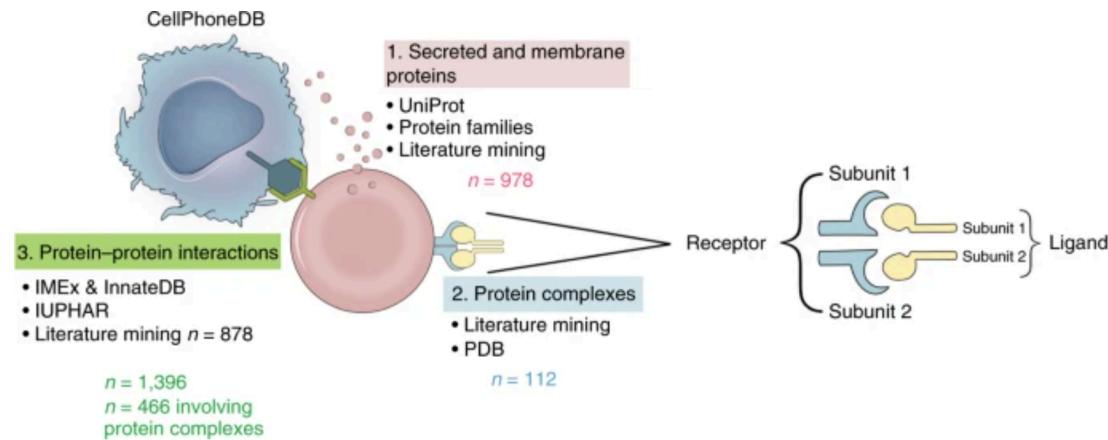


(Tirosh et al. *Science* 2016)

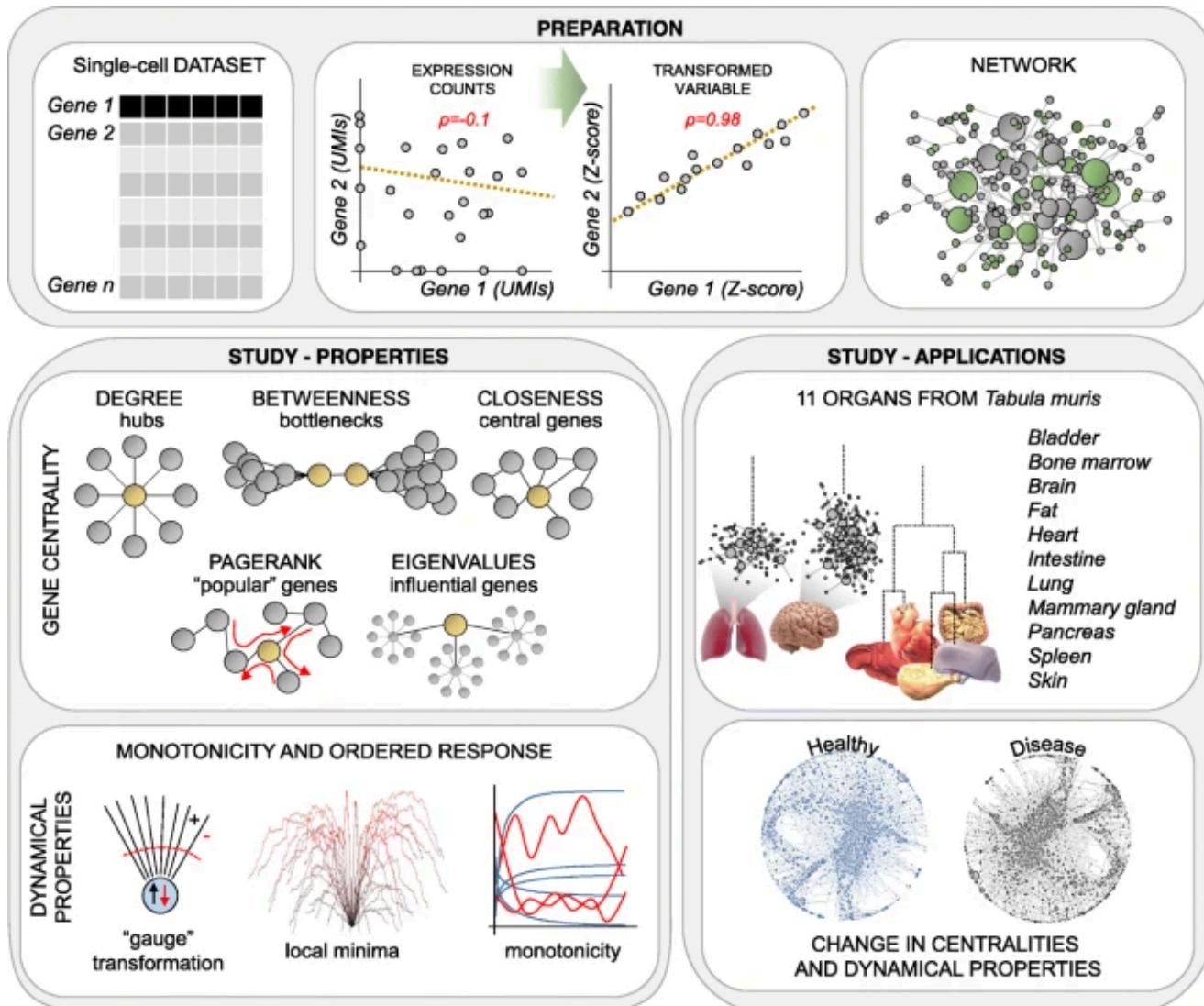
Allele and isoform information with SmartSeq3



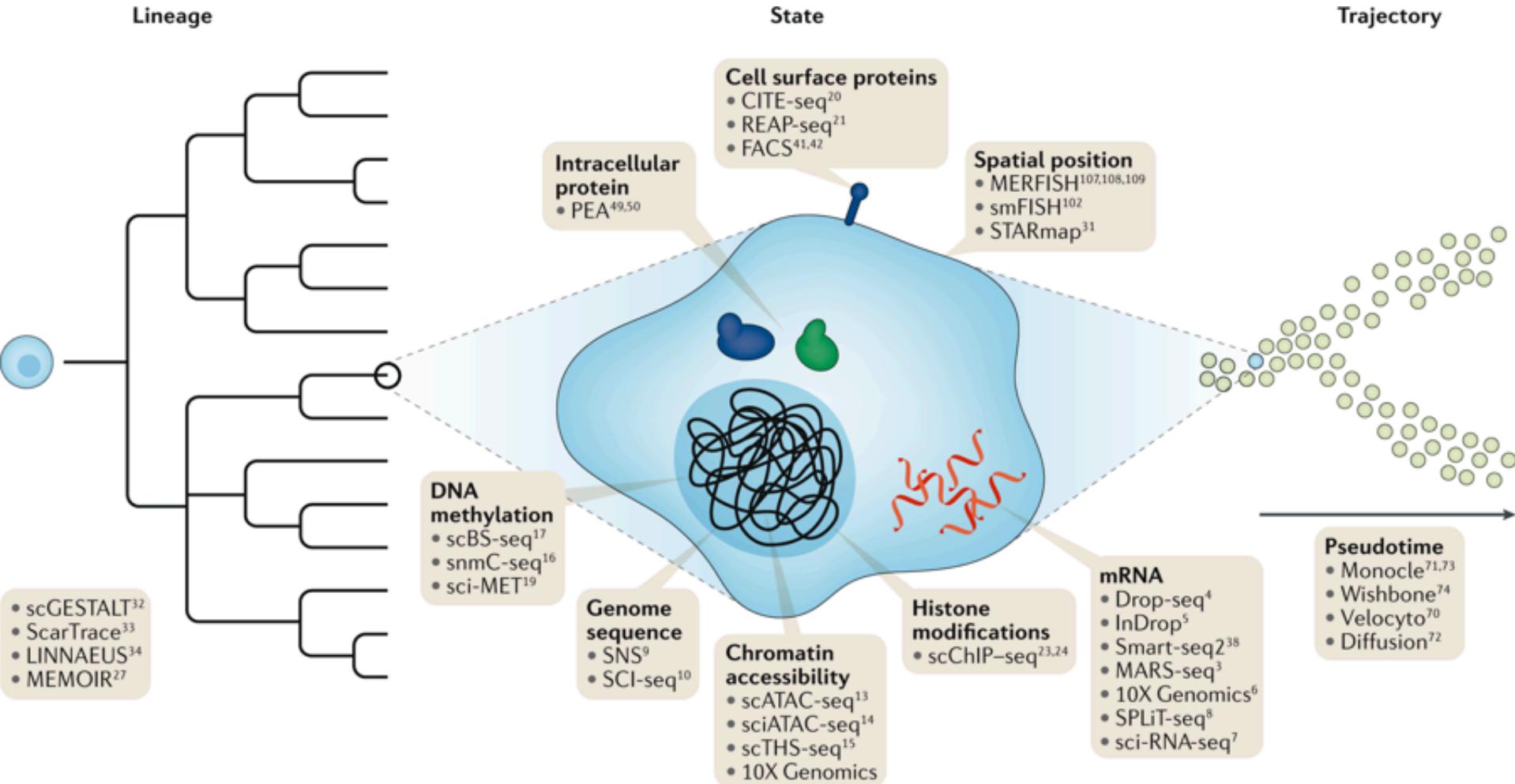
Receptor ligand interaction



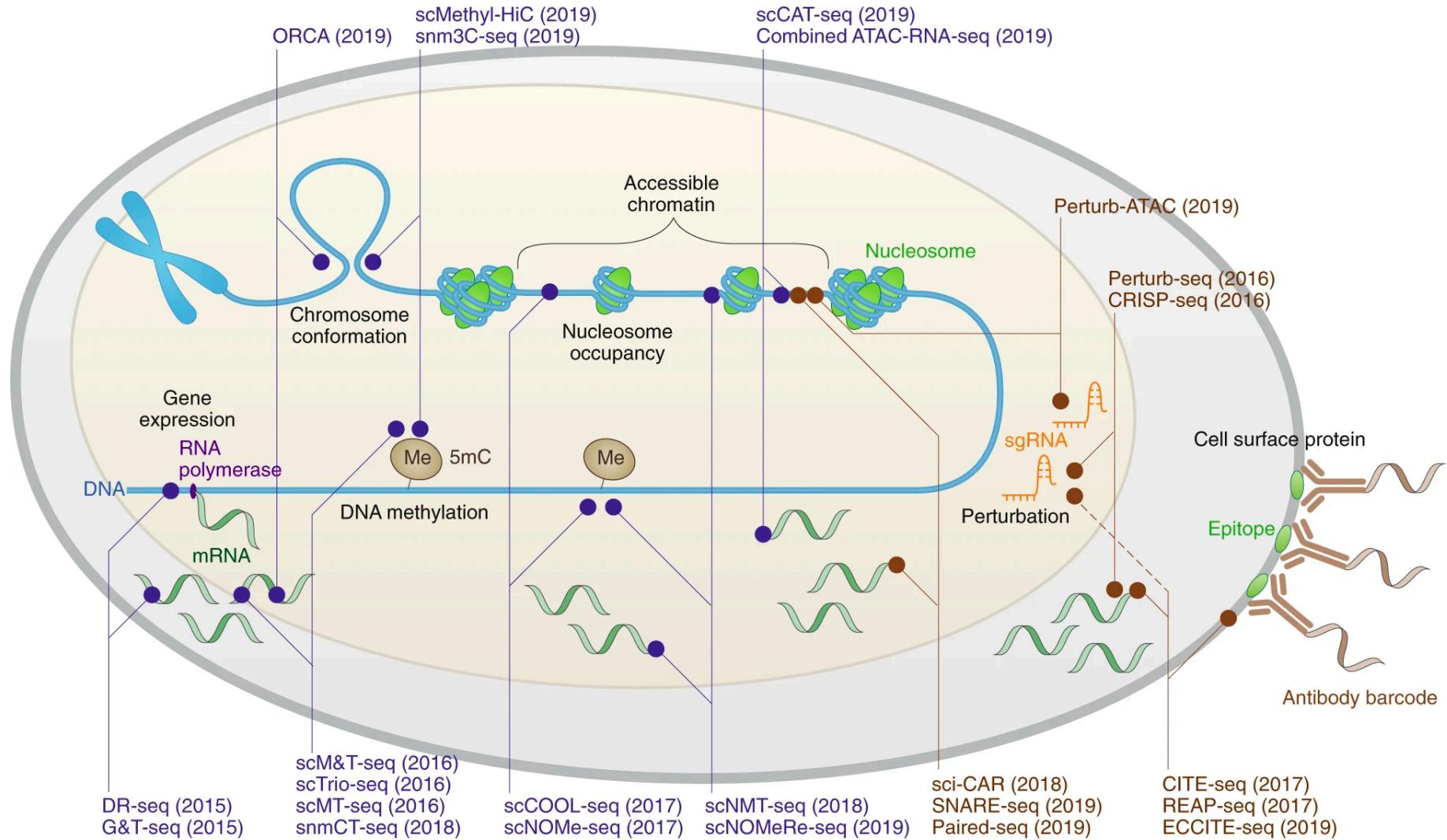
Gene regulatory networks



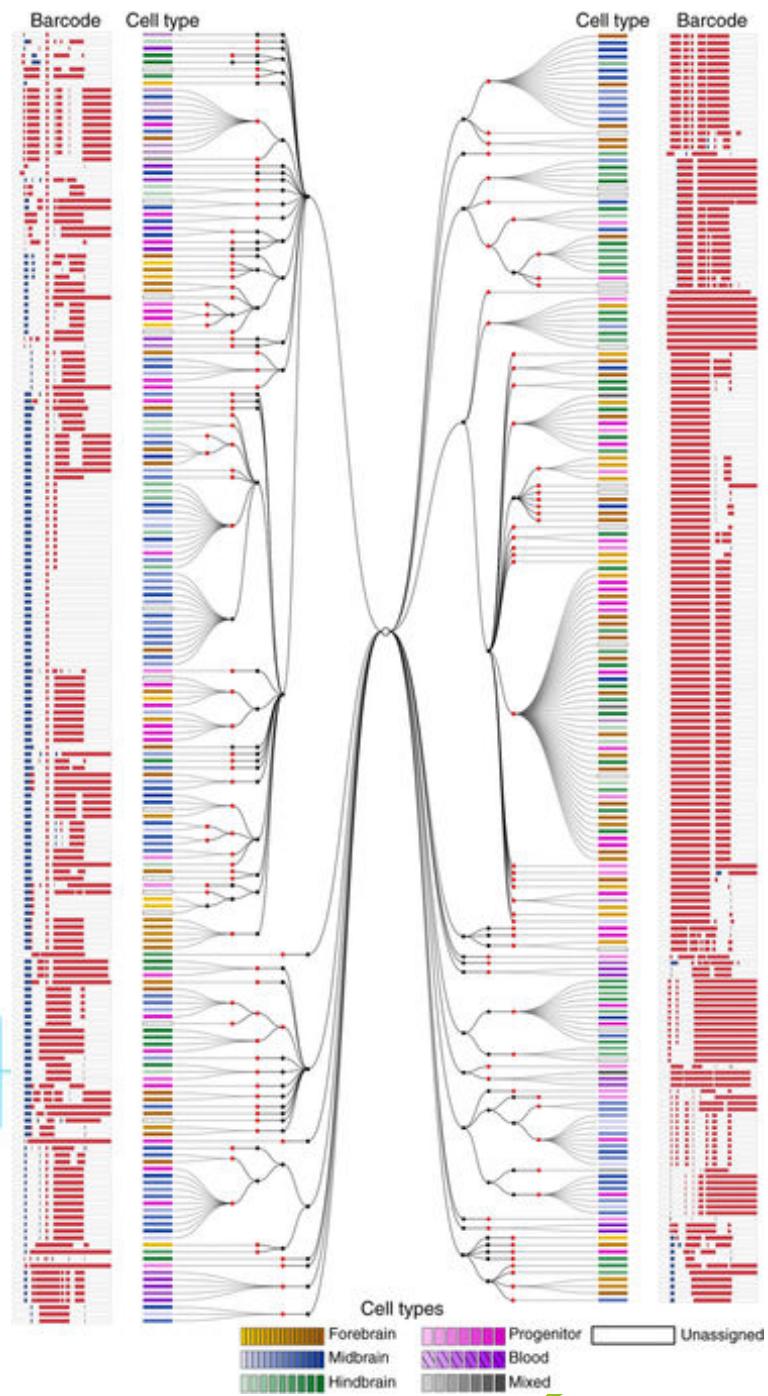
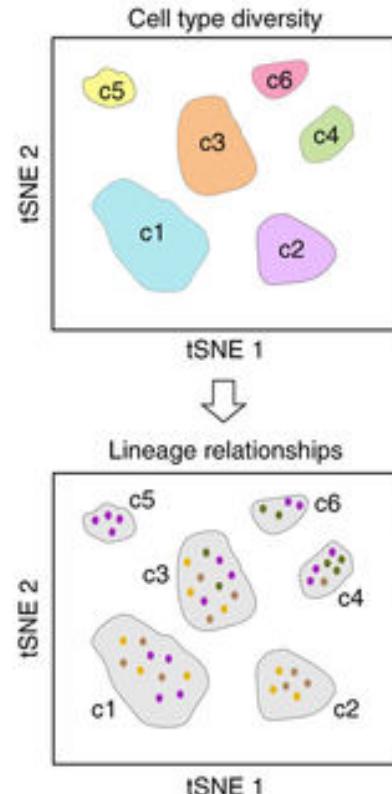
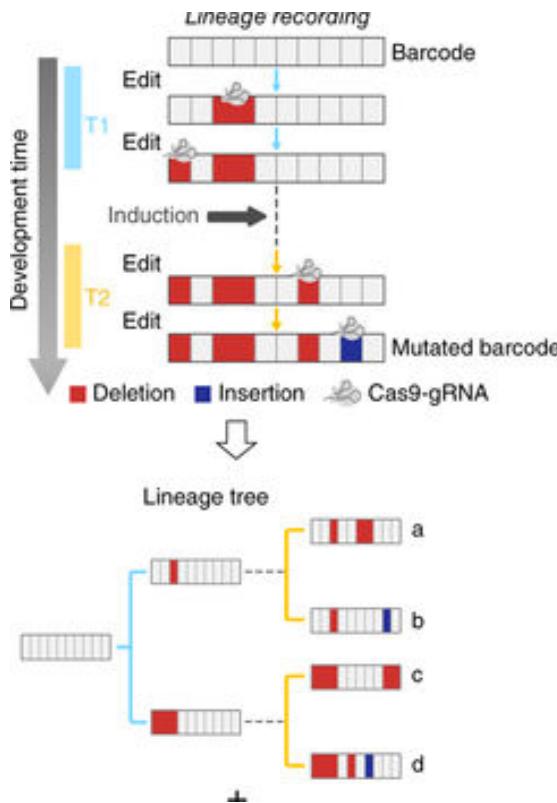
Single cell omics



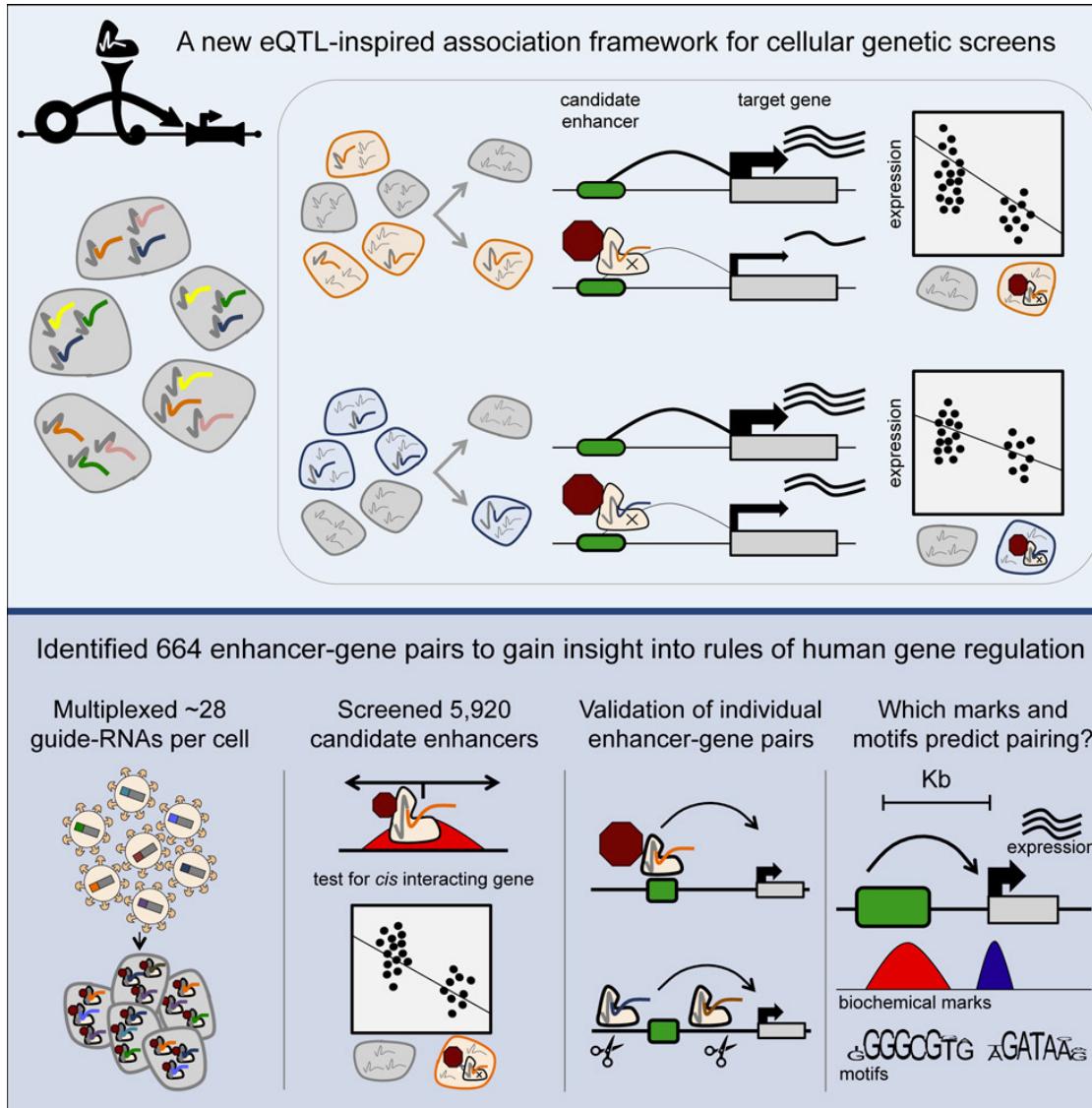
SC Multimodal omics



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



crisprQTL mapping for enhancer-gene pairs



Resources

- Good course at: <https://hemberg-lab.github.io/scRNA.seq.course/>
- Many of the packages have very thorough 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/>
- Conquer datasets - salmon pipeline to many different datasets: <http://imlspenticton.uzh.ch:3838/conquer/>
- 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 weekly/bi-weekly at several sites across Sweden
 - SciLifeLab Stockholm – Gamma 2, lunch room, Tuesdays 10.30
 - SciLifeLab Uppsala – Navet floor 3, Thursdays 10.00
 - Umeå, Linköping, Stockholm University, Lund and Göteborg as well.
- More info at: <http://nbis.se/>

Please fill in the Evaluation Form

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

Good luck with your analyses!