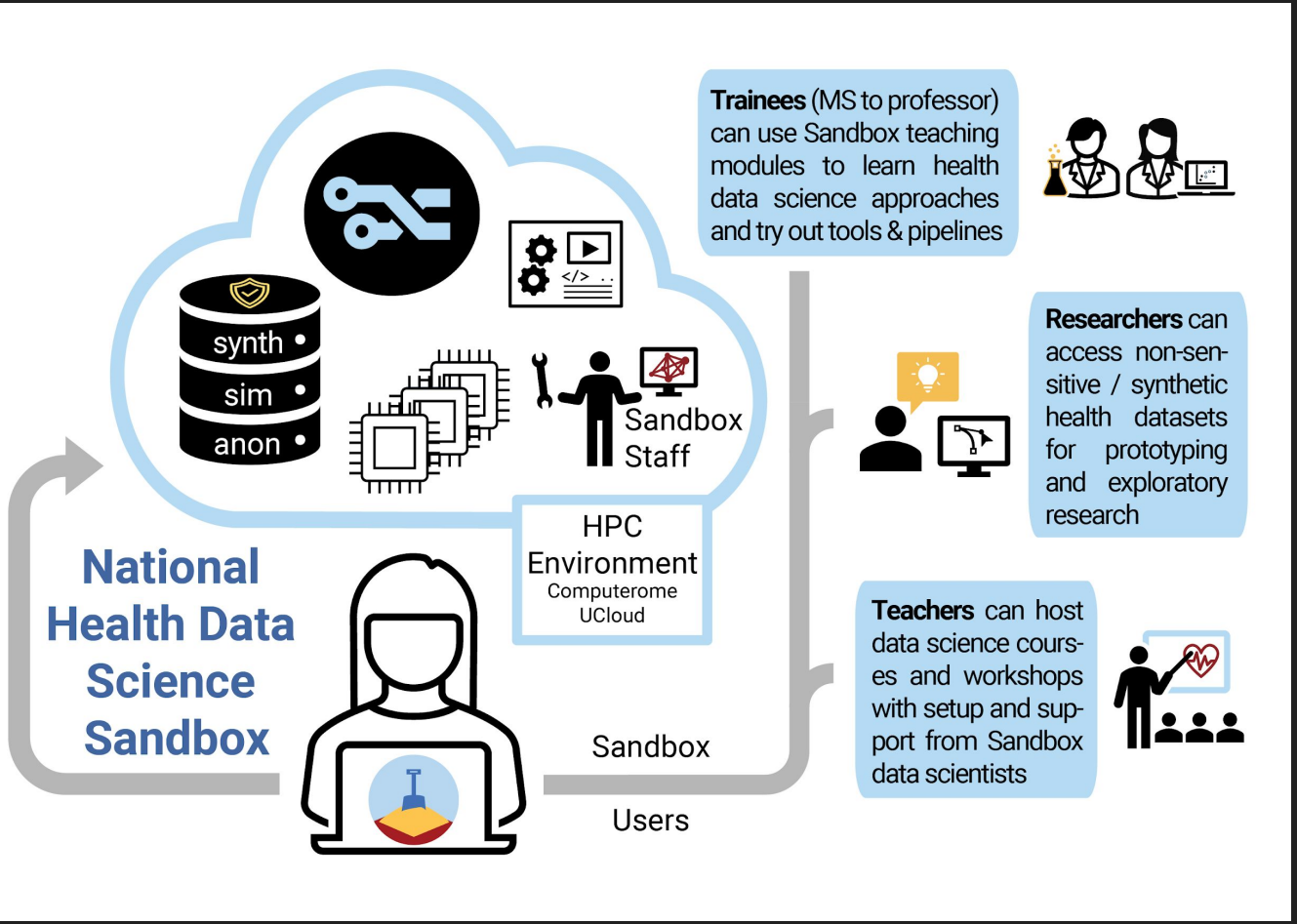
Transcriptomics

OMICS Workshop 30.august.2023

Samuele Soraggi
Health Data Science Sandbox

Health data science sandbox

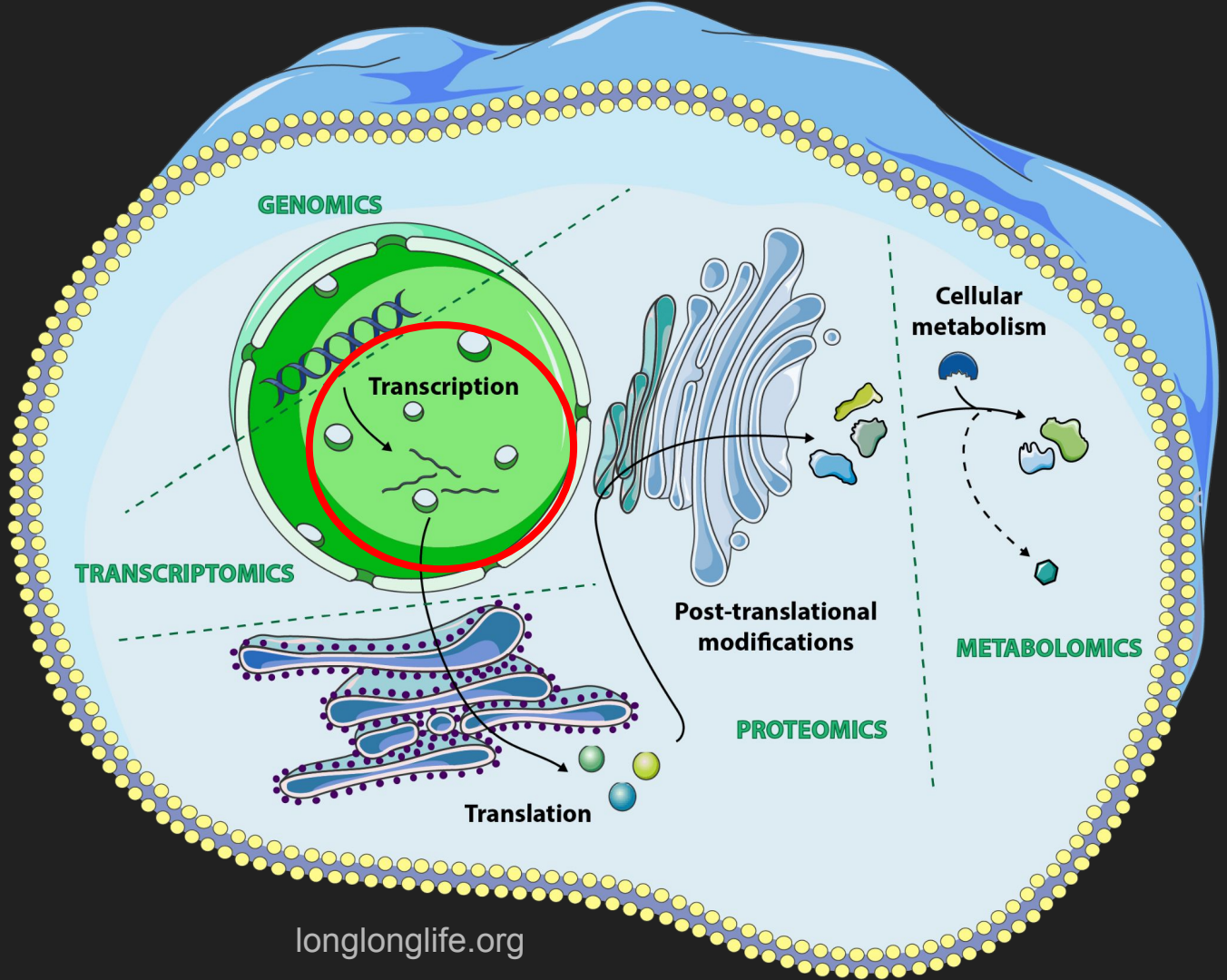


Home Page:

hds-sandbox.github.io

Contact:

samuele@birc.au.dk



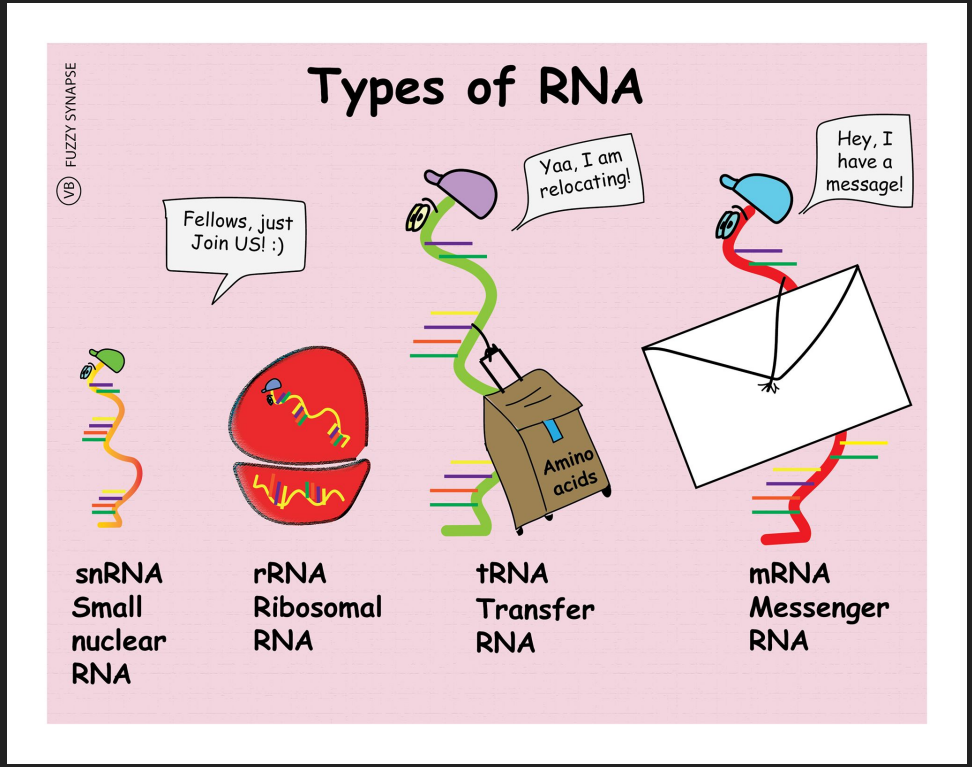
Program

9-9.30	Introduction to transcriptomics and tutorial format
9.30-9.45	Questions/Small break
9.45 - X	Log into uCloud, start the alignment part of the tutorial
X - X+15	Discussion and questions
X+15 - 13.00	Continuing with the variants analysis tutorial (If X small enough, we use the final time for discussion again)

Studying (mainly) the mRNA set at a specific point in time

tRNA and rRNA can be however sequenced and profiled if needed

lncRNA and snRNA are also of interest - some reference transcriptomes have many of them assembled

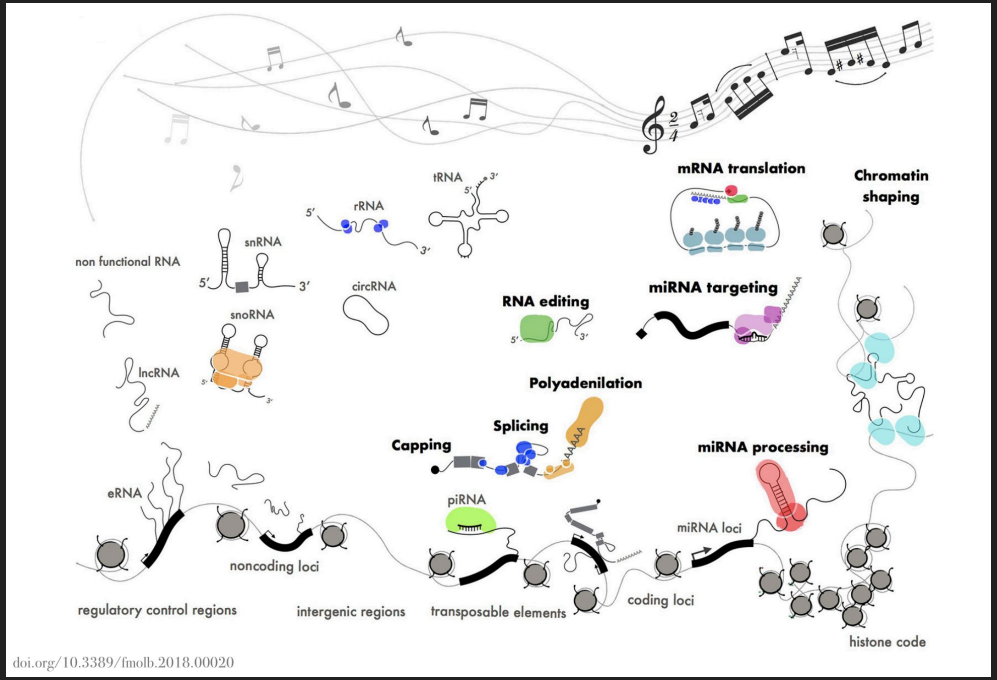


Transcriptomics

Studying (mainly) the mRNA set at a specific point in time

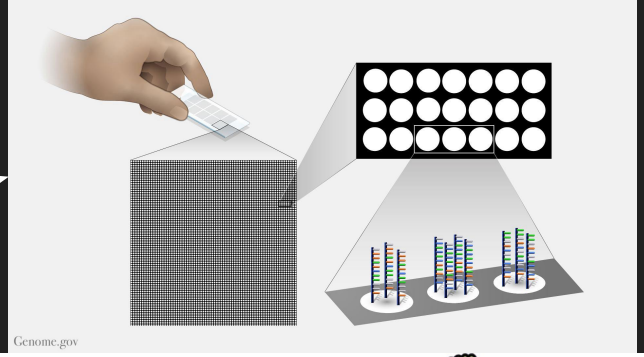
tRNA and rRNA can be however sequenced and profiled if needed

lncRNA and snRNA are also of interest - some reference transcriptomes have many of them assembled



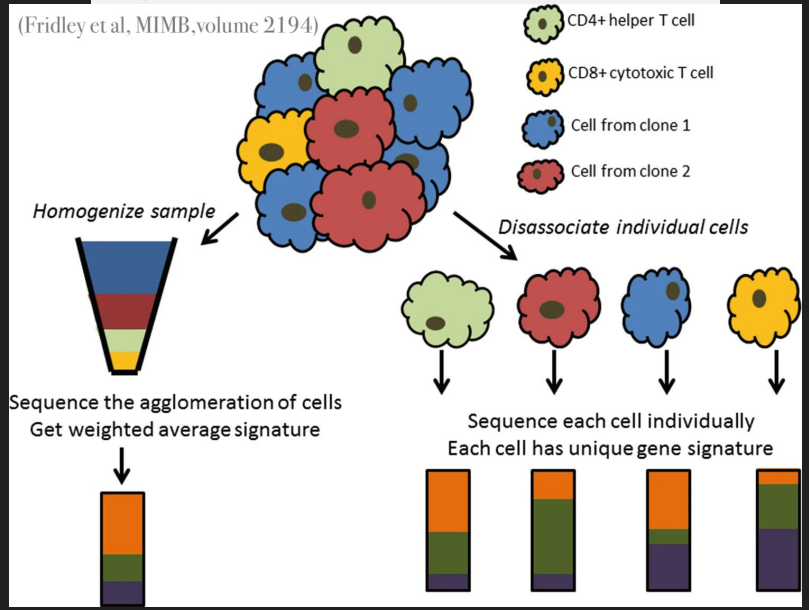
Types of transcriptomics Data

- Microarray



- Bulk RNA

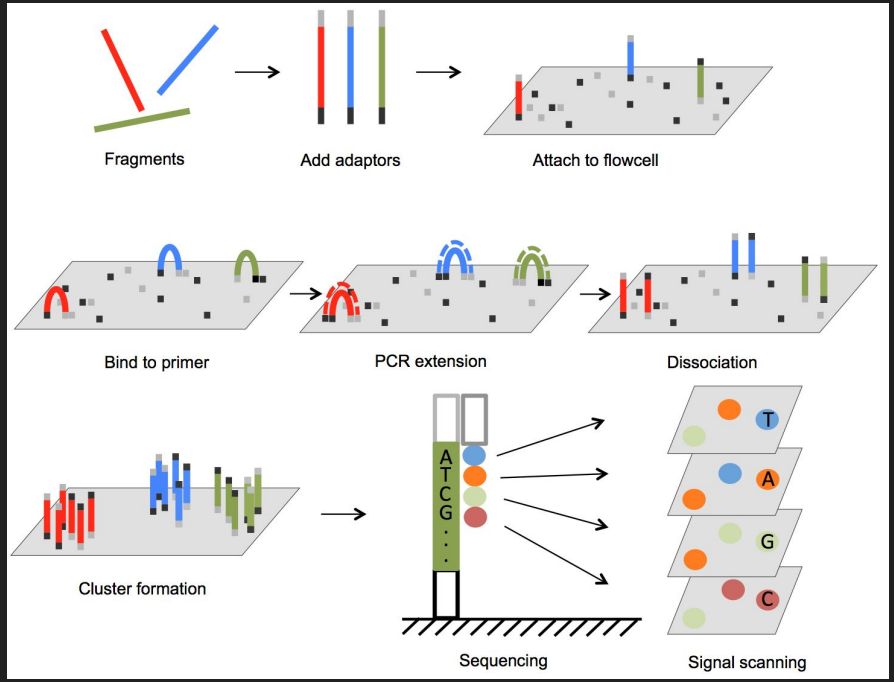
- single cell RNA (scRNA)



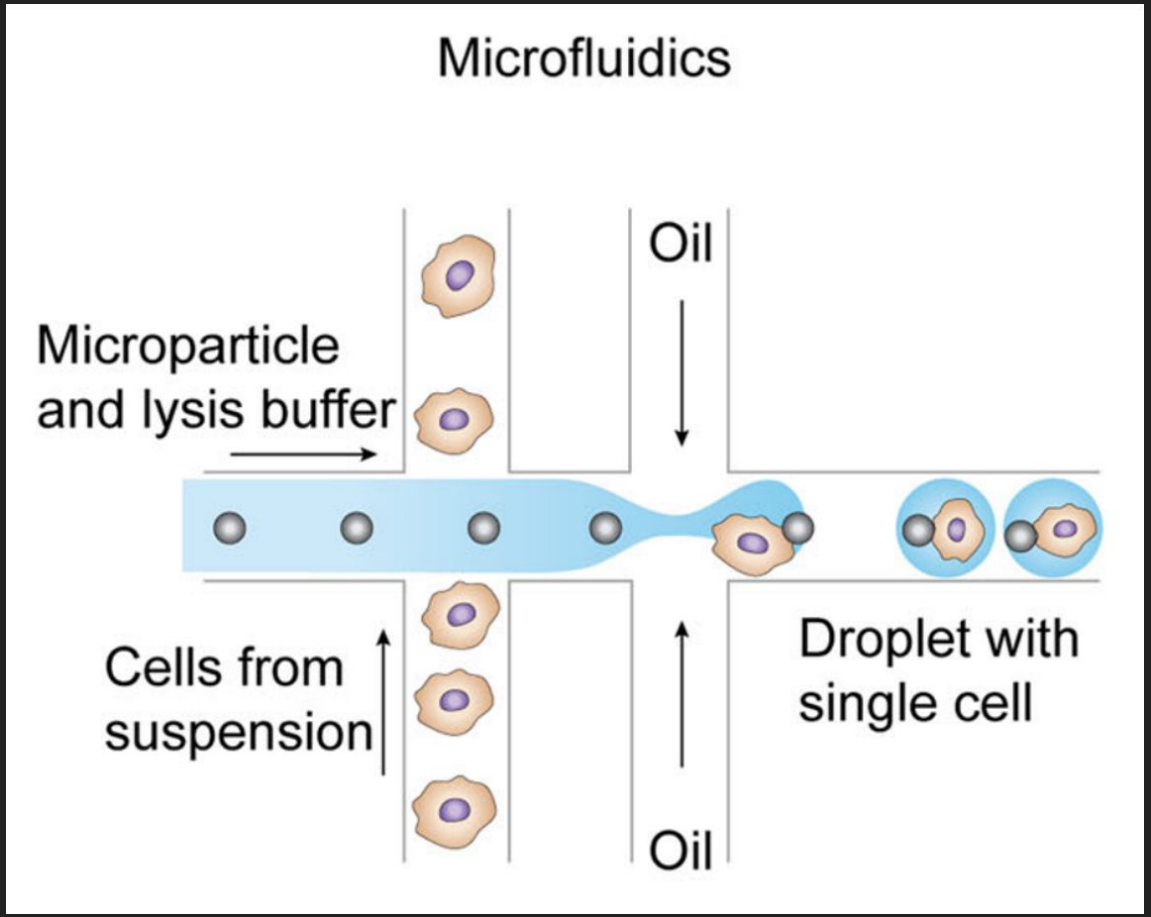
Types of transcriptomics Data

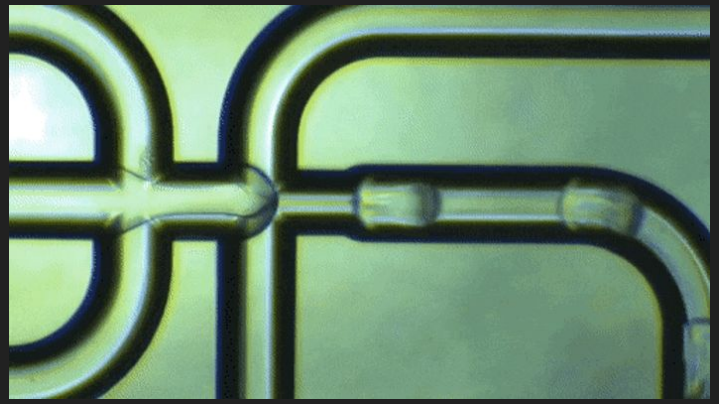
- Microarray
- Bulk RNA
- single cell RNA (scRNA)

Next Generation Sequencing (NGS) is essential for high throughput and high quality bulk and scRNA (*illumina* is the most used)



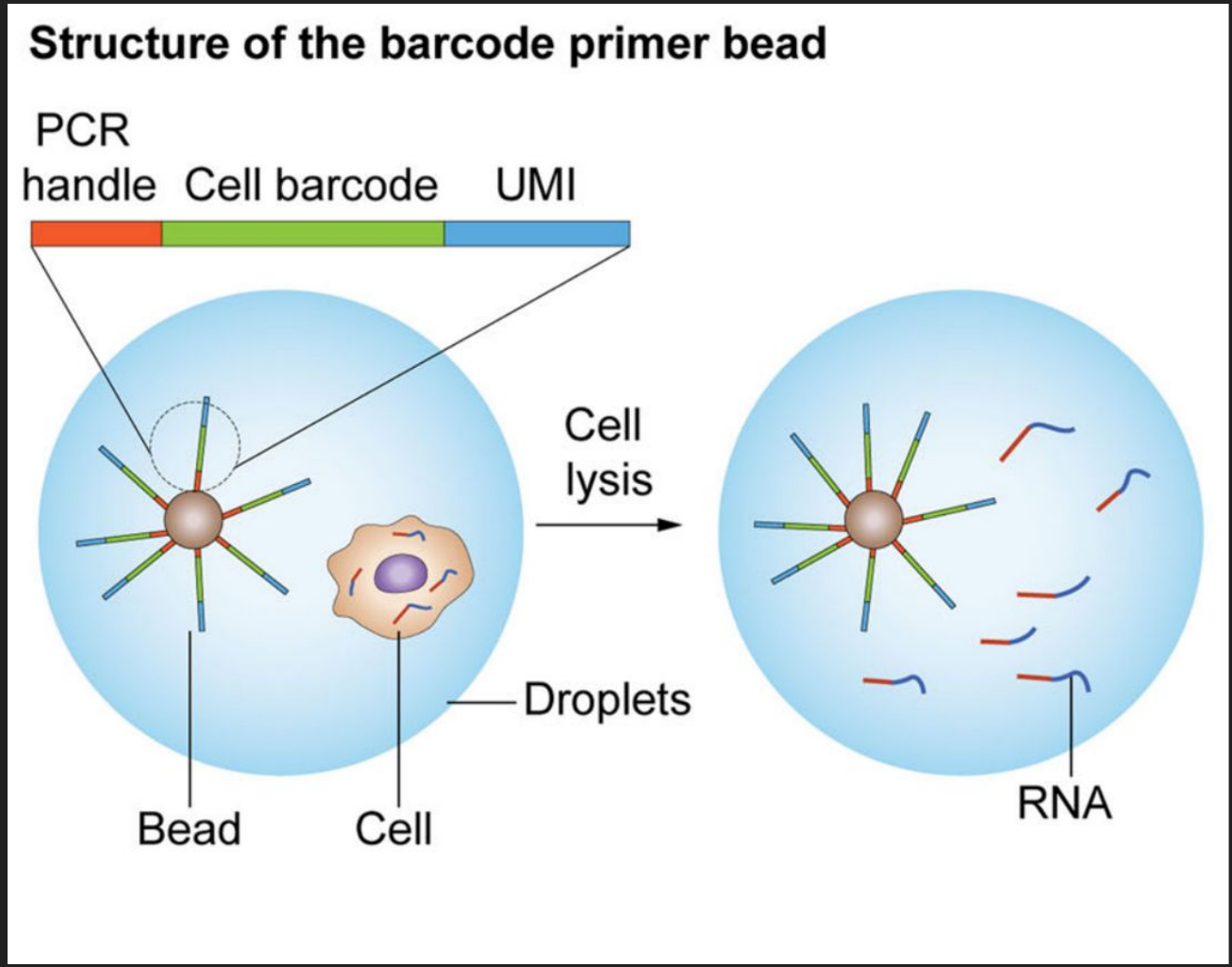
Data - single cell RNA (scRNA) sequencing
Cell isolation





Data - scRNA
Barcodes and UMIs

- Barcode: identifies the cell
- UMI: identifies the mRNA transcript



Data - scRNA
Preamplification and sequencing

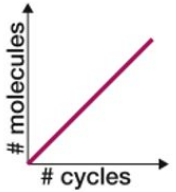
PCR

- exponential amplification
- PCR base specific biases

Tang protocol (Tang et al. 2009)
STRT (Islam et al. 2011)
SmartSeq/SmartSeq2 (Ramskold et al. 2012, Deng et al. 2014)

IVT

- linear amplification
- 3' bias due to two rounds of reverse transcription



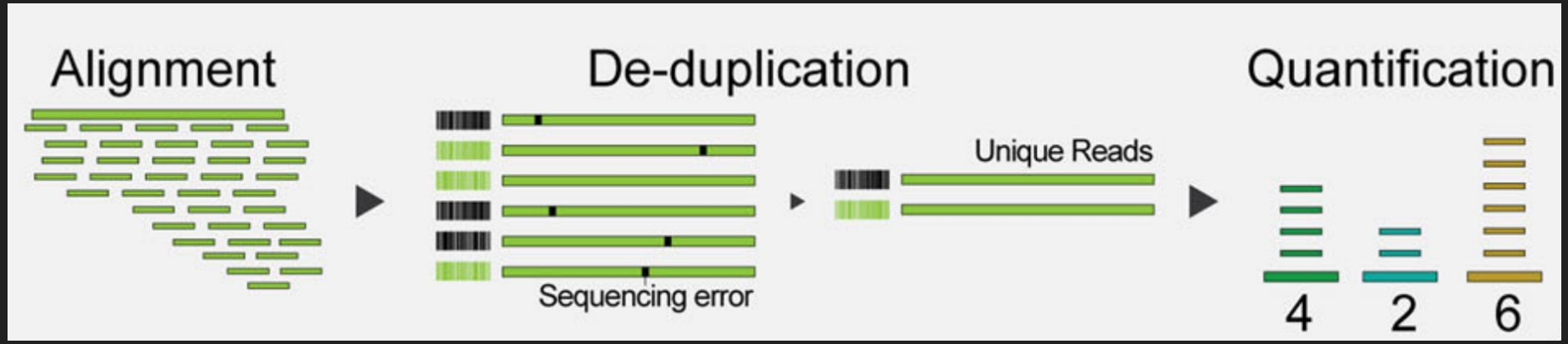
CELseq/MARSseq (Hashimony et al. 2013, Jaitin et al. 2014)

Illumina

AB SOLID

PacBio

Data - scRNA
Alignment



Align to a reference transcriptome

Remove errors by comparing reads with the same UMI

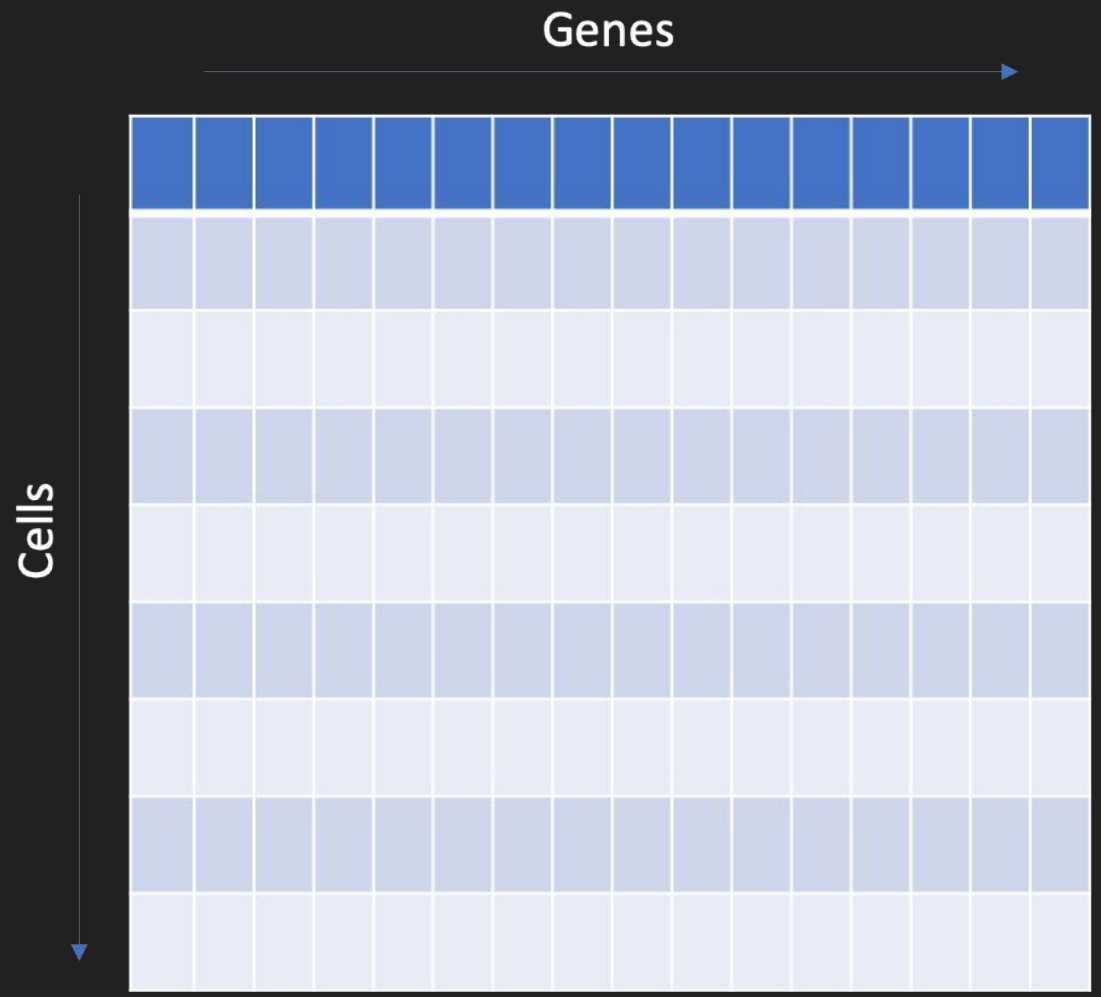
Consider reads with same UMI only once

Quantify with barcodes the gene abundances

Data - scRNA
Expression Matrix

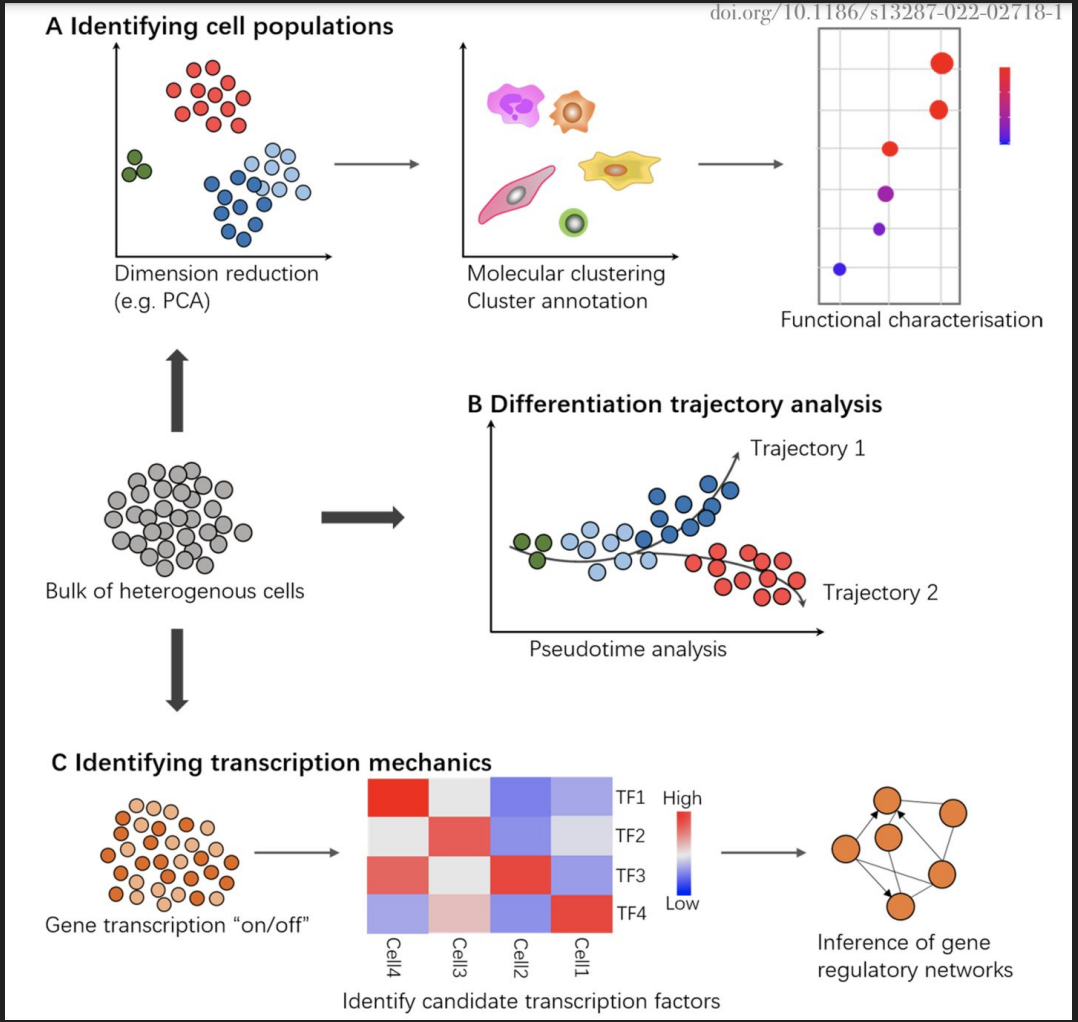
The final data is a cell**x**genes matrix which is

- large 1000s x 1000s in size
- noisy (seq errors, RNA contamination, 3' bias)
- sparse (90%-99% zeros)

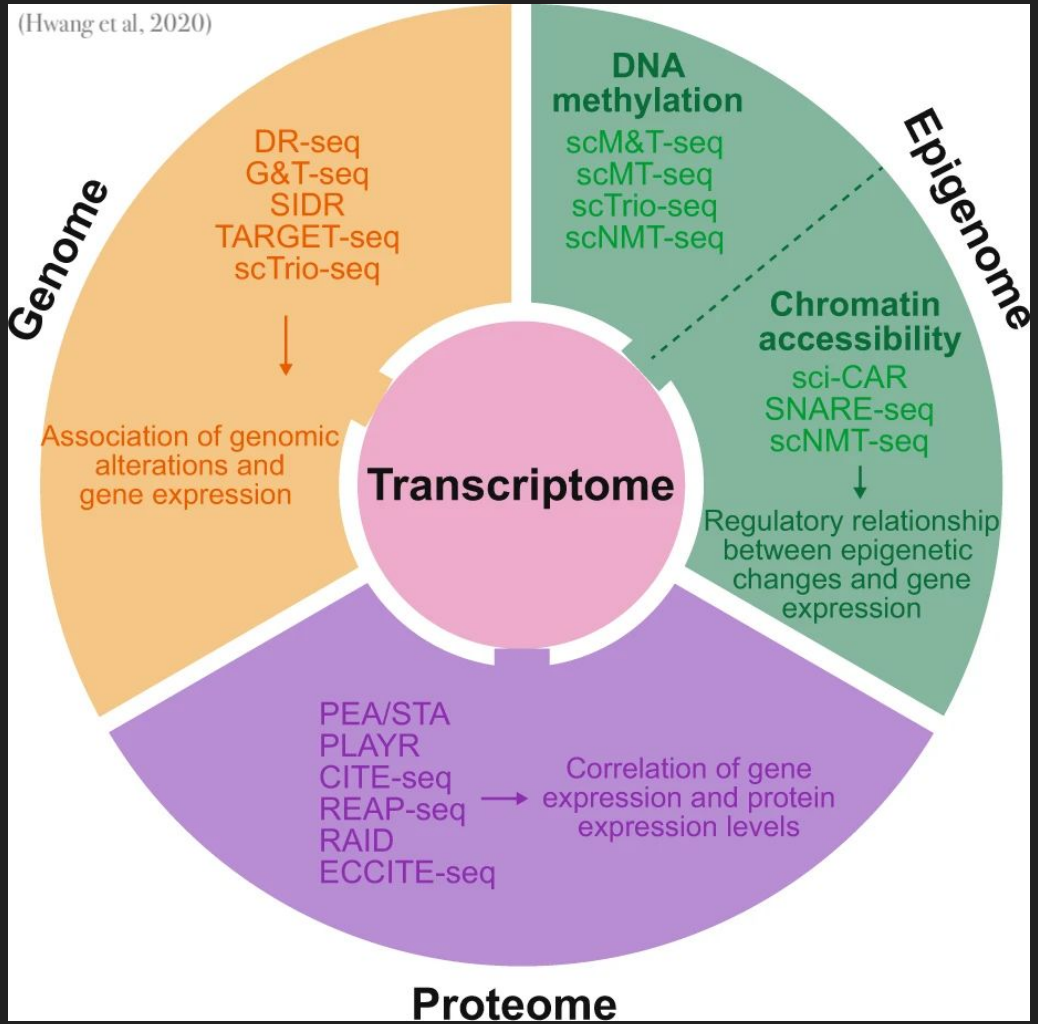


scRNA analysis - Applications

Those applications hold partially for bulk, exception for those at cell resolution



scRNA Data
Beyond mRNA

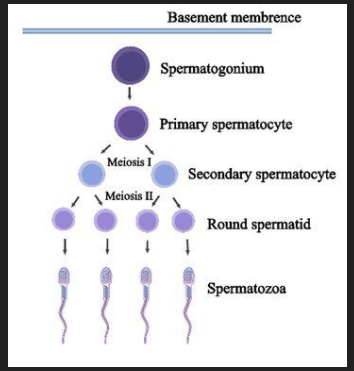


Tutorial

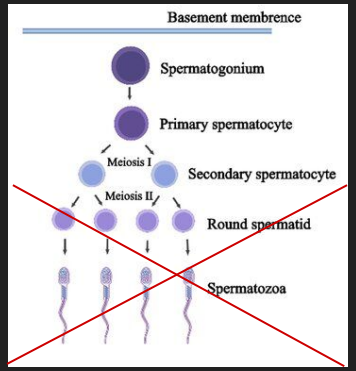
scRNA analysis - Tutorial

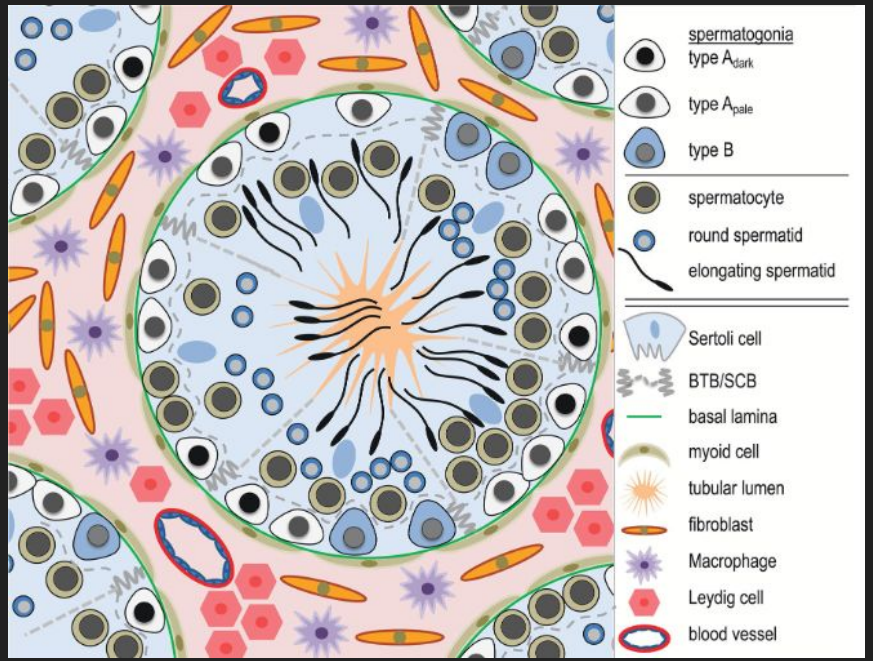
- Aligned matrices from 5 samples: 2fertile and 3 infertile men using droplet isolation and Illumina Sequencing (10X chromium technology)

- Filtering and doublets removal
- Normalization
- Clustering and subclustering
- Dimensionality reduction
- Differential Gene expression



VS

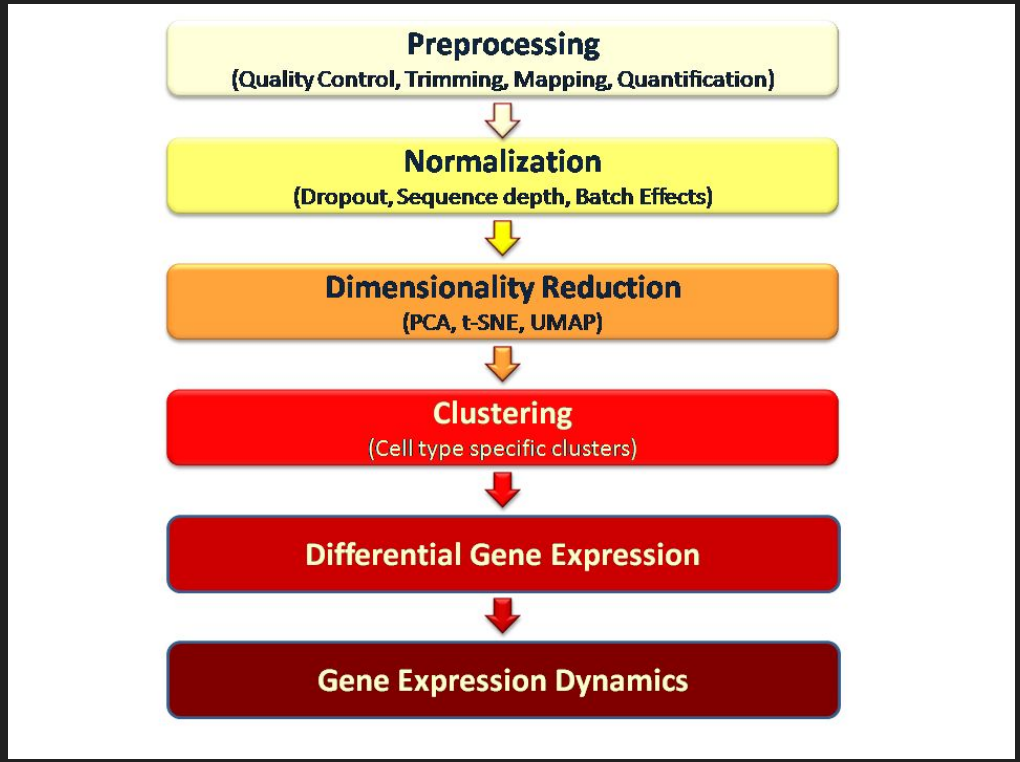




scRNA analysis - Tutorial

- Aligned matrices from 5 samples: 2fertile and 3 infertile men using droplet isolation and Illumina Sequencing (10X chromium technology)

- Filtering and doublets removal
- Normalization
- Dimensionality reduction
- Clustering and subclustering
- Differential Gene expression
- Go terms and other database access



- Go to https://hds-sandbox.github.io/OMICS-workshop/
- Follow the "**uCloud access**" instructions in the menu if it is your first access on uCloud
- Go on "**Day 2 - Transcriptomics**" to follow the tutorial instructions

More material:

- We have an introduction to bulkRNA sequencing course and a longer version of this tutorial at the **Transcriptomics Sandbox App on uCloud**
- **singlecellcourse.org** is a very good reference for single cell analysis in R
- The home page of the analysis tools **scanpy** and **seurat** (in python and R, respectively) include a lot of pedagogical tutorials
- At the Sandbox project we organize tutorials or courses in-person whenever we are able to. Keep an eye on our event list.