

# KEEP SCIENCE GOING

Let's make the best of your work-at-home time

<http://keepsceincegoing.org/tutorials.html>

**A Short Guide to dive into scRNA-seq analysis**

Arvind Iyer

Computational Systems Oncology  
Department of Computational Biology.

The logo for the University of Lausanne (UNIL), featuring the word "Unil" in a white, cursive script font.

UNIL | Université de Lausanne

# WHAT WE GONNA DO!!

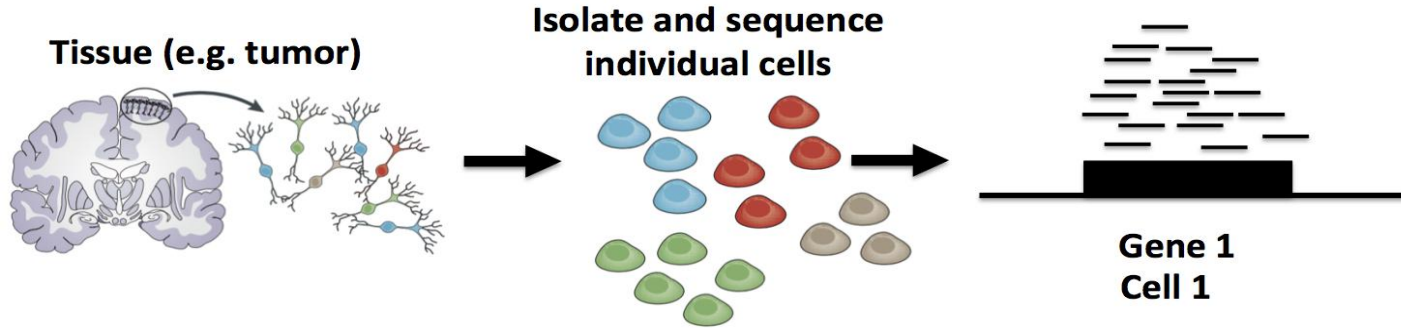
- Motivation
- Prerequisites
- Basic computational pipeline of scRNA-seq analysis
- Environment Setup
- Case Study with a lung cancer dataset.
- Courses/Materials to follow
- Get in touch



# PREREQUISITES

- Basic Python programming
- Basic shell scripting
  
- **Target Audience**
  - Looking to Explore computational scRNA seq analysis
  - No idea about scRNA seq.

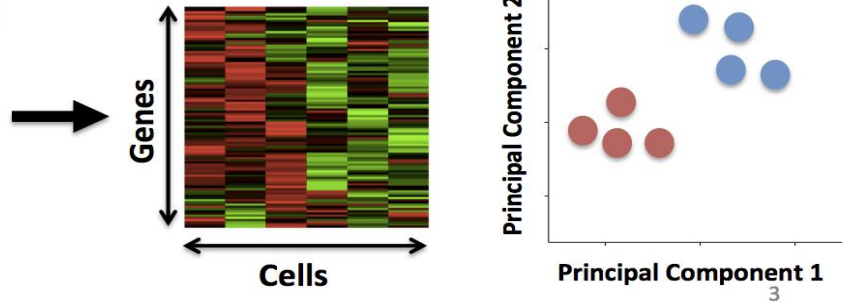
# Single-cell RNA-Seq (scRNA-Seq)

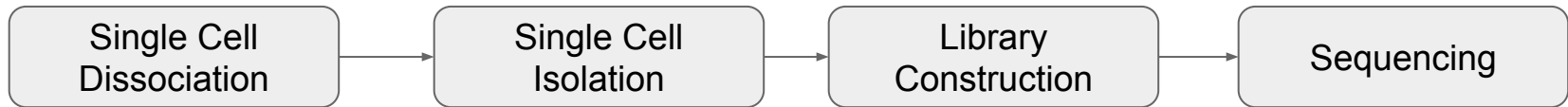
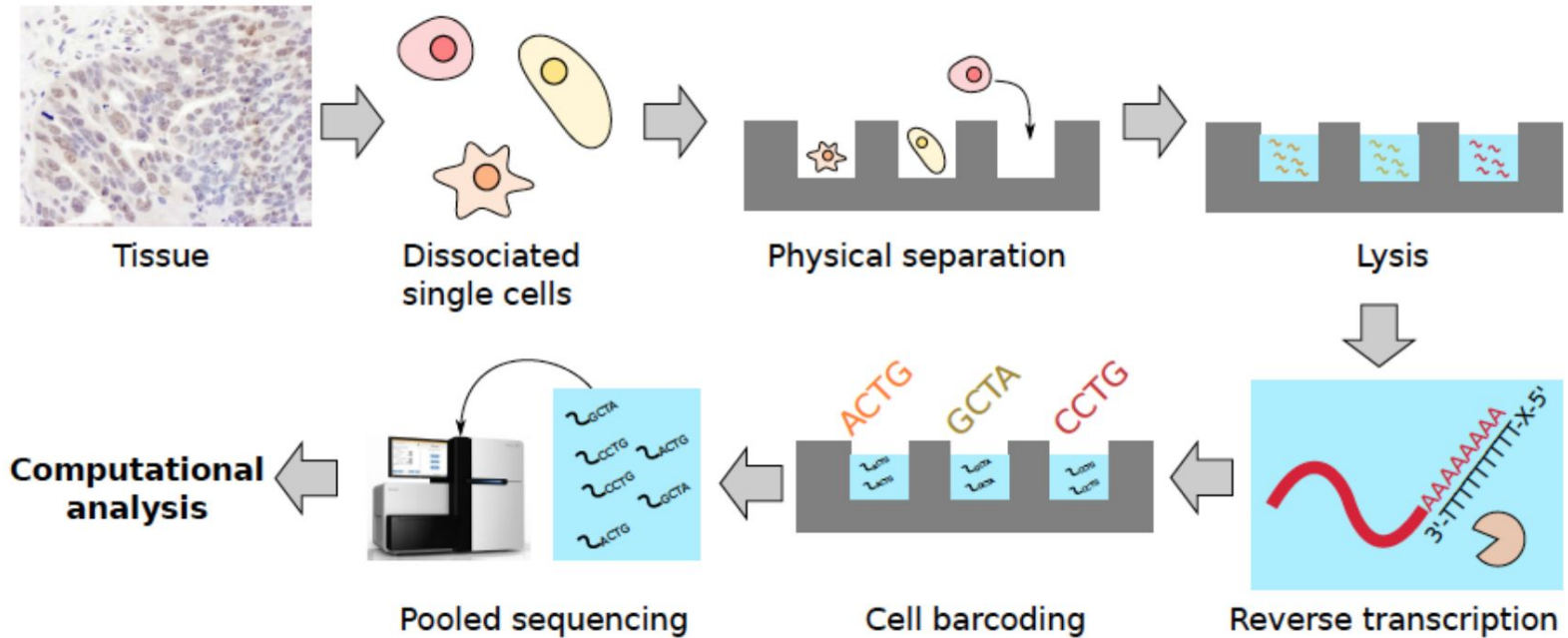


**Read Counts**

	Cell 1	Cell 2	...
Gene 1	18	0	
Gene 2	1010	506	
Gene 3	0	49	
Gene 4	22	0	
...			

**Compare gene expression profiles of single cells**





Taken From:

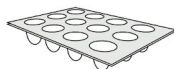
[https://bioinformatics-core-shared-training.github.io/cruk-summer-school-2018/SingleCell/slides/2018-07-25\\_CRUK\\_CI\\_summer\\_school-scRNAseq.pdf](https://bioinformatics-core-shared-training.github.io/cruk-summer-school-2018/SingleCell/slides/2018-07-25_CRUK_CI_summer_school-scRNAseq.pdf)

Manual



Tang et al 2009

Multiplexing



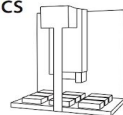
Islam et al 2011

Integrated Fluidic Circuits



Brennecke et al 2013

Liquid Handling Robotics



Jaitin et al 2014

Nanodroplets



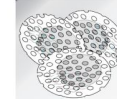
Klein et al 2015  
Macosko et al 2015

Picowells



Bose et al 2015

In situ barcoding



Cao et al 2017  
Rosenberg et al 2017

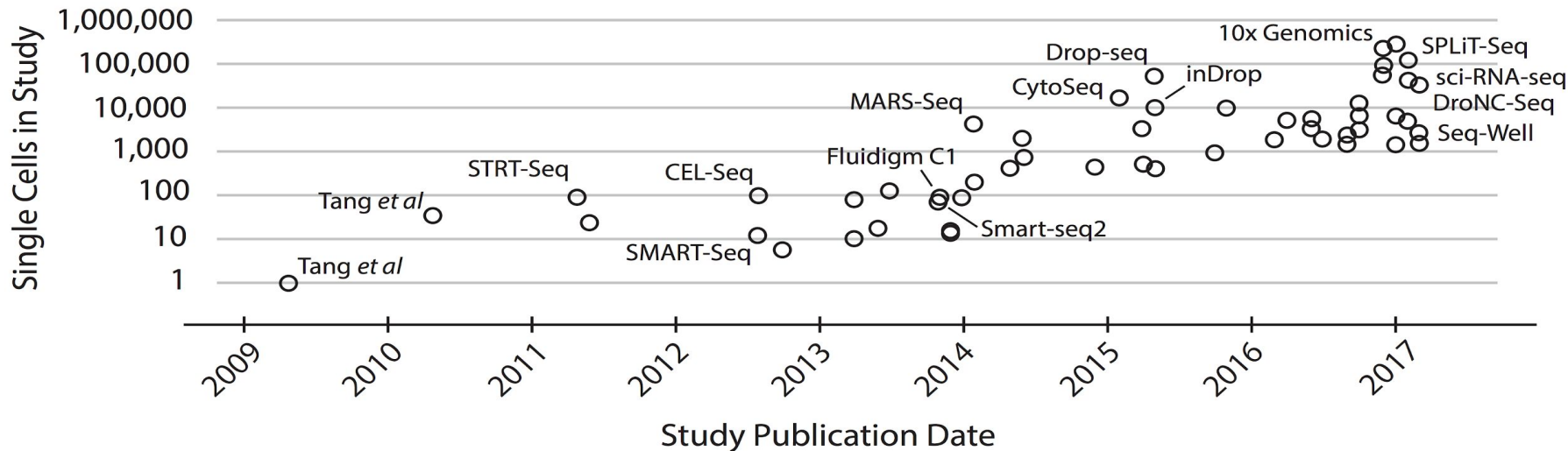
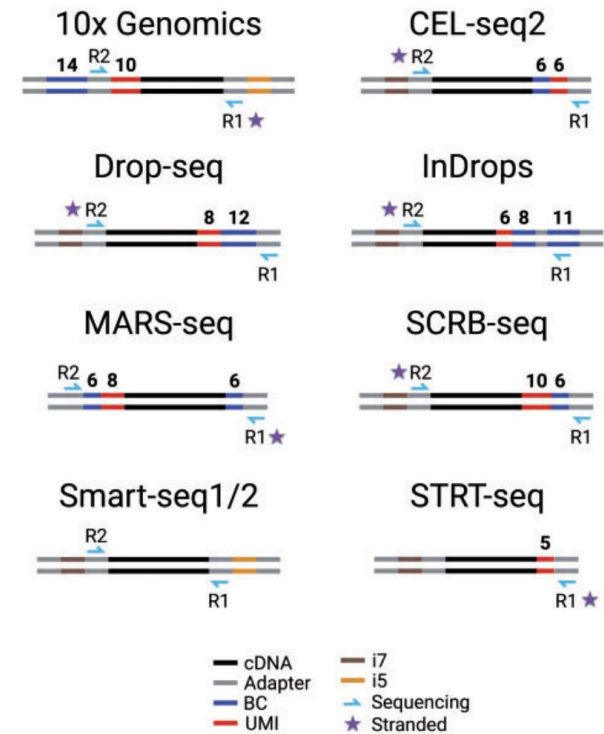
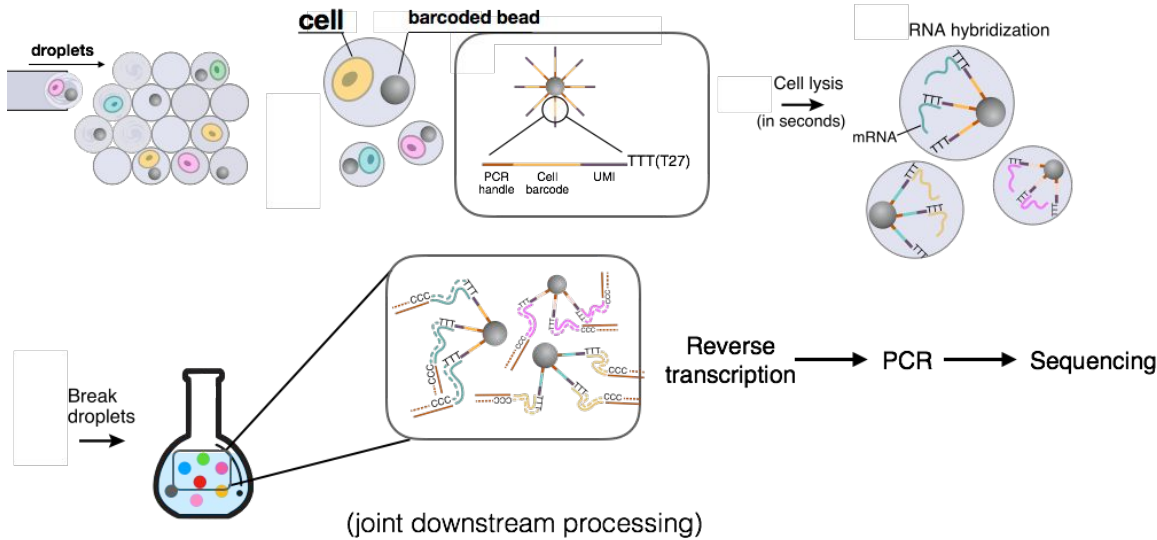


Plate Based  
Smart-Seq 2, CEL-seq,  
Mars-seq

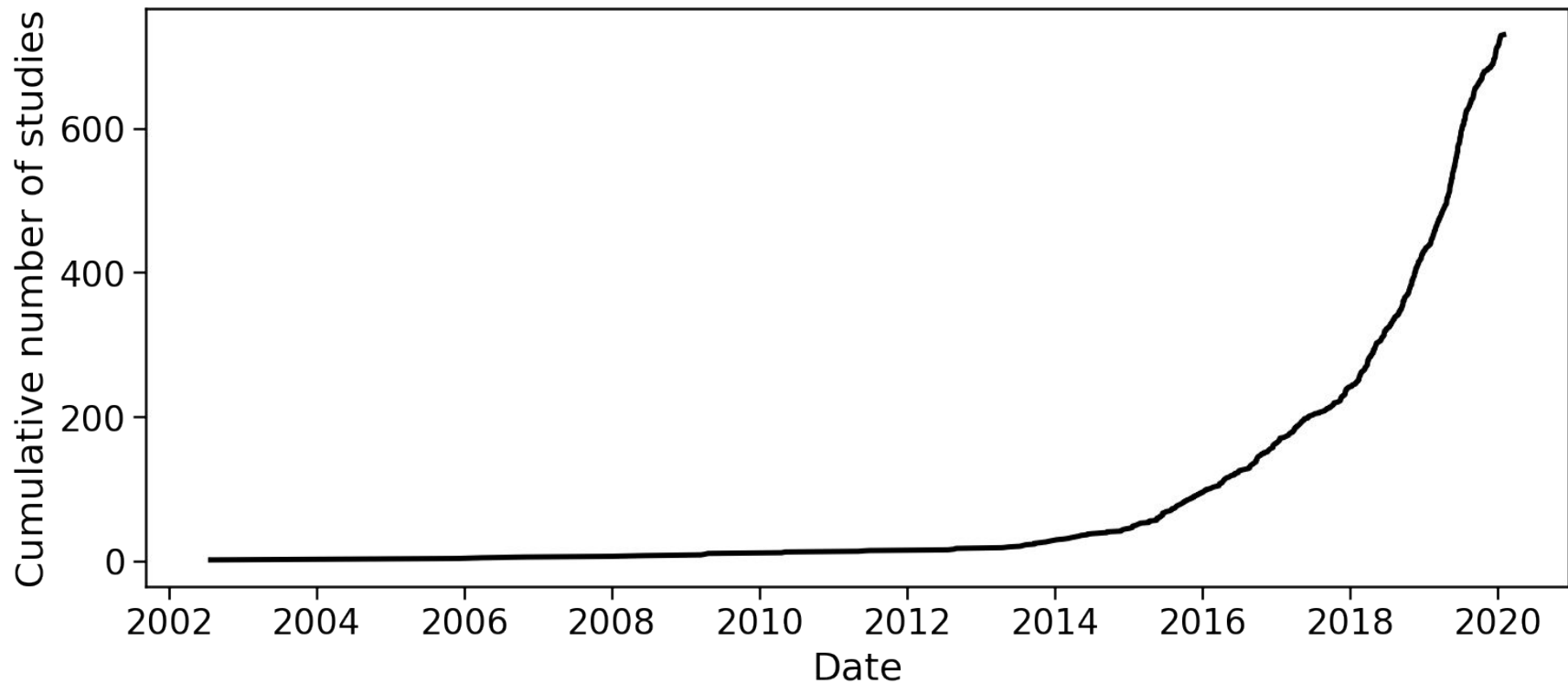
Droplet Based  
10x, Drop-Seq,  
inDrops etc

# DROP-SEQ



- Drop-Seq aims to trace molecules back to each cell, and this is achieved by using a 12bp cell barcode to distinguish different cells and an 8bp UMI to differentiate molecules within the same cell.
- UMIs allow us to distinguish between amplified copies of the same mRNA molecule and reads from separate mRNA molecules transcribed from the same gene.



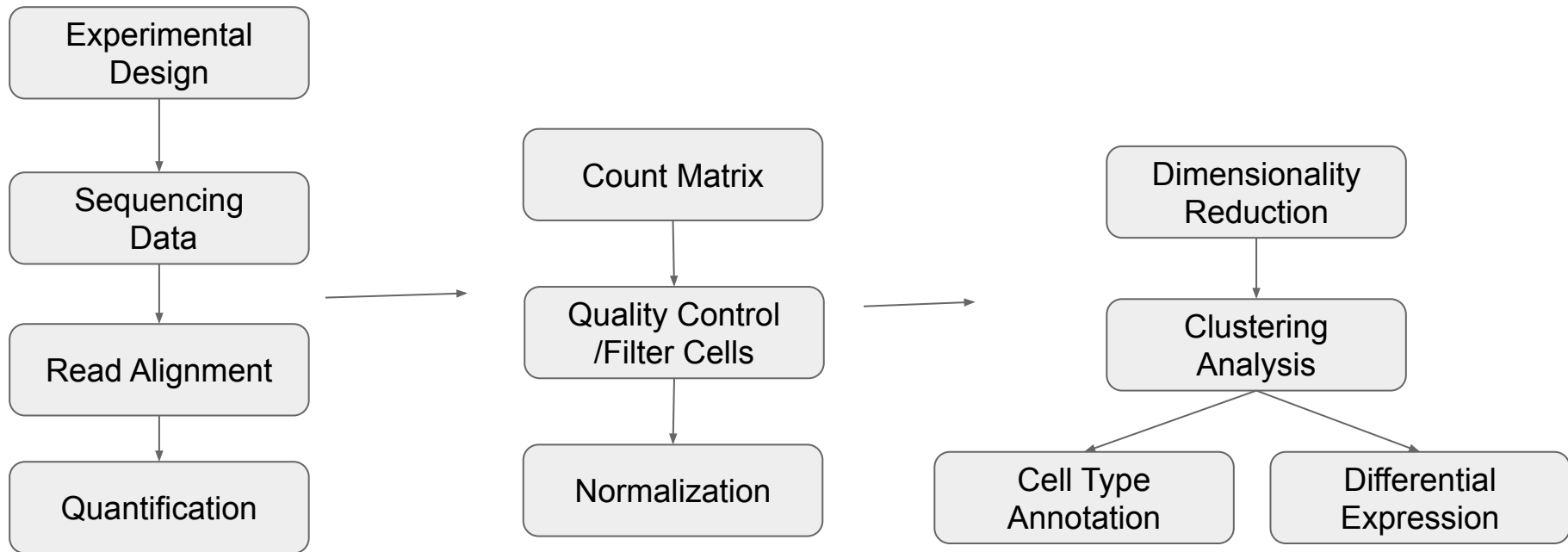


Taken: [https://github.com/pachterlab/kallistobustools/blob/master/notebooks/Introduction\\_single\\_cell\\_RNA\\_seq.ipynb](https://github.com/pachterlab/kallistobustools/blob/master/notebooks/Introduction_single_cell_RNA_seq.ipynb)

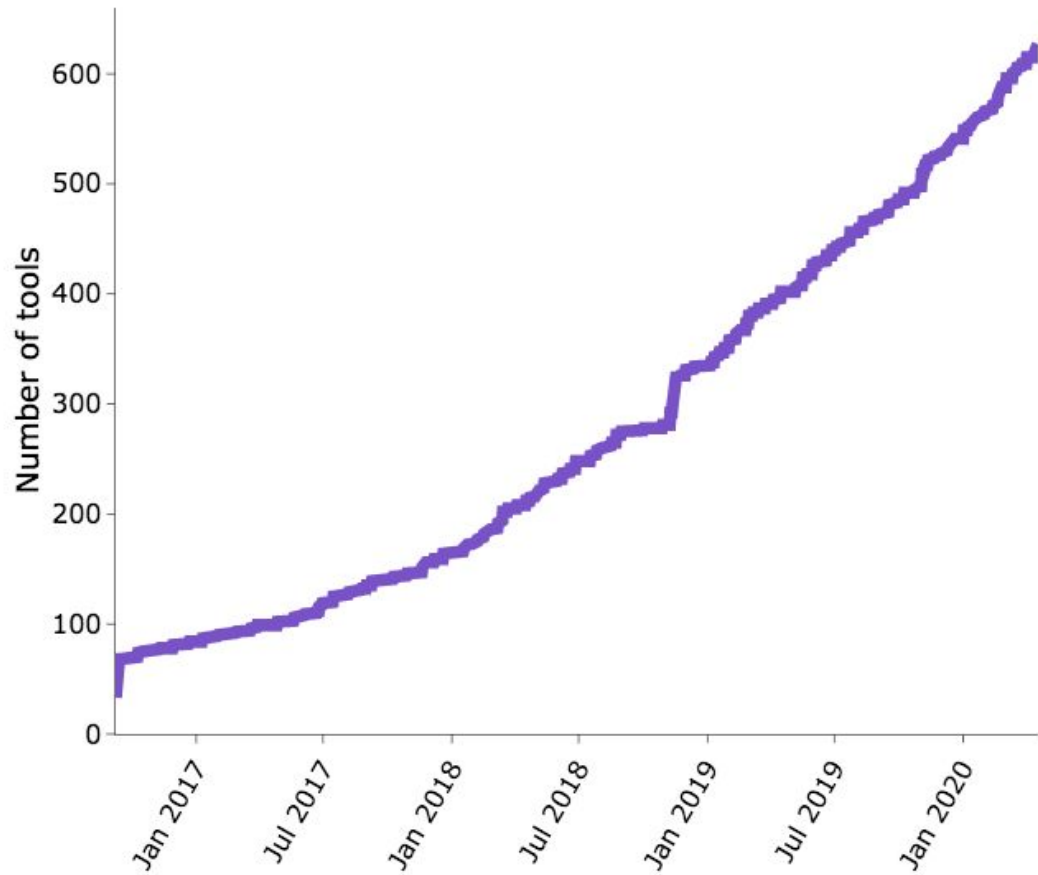
# TRANSCRIPTOMICS AT A GLANCE!!

High dimensional snapshot of transcriptomic activity of all RNA species in a sample.

**Data: Genes \* Samples (Single Cells)**



- Experimental Design & Methods matter a lot!!
- There are multiple tools/software/algorithm at each step.
- There are different protocols.

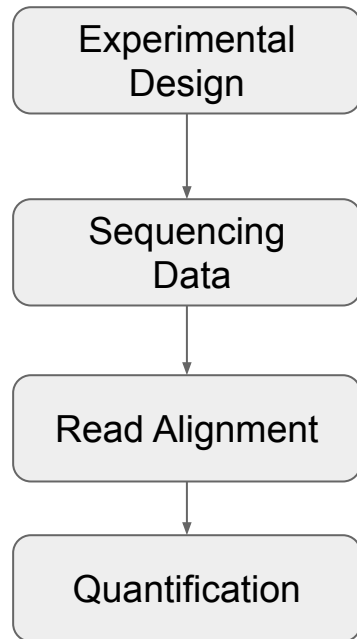


Taken from: <https://www.scrna-tools.org/>

HANDS ON SESSION

# LET SET UP THE ENVIRONMENT

- We gone use Python for handson session
- We will be going through downloading the packages/tools one after the other.
- We will do an analysis on Lung Cancer Dataset.
- It will be full on computational from here on.



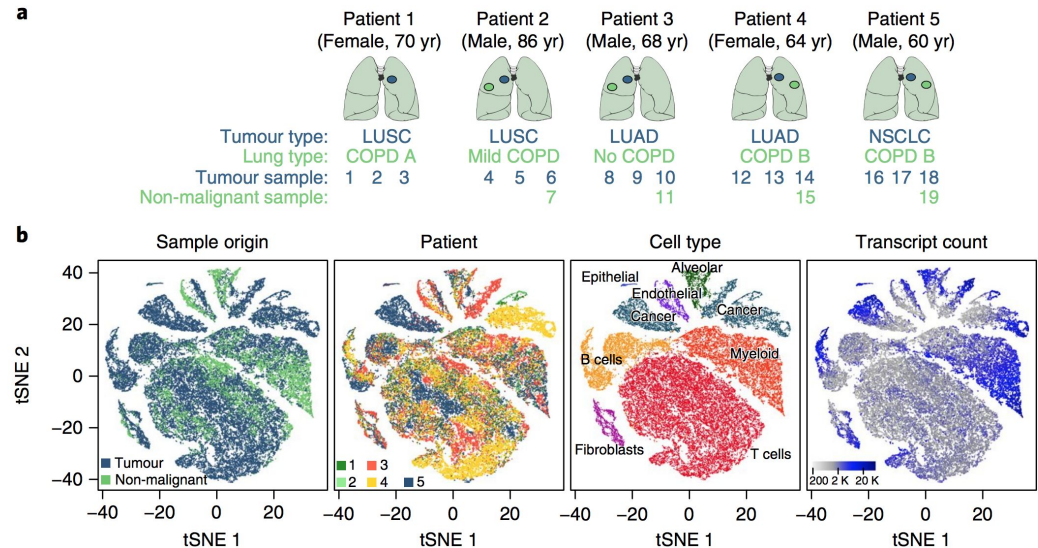
# DATA

- Cancer cells are embedded in the tumor microenvironment (TME)

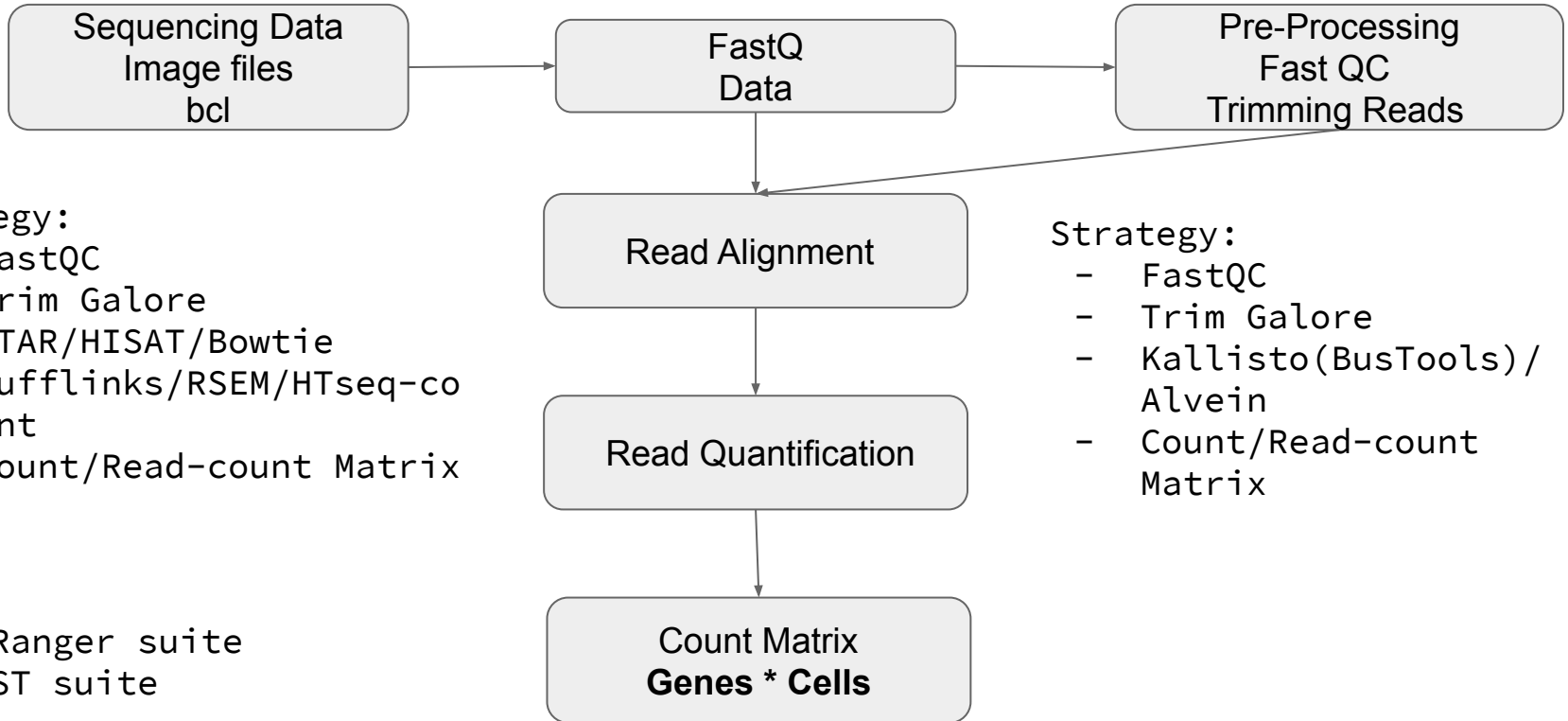
- The paper presents a 52,698-cell catalog of the TME transcriptome in human lung tumors at single-cell resolution.

## Phenotype molding of stromal cells in the lung tumor microenvironment

Diether Lambrechts<sup>1,2\*</sup>, Els Wauters<sup>3,4</sup>, Bram Boeckx<sup>1,2</sup>, Sara Aibar<sup>5,6</sup>, David Nittner<sup>7,8</sup>, Oliver Burton<sup>6,9</sup>, Ayse Bassez<sup>1,2</sup>, Herbert Decaluwé<sup>10,11</sup>, Andreas Pircher<sup>1,12</sup>, Kathleen Van den Eynde<sup>13</sup>, Birgit Weynand<sup>13</sup>, Erik Verbeken<sup>13</sup>, Paul De Leyn<sup>11</sup>, Adrian Liston<sup>13,6,9</sup>, Johan Vansteenkiste<sup>13,4</sup>, Peter Carmeliet<sup>1,12,14</sup>, Stein Aerts<sup>5,6</sup> and Bernard Thienpont<sup>13,15\*</sup>



# PART-1!! RAW DATA TO COUNT DATA



## Strategy:

- FastQC
- Trim Galore
- STAR/HISAT/Bowtie
- Cufflinks/RSEM/HTseq-count
- Count/Read-count Matrix

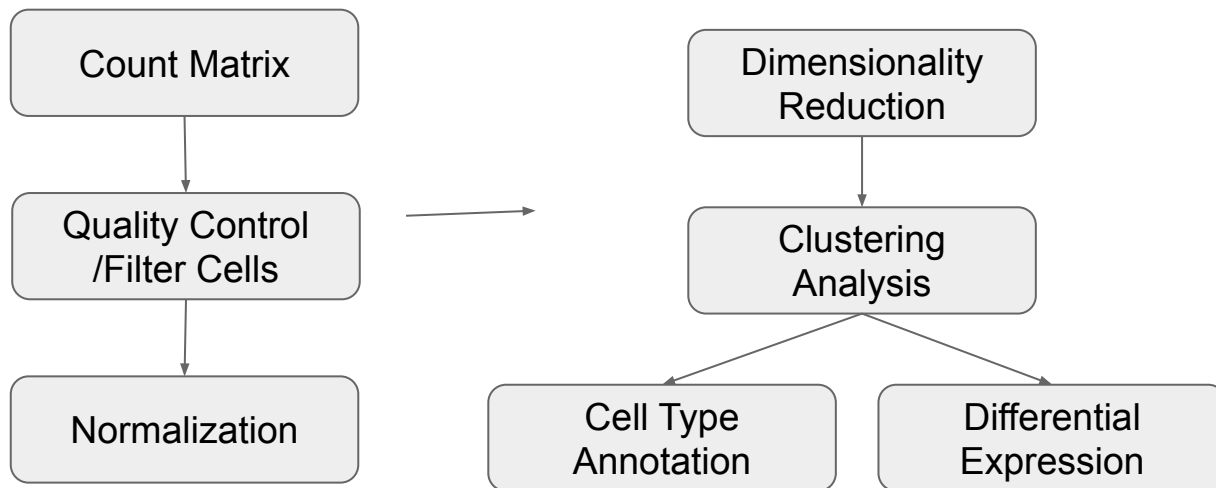
Cell-Ranger suite  
DropEST suite  
etc

## Strategy:

- FastQC
- Trim Galore
- Kallisto(BusTools)/Alvein
- Count/Read-count Matrix



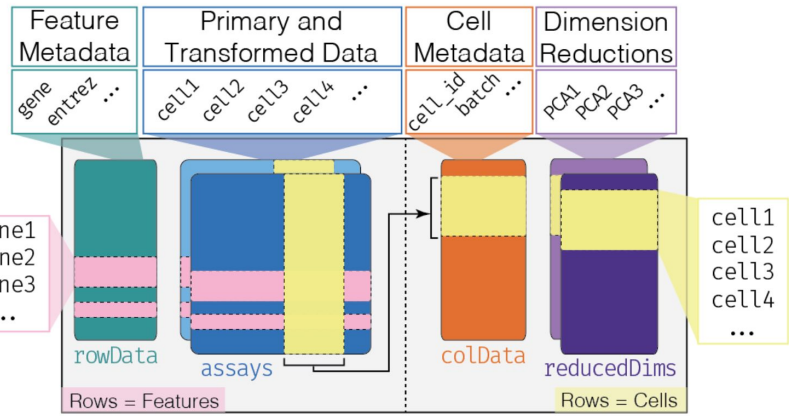
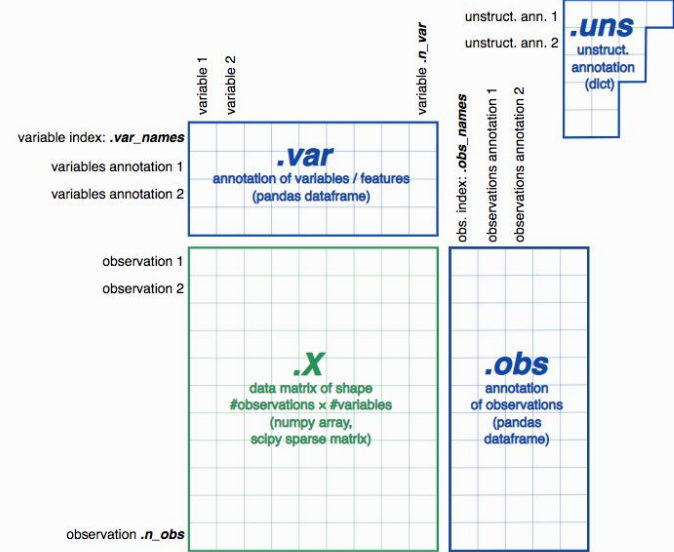
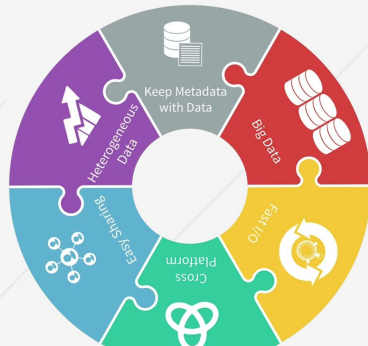
# PART-2!! LET'S PLAY WITH EXPRESSION DATA



# HUGE DATA!!

What is HDF5@?

~	HETEROGENEOUS DATA	+
🔗	EASY SHARING	+
↔	CROSS PLATFORM	+
⚡	FAST I/O	+
📦	BIG DATA	+
🔄	KEEP METADATA WITH DATA	+



SingleCellExperiment

# WEAPONS



## Orchestrating Single-Cell Analysis with Bioconductor

2020-05-04

### Welcome

This is the website for "Orchestrating Single-Cell Analysis with Bioconductor", a book that teaches users some common workflows for the analysis of single-cell RNA-seq data (scRNA-seq). This book will teach you how to make use of cutting-edge Bioconductor tools to process, analyze, visualize, and explore scRNA-seq data. Additionally, it serves as an online companion for the manuscript "Orchestrating Single-Cell Analysis with Bioconductor".

While we focus here on scRNA-seq data, a newer technology that



Custom Scripts

Programming  
Languages  
highly found:

R & Python!!

# RESOURCES!! (MUST READ)

A paper to read <https://www.embopress.org/doi/10.15252/msb.20188746>

## Courses to follow:

<https://scrnaseq-course.cog.sanger.ac.uk/website/index.html> (R version)(little knowledge)

<https://chanzuckerberg.github.io/scRNA-python-workshop/intro/about> (Python version)  
(little knowledge)

[https://broadinstitute.github.io/2020\\_scWorkshop/](https://broadinstitute.github.io/2020_scWorkshop/) (very beginner) (start here)

# CONTACT ME!!

 [arvind.iyer@unil.ch](mailto:arvind.iyer@unil.ch)

 [www.arvindkiyer.com](http://www.arvindkiyer.com)

 [@Arvind K Iyer](https://twitter.com/Arvind_K_Iyer)



Thank  
you!