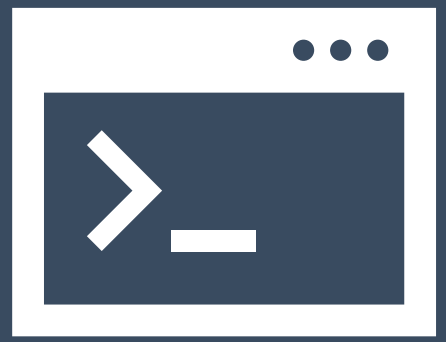
HBC
Harvard Chan Bioinformatics Core

Introduction to Single-cell RNA-seq Analysis

GEN349: Current tools for Gene Analysis



Harvard Chan Bioinformatics Core
February 25th, 2025

https://tinyurl.com/hbc-GEN349-scRNAseq

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

Heather Wick

Will Gammerdinger

Noor Sohail

Elizabeth Partan

Alex Bartlett

Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

HARVARD MEDICAL SCHOOL

DF/HCC
DANA-FARBER / HARVARD CANCER CENTER

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

Training

- **Basic Data Skills**
  - Shell
  - R
- **Advanced Topics: Analysis of high-throughput sequencing data**
  - Chromatin Biology
  - Bulk RNA-seq
  - Differential Gene Expression
  - scRNA-seq
  - Variant Calling
- **Current Topics in Bioinformatics**

Consulting

- **Transcriptomics:** RNA-seq, small RNA-seq, scRNA-seq
- **Epigenetics:** ChIP-seq, genome wide methylation, ATAC-seq
- **DNA Variation:** WGS, resequencing, exome-seq, CNV studies
- **Functional enrichment analysis**
- **Experimental design help**
- **Grant support**

https://bioinformatics.sph.harvard.edu/

Shannan Ho Sui
Director

Meeta Mistry
Associate Director

Lorena Pantano
Director of Bioinformatics Platform

John Quackenbush
Faculty Advisor

Upen Bhattarai

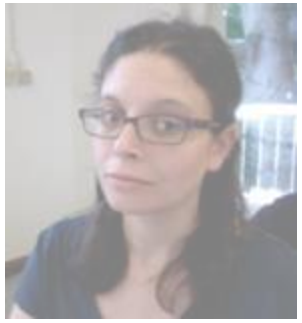
Heather Wick

Will Gammerdinger

Noor Sohail

Elizabeth Partan

Alex Bartlett



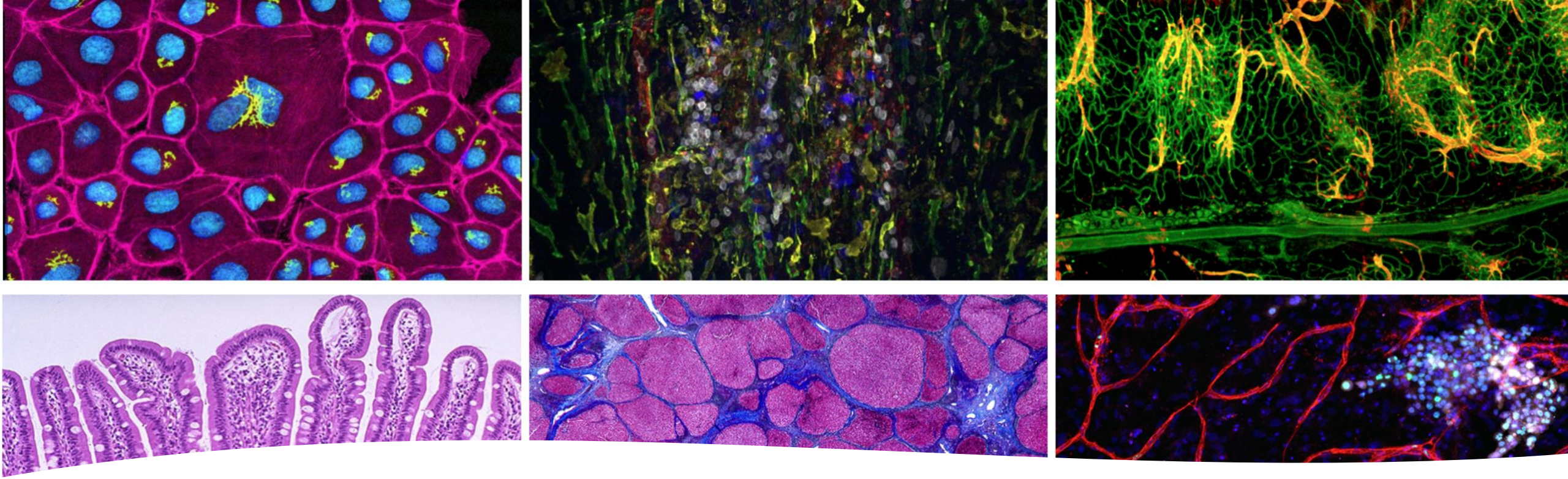
Emma Berdan

James Billingsley

Zhu Zhuo

Maria Simoneau

Introductions!



Why single cell RNA-seq?

Single-cell RNA-seq (scRNA-seq) allows us to evaluate the transcriptome at the level of individual cells. This offers a glimpse into the incredible diversity of cell types, states, and interactions.

Why single cell RNA-seq?

- To explore which cell types are present in a tissue
- To identify unknown/rare cell types or states
- To elucidate the changes in gene expression during differentiation processes or across time or states
- To identify genes that are differentially expressed in particular cell types between conditions (e.g. treatment or disease)
- To explore changes in expression among a cell type while incorporating spatial, regulatory, and/or protein information

scRNA-seq technologies

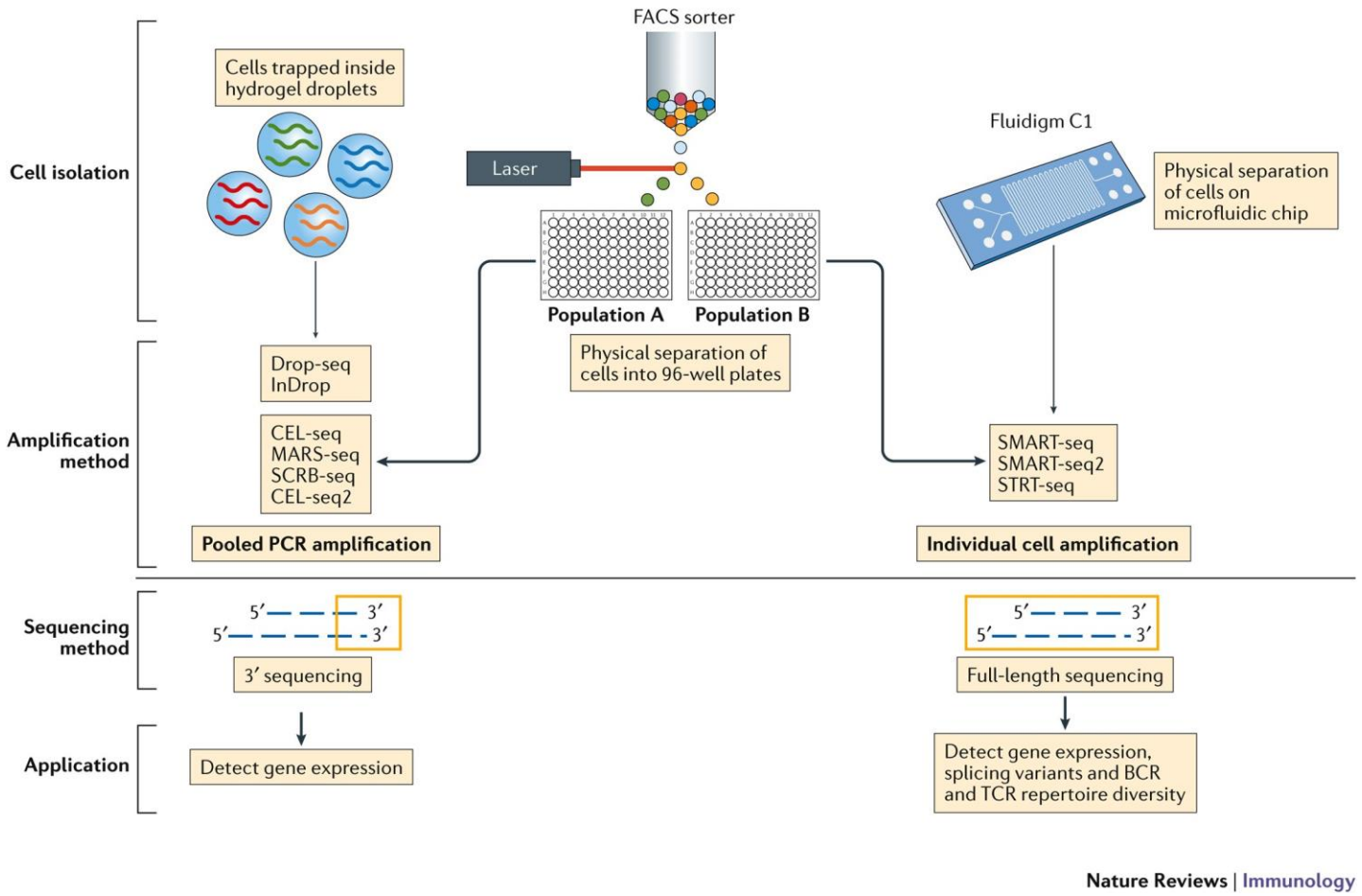
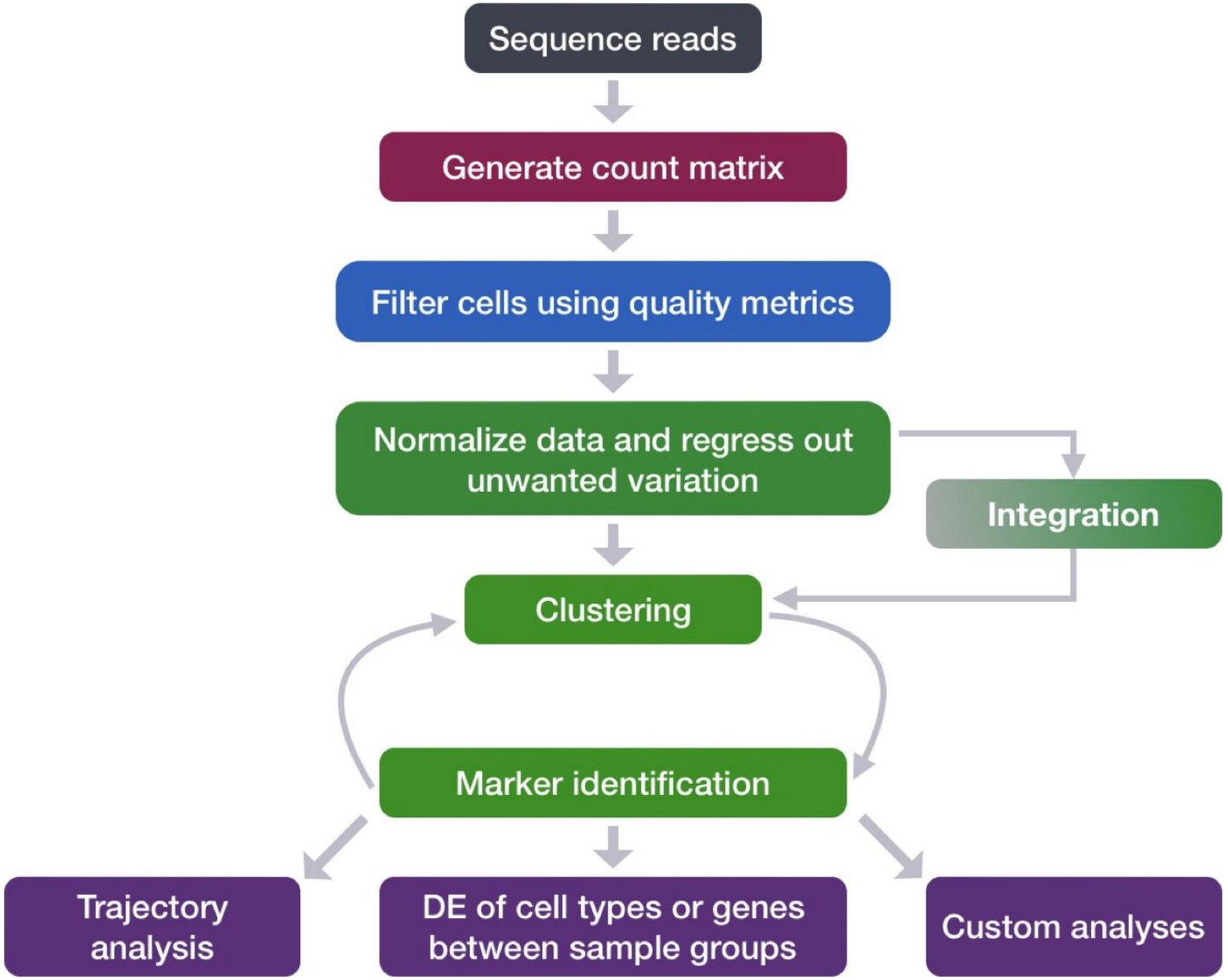
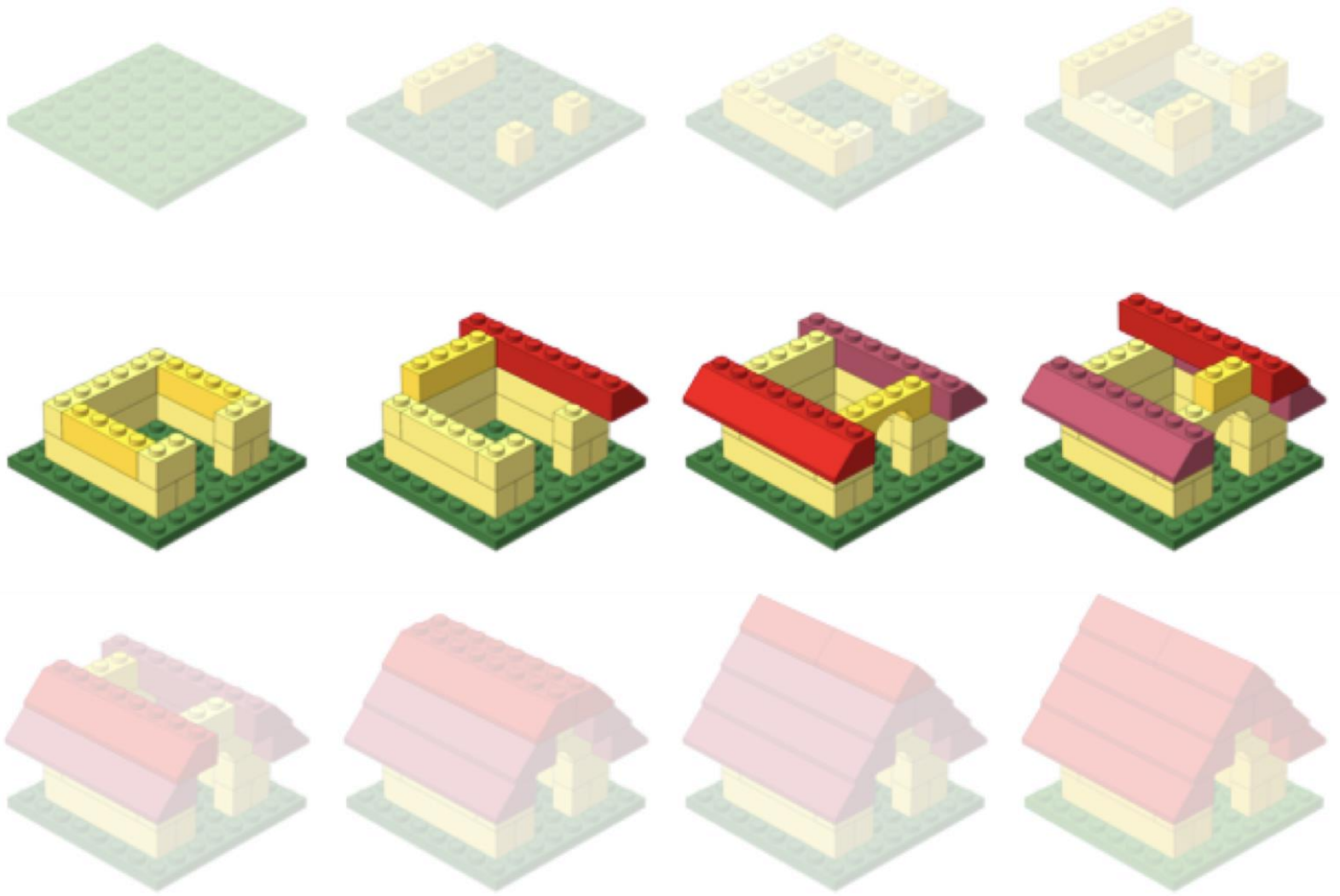


Image source: Papalexi E and Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity, Nature Reviews Immunology 2018

Analysis workflow





Bioinformatic Data Analysis

Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our install page. This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

  We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

  - Paper: Dictionary learning for integrative, multimodal, and scalable single-cell analysis
  - Vignette: Streamlined integration of scRNA-seq data
  - Vignette: Cross-modality bridge integration
  - Website: Azimuth-ATAC, reference-mapping for scATAC-seq datasets

- **Flexible, interactive, and highly scalable analsyis:** The size and scale of single-cell sequencing datasets is rapidly increasing, outpacing even Moore's law. In Seurat v5, we introduce new infrastructure and methods to analyze, interpret, and explore exciting datasets spanning millions of cells, even if they cannot be fully loaded into memory. We introduce support for 'sketch'-based analysis, where representative subsamples of a large dataset are stored in-memory to enable rapid and iterative analysis - while the full dataset remains accessible via on-disk storage.

Links

View on CRAN

Browse source code

Report a bug

License

Full license

MIT + file LICENSE

Community

Code of conduct

Citation

Citing Seurat

Developers

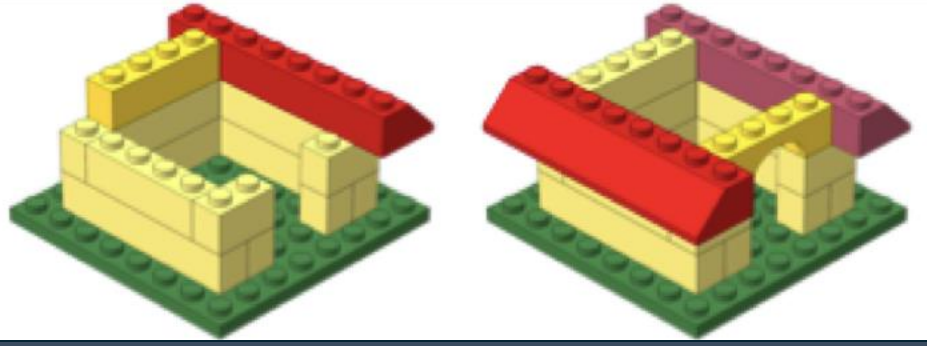
Rahul Satija
Author, maintainer

Satija Lab and Collaborators
Funder

More about authors...

https://satijalab.org/seurat/

Learning objectives



- Describe best practices for designing a single-cell RNA-seq experiment
- Describe steps in a single-cell RNA-seq analysis workflow
- Use Seurat and associated tools to perform analysis of single-cell expression data, including data filtering, QC, integration, clustering, and marker identification
- Understand practical considerations for performing scRNA-seq, rather than in-depth exploration of algorithm theory

Logistics

Course schedule

Day 1

Time	Topic	Instructor
13:00 - 13:15	Workshop introduction	Meeta
13:15 - 13:45	scRNA-seq pre-reading discussion	All
13:45 - 13:50	Break	
13:50 - 14:20	Quality control of Cellranger counts	Noor
14:20 - 14:55	Quality control setup in R	Meeta
14:55 - 15:00	Overview of self-learning materials and homework submission	Meeta

Before the next class:

I. Please **study the contents** and **work through all the code** within the following lessons:

1. Quality control with additional metrics
   Click here for a preview of this lesson
2. Theory of PCA
   Click here for a preview of this lesson

II. **Submit your work**:

https://tinyurl.com/hbc-GEN349-scRNAseq

Course materials

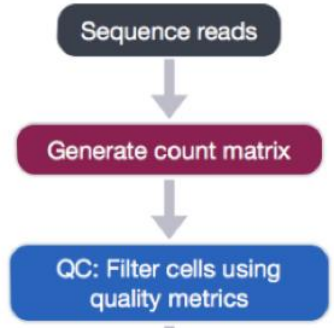
- We continuously update our materials to reflect changes in the field/software

Approximate time: 90 minutes

Learning Objectives:

- Understand how to bring in data from single-cell RNA-seq experiments
- Construct QC metrics and associated plots to visually explore the quality of the data
- Evaluate the QC metrics and set filters to remove low quality cells

Single-cell RNA-seq: Quality control



https://tinyurl.com/hbc-GEN349-scRNAseq

Course participation

- Mandatory review of self-learning lessons and assignments
- At-home lessons and exercises after each session
- Your questions and active participation drive learning
- **We look forward to all of your questions!**

Using AI for Assignments

- Do
  - Try to resolve error messages with it
  - Test code written by AI on a dataset where you have expected results
  - Take the time to review the generated code line-by-line
- Don’t
  - Implement it in replacement to learning
  - Write code that you don’t understand
  - Assume the output from an AI process is correct

Odds & Ends

- Quit/minimize all applications that are not required for class
- Post-its
  -  green - I am all set
  -  red - I need time/help
- Phones on vibrate/silent

Contact Us

HBC
Harvard Chan Bioinformatics Core

- *HBC training team:* hbctraining@hsph.harvard.edu
- *HBC consulting:* bioinformatics@hsph.harvard.edu