Single Cell RNA Seq

Introduction

Single cell omics is rapidly becoming the most essential raw material for single AI backed innovations in life science. It by far provides the quickest route to large scale biomedical data, accounting for heterogeneity, high dimensionality and complexity of data that is needed for training highly efficient neural networks. In this course, we will focus on single cell RNA seq (scRNA-seq) because of its broad applications.

Traditional RNA-seq has been used in samples composed of a mixture of cells, referred to as bulk RNA-seq, and has many applications. However, with bulk RNA-seq we can only estimate the average expression level for each gene across a population of cells, without regard for the heterogeneity in gene expression across individual cells of that sample. Therefore, it is insufficient for studying heterogeneous systems, e.g. early development studies or complex diseases and tissues such as the brain and immune system.

Single cell RNA seq experiments give us more depth and resolution into the cell specific transcriptome signatures. Through scRNA-seq, we can discover new cell types, cell composition, and track the trajectory of cell development.

Available Mentors

Core Resource: https://www.nature.com/articles/s41596-020-00409-w

Course Outline

- Stage Zero: Introduction to python
  - Syntax
  - Variable Assignment
  - Loops
  - Conditions
  - Functions
- Stage One: Python data libraries (crash course)
  - Numpy
  - Pandas
  - Seaborn
  - Matplotlib
  - Anndata (specifically for biological data)
  - Mention Scanpy
- Stage Two: The Single Cell Data Pipeline
  - Introduction to Single Cell Experiments
  - The Single Cell Data Analysis Pipeline
  - The Core Pipeline: Single Cell Pipeline with ScanPy
  - Task Bone Marrow Dataset
- Stage Three: Applications
  - Key Data Sources in SC
  - Trajectory Analysis (Psoriasis): Trajectory_Analysis_PSORIASIS
  - Pseudotime Prediction
  - Perturbation Analysis: Trajectory_Analysis_Perturbations_PBMC
- Stage Four: Projects

Projects

Stage Zero: HackBio Landing (Technical Writing and appreciation of Single Cell Genomics). In this introductory stage, participants will do two things:

- Develop essential technical writing skills, focusing on clear and concise communication of complex scientific concepts. Interns will learn how to document research findings, write reports, and create educational content that is accessible to both scientific and general audiences.
- Write a simple Python script for printing the names, slack username, country, 1 hobby, affiliations of people on your team and the DNA sequence of the genes they love most.
- Note: You are allowed to work in teams but submission is individual and promotion is automatic.
- Learning Resource (For coding alone): Python Course

S/N	Topics	Description
1	Single Cells, Big Data: How One Cell Can Tell a Thousand Stories	Explore the revolution of single-cell RNA-seq and what it reveals about cell identity and disease.
2	Why Biology Became a Data Science Field	Explain how Python, statistics, and machine learning are transforming the life sciences.
3.	The Art of Clustering Cells: How Computers Discover New Cell Types	Break down clustering algorithms (e.g., Leiden, Louvain) in simple terms.
4.	Why Two Scientists Can Analyze the Same Genome and Get Different Answers	Talk about variability in pipelines, parameters, and reference genomes.
5.	Why Personalized Medicine Isn't Just Hype—But Still Isn't Routine	Address where bioinformatics is essential and where the healthcare system lags behind.
6	From Bulk to Single-Cell: Advancing Transcriptomics with Python	Compare traditional RNA-seq to single-cell methods, highlighting Python scripts for clustering and trajectory inference in studying stem cell differentiation.
7	What Single-Cell Data Is Teaching Us About Cancer Evolution	Connect scRNA-seq to tumor heterogeneity and therapy resistance.
8	Can We Reconstruct a Human Body from Its Cells' RNA?	Discuss the frontier of spatial transcriptomics and multimodal single-cell analysis.

Stage One: **Python deep dive**

You are going to complete the following tasks to reinforce your python skills in bioinformatics.

- Write a python function for translating DNA to protein
- Write a python function for calculating the hamming distance between your slack username (e.g *josoga*) and twitter/X (*joseph*) handle (synthesize one if you don't have one). Feel free to pad it with extra words if they are not of the same length.

Furthermore, you are going to work with your team to review one of the following papers.

S/N	Topics	Description
1	*Tang et al., mRNA-Seq whole-transcriptome analysis of a single cell (Nature Methods, 2009).* https://pubmed.ncbi.nlm.nih.gov/19349980/	Details the original practical demonstration that you can sequence one cell's transcriptome. Basically, where it all started from
2	Picelli et al., Smart-seq2 for sensitive full-length transcriptome profiling (Nat Methods, 2013) https://www.nature.com/articles/nmeth.2639	The development of full-length single-cell libraries for isoforms and SNPs
3.	Ziegenhain et al., Comparative analysis of single-cell RNA-sequencing methods (Mol Cell, 2017) https://pubmed.ncbi.nlm.nih.gov/28212749/	Side-by-side benchmarking of major protocols — indispensable for understanding sensitivity, noise and cost tradeoffs
4.	Wolf et al., SCANPY: large-scale single-cell gene expression data analysis (Genome Biology, 2018). https://pubmed.ncbi.nlm.nih.gov/29409532/	The Python toolkit (AnnData + Scanpy) for single cell data analysis.
5.	Lun, Bach & Marioni, *Pooling across cells to normalize scRNA-seq data* (Genome Biol, 2016) https://pubmed.ncbi.nlm.nih.gov/27122128/	How normalization of data works
6	Current best practices in single-cell RNA-seq analysis: a tutorial https://www.embopress.org/doi/full/10.15252/msb.20188746	Essential tutorial for the computational workflow of single cell datasets.
7	Single-Cell RNA-Seq of the Pancreatic Islets—a Promise Not yet Fulfilled? https://www.cell.com/cell-metabolism/fulltext/S1550-4131(18)30735-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1550413118307356%3Fshowall%3Dtrue	Solving type 2 diabetes mellitus
8	A single-cell multiomics roadmap of zebrafish spermatogenesis reveals regulatory principles of male germline formation https://www.embopress.org/doi/full/10.1038/s44320-025-00157-7	Developmental biology
9	Dissecting microbial communities with single-cell transcriptome analysis https://www.science.org/doi/10.1126/science.adp625	Microbiology and metagenomics

	2	
10	Single-Cell Genomics: A Stepping Stone for Future Immunology Discoveries https://www.cell.com/cell/fulltext/S0092-8674(17)31320-X?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS009286741731320X%3Fshowall%3Dtrue	Immunology
11	The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans https://www.science.org/doi/10.1126/science.abl4896	A single cell atlas for humans

Stage Two:

Reproduce the complete single cell sequencing pipeline for the Bone Marrow Dataset: From QC to Cell Type Detection

Stage Three: Reproduce a paper

Trajectory Analysis

Project 1: SARS-Cov-2 Infection Dynamics

Reference: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001143

Data Source: MTX,TSV here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166766

This paper performs trajectory analysis on scRNA-seq data from SARS-CoV-2-infected human bronchial epithelial cells.

You are to reproduce the neighbourhood clustering and cell type identification in this paper (Figures 3A & 3B). You should also perform pseudotime analysis to order the differentiation of cells.

Project 2: Limb regeneration in axolotl

Reference: https://www.science.org/doi/10.1126/science.aaq0681

Data Source: MTX,TSV here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106269

This paper uses scRNA-Seq to understand how axolotl regenerates their limbs, the only tetrapod to do so. You are to reproduce the trajectory analysis showing that diverse connective tissue cell types converge into a multipotent blastema progenitor state and then redifferentiate (lineage funneling and re-patterning). (Figures 2B, C, D, E)

Project 3: Pseudotime + Trajectory Analysis in Liver Regeneration

Reference:

Data Source: MTX,TSV here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151309

This project analyzes scRNA-seq from mouse liver post-partial hepatectomy (PHx) to identify regeneration-associated hepatocytes. Using pseudotime and trajectory analysis, Reproduce the Figures 1-D,E,F,G

Perturbation Analysis:

Project 4: Cell type specific cytokine stimulation

Reference: https://www.nature.com/articles/nbt.4042

Data source: MTX, TSV here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96583

Reproduce an analysis of cell-type specific transcriptional responses to cytokine stimulation (IFN-β) across donors, and infer trajectories of activation/differentiation within immune compartments. This mixes *perturbation* (stimulus) + *trajectory* (activation pseudotime). Figures 2D, 3A, 3D

Final Projects will be published soon