



THE UNIVERSITY OF  
**CHICAGO**

Biological Sciences  
**Bioinformatics Core**

CENTER FOR RESEARCH INFORMATICS

BIOINFORMATICS CORE

**Bioinformatics Workshop Learning Series**

# single cell RNA-seq (scRNA-seq) data analysis

Yan Li, Ph.D.

Bioinformatics Core

CRI, University of Chicago

November 7<sup>th</sup>, 2023

# Who We Are

Center for Research Informatics  
(CRI)

CRDW  
(clinical Research  
Data Warehouse)

Bioinformatics  
Core

HPC

Data Storage

- Cohort Discovery
- REDCap
- NGS data analysis
  - Bulk \*-seq
  - sc \*-seq
- Gardner
- Randi
- Lab share
- VM
- Support grant applications
- Support research manuscript publications
- Bioinformatics workshop series from bioinformatics core
  - Bioinformatics concepts
  - Bioinformatics tools
  - Suggestions: [liyan@uchicago.edu](mailto:liyan@uchicago.edu); [bioinformatics@bsd.uchicago.edu](mailto:bioinformatics@bsd.uchicago.edu)

# Bioinformatics Core Workshop Series

- Autumn 2023 Quarter
  - November 2023: 2 scRNA-Seq workshops
    - Nov 7<sup>th</sup>: Introduction of 10x genomics sequencing protocols
    - Nov 21<sup>st</sup>: scRNA-seq data integration, downstream data analysis, and visualizations
  - December 2023: basics bioinformatics concept workshop
    - Introduction of shell scripts and HPC
- Winter 2024 Quarter: plan to launch 3 workshops
- **Provide your feedbacks regarding to topics to us**
  - [bioinformatics@bsd.uchicago.edu](mailto:bioinformatics@bsd.uchicago.edu)

# scRNA-seq Part1 Objectives

- Overview of 10x Genomics scRNA-seq protocols.
- Engage in hands-on training with the 10x Genomics software, CellRanger, for single cell data preprocessing.
- Learn to interpret the outputs from the CellRanger pipeline.

# scRNA-seq Part1 Session

- 1<sup>st</sup> half session: technique introduction
- 2<sup>nd</sup> half session: hands on practice + output interpretations
  - HPC randi access: only on-site person can obtain the HPC access (username and password), which will expire at 11/17/2023
  - Hands out practice material:  
2023workshop\_Nov\_scRNAseq\_handson\_notes.html
  - Hands on practice software, reference and test data are accessible at:  
'/gpfs/data/cri-training/Nov\_scRNAseq'

# Objective 1

## Overview of 10x Genomics scRNA-seq protocols

# scRNA-seq – 10x Genomics

Sample preparation

Single cell library construction

Sequencing

Data analysis  
Data visualization

Tissue or cell lines



Live cells



Cell suspension



Single cells / nuclei

GEM generation



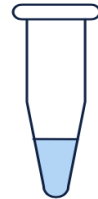
Chromium X Series



Chromium Connect



Chromium Controller



Sequencing ready libraries

Compatible sequencer

- **ILLUMINA® NovaSeq 6000**
- Illumina® HiSeq 3000/4000
- Illumina® HiSeq 2500 Rapid Run
- Illumina® NextSeq 500/550
- Illumina® NextSeq 1000/2000
- Illumina® MiSeq

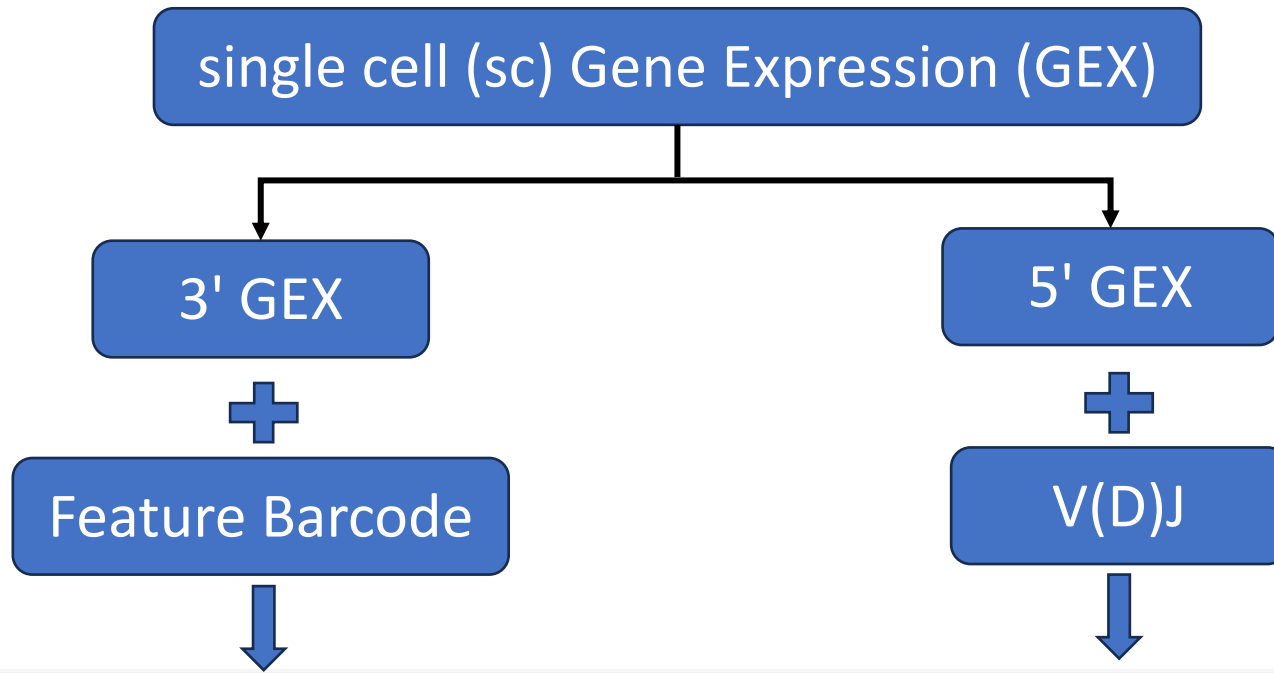
- 1
- Fastq to count matrix
  - cellranger
    - count
    - arc
    - aggregation
- 

• 10x genomics visualization software: Loupe Browser

- 2
- Count to integration/visualization
    - Seurat
- 



# 10x Genomics Single Cell Transcriptome

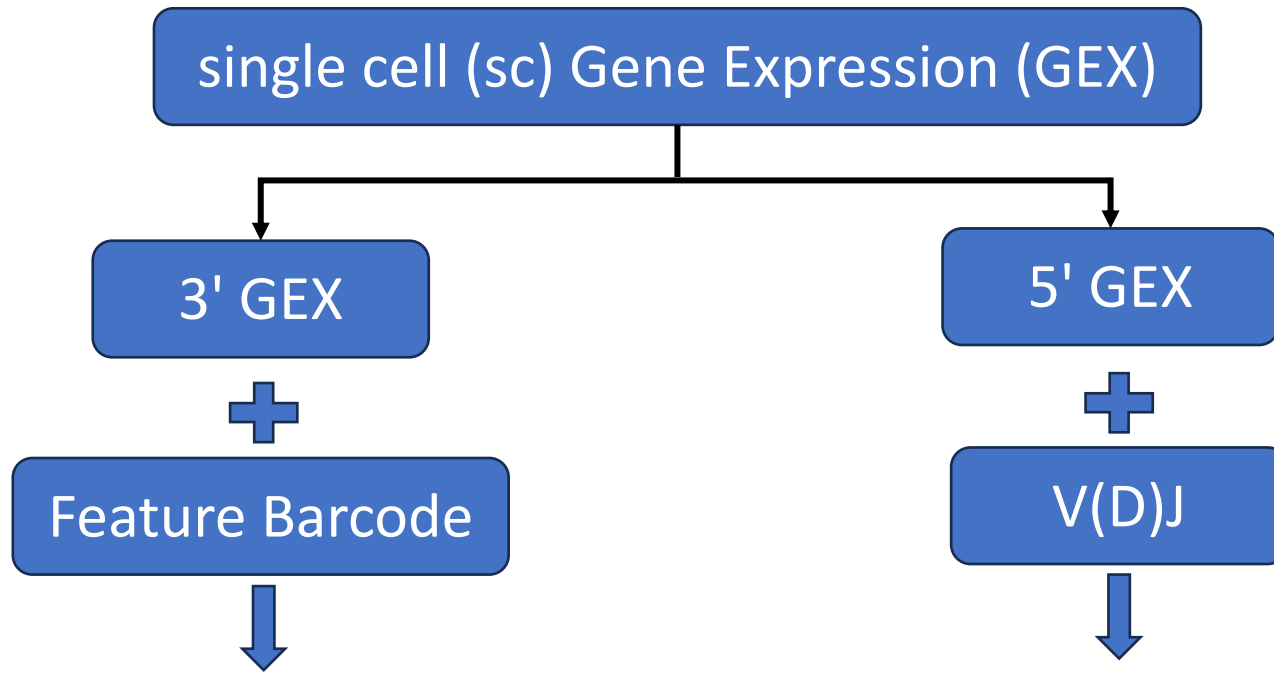


10x Solution	Gene Expression	Cell Surface Protein	CRISPR Screening
Single Cell Gene Expression 3' v3.1	✓	✓	✓
Single Cell Gene Expression 3' v3.1 (Dual Index)	✓	✓	✓
Single Cell Gene Expression 3' HT v3.1 (Dual Index)	✓	✓	✓
Single Cell Gene Expression 3' LT v3.1 (Dual Index)	✓	✓	-





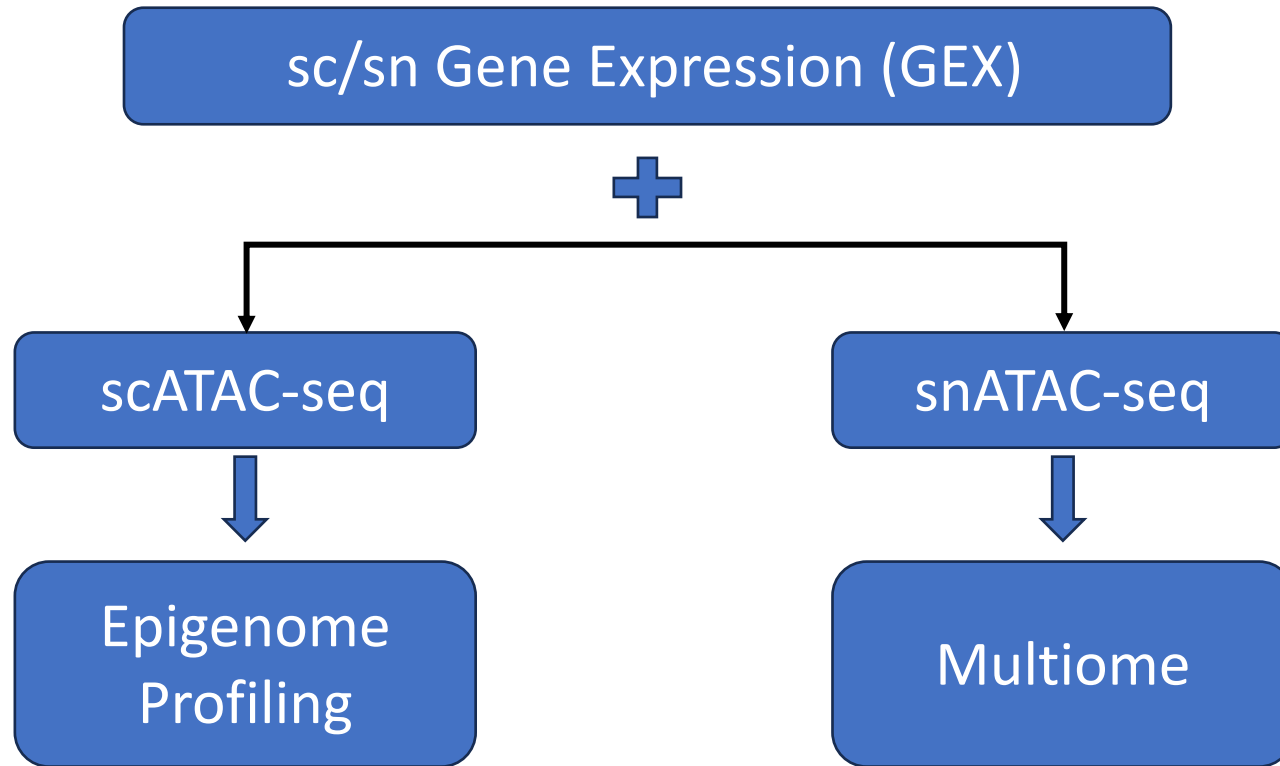
# 10x Genomics Single Cell Transcriptome



The Chromium™ Single Cell Immune Profiling Solution produces up to four different Illumina® sequencer-ready libraries:

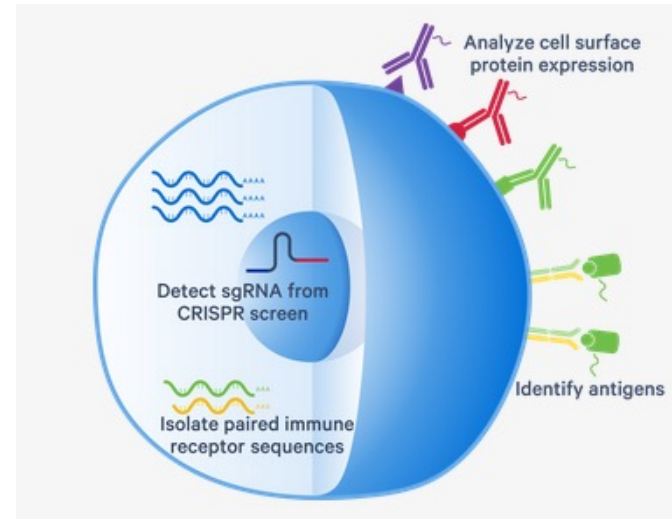
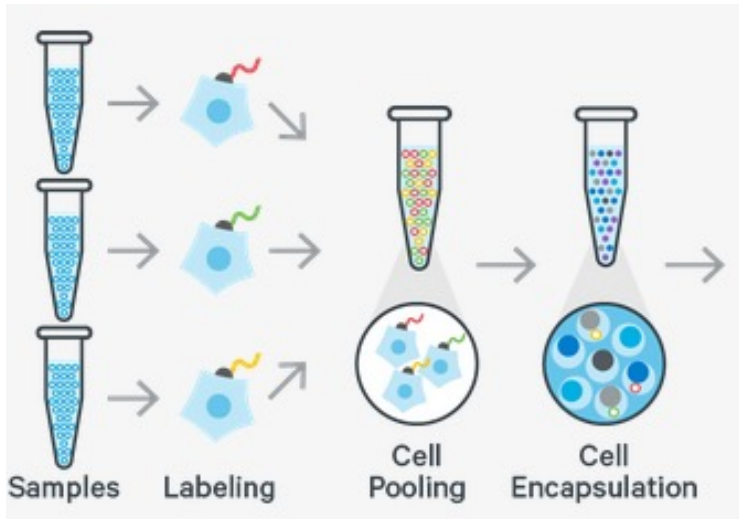
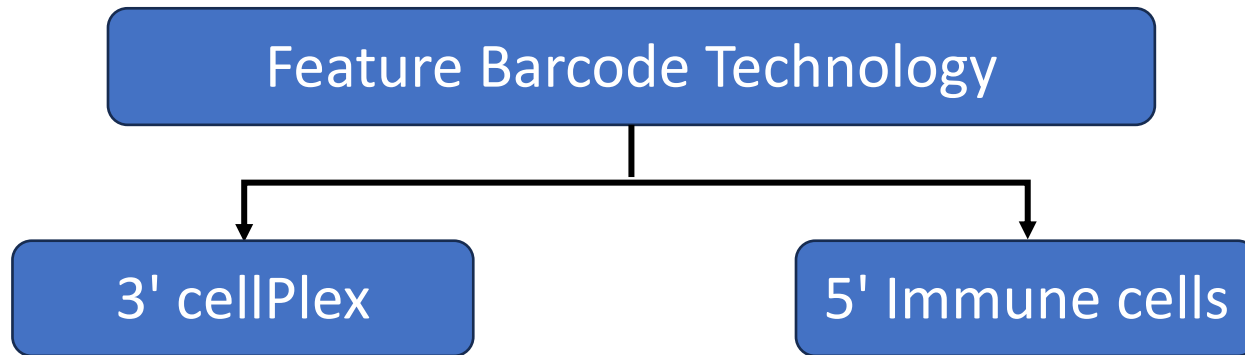
- V(D)J Enriched library (TCR and/or Ig)
- 5' Gene Expression library
- Cell Surface Protein library
- 5' CRISPR Screening library
- Barcode Enabled Antigen Mapping (BEAM) library

# 10x Genomics Single Cell Multiome

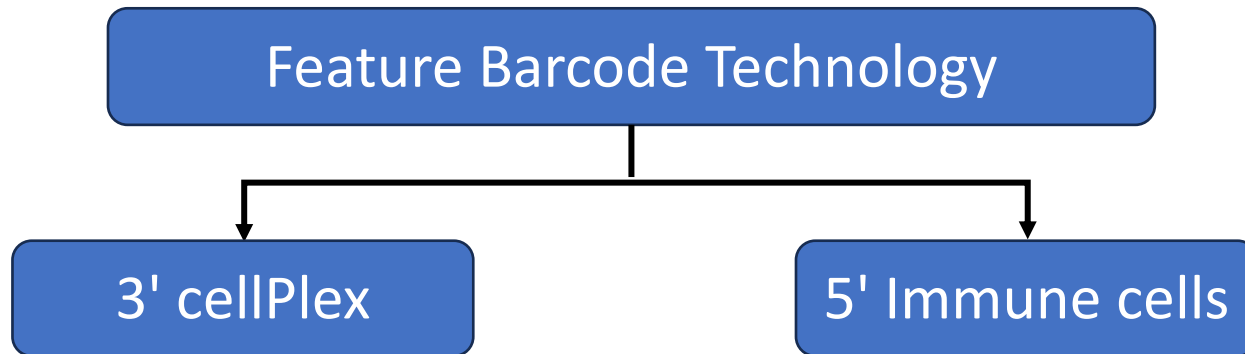


- Simultaneous detection of mRNA and chromatin accessibility
- Discover new gene regulatory interaction

# 10x Genomics Feature Barcode Technology



# 10x Genomics Feature Barcode Technology



- Chromium Next GEM Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry Dual Index) with Feature Barcoding technology for Cell Surface Protein (CG000317)
- Chromium Next GEM Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry Dual Index) with Feature Barcoding technology for Cell Surface Protein and Cell Multiplexing (CG000390)
- Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide with Feature Barcoding technology for Cell Surface Protein (CG000206)
- Chromium Next GEM Single Cell 5' Reagent Kits User Guide (v2 Chemistry Dual Index) with Feature Barcoding technology for Cell Surface Protein and Immune Receptor Mapping (CG000330)
- Chromium Next GEM Single Cell V(D)J Reagent Kits User Guide v1.1 with Feature Barcoding technology for Cell Surface Protein (CG000208)
- Chromium Single Cell 3' Reagent Kits v3 User Guide with Feature Barcoding technology for Cell Surface Protein (CG000185)
- Chromium Single Cell V(D)J Reagent Kits User Guide with Feature Barcoding technology for Cell Surface Protein (CG000186)
- Chromium Next GEM Single Cell 3' LT Reagent Kits (v3.1 Dual Index) with Feature Barcode technology for Cell Surface Protein (CG000400)

# Cellranger pipelines

- A set of analysis pipelines that process Chromium single cell data to align reads, generate feature-barcode matrices, perform clustering and other secondary analysis, and more
  - cellranger mkfastq
    - demultiplexes raw base call (BCL) files into FASTQ files
  - **cellranger count**
    - uses the Chromium cellular barcodes to generate feature-barcode matrices, additionally it also makes clustering analysis and performs gene expression analysis
  - **cellranger multi**
    - analyze Cell Multiplexing and Fixed RNA Profiling data
  - cellranger aggr
    - aggregates outputs from multiple runs of cellranger count or cellranger multi
  - cellranger reanalyze
    - reruns the dimensionality reduction, clustering, and gene expression algorithms using tunable parameter settings
- **cellranger-arc**: except multi, all other 4 pipelines to handle multiome data

## Objective 2

Hands-on training with the 10x Genomics software, Cell Ranger, for single cell data preprocessing

# HPC randi login



handout username

```
CRIs-MacBook-Pro-2:~ yanli$ ssh t.cri.biowksp01@randi.cri.uchicago.edu
** Unauthorized use/access is prohibited. **

This computer system is owned by the University of Chicago Biological Sciences
Division and is for authorized use only. Logging onto this computer verifies
you have read and agree both to the statement below and to use BSD computer
networks and systems in accordance with the BSD Eligibility and Acceptable Use
policy and related policies.

Individuals using this computer system are subject to having all of their
activities on this system monitored and recorded by system personnel. Anyone
using this system expressly consents to such monitoring and is advised that if
such monitoring reveals possible criminal activity or policy violation, system
personnel may provide the evidence of such monitoring to law enforcement or
other officials.

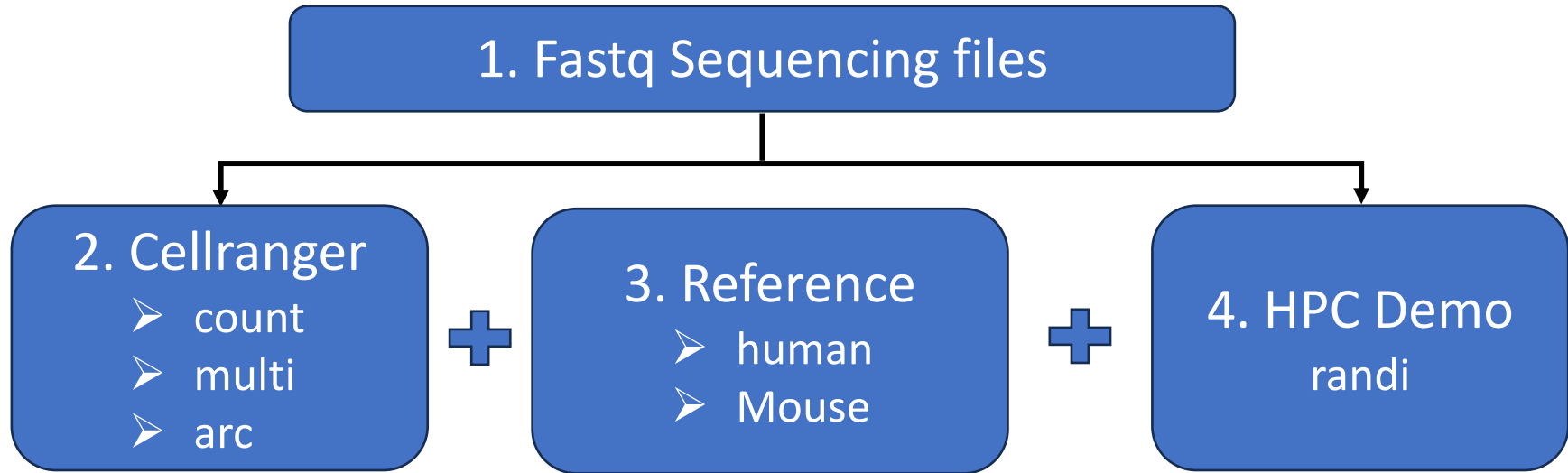
University of Chicago Acceptable Use Policy:
https://itservices.uchicago.edu/policies/acceptable-use-policy

(t.cri.biowksp01@randi.cri.uchicago.edu) Password:
Last login: Fri Nov  3 16:10:58 2023 from 205.208.121.120
[t.cri.biowksp01@cri22in002 ~]$
```

home directory

handout password

# cellranger pipeline



```
[t.cri.biowksp01@cri22in002 ~]$ ll /gpfs/data/cri-training/Nov_scrNAseq
total 1383683
drwxr-sr-x 2 yli22 cri-bioinformatics 4096 Nov 3 18:35 analysis_demo
drwxr-sr-x 6 yli22 cri-bioinformatics 4096 Aug 31 16:19 cellranger-7.2.0
-rwxr-xr-x 1 yli22 cri-bioinformatics 683925475 Sep 1 14:08 cellranger-7.2.0.tar.gz
drwxr-sr-x 6 yli22 cri-bioinformatics 4096 Aug 16 2022 cellranger-arc-2.0.2
-rwxr-xr-x 1 yli22 cri-bioinformatics 732879039 Aug 16 2022 cellranger-arc-2.0.2.tar.gz
drwxr-sr-x 4 yli22 cri-bioinformatics 4096 Nov 1 19:19 ref
drwxr-sr-x 5 yli22 cri-bioinformatics 4096 Nov 1 15:32 test_data
drwxr-sr-x 5 yli22 cri-bioinformatics 4096 Nov 1 18:51 test_data_lite
```

4

2

3

1



# Test data: lite version

```
-- testData1_count
|-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_I1_001.fastq.gz
|-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_R1_001.fastq.gz
|-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_R2_001.fastq.gz
-- testData2_cellPlex
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_I1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_I2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_R1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_R2_001.fastq.gz
-- testData3_arc
|-- ATAC
|-- AB-PB051523-10X-scA-EB1-A_S1_L001_I1_001.fastq.gz
|-- AB-PB051523-10X-scA-EB1-A_S1_L001_R1_001.fastq.gz
|-- AB-PB051523-10X-scA-EB1-A_S1_L001_R2_001.fastq.gz
|-- AB-PB051523-10X-scA-EB1-A_S1_L001_R3_001.fastq.gz
|-- RNA
|-- AB-PB051523-EB1-R_S1_L001_I1_001.fastq.gz
|-- AB-PB051523-EB1-R_S1_L001_I2_001.fastq.gz
|-- AB-PB051523-EB1-R_S1_L001_R1_001.fastq.gz
|-- AB-PB051523-EB1-R_S1_L001_R2_001.fastq.gz
```

## 1. cellranger count

- 3' GEX data

## 2. cellranger multi

- 3' cellPlex
- CMO feature barcode multiplexing

## 3. cellranger arc

- multiome
- RNA + ATAC



# Test data: full version

```
-- testData1_count
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_I1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_R1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L001_R2_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L002_I1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L002_R1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L002_R2_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L003_I1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L003_R1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L003_R2_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L004_I1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L004_R1_001.fastq.gz
-- AB-EL-HE-HE200915FCARNA-RNA3203A_S1_L004_R2_001.fastq.gz
```

```
-- testData3_arc
-- ATAC
-- AB-PB051523-10X-scA-EB1-A_S1_L001_I1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L001_R1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L001_R2_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L001_R3_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L002_I1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L002_R1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L002_R2_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L002_R3_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L003_I1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L003_R1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L003_R2_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L003_R3_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L004_I1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L004_R1_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L004_R2_001.fastq.gz
-- AB-PB051523-10X-scA-EB1-A_S1_L004_R3_001.fastq.gz
```

```
-- RNA
-- AB-PB051523-EB1-R_S1_L001_I1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L001_I2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L001_R1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L001_R2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L002_I1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L002_I2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L002_R1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L002_R2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L003_I1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L003_I2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L003_R1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L003_R2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L004_I1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L004_I2_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L004_R1_001.fastq.gz
-- AB-PB051523-EB1-R_S1_L004_R2_001.fastq.gz
```

```
-- testData2_cellPlex
-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L002_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L002_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L002_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1-CP_S6_L002_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L001_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L001_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L001_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L001_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L002_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L002_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L002_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-1_S1_L002_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L001_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L001_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L001_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L001_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L002_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L002_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L002_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2-CP_S7_L002_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L001_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L001_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L001_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L001_R2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L002_I1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L002_I2_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L002_R1_001.fastq.gz
-- CO-NS-10X-5S-OBE-2_S2_L002_R2_001.fastq.gz
```

# cellranger installation

Cellranger downloading page: <https://www.10xgenomics.com/support/software/cell-ranger/downloads>

The screenshot shows the 10x Genomics Support page for Cell Ranger. The main heading is "Download Center". A central text box provides instructions: "To set up and run Cell Ranger on Linux, please follow the steps below, or skip download and installation and get started with 10x Cloud Analysis (currently available in the United States and Canada):". A numbered list follows: 1. Check system requirements, 2. Download Cell Ranger, 3. Download References (Optional), 4. Download Probe Sets (Optional), 5. Install Cell Ranger, 6. 10x Genomics Cloud Analysis (Optional). A sidebar on the left lists navigation options: Overview, Getting Started, Downloads & Installation, Cell Ranger Commands, Analysis (with sub-items: Inputs, Running Pipelines, Outputs), Tutorials, Algorithms, and Advanced. A search bar and "Contact us" link are in the top right.

## tar.gz compression

[Download for Linux 64-bit \(tar.gz\)](#)

File size: 653 MB  
md5sum:  
85e2573e80a6f8656a42ff09460463e6

curl wget

```
wget -O cellranger-7.2.0.tar.gz "https://cf.10xgenomics.com/re"
Copy
```

# cellranger installation

```
[t.cri.biowksp01@cri22in002 ~]$ wget -O cellranger-7.2.0.tar.gz "https://cf.10xgenomics.com/releases/cell-exp/cellranger-7.2.0.tar.gz?Expires=1699091872&Key-Pair-Id=APKAI7S6A5RYOXBWRPDA&Signature=CkUr-y4Xwbz3uhCThK68LN1LeRow7PgqK3VDj-lm7twpK9I7mVEzplcWu6ZuOWQ3-8KyuwKBjqGVqwr4UdDxsIrCwK2fchDLtUgQUdZFlWHeYcqkGBwV2ugAB~5pnnxz-4BxhRsrc3oo6z6-wJNjSHeTBaPTzj~LX1qY6-l1Z42tuwBIgH-GgUfyquqwDw6uU5DHiAKnAmgvAEs3aeSDKmdIKcKt7CD2PBFHPBKodU4o8qvpxsf9-W9C77ylAi17WwylfsXxODOP0KsCpcriyiArbTuRSA3En~RrQyHNBjz71u6HD1vcUfizpCo5BRVoS1czmZjLHhYqEVe8fouw__"
--2023-11-03 17:06:51-- https://cf.10xgenomics.com/releases/cell-exp/cellranger-7.2.0.tar.gz?Expires=1699091872&Key-Pair-Id=APKAI7S6A5RYOXBWRPDA&Signature=CkUr-y4Xwbz3uhCThK68LN1LeRow7PgqK3VDj-lm7twpK9I7mVEzplcWu6ZuOWQ3-8KyuwKBjqGVqwr4UdDxsIrCwK2fchDLtUgQUdZFlWHeYcqkGBwV2ugAB~5pnnxz-4BxhRsrc3oo6z6-wJNjSHeTBaPTzj~LX1qY6-l1Z42tuwBIgH-GgUfyquqwDw6uU5DHiAKnAmgvAEs3aeSDKmdIKcKt7CD2PBFHPBKodU4o8qvpxsf9-W9C77ylAi17WwylfsXxODOP0KsCpcriyiArbTuRSA3En~RrQyHNBjz71u6HD1vcUfizpCo5BRVoS1czmZjLHhYqEVe8fouw__
Resolving cf.10xgenomics.com (cf.10xgenomics.com)... 104.18.1.173, 104.18.0.173, 2606:4700::6812:1ad, ..
Connecting to cf.10xgenomics.com (cf.10xgenomics.com)|104.18.1.173|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 683925475 (652M) [application/x-tar]
Saving to: 'cellranger-7.2.0.tar.gz'

cellranger-7.2.0.tar.gz  100%[=====] 652.24M  179MB/s  in 3.8s

2023-11-03 17:06:55 (171 MB/s) - 'cellranger-7.2.0.tar.gz' saved [683925475/683925475]

[t.cri.biowksp01@cri22in002 ~]$ ll
total 667648
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 683925475 Sep  1 14:08 cellranger-7.2.0.tar.gz
```

Unzip the downloaded \*.tar.gz file

```
[t.cri.biowksp01@cri22in002 ~]$ tar -xvzf cellranger-7.2.0.tar.gz
```

```
cellranger-7.2.0
|-- ITCFNSF
[t.cri.biowksp01@cri22in002 ~]$ /home/t.cri.biowksp01/cellranger-7.2.0/cellranger --help
cellranger cellranger-7.2.0

Process 10x Genomics Gene Expression, Feature Barcode, and Immune Profiling data

Usage: cellranger <COMMAND>

Commands:
  count      Count gene expression and/or feature barcode reads from a single sample and GEM
             well
  multi      Analyze multiplexed data or combined gene expression/immune profiling/feature
             barcode data

-- SOURCE CODE
```

# cellranger usage

```
[yli22@cri22in001 Nov_scrNAseq]$ /gpfs/data/bioinformatics/biocore_workshop_2023/Nov_scrNAseq/cellranger-7.2.0/cellranger --help
cellranger cellranger-7.2.0

Process 10x Genomics Gene Expression, Feature Barcode, and Immune Profiling data

Usage: cellranger <COMMAND>

Commands:
count ✓ Count gene expression and/or feature barcode reads from a single sample
and GEM well
multi ✓ Analyze multiplexed data or combined gene expression/immune
profiling/feature barcode data
multi-template Output a multi config CSV template
vdj Assembles single-cell VDJ receptor sequences from 10x Immune Profiling
libraries
aggr Aggregate data from multiple Cell Ranger runs
reanalyze Re-run secondary analysis (dimensionality reduction, clustering, etc)
mkvdjref Prepare a reference for use with CellRanger VDJ
mkfastq Run Illumina demultiplexer on sample sheets that contain 10x-specific
sample index sets
testrun Execute the 'count' pipeline on a small test dataset
mat2csv Convert a gene count matrix to CSV format
mkref Prepare a reference for use with 10x analysis software. Requires a GTF and
FASTA
mkgtf Filter a GTF file by attribute prior to creating a 10x reference
upload Upload analysis logs to 10x Genomics support
sitecheck Collect linux system configuration information
help Print this message or the help of the given subcommand(s)

Options:
-h, --help Print help
-V, --version Print version
```



# Reference and cellranger-arc installation

cellranger-arc downloading: <https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/downloads/latest>

## Cell Ranger ARC - 2.0.2 (August 18, 2022)

- Self-contained, relocatable tar file. Does not require centralized installation.
- Contains binaries pre-compiled for CentOS/RedHat 6.0+ and Ubuntu 12.04+.
- [Linux 64-bit - 699 MB](#) - md5sum: 7303f8ceee7b60113c9a0087268830cd

curl

wget

```
wget -O cellranger-arc-2.0.2.tar.gz "https://cf.10xgenomics.com/releases/cell-arc/cellranger-arc-2.0.2.tar.gz?Expires=1698925537&Policy=eyJTdGF0ZW1lbnQiOiIlt7IlJlc291cmNlIjoiaHR0cHM6Ly9jZi4xMHhnZW5vbWljcy5jb20vcmlvdGFzZXZmY2VsbC1hcmMvY2VsbHJhbmdlci1hcmMtMi4wLjIudGFyLmd6IiwiaWwiOiIwIiwiaW9uIj07IikRhdGVmZXNzVGhbiI6eyJBV1M6RXBvY2hUaW1lIj0xNjk4OTI1NTM3fX19XX0_&Signature=byXS0aeX0D0mzrEKCIlDeL4F1V~qYgmFbpqoLUT0bYlzuKWHdi2gS4xzcm-2Gzt7AcPtWcWoYx2cgnk09MNHlZxGsuBDayFZam5Re3T2~ddojYl0A9fwbWqwptZhU0ZmbqSyW8cdTJjDh~1K-riP~dlZse3DprWgJIqvUtaehqALNDT7qNqmp4Ng1mp0FtnMTs5BaLXzQqa3B80vEroFSpw8Rp6xHcvl8gFtMPP6up7UHqGgqKx6xbSXpiVQwwxERDZFpJJH23KpEieiaEIVrZKw0Go5HZCtqcU-KMkA1qbRXYtIaVsNXi84D3QEEmFRONFQWmxGfiLi8BKs-Na0g__&Key-Pair-Id=APKAI7S6A5RYOXBWRPDA"
```

cellranger-arc human reference file downloading:

## GRCh38 Reference - 2020-A-2.0.0 (May 3, 2021)

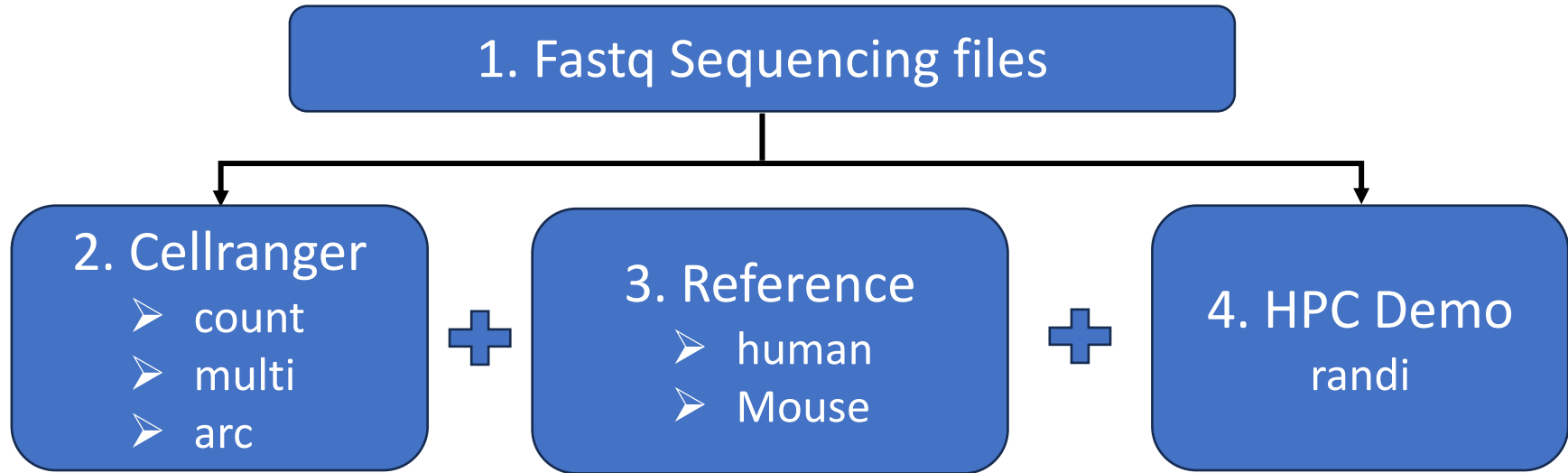
- Human reference (GRCh38) dataset required for Cell Ranger ARC.
- [Download - 14 GB](#) - md5sum: 2f12f6016197869e21e5559827002716
- [Build steps](#)

curl

wget

```
wget https://cf.10xgenomics.com/supp/cell-arc/refdata-cellranger-arc-GRCh38-2020-A-2.0.0.tar.gz
```

# cellranger pipeline



```
[yli22@cri22in001 ~]$ cd /gpfs/data/bioinformatics/biocore_workshop_2023/Nov_scRNAseq/
[yli22@cri22in001 Nov_scRNAseq]$ ll
total 1383715
drwxr-s--- 5 yli22 cri-bioinformatics 32768 Nov 1 20:28 analysis_demo
drwxr-s--- 6 yli22 cri-bioinformatics 4096 Aug 31 16:19 cellranger-7.2.0
-rwxr-x--- 1 yli22 cri-bioinformatics 683925475 Sep 1 14:08 cellranger-7.2.0.tar.gz
drwxr-s--- 6 yli22 cri-bioinformatics 4096 Aug 16 2022 cellranger-arc-2.0.2
-rwxr-x--- 1 yli22 cri-bioinformatics 732879039 Aug 16 2022 cellranger-arc-2.0.2.tar.gz
drwxr-s--- 4 yli22 cri-bioinformatics 4096 Nov 1 19:19 ref
drwxr-s--- 5 yli22 cri-bioinformatics 4096 Nov 1 15:32 test_data
drwxr-s--- 5 yli22 cri-bioinformatics 4096 Nov 1 18:51 test_data_lite
```

- 4
- 2
- 3
- 1

# Hands Practice 1: cellranger count

```
[t.cri.biowksp01@cri22in002 ~]$ cp /gpfs/data/cri-training/Nov_scrRNAseq/analysis_demo/* ./
[t.cri.biowksp01@cri22in002 ~]$ ll
total 667907
drwxr-x--- 6 t.cri.biowksp01 t.cri.biowksp01      4096 Aug 31 16:19 cellranger-7.2.0
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 683925475 Sep  1 14:08 cellranger-7.2.0.tar.gz
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01      775 Nov  3 17:47 testData1_count.slurm
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01      503 Nov  3 17:47 testData2_cellPlex_OBE-1.csv
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01      612 Nov  3 17:47 testData2_cellPlex_multi.slurm
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01      252 Nov  3 17:47 testData3_arc.csv
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01      705 Nov  3 17:47 testData3_arc.slurm
```

## testData1\_count.slurm

```
#!/bin/bash

#SBATCH --job-name=testData1_cellranger_count
#SBATCH --account=t.cri.biowksp01
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --time=10:00:00
#SBATCH --cpus-per-task=4
#SBATCH --mem=64gb
#SBATCH --output=%x.%j.out
#SBATCH --error=%x.%j.err

export MRO_DISK_SPACE_CHECK=disable

echo hostname = $HOSTNAME

echo `date`
echo -----
cd ~/analysis_results

/gpfs/data/cri-training/Nov_scrRNAseq/cellranger-7.2.0/cellranger count --id=samp1_ln1 \
--transcriptome=/gpfs/data/cri-training/Nov_scrRNAseq/ref/refdata-gex-GRCh38-2020-A \
--fastqs=/gpfs/data/cri-training/Nov_scrRNAseq/test_data_lite/testData1_count \
--sample=AB-EL-HE-HE200915FCARNA-RNA3203A > testData1_cellranger_count.log 2>&1

echo -----
echo `date`

[t.cri.biowksp01@cri22in002 ~]$ mkdir ~/analysis_results1
```

Handout username

Maximum execution time

Memory for execution

Execution/output folder, need to create this folder

Any name to save output

Reference folder

Fastq sequencing folder

1. prefix file name  
2. execution log file



# cellranger count help page

Usage: cellranger count [OPTIONS] --id <ID> --transcriptome <PATH>

## Options:

--id <ID> A unique run id and output folder name [a-zA-Z0-9\_-]+  
--description <TEXT> Sample description to embed in output files [default: ]  
--transcriptome <PATH> Path of folder containing 10x-compatible transcriptome reference  
--fastqs <PATH> Path to input FASTQ data  
--project <TEXT> Name of the project folder within a mkfastq or bcl2fastq-generated folder from which to pick FASTQs  
--sample <PREFIX> Prefix of the filenames of FASTQs to select  
--lanes <NUMS> Only use FASTQs from selected lanes  
--libraries <CSV> CSV file declaring input library data sources  
--feature-ref <CSV> Feature reference CSV file, declaring Feature Barcode constructs and associated barcodes  
--expect-cells <NUM> Expected number of recovered cells, used as input to cell calling algorithm  
--force-cells <NUM> Force pipeline to use this number of cells, bypassing cell calling algorithm. [MINIMUM: 10]  
--no-bam Set --no-bam to not generate the BAM file. This will reduce the total computation time for the pipeline and the size of the output directory. If unsure, we recommend not to use this option. BAM file could be useful for troubleshooting and downstream analysis  
--nosecondary Disable secondary analysis, e.g. clustering. Optional  
--r1-length <NUM> Hard trim the input Read 1 to this length before analysis  
--r2-length <NUM> Hard trim the input Read 2 to this length before analysis  
--include-introns <true|false> Include intronic reads in count [default: true] [possible values: true, false]  
--chemistry <CHEM> Assay configuration. NOTE: by default the assay configuration is detected automatically, which is the recommended mode. You usually will not need to specify a chemistry. Options are: 'auto' for autodetection, 'threeprime' for Single Cell 3', 'fiveprime' for Single Cell 5', 'SC3Pv1' or 'SC3Pv2' or 'SC3Pv3' for Single Cell 3' v1/v2/v3, 'SC3Pv3LT' for Single Cell 3' v3 LT, 'SC3Pv3HT' for Single Cell 3' v3 HT, 'SC5P-PE' or 'SC5P-R2' for Single Cell 5', paired-end/R2-only, 'SC-FB' for Single Cell Antibody-only 3' v2 or 5'. To analyze the GEX portion of multiome data, chemistry must be set to 'ARC-v1'; 'ARC-v1' chemistry cannot be autodetected [default: auto]  
--no-libraries Proceed with processing using a --feature-ref but no Feature Barcode libraries specified with the 'libraries' flag  
--check-library-compatibility <true|false> Whether to check for barcode compatibility between libraries. [default: true] [possible values: true, false]  
--dry Do not execute the pipeline. Generate a pipeline invocation (.mro) file and stop job manager to use. Valid options: local (default), sge, lsf, slurm or path to a .template file. Search for help on "Cluster Mode" at support.10xgenomics.com for more details on configuring the pipeline to use a compute cluster [default: local]  
--jobmode <MODE>  
--localcores <NUM> Set max cores the pipeline may request at one time. Only applies to local jobs  
--localmem <NUM> Set max GB the pipeline may request at one time. Only applies to local jobs  
--localvmem <NUM> Set max virtual address space in GB for the pipeline. Only applies to local jobs  
--mempercore <NUM> Reserve enough threads for each job to ensure enough memory will be available, assuming each core on your cluster has at least this much memory available. Only applies to cluster jobmodes  
--maxjobs <NUM> Set max jobs submitted to cluster at one time. Only applies to cluster jobmodes  
--jobinterval <NUM> Set delay between submitting jobs to cluster, in ms. Only applies to cluster jobmodes



# Hands Practice 1: randi job submission

```
[t.cri.biowksp01@cri22in002 ~]$ sbatch testData1_count.slurm
```

Submit job for execution

```
[t.cri.biowksp01@cri22in002 ~]$ squeue -u t.cri.biowksp01
```

Check jobs execution in your account

```
      JOBID PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
      3078143      tier1q testData t.cri.bi R      1:25      1 cri22cn070
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results1/
```

```
total 4097
```

```
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1483 Nov  3 17:59 __samp1_ln1.mro
drwxr-x---  5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 17:59 samp1_ln1
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 15773 Nov  3 18:08 testData1_cellranger_count.log
```



# Hands Practice 2: cellranger multi

testData2\_cellPlex\_multi.slurm

```
#!/bin/bash

#SBATCH --job-name=testData2_cellranger_multi
#SBATCH --account=t.cri.biowksp01
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --time=10:00:00
#SBATCH --cpus-per-task=4
#SBATCH --mem=64gb
#SBATCH --output=%x.%j.out
#SBATCH --error=%x.%j.err

export MRO_DISK_SPACE_CHECK=disable

echo hostname = $HOSTNAME

echo `date`
echo -----
cd ~

/gpfs/data/cri-training/Nov_scRNAseq/cellranger-7.2.0/cellranger multi --id=samp1_obe1 \
--csv=testData2_cellPlex_OBE-1.csv > testData2_cellranger_multi_cellPlex.log 2>&1

echo -----
echo `date`
```

testData2\_cellPlex\_OBE-1.csv (corresponding config file for cellranger multi)

```
[gene-expression]
reference,/gpfs/data/cri-training/Nov_scRNAseq/ref/refdata-gex-GRCh38-2020-A
expect-cells,9580
include-introns,TRUE

[libraries]
fastq_id,fastqs,feature_types
CO-NS-10X-5S-OBE-1,/gpfs/data/cri-training/Nov_scRNAseq/test_data_lite/testData2_cellPlex,Gene Expression
CO-NS-10X-5S-OBE-1-CP,/gpfs/data/cri-training/Nov_scRNAseq/test_data_lite/testData2_cellPlex,Multiplexing Capture

[samples]
sample_id,cmo_ids
13080_D,CM0301
13801_C,CM0302
13801_D,CM0303
13853_C,CM0305
13862_D,CM0306

-- testData2_cellPlex
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_I2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1-CP_S6_L001_R2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_I1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_I2_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_R1_001.fastq.gz
|-- CO-NS-10X-5S-OBE-1_S1_L001_R2_001.fastq.gz
```

# Hands Practice 2: cellranger multi

```
[t.cri.biowksp01@cri22in002 ~]$ ll
total 667907
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:24 analysis_results1
drwxr-x--- 6 t.cri.biowksp01 t.cri.biowksp01 4096 Aug 31 16:19 cellranger-7.2.0
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 683925475 Sep 1 14:08 cellranger-7.2.0.tar.gz
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 0 Nov 3 17:59 testData1_cellranger_count.3078143.err
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 98 Nov 3 18:18 testData1_cellranger_count.3078143.out
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 740 Nov 3 17:49 testData1_count.slurm
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 503 Nov 3 18:27 testData2_cellPlex_OBE-1.csv
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 559 Nov 3 18:24 testData2_cellPlex_multi.slurm
```

```
[t.cri.biowksp01@cri22in002 ~]$ sbatch testData2_cellPlex_multi.slurm
Submitted batch job 3078144
```

```
[t.cri.biowksp01@cri22in002 ~]$ queue -u t.cri.biowksp01
      JOBID PARTITION   NAME     USER ST   TIME  NODES NODELIST(REASON)
      3078147      tier1q  testData t.cri.bi PD    0:00      1 (None)
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll
total 667909
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 876 Nov 3 18:31 __samp1_obel.mro → Temporary folder
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:24 analysis_results1
drwxr-x--- 6 t.cri.biowksp01 t.cri.biowksp01 4096 Aug 31 16:19 cellranger-7.2.0
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 683925475 Sep 1 14:08 cellranger-7.2.0.tar.gz
drwxr-x--- 5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:31 samp1_obel → output folder
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 0 Nov 3 17:59 testData1_cellranger_count.3078143.err
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 98 Nov 3 18:18 testData1_cellranger_count.3078143.out
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 740 Nov 3 17:49 testData1_count.slurm
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 503 Nov 3 18:27 testData2_cellPlex_OBE-1.csv
-rwxr-x--- 1 t.cri.biowksp01 t.cri.biowksp01 559 Nov 3 18:24 testData2_cellPlex_multi.slurm
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 0 Nov 3 18:31 testData2_cellranger_multi.3078147.err
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 60 Nov 3 18:31 testData2_cellranger_multi.3078147.out
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 296 Nov 3 18:31 testData2_cellranger_multi_cellPlex.log → Log files
```

# Hands Practice 3: cellranger arc

```
[t.cri.biowksp01@cri22in002 ~]$ cp /gpfs/data/cri-training/Nov_scrNAseq/analysis_demo/testData3_arc.slurm ./
```

```
[t.cri.biowksp01@cri22in002 ~]$ vim testData3_arc.slurm
```

```
#!/bin/bash
```

```
#SBATCH --job-name=testData3_cellrange_arc
#SBATCH --account=t.cri.biowksp01
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --time=10:00:00
#SBATCH --cpus-per-task=4
#SBATCH --mem=64gb
#SBATCH --output=%x.%j.out
#SBATCH --error=%x.%j.err
```

```
export MRO_DISK_SPACE_CHECK=disable
```

```
echo hostname = $HOSTNAME
```

```
echo `date`
```

```
echo -----
```

```
cd ~/analysis_results_arc_v123
```

```
/gpfs/data/cri-training/Nov_scrNAseq/cellranger-arc-2.0.2/cellranger-arc count --id=samp1_arc \  
--reference=/gpfs/data/cri-training/Nov_scrNAseq/ref/refdata-cellranger-arc-GRCh38-2020-A-2.0.0 \  
--libraries=testData3_arc.csv > testData3_cellranger_arc_count.log 2>&1
```

```
echo -----
```

```
echo `date`
```

```
[t.cri.biowksp01@cri22in002 ~]$ mkdir analysis_results_arc_v123
```

```
[t.cri.biowksp01@cri22in002 ~]$ cp /gpfs/data/cri-training/Nov_scrNAseq/analysis_demo/testData3_arc.csv analysis_results_arc_v123/
```

```
[t.cri.biowksp01@cri22in002 ~]$ vim analysis_results_arc_v123/testData3_arc.csv
```

```
fastqs,sample,library_type
```

```
/gpfs/data/cri-training/Nov_scrNAseq/test_data_lite/testData3_arc/RNA,AB-PB051523-EB1-R,Gene Expression
```

```
/gpfs/data/cri-training/Nov_scrNAseq/test_data_lite/testData3_arc/ATAC,AB-PB051523-10X-scA-EB1-A,Chromatin Accessibility
```

```
[t.cri.biowksp01@cri22in002 ~]$ sbatch testData3_arc.slurm
```

```
Submitted batch job 3078148
```

```
[t.cri.biowksp01@cri22in002 ~]$ squeue -u t.cri.biowksp01
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
3078148	tier1q	testData	t.cri.bi	R	0:16	1	cri22cn070
3078147	tier1q	testData	t.cri.bi	R	13:16	1	cri22cn070

# Hands Practice 3: cellranger arc

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123
total 4098
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1662 Nov  3 18:44 __samp1_arc.mro
drwxr-x---  5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 18:44 samp1_arc
-rwxr-x---  1 t.cri.biowksp01 t.cri.biowksp01  253 Nov  3 18:36 testData3_arc.csv
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 4500 Nov  3 18:46 testData3_cellranger_arc_count.log
```

Temporary folder  
output folder  
Log files

## Objective 3

Learn to interpret the outputs from the CellRanger pipeline.

# Output: cellranger count

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results1/samp1_ln1/
total 7813
drwxr-x--- 10 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:15 SC_RNA_COUNTER_CS
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 288 Nov 3 18:18 _cmdline
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 159136 Nov 3 18:18 _filelist
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2260877 Nov 3 18:18 _finalstate
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1483 Nov 3 17:59 _invocation
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5 Nov 3 17:59 _jobmode
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 96163 Nov 3 18:18 _log
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 298731 Nov 3 17:59 _mresource
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 886961 Nov 3 18:18 _perf
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 19086 Nov 3 18:18 _sitecheck
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2 Nov 3 17:59 _tags
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 51 Nov 3 18:18 _timestamp
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 36 Nov 3 17:59 _uuid
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 31670 Nov 3 18:18 _vdrkill
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 65 Nov 3 17:59 _versions
drwxr-x--- 5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:18 outs
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3627848 Nov 3 18:18 samp1_ln1.mri.tgz
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results1/samp1_ln1/outs/
total 2259074
drwxr-x--- 7 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:18 analysis
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 44305818 Nov 3 18:17 cloupe.cloupe
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:18 filtered_feature_bc_matrix
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 9545048 Nov 3 18:14 filtered_feature_bc_matrix.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 647 Nov 3 18:16 metrics_summary.csv
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 64848066 Nov 3 18:15 molecule_info.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2170486930 Nov 3 18:18 possorted_genome_bam.bam
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3175536 Nov 3 18:18 possorted_genome_bam.bam.bai
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:18 raw_feature_bc_matrix
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 15464629 Nov 3 18:13 raw_feature_bc_matrix.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5104757 Nov 3 18:16 web_summary.html
```

raw feature-  
barcode matrices  
metrics summary

alignment bam  
filtered feature-  
barcode matrices  
web summary

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results1/samp1_ln1/outs/filtered_feature_bc_matrix
total 23680
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 29288 Nov 3 18:14 barcodes.tsv.gz
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 333437 Nov 3 18:14 features.tsv.gz
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 23600928 Nov 3 18:14 matrix.mtx.gz
```

Input for downstream  
analysis



# Output: cellranger multi

```
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obe1/  
total 24453  
drwxr-x--- 12 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 SC_MULTI_CS  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 122 Nov 3 19:05 _cmdline  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 412546 Nov 3 19:05 _filelist  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5810573 Nov 3 19:04 _finalstate  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 876 Nov 3 18:31 _invocation  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5 Nov 3 18:31 _jobmode  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 374777 Nov 3 19:05 _log  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 287853 Nov 3 18:31 _mresource  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2253883 Nov 3 19:04 _perf  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 11080 Nov 3 19:05 _perf._truncated_  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 19088 Nov 3 19:05 _sitecheck  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2 Nov 3 18:31 _tags  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 51 Nov 3 19:04 _timestamp  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 36 Nov 3 18:31 _uuid  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 163876 Nov 3 19:04 _vdrkill  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 65 Nov 3 18:31 _versions  
drwxr-x--- 4 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 outs  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 14862520 Nov 3 19:05 samp1_obe1.mri.tgz
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obe1/outs/  
total 2  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 503 Nov 3 18:31 config.csv  
drwxr-x--- 4 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 multi  
drwxr-x--- 7 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 per_sample_outs
```

General multiplexing  
output  
Demultiplexed  
outputs per sample

```
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obe1/outs/multi/count/  
total 4368385  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 775 Nov 3 19:04 feature_reference.csv  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 108116871 Nov 3 19:04 raw_cloupe.cloupe  
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 raw_feature_bc_matrix  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 49979585 Nov 3 18:53 raw_feature_bc_matrix.h5  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 342869247 Nov 3 18:56 raw_molecule_info.h5  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3968688657 Nov 3 19:01 unassigned_alignments.bam  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3363016 Nov 3 19:02 unassigned_alignments.bam.bai  
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obe1/outs/multi/multiplexing_analysis/  
total 3201  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2627634 Nov 3 18:56 assignment_confidence_table.csv  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 145036 Nov 3 18:57 cells_per_tag.json  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 152704 Nov 3 18:57 tag_calls_per_cell.csv  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 481 Nov 3 18:57 tag_calls_summary.csv
```

# Output: cellranger multi

Demultiplexed outputs per sample

```
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obel/outs/per_sample_outs/  
total 3  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:20 13080_D  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 13801_C  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 13801_D  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 13853_C  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 13862_D
```

One sample demultiplexed output results

```
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obel/outs/per_sample_outs/13080_D/  
total 5633  
drwxr-x--- 4 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 count  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 6817 Nov 3 19:04 metrics_summary.csv  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5583636 Nov 3 19:04 web_summary.html  
[t.cri.biowksp01@cri22in002 ~]$ ll samp1_obel/outs/per_sample_outs/13080_D/count/  
total 448386  
drwxr-x--- 7 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 analysis  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 775 Nov 3 19:02 feature_reference.csv  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 410079488 Nov 3 19:01 sample_alignments.bam  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 2557992 Nov 3 19:01 sample_alignments.bam.bai  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 13612837 Nov 3 19:03 sample_cloupe.cloupe  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 19818 Nov 3 18:57 sample_filtered_barcode.csv  
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:04 sample_filtered_feature_bc_matrix  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 2259118 Nov 3 18:57 sample_filtered_feature_bc_matrix.h5  
-rw-r----- 2 t.cri.biowksp01 t.cri.biowksp01 30175935 Nov 3 18:59 sample_molecule_info.h5
```



# Output: cellranger-arc count

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123/samp1_arc/
total 7813
drwxr-x--- 5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 18:44 SC_ATAC_GEX_COUNTER_CS
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 220 Nov 3 19:20 _cmdline
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 192252 Nov 3 19:20 _filelist
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1537167 Nov 3 19:20 _finalstate
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1662 Nov 3 18:44 _invocation
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 5 Nov 3 18:44 _jobmode
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 152160 Nov 3 19:20 _log
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 146038 Nov 3 18:44 _mrosource
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1126835 Nov 3 19:20 _perf
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 9925 Nov 3 19:20 _perf_truncated_
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 19102 Nov 3 19:20 _sitecheck
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2 Nov 3 18:44 _tags
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 51 Nov 3 19:20 _timestamp
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 36 Nov 3 18:44 _uuid
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 205905 Nov 3 19:20 _vdrkill
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 68 Nov 3 18:44 _versions
drwxr-x--- 5 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:20 outs
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3882713 Nov 3 19:20 samp1_arc.mri.tgz
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123/samp1_arc/outs/
total 5921922
drwxr-x--- 6 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:20 analysis
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 342603530 Nov 3 19:03 atac_cut_sites.bigwig
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 241496117 Nov 3 19:00 atac_fragments.tsv.gz
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 599501 Nov 3 19:00 atac_fragments.tsv.gz.tbi
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3666335 Nov 3 19:08 atac_peak_annotation.tsv
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1572779 Nov 3 19:03 atac_peaks.bed
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3198462792 Nov 3 19:03 atac_possorted_bam.bam
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 2745304 Nov 3 19:03 atac_possorted_bam.bam.bai
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 198607442 Nov 3 19:20 cloupe.cloupe
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:20 filtered_feature_bc_matrix
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 20817222 Nov 3 19:08 filtered_feature_bc_matrix.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 67339148 Nov 3 19:06 gex_molecule_info.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1900615306 Nov 3 19:05 gex_possorted_bam.bam
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 3803768 Nov 3 19:05 gex_possorted_bam.bam.bai
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 50726847 Nov 3 19:17 per_barcode_metrics.csv
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov 3 19:20 raw_feature_bc_matrix
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 25866298 Nov 3 19:07 raw_feature_bc_matrix.h5
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1808 Nov 3 19:17 summary.csv
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 4290262 Nov 3 19:19 web_summary.html
```

ATAC peak outputs

RNA GEX outputs

# Output: cellranger arc

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123/samp1_arc/outs/analysis/  
total 2  
drwxr-x--- 4 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 19:20 clustering  
drwxr-x--- 4 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 19:20 dimensionality_reduction  
drwxr-x--- 2 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 19:20 feature_linkage  
drwxr-x--- 3 t.cri.biowksp01 t.cri.biowksp01 4096 Nov  3 19:20 tf_analysis
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123/samp1_arc/outs/filtered_feature_bc_matrix  
total 47360  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 28597 Nov  3 19:08 barcodes.tsv.gz  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1999463 Nov  3 19:08 features.tsv.gz  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 46156799 Nov  3 19:08 matrix.mtx.gz
```

```
[t.cri.biowksp01@cri22in002 ~]$ ll analysis_results_arc_v123/samp1_arc/outs/raw_feature_bc_matrix  
total 57728  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1584261 Nov  3 19:07 barcodes.tsv.gz  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 1999463 Nov  3 19:07 features.tsv.gz  
-rw-r----- 1 t.cri.biowksp01 t.cri.biowksp01 55227506 Nov  3 19:07 matrix.mtx.gz
```



# OUR AWESOME TEAM

Bioinformaticians



**Mengjie Chen, PhD**  
Faculty Director



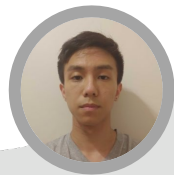
**Wenjun Kang, MS**  
Technical Director



**Jason Shapiro, PhD**



**Diana Vera Cruz, PhD**



**Evan Wu**



**Zhongyu Li, MS**



**David Tieri, PhD**



**Yan Li, PhD**  
Associate Director



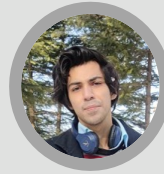
**Houxiang Zhu, PhD**



**Qiaoshan Lin, PhD**



**Yildiz Koca, PhD**



**Mannan Bhola, MS**



**Geetha Priyanka, MS**



## Genomics and proteomics data analysis

BiCF applies appropriate and state-of-the-arts statistical and bioinformatic methodologies to analyze genomics data generated from standard and emerging assays.



## Consulting, grant writing and training

BiCF provides consulting services for experimental design or data analysis; grant writing assistance, including bioinformatics development, cost analysis, and documentation of tools to complete the research.



## Data management system development

BiCF offers enterprise solutions for project and study management, for data production, sharing and integration.

Contact us: [bioinformatics@bsd.uchicago.edu](mailto:bioinformatics@bsd.uchicago.edu)

Submit a project request: <https://biocore.cri.uchicago.edu/>



THE UNIVERSITY OF  
**CHICAGO**

Biological Sciences  
**Bioinformatics Core**

CENTER FOR RESEARCH INFORMATICS

BIOINFORMATICS CORE

**Bioinformatics Workshop Learning Series**

# Thanks for your attention!

Part 2 (Nov 21<sup>st</sup>): scRNA-seq data integration, downstream data analysis, and visualizations

Yan Li, Ph.D.

yli22@bsd.uchicago.edu

bioinformatics@bsd.uchicago.edu

Feedback survey: <https://redcap.link/CRISeminarSeriesFeedback>

<https://biocore.cri.uchicago.edu/cgi-bin/survey.cgi?id=70>

