Single-cell RNA Sequence Analysis	The Jackson Laboratory
9am-4pm	Thursday & Friday, Dec 1-2
Workshop website	https://smcclatchy.github.io/2022-12-01-single-cell-rnaseq/
Lesson website	https://dmgatti.github.io/SingleCellRNAseq/
Etherpad	https://pad.carpentries.org/2022-12-01-single-cell-rnaseq

Welcome to Single-cell RNA Sequence Analysis
We will use this Etherpad to share links and snippets of code, take notes, ask and answer questions, and whatever else comes to mind.
The page displays a screen with three major parts:

- The left side holds today's notes: please edit these as we go along.
- The top right side shows the names of users who are logged in: please add your name and pick the color that best reflects your mood and personality.
- The bottom right is a real time chat window for asking questions of the instructor and your fellow learners.

Instructors

- Dan Gatti, Computational Sciences
- Dan Skelly, Computational Sciences

Helpers

- Sue McClatchy, Churchill Lab
- Michael Saul, Computational Sciences

Participants
Please sign in below by name and affiliation.

- Colleen Mayberry - Chang Lab (JAX)
- Michael Saul -- CS (JAX MG)
- Kourtney Graham-Bloss Lab (JAX)
- Jiayuan Shi, Ren lab
- Jee Young Kwon - Lee Lab (JGM)
- Jake Beierle (Kumar Lab, JAX)
- Tionna Ouellette- OConnell Lab (JAX)
- Greg Perry- Single Cell Biology Core (JAX BH)
- Austin Korgan (O'Connell lab)
- Dan Cortes (Pera)
- Abdulfatai Tijjani Munger lab
- Cat Witmeyer (Kaczorowski lab)
- Bill Buaas (JMCRS - OEPD)
- Asli Uyar (Carter Lab - JGM)
- Stephanie Boas (Kaczorowski Lab)
- Ashley Olson (Chesler Lab)

- Yehya Barakat (Kumar Lab)

Icebreaker
Turn to a person near you, tell them your name and describe one kind thing you have done for someone recently (this someone could be yourself). Write your partner's name and kind thing that they did below.

- Tionna met Greg-bought land recently!
- Jee Young Kwon - Cooked for her daughter!
- Jake gave feedback to a coworker
- Greg met Tianna- Tianna gave away a cross stich project for no charge because of the holodays.
- Colleen - bought someone's coffee
- Kourtney- Made a cake for a Master's thesis defense
- Stephanie - made some PowerPoint slides for a committee

Austin met Dan G. who put up our help signs to the training room
Dan Kotes - visited someone to grab dinner tgether and to cheer them up
Abdulfatai - Going to bed earlier
Asli, took care of herself
Jiayuan helped a lab member move to a new apartment.
Bill met Cat.. Cat did a nice thing by talking to me :)
Yehya - Helped my brother with his project

Setup instructions
https://dmgatti.github.io/SingleCellRNAseq/setup.html

Introduction to Single Cell RNA-seq (
https://dmgatti.github.io/SingleCellRNAseq/01-introduction/index.html)

Challenge
For each of these scenarios, choose between using bulk RNA-Seq and scRNA-Seq to address your problem.

1. Differentiation of embryonic stem cells to another cell type

+1 +1single cell RNA seq
scRNAseq+1+1
ScRNAseq+1
scRNA
scRNAseq
scRNAseq
scRNA seq
scRNA
scRNAseq

1. Studying aging with a specific focus on the senescence-involved (e.g. ref) gene *Cdkn2a*

+1 Bulk RNAseq
Bulk RNAseq
bulk+1
Bulk
Bulk RNAseq

Bulk RNA seq
RNAseq
Depends on the question; if you want to separate by cell type, scRNA, but if you just want to measure changes in expression, bulk, +1

1. Studying variation in vaccine response by profiling peripheral blood mononuclear cells (PBMCs)
   –

+1Bulk+1
+1 scRNAseq
Bulk
scRNAseq
sc
scRNA seq
scRNA
scRNA
scRNA

1. Doing functional genomics in a non-model species

+1+1+1+1Bulk
Bulk
Bulk RNAseq
Bulk RNA seq
Bulk

1. Studying micro RNAs
    1. +1+1+1Bulk

Bulk
Bulk
Bulk
Bulk RNAseq
Bulk +1+1+1
not sure
bulk

1. Performing gene expression quantitative trait locus (eQTL) mapping
    1. +1+1Bulk
    2. +1, scRNA (depends on question)
    3. Bulk+1
    4. Bulk

Bulk
Bulk RNAseq
Bulk RNA seq

Overview of scRNA-seq Data (https://dmgatti.github.io/SingleCellRNAseq/03-Overview-scRNA-seq-Data/index.html)

Challenge 1
1). R has a function called file.size. Look at the help for this function and get the size of each of the files in the mouseStSt_invivo directory. Which one is the largest? Copy-paste your code and answers below.

346524073 - file.size("data/mouseStSt_invivo/matrix.mtx.gz")

346524073 matrix
file.size('data/mouseStSt_invivo/annot_metadata.csv')
file.size('data/mouseStSt_invivo/annot_metadata_first.csv')
file.size('data/mouseStSt_invivo/barcodes.tsv.gz')
file.size('data/mouseStSt_invivo/features.tsv.gz')
file.size('data/mouseStSt_invivo/matrix.mtx.gz') -->

Challenge 2
How many mice were used to produce this data? The “sample” column contains the mouse identifier for each cell.

Challenge 3
How many cells are there from each mouse?

Quality Control of scRNA-Seq Data (
https://dmgatti.github.io/SingleCellRNAseq/04-Quality-Control/index.html#doublet-detection)

Challenge 1:
How would filtering genes too strictly affect your results? How would this affect your ability to discriminate between cell types?
It may elimenate any potential small populations that can express these genes. It may also create better sptatial maps.
If you have small cell population that is the only one that expresses certain genes, you could filter this population out if you're criteria are too stringent
If you filter too strictly, you might miss out on rarer and important cell types. It would affect ability to discriminate between cell types by decreasing the overall number of cell types that you identify.
It could eliminate many rare cell types.I think it could also be a detriment to determining the overall cell types because it may be more general with fewer specific markers.
In studies of cells with rare populations, like resident immune cells, you risk missing informative markers of celltypes that can be relevant to your study.
depending on what you want, being too strict might remove rare populations

Challenge 2:
What total count threshold would you choose to filter genes? Remember that there are 47743 cells.
I think it depends on the type of tissue you are studying and what others in the field have discovered working with cells in that region.
Do we know what percentage breakdown we would predict between distinct populations in our dataset? This could potentially be informed by prior studies, ie FACS/Flocytometery data to estimate cell population sizes.
I might set a threshold of .1% of the total numbers of cells (47 or 48).

Day 2

Sign In:

Austin Korgan
Kourtney Graham
Jake Beierle
Bill Buaas
Colleen Mayberry
Michael Saul
Jee Young Kwon
Tionna Ouellette
Greg Perry
Asli Uyar
Abdulfatai Tijjani
Cat Witmeyer
Stephanie Boas