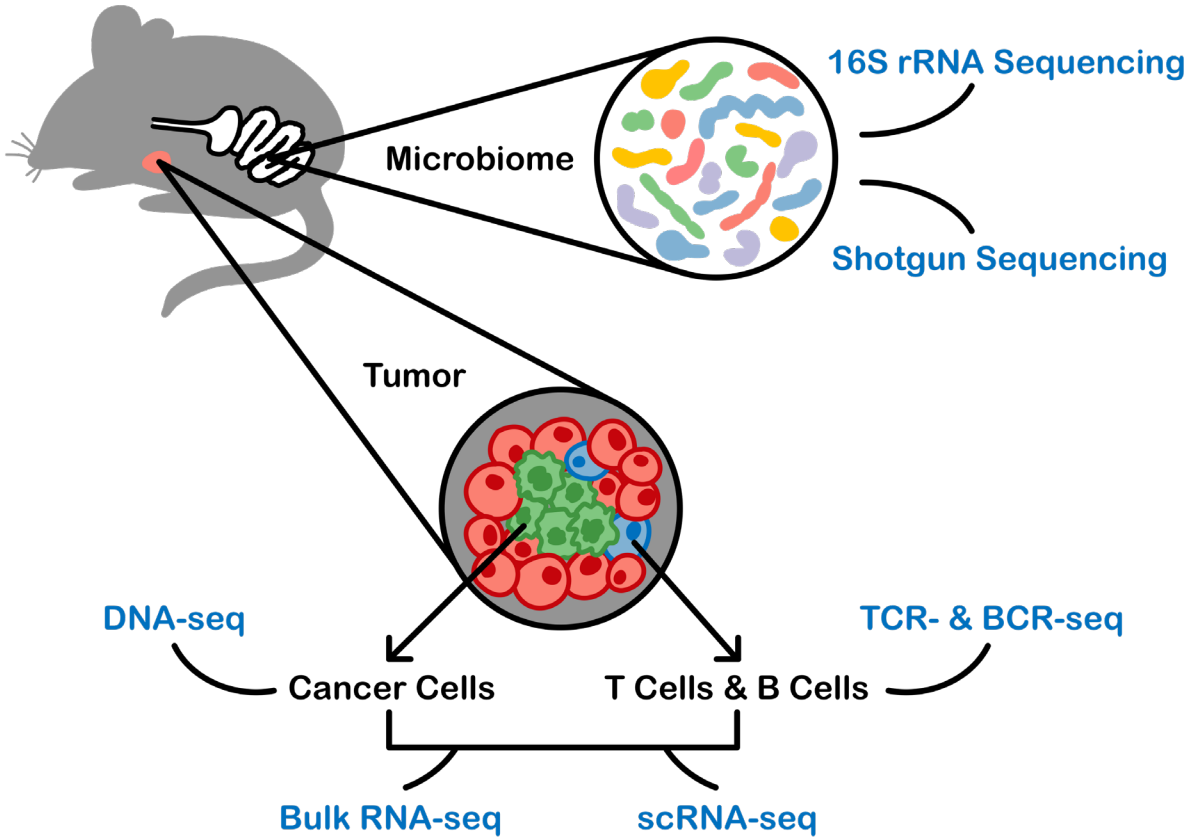
Duke 2022 MIC Course



Course faculty and staff

- **Faculty**
  - Georgia Marie Beasley, MD, MHS
  - Cliburn Chan, PhD
  - Raluca Gordan, PhD
  - Josh Granek, PhD
  - Janice McCarthy, PhD
  - Smita Nair, PhD
  - Kouros Owzar, PhD
  - Greg Palmer, PhD
  - Pixu Shi, PhD
  - Jichun Xie, PhD
- **Teaching Assistants**
  - Marissa Lee, PhD
  - Xiaodi Qin, MB
  - Tyler Schappe, MS
- **Course Evaluation**
  - Jennifer Hill, Ed.D.
- **Experimental Work**
  - Ken Young
  - Keith Laemont
- **Administration**
  - Tasha Allison
  - Tim Durning
  - James Thomas

What biological knowledge will you gain?

- Background knowledge of cancer immunology
- Background knowledge of the microbiome and its role in cancer
- Principles of common assays in cancer immunotherapy research
- Role of high-throughput sequencing in modern cancer immunology research
- How to read a research paper in cancer immunotherapy

What bioinformatics skills will you learn?

- Pre-process sequencing reads from single-cell sequencing RNA (scRNA) assays (raw reads to cell level gene counts)
- Quality control/assessment of scRNA data
- Clustering and cell type inference
- Differential expression analysis (gene and pathway level)
- Pseudo-time trajectory and cell differentiation analysis
- Interpret the results within a proper statistical framework

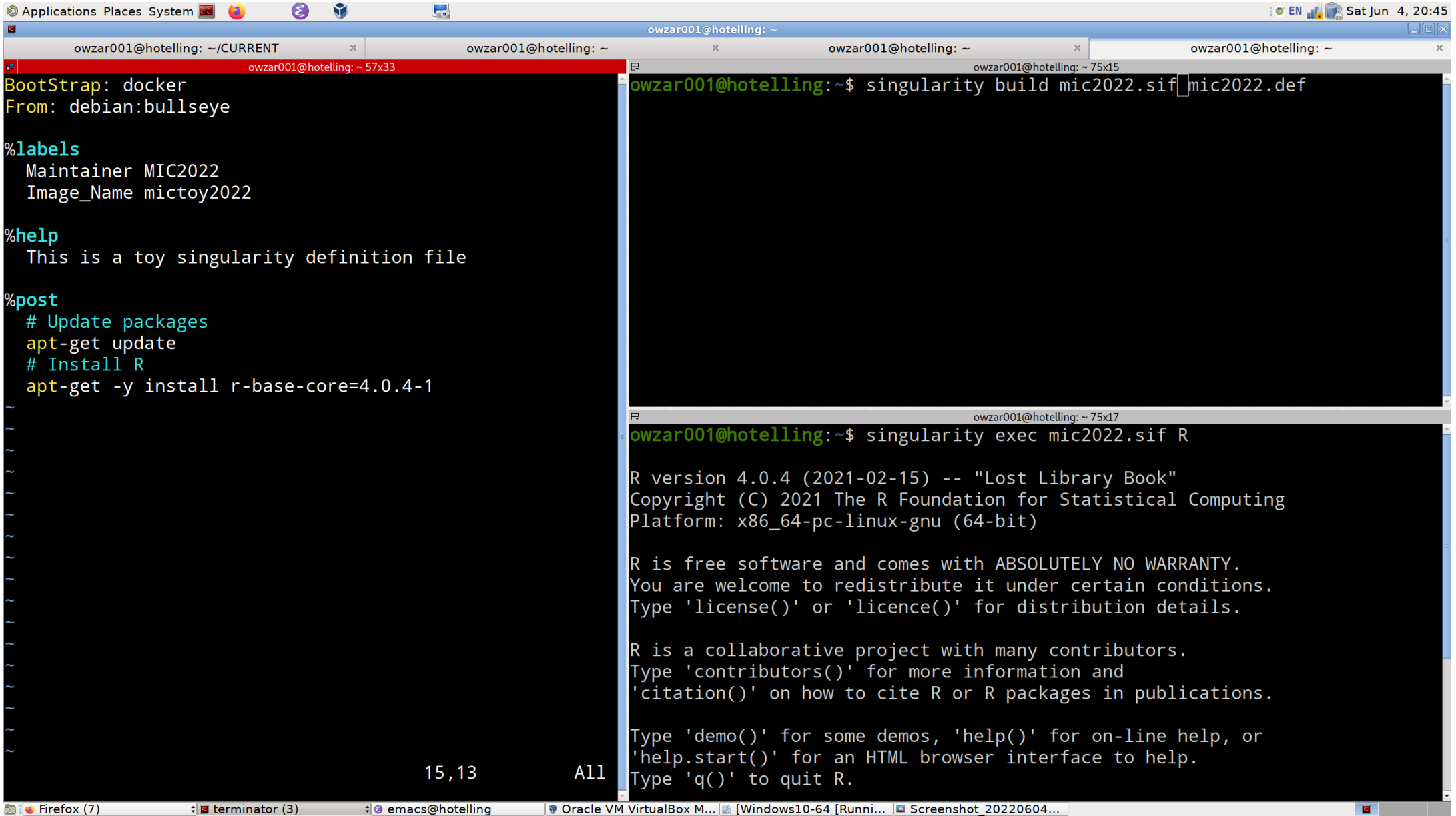
What computing skills will you learn?

- RStudio notebooks to facilitate literate programming
- Use of computing containers
- git and gitlab for source code management
- UNIX command line
- Idiomatic modern R and ”tidy data” to manipulate and visualize data
- R/Bioconductor genomics packages to analyze data

Beyond the mechanics of data analysis

- Computational biology concepts and algorithms
- Statistics: Concepts, limitations, abuse
- Reproducible analysis
- Working in a virtual and containerized computing through Open OnDemand

Containerized computing (toy example)



Data analysis task

Task: Summarize the mean expression levels for genes 1 and 2 by mutation status (WT vs MT)

```
i geno  gene1  gene2
1 WT    0.497  0.445
2 WT    0.561  0.621
3 WT    0.956  1.2
4 MT    2.53  12.4
5 MT    9.92   1.33
6 MT    1.31   0.568
```

A messy approach (messy programming)

```
# Subset data set
x0 <- mydat[mydat$geno == "WT",]
x1 <- mydat[mydat$geno == "MT",]
```

Find the error!

```
# Mean expression of gene 1 in WT
mean(x0$gene1)
 0.6713333
# Mean expression of gene 1 in MT
mean(x1$gene1)
 4.586667
# Mean expression of gene 2 in WT
mean(x0$gene2)
 0.7553333
# Mean expression of gene 2 in MT
mean(x1$gene1)
 4.586667
```

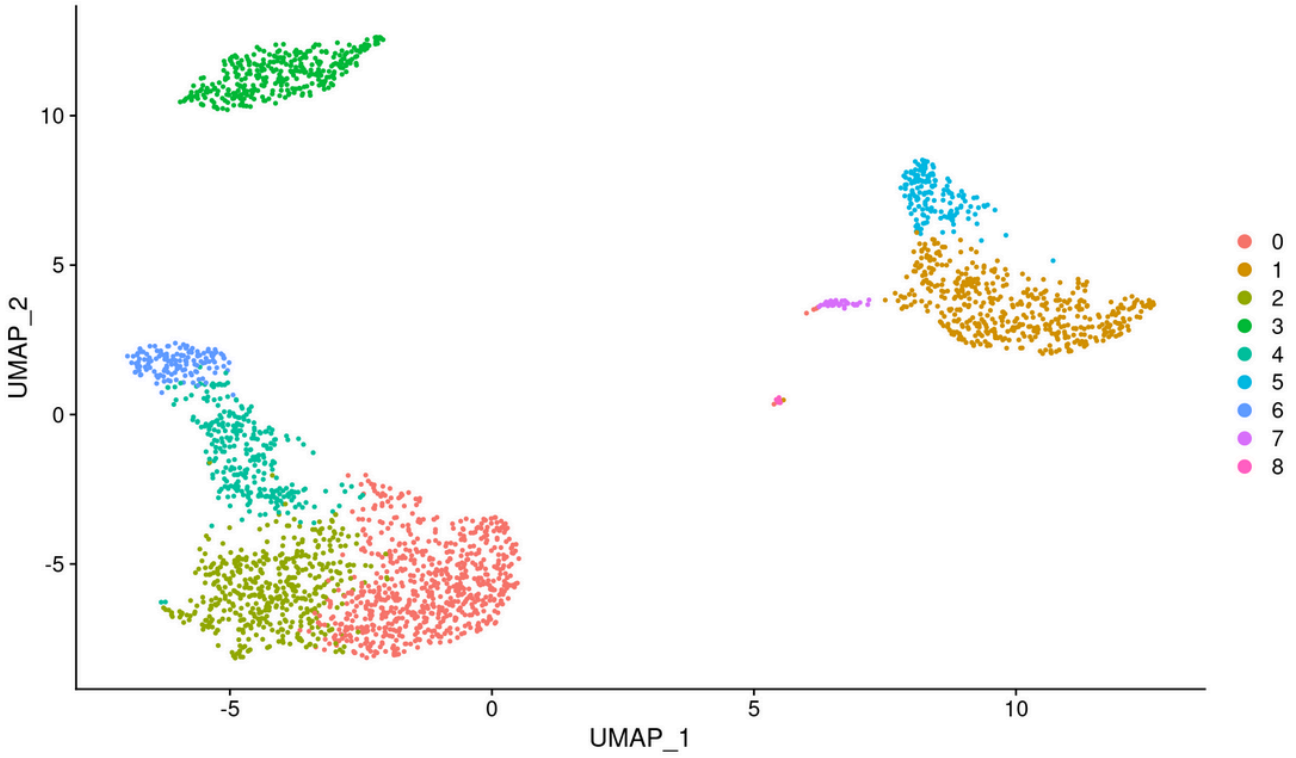
The tidyverse approach (tidy programming)

```
mydat %>%
     dplyr::group_by(geno) %>%
     dplyr::summarize_at(vars(gene1, gene2), mean)

# A tibble: 2 × 3
geno  gene1 gene2
<chr> <dbl> <dbl>
1 MT    4.59  4.77
2 WT    0.671 0.755
```

Basic illustration

```
DimPlot(pbmc, reduction = "umap")
```

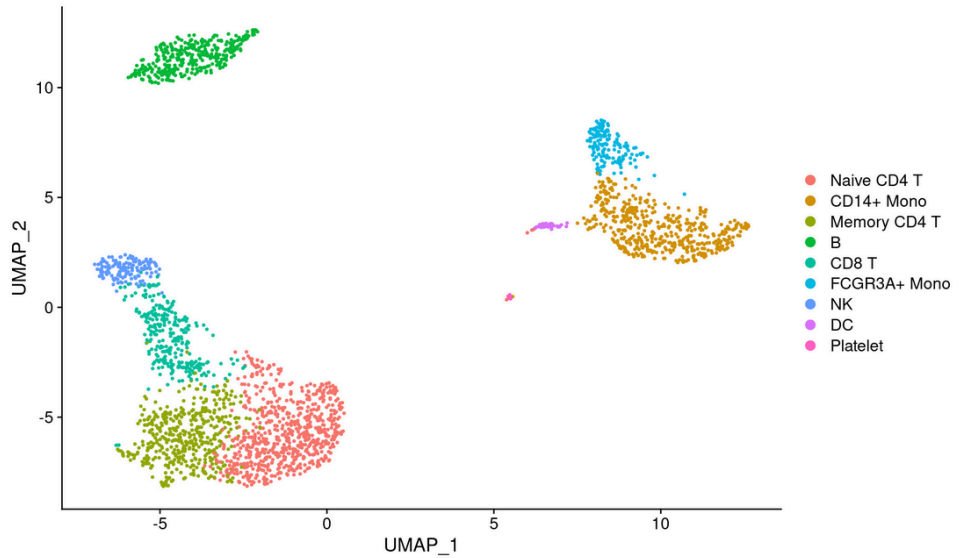


Customized illustration

```
# Assign cell type identity to clusters
new.cluster.ids <- c("Naive CD4 T", "CD14+ Mono", "Memory CD4 T", "B", "CD8 T",
                     "FCGR3A+ Mono", "NK", "DC", "Platelet")
names(new.cluster.ids) <- levels(pbmc)
pbmc$cell_types <- new.cluster.ids[pbmc$seurat_clusters]
# Rename cell identity
pbmc <- RenameIdents(pbmc, new.cluster.ids)

# Alternatively, we can rename the idents in the following way,
# Idents(pbmc) <- "cell_types"

# Plot
DimPlot(pbmc, reduction = "umap")
```



What do these results tell me (other than P-value column)?

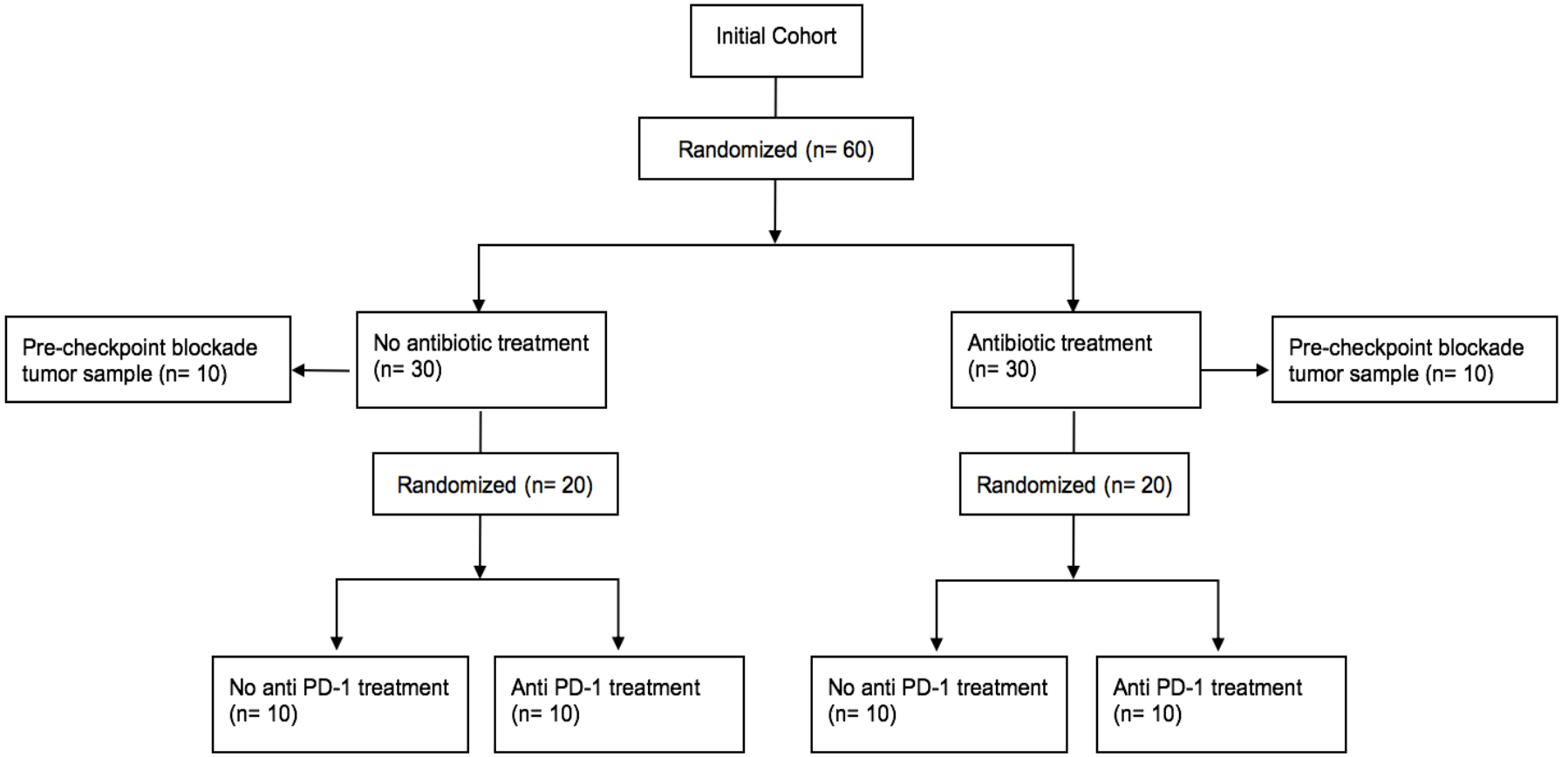
```
Idents(pbmc) <- "seurat_clusters"
cluster2.markers <- FindMarkers(pbmc, ident.1 = 2, ident.2 = NULL, min.pct = 0.25)
head(cluster2.markers)
```

```
##              p_val avg_log2FC pct.1 pct.2    p_val_adj
## IL32 2.593535e-91  1.2154360 0.949 0.466 3.556774e-87
## LTB  7.994465e-87  1.2828597 0.981 0.644 1.096361e-82
## CD3D 3.922451e-70  0.9359210 0.922 0.433 5.379250e-66
## IL7R 1.130870e-66  1.1776027 0.748 0.327 1.550876e-62
## LDHB 4.082189e-65  0.8837324 0.953 0.614 5.598314e-61
## CD2  5.526492e-61  1.2392186 0.657 0.245 7.579031e-57
```

COVID-19 Pandemic: impact

- The program was designed to be held in-person on the Duke campus
- The program was designed to teach integrative genomics from data generated from a prospective experiment
- The impact of the pandemic:
  - Transition to a remote teaching model (through zoom)
  - Use of public data rather than course experiment data

Course experiment: Design



Course Experiment: Assays

	Stool	Tumor	Spleen	Tail
Bulk RNA-Seq		💥		
scRNA-Seq		💥		
Microbiome	💩			
Whole Exome		💥		🫘
TCR/BCR-Seq		💥	🫘	

Table 5: Usage for Assays.

Week	Stool	Tumor	Spleen
1*	💩		
2	💩		
3	💩		
4	💩		
5	💩		
6	💩		
7	💩		
8	💩		
9	💩		
10	💩		
11	💩		
12	💩	💥	🫘
* first collection before antibiotic administration			

Table 6: Sample collection schedule.

2022 Course Data Set

› Cell Rep. 2021 May 11;35(6):109118. doi: 10.1016/j.celrep.2021.109118.

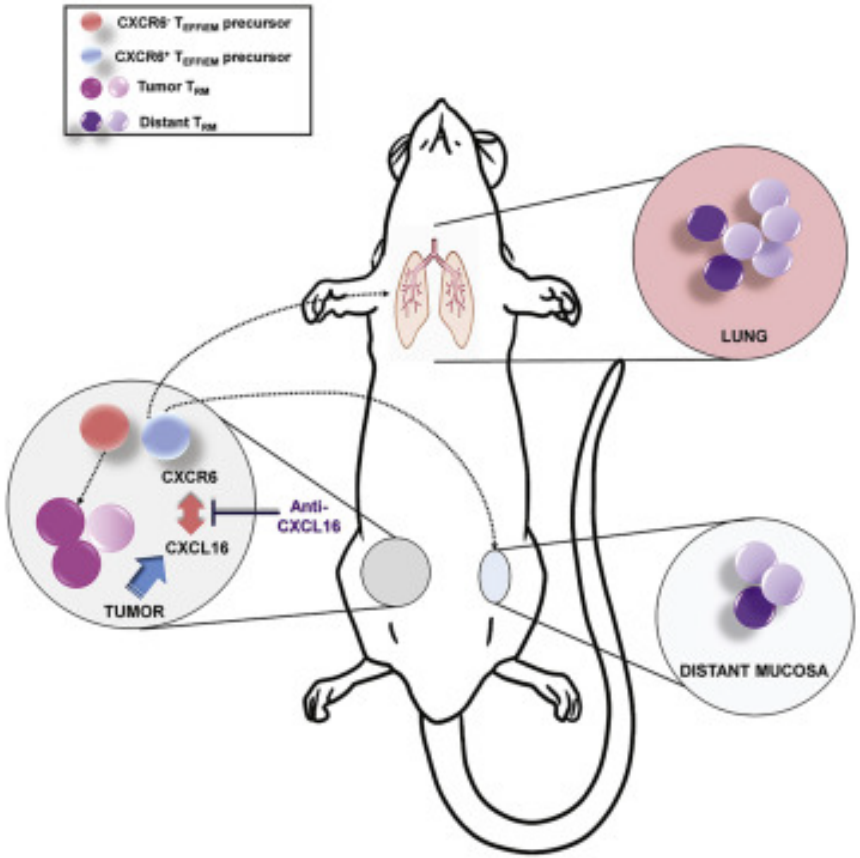
Resident memory T cells in tumor-distant tissues fortify against metastasis formation

Laura S Christian [1], Liuyang Wang [2], Bryan Lim [1], Dachuan Deng [3], Haiyang Wu [3], Xiao-Fan Wang [4], Qi-Jing Li [5]

Affiliations + expand

PMID: 33979626 PMCID: PMC8204287 DOI: 10.1016/j.celrep.2021.109118

Free PMC article



Abstract

As a critical machinery for rapid pathogen removal, resident memory T cells (T_{RM}s) are locally generated after the initial encounter. However, their development accompanying tumorigenesis remains elusive. Using a murine breast cancer model, we show that T_{RM}s develop in the tumor, the contralateral mammary mucosa, and the pre-metastatic lung. Single-cell RNA sequencing of T_{RM}s reveals two phenotypically distinct populations representing their active versus quiescent phases. These T_{RM}s in different tissue compartments share the same TCR clonotypes and transcriptomes with a subset of intratumoral effector/effector memory T cells ($T_{Eff/EM}$s), indicating their developmental ontogeny. Furthermore, CXCL16 is highly produced by tumor cells and CXCR6$^-$ $T_{Eff/EM}$s are the major subset preferentially egressing the tumor to form distant T_{RM}s. Functionally, releasing CXCR6 retention in the primary tumor amplifies tumor-derived T_{RM}s in the lung and leads to superior protection against metastases. This immunologic fortification suggests a potential strategy to prevent metastasis in clinical oncology.

Keywords: TCR-β repertoire sequencing; breast cancer; metastasis; ontogeny; resident memory T cells; single-cell RNA sequencing; tumor immunology.

2022 Course Schedule

- Two-week course
  - 06/06/2022-06/17/2022 (Mon-Fri)
  - Morning sessions: 0900-1015; 1030-1145
  - Afternoon sessions:1315-1430; 1445-1600
  - Lunch 1145-1315
  - Optional office hours (1600-1700)
- Week 1 (preparatory material and methods)
- Week 2 (bioinformatics, statistics and computing specific to analysis single-cell RNA-Seq data; data analysis and presentation)

Week 1 (Preparatory)

- Introduction to computing environment
- Introduction to the UNIX environment
- Introduction to R environment and programming
- Elements of statistical inference
- High-throughput sequencing background
- Primers on microbiome and cancer immunology
- Journal article (Christian *et al*; 2021) review
- https://gitlab.oit.duke.edu/mic-course/2022-mic/-/blob/main/admin/week1_schedule.md

Week 2 (Assay Specific)

- Computational biology and bioinformatics for sequencing data
- Pre-processing of single-cell sequencing data (demultiplexing, alignment, mapping; raw reads to cell level gene counts)
- Quality control and assessment
- Clustering and cell type inference
- Differential expression analysis
- Pseudo-time trajectory and cell type differentiation
- Data Analysis/Presentations (Friday; will ask for volunteers)
- https://gitlab.oit.duke.edu/mic-course/2022-mic/-/blob/main/admin/week2_schedule.md

Future years

- Year 3: 16S and shotgun metagenomics sequencing
- Year 4: Paired tumor/germline whole-exome DNA-Seq
- Year 5: TCR single-cell sequencing

Plan for today

- Questions
- Quick Introduction (all)
- Introduction to the course computing environment (Janice McCarthy)
- Lecture on reproducible analysis
- Introduction to UNIX

Questions

- In Zoom
  - Use chat feature
  - Unmute yourself
- By email
  - miccourse@duke.edu
- TA Office hours
  - Times will be posted on gitlab page

Be sure to regularly check your email for announcements

Acknowledgements

- Funding:
  - Training Program in Bioinformatics at the Intersection of Cancer Immunology and Microbiome
  - National Cancer Institute (NCI)
  - Education Projects (R25)
  - 1R25CA244070
- Administrative:
  - Office of Cancer Research Career Development, Duke Cancer Institute
  - Duke Department of Biostatistics and Bioinformatics
- Computing resources:
  - Duke Center for Human System Immunology (CHSI)
- Data resource:
  - Liuyang Wang, PhD
  - Qi-Jing Li, PhD

Welcome