RNA-seq miniCURE

January, 2023

Table of Contents

# About this Course

This miniCURE allows students to develop a project using RNA-seq datasets to introduce how model organisms, high-throughput sequencing, and the scientific process are used in both basic and clinical research.



Skills Level

*Genetics*  
**Novice**: Introduction to [central dogma of molecular biology](https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology)

*Programming skills*  
**Novice**: No programming experience needed

Learning Objectives

* Learn about model organisms
* Explore high-throughput sequencing datasets
* Practice the [scientific process](https://www.biointeractive.org/classroom-resources/how-science-works)

C-MOOR Collection

More coming soon!

# Onboarding

## 0.1 Join SciServer

*Providing access to big data resources to researchers worldwide*

### 0.1.1 Purpose

In this course we will use the online SciServer platform to do some data analysis for your research project. The purpose of this assignment is to register for a SciServer account, and then to inform the instructor of your username so that you can be added to the SciServer group for this course and access course materials.



### 0.1.2 Learning Objectives

1. Create an account on SciServer
2. Confirm your email address
3. Share your username with your instructor

### 0.1.3 Introduction

SciServer is an online platform for doing scientific data analysis. It is used by scientists studying astronomy, biology, oceanography, and more, and is free as long as you are using it for scientific research. Using SciServer means you do not need a fancy computer or need to install any special programs on your computer, you can just log in with your internet browser to start doing research. For this course, we have set up SciServer with customized collections of programs for RNA-seq analysis, as well as the data that we’ll be analyzing. Once you sign up for SciServer and are added to the group for this course, you will be able to access these tools and begin your data analysis journey!

### 0.1.4 Instructions

#### 0.1.4.1 Create an account on SciServer

This video ([video](https://link.c-moor.org/video-join-sciserver))([slides](https://docs.google.com/presentation/d/1kxbnBLoRsdPW4ZkjwNsAHS1XFPuJpQZ8I1aVqyZISW0)) shows you how to create a SciServer account. You can follow along with the video, or follow the steps below.

1. Open [sciserver.org](https://www.sciserver.org/) in a web browser
   1. It is a good idea to bookmark this page so that you can easily access it throughout the course.
2. Click “Login to SciServer”
3. Click “Create a new account”
4. Enter a User name, Email, etc. and click “Create account”
   1. Note that you cannot change your username

#### 0.1.4.2 Confirm your email address

1. **Important!**: Click the verification link in your email inbox.
   1. If you do not verify your account you will get locked out and will need to contact your instructor to unlock your account.
   2. If you do not see an email, try checking your spam.
2. After clicking the verification link, confirm that your username appears in the upper right hand corner.

##### 0.1.4.2.1 Resources

* [sciserver.org](https://www.sciserver.org/)
* [How to add a bookmark in Chrome](https://support.google.com/chrome/answer/188842)
* [SciServer help page](https://www.sciserver.org/support/how-to-use-sciserver/)

#### 0.1.4.3 Share your username with your instructor

1. Fill out [this form](https://docs.google.com/forms/d/e/1FAIpQLSdJva363KdHIVxI0jWBZzNdhz2M-u8Be3viKiy0Rboyzy4PPQ/viewform) with your SciServer username.

### 0.1.5 Footnotes

#### 0.1.5.1 Contributions and Affiliations

* Katherine Cox, Johns Hopkins University
* Frederick Tan, Carnegie Institution

Last Revised: January 25, 2022

## 0.2 First LearnR

*Interactive tutorials introducing various data science concepts*

### 0.2.1 Purpose

The purpose of this assignment is to (1) join the class SciServer group so you can access course materials, and (2) learn how to access the tutorials for this course on SciServer.



### 0.2.2 Learning Objectives

1. Accept invitation to join class SciServer Group
2. Start up a **C-MOOR LearnR** compute container
3. Complete your first LearnR tutorial
4. Delete your C-MOOR LearnR compute container

### 0.2.3 Introduction

Before beginning this assignment, you should have already created a SciServer account and submitted your SciServer username to your instructor. In this assignment you will learn how to set up a “compute container” on SciServer. Compute containers are how you use programs on SciServer. In this course you will use two compute containers: “C-MOOR LearnR” has tutorials that will teach you how to run data analyses, and “C-MOOR R-Studio” is where you will work on your own data analysis projects. This assignment shows you how to set up the C-MOOR LearnR compute container and start up your first tutorial.

### 0.2.4 Instructions

#### 0.2.4.1 Accept invitation to join class SciServer group

This video ([video](https://link.c-moor.org/video-join-sciserver-group))([slides](https://docs.google.com/presentation/d/1codot9UeUO7l0EDcEre7dJgyXurD_xyxpw6IJL_aEjM)) shows you how to join a SciServer group. You can follow along with the video, or follow the steps below.

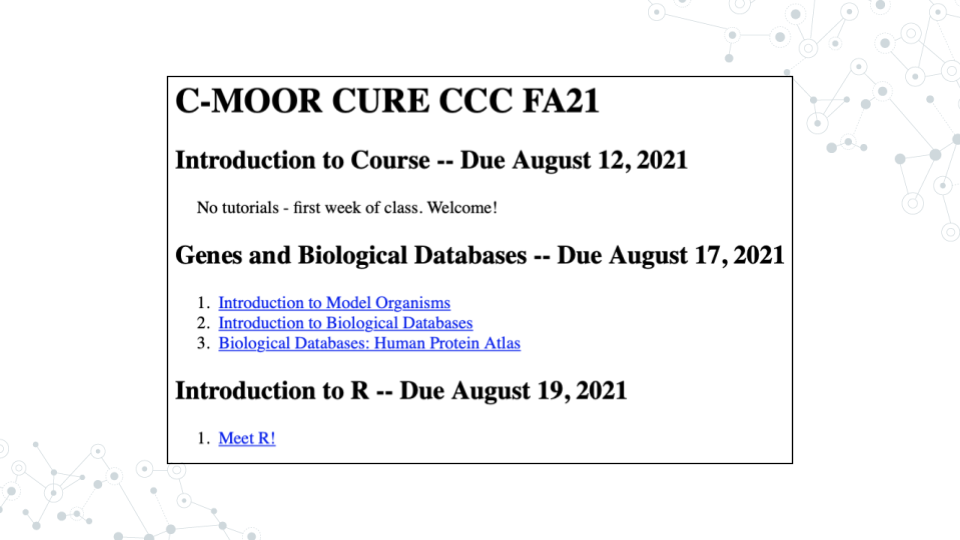
1. Open [sciserver.org](https://www.sciserver.org/) in a web browser and log in to your account.
2. Click “Groups”
3. On the left, you should see a list of all the groups you have joined or been invited to. Click on the name of the group for this course, then click “Accept invitation”.
   1. Your instructor must have your username to invite you to the group. If you do not see an invitation, contact your instructor with your SciServer username.
4. Confirm that you can access course data
   1. On the top menu bar, click “Files”
   2. On the left-hand menu, click “Data Volumes”
   3. Confirm that you see “C-MOOR-Data”
5. Confirm that you can access course computing resources
   1. Click “Home” in the top menu to return to the home page.
   2. Scroll down to the second set of boxes and click “Compute”
   3. Click “Create container”
   4. In the “Compute Image” drop-down menu, confirm that you can see “C-MOOR LearnR” and “C-MOOR R-Studio”
   5. Under “Data Volumes”, confirm that you can see “C-MOOR Data”
   6. You can close the Create Container dialog box (by clicking the “X” in the top right) once you’ve confirmed that you can see the C-MOOR content

##### 0.2.4.1.1 Resources

* [sciserver.org](https://www.sciserver.org/)
* [Get help with SciServer on the C-MOOR Discourse](https://help.c-moor.org/c/help/)

#### 0.2.4.2 Start up a “C-MOOR LearnR” compute container

This video ([video](https://link.c-moor.org/video-sciserver-create-learnr-container))([slides](https://docs.google.com/presentation/d/1Oaq8RzhaDANxkNh-tTKwme7e095pGgoiq5iZHbt7PLg)) shows you how to create and start up a C-MOOR LearnR compute container. You can follow along with the video, or follow the steps below.

1. Open [sciserver.org](https://www.sciserver.org/) in a web browser and log in to your account.
   1. If you are already logged in, click “Home” in the top menu to return to the home page.
2. Scroll down to the second set of boxes and click “Compute”
3. Click “Create container”
   1. Give your container a name. This can be anything you like, but it’s useful if it says something about the purpose of the container so you can tell your containers apart. You could name this container “Tutorials”, since you’ll be using it to access tutorials.
   2. In the “**Compute Image**” drop-down menu, select “**C-MOOR LearnR**”
   3. Under “**Data Volumes**”, check the box next to “**C-MOOR Data**”
   4. Click “Create”. This may take a moment.
4. You should now see a new entry in your list of containers
   1. “Created At” should be a few moments ago.
   2. “Name” should be the name you chose
   3. “Image” should be “C-MOOR LearnR”
5. Start your C-MOOR LearnR container by clicking on its name (whatever name you chose when you created it). This will open in a new tab.
   1. You should see a list of tutorials, organized by date. 
   2. If instead you see an error message, you most likely forgot to check the box next to “C-MOOR Data” when you created the container. 
   3. If you see something else, you may have picked the wrong “Compute Image” from the drop-down menu.

**If anything goes wrong, you can always delete your container by clicking the red “X” in the last column, and create a new container.**

##### 0.2.4.2.1 Resources

* [sciserver.org](https://www.sciserver.org/)
* [Get help with SciServer on the C-MOOR Discourse](https://help.c-moor.org/c/help/)

#### 0.2.4.3 Complete your first LearnR tutorial

1. If you’re not there already, go to the SciServer compute page and start up the C-MOOR LearnR container.
   1. Open [sciserver.org](https://www.sciserver.org/) in a web browser and log in to your account.
   2. If you are already logged in, click “Home” in the top menu to return to the home page.
   3. Scroll down to the second set of boxes and click “Compute”.
   4. Start your C-MOOR LearnR container by clicking on its name.
2. Click on “Biological Databases”. The tutorial will open in a new tab.
3. Complete the tutorial.

##### 0.2.4.3.1 Resources

* [sciserver.org](https://www.sciserver.org/)
* [Get help with SciServer on the C-MOOR Discourse](https://help.c-moor.org/c/help/)

#### 0.2.4.4 Delete your C-MOOR LearnR compute container

Compute containers are meant to be temporary, and you can only have 3 containers total on SciServer. So it’s generally a good idea to clean up after yourself and delete your containers when you’re done using them. Also, in this course, we will be updating the tutorials on the C-MOOR LearnR container, and **you will need to create a new container to get the latest updates.**

**Deleting your container will delete your progress in a tutorial**, so don’t delete the container until you have completed the tutorial and submitted any required items to your instructor. Later on in the course you will learn how to save things permanently on SciServer.

To delete a container:

1. If you’re not there already, go to the SciServer compute page.
   1. Open [sciserver.org](https://www.sciserver.org/) in a web browser and log in to your account.
   2. If you are already logged in, click “Home” in the top menu to return to the home page.
   3. Scroll down to the second set of boxes and click “Compute”.
   4. Start your C-MOOR LearnR container by clicking on its name.
2. Find the container you want to delete.
3. Click on the red “X” in the last column.

### 0.2.5 Footnotes

#### 0.2.5.1 Contributions and Affiliations

* Katherine Cox, Johns Hopkins University
* Frederick Tan, Carnegie Institution

Last Revised: May 13, 2021

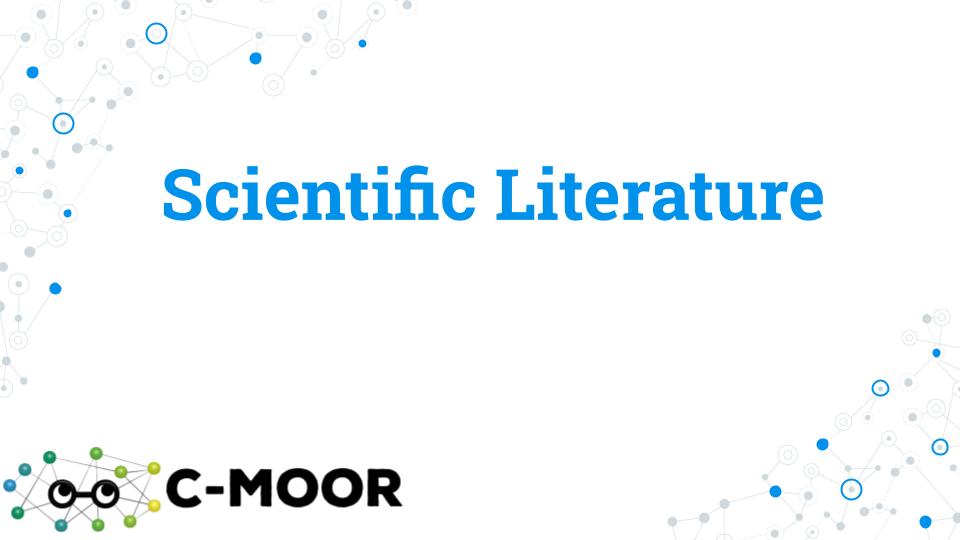
# 1 Scientific Literature

## 1.1 Pre-lab

Familiarize yourself with the paper “[A high-sugar diet produces obesity and insulin resistance in wild-type Drosophila](https://pubmed.gov/21719444)” by Musselman et al., 2011 Dis Model Mech



## 1.2 Lab Slides



[[slides](https://docs.google.com/presentation/d/1uuuAbg_rcfCVohaVrrxfUfszwTDqigzg34qEWfTNeR0)]

## 1.3 Lab Activity: Scientific Literature

### 1.3.1 Instructions

Follow your instructor’s instructions and answer the following questions with your group. In your responses, properly paraphrase (do not copy/paste or inappropriately paraphrase). To receive credit, you must use your own words and sentence structure. Using the same sentence structure (and changing a few words only) is considered plagiarism.

[Link to Musselman et al. 2011](https://journals.biologists.com/dmm/article/4/6/842/3157/A-high-sugar-diet-produces-obesity-and-insulin). You can also download a pdf from the website.

### 1.3.2 Part 1. Overview of the paper

Determine the main objectives and purpose of the paper. Re-read the Abstract (or Summary) and the introduction with your group.

| 1. What is the purpose of the study? |
| --- |
|  |

| 2. What is the hypothesis? |
| --- |
|  |

| 3. Describe the knowledge gap. In essence, what did the scientific community not know that this study was trying to answer? |
| --- |
|  |

### 1.3.3 Part 2. Figure Analysis - Methods and Findings

**Methods.** How did the researchers test their hypothesis? Explain in your own words the methods in each figure.

| Figure | Methods |
| --- | --- |
| Fig. 1A-F |  |
| Fig. 2A-D |  |
| Fig. 2E-2I |  |
| Fig. 3A-E |  |
| Fig. 4A-B |  |

**Results.** What are the main findings from each figure?

| Figure | Main Findings |
| --- | --- |
| Fig. 1A-F |  |
| Fig. 2A-D |  |
| Fig. 2E-2I |  |
| Fig. 3A-E |  |
| Fig. 4A-B |  |

### 1.3.4 Part 3. Your Conclusions

| 1. Work with your group and write 3-4 sentences summarizing the main findings/conclusions of the paper based one what you came up with in Part 3. |
| --- |
|  |

| 2. Talk with your group and come up with a list of 2-3 limitations. |
| --- |
|  |

### 1.3.5 Part 4. The Author’s Conclusions (Discussion Section)

| 1. Read the discussion section and compare it to what your group came up with in Part 4. How similar are the conclusions you came up with to the authors? Were there any differences in interpreting the data? Different conclusions based on those data? |
| --- |
|  |

| 2. What do you think about the conclusions made by the authors? Do you think they are over-stating their findings? Do you think their conclusions are accurate and appropriate? |
| --- |
|  |

### 1.3.6 Part 5. The Future

| 1. Scientific work builds on previous studies. What do you believe could be the next step to further the work these researchers did? |
| --- |
|  |

| 2. Do you believe further research in this area may benefit society? Can we build on what this study found? |
| --- |
|  |

### 1.3.7 Footnotes

#### 1.3.7.1 Resources

* [Google Doc](https://docs.google.com/document/d/1QmuNMHKiHRGXohS0P9Pojc3zBvUZDG8L)

#### 1.3.7.2 Contributions and Affiliations

* Stephanie Coffman, Clovis Community College

Last Revised: February 15, 2022

# 2 Model Organisms and Databases

## 2.1 Pre-lab: Model Organisms



Image credit: [Max Westby](http://cubocube.com/dashboard.php?a=1179&b=1228&c=10). Some of the most important genetic model organisms in use today. Clockwise from top left: yeast, fruit fly, arabidopsis, mouse, roundworm, zebrafish. License: [CC ANS 2.5](https://creativecommons.org/licenses/by-nc-sa/2.5/)

### 2.1.1 Purpose

The two tutorials in this pre-lab will familiarize you with the concepts of model organisms, with an emphasis on Drosophila. In Lab 8, this information will help you look up information about Drosophila genes.

### 2.1.2 Learning Objectives

1. Explain the importance of model organisms and identify some of their desirable characteristics.
2. Describe the usefulness of Drosophila as a model system.

### 2.1.3 Introduction

Scientists frequently use a few specific organisms, called “model organisms” for their experiments. This tutorial will introduce you to a few of the most popular model organisms and will discuss why these organisms were chosen and what they are useful for.

The fruit fly (Drosophila melanogaster) is a popular model organism used to study a wide range of biological questions. The second tutorial will introduce you to some of the types of research being conducted with fruit flies, give a brief overview of Drosophila biology, and show you what it’s like to work with Drosophila in the lab.

### 2.1.4 Activity 1 - Model Organisms

Estimated time: 15 min

#### 2.1.4.1 Instructions

1. [Click here to open the Model Organisms Tutorial.](https://clovis.shinyapps.io/BIOL11A_Model_Organisms)
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.
3. As you complete the tutorial, answer the questions below.

#### 2.1.4.2 Questions

| 1A. Explain what a “model organism” is and why they are useful. |
| --- |
|  |

| 1B. Define ortholog and explain how model organisms can be used to understand human genes. |
| --- |
|  |

| 1C. Name 4 commonly used model organisms |
| --- |
|  |

### 2.1.5 Activity 2 - Drosophila melanogaster

Estimated time: 15 min

#### 2.1.5.1 Instructions

1. [Click here to open the Drosophila melanogaster Tutorial.](https://clovis.shinyapps.io/BIOL11A_Drosophila/)
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.
3. As you complete the tutorial, answer the questions below.

#### 2.1.5.2 Questions

| 2A. Provide 3 reasons why fruit flies are useful for scientific research. |
| --- |
|  |

| 2B. List 3 ways in which fruit flies are similar to humans. |
| --- |
|  |

| 2C. Compare and contrast the fruit fly genome to the human genome. |
| --- |
|  |

| 2D. Briefly describe the fruit fly life cycle. |
| --- |
|  |

### 2.1.6 Footnotes

#### 2.1.6.1 Resources

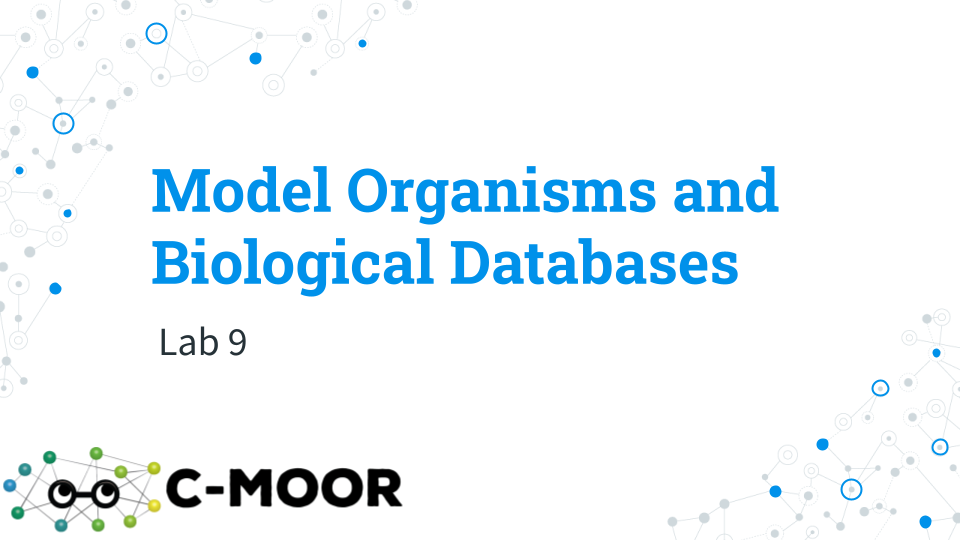
* [Google Doc](https://docs.google.com/document/d/1vFhm2XLMO9vjMDNT6CxZJ4VJuinw0BEo)

#### 2.1.6.2 Contributions and Affiliations

* Stephanie R. Coffman, Ph.D.

Last Revised: February 2022

## 2.2 Lab Slides



[[slides](https://docs.google.com/presentation/d/1kt0lW4D8AWqQm1j6FMo0rRDGM_G3zkLmPy3EZNDg3s0)]

## 2.3 Lab Activity: Biological Databases

### 2.3.1 Purpose

In this activity, students will learn to search the same online databases used by scientists to collect information about a set of genes and present them to your group. This will give you an opportunity to engage in inquiry-based learning and apply the concepts in molecular biology and genetics from this course.

### 2.3.2 Learning Objectives

1. Use online databases to look up information about a gene.

### 2.3.3 Activity 1 - Databases

Estimated time: 15 min

#### 2.3.3.1 Instructions

1. Follow [these directions](https://docs.google.com/document/d/1nQ-wd4hX0Xtd4mfloXcpbbtK_CqDzY874oqmBYql1M8) to launch your first LearnR tutorial: Biological Databases Tutorial.
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.
3. As you complete the tutorial, fill in the table below. This will help you know which database to go back to later on.

#### 2.3.3.2 Questions

Table 1. Databases

| Database | Description |
| --- | --- |
| GenBank |  |
| OMIM |  |
| Human Protein Atlas |  |
| PDB |  |

### 2.3.4 Activity 2 - FlyBase

Estimated time: 15 min

#### 2.3.4.1 Instructions

1. Work through the FlyBase Tutorial in SciServer.
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.

#### 2.3.4.2 Questions

| 2A. What is one question you have about using FlyBase? |
| --- |
|  |

| 2B. What is something that surprised you or that you found interesting about using FlyBase? |
| --- |
|  |

### 2.3.5 Activity 3 - Human Protein Atlas

Estimated time: 15 min

#### 2.3.5.1 Instructions

1. Work through the Human Protein Atlas in SciServer.
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.

#### 2.3.5.2 Questions

| 3A. What is one question you have about using HPA? |
| --- |
|  |

| 3B. What is something that surprised you or that you found interesting about using HPA? |
| --- |
|  |

### 2.3.6 Activity 4 - Research a Gene!

Estimated time: 45 min

#### 2.3.6.1 Instructions

1. Before getting started on this activity, your instructor will assign your group a letter that corresponds to a group of 4 genes.

| Group Assigned Letter |
| --- |
|  |

1. [Look up your letter here](https://docs.google.com/spreadsheets/d/1GZtHz2GU3B4KMOd9yuKrcQLJJLn7qlnZgPN5c6SNqGg) and write the names of the four genes your group is assigned at the top of each column in the table below.
2. In your group, assign each student one of the four genes to research.

| Individual Assigned Gene |
| --- |
|  |

1. Use FlyBase to look up the information in Table 2 below.
2. Use HPA to look up the information in Table 3 below.

Table 2. FlyBase Information

| Category | Information |
| --- | --- |
| **General Information** |  |
| Full Gene Name |  |
| FlyBase ID |  |
| Sequence Location |  |
| **Function** |  |
| Biological Process |  |
| Cellular Component |  |
| **Expression Data** |  |
| Anatomical Expression |  |
| Developmental Stage |  |
| **Orthologs** |  |
| Orthologs in other species |  |
| Human Orthologs |  |

Table 3. Human Protein Atlas

| Category | Information |
| --- | --- |
| Function |  |
| Is the gene tissue specific? Which tissue? |  |
| Where is it localized in cells? |  |

### 2.3.7 Activity 5 - Present to your Group

Estimated time: 15 min

#### 2.3.7.1 Instructions

1. Present your gene to your group.
2. Take turns presenting your genes among your group and decide on one gene that you think is the most interesting.

### 2.3.8 Activity 6 - Class Presentation

Estimated time: 30 min

#### 2.3.8.1 Instructions

1. With your group, create a short presentation to present your chosen gene to the class.
2. Your presentation should have about four slides and be thorough:
   1. Slide 1: The GENE you picked to share with your group, your name and date
   2. Slide 2 - 4: Present the information you collected about the gene. For full credit, include relevant images/ diagrams on your slides.
3. One student in the group should post your slides on Canvas to the discussion board. Make sure you mention everyone in your group by name so they also get credit for the presentation.

### 2.3.9 Footnotes

#### 2.3.9.1 Resources

* [Google Doc](https://docs.google.com/document/d/1M7rtDzyGVUxO2GVBp09bTBu4fMDxocTs)

#### 2.3.9.2 Contributions and Affiliations

* Rosa Alcazar, Ph.D., Clovis Community College
* Katherine Cox, Ph.D., John Hopkins
* Stephanie R. Coffman, Ph.D., Clovis Community College

Last Revised: September 2021

# 3 RNA-seq Analysis

## 3.1 Pre-lab: Intro to RNA-seq

### 3.1.1 Purpose

In this pre-lab, students learn about RNA-sequencing so that we can take a closer look at some RNA-seq data in class.

### 3.1.2 Learning Objectives

1. Compare and contrast the genome and the transcriptome
2. Describe the steps involved in RNA-seq
3. Define bioinformatics and its role in biology

### 3.1.3 Introduction

Next-generation DNA sequencing has revolutionized biological research. This tutorial will explain the basic process of next-gen sequencing and will discuss some of the ways it is used in research.

### 3.1.4 Activity 1 - Biotechnology: Next-Gen Sequencing

*Estimated time: 20 min*

#### 3.1.4.1 Instructions

1. Log into SciServer, click on compute and open your C-MOOR LearnR” container.
2. Start the “Biotechnology: Next-Gen Sequencing” tutorial. Visit [SciServer Guides and FAQs](https://help.c-moor.org/t/sciserver-guides-and-faqs/22). if you need to jog your memory on how to do this.
3. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.

#### 3.1.4.2 Questions

| What is Bioinformatics? |
| --- |
|  |

| Briefly describe each of the following steps of next-gen sequencing: |  |
| --- | --- |
| in vivo |  |
| in vitro |  |
| in silico |  |

### 3.1.5 Activity 2 - Biotechnology: RNA-Seq

*Estimated time: 10 min*

#### 3.1.5.1 Instructions

1. Start the “Biotechnology:RNA-Seq” tutorial.
2. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.

#### 3.1.5.2 Questions

| What is Differential Gene Expression? |
| --- |
|  |

| What feature of mRNA allows scientists to specifically isolate mRNA for RNA-seq? |
| --- |
|  |

| Describe the steps in making cDNA from an mRNA. |
| --- |
|  |

| Explain how the number of reads related to gene expression. |
| --- |
|  |

### 3.1.6 Footnotes

#### 3.1.6.1 Resources

* [Google Doc](https://docs.google.com/document/d/1zqTb5HacEqEVhFv4nmRsCB8FSCEMDkP_)

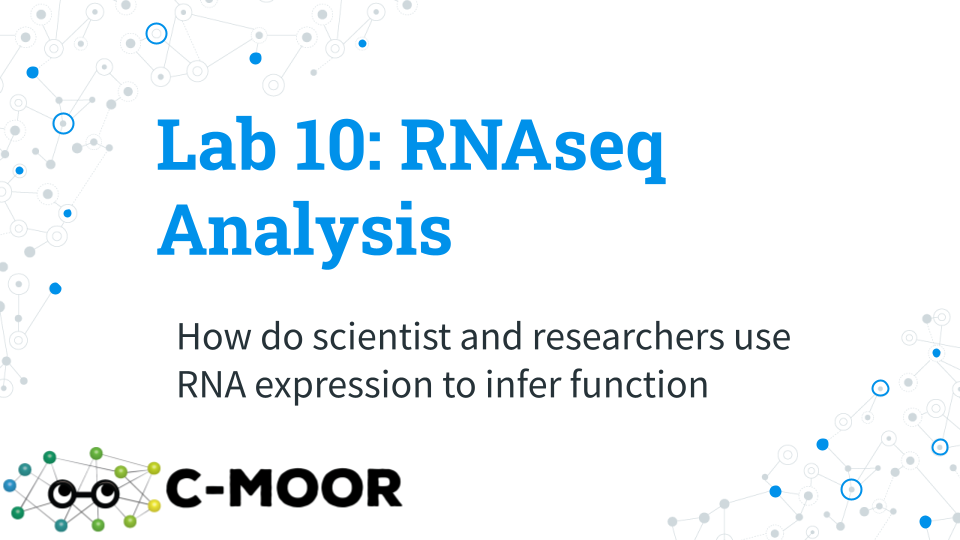
#### 3.1.6.2 Contributions and Affiliations

* Stephanie R. Coffman, Ph.D. Clovis Community College
* Katie Cox, Ph.D. Carnegie Institute at John Hopkins University

Last Revised: March 2022

## 3.2 Lab Slides and Demo

### 3.2.1 Lab Slides



[[slides](https://docs.google.com/presentation/d/1-bQsuEIWO6e0ekb98byMAHR8_l45pp_fsl7NxJIXkbM)]

### 3.2.2 Sort Gene Expression Data Using Spreadsheets



[[slides](https://docs.google.com/presentation/d/1oNO9JFmC8itk3eq6amT95MWvOsr3zQb6LNwZsQ7-i-g)][[test-driveR.tsv](https://drive.google.com/file/d/1kMWQReRTg0FDZDA5abxmkgKxp-7o5IeM)]

## 3.3 Lab Activity: RNA-seq Analysis

### 3.3.1 Purpose

In this lab, students will complete a tutorial on RNA-seq data and learn how to analyze, graph and interpret the data. In the following lab, we will use these skills to compare two RNA-seq data sets to investigate gene expression patterns.

### 3.3.2 Learning Objectives

1. Use R to analyze HTSeq files
2. Create and analyze histograms from HTSeq files

### 3.3.3 Introduction

Today’s lab will investigate how scientists use computer science to analyze RNA-seq data. In general, the sequences are first aligned to a reference genome. For RNA-seq, the sequences will align to exons of the expressed genes. The data you will look at today has already been processed using a program called HTSeq. This program aligns the sequences to the reference genome and counts how many sequences align to each gene, producing files known as HTSeq files. The more sequences that align to the gene, the higher the expression level of the gene. The following tutorial will walk you through how to analyze an HTSeq file using the programming language R.

The RNA-seq libraries from today’s lab are from: [eLife 2013;2:e00886 DOI: 10.7554/eLife.00886](https://elifesciences.org/articles/00886). The paper analyzes genes expression in the drosophila midgut.

### 3.3.4 Activity 1 - Introduction to RNA-seq Data Tutorial

*Estimated time: 20 min*

#### 3.3.4.1 Instructions

1. Log into SciServer, click on compute and open your C-MOOR LearnR” container.
2. Start the “Introduction to RNA-seq Data” tutorial. Visit [SciServer Guides and FAQs](https://help.c-moor.org/t/sciserver-guides-and-faqs/22). if you need to jog your memory on how to do this.
3. To move through the activities click “Continue” at the bottom of the screen. When you are done with a topic, click “Next Topic” to move on.
4. This tutorial has small boxes in which you can enter and run short lines of code to analyze the data.
5. As you work through the tutorial, answer the questions below. When you get to “Try it Out!” move on to Activity 2.

#### 3.3.4.2 Questions

| What are the two columns (V1 and V2) in an HT-seq file? What data is stored in each column? |
| --- |
|  |

| Explain what is readCount and what is GeneID |
| --- |
|  |

| Share a screenshot of the row showing the readCount of the lab gene in the “Reproduce Results for a Single Gene” section and explain in your own words what the code in your screenshot is doing. |
| --- |
|  |

### 3.3.5 Activity 2 - Analyze an HT-seq file

*Estimated time: 15-20 min*

#### 3.3.5.1 Instructions

1. In groups of two, analyze the HTSeq samples assigned to you.

|  | Assigned Sample |
| --- | --- |
| Name |  |
| Name |  |

1. Use the code blocks on the “Try it out!” page to analyze the data.
   1. The codeblocks on the “Try it out!” page, has this code typed out for you:
      1. readCounts <- read.table( "data/FILENAMEHERE.htseq")
   2. Change FILENAMEHERE to the filename for your file. Once you have done this, **readCounts** will have the new HTSeq file loaded into it
      1. Example: readCounts <- read.table( "data/SRR891602.htseq")
2. The code above loads the data set into readCounts. If you run this code alone, nothing will happen. Try requesting some analysis to get a look at the data and answer the questions below.
3. Answer the questions below as you analyze the data. Consult the “Cheat Sheet” to figure out which code to use.

#### 3.3.5.2 Questions

Determine the total reads across all genes, and the mean, median and max read counts for a single sample. Each student reports on one of the samples analyzed.

| Assigned Sample |  |
| --- | --- |
| Total Reads |  |
| Mean Read Count |  |
| Median Read Count |  |
| Max Read Count |  |

Look up the GeneID of the gene you presented on from the Biological Databases Lab. Use the filter command to find the readCount in both samples assigned to your group.

| How many reads does the gene have in your assigned dataset? |
| --- |
| Share a screenshot. |

| How many reads does the gene have in your partner’s dataset? |
| --- |
| Share a screenshot. |

| Compare this number to the mean; is it average, high or low? |
| --- |
|  |

### 3.3.6 Footnotes

#### 3.3.6.1 Resources

* [Google Doc](https://docs.google.com/document/d/1ABmBmDwDtxipxmlfWuVypjIVt2AmTZtY)

#### 3.3.6.2 Contributions and Affiliations

* Stephanie R. Coffman, Ph.D. Clovis Community College
* Rosa Alcazar, PH.D. Clovis Community College
* Katie Cox, Ph.D. Carnegie Institute at John Hopkins University

Last Revised: March 2022

# 4 Differential Gene Expression

# 5 Project Work

Purpose

The purpose of this lab is to gain practice in scientific communication, project exploration and work on your scientific poster!

Learning Objectives

1. Engage with the Scientific Process
2. Communicate your findings
3. Create an outline for further exploratory research
4. Develop your plot images using R code

Introduction

In today’s lab we will learn to develop a scientific idea and convert it into a scientific hypothesis. Remember that a hypothesis needs to be testable and falsifiable. Since your project is *in silico*, your experiments will be *in silico*. You might look at a gene, do some background research on that gene, wonder about the gene function and its relationship to gene expression. Then use this information to develop a hypothesis about its expression and test your hypothesis by analyzing gene expression. The most difficult part of your research will be identifying a gene that is differentially expressed and interesting enough to build a narrative around it.

A few ideas to get you started:

* Research a disease – For example: Look up genes involved in digestive diseases at [NIDDK](https://www.niddk.nih.gov/health-information/digestive-diseases) or [GARD](https://rarediseases.info.nih.gov/diseases/diseases-by-category/6/digestive-diseases)
* Research a gene or gene family – For example: Trypsin, how many different trypsins are in flies?
* Research a gene pathway – For example: How many other genes work with trypsin to digest protein?

Another interesting area may be due to other absorption problems in the gut besides glucose, like issues in lipid or protein metabolism. We read a paper about how flies can be used as a model for diabetes, could it be used as a model for other diseases? Use data to support your conclusions. This is just scratching the surface of the types of projects that can be developed or begun with the information you now have at your fingertips. All good research begins with an observation or question but the best ones have to do with the follow up research.

Instructions

### 5.0.1 Activity 1 - Search for an interesting differentially expressed gene

*Estimated time: 45 min (this is at a minimum to find an interesting gene to pursue)*

1. Log into SciServer, click on compute and open your “C-MOOR LearnR” container.
2. Start the “Differential Gene Expression” tutorial. Visit [SciServer Guides and FAQs](https://help.c-moor.org/t/sciserver-guides-and-faqs/22). if you need to jog your memory on how to do this.
3. This tutorial has small boxes in which you can enter and run short lines of code to analyze the data. Since we are not working directly in RStudio but in a LearnR tutorial environment, you don’t need to know more than a few lines of code to be able to ask some very interesting questions.
4. **Use the Notes section below to copy and paste important blocks of code that you can refer back to later.**
5. We will go through an example and you can use this to get other ideas for your project.

#### 5.0.1.1 Questions

| Here is an example and therefore this is not available for your project as we want different projects. Look up the gene trypsin. How many trypsin genes are in drosophila? What are their FBgn IDs? |
| --- |
|  |

| Use the code to find the expression of trypsin across all regions. What can you conclude? What do you think it suggests about the function of the genes and the regions? |
| --- |
|  |

| Compare the expression between the two highest regions. Record your observations and write two short paragraphs. |
| --- |
| I was surprised by … |
|  |
| I’m curious about … |
|  |

| Notes |
| --- |
|  |

### 5.0.2 Activity 2 - Explore your gene further

*Estimated time: 45 min*

1. Work with a partner to complete the following analysis using the Try it Out section of the tutorial.

#### 5.0.2.1 Questions

| Graph your gene of interest across all the regions of the gut |
| --- |
| **INSERT YOUR GRAPH** |

| Create a clusterProfiler graph comparing two regions. If the number of pathways are too many, increase the stringency of your adjusted p-value to reduce the clusters. Pick a cluster to explore, copy and paste the image and write a description of what the image shows. |
| --- |
|  |

| Compare two regions for your gene of interest. What does this suggest? |
| --- |
| **INSERT YOUR GRAPH** |

| Look up possible genes that might work with your gene of interest. Hint: what is in the clusterProfiler you performed? |
| --- |
|  |

### 5.0.3 Activity 3 - Begin developing your poster!

*Estimated time: 30 min*

1. Turn in at least one graph that you can use in your poster and include a legend for your figure. Your legend should describe in prose what the figure is showing.
2. Post your tentative title and project idea on the class padlet, include the names of everyone in your group.

| Notes |
| --- |
|  |

Footnotes

### 5.0.4 Resources

* [Google Doc](https://docs.google.com/document/d/17nQ6OYVCHeLCeXrDYcNjZvj7-fdAGPZQ)

### 5.0.5 Contributions and Affiliations

* Rosa Alcazar, Ph.D., Clovis Community College
* Stephanie Coffman, Ph.D., Clovis Community College

Last Revised: March 29, 2022

# 6 Scientific Communication

## 6.1 Group Poster

### 6.1.1 Introduction

An important part of scientific research is presenting your findings. This might be in the form of a peer-reviewed journal article or a more informal poster presentation at a conference or symposium. In this assignment, students will work with their groups to put together their work from the C-MOOR Labs into a Digital Poster.

### 6.1.2 Part 1 - Choose a Template

1. Read through the [grading criteria](#grading-criteria) at the end of this document.
2. With your group, choose a template to use for your poster. You can use one of the ones provided below or another one you find on the internet. Just make sure it has a space for an abstract, introduction, methods, results, discussion, references and acknowledgments.
3. Discuss with your group how you will divide up the work and exchange important information (e.g. phone numbers, email).

### 6.1.3 Part 2 - Make an Academic Research Poster

1. Complete the following components of your research poster with your group. You might not do them in this order, but these are the components you are being graded on. For more details on each of these sections and their role in a scientific paper, see the Scientific Literature Lab.

#### 6.1.3.1 Title, Authors and Affiliations

Your poster should include: (1) a title for your work, (2) authors who contributed to the work and (3) author affiliations. Since you completed this work as a team of scientists, all of your team members are considered authors. List your team members in alphabetical order. The author’s affiliation is the university, college, research institution or company that the work was conducted at. For us, the affiliation would be “Department of Biology, Clovis Community College, California, United States.”

#### 6.1.3.2 Abstract

An abstract is a concise summary of your paper. An effective abstract will inform the reader of the scientific hypothesis being tested, the purpose, or “why”, of the study, the main methods, important results and conclusions in only **one paragraph**. When writing an abstract for a publication or presentation, there is always a maximum word or character count. For your poster, your abstract can have a **maximum length of 200 words**. Many scientists choose to write the abstract last.

Different fields of science have slightly different requirements and formats for abstracts. Here is a general guideline:

* Background: In one sentence, introduce your work. An effective introduction tells the reader what is known in the field (context) and identifies the gap of knowledge being addressed in the paper.
* Methods and Results: This section should be the longest part of your abstract, but no more than two or three sentences. This section is arguably the most important part of the abstract, because other scientists seek out a paper when they are interested in the results. The details for your methods will be contained in your paper, so in your abstract you can keep it brief. For your results, pick out the most important results and summarize them. Depending on your research, you may want to address the methods and results separately, but often they are intertwined (e.g. the example abstract below).
* Conclusions: In one sentence, concisely state what you learned from your research.

#### 6.1.3.3 Introduction

A good introduction section should do two things. First, it provides context for your work by describing what is already known in the field, as well as an unknown that your research is addressing. The latter is often called the gap in knowledge. Second, it should identify your scientific question and hypothesis. Usually, the introduction starts broadly, describing the work of other scientists. It is important to summarize this work (do not quote) and to properly cite the work.

When writing your introduction, it is often helpful to start at the end. Identify your scientific question, your hypothesis and the gap of knowledge first. Then brainstorm what you will need to tell your readers in terms of context and background. For this project, be sure to include background information on Drosophila melanogaster and why it is a good model system for this research, because your audience might not know.

#### 6.1.3.4 Materials and Methods

For your lab poster, your materials and methods section will detail your analysis of the data. Since this is a poster and not an article, you do not need to worry about including all the details and can keep it pretty brief. Don’t provide any of your results, just the methods. Because we did not make the RNA-seq libraries ourselves, you will simply cite the paper that made them instead of detailing their construction. Scientists usually write this section of their paper first, followed by the results section. Some other things you might include would be what type of analysis you decided to do (type of plot, which parts of the midgut you analyzed, sets of genes, what p-value you used, etc.).

#### 6.1.3.5 Results

The Results section is where you will detail your data in the form of figures, tables and written text. Begin by creating your tables and figures. Place the figures and tables in order of how you want to present them and name them Figure 1, Figure 2, Table 1, Table 2, etc.

In your written narrative of the results, you should go through each figure in order, emphasizing any important results from each one. As you discuss each figure, you will reference the figure or table in parentheses. For example: “RT-PCR analysis shows an increase in gene expression for gene X (Fig 2).” It is important that you present your data clearly and in a logical manner.

Have fun playing around with how to organize your figures with the text to make the poster look professional. You need a minimum of 2 figures for your poster.

#### 6.1.3.6 Discussion

The discussion section of the paper is your chance to analyze and interpret your results. For your lab report, make sure your discussion section includes all of the following:

1. What do your results mean?
2. How do they fit into the bigger picture?
3. If any experiments did not give expected results, hypothesize why that might have been the case and propose alternate experiments that could confirm or clarify your results.
4. Include at least one sentence of future work that you would do if you had more time or what students in upcoming semesters could do to continue to answer your questions.

#### 6.1.3.7 References

All the references that you cite on your poster must be present in a References section. For your lab report, make your reference section in alphabetical order by the first author’s last name. For your lab report, all of your sources will be scientific journals and should use the following format:

Authors (year) “Title.” Journal Name, vol. #, page #s, DOI

Online article that is also in print:

Haussecker D., Huang Y., Lau A., Parameswaran P., Fire A. Z. and M. A. Kay (2010) “Human tRNA-derived small RNAs in the global regulation of RNA silencing.” RNA, Vol. 16, page 637-695, doi:10.1261/rna.2000810

Online article only:

Marianes, A. and A. C. Spradling (2013) “Physiological and stem cell compartmentalization within the Drosophila midgut.” eLife, doi:10.7554/eLife.00886

#### 6.1.3.8 In-text Citations

To save space on our posters, we will number our references 1 through 5 and use the numbers as citations throughout the text of your poster.

### 6.1.4 Part 3 - Proofread and Add Final Touches

*Estimated time: 30 min to an hour*

1. Each group member should re-read the poster from beginning to end and fix any typos or grammatical errors.
2. Check the alignment of figures, text boxes, titles, etc.
3. Add some finishing touches. You can play with the color, the font, add additional images if it’s relevant.

### 6.1.5 Part 4 - Canvas Discussion

*Estimated time: 30 min*

You will turn in your poster to be graded as a group in a Canvas Assignment and post it to a Canvas Discussion to be viewed by the class.

1. Convert your poster to a pdf.
2. Have one member of your group turn in the pdf of your poster to the Graded Canvas Assignment. This assignment is already set up so that if one group member turns it in, it will show as submitted for all students in the group. This is where your instructor will grade you poster as a group.
3. Have one member of your group post a pdf of your poster in the Canvas C-MOOR Poster Discussion.
   1. With your poster, introduce your group members and copy and paste your abstract into the post.
   2. Insert your pdf into the post and edit the link so that it automatically shows the inline preview. This will make it easier for students to view your poster.
4. As an individual, read through the other posters from different groups.
5. Post comments

### 6.1.6 Grading Criteria

Everyone in the group earns the same grade, so it is important to work together.

| Points | Category |
| --- | --- |
| 2 points | Title, Authors and Affiliations |
| 3 points | Abstract |
| 4 points | Introduction 2 - relevant background 2 - drosophila as a model system |
| 5 points | Methods |
| 5 points | Results |
| 4 points | Discussion |
| 5 points | References ( 2 primary sources, 5 sources total) |
| 2 points | Poster Organization |

### 6.1.7 Footnotes

#### 6.1.7.1 Resources

* [Google Doc](https://docs.google.com/document/d/1xMewa-RktOWeGfSYjlSy9TB7yd_w8x0b)

#### 6.1.7.2 Contributions and Affiliations

* Stephanie R. Coffman, Clovis Community College

Last Revised:September 2021

## 6.2 Poster Template



[slide](https://docs.google.com/presentation/d/1NYfLr0sghb_nUwJZC_NJM4HoXIsLu1-n)

# 7 Example miniCURE Projects

Read more about what these students did and how you can help

Zellweger Spectrum Disorder

[Drosophila Melanogaster a Good Model System of Zellweger Spectrum Disorder BIO11A SP2022](https://help.c-moor.org/t/326)



Look at This!

Explore other miniCURE and CURE projects in our [Look at This!](https://help.c-moor.org/c/look-at-this/8) category

# 8 C-MOOR Scholars

Meet the C-MOOR Scholars and learn how you can support them

* <https://www.cloviscollege.edu/alumni-and-community/c-moor/c-moor-scholars.html>

# 9 Online Community

Join the discussion at <https://help.c-moor.org>

# About the Authors

These credits are based on our [course contributors table guidelines](https://github.com/jhudsl/OTTR_Template/wiki/How-to-give-credits).

| Credits | Names |
| --- | --- |
| **Pedagogy** |  |
| Lead Content Instructor(s) | [FirstName LastName](link%20to%20personal%20website) |
| Lecturer(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | Delivered the course in some way - video or audio |
| Content Author(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | If any other authors besides lead instructor |
| Content Contributor(s) (include section name/link in parentheses) - make new line if more than one section involved | Wrote less than a chapter |
| Content Editor(s)/Reviewer(s) | Checked your content |
| Content Director(s) | Helped guide the content direction |
| Content Consultants (include chapter name/link in parentheses or word “General”) - make new line if more than one chapter involved | Gave high level advice on content |
| Acknowledgments | Gave small assistance to content but not to the level of consulting |
| **Production** |  |
| Content Publisher(s) | Helped with publishing platform |
| Content Publishing Reviewer(s) | Reviewed overall content and aesthetics on publishing platform |
| **Technical** |  |
| Course Publishing Engineer(s) | Helped with the code for the technical aspects related to the specific course generation |
| Template Publishing Engineers | [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Candace Savonen](https://www.cansavvy.com/) |
| Package Developers ([ottrpal](https://github.com/jhudsl/ottrpal)) [Candace Savonen](https://www.cansavvy.com/), [John Muschelli](https://johnmuschelli.com/), [Carrie Wright](https://carriewright11.github.io/) |  |
| **Art and Design** |  |
| Illustrator(s) | Created graphics for the course |
| Figure Artist(s) | Created figures/plots for course |
| Videographer(s) | Filmed videos |
| Videography Editor(s) | Edited film |
| Audiographer(s) | Recorded audio |
| Audiography Editor(s) | Edited audio recordings |
| **Funding** |  |
| Funder(s) | Institution/individual who funded course including grant number |
| Funding Staff | Staff members who help with funding |

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.3 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-01-20   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.3)   
## bookdown 0.24 2022-02-15 [1] Github (rstudio/bookdown@88bc4ea)   
## callr 3.4.4 2020-09-07 [1] RSPM (R 4.0.2)   
## cli 2.0.2 2020-02-28 [1] RSPM (R 4.0.0)   
## crayon 1.3.4 2017-09-16 [1] RSPM (R 4.0.0)   
## desc 1.2.0 2018-05-01 [1] RSPM (R 4.0.3)   
## devtools 2.3.2 2020-09-18 [1] RSPM (R 4.0.3)   
## digest 0.6.25 2020-02-23 [1] RSPM (R 4.0.0)   
## ellipsis 0.3.1 2020-05-15 [1] RSPM (R 4.0.3)   
## evaluate 0.14 2019-05-28 [1] RSPM (R 4.0.3)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.6.1 2022-01-22 [1] CRAN (R 4.0.2)   
## htmltools 0.5.0 2020-06-16 [1] RSPM (R 4.0.1)   
## knitr 1.33 2022-02-15 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.0 2021-02-15 [1] CRAN (R 4.0.2)   
## magrittr 2.0.2 2022-01-26 [1] CRAN (R 4.0.2)   
## memoise 1.1.0 2017-04-21 [1] RSPM (R 4.0.0)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgload 1.1.0 2020-05-29 [1] RSPM (R 4.0.3)   
## prettyunits 1.1.1 2020-01-24 [1] RSPM (R 4.0.3)   
## processx 3.4.4 2020-09-03 [1] RSPM (R 4.0.2)   
## ps 1.3.4 2020-08-11 [1] RSPM (R 4.0.2)   
## purrr 0.3.4 2020-04-17 [1] RSPM (R 4.0.3)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 0.4.10 2022-02-15 [1] Github (r-lib/rlang@f0c9be5)   
## rmarkdown 2.10 2022-02-15 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.0.2)   
## sessioninfo 1.1.1 2018-11-05 [1] RSPM (R 4.0.3)   
## stringi 1.5.3 2020-09-09 [1] RSPM (R 4.0.3)   
## stringr 1.4.0 2019-02-10 [1] RSPM (R 4.0.3)   
## testthat 3.0.1 2022-02-15 [1] Github (R-lib/testthat@e99155a)   
## usethis 2.1.5.9000 2022-02-15 [1] Github (r-lib/usethis@57b109a)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2022-02-15 [1] Github (yihui/xfun@74c2a66)   
## yaml 2.2.1 2020-02-01 [1] RSPM (R 4.0.3)   
##   
## [1] /usr/local/lib/R/site-library  
## [2] /usr/local/lib/R/library