CoGAPS on SciServer

July, 2023

Table of Contents

# About this Course

This introductory course will provide a quick overview of how the Bayesian NMF algorithm, CoGAPS (Coordinated Gene Activity across Pattern Subsets), can provide new insights into single cell datasets. Through these exercises you will analyze a real dataset using the SciServer compute platform.

## 0.1 Available course formats

This course is available in multiple formats which allows you to take it in the way that best suits your needs. You can take it for certificate (for free) using Leanpub.

* The material for this course can be viewed without login requirement on this [Bookdown website](https://practicalgenomics.github.io/cogaps-on-sciserver/). This format might be most appropriate for you if you rely on screen-reader technology.
* This course can be taken for [free certification through Leanpub](LINK%20HERE).
* Our courses are open source, you can find the [source material for this course on GitHub](https://github.com/PracticalGenomics/cogaps-on-sciserver).

# 1 Introduction

This introductory course will provide a quick overview of how the Bayesian NMF algorithm, CoGAPS (Coordinated Gene Activity across Pattern Subsets), can provide new insights into single cell datasets. Through these exercises you will analyze a real dataset using the SciServer compute platform.

## 1.1 Motivation

If you would like to perform sparse matrix factorization on any data. And when this data represents biomolecules, to do gene set analysis. This can be done with CoGAPS, which can be used by anyone; no machine learning experience is required.

## 1.2 Target Audience

The course is intended for anyone! No software or prior coding experience is required.

## 1.3 Curriculum

The course covers:

* How to join the compute platform, SciServer
* How to access and launch cellxgene
* How to load packages, data, configure/run CoGAPS, visualize patterns, find pattern markers, and document software in RStudio

devtools::session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## hms 0.5.3 2020-01-08 [1] RSPM (R 4.0.0)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.3 2022-10-07 [1] CRAN (R 4.0.2)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## ottrpal 1.0.1 2023-03-28 [1] Github (jhudsl/ottrpal@151e412)   
## pillar 1.9.0 2023-03-22 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] RSPM (R 4.0.3)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## readr 1.4.0 2020-10-05 [1] RSPM (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## tibble 3.2.1 2023-03-20 [1] CRAN (R 4.0.2)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## utf8 1.1.4 2018-05-24 [1] RSPM (R 4.0.3)   
## vctrs 0.6.1 2023-03-22 [1] CRAN (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# 2 Getting Started with SciServer

## 2.1 What is SciServer?

SciServer is an online platform that provides access to big data resources to researchers worldwide. It is used by scientists studying astronomy, biology, oceanography, and more. It is free to use as long as you are using it for scientific research. Through using SciServer, you don’t need a fancy computer or need to install any special programs on your computer, you can simply 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!

## 2.2 Why use SciServer?

In this course, we will use the online SciServer platform in order to perform data analysis. 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.

## 2.3 Learning Objectives

This chapter will cover:

* How to create an account on SciServer
* How to confirm your email address
* How to share your username with your instructor

## 2.4 Create an Account on SciServer

1. Open [sciserver.org](https://sciserver.org) in a web browser.       a. TIP: 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 username, email, etc. and click “**Create account**”       a. Note that you cannot change your username.

## 2.5 Confirm Your Email Address

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

### 2.5.1 Resources

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

### 2.5.2 References

Cox, K., & Tan, F. (2022, January 25). Join SciServer. C-MOOR. Retrieved June 27, 2023, from <http://www.c-moor.org/miniCURE-RNA-seq/join-sciserver.html>

devtools::session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## hms 0.5.3 2020-01-08 [1] RSPM (R 4.0.0)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.3 2022-10-07 [1] CRAN (R 4.0.2)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## ottrpal 1.0.1 2023-03-28 [1] Github (jhudsl/ottrpal@151e412)   
## pillar 1.9.0 2023-03-22 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] RSPM (R 4.0.3)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## readr 1.4.0 2020-10-05 [1] RSPM (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## tibble 3.2.1 2023-03-20 [1] CRAN (R 4.0.2)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## utf8 1.1.4 2018-05-24 [1] RSPM (R 4.0.3)   
## vctrs 0.6.1 2023-03-22 [1] CRAN (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# 3 cellxgene

## 3.1 Learning Objectives

* Confirm access to STAC Administrators Group
* Launch cellxgene
* Edit SciServer Dashboard

## 3.2 Instructions

### 3.2.1 Confirm Access

1. Go to [sciserver.org](https://sciserver.org) and click on “**Login to SciServer**”.
2. Log in with the SciServer account you created in Part 1.
3. Once logged in, you will see options on the top menu bar of the homepage/Dashboard (Home, Files, Groups, and various options for Compute). Confirm that you received and accepted the invitation to the STAC Administrators Group by clicking on “**Groups**” in the top menu bar.
4. Click on “**STAC Administrators**” in the left sidebar menu. You should see your username in the Members list on the right sidebar. You should also have access to the Shared Data Volume “STAC” and Share Compute Image “single-cell-explorer”.

### 3.2.2 Launch cellxgene

1. In order to explore the available datasets, click on “**Files**” in the top menu bar.
2. Click on “**Data Volumes**” in the left sidebar menu.
3. Scroll down the page to find the data volume “**STAC**”. Click on the name to access the dashboard.
4. Under the “Available datasets” heading, click on “**PDAC** [**cellxgene**](#cellxgene)”.
5. cellxgene should now launch with your dataset of interest ready to explore!

### 3.2.3 Edit Dashboard

1. Click on “**Files**” in the top menu bar.
2. Click on “**Data Volumes**” in the left sidebar menu.
3. Scroll down the page to find the data volume “**STAC**”. Click on the name to access the dashboard.
4. Scroll to the bottom of the page and find the “README.md” file. Click on the three dots, then click “**View/Edit**”.
5. You should now be able to edit and improve the README file using standard Markdown syntax.

devtools::session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## hms 0.5.3 2020-01-08 [1] RSPM (R 4.0.0)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.3 2022-10-07 [1] CRAN (R 4.0.2)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## ottrpal 1.0.1 2023-03-28 [1] Github (jhudsl/ottrpal@151e412)   
## pillar 1.9.0 2023-03-22 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] RSPM (R 4.0.3)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## readr 1.4.0 2020-10-05 [1] RSPM (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## tibble 3.2.1 2023-03-20 [1] CRAN (R 4.0.2)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## utf8 1.1.4 2018-05-24 [1] RSPM (R 4.0.3)   
## vctrs 0.6.1 2023-03-22 [1] CRAN (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# 4 R/RStudio

## 4.1 Learning Objectives

* Start up a C-MOOR RStudio compute container
* Find and use features of RStudio: R console, help window, viewer, environment window, history.
* Complete your first “swirl” tutorial

## 4.2 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 the “C-MOOR RStudio” compute container on SciServer. You will learn the basics of how to use RStudio, and will practice doing R coding within RStudio. You will also do your first “swirl” lesson. Swirl is a set of R tutorials that run inside RStudio.

## 4.3 Instructions

### 4.3.1 Start up a “C-MOOR RStudio” compute container

1. Open [sciserver.org](https://sciserver.org) in a web browser and log in to your account.

      a. If you’re already logged in, click “**Home**” in the top menu bar to return to the home page.

1. Scroll down to the second set of boxes and click “**Compute**”.
2. Click “**Create container**”.

      a. 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 that you can tell your containers apart. You could name this container “RStudio”, since you’ll be using it to access RStudio.

      b. In the “**Compute Image**” drop-down menu, select “**C-MOOR RStudio**”.

      c. Under “**Data Volumes**”, check the box next to “**C-MOOR Data**”.

      d. Click “**Create**”. This may take a moment.

1. You should now see a new entry in your list of containers.

      a. “Created At” should be a few moments ago.

      b. “Name” should be the name you chose.

      c. “Image” should be “C-MOOR RStudio”.

1. Start your C-MOOR RStudio container by clicking on its **name** (whatever name you chose when you created the container). This will open in a new tab.

      a. You should see RStudio.

RStudio startup screen

Figure : RStudio startup screen

      b. 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.**

### 4.3.2 Complete your first swirl tutorial

1. Watch this 90 second video tour of RStudio:
2. If you’re not there already, go to the SciServer compute page and start up the C-MOOR RStudio container.

      a. Open [sciserver.org](https://sciserver.org) in a web browser and log in to your account.

      b. If you’re already logged in, click “**Home**” in the top menu bar to return to the home page.

      c. Scroll down to the second set of boxes and click “**Compute**”.

      d. Start your C-MOOR container by clicking on its name.

1. In the **R console** window of RStudio (bottom left, or if you have no files open, it may take up the whole left side of the screen) type these commands to start up swirl:

      a. library(swirl)

      b. swirl()

1. Install the course, following the instructions provided by swirl:

      a. Enter your name

      b. Press ENTER

      c. Select 1, 2, or 3

      d. Install the course: “**R Programming: The basics of programming in R**”.

1. Complete your first swirl lesson.

      a. Choose the “**R programming**” course.

      b. Choose Lesson 1 “**Basic Building Blocks**”.

      c. Follow the instructions provided by swirl to complete the lesson.

      d. When you get to the end, it will ask if you want credit on Coursera. Choose “**No**” (we are not using Coursera for this course).

### 4.3.3 Managing your C-MOOR RStudio compute container

If you delete your container now, you will lose **all** your progress in swirl. If you need to return to any part of the tutorial later, it is a good idea to keep the container until you are sure that you’re finished using it.

devtools::session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## hms 0.5.3 2020-01-08 [1] RSPM (R 4.0.0)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.3 2022-10-07 [1] CRAN (R 4.0.2)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## ottrpal 1.0.1 2023-03-28 [1] Github (jhudsl/ottrpal@151e412)   
## pillar 1.9.0 2023-03-22 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] RSPM (R 4.0.3)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## readr 1.4.0 2020-10-05 [1] RSPM (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## tibble 3.2.1 2023-03-20 [1] CRAN (R 4.0.2)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## utf8 1.1.4 2018-05-24 [1] RSPM (R 4.0.3)   
## vctrs 0.6.1 2023-03-22 [1] CRAN (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# 5 CoGAPS

## 5.1 Learning Objectives

* How to load packages in RStudio
* How to load data in RStudio
* How to configure CoGAPS
* How to run CoGAPS
* How to visualize patterns
* How to find pattern markers
* How to document software

## 5.2 Instructions

### 5.2.1 Start up a “CoGAPS RStudio” compute container

1. Open [sciserver.org](https://sciserver.org) in a web browser and log in to your account.

      a. If you’re already logged in, click “**Home**” in the top menu bar to return to the home page.

1. Scroll down to the second set of boxes and click “**Compute**”.
2. Click “**Create container**”

      a. 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 that you can tell your containers apart. You could name this container “CoGAPS”, since you’ll be using it to run CoGAPS.

      b. In the “**Compute Image**” drop-down menu, select “**R [version #] (RStudio)**” (there may be multiple versions of R in this drop-down menu; click on the latest version listed).

      c. Click “**Create**”. This may take a moment.

1. You should now see a new entry in your list of containers.

      a. “Created at” should be a few moments ago.

      b. “Name” should be the name you chose.

      c. “Image” should be "R[version #](RStudio)"

1. Start your CoGAPS RStudio container by clicking on its **name** (whatever name you chose when you created the container). This will open in a new tab.

      a. You should see RStudio.

      b. 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.**

### Update rlang

1. Enter and run these commands into RStudio to make sure that you have the current version of rlang (you will receive an error later on otherwise):

packageVersion("rlang") # ‘1.0.6’  
devtools::install\_github("r-lib/rlang")  
packageVersion("rlang") # ‘1.1.0.9000’

1. Once you see the output: [1] '1.1.1.9000' rlang has been updated.

### 5.2.2 Install Packages

1. Enter and run this command:

devtools::install\_github("FertigLab/CoGAPS")

1. This will take a while. Once the red “STOP” symbol is no longer visible in the top right hand corner of the Console, the installation is complete.
2. Enter and run this command:

devtools::install\_github("sjmgarnier/viridis")

1. Once the red “STOP” symbol is no longer visible in the top right hand corner of the Console, the installation is complete.
2. Enter and run this command:

remotes::install\_github("satijalab/seurat", "seurat5", quiet = TRUE)

1. This will take a while. Once the red “STOP” symbol is no longer visible in the top right hand corner of the Console, the installation is complete.

### 5.2.3 Load Packages

1. Enter and run these commands:

library( "CoGAPS" )  
library( "Seurat" )  
library( "viridis" )

1. Once the red “STOP” symbol is no longer visible in the top right hand corner of the Console, the packages have been loaded.

### 5.2.4 Load Data

1. Enter and run this command:

url <- "https://github.com/FertigLab/CoGAPS/raw/master/data/inputdata.Rds"  
download.file( url, "inputdata.Rds" )

1. You should see this output if the run is successful:

trying URL 'https://github.com/FertigLab/CoGAPS/raw/master/data/inputdata.Rds'  
Content type 'application/octet-stream' length 433262849 bytes (413.2 MB)  
==================================================  
downloaded 413.2 MB

1. Enter and run this command:

pdac\_data <- readRDS( "inputdata.Rds" )  
pdac\_data

1. You should see this output if the run is successful:

An object of class Seurat   
15184 features across 25442 samples within 2 assays   
Active assay: originalexp (15176 features, 2000 variable features)  
 1 other assay present: CoGAPS  
 5 dimensional reductions calculated: PCA, Aligned, UMAP, pca, umap

1. Enter and run this command:

pdac\_epi\_counts <- as.matrix( pdac\_data@assays$originalexp@counts )

1. You should see this output if the run is successful:

Warning: sparse->dense coercion: allocating vector of size 2.9 GiB

### Configure CoGAPS

1. Enter and run this command:

pdac\_params <- CogapsParams(  
 nIterations=100, # run for 100 iterations   
 seed=42, # for consistency across stochastic runs  
 nPatterns=8, # each thread will learn 8 patterns  
 sparseOptimization=TRUE, # optimize for sparse data  
 distributed="genome-wide" # parallelize across sets  
)

1. Enter and run this command:

pdac\_params <- setDistributedParams( pdac\_params, nSets=7 )

1. You should see this output if the run is successful:

setting distributed parameters - call this again if you change nPatterns

### 5.2.5 Run CoGAPS

1. Enter and run this command:

Sys.time()

1. Your output should include today’s date and time.
2. Enter and run this command:

pdac\_epi\_result <- CoGAPS( pdac\_epi\_counts, pdac\_params )

1. You should see this output if the run is successful (This will take a while, most likely around 20+ minutes):

This is CoGAPS version 3.19.1   
Running genome-wide CoGAPS on pdac\_epi\_counts (15176 genes and 25442 samples) with parameters:  
  
-- Standard Parameters --  
nPatterns 8   
nIterations 100   
seed 42   
sparseOptimization TRUE   
distributed genome-wide   
  
-- Sparsity Parameters --  
alpha 0.01   
maxGibbsMass 100   
  
-- Distributed CoGAPS Parameters --   
nSets 7   
cut 8   
minNS 4   
maxNS 11   
  
Creating subsets...  
set sizes (min, mean, max): (2168, 2168, 2168)  
Running Across Subsets...  
  
Data Model: Sparse, Normal  
Sampler Type: Sequential  
Loading Data...  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
Done! (00:00:36)  
 worker 1 is starting!  
-- Equilibration Phase --  
  
Warning: Large values detected, is data log transformed?  
 worker 6 is starting!  
  
Warning: Large values detected, is data log transformed?  
 worker 4 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 5 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 3 is starting!  
  
Warning: Large values detected, is data log transformed?  
 worker 7 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 2 is starting!  
-- Sampling Phase --  
 worker 2 is finished! Time: 00:18:04  
 worker 7 is finished! Time: 00:18:20  
 worker 4 is finished! Time: 00:18:49  
 worker 3 is finished! Time: 00:18:42  
 worker 6 is finished! Time: 00:18:59  
 worker 1 is finished! Time: 00:19:09  
 worker 5 is finished! Time: 00:19:19  
  
Matching Patterns Across Subsets…  
Running Final Stage...  
  
Connected to your session in progress, last started 2023-Jun-29 16:26:07 UTC (3 hours ago)  
Data Model: Sparse, Normal  
Sampler Type: Sequential  
Loading Data...  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
Done! (00:00:29)  
 worker 1 is starting!  
-- Equilibration Phase --  
  
Warning: Large values detected, is data log transformed?  
 worker 5 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 4 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 6 is starting!  
  
Warning: Large values detected, is data log transformed?  
 worker 2 is starting!  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
  
Warning: Large values detected, is data log transformed?  
 worker 3 is starting!  
  
Warning: Large values detected, is data log transformed?  
 worker 7 is starting!  
-- Sampling Phase --  
 worker 1 is finished! Time: 00:17:04  
 worker 6 is finished! Time: 00:17:04  
 worker 5 is finished! Time: 00:17:16  
 worker 4 is finished! Time: 00:17:15  
 worker 7 is finished! Time: 00:17:07  
 worker 2 is finished! Time: 00:17:17  
 worker 3 is finished! Time: 00:17:11

1. Enter and run this command:

Sys.time()

1. Your output should include today’s date and time.
2. Enter and run this command:

saveRDS( pdac\_epi\_result, "../data/pdac\_epi\_cogaps\_result" )

1. You should see this output if the run is successful:

Warning: cannot open compressed file '../data/pdac\_epi\_cogaps\_result', probable reason 'No such file or directory'Error in gzfile(file, mode) : cannot open the connection

1. Enter and run this command:

pdac\_epi\_result

1. You should see this output if the run is successful:

[1] "CogapsResult object with 15176 features and 25442 samples"  
[1] "7 patterns were learned"

1. Enter and run this command in order to save your results:

saveRDS( pdac\_epi\_result, "pdac\_epi\_cogaps\_result.rds" )

### 5.2.6 Visualize Patterns

1. Enter and run this command:

cogapsresult <- readRDS( "pdac\_epi\_cogaps\_result.rds" )

1. Enter and run this command:

patterns\_in\_order <- t( cogapsresult@sampleFactors[colnames(pdac\_data),] )  
pdac\_data[["CoGAPS"]] <- CreateAssayObject( counts = patterns\_in\_order )

1. You should receive this output:

Warning: Feature names cannot have underscores ('\_'), replacing with dashes ('-')

1. Enter and run this command:

inputdata <- pdac\_data  
DefaultAssay(inputdata) <- "CoGAPS"  
pattern\_names = rownames( inputdata@assays$CoGAPS )

1. Enter and run this command:

color\_palette <- viridis(n=10)  
FeaturePlot(inputdata, pattern\_names, cols=color\_palette, reduction = "umap") & NoLegend()

1. Your output should look like this if the run was successful (visible in the bottom right corner of your screen):

### 5.2.7 Find Pattern Markers

1. Enter and run this command:

pm <- patternMarkers( cogapsresult, threshold="cut" )

1. You should see this output if the run was successful:

Warning: STATS is longer than the extent of 'dim(x)[MARGIN]'

1. Enter and run this command:

# hallmarks <- PatternHallmarks( cogapsresult )

1. Enter and run this command:

# plotPatternHallmarks(hallmarks, whichpattern = 7)

### 5.2.8 Document Software

1. Enter and run this command:

sessionInfo()

1. Your output should include all the information about your RStudio session.

**Troubleshooting/Reminders:**

* If you restart RStudio, you must repeat the *Load Packages* step, otherwise errors will occur and you will not be able to successfully run your code.

devtools::session\_info()

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fansi 0.4.1 2020-01-08 [1] RSPM (R 4.0.0)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## hms 0.5.3 2020-01-08 [1] RSPM (R 4.0.0)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.3 2022-10-07 [1] CRAN (R 4.0.2)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## ottrpal 1.0.1 2023-03-28 [1] Github (jhudsl/ottrpal@151e412)   
## pillar 1.9.0 2023-03-22 [1] CRAN (R 4.0.2)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] RSPM (R 4.0.3)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] RSPM (R 4.0.0)   
## readr 1.4.0 2020-10-05 [1] RSPM (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## tibble 3.2.1 2023-03-20 [1] CRAN (R 4.0.2)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## utf8 1.1.4 2018-05-24 [1] RSPM (R 4.0.3)   
## vctrs 0.6.1 2023-03-22 [1] CRAN (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# About the Authors

These credits are based on our [course contributors table guidelines](https://www.ottrproject.org/more_features.html#giving-credits-to-contributors).

| 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/), [Ava Hoffman](https://www.avahoffman.com/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Ava Hoffman](https://www.avahoffman.com/), [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 20.04.5 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2023-07-05   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] RSPM (R 4.0.5)   
## bookdown 0.24 2023-03-28 [1] Github (rstudio/bookdown@88bc4ea)   
## cachem 1.0.7 2023-02-24 [1] CRAN (R 4.0.2)   
## callr 3.5.0 2020-10-08 [1] RSPM (R 4.0.2)   
## cli 3.6.1 2023-03-23 [1] CRAN (R 4.0.2)   
## 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 0.20 2023-01-17 [1] CRAN (R 4.0.2)   
## fastmap 1.1.1 2023-02-24 [1] CRAN (R 4.0.2)   
## fs 1.5.0 2020-07-31 [1] RSPM (R 4.0.3)   
## glue 1.4.2 2020-08-27 [1] RSPM (R 4.0.5)   
## htmltools 0.5.5 2023-03-23 [1] CRAN (R 4.0.2)   
## knitr 1.33 2023-03-28 [1] Github (yihui/knitr@a1052d1)   
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.0.2)   
## memoise 2.0.1 2021-11-26 [1] CRAN (R 4.0.2)   
## 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 1.4.0 2020-10-07 [1] RSPM (R 4.0.2)   
## 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 1.1.0 2023-03-14 [1] CRAN (R 4.0.2)   
## rmarkdown 2.10 2023-03-28 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.3 2022-04-02 [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 2023-03-28 [1] Github (R-lib/testthat@e99155a)   
## usethis 1.6.3 2020-09-17 [1] RSPM (R 4.0.2)   
## withr 2.3.0 2020-09-22 [1] RSPM (R 4.0.2)   
## xfun 0.26 2023-03-28 [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

# 6 References