

# TRANSCRIPTOMICS

## Setting up

Day 01

<https://tttorres.github.io/transcriptomics/>

# Transcriptomics

## Setting up

- Brief overview of the tools we'll use throughout the course  
(R/Bioconductor, RStudio, Trinity, Salmon, etc.)
- Set-up your laptop with all the software needed for the course

# Transcriptomics

## Tools



# Transcriptomics

Tools: all free/open source!



# Transcriptomics

## Tools: commercial platforms

- QIAGEN CLC Genomics Workbench
- Geneious by Dotmatics
- GeneSpring GX, Agilent

- ⚠ Expensive (\$1280-\$5500/year)
- ⚠ Not open-source
- ⚠ issues with reproducibility

# Transcriptomics

## Open source tools: Pros

- ➔ **Cost-Effective:** generally free to use
- ➔ **Flexibility and Customization:** possibility to modify the code
- ➔ **Large Community Support:** active communities that provide extensive documentation, tutorials, and forums.
- ➔ **Rapid Updates:** frequently updated
- ➔ **Transparency :** open-source code is available for review
- ➔ **Interoperability:** often designed to integrate well with other open-source software.
- ➔ **No Vendor Lock-In:** avoids dependence on proprietary systems

# Transcriptomics

## Open source tools: Cons

- ➔ **Limited Graphical User Interfaces (GUIs)**: most lack user-friendly interfaces
- ➔ **Steep Learning Curve**: require proficiency in command-line interfaces
- ➔ **Limited User Support**: typically do not offer dedicated customer support, relying instead on community assistance.
- ➔ **Compatibility Issues**: often depend on specific versions of libraries, packages, or system configurations
- ➔ **Time-Consuming**: setting up and troubleshooting pipelines can take time, especially for complex analyses

# Transcriptomics

## Systems

my computer



# Transcriptomics

## Systems

### other options

#### workstation

**Pros**

- ✓ unrestricted access
- ✓ all the software you need
- ✓ secure

**Cons**

- high initial investment
- need a SysAdmin
- quickly outdated

#### 'cloud'

**Pros**

- ✓ scalable
- ✓ easy set-up
- ✓ software installation made easy by docker and conda

**Cons**

- less secure
- data transfer charges

#### cluster

**Pros**

- ✓ scalable
- ✓ highly cost efficient
- ✓ SysAdmin taken care of
- ✓ secure

**Cons**

- may have to request software
- steep learning curve
- frequent server downtime

# Installing software

## Windows Subsystem for Linux

1. Search for "Command Prompt"
2. Choose the option "Run as administrator"
3. Allow it to make changes to your device
4. Type:

```
wsl --install
```

5. Close the prompt
6. Open "Ubuntu"

# Installing software: R



[Home]

## Download

CRAN

## R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Conferences

Search

Get Involved: Mailing Lists

Get Involved: Contributing

Developer Pages

R Blog

## R Foundation

Foundation

Board

Members

Donors

Donate

## Help With R

Getting Help

## Documentation

Manuals

FAQs

The R Journal

# The R Project for Statistical Computing

## Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

## News

- The [useR! 2025](#) conference will take place at Duke University, in Durham, NC, USA, August 8-10.
- [R version 4.4.2 \(Pile of Leaves\)](#) has been released on 2024-10-31.
- We are deeply sorry to announce that our friend and colleague Friedrich (Fritz) Leisch has died. [Read our tribute to Fritz here](#).
- [R version 4.3.3 \(Angel Food Cake\)](#) (wrap-up of 4.3.x) was released on 2024-02-29.
- You can support the R Foundation with a renewable subscription as a [supporting member](#).

## News via Mastodon



zeileis

@ResearchOrgs The CRAN team is currently working on supporting ROR information in package metadata.

The idea is to display it similar to ORCID information @ORCID\_Org

And also include it in the Crossref data associated with the DOI @crossref

Dec 18, 2024



zeileis

:rstats: 🎉

Two #rstats related research oras iust joined @ResearchOras 🎉

# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>

# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>
2. In the R home, click on download R

# Installing software: R

## CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: [main page](#), [windows release](#), [windows old release](#).

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

0-Cloud

<https://cloud.r-project.org/>

Argentina

<http://mirror.fcaglp.unlp.edu.ar/CRAN/>

Australia

<https://cran.csiro.au/>

<https://mirror.aarnet.edu.au/pub/CRAN/>

<https://cran.ms.unimelb.edu.au/>

Austria

<https://cran.wu.ac.at/>

Belgium

<https://www.freestatistics.org/cran/>

<https://ftp.belnet.be/mirror/CRAN/>

Brazil

<https://cran-r.c3sl.ufpr.br/>

<https://vps.fmvz.usp.br/CRAN/>

<https://briger.esalq.usp.br/CRAN/>

Bulgaria

<https://ftp.uni-sofia.bg/CRAN/>

Canada

<https://muug.ca/mirror/cran/>

<https://mirror.csclub.uwaterloo.ca/CRAN/>

<https://cran.mirror.rafal.ca/>

Chile

<https://cran.dcc.uchile.cl/>

China

<https://mirrors.tuna.tsinghua.edu.cn/CRAN/>

<https://mirrors.bfsu.edu.cn/CRAN/>

<https://mirrors.pku.edu.cn/CRAN/>

<https://mirrors.ustc.edu.cn/CRAN/>

Automatic redirection to servers worldwide, currently sponsored by Posit

Universidad Nacional de La Plata

CSIRO

AARNET

School of Mathematics and Statistics, University of Melbourne

Wirtschaftsuniversität Wien

Patrick Wessa

Belnet, the Belgian research and education network

Universidade Federal do Paraná

University of São Paulo, São Paulo

University of São Paulo, Piracicaba

Sofia University

Manitoba Unix User Group

University of Waterloo

Rafal Rzeczkowski

Departamento de Ciencias de la Computación, Universidad de Chile

TUNA Team, Tsinghua University

Beijing Foreign Studies University

Peking University

University of Science and Technology of China

# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>
2. In the R home, click on download R
3. CRAN (Comprehensive R Archive Network) mirrors, choose a server

# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>
2. In the R home, click on download R
3. CRAN (Comprehensive R Archive Network) mirrors, choose a server
4. Choose appropriate distribution

# Installing software: R



[CRAN](#)  
[Mirrors](#)  
[What's new?](#)  
[Search](#)  
[CRAN Team](#)

[About R](#)  
[R Homepage](#)  
[The R Journal](#)

[Software](#)  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Task Views](#)  
[Other](#)

[Documentation](#)  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

[Donations](#)  
[Donate](#)

This directory contains binaries for the base distribution and of R and packages to run on macOS. R and package binaries for R versions older than 4.0.0 are only available from the [CRAN archive](#) so users of such versions should adjust the CRAN mirror setting (<https://cran-archive.r-project.org>) accordingly.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

## R 4.4.2 "Pile of Leaves" released on 2024/10/31

Please check the integrity of the downloaded package by checking the signature:

`pkutil --check-signature R-4.4.2-arm64.pkg`

in the *Terminal* application. If Apple tools are not available you can check the SHA1 checksum of the downloaded image:

`openssl sha1 R-4.4.2-arm64.pkg`

### Latest release:

For Apple silicon (M1,2,...) Macs:

[R-4.4.2-arm64.pkg](#)

SHA1-hash: 7832cb5d6cd686fd3cc54c8ab4c93c464540a944  
(ca. 94MB, notarized and signed)

For older Intel Macs:

[R-4.4.2-x86\\_64.pkg](#)

SHA1-hash: f49ad56ce3a0ac569fd8f9668749bc861b965b5e  
(ca. 96MB, notarized and signed)

**R 4.4.2** binary for macOS 11 (**Big Sur**) and higher, signed and notarized packages.

Contains R 4.4.2 framework, R.app GUI 1.81, Tcl/Tk 8.6.12 X11 libraries and Texinfo 6.8. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `tcltk` R package or build package documentation from sources.

macOS Ventura users: there is a known bug in Ventura preventing installations from some locations without a prompt. If the installation fails, move the downloaded file away from the *Downloads* folder (e.g., to your home or Desktop).

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) (version 2.8.5 or later). Always re-install XQuartz when upgrading your macOS to a new major version.

This release uses Xcode 14.2/14.3 and GNU Fortran 12.2. If you wish to compile R packages which contain Fortran code, you may need to download the corresponding GNU Fortran compiler from <https://mac.R-project.org/tools>. Any external libraries and tools are expected to live in `/opt/R/arm64` (Apple silicon) or `/opt/R/x86_64` (Intel).

[NEWS](#) (for Mac GUI)

News features and changes in the R.app Mac GUI

[Mac-GUI-1.81.tar.gz](#)

SHA1-hash: 2980af0d8fa91660c1df44cc200c8914df6451fd

Sources for the R.app GUI 1.81 for macOS. This file is only needed if you want to join the development of the GUI (see also [Mac-GUI repository](#)), it is not intended for regular users. Read the `INSTALL` file for further instructions.

# Installing software: R

3.2.0 (2015-04-16) Full of Ingredients



3.0.2 (2013-09-25)  
Frisbee Sailing



4.4.2 (2024-10-31)  
Pile of Leaves



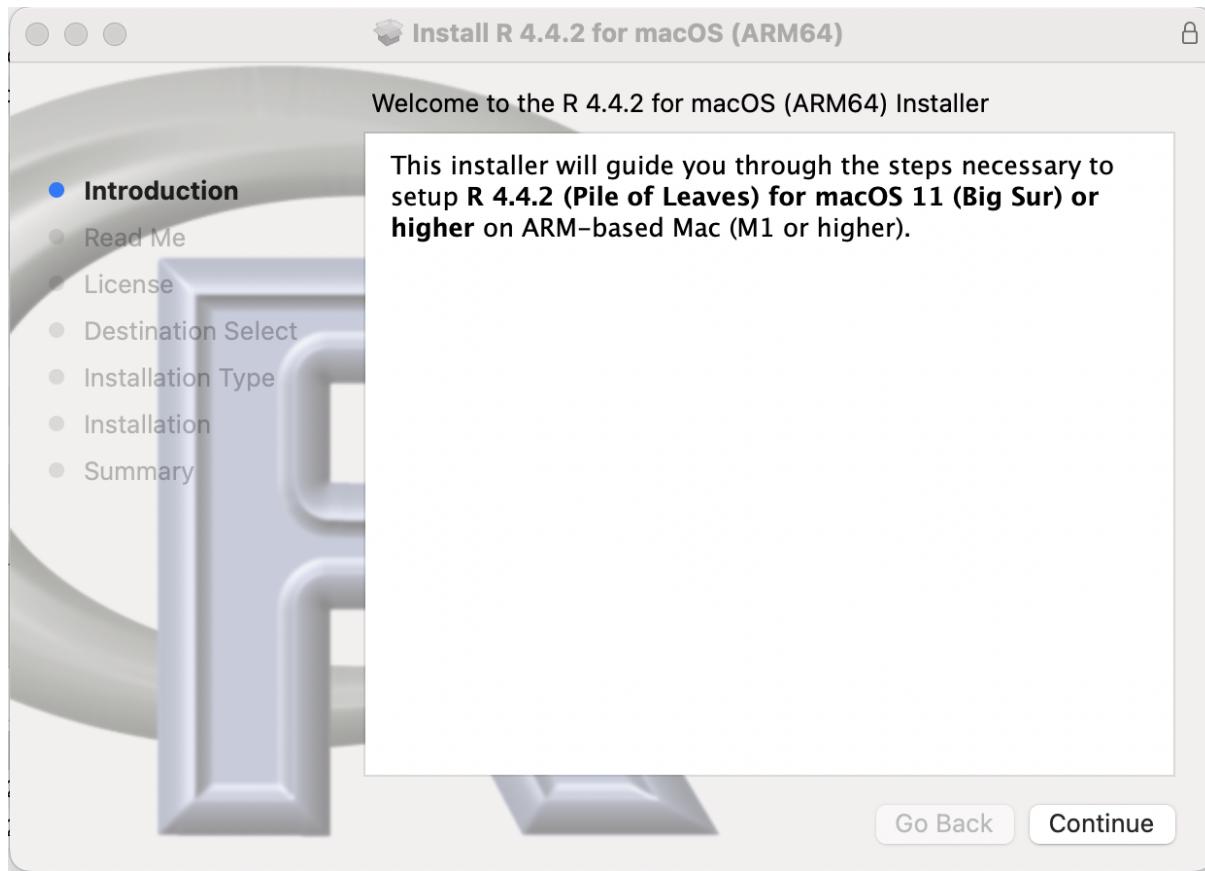
4.1.2 (2021-11-01)  
Bird Hippie

# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>
2. In the R home, click on download R
3. CRAN (Comprehensive R Archive Network) mirrors, choose a server
4. Choose appropriate distribution
5. Install R with double click

# Installing software: R

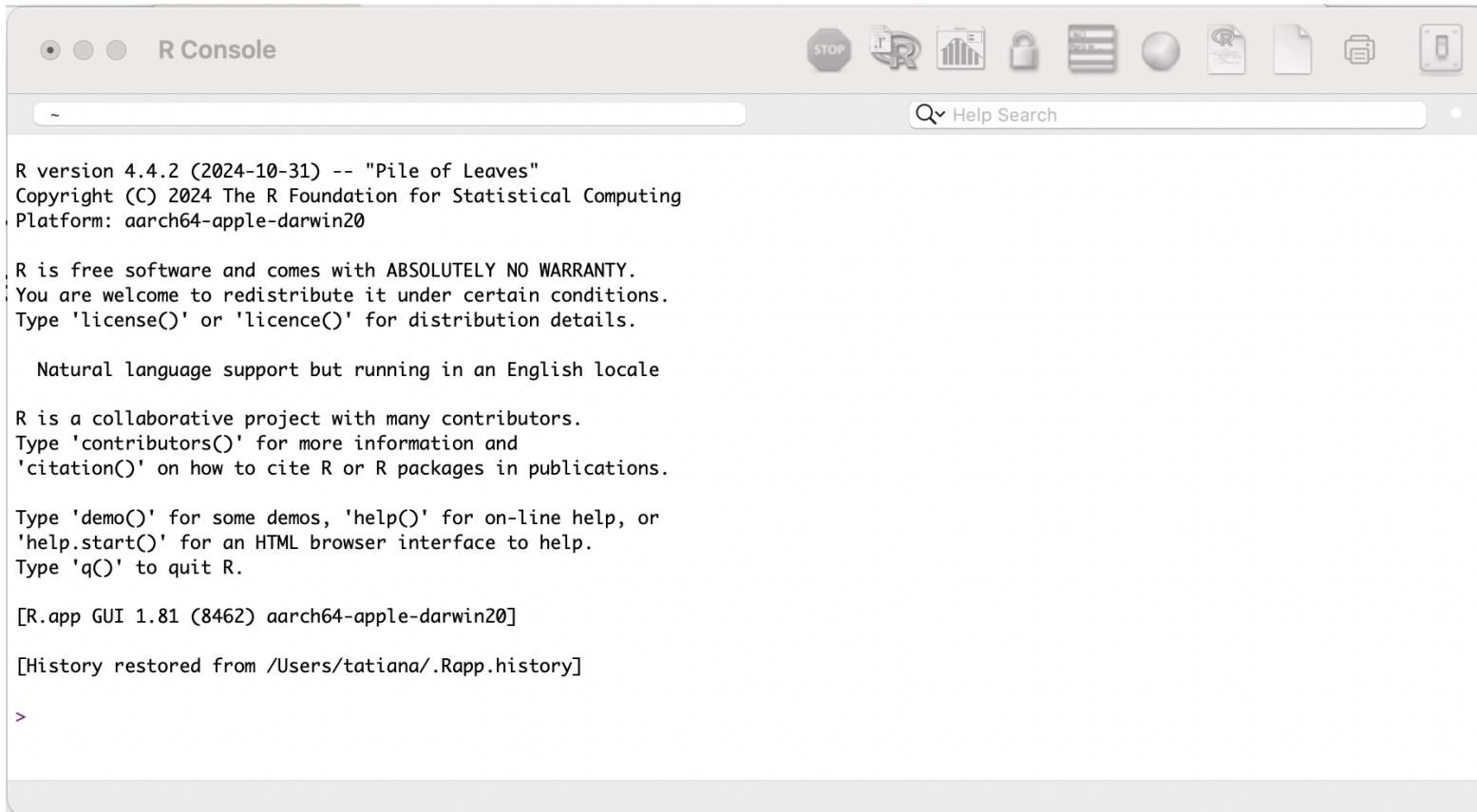


# Installing software

R

1. In the course webpage, click on <https://www.r-project.org/>
2. In the R home, click on download R
3. CRAN (Comprehensive R Archive Network) mirrors, choose a server
4. Choose appropriate distribution
5. Install R with double click
6. Open R (double click)

# Installing software: R



R version 4.4.2 (2024-10-31) -- "Pile of Leaves"  
Copyright (C) 2024 The R Foundation for Statistical Computing  
Platform: aarch64-apple-darwin20

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

[R.app GUI 1.81 (8462) aarch64-apple-darwin20]  
[History restored from /Users/tatiana/.Rapp.history]

>

# Installing software

## R-studio



The image displays three side-by-side screenshots of software interfaces:

- Terminal Window (BASH):** Shows a terminal session on a Mac OS X system. The user has just logged in as 'tatiana'. The screen shows the standard BASH welcome message, including the license information and the command to run 'chsh -s /bin/zsh' to update the default interactive shell.
- R Console:** Shows the R console interface. It displays the R version 4.4.2 startup message, which includes the copyright notice, the name 'R', and the statement 'ABSOLUTELY NO WARRANTY'. It also shows the standard R command-line interface with prompts like '>' and '[1] <-'.
- RStudio Script Editor:** Shows an RStudio workspace. The script editor tab is open with the file 'DCEmtDNA.R'. The code in the editor is for differential gene expression analysis using DESeq2, specifically for mtDNA genes. The code includes loading data, listing directories, and replacing file names.

# Installing software

## R-studio

1. In the course webpage, click on

<https://posit.co/download/rstudio-desktop/>

# Installing software: R-studio

[PRODUCTS](#) ▾[OPEN SOURCE](#) ▾[USE CASES](#) ▾[PARTNERS](#) ▾[LEARN & SUPPORT](#) ▾[ABOUT](#) ▾[DOWNLOAD](#)

## RStudio Desktop

Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python.

Don't want to download or install anything? Get started with RStudio on [Posit Cloud for free](#). If you're a professional data scientist looking to download RStudio and also need common enterprise features, don't hesitate to [book a call with us](#).

Want to learn about core or advanced workflows in RStudio? Explore the [RStudio User Guide](#) or the [Getting Started](#) section.

# Installing software

## R-studio

1. In the course webpage, click on

<https://posit.co/download/rstudio-desktop/>

2. Click on download RStudio Desktop

# Installing software: R-studio

[PRODUCTS](#) ▾[OPEN SOURCE](#) ▾[USE CASES](#) ▾[PARTNERS](#) ▾[LEARN & SUPPORT](#) ▾[ABOUT](#) ▾

## 1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

*R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.*

[DOWNLOAD AND INSTALL R](#)[All Installers and Tarballs](#)

## 2: Install RStudio

[DOWNLOAD RSTUDIO DESKTOP FOR MACOS 12+](#)

This version of RStudio is only supported on macOS 12 and higher. For earlier macOS environments, please [download a previous version](#).

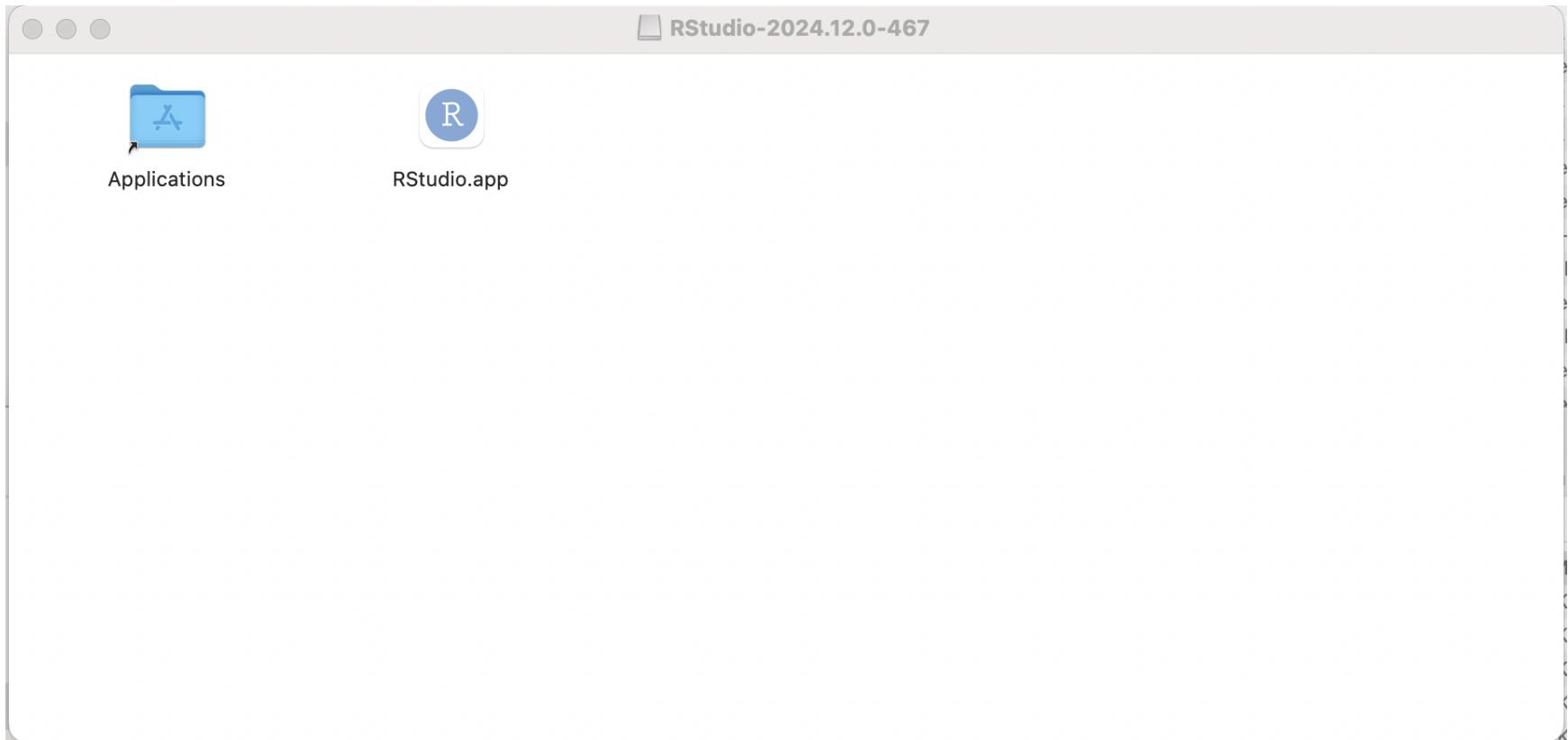
Size: 617.71 MB | [SHA-256: 46958FB4](#) | Version: 2024.12.0+467 | Released: 2024-12-16

# Installing software

## R-studio

1. In the course webpage, click on  
<https://posit.co/download/rstudio-desktop/>
2. Click on download RStudio Desktop
3. Drag to the Applications folder (macOS)

# Installing software: R-studio



# Installing software

## R-studio

1. In the course webpage, click on

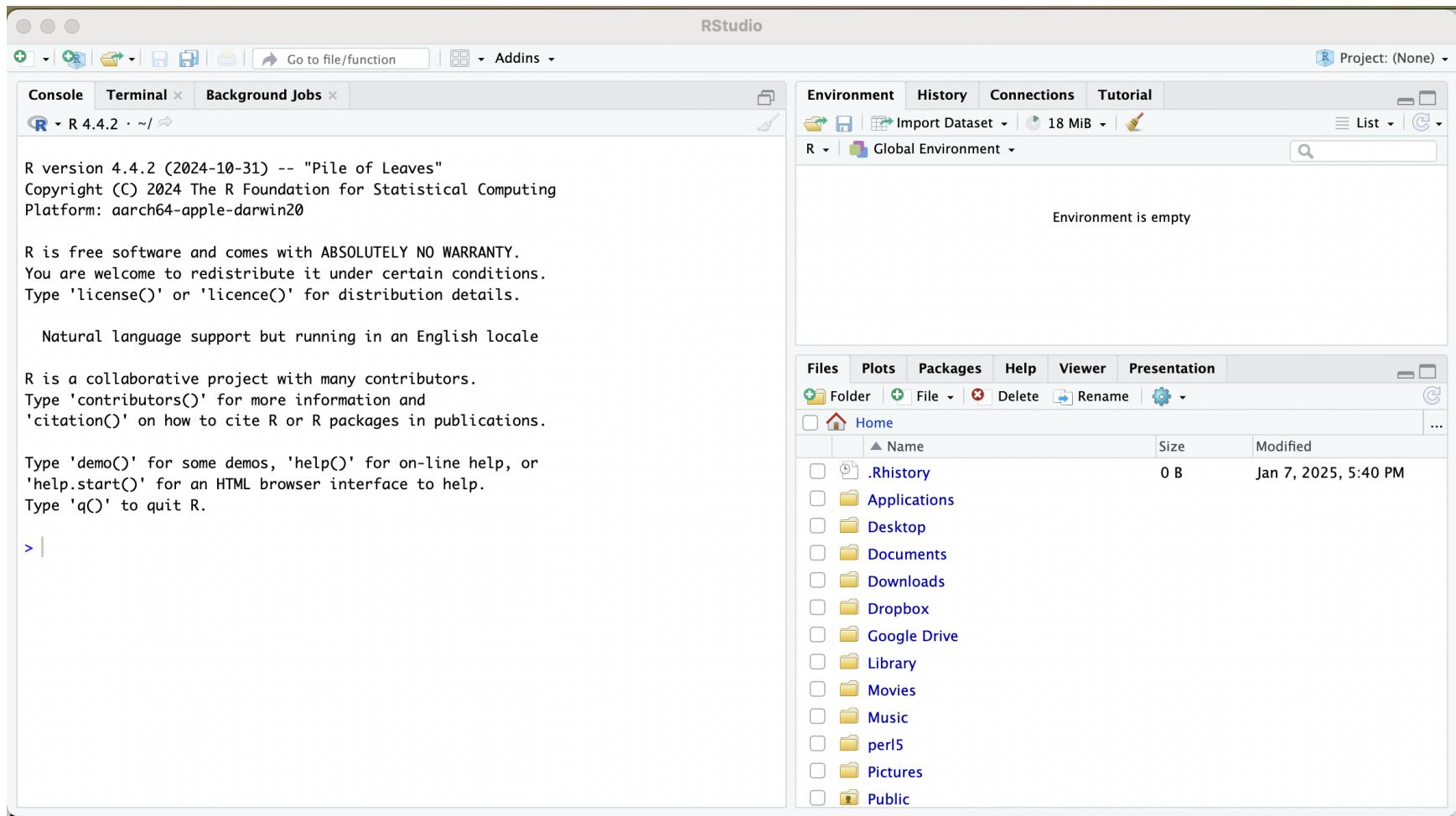
<https://posit.co/download/rstudio-desktop/>

2. Click on download RStudio Desktop

3. Drag to the Applications folder (macOS)

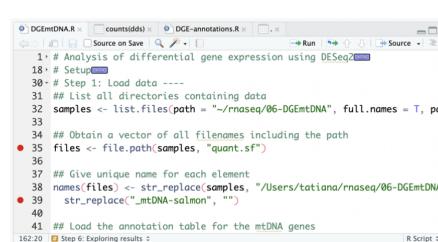
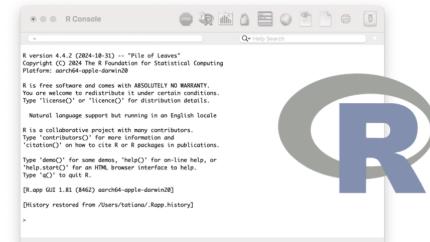
4. Open RStudio (double click)

# Installing software: R-studio



# Installing software

RStudio = IDE



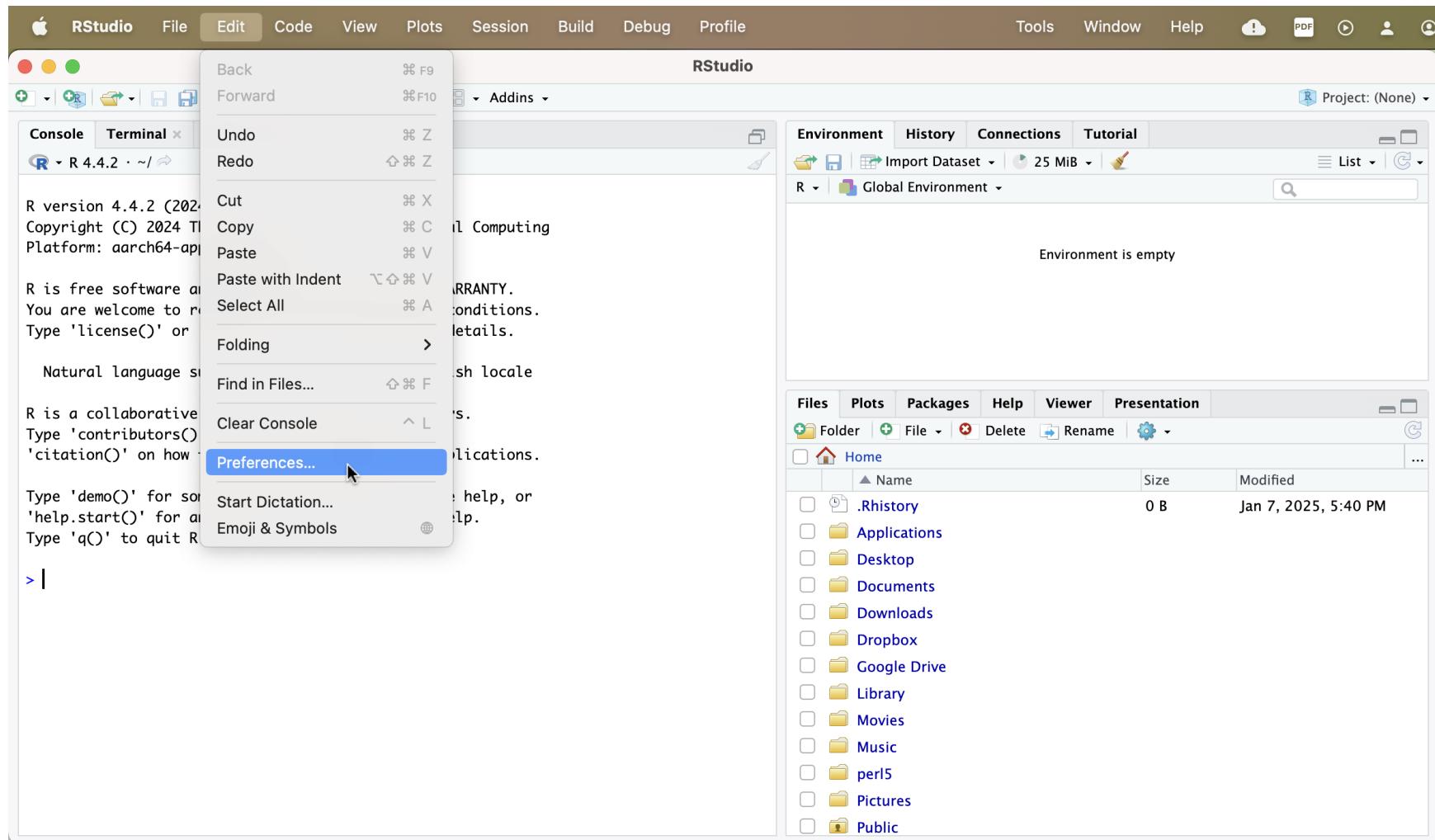
Last login: Mon Jan 20 00:46:28 on ttys000  
The default interactive shell is now zsh.  
To update your account to use zsh, please run 'chsh -s /bin/zsh'.  
For more details, please visit <https://support.apple.com/kb/HT200050>.  
(base) MacBook-Air:~ tatianna\$

R version 4.4.2 (2024-03-31) -- "Pile of Leaves"  
Copyright (C) 2024 The R Foundation for Statistical Computing  
Platform: darwin-arm64 (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type "license()" or "licence()" for distribution details.  
  
Natural language support not running in an English locale  
R is a collaborative project with many contributors.  
Type "contributors()" for more information and  
"citation()" on how to cite R or R packages in publications.  
Type "demo()" for some demos, "help()" for on-line help, or  
"help.start()" to launch a web browser interface to help.  
Type "q()" to quit R.  
  
[R.app GUI 1.1 (4462) arm64-64-apple-darwin20]  
[History restored from /Users/tatianna/Rapp.history]  
>

```
1 # Analysis of differential gene expression using DESeq2
18 ## Step 1: Load data ----
30 samples <- list.files(path = "~/rnaseq/06-DGEmtDNA", full.names = T, recursive = TRUE)
31 ## List all directories containing data
32 samples <- list.files(path = "~/rnaseq/06-DGEmtDNA", full.names = T, recursive = TRUE)
33
34 ## Obtain a vector of all filenames including the path
35 files <- file.path(samples, "quant.sf")
36
37 ## Give unique name for each element
38 names(files) <- str_replace(samples, "Users/tatianna/rnaseq/06-DGEmtDNA",
39 str_replace("mtDNA-salmon", ""))
40
41 ## Load the annotation table for the mtDNA genes
```

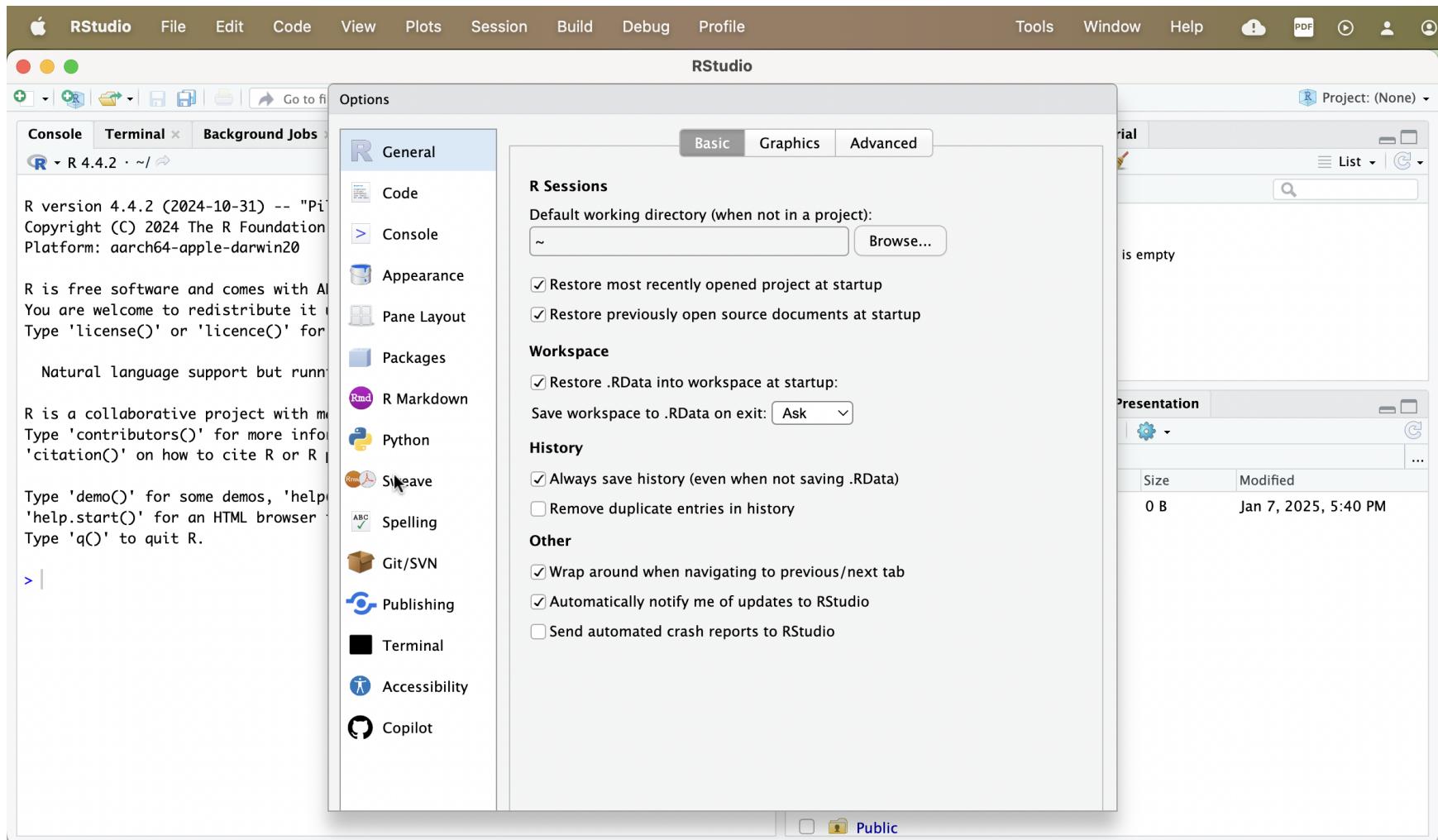
# Installing software

## RStudio = IDE



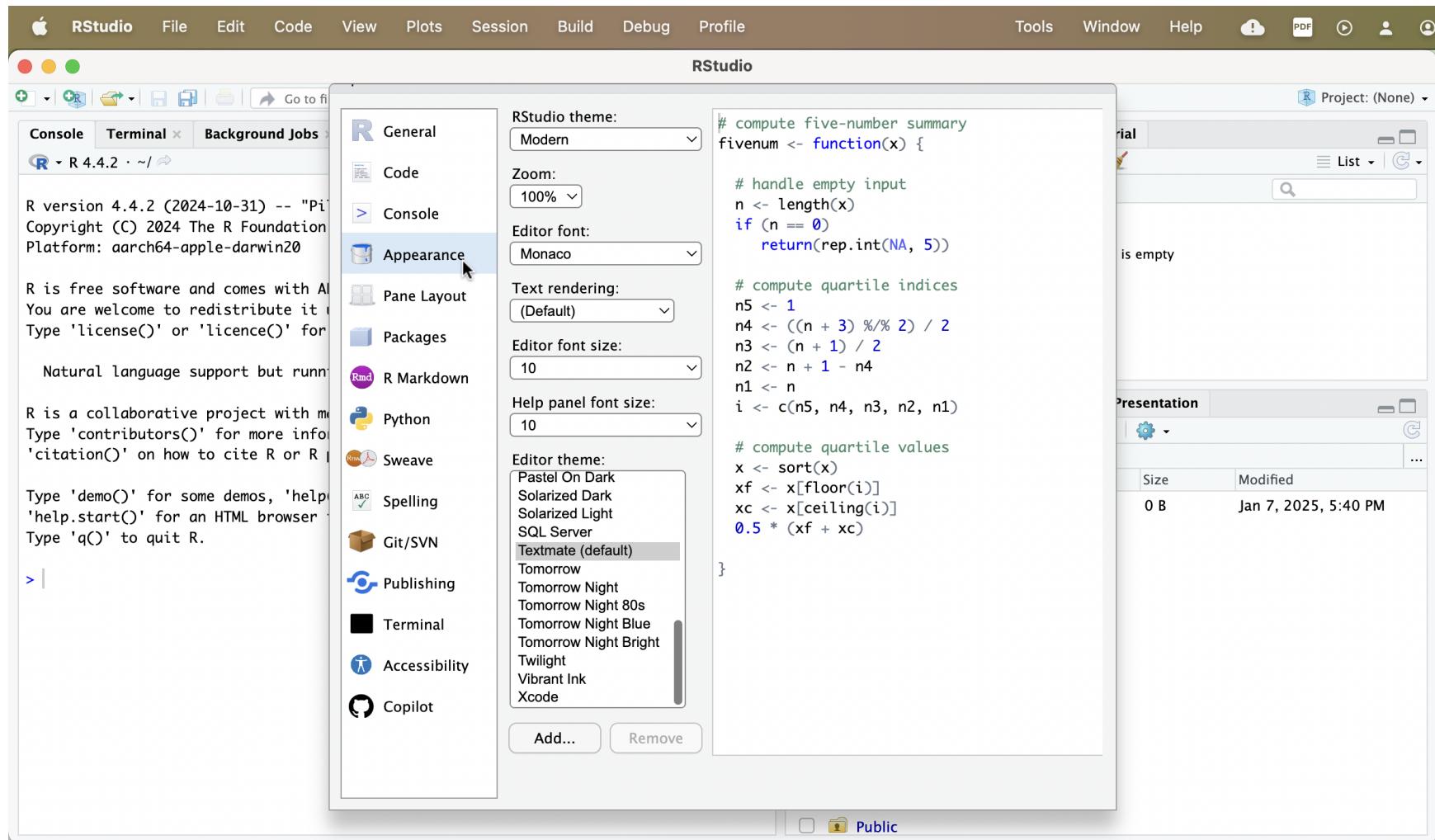
# Installing software

## RStudio = IDE



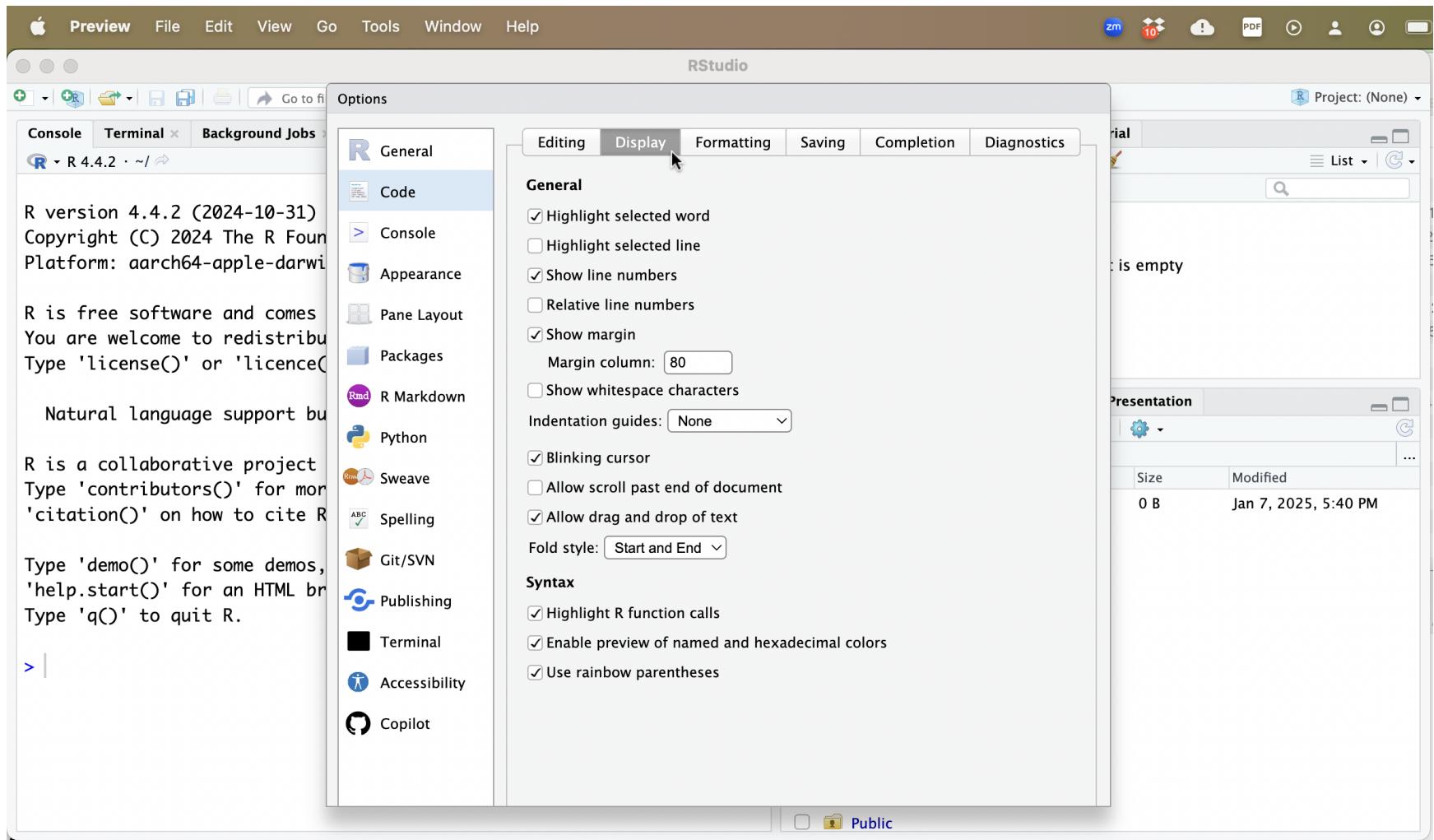
# Installing software

## RStudio = IDE



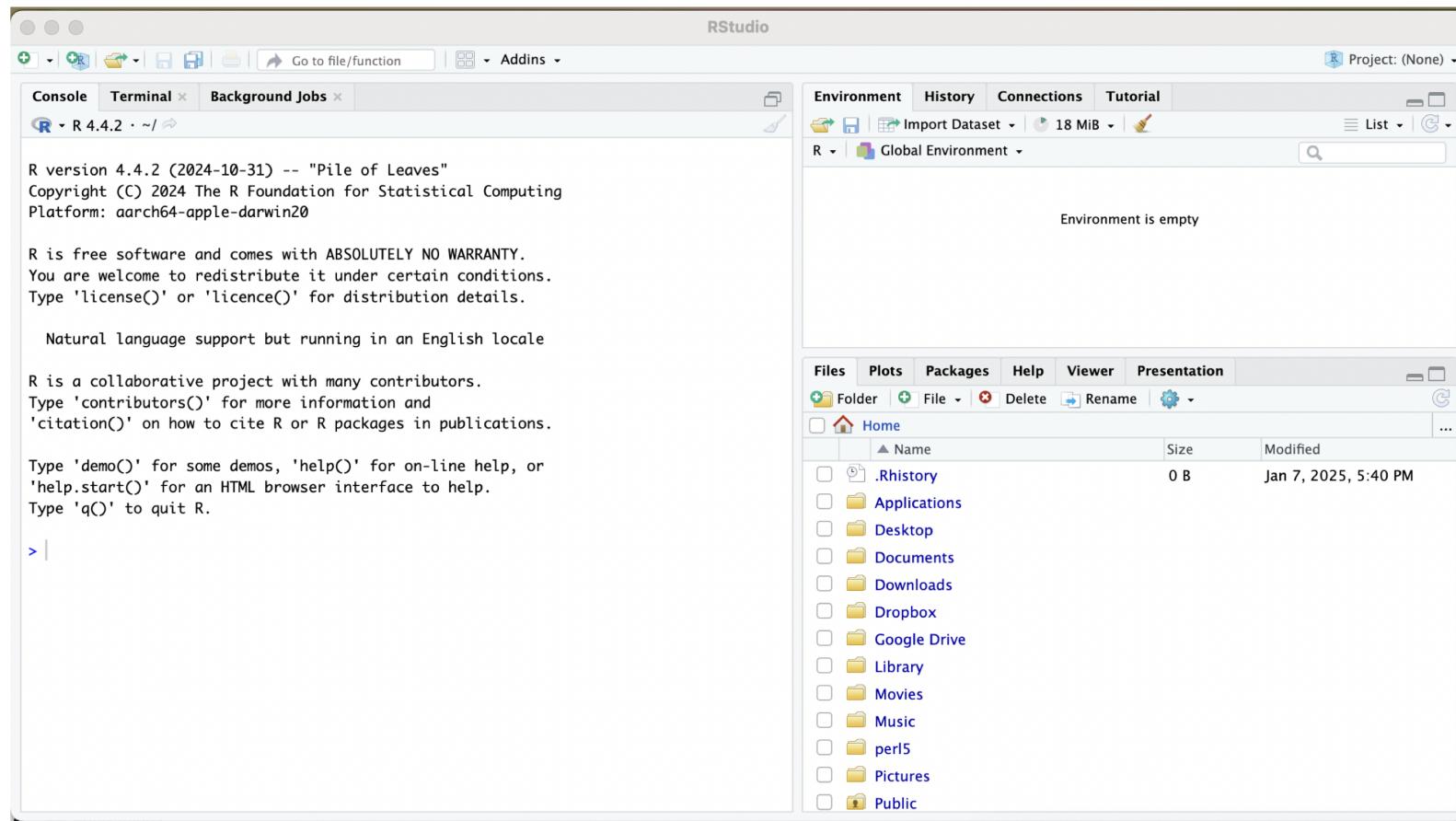
# Installing software

## RStudio = IDE



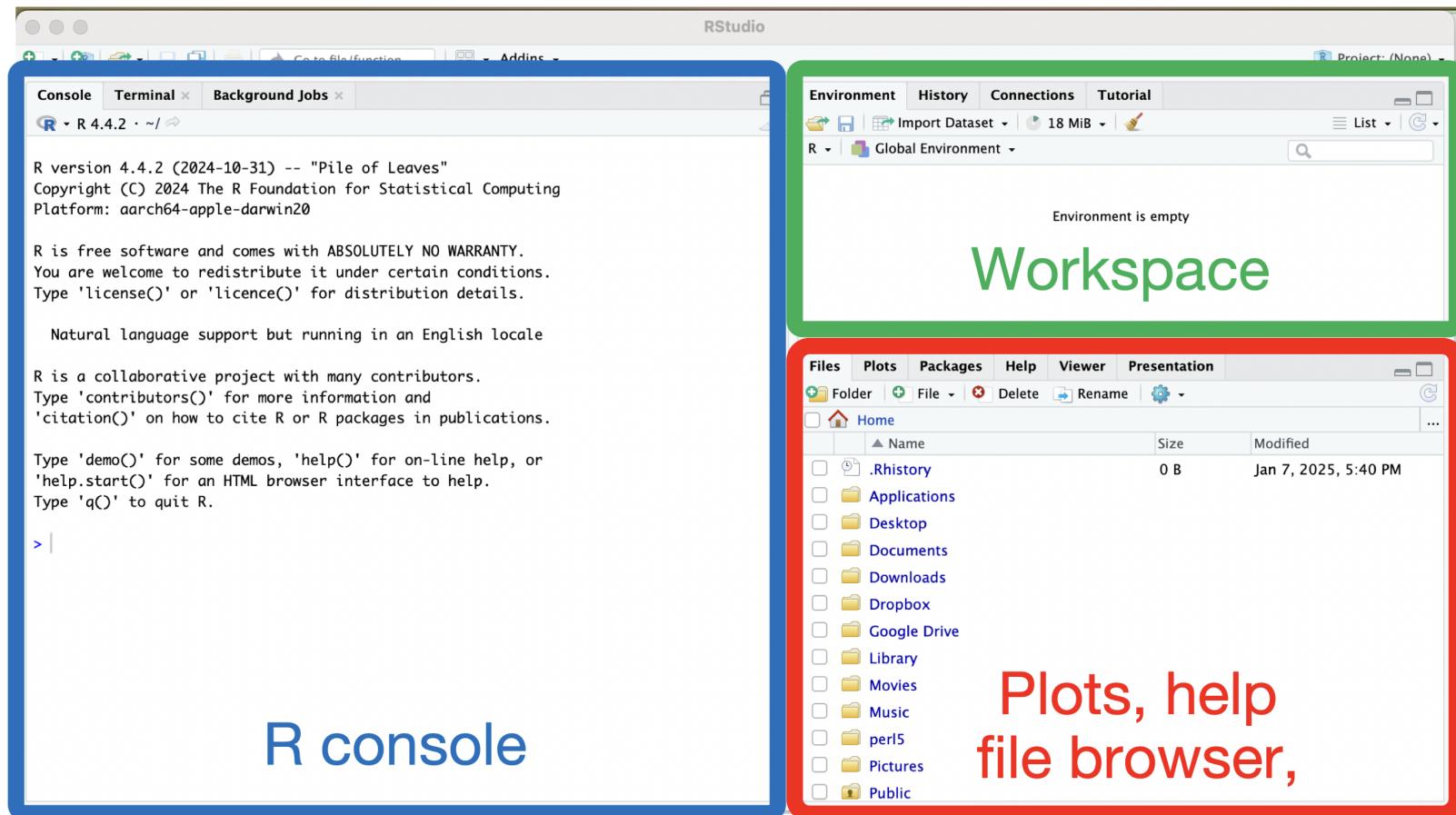
# Installing software

## RStudio



# Installing software

## RStudio



# R for differential expression analysis

## 1. Comprehensive Statistical Framework

- **Sophisticated DE Tools:** R hosts leading packages like **DESeq2**, **edgeR**, and **limma**, which are specifically designed for RNA-seq DE analysis.
- **Customizability:** R allows users to tailor statistical models, such as using alternative normalization methods or fitting custom covariates to account for complex experimental designs.

# R for differential expression analysis

## 2. Extensive Visualization Capabilities

- **Built-in Plotting Tools:** R includes libraries like `ggplot2`, `base R graphics`, and `ComplexHeatmap`, enabling the generation of high-quality visualizations such as:
  - MA plots
  - Volcano plots
  - Heatmaps of expression patterns
  - Principal Component Analysis (PCA)
- **Customizable Graphics:** Users can fully customize visualizations to fit publication or presentation standards.

# R for differential expression analysis

## 3. Robust Handling of RNA-seq Data

- **Normalization Methods:** Packages in R implement state-of-the-art normalization techniques, such as variance-stabilizing transformation (VST) and trimmed mean of M values (TMM), which account for library size and sequencing depth.
- **Batch Effect Correction:** Tools like `sva` and `removeBatchEffect` in `limma` are available for correcting batch effects, ensuring reliable downstream analysis.

# R for differential expression analysis

## 4. Open-Source and Well-Documented

- **Cost-free:** R is free and open-source, making it accessible to researchers worldwide.
- **Rich Documentation:** Most RNA-seq packages are accompanied by detailed vignettes, tutorials, and example datasets.
- **Active Community:** The Bioconductor project and broader R community provide extensive forums, mailing lists, and user support.

# R for differential expression analysis

## 5. Breadth of Bioconductor Ecosystem

- **RNA-seq-Specific Packages:** The Bioconductor project includes hundreds of packages for preprocessing, quality control, differential expression, and downstream functional analysis .
- **Integration:** Bioconductor tools are designed to work seamlessly together, minimizing data compatibility issues.

# Installing software

## Bioconductor

1. In the course webpage, click on the Bioconductor website
2. Click on "Get started"

# Installing software

## Bioconductor



About   Learn   Packages   Developers

Search

Get Started >

### Get started

#### 1. Install R

The current release of Bioconductor is version 3.20; it works with R version 4.4.0. Users of older R and Bioconductor must update their installation to take advantage of new features and to access packages that have been added to Bioconductor since the last release.

The development version of Bioconductor is version 3.21; it works with R version 4.5.0. More recent ‘devel’ versions of R (if available) will be supported during the next Bioconductor release cycle.

✓ Step 1  
Install R

1. Download the most recent version of R. The R FAQs and the R Installation and Administration Manual contain detailed instructions for installing R on various platforms (Linux, OS X, and Windows being the main ones).
2. Start the R program; on Windows and OS X, this will usually mean double-clicking on the R application, on UNIX-like systems, type “R” at a shell prompt.

#### 2. Get the latest version of Bioconductor

Once R has been installed, get the latest version of Bioconductor by starting R and entering the following commands.

It may be possible to change the Bioconductor version of an existing installation; [see the ‘Changing version’ section of the BiocManager vignette](#).

Details, including instructions to [install additional packages](#) and to [update](#), [find](#), and [troubleshoot](#) are provided below. A [devel](#) version of Bioconductor is available. There are good [reasons for using BiocManager::install\(\)](#) for managing Bioconductor resources.

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.20")
```

Step 2  
Get Bioconductor

# Installing software

## Bioconductor

1. In the course webpage, click Bioconductor
2. Click on "Get started"
3. Copy code

# Installing software

## Bioconductor



About   Learn   Packages   Developers

Search

Get Started >

### Get started

#### 1. Install R

The current release of Bioconductor is version 3.20; it works with R version 4.4.0. Users of older R and Bioconductor must update their installation to take advantage of new features and to access packages that have been added to Bioconductor since the last release.

The development version of Bioconductor is version 3.21; it works with R version 4.5.0. More recent ‘devel’ versions of R (if available) will be supported during the next Bioconductor release cycle.

✓ Step 1  
Install R

1. Download the most recent version of R. The R FAQs and the R Installation and Administration Manual contain detailed instructions for installing R on various platforms (Linux, OS X, and Windows being the main ones).
2. Start the R program; on Windows and OS X, this will usually mean double-clicking on the R application, on UNIX-like systems, type “R” at a shell prompt.

#### 2. Get the latest version of Bioconductor

Once R has been installed, get the latest version of Bioconductor by starting R and entering the following commands.

It may be possible to change the Bioconductor version of an existing installation; see the ‘[Changing version](#)’ section of the [BiocManager vignette](#).

Details, including instructions to [install additional packages](#) and to [update](#), [find](#), and [troubleshoot](#) are provided below. A [devel](#) version of Bioconductor is available. There are good [reasons for using](#) [\*\*BiocManager::install\(\)\*\*](#) for managing Bioconductor resources.

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.20")
```

Step 2  
Get Bioconductor

# Installing software

## Bioconductor

1. In the course webpage, click Bioconductor
2. Click on "Get started"
3. Copy code

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install(version = "3.20")
```

5. Paste code on Rstudio prompt

# Bioconductor

5. Paste code on Rstudio prompt

```
Upgrade 47 packages to Bioconductor version '3.20'? [y/n]:
```

# Bioconductor

## 5. Paste code on Rstudio prompt

```
Upgrade 47 packages to Bioconductor version '3.20'? [y/n]: y
```

# Bioconductor

5. Paste code on Rstudio prompt

```
Upgrade 47 packages to Bioconductor version '3.20'? [y/n]: y
```

```
Do you want to install from sources the package which  
needs compilation? (Yes/no/cancel)
```

# Bioconductor

5. Paste code on Rstudio prompt

```
Upgrade 47 packages to Bioconductor version '3.20'? [y/n]: y
```

```
Do you want to install from sources the package which  
needs compilation? (Yes/no/cancel) yes
```

# Installing software

Conda for installing multiple softwares



# Installing software

## Conda

Conda is an open source package management system and environment management system for installing multiple versions of software packages and their dependencies and switching easily between them. It works on Linux, OS X and Windows, and was created for Python programs but can package and distribute any software.

# Installing software

## Conda

### Why should I use conda?

- Simplified Package Management
- Cross-Platform Compatibility
- Extensive Package Ecosystem
- Environment Management
- Conda Channels

<https://conda.org/learn/faq>

# Installing software

## CONDA



[Learn](#) [Community](#) [Blog](#)

[GitHub](#)



[Search](#)



# CONDA COMMUNITY

A community supporting a language-agnostic, multi-platform package management ecosystem for projects of any size and complexity.

[Learn More](#)



# Installing software

## CONDA

### Should I use Anaconda Distribution or Miniconda?

Both the Anaconda Distribution and Miniconda installers include the conda package and environment manager, but how you plan to use the software will determine which installer you want to choose.

[Install Anaconda Distribution](#)

[Install Miniconda](#)

	Anaconda Distribution	Miniconda
Created and published by Anaconda	Yes	Yes
Has conda	Yes	Yes
Has <a href="#">Anaconda Navigator</a>	Yes	No
# of packages	300+	< 70
Install space required	~4.4 GB	~480 MB

# Installing software

## Miniconda

1. In the course webpage, click Miniconda
2. On "Install", download the appropriate distribution

<https://docs.conda.io/projects/conda/en/stable/>

# Installing software

## Miniconda

**CONDA**

Conda Conda-build Miniconda conda.org

Search K ⌂ ⌂ ⌂ ⌂ ⌂ ⌂

### Install ↓

We recommend the following conda distributions to install conda:

**Miniconda**

Miniconda is an installer by Anaconda that comes preconfigured for use with the Anaconda Repository. See the notes about Anaconda's [Terms of Service](#).

Windows x86\_64 ↓

macOS arm64 (Apple Silicon) ↓

macOS x86\_64 (Intel) ↓

Linux x86\_64 (amd64) ↓

Linux aarch64 (arm64) ↓

**Miniforge**

Miniforge is an installer maintained by the [conda-forge community](#) that comes preconfigured for use with the conda-forge channel.

Windows x86\_64 ↓

macOS arm64 (Apple Silicon) ↓

macOS x86\_64 (Intel) ↓

Linux x86\_64 (amd64) ↓

Linux aarch64 (arm64) ↓

On this page

Install ↓

New to conda? ↗

Other useful resources ↗

Contributors welcome ↗

Edit on GitHub

Show Source

stable

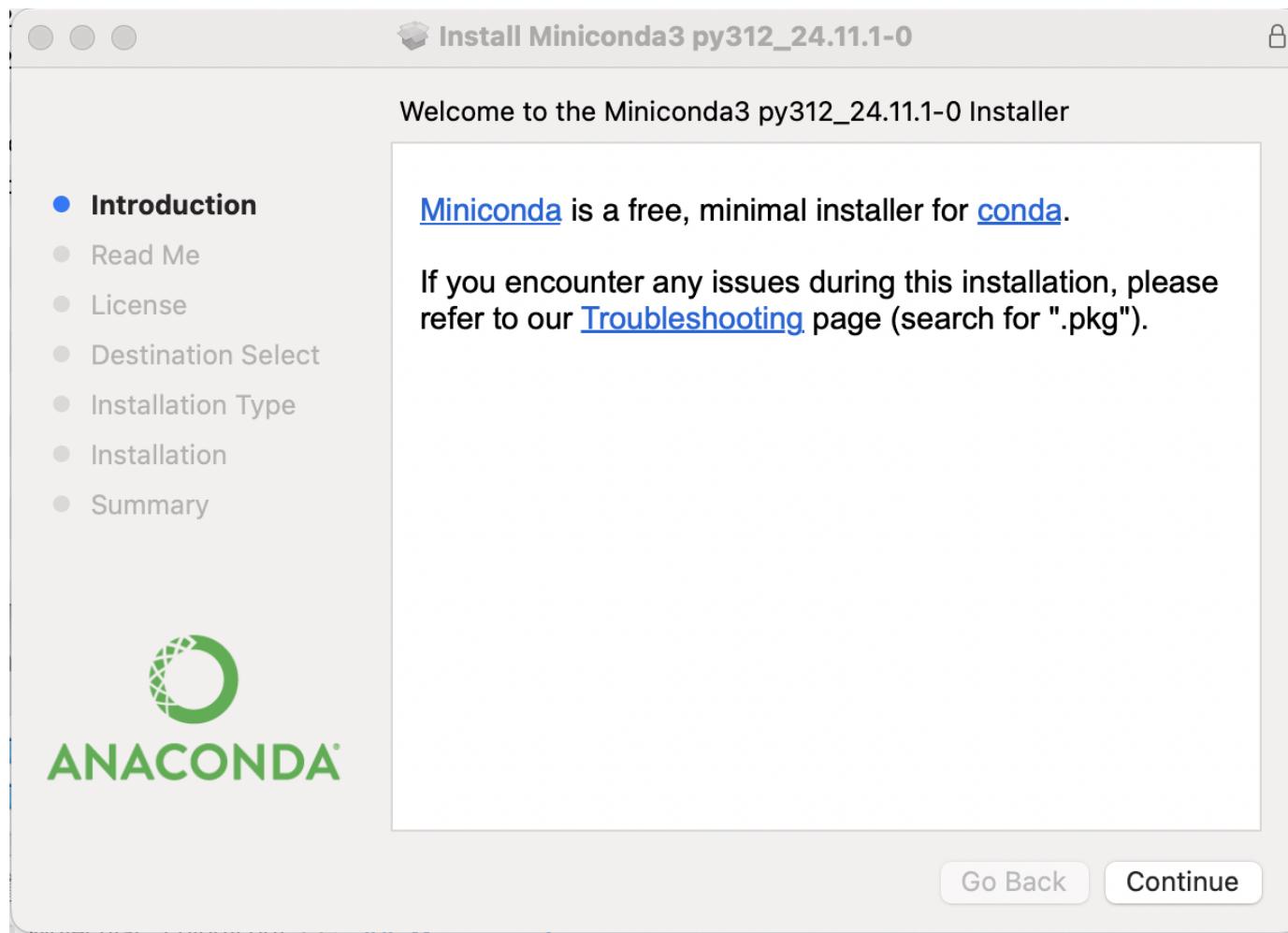
# Installing software

## Miniconda: Mac OS

1. In the course webpage, click Miniconda
2. On "Install", download the appropriate distribution
3. Double-click the downloaded package `.pkg`
4. Follow the instructions on the screen.

# Installing software

## Miniconda: Mac OS



# Installing software

## Miniconda: Windows

1. In the course webpage, click Miniconda
2. On "Install", download the appropriate distribution
3. Double-click the installer `.exe`
4. Follow the instructions on the screen.

# Installing software

## Miniconda: Windows

1. In the course webpage, click Miniconda
2. On "Install", download the appropriate distribution
3. Double-click the installer `.exe`
4. Follow the instructions on the screen.

# Installing software

## Miniconda: Linux

1. In the course webpage, click Miniconda
2. On "Install", download the appropriate distribution
3. Run the following command:

```
bash ~/Miniconda3-latest-Linux-x86_64.sh
```

4. Press Return to review the license agreement. Then press and hold Return to scroll.

# Installing software

## Miniconda: Linux

5. Enter `yes` to agree to the license agreement.
6. Press Return to accept the default install location, or enter another file path to specify an alternate installation directory.
7. Choose an initialization option: `Yes`
8. Close and re-open your terminal window

# Installing software

## Miniconda: all

- Test your installation by running `conda info`. If conda has been installed correctly, a list of installed packages appears.
- This will display view all the details about your conda set-up

# Installing software

## Configuring Conda installation

```
conda info --envs #to view all the environments  
# available to you (note, since you just installed  
# miniconda, you'll only have a 'base' environment  
# available)
```

# Installing software

## Configuring Conda installation

- Channels with many pre-packaged bioinformatics programs can be downloaded with all their dependencies

```
conda config --add channels defaults
conda config --add channels bioconda
conda config --add channels conda-forge
conda config --set offline false
```

# Installing software

## Conda environments

In Conda, environments are isolated spaces where you can install and manage specific packages. Each environment operates independently, allowing you to work on multiple projects with distinct dependencies and configurations without conflicts.

# Installing software

## Conda environments

- Dependency Isolation: Prevents conflicts between projects by isolating libraries and configurations.
- Experimentation: Enables safe testing of new tools or configurations in disposable environments.
- Organization: Keeps projects separate and manageable with intuitive environment names.
- Collaboration: Provides consistency for teams, ensuring everyone works in identical setups.
- System Safety: Protects the base system or global installations by confining changes to the environment.

# Installing software

## Creating a Conda environment

1. Creating an environment called 'rnaseq'

```
conda create --name rnaseq
```

2. Activate the new environment

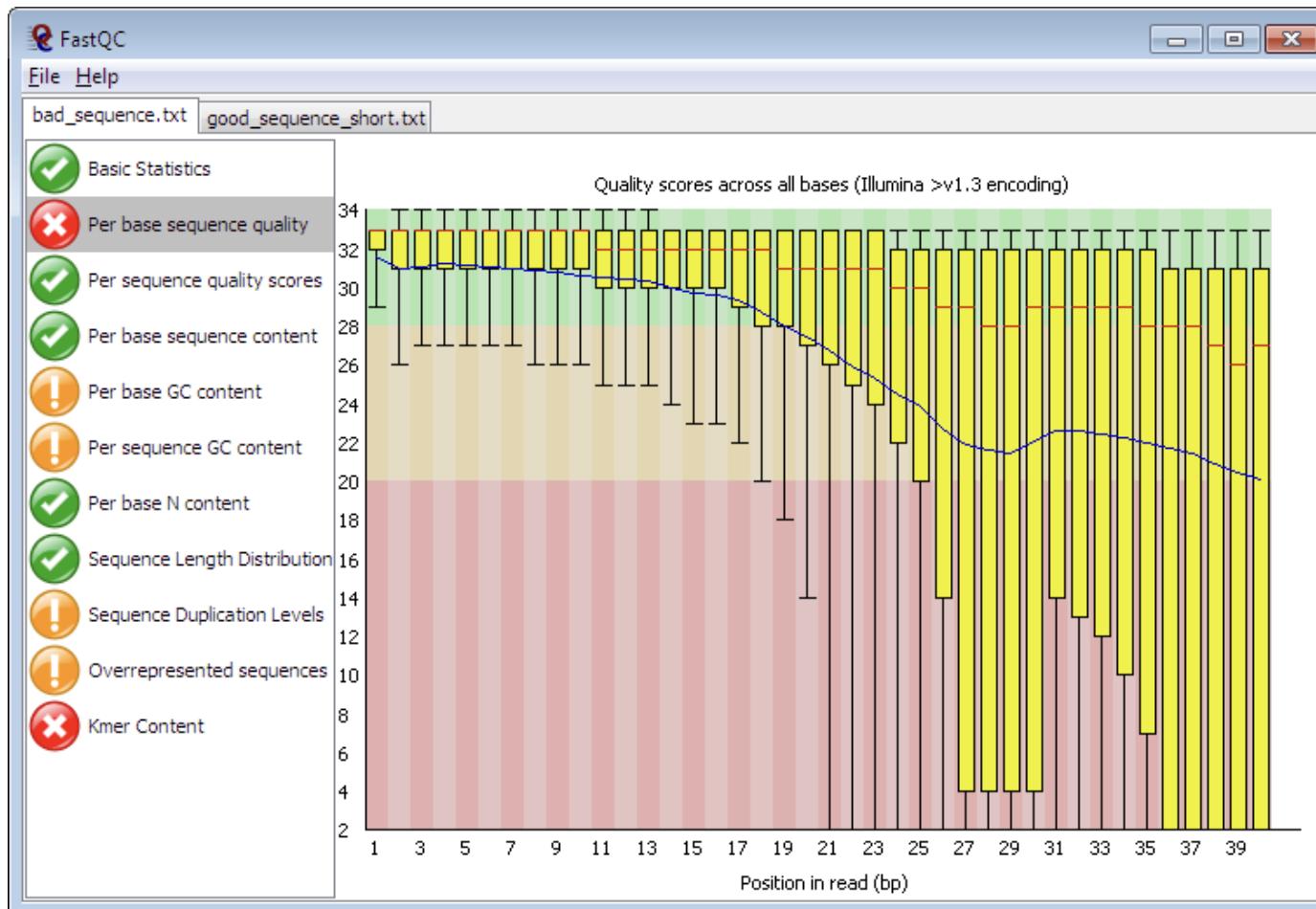
```
conda activate rnaseq
```

3. Deactivate to return to the `base` environment

```
conda deactivate
```

# Installing software using CONDA

## FastQC



<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

# Installing software using CONDA

## FastQC

The main functions of FastQC are:

- Import of data from BAM, SAM or FastQ files
- Providing a quick overview to tell you in which areas there may be problems
- Summary graphs and tables to quickly assess your data
- Export of results to an HTML based permanent report
- Offline operation to allow automated generation of reports without running the interactive application

# Installing software using CONDA

## FastQC

1. In the 'rnaseq' environment, run:

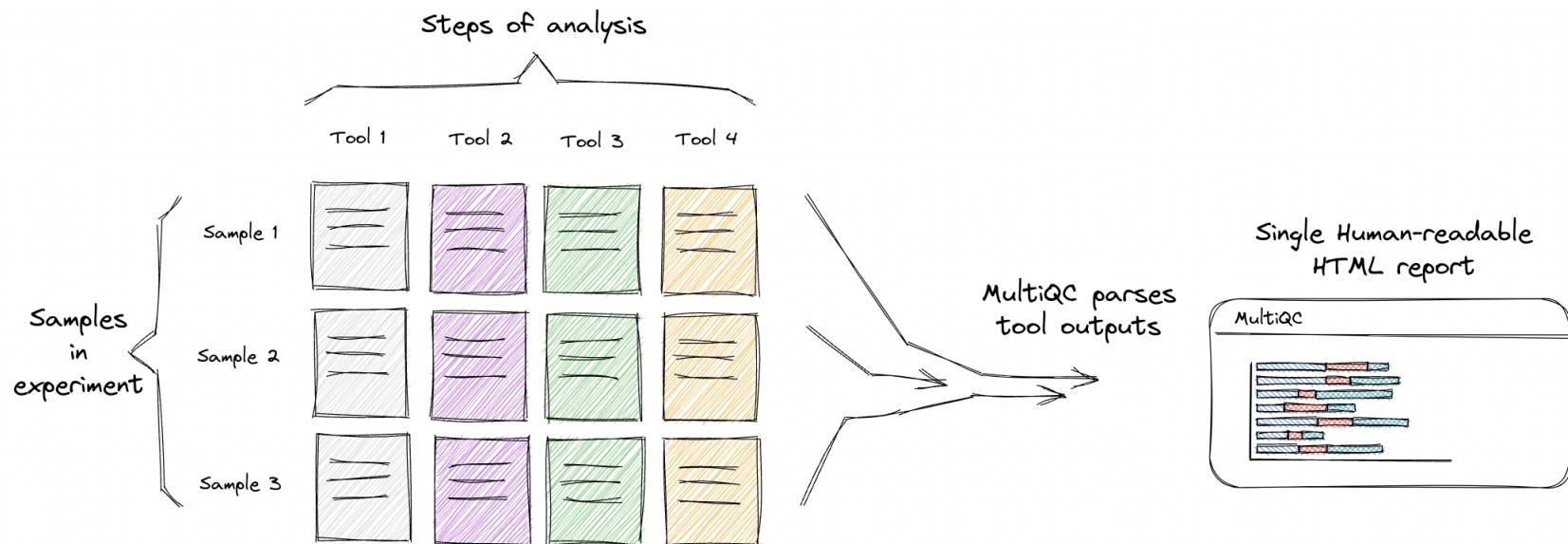
```
conda install -c bioconda fastqc
```

2. Test installation

```
fastqc
```

# Installing software using CONDA

## MultiQC



<https://docs.seqera.io/multiqc>

# Installing software using CONDA

## MultiQC

Typical uses of MultiQC are:

- Quality Control Summaries: combine results from FastQC, trimming tools, and alignment software into one report to streamline QC checks.
- Pipeline Monitoring: monitor and report on all pipeline steps, ensuring consistency and data integrity.
- Project Reports: quickly generate comprehensive summaries for collaborators or publication.

# Installing software using CONDA

## MultiQC

1. In the 'rnaseq' environment, run:

```
conda install -c bioconda multiqc
```

2. Test installation

```
multiqc --help
```

# Installing software using CONDA

## Trimmomatic

USADELLAB.org

Home      Research      Projects      Education      Service & Software      Publications  
Supporting Info      About Us      NGS, DE and other things      Data Protection

---

### Trimmomatic: A flexible read trimming tool for Illumina NGS data

#### Citations

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

#### Downloading Trimmomatic

starting on version 0.40 we also offer a [github page](#) (as well as older versions)

Version 0.39: [binary](#), [source](#) and [manual](#)

Version 0.36: [binary](#) and [source](#)

#### Quick start

##### Paired End:

With most new data sets you can use gentle quality trimming and adapter clipping.

<http://www.usadellab.org/cms/?page=trimmomatic>

# Installing software using CONDA

## Trimmomatic

Uses of Trimmomatic are:

- Cut adapter and other illumina-specific sequences from the read.
- Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold.
- Cut bases off the start of a read, if below a threshold quality
- Cut bases off the end of a read, if below a threshold quality
- Cut the read to a specified length
- Cut the specified number of bases from the start of the read
- Drop the read if it is below a specified length
- Convert quality scores to Phred-33/Phred-64

# Installing software using CONDA

## Trimmomatic

1. In the 'rnaseq' environment, run:

```
conda install -c bioconda trimmomatic
```

2. Test installation

```
trimmomatic
```

# Installing software using CONDA

## Trinity

---

### Home

Brian Haas edited this page on May 6, 2024 · [41 revisions](#)

---

### RNA-Seq De novo Assembly Using Trinity

As of 2024, Trinity RNA-seq project no longer has dedicated funding for future development, maintenance, or support.  
Please kindly consider this as you post issues or are in search of assistance.



#### Quick Guide for the Impatient

---

Trinity assembles transcript sequences from Illumina RNA-Seq data.

Download Trinity [here](#).

Build Trinity by typing 'make' in the base installation directory.

<https://github.com/trinityrnaseq/trinityrnaseq/wiki>

# Installing software using CONDA

## Trinity

Uses of Trinity are:

- Inchworm assembles the RNA-seq data into the unique sequences of transcripts, often generating full-length transcripts for a dominant isoform.
- Chrysalis clusters the Inchworm contigs into clusters and constructs complete de Bruijn graphs for each cluster. Each cluster represents the full transcriptonal complexity for a given gene.
- Butterfly reports full-length transcripts for alternatively spliced isoforms, and teasing apart transcripts that corresponds to paralogous genes.

# Installing software using CONDA

## Trinity

1a. Unix: Create a new environment `trinity` :

```
conda create --name trinity
conda activate trinity
```

1b. MacOS: Create a new environment `trinity` :

```
CONDA_SUBDIR=osx-64 conda create -n trinity
conda activate trinity
conda env config vars set CONDA_SUBDIR=osx-64
```

# Installing software using CONDA

## Trinity dependencies

2. Install trinity

```
conda install trinity
```

3. Test installation

```
Trinity
```

# Installing software using CONDA

## BUSCO

### Benchmarking Universal Single-Copy Orthologue (BUSCO)

Provides evolutionarily sound measures of completeness and redundancy in terms of expected gene content

This is achieved by considering the presence of a predefined and expected set of single-copy marker genes as a proxy for genome-wide completeness.

<https://busco.ezlab.org>

# Installing software using CONDA

## BUSCO

1. Create a new environment (BUSCO has conflicts with Trinity):
  - a. Unix: Create a new environment `annotation`:

```
conda create --name annotation  
conda activate annotation
```

- b. MacOS: Create a new environment `annotation`:

```
CONDA_SUBDIR=osx-64 conda create -n annotation  
conda activate annotation  
conda env config vars set CONDA_SUBDIR=osx-64
```

2. Install BUSCO

```
conda install busco
```

# Installing software using CONDA

## BUSCO

### 3. Install BUSCO

```
conda install busco
```

### 4. Test installation

```
busco --help
```

# Installing software using CONDA

## Trinotate

Annotation suite designed for automatic functional annotation of transcriptomes, particularly de novo assembled transcriptomes, from model or non-model organisms

# Installing software using CONDA

## Trinotate

Trinotate: Transcriptome Functional Annotation and Analysis



# Installing software using CONDA

## TransDecoder

TransDecoder predicts open reading frames (ORFs) in transcript sequences. These ORFs represent potential protein-coding regions.

### Key Steps:

1. **LongOrfs:** Identifies likely coding regions based on sequence properties.
2. **Predict:** Refines ORF predictions by incorporating similarity and coding potential.

<https://github.com/TransDecoder/TransDecoder/wiki>

# Installing software using CONDA

## TransDecoder

1. Activate environment:

```
conda activate annotation
```

2. Install TransDecoder

```
conda install transdecoder
```

# Installing software using CONDA

# TransDecoder

### 3. Test installation:

## TransDecoder.LongOrfs

# Installing software using CONDA

## SignalP

Identifying signal peptides during genome or transcriptome annotation is crucial for understanding the functionality and localization of proteins.

**But, this step will not be done during the course**

SignalP is licensed software. It is necessary to provide e-mail and affiliation and to agree to the terms of the academic license before downloading.

# Installing software using CONDA

## SignalP

1. Create and activate environment:

```
conda create -n signalp python=3.7 numpy=1.16.4  
conda activate signalp
```

2. Visit the [SignalP official website](#).
3. Download the macOS version of SignalP.
4. Locate the `.tar.gz` file.
5. Extract the package:

```
tar -xvzf signalp-*.tar.gz  
cd signalp-*
```

# Installing software using CONDA

## SignalP

6. Activate environment:

```
conda activate annotation
```

2. Visit the [SignalP official website](#).

3. Download the macOS version of SignalP.

4. Locate the `.tar.gz` file.

5. Extract the package:

```
tar -xvzf signalp-*.tar.gz  
cd signalp-*
```

# RNA-Seq: Annotation

## SignalP

6. Install SignalP with the package manager for Python.

```
pip install signalp-6-package/
```

7. Copy the model files to the location where the signalp got installed.

```
SIGNALP_DIR=$(python3 -c "import signalp; import os; \
print(os.path.dirname(signalp.__file__))" )
cp -r signalp-6-package/models/* $SIGNALP_DIR/model_weights
```

8. Test installation

```
signalp6 -h
```

# Installing software using CONDA

## EggnoG-Mapper

1. Activate environment:

```
conda activate annotation
```

2. Install EggnoG-Mapper

```
conda install -c bioconda -c conda-forge eggnog-mapper
```

<http://eggnog-mapper.embl.de>

# Installing software using CONDA

## EggnoG-Mapper

3. Download the eggNOG-mapper databases

```
download_eggnog_data.py -H -d 6656 --dbname Arthropoda
```

**6,31GB just the Arthropoda database**

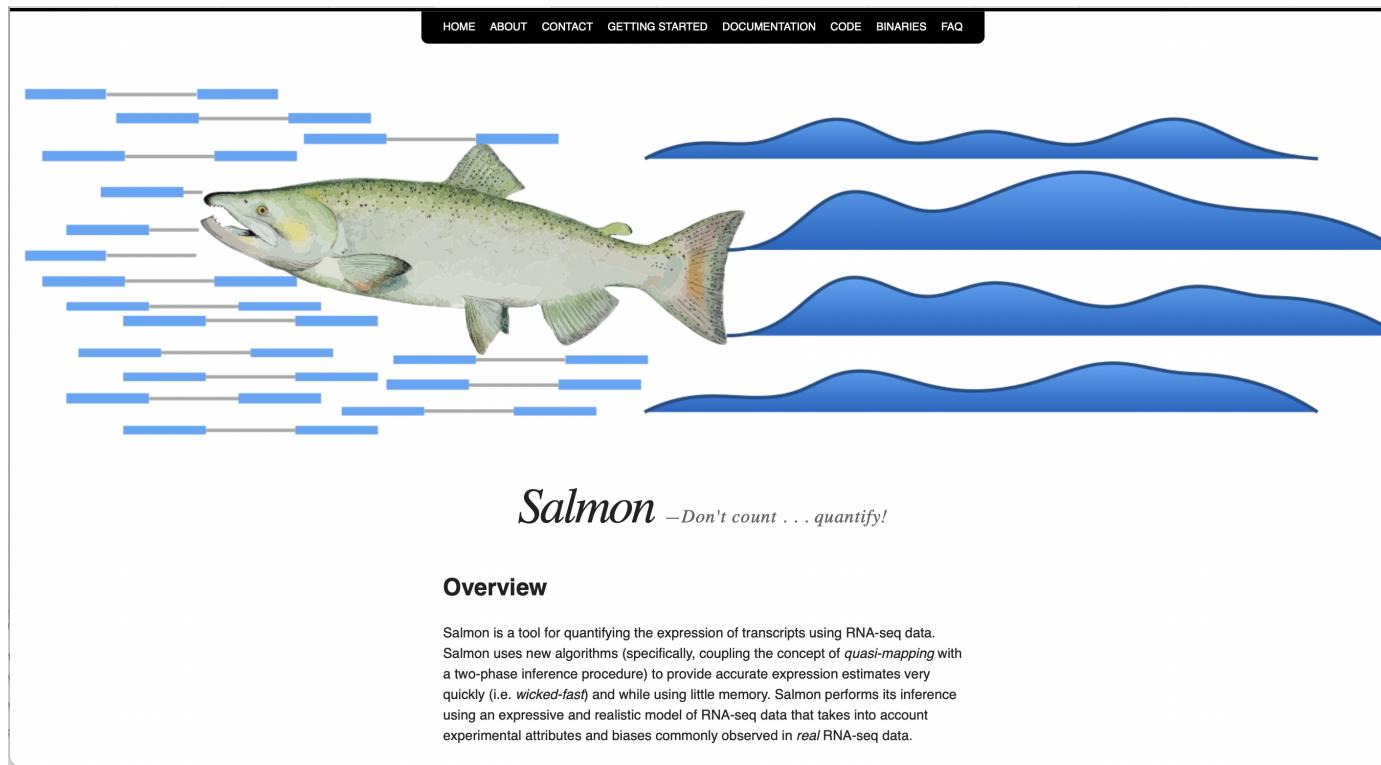
- -H:install the HMMER database
- -d:Tax ID of eggNOG HMM database to download

6656: arthropoda

--dbname: Tax ID of eggNOG HMM database to download

# Installing software using CONDA

## Salmon



<https://combine-lab.github.io/salmon/>

# Installing software using CONDA

## Salmon

1. Create a new environment:

a. Unix: Create a new environment `quant` :

```
conda create --name quant
conda activate quant
```

b. MacOS: Create a new environment `annotation` :

```
CONDA_SUBDIR=osx-64 conda create -n quant
conda activate quant
conda env config vars set CONDA_SUBDIR=osx-64
```

# Installing software using CONDA

## Salmon

### 2. Install Salmon

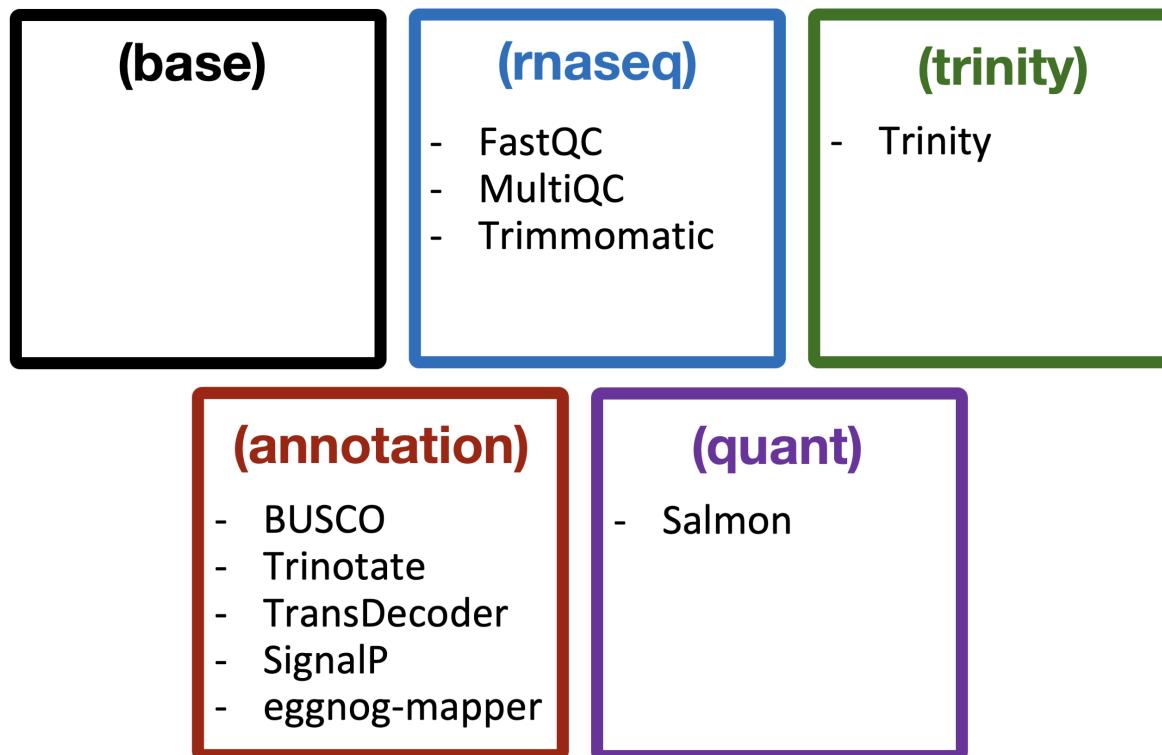
```
conda install -c bioconda salmon
```

### 2. Test installation

```
salmon -h
```

# Installing software using CONDA

## CONDA environments



# Installing software using Docker

## Issues with any installation

Trinotate in Conda does not work for me

The screenshot shows the Docker website homepage. At the top, there is a dark blue header bar with the text "Incident Update: Docker Desktop for Mac. Learn more →" on the left, and "Docs", "Get support", and "Contact sales" on the right. Below the header is a navigation bar with the Docker logo on the left, followed by "Products ▾", "Developers ▾", "Pricing", "Support", "Blog", and "Company ▾". To the right of the navigation bar are a search icon, a "Sign In" button, and a blue "Get started" button. The main content area features a large, bold, dark blue headline: "Develop faster. Run anywhere." Below the headline is the subtext "Build with the #1 most-used developer tool". At the bottom of the main content area are two buttons: a blue "Download Docker Desktop" button and a white "Learn more about Docker" button. The background of the main content area has a subtle gradient and some abstract blue shapes.

Incident Update: Docker Desktop for Mac. Learn more →

Docs Get support Contact sales

docker Products ▾ Developers ▾ Pricing Support Blog Company ▾

Sign In Get started

# Develop faster. Run anywhere.

Build with the #1 most-used developer tool

Download Docker Desktop

Learn more about Docker

# Installing software using Docker

## Issues with Trinotate installation

1. Download docker.
2. Follow the instructions for installation
3. Open Docker Desktop and create a login
4. On Terminal:

```
docker --version
```

# Installing software using Docker

## Issues with Trinotate installation

5. Download Trinotate docker image

```
docker pull trinityrnaseq/trinotate
```

6. Run Trinotate interactively

```
docker run --rm -it \
    -v `pwd`:/data \
    -v /tmp:/tmp \
    -e TRINOTATE_HOME=/usr/local/src/Trinotate \
    trinityrnaseq/trinotate bash
```

7. Test Trinotate

```
$TRINOTATE_HOME/Trinotate
```