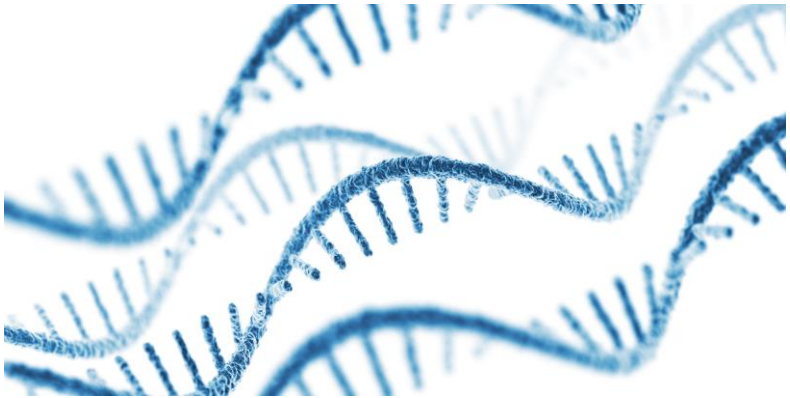
Setup for RNA-seq Workshop

24/04/2025

by Jiajia Li

ANU Biological Data Science Institute



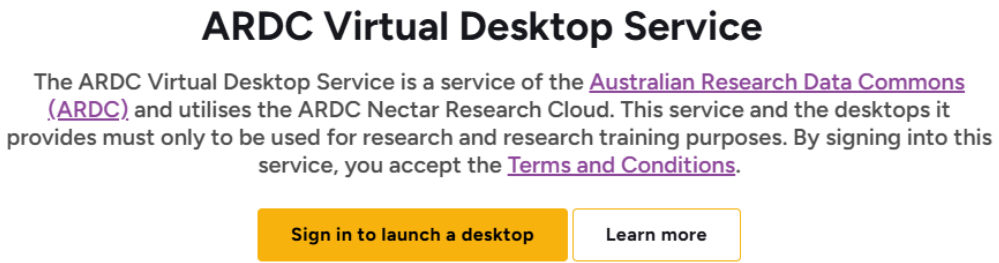
Australian National University

Software

- Unix-like Operating System (Linux, WSL, or MacOS)
- Command-line tools:
  - hisat2, fastqc, fastp, multiqc, samtools, java, picard tools, stringtie, htseq-count
- Integrative Genome Browser (IGV)
- R, Rstudio, and libraries:
  - Bioconductor, ggplot2, data.table, ballgown, genefilter, dplyr, devtools, edgeR, DESeq2, …

Device

We suggest using ARDC's Virtual Desktop Service https://desktop.rc.nectar.org.au/



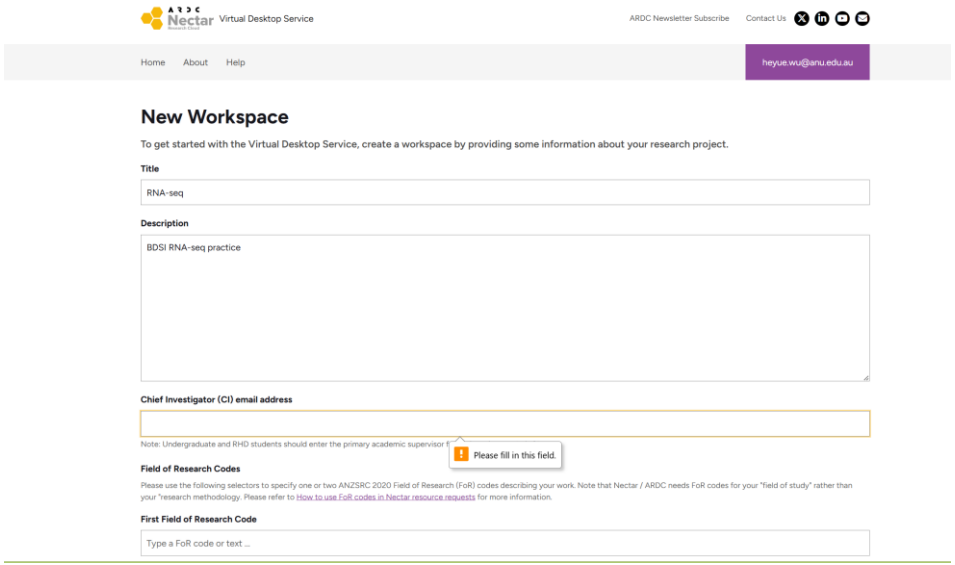
Use your university email address to sign in. Then you will be guided to the home page.

If you prefer using your own device, and you are confident in your computer skills, you can refer to this tutorial and adapted to your own system.

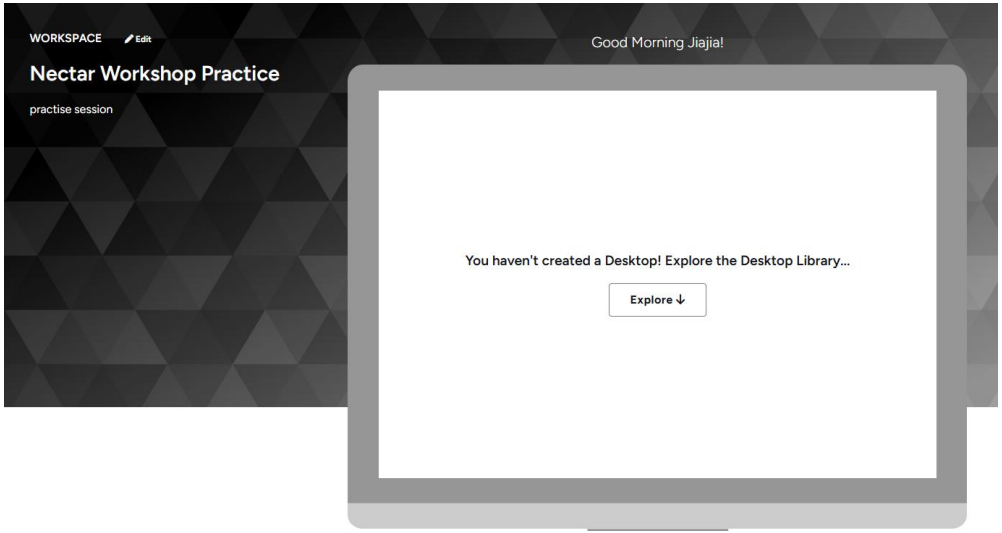
ARDC Nectar Virtual Desktop Service

You will be asked to create a new workspace when you first log in. It doesn't matter what information you put in; you could put yourself as the **Chief Investigator**.

You can use 460601 Cloud Computing for the **Field of Research Codes**.

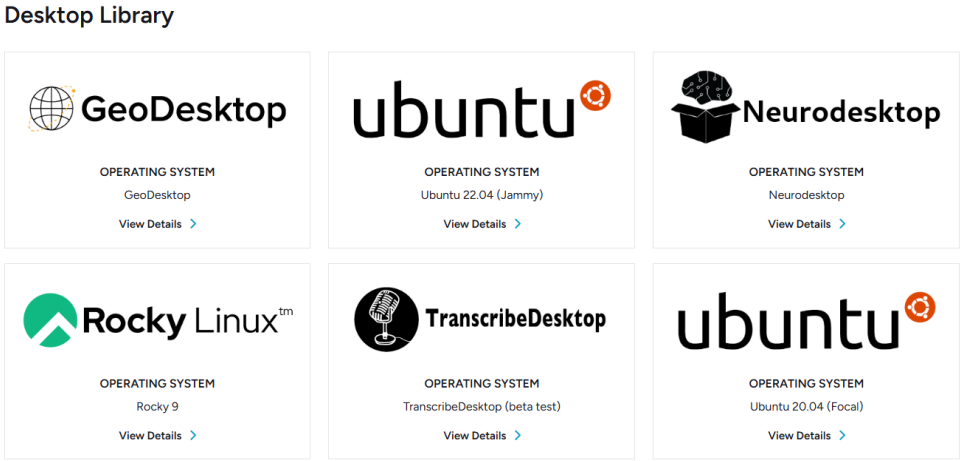


ARDC Nectar Virtual Desktop Service



Scroll down you can see a list of virtual desktops you can create.

ARDC Nectar Virtual Desktop Service



Pick ubuntu 22.04 (Jammy). Select default zone when getting prompted.

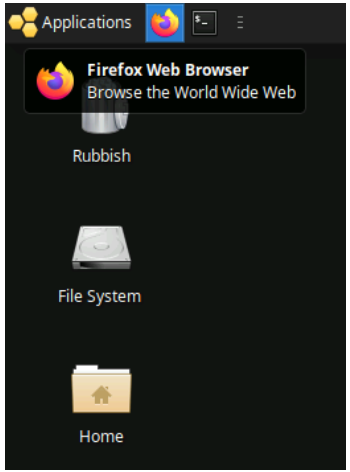
Wait for it to deploy then you can launch.

ARDC Nectar Virtual Desktop Service

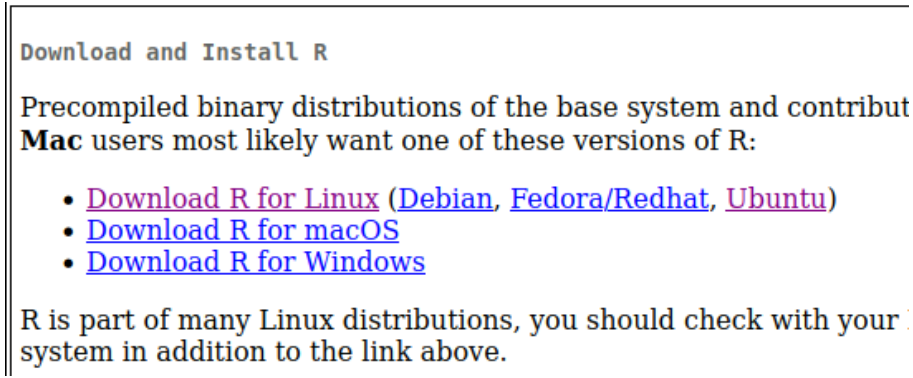
Your Ubuntu machine will look like this. On the top left, there are 2 icons, where you have the Firefox browser and the Terminal application.

This Ubuntu machine is new and doesn't have the software we need for our RNA-seq pipeline, we will still need to set up it.

Install R on ARDC virtual desktop



Open your Firefox browser and go to https://cloud.r-project.org/



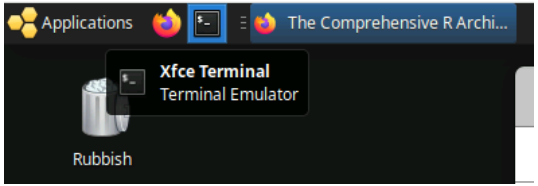
Click “Ubuntu”

Install R on ARDC virtual desktop

Installation Steps

> User | Root

```
# update indices
sudo apt update -qq
# install two helper packages we need
sudo apt install --no-install-recommends software-properties-common dirmngr
# add the signing key (by Michael Rutter) for these repos
# To verify key, run gpg --show-keys /etc/apt/trusted.gpg.d/cran_ubuntu_key.asc
# Fingerprint: E298A3A825C0D65DFD57CBB651716619E084DAB9
wget -qO- https://cloud.r-project.org/bin/linux/ubuntu/marutter_pubkey.asc | sudo t
# add the repo from CRAN -- lsb_release adjusts to 'noble' or 'jammy' or ... as nee
sudo add-apt-repository "deb https://cloud.r-project.org/bin/linux/ubuntu $(lsb_rel
# install R itself
sudo apt install --no-install-recommends r-base
```



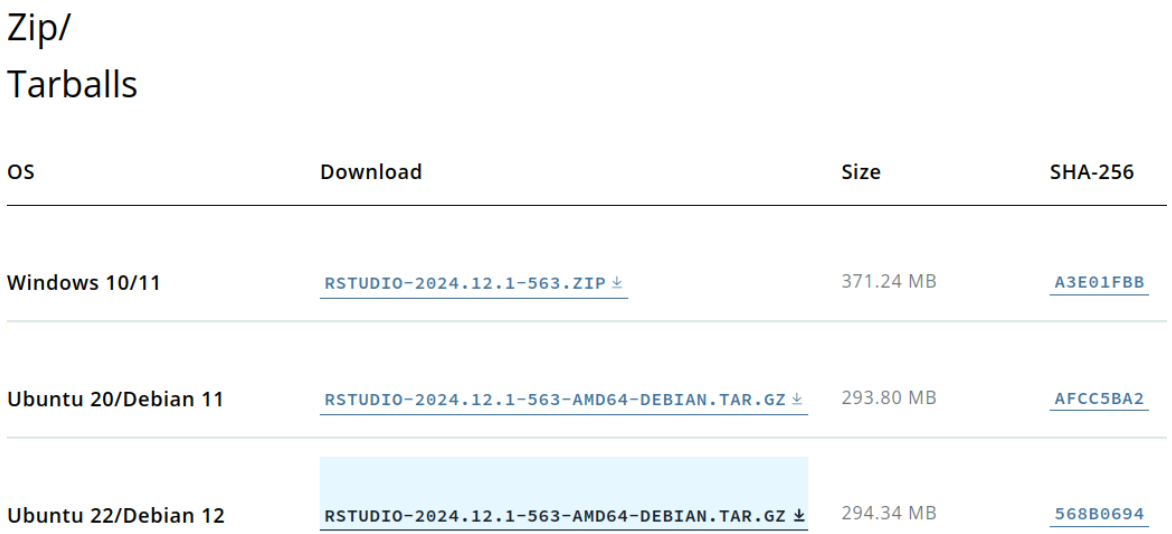
Open your “Xfce Terminal” and follow the installation steps to install R on your virtual desktop.

Follow the standard installation process. You should always press “Y” and “Enter” when the system asks you questions.

Install RStudio

Go to https://posit.co/download/rstudio-desktop/

Scroll down to find the “Zip/Tarballs”, choose the Ubuntu 22.tar.gz and click to download.



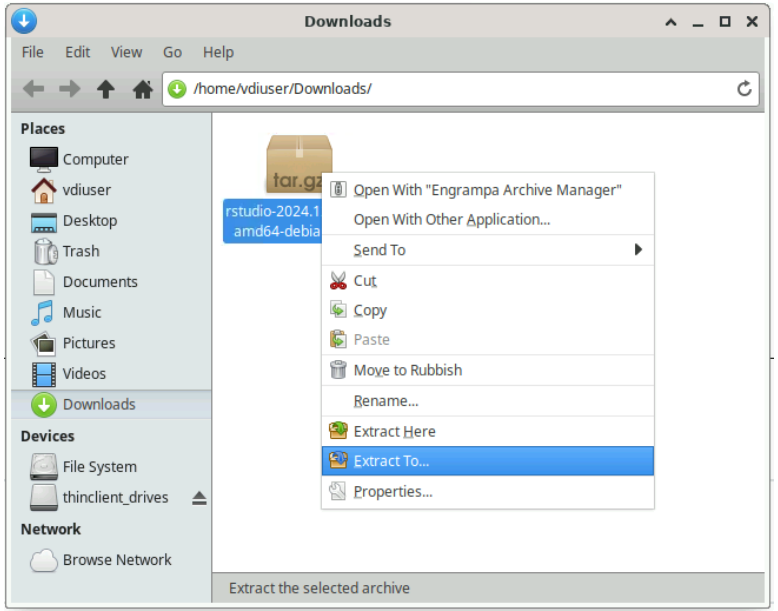
Zip/
Tarballs

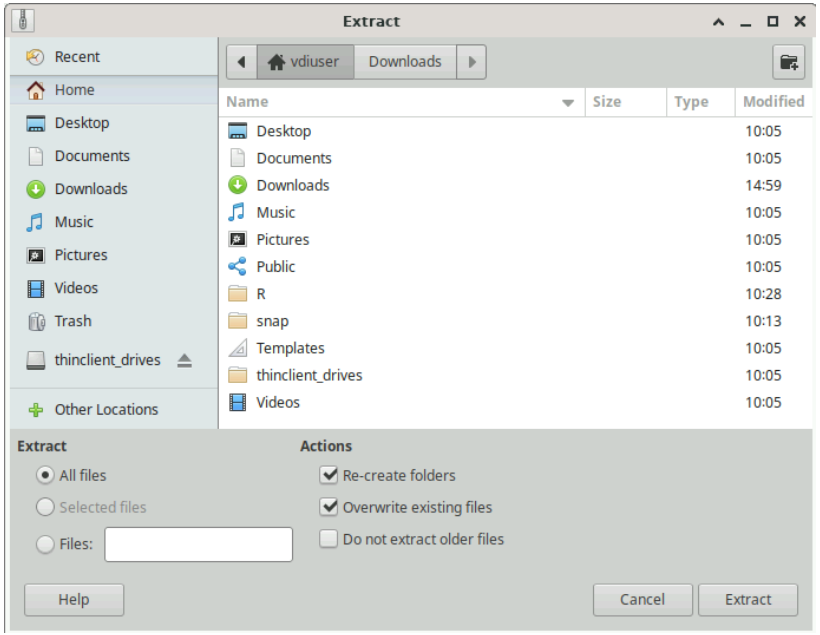
OS	Download	Size	SHA-256
Windows 10/11	RSTUDIO-2024.12.1-563.ZIP	371.24 MB	A3E01FBB
Ubuntu 20/Debian 11	RSTUDIO-2024.12.1-563-AMD64-DEBIAN.TAR.GZ	293.80 MB	AFCC5BA2
Ubuntu 22/Debian 12	RSTUDIO-2024.12.1-563-AMD64-DEBIAN.TAR.GZ	294.34 MB	568B0694

Install RStudio

Click open in folder when it finishes download. It will open the “Downloads” folder.

Right click choose “Extract to”, then select “Home” on left pane, then click “Extract”.

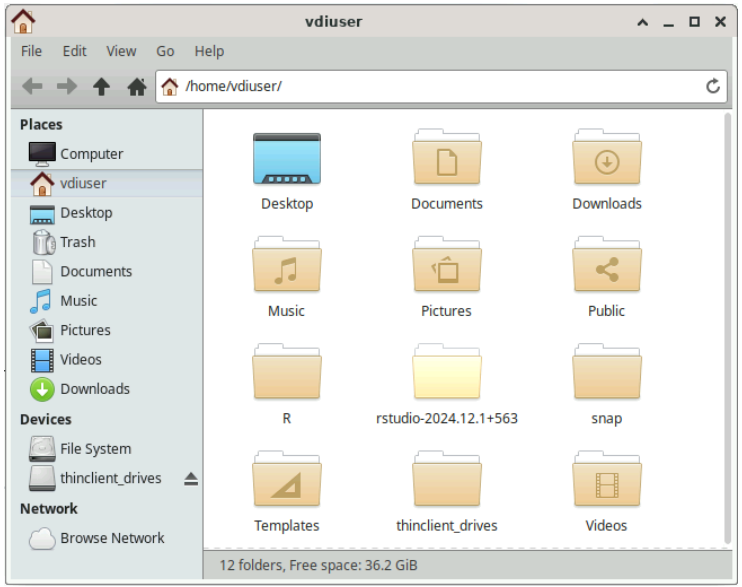


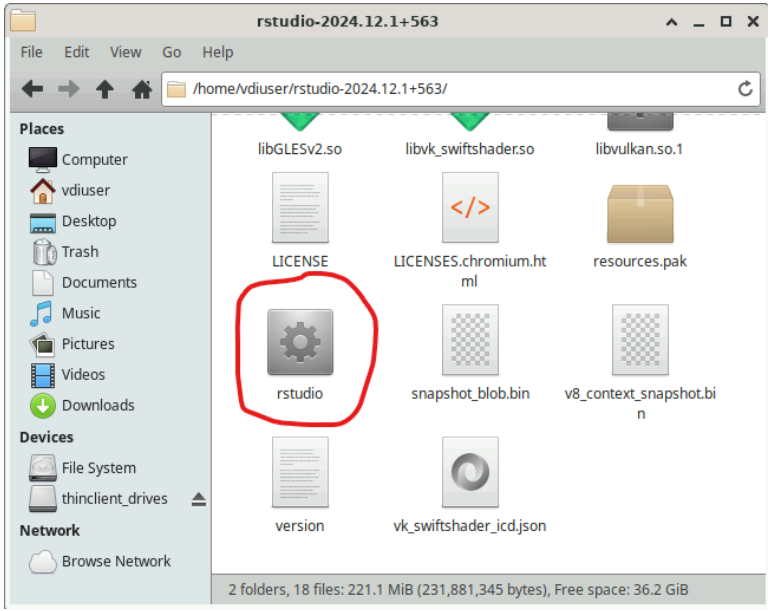


Install RStudio

Then you can close this “Downloads” folder, and on your Desktop, there is a Home folder, open it. You will see a folder “rstudio-2024.12.1+563”, open it.

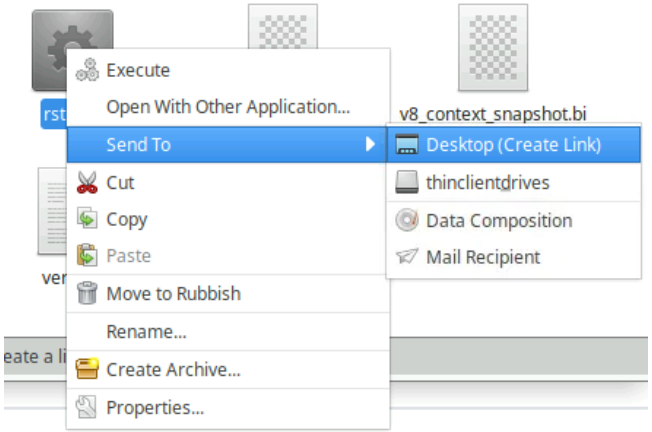
Scroll down to find “rstudio”, then double click to run. Your Rstudio is ready.





Install RStudio

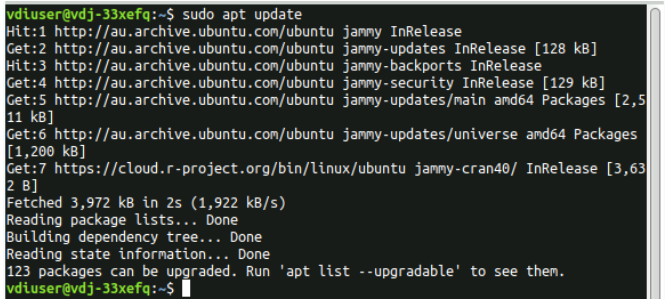
You can also create a link on Desktop for easy access.



Install Linux software

There are many need to install.

Open Terminal and run `sudo apt update` first.



Then run the following code, all the names after sudo apt install are the package names we are going to install. After it installed all, you will be asked to reboot your system, just press ENTER. Then you will be logged out. One issue about the ARDC virtual machine is that it takes a long time for rebooting. So you might want to take a break and come back. Or you can reboot manually by go back to the https://desktop.rc.nectar.org.au/home/ page, and click REBOOT then SOFT REBOOT for your machine.

```
`sudo apt install build-essential zlib1g-dev libssl-dev libfontconfig1-dev libfreetype6-dev \
                  libcurl4-openssl-dev libharfbuzz-dev libfribidi-dev libxml2-dev libfreetype6-dev \
                  libpng-dev libtiff5-dev libjpeg-dev gfortran libbz2-dev liblapack-dev libblas-dev \
                  openjdk-17-jdk`
```

Make sure you do not copy the `.

Install Bioconductor

Go to https://www.bioconductor.org/install/

```
if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")
BiocManager::install(version = "3.21")
```

Open Rstudio and copy/paste the code and run it in your Console.

Install libraries in R

In your Rstudio Console, run `install.packages("ggplot2")` to install ggplot2.

Run `library(ggplot2)` afterwards to make sure it's installed successfully.

Use the same method to install:

- dplyr
- data.table
- devtools

```
> install.packages("ggplot2")
Installing package into '/home/vdiuser/R/x86_64-pc-linux-gnu-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cloud.r-project.org/src/contrib/ggplot2_3.5.2.tar.gz'
Content type 'application/x-gzip' length 3580451 bytes (3.4 MB)
==================================================
downloaded 3.4 MB
```

```
> library(ggplot2)
>
```


Install packages through Bioconductor

There are some bioinformatics packages are only available via Bioconductor. To install a package using Bioconductor, we run:

```
> BiocManager::install("ballgown")
'getOption("repos")' replaces Bioconductor
package = "BiocManager")' for details.
Replacement repositories:
    CRAN: https://cloud.r-project.org
Bioconductor version 3.21 (BiocManager 1.30
```

`BiocManager::install("ballgown")`

Then you can use `library(ballgown)` to check if it's installed successfully.

```
> library(ballgown)

Attaching package: 'ballgown'

The following object is masked from 'package:base':

    structure
```

Using the same method, please install:

- DESeq2
- edgeR
- genefilter (though I think it's already included in ballgown, you could try loading it first to check)

Install HISAT2

In your Ubuntu virtual desktop, go to https://daehwankimlab.github.io/hisat2/download/

Click on Binaries – Version: HISAT2 2.2.1

- Binaries
  - Version: HISAT2 2.2.1
  - Version: HISAT2 2.2.0
  - Version: HISAT2 2.1.0

Then click on the Linux_x86_64 to download the binaries

Version: HISAT2 2.2.1

Release Date: 7/24/2020

Source	https://cloud.biohpc.swmed.edu/index.php/s/fE9QCsX3NH4QwBi/download
OSX_x86_64	https://cloud.biohpc.swmed.edu/index.php/s/zMgEtnF6LjnjFrr/download
Linux_x86_64	https://cloud.biohpc.swmed.edu/index.php/s/oTtGWbWjaxsQ2Ho/download

hisat2-2.2.1-Linux_x86_64.zip

Then follow the same strategy you did with Rstudio, extract the zip file to your /home directory, then you will be able to use it.

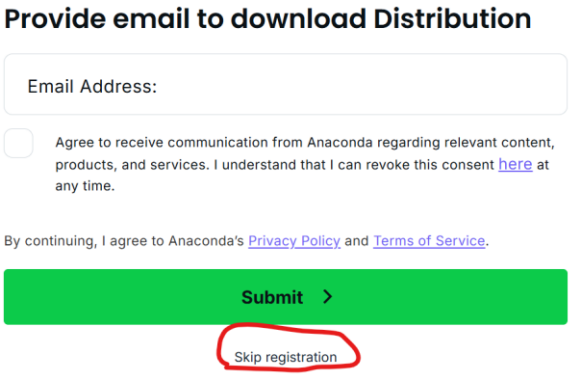
Install Conda

Conda is a software package manager where you can install many packages through it.

Except HISAT2 which we need to use a few python scripts from it and install the binaries is easy for us to locate the scripts. For other command-line tools we will use conda to install it.

In your Ubuntu virtual machine, go to https://www.anaconda.com/download

Then click “Skip registration”



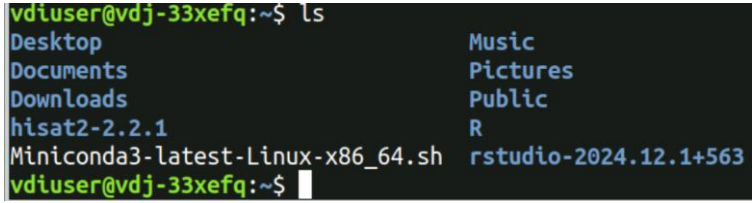
Install Conda

Scroll down to “Miniconda Installers”, click on the 64-Bix Installer for Linux. And right click copy the downloaded file to your home directory.

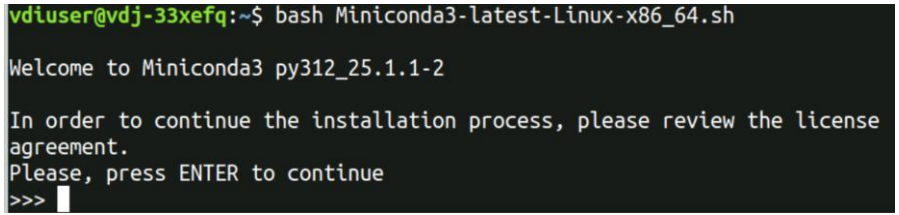
Miniconda3-latest-Linux-x86_64.sh

Install Conda

After you paste the .sh file to your home directory, open your Terminal. Run command `ls` and you can see the file in your terminal as well.



To install, we can run `bash Miniconda3-latest-Linux-x86_64.sh`. Press the space bar to read through the license agreement. Then say “YES” and press “ENTER” for the questions that system asks.



Install Conda

Then after it finishes installing. Close this terminal and reopen a new one.

Run `conda --version` to check if it’s installed successfully.

Install Conda

Create a conda environment, we name it “RNAseq_env”.

`conda create –n RNAseq_env`

```
(base) vdiuser@vdj-33xefq:~$ conda create -n RNAseq_env
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done
```

Then activate the conda channel: `conda activate RNAseq_env`

```
(base) vdiuser@vdj-33xefq:~$ conda activate RNAseq_env
(RNAseq_env) vdiuser@vdj-33xefq:~$
```

Then we need to add a few conda channels where we download and install our package from. Run the commands line by line.

`conda config --add channels defaults`
`conda config --add channels conda-forge`
`conda config --add channels bioconda`
`conda config --set channel_priority strict`

```
(base) vdiuser@vdj-33xefq:~$ conda activate RNAseq_env
(RNAseq_env) vdiuser@vdj-33xefq:~$ conda config --add channels defaults
(RNAseq_env) vdiuser@vdj-33xefq:~$ conda config --add channels conda-forge
(RNAseq_env) vdiuser@vdj-33xefq:~$ conda config --add channels bioconda
(RNAseq_env) vdiuser@vdj-33xefq:~$ conda config --set channel_priority strict
(RNAseq_env) vdiuser@vdj-33xefq:~$
```

Then run `conda config --show channels` to check the result.

```
(RNAseq_env) vdiuser@vdj-33xefq:~$ conda config --show channels
channels:
  - bioconda
  - conda-forge
  - defaults
  - https://repo.anaconda.com/pkgs/main
  - https://repo.anaconda.com/pkgs/r
```

Install RNAseq command line tools using Conda

Run below to install the needed software.

`conda install -c bioconda fastqc fastp multiqc samtools picard stringtie htseq`

And then you are ready for the workshop!!!!

Thank you

Contact us

Jiajia Li
ANU Biological Data Science Institute

RN Robertson Building
46 Sullivan's Creek Road
The Australian National University
Canberra ACT 2600

E jiajia.li1@anu.edu.au

Australian National University

TEQSA PROVIDER ID: PRV12002 (AUSTRALIAN UNIVERSITY)
CRICOS PROVIDER CODE: 00120C