@Author: James Talwar

Installations:

Summary: This notebook will walk you through the process of installing the necessary packages/programs/software you will need for the MSTP bioinformatics bootcamp

NB: Please be sure to follow these instructions thoroughly and in order. Installations can be finicky and package incompatibility issues can cause a bunch of headaches to fix.

1. Log into TSCC:

1.1 Getting onto the Login Node

Follow the instructions as outlined in the /Day_0_Setup/Generate_Keys folder on the course website for logging in. Make sure you have ran through everything in the Generate_Keys notebook before proceeding with installations.

Mac users in your terminal window enter:

ssh ucsd-train##@tscc-login.sdsc.edu <-- replace ## with the account number James sent you

For **Windows** users open PuTTY, load your saved "tscc" settings along with the appropriate private key and log on.

After logging in your terminal should show something along the lines of [ucsd-train##@tscc-login12 ~]$ where ## is your account number. Here you are on the TSCC login node which essentially is your access point to the TSCC cluster (There are multiple login nodes so your number may not be 12).

To actually run things properly (and not evoke SDSCs ire) you will need to start an interactive session.

1.2 Starting an Interactive Session

TSCC (a helpful guide to navigating TSCC can be found here: https://www.sdsc.edu/support/user_guides/tscc.html) operates through SGE. So we want to start an interactive session - great how do we do that? Well for those of you familiar with SGE you likely would assume a simple qlogin would be good enough to get us up and running, but sadly for TSCC that is not the case...

TSCC has certain compute partitions that you will access when submitting jobs and starting interactive session. The main one you will be interacting with is **hotel**, however to do so you will need TSCC credits (or SUs). Fortunately we all have plenty of credits (~770 per user). These credits equate to one CPU hour.

Okay so now that the brief logistical detour is complete let's figure out how to get an interactive job submitted so we can install what we need.

To start an interactive session for INSTALLATION copy the following command to your terminal:

qsub -I -l walltime=5:30:00 -l nodes=2:ppn=2 -q hotel

Here we are setting the walltime - the amount of time TSCC lets the job run before kicking us off - to 5.5 hours. Installation should take less than this so don't panic), but in case things get bogged down on TSCC, it's better to request more than less resources. We can always exit out early when done.

If successful you should get something like this in your terminal:

```
[ucsd-train58@tscc-login12 ~]$ qsub -I -l walltime=5:30:00 -l nodes=2:ppn=2 -q hotel
qsub: waiting for job 26321109.tscc-mgr7.local to start
qsub: job 26321109.tscc-mgr7.local ready

[ucsd-train58@tscc-4-52 ~]$
```

2. Installing Miniconda:

To start we need to install Miniconda - a lightweight version of anaconda that will serve as our python/R package manager.

2.1 Copy the installation script to your home directory

The miniconda installation script needed can be found in the shared directory for the course(/oasis/tscc/scratch/mstp2020) we are going to be using for bootcamp. To start move the script from the shared directory to your home directory by copying and pasting the following command:

scp /oasis/tscc/scratch/mstp2020/miniconda/Miniconda3-py37_4.8.3-Linux-x86_64.sh .

```
[ucsd-train58@tscc-4-52 ~]$ scp /oasis/tscc/scratch/mstp2020/miniconda/Miniconda3-py37_4.8.3-Linux-x86_64.sh .
[ucsd-train58@tscc-4-52 ~]$ ls
Miniconda3-py37_4.8.3-Linux-x86_64.sh
[ucsd-train58@tscc-4-52 ~]$
```

- Yes I know the shared folder has it as 2020 and not 2021, don't bite my head off - that was TSCC's doing.

2.2 Install Miniconda

To install Miniconda enter the following command from your home directory:

bash Miniconda3-py37_4.8.3-Linux-x86_64.sh

Press enter when prompted

```
[ucsd-train58@tscc-4-52 ~]$ bash Miniconda3-py37_4.8.3-Linux-x86_64.sh

Welcome to Miniconda3 py37_4.8.3

In order to continue the installation process, please review the license
agreement.
Please, press ENTER to continue
>>>
```

You will then be presented with the license/terms and conditions. If you want to skip to the end hit q and then accept the license terms by typing *yes*.

Miniconda will then present you with an installation location. It should be /home/ucsd-train##/miniconda3. Press enter to confirm the location

```
Do you accept the license terms? [yes|no]
[no] >>> yes

Miniconda3 will now be installed into this location:
/home/ucsd-train58/miniconda3

  - Press ENTER to confirm the location
  - Press CTRL-C to abort the installation
  - Or specify a different location below

[/home/ucsd-train58/miniconda3] >>>
```

Miniconda is now installing! This may take a bit so don't get frustrated. Leave your terminal open and let this run.

2.3 Setting the path

In order for things to work seamlessly TSCC needs to know where to look for programs and packages. To do that it looks at your .bashrc. After miniconda installation we move to update our .bashrc to point to our miniconda installation. To do this respond with yes after miniconda finishes installing and gives the prompt "Do you wish the installer to initialize Miniconda3 by running conda init?"

If you don't hit yes though don't worry you can tell conda to add it to your .bashrc after the fact. To do this copy and run the following into your terminal:

~/miniconda3/bin/conda init

After this you need source your .bashrc, which at a high-level tells the cluster to reload your .bashrc with the updated changes. To do this copy the following into your terminal:

source .bashrc

2.4 Sanity Check:

Okay moment of truth - time to check if this worked. Type **which python** into your terminal and you should see it point to the python in your miniconda

```
[ucsd-train58@tscc-4-52 ~]$ source .bashrc
(base) [ucsd-train58@tscc-4-52 ~]$ which python
~/miniconda3/bin/python
```

3. Installing Necessary Python Packages:

Since miniconda is lightweight, we need to add the packages we will be using ourselves.

3.1 Install Numpy:

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c anaconda numpy

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c anaconda numpy
Collecting package metadata (current_repodata.json): done
Solving environment: done

==> WARNING: A newer version of conda exists. <==
  current version: 4.8.3
  latest version: 4.10.1

Please update conda by running

    $ conda update -n base -c defaults conda


## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - numpy


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    blas-1.0                   |              mkl           6 KB  anaconda
    ca-certificates-2020.10.14 |                0         128 KB  anaconda
    certifi-2020.6.20          |           py37_0         159 KB  anaconda
    conda-4.9.0                |           py37_0         3.1 MB  anaconda
    intel-openmp-2020.2        |              254         947 KB  anaconda
    mkl-2019.4                 |              243       204.1 MB  anaconda
    mkl-service-2.3.0          |   py37he904b0f_0         208 KB  anaconda
    mkl_fft-1.2.0              |   py37h23d657b_0         164 KB  anaconda
    mkl_random-1.0.4           |   py37hd81dba3_0         390 KB
    numpy-1.19.1               |   py37hbc911f0_0          20 KB  anaconda
    numpy-base-1.19.1          |   py37hfa32c7d_0         5.2 MB  anaconda
    openssl-1.1.1h             |       h7b6447c_0         3.8 MB  anaconda
    ------------------------------------------------------------
                                           Total:       218.1 MB

The following NEW packages will be INSTALLED:

  blas               anaconda/linux-64::blas-1.0-mkl
  intel-openmp       anaconda/linux-64::intel-openmp-2020.2-254
  mkl                anaconda/linux-64::mkl-2019.4-243
  mkl-service        anaconda/linux-64::mkl-service-2.3.0-py37he904b0f_0
  mkl_fft            anaconda/linux-64::mkl_fft-1.2.0-py37h23d657b_0
  mkl_random         pkgs/r/linux-64::mkl_random-1.0.4-py37hd81dba3_0
  numpy              anaconda/linux-64::numpy-1.19.1-py37hbc911f0_0
  numpy-base         anaconda/linux-64::numpy-base-1.19.1-py37hfa32c7d_0

The following packages will be UPDATED:

  ca-certificates    pkgs/main::ca-certificates-2020.1.1-0 --> anaconda::ca-certificates-2020.10.14-0
  certifi            pkgs/main::certifi-2020.4.5.1-py37_0 --> anaconda::certifi-2020.6.20-py37_0
  conda              pkgs/main::conda-4.8.3-py37_0 --> anaconda::conda-4.9.0-py37_0
  openssl            pkgs/main::openssl-1.1.1g-h7b6447c_0 --> anaconda::openssl-1.1.1h-h7b6447c_0


Proceed ([y]/n)? y
```

3.2 Install Pandas:

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c anaconda pandas

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c anaconda pandas
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - pandas


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    pandas-1.1.3               |   py37he6710b0_0        10.5 MB  anaconda
    python-dateutil-2.8.1      |             py_0         224 KB  anaconda
    pytz-2020.1                |             py_0         239 KB  anaconda
    ------------------------------------------------------------
                                           Total:        11.0 MB

The following NEW packages will be INSTALLED:

  pandas             anaconda/linux-64::pandas-1.1.3-py37he6710b0_0
  python-dateutil    anaconda/noarch::python-dateutil-2.8.1-py_0
  pytz               anaconda/noarch::pytz-2020.1-py_0


Proceed ([y]/n)? y
```

3.3 Install Jupyter

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c anaconda jupyter

There will be a lot of packages installed here and this should take longer than the previous two installations (numpy and pandas).

3.4 Install Matplotlib

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c conda-forge matplotlib

```
(base) [ucsd-train58@tscc-login2 ~]$ conda install -c conda-forge matplotlib

Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - matplotlib


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    cycler-0.10.0              |             py_2           9 KB  conda-forge
    kiwisolver-1.3.1           |   py37hc928c03_0          86 KB  conda-forge
    lcms2-2.12                 |       h3be6417_0         312 KB
    matplotlib-3.3.4           |   py37h06a4308_0          26 KB
    matplotlib-base-3.3.4      |   py37h62a2d02_0         5.1 MB
    olefile-0.46               |     pyh9f0ad1d_1          32 KB  conda-forge
    pillow-8.2.0               |   py37he98fc37_0         622 KB
    ------------------------------------------------------------
                                           Total:         6.1 MB

The following NEW packages will be INSTALLED:

  cycler             conda-forge/noarch::cycler-0.10.0-py_2
  kiwisolver         conda-forge/linux-64::kiwisolver-1.3.1-py37hc928c03_0
  lcms2              pkgs/main/linux-64::lcms2-2.12-h3be6417_0
  matplotlib         pkgs/main/linux-64::matplotlib-3.3.4-py37h06a4308_0
  matplotlib-base    pkgs/main/linux-64::matplotlib-base-3.3.4-py37h62a2d02_0
  olefile            conda-forge/noarch::olefile-0.46-pyh9f0ad1d_1
  pillow             pkgs/main/linux-64::pillow-8.2.0-py37he98fc37_0

The following packages will be UPDATED:

  ca-certificates    pkgs/main::ca-certificates-2021.5.25-~ --> conda-forge::ca-certificates-2021.5.30-ha878542_0

The following packages will be SUPERSEDED by a higher-priority channel:

  certifi            pkgs/main::certifi-2021.5.30-py37h06a~ --> conda-forge::certifi-2021.5.30-py37h89c1867_0


Proceed ([y]/n)? y
```

3.5 Install Seaborn

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install seaborn

```
(base) [ucsd-train58@tscc-login1 ~]$ conda install seaborn
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - seaborn


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    conda-4.10.3               |   py37h06a4308_0         2.9 MB
    scipy-1.6.2                |   py37h91f5cce_0        15.4 MB
    seaborn-0.11.1             |     pyhd3eb1b0_0         212 KB
    ------------------------------------------------------------
                                           Total:        18.5 MB

The following NEW packages will be INSTALLED:

  scipy              pkgs/main/linux-64::scipy-1.6.2-py37h91f5cce_0
  seaborn            pkgs/main/noarch::seaborn-0.11.1-pyhd3eb1b0_0

The following packages will be SUPERSEDED by a higher-priority channel:

  ca-certificates    conda-forge::ca-certificates-2021.5.~ --> pkgs/main::ca-certificates-2021.5.25-h06a4308_1
  certifi            conda-forge::certifi-2021.5.30-py37h8~ --> pkgs/main::certifi-2021.5.30-py37h06a4308_0
  conda              conda-forge::conda-4.10.3-py37h89c186~ --> pkgs/main::conda-4.10.3-py37h06a4308_0


Proceed ([y]/n)? y
```

4. External Genomic/RNA-Seq Program Installations:

We also need to install the following programs for bootcamp

4.1 Install STAR

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c bioconda STAR

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c bioconda STAR
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - star


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    ca-certificates-2021.5.25  |       h06a4308_1         112 KB
    certifi-2021.5.30          |   py37h06a4308_0         139 KB
    conda-4.10.1               |   py37h06a4308_1         2.9 MB
    openssl-1.1.1k             |       h27cfd23_0         2.5 MB
    star-2.7.9a                |       h9ee0642_0         3.8 MB  bioconda
    ------------------------------------------------------------
                                           Total:         9.5 MB

The following NEW packages will be INSTALLED:

  star               bioconda/linux-64::star-2.7.9a-h9ee0642_0

The following packages will be UPDATED:

  ca-certificates    anaconda::ca-certificates-2020.10.14-0 --> pkgs/main::ca-certificates-2021.5.25-h06a4308_1
  certifi            anaconda::certifi-2020.6.20-py37_0 --> pkgs/main::certifi-2021.5.30-py37h06a4308_0
  conda              anaconda::conda-4.9.0-py37_0 --> pkgs/main::conda-4.10.1-py37h06a4308_1
  openssl            anaconda::openssl-1.1.1h-h7b6447c_0 --> pkgs/main::openssl-1.1.1k-h27cfd23_0


Proceed ([y]/n)? y
```

4.2 Install fastqc

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c bioconda fastqc

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c bioconda fastqc
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - fastqc


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    fastqc-0.11.9              |       hdfd78af_1         9.7 MB  bioconda
    font-ttf-dejavu-sans-mono-2.37|       h6964260_0         335 KB
    openjdk-8.0.152            |       h7b6447c_3        57.4 MB
    perl-5.26.2                |       h14c3975_0        10.5 MB
    ------------------------------------------------------------
                                           Total:        77.9 MB

The following NEW packages will be INSTALLED:

  fastqc             bioconda/noarch::fastqc-0.11.9-hdfd78af_1
  font-ttf-dejavu-s~ pkgs/main/noarch::font-ttf-dejavu-sans-mono-2.37-h6964260_0
  openjdk            pkgs/main/linux-64::openjdk-8.0.152-h7b6447c_3
  perl               pkgs/main/linux-64::perl-5.26.2-h14c3975_0


Proceed ([y]/n)? y
```

4.3 Install samtools and bzip2

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c conda-forge -c bioconda samtools bzip2

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c conda-forge -c bioconda samtools bzip2
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - bzip2
    - samtools


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    bzip2-1.0.8                |       h516909a_3         398 KB  conda-forge
    ca-certificates-2021.5.30  |       ha878542_0         136 KB  conda-forge
    certifi-2021.5.30          |   py37h89c1867_0         141 KB  conda-forge
    conda-4.10.3               |   py37h89c1867_0         3.1 MB  conda-forge
    python_abi-3.7             |          2_cp37m           4 KB  conda-forge
    samtools-1.3.1             |                0         1.5 MB  bioconda
    ------------------------------------------------------------
                                           Total:         5.2 MB

The following NEW packages will be INSTALLED:

  bzip2              conda-forge/linux-64::bzip2-1.0.8-h516909a_3
  python_abi         conda-forge/linux-64::python_abi-3.7-2_cp37m
  samtools           bioconda/linux-64::samtools-1.3.1-0

The following packages will be UPDATED:

  ca-certificates    pkgs/main::ca-certificates-2021.5.25-~ --> conda-forge::ca-certificates-2021.5.30-ha878542_0
  conda              pkgs/main::conda-4.10.1-py37h06a4308_1 --> conda-forge::conda-4.10.3-py37h89c1867_0

The following packages will be SUPERSEDED by a higher-priority channel:

  certifi            pkgs/main::certifi-2021.5.30-py37h06a~ --> conda-forge::certifi-2021.5.30-py37h89c1867_0


Proceed ([y]/n)? y
```

4.4 Install subread

Copy the following into your terminal, hit enter, and then type *y* when prompted:

conda install -c bioconda subread

```
(base) [ucsd-train58@tscc-4-52 ~]$ conda install -c bioconda subread
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/ucsd-train58/miniconda3

  added / updated specs:
    - subread


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    subread-2.0.1              |       hed695b0_0        20.7 MB  bioconda
    ------------------------------------------------------------
                                           Total:        20.7 MB

The following NEW packages will be INSTALLED:

  subread            bioconda/linux-64::subread-2.0.1-hed695b0_0

The following packages will be SUPERSEDED by a higher-priority channel:

  certifi            conda-forge::certifi-2021.5.30-py37h8~ --> pkgs/main::certifi-2021.5.30-py37h06a4308_0


Proceed ([y]/n)? y
```

4.5 Sanity Checks:

To quickly sanity check everything installed properly run the *which* command for the following 4 programs:

```
(base) [ucsd-train59@tscc-login1 ~]$ which STAR
~/miniconda3/bin/STAR
(base) [ucsd-train59@tscc-login1 ~]$ which fastqc
~/miniconda3/bin/fastqc
(base) [ucsd-train59@tscc-login1 ~]$ which samtools
~/miniconda3/bin/samtools
(base) [ucsd-train59@tscc-login1 ~]$ which featureCounts
~/miniconda3/bin/featureCounts
```

If all 4 point to the same locations as above (i.e., they should all point to your home directory's miniconda/bin/PROGRAM) you should be good to go and move onto R installation

5. Install R

Now we need to install R. To install R and the necessary packages for everything to work properly copy the following into your terminal, hit enter, and then type y when prompted:

conda install r r-essentials

There are quite a few things that need to install here so this likely will take the longest of all the installations you will perform. **Let this complete fully - don't kill this if it seems stuck. Let this run until you see your usual** (base) [ucsd-train58@tscc-#-## ~]

After this finishes perform a quick sanity check with which R which should point you to your ~/miniconda3/bin/R

6. Install DESeq2

Okay homestretch - we now need to install DESeq2. To do this copy the following into your terminal, hit enter, and then type y when prompted:

conda install -c bioconda bioconductor-deseq2

Again there are quite a few things that need to install here so this may take longer than usual. **Again let this complete fully!**

Before wrapping up we need to perform a quick sanity check. Specifically we need to make sure we can load the DESeq2 library in R. To do this type and run R and then when R loads type and run library(DESeq2). Several loading messages should show up.

```
(base) [ucsd-train59@tscc-4-40 ~]$ R

R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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.

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.

> library(DESeq2)
Loading required package: S4Vectors
Loading required package: stats4
Loading required package: BiocGenerics
Loading required package: parallel

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:parallel':

    clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
    clusterExport, clusterMap, parApply, parCapply, parLapply,
    parLapplyLB, parRapply, parSapply, parSapplyLB

The following objects are masked from 'package:stats':

    IQR, mad, xtabs

The following objects are masked from 'package:base':

    Filter, Find, Map, Position, Reduce, anyDuplicated, append,
    as.data.frame, as.vector, cbind, colnames, do.call, duplicated,
    eval, evalq, get, grep, grepl, intersect, is.unsorted, lapply,
    lengths, mapply, match, mget, order, paste, pmax, pmax.int, pmin,
    pmin.int, rank, rbind, rownames, sapply, setdiff, sort, table,
    tapply, union, unique, unlist, unsplit

Loading required package: IRanges
Loading required package: GenomicRanges
```

If they all pass without any error messages you should be all set! Escape R with q() then type n (no need to save the workspace image) and then type exit to end your interactive session. Congratulations on a job well done.