

TUTORIAL 1 - Getting Started

Unix and “flavors” of Unix (i.e. Linux) are powerful multi-user operating systems. The flexibility and power of Unix has resulted in genetic epidemiologists, statistical geneticists and statisticians to write the majority of genetic analysis packages to run preferentially on Unix platforms. At first, Unix can appear cryptic and difficult to use. However, there are a few easy-to-learn features of Unix that will allow us to run a host of genetic analysis software packages. For this workshop, we will be using the Ubuntu Linux distribution through a VM server. The advantage of this approach is that you the VM can be made available after the workshop as well.

Below, you will see,

text written like this

That refers to what you should type at the command line prompt, followed by the <ENTER> key.

Follow these steps to get started with exploring the data and the file system.

1) Log into the desktop using provided credentials.

2) Log into the server by opening up a *SSH Secure Shell Client* such as PUTTY. The following information will be necessary to access the server:

- Host Name: courses.cbio.uct.ac.za
- Connection Type: SSH
- username : student#
- password : TBD

You should now have a terminal window with a similar command line prompt:

```
student@courses:~$
```

GWAS, Quality Control and Family based analysis workshop February 2016 – Matthew McQueen

TUTORIAL 1 - Getting Started

3) To see what is in your home directory, use the list [`ls`] command at the prompt:

```
student@courses:~$ ls
```

4) Let's move into data directory by using the change directory [`cd`] command:

```
student@courses:~$ cd /student_data/cbio2016
```

5) Issue another [`ls`] command to see the directories listed here.

6) To view files in a Unix/Linux environment, there are a few options. Some of the more popular editors are *vi* and *emacs*, but these can be difficult to learn and even harder to master. If you want to see what is in a particular file, you can use the [`more`] or [`less`] commands. Let's have a look at a file used for a PLINK analysis.

```
student@courses:~$ cd clean
student@courses:~$ more bmi.fam
```

You can scroll through the contents of the file by hitting the *Space bar* and exit by typing, [`q`]

Alternatively, we can use [`less`]

```
student@courses:~$ less bmi.fam
```

7) To make edits to a file (not recommended!) you need an editor. Sometimes it is easier to view the contents of a file using an editor. If you are not familiar with *vi* or *emacs*, an easy-to-use editor is *pico*.

```
student@courses:~$ pico bmi.bim
```

Note the commands at the bottom of the page, including the **^X** (control-X) that will exit you out of the editor.

8) Now if you change directory again, this will move you back to your home directory

```
student@courses:~$ cd
```

9) Finally, let's open RStudio to become familiar with that interface to R. For this course, we will be using RStudio through the server. Point your browser to the following URL:

<http://courses.cbio.uct.ac.za:8787>

Moving around in the file system

<i>Command</i>	<i>Action</i>
pwd	"Print working directory" - show what directory you're in.
ls	List the contents of a directory
ls -l	List the contents of a directory and show additional info of the files.
ls -a	List all files, including hidden files.
cd	Change directory.
cd ..	Go to the parent directory.
cd ../	Go back to the previous directory.

Examining files

<i>Command</i>	<i>Action</i>
file	Determine the type of a file.
cat	Concatenate a file.
less	View text files and paginate them if needed.

Manipulating files and directories

<i>Command</i>	<i>Action</i>
cp	Copy a file.
cp -i	Copy a file and ask before overwriting.
cp -r	Copy a directory with its contents.
mv	Move or rename a file.
mv -i	Move or rename a file and ask before overwriting.
rm	Remove a file.
rm -r	Remove a directory with its contents.
rm -i	Ask before removing a file. Good to use with the -r option.
mkdir	Make a directory.
rmdir	Remove an empty directory.

Useful UNIX/LINUX Commands