

genomics workshop runsheet

📅 Created	
🕒 Last Edited Time	@August 24, 2021 3:15 PM
☰ Property	
⋮ Type	Notes

Getting started

1. Rstudio working environment (console vs. terminal)

The R environment exists within the UNIX machine. The console is the user interface of R, and terminal is the user interface of the UNIX operating system.

If you're not familiar with this concept, it is in a way very similar to how you use your personal computer. For example if you want to search for a document that exists somewhere on your computer you will do an operating system based command to search for that document. After you locate that document you will need to edit or view the document using the suitable software.

R is one of the many software that exist in the UNIX machine, and you can use the console to run functions that are designed to work with R. Rstudio is an IDE (integrated development environment) designed primarily to work with R code and packages, and it also contains the terminal for convenience.

Not every software we will need today runs within R. Some of them are independent softwares outside of R, like the text editor is independent from the web browser, that run on themselves in the command line interface. That's what the terminal is for.

2. Data directory

Explain the data types. Store data in separate folders by tasks.

```
~/home/rstudio/colo829`
```

3. Working with UNIX commands (live demo)

Introduce simple cmd, such as

- ls

- cd
- man (q for quit man page)
- Running cmd tools
- Piping

Task 1

explain the usage of paired tumor/normal samples and how it helps with pathogenicity analysis

- Locate the reads `.fastq` files & the reference `.fa` file in demo (e.g. `/data/task1/colo829TM.r1.fastq`)
- Explain what `bwa` and `samtools` do for alignment, sorting and indexing. Use `man` and home pages as resources.
- Explain the file formats: fasta, fastq, SAM, BAM (why BAM needs to be sorted and indexed)
- Give one example of running the commands in terminal, e.g.,
`bwa mem hg19.fa colo829TM.r1.fq colo829TM.r2.fq > colo829TM.sam`
- Let the students figure out the rest

Task 2

Task 3

Summary

Live demo using IGV (web version) to look at the interesting regions called and analysed in the tasks.

Re-address the take home messages to the students.

