Module – practical bioinformatics

Often in bioinformatics, the problem is not to run a tool, but to find data or to transform then into the correct format. In this practical we are going to practice that a little bit.

This will be followed on how to submit jobs on the HPC and finally we can have a little catchup discussion round.

Exercise 1: From Excel to R

- Download the file
ftp://ftp.sanger.ac.uk/pub/project/pathogens/tdo/Exercise/readcounts.xlsx
- load it into R
- get the correlation of the first and second data column

Download data from EBI

- Load the chromosome one of *Plasmodium falciparum* into artemis or igv

- Load the read counts from ArrayExpress/ Gene Expression Omnibus accession GSE75795 into R

- How long are the reads from the study ERX001048 on the European Nucleotide Archive?

Merge three lists

- Download the files MergingExercise.tar.gz from ftp://ftp.sanger.ac.uk/pub/project/pathogens/tdo/Exercise/

- untar the file (tar xvzf filename)

- Merge somehow the three lists… how could you do that?

- What about the perl script ~to16r/Bin/list.join2List.pl ?

- What is the problem with those lists? Do you get all the information maintained?

How to submit jobs to the HPC?

- Go and visit one of Mark Meenan's courses!!!!

- https://www.gla.ac.uk/myglasgow/it/hpcc/usergui de/howtousetorquepbs/

- Submit a job: qsub
- Check on running: qstat –u username
- Other commands: qdel

- What does this script?
  ~to16r/bioinfo/scriptMap.sh (use cat or less)

- Submit the jobs qsub –q short
  ~to16r/bioinfo/scriptMap.sh
- Check if it is running qstat
- What files are generated once done?
- Submit this job and look what happens qsub
  ~to16r/bioinfo/notworking.sh

Time from discussion

Do you have other questions?
Do you want to spend more time on specific exercises?