

Analysis of RNA-Seq Data: Preparation Questions

The following questions aim to introduce you into the cloud environment, bash, R and the pipelines that we are going to use during this course. Complete the following two tasks. The first task (registration to the cloud service) is due **Sunday, March 8th**. Second, answer the preparation questions and **bring them** to either Abdul or Michael (A2.4. Ground Floor, R.04 or R.20) **before 18th of March** or **send them** by e-mail to **Abdul** or **Michael**. These tasks are mandatory and the failure to comply with them will exclude you from the workshop.

1 Task1

1.1 Registration

We will use cloud resources from the German Network for Bioinformatics Infrastructure [de.NBI](https://cloud.denbi.de/) for this course. In order to use these, you must create an [Elixir](https://elixir-europe.org/) account and link it to de.NBI. Mind that you have to **create the account linked to your university account**. Saarland University needs to be specified as the Identity Provider during the registration process.

Instructions for how to do this can be found here:

<https://cloud.denbi.de/wiki/registration/>

1.2 Generate a SSH keypair

In order to connect to the cloud, you will need a valid SSH keypair that authorizes your access to the services. Please create a SSH keypair. One way of doing this is described here: <https://help.github.com/en/articles/generating-a-new-ssh-key-and-adding-it-to-the-ssh-agent#generating-a-new-ssh-key>.

You have to add the SSH keypair (public key) in de.NBI portal (<https://cloud.denbi.de/portal>, Figure 1). Last, send your ELIXIR Login and your SSH public key to **Abdul**. We will then add your key to the respective cloud instance and invite you to our project (bioquantSB, accept the invitation sent to you via e-mail) such that you can connect (see next step).

1.3 Connect to de.NBI cloud

This step is only possible after you

- sent your Elixir login and public key to Abdul
- added your public key to de.NBI portal
- accepted the invitation
- Abdul added the key to the cloud instance (your account may take few hours to get activated from the time you send the information to Abdul. Contact him if you couldn't access after two days)

A normal SSH connection will allow you to command the machine you have connected to. To make the rest of the course easier, we must start by complicating things. We want graphics, which in Linux-speech is called X11-forwarding and we want to easily transfer files. The problem with the latter is that we need to connect through a middle machine - a jumhost. For this, we will setup a SSH tunnel. Depending on the operation system of your laptop follow these instructions:

- **Linux & Mac:** Here, we can use the built in ssh client and configure it to our liking.
 - **Graphics:** This is easy, simply add the `-X` flag
 - **Tunnel:** The easiest way to do this is to follow the instructions here: [Setting up a SOCKS proxy](#)
Open the file `~/.ssh/config` with your favorite text editor. Paste the excerpt from the link. Mind that you have to edit the instances of `ELIXIR_USERNAME` and `YOUR-KEY-FILE` to your settings. The latter should point to your private key.

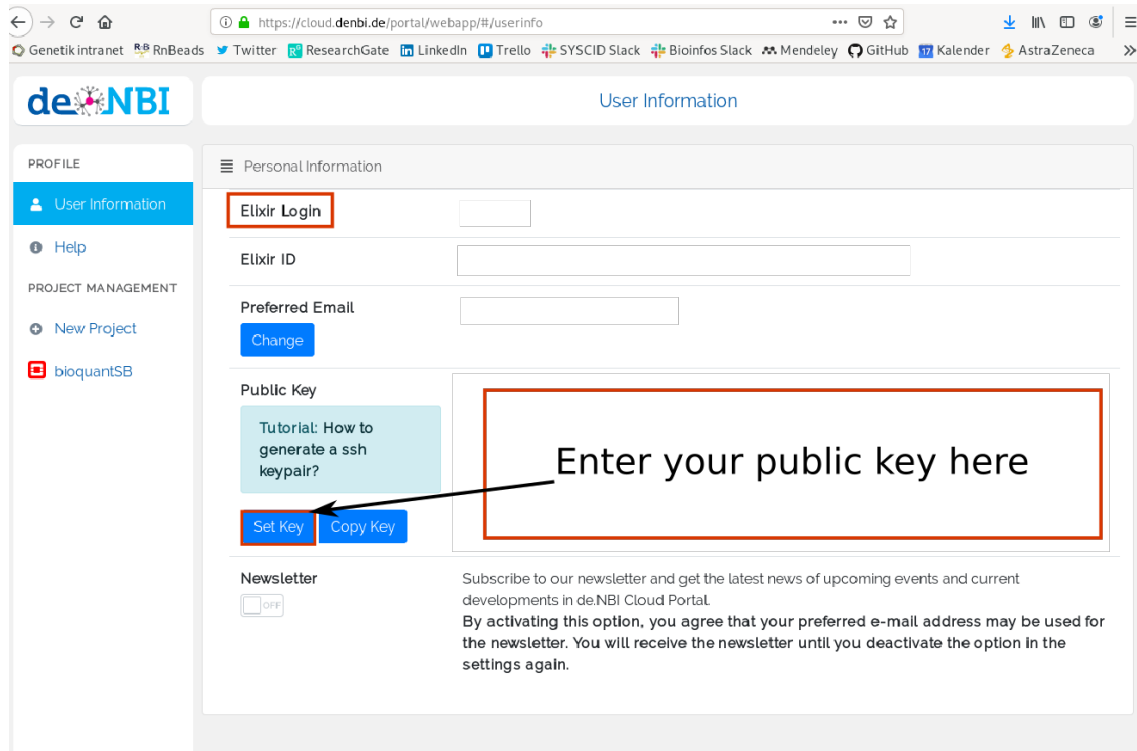
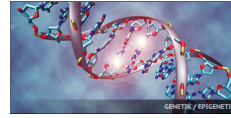


Figure 1: The de.NBI portal interface.

You will need two connections: one to the jumphost (`ssh -X -A denbi-jumphost-01.bioquant.uni-heidelberg.de`, ust leave this running in a minimized/hidden window) and directly to the computer that we use for computations (`ssh -X <username>@172.16.72.38`). Here, the <username> is again your ELIXIR ID.

- **Windows:** We recommend you to use a software called [MobaXterm](#). It is commercial, but there is a free version that well suffices for our use. It will automatically handle X11 forwarding and also allows file transfer. Just follow these steps to configure the connection:
 1. Click on Session and select SSH ([Figure 2](#))
 2. As remote host, give the end node, the machine you are going to work on. Also specify your username (ELIXIR login).
 3. Under "Advanced SSH settings", make sure X11 forwarding is ticked and add your private key ([Figure 3](#))
 4. Under "Network settings", activate the SSH gateway, specify server address, user and private key ([Figure 4](#)).

1.4 Test your configuration

After you logged into the cloud instance, create a folder with your username in your home folder. How many disks are available under `/vol`? To which groups do you belong?

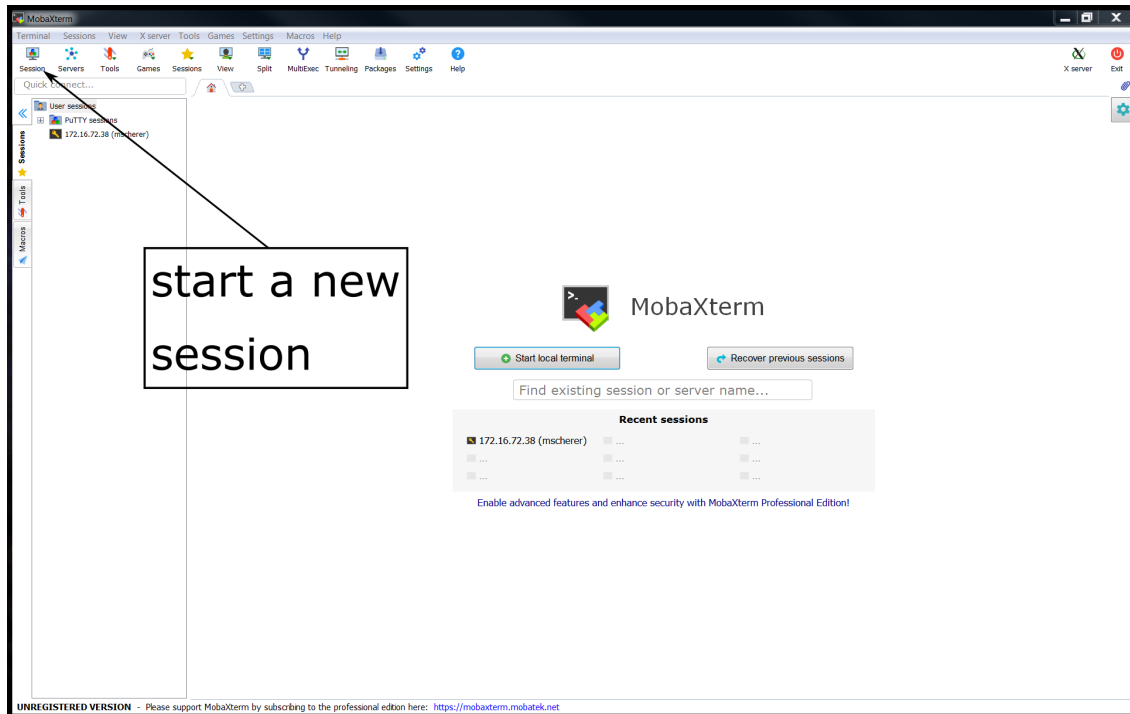


Figure 2: Start screen for MobaXterm.

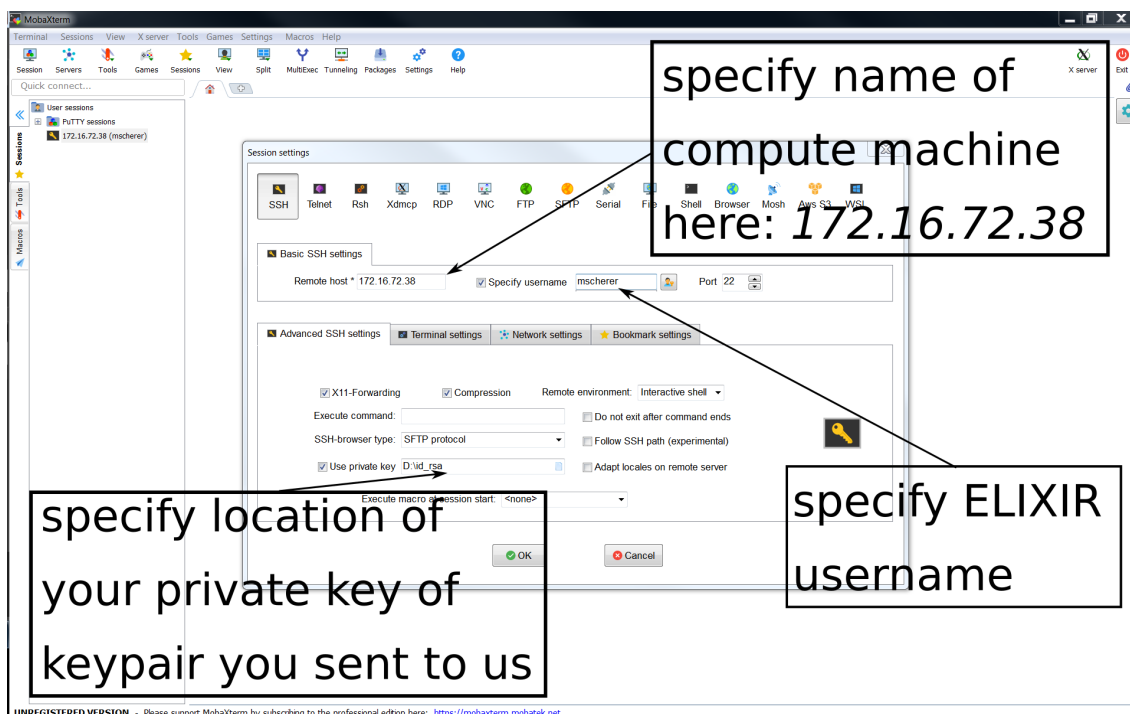


Figure 3: Advanced SSH settings for MobaXterm.

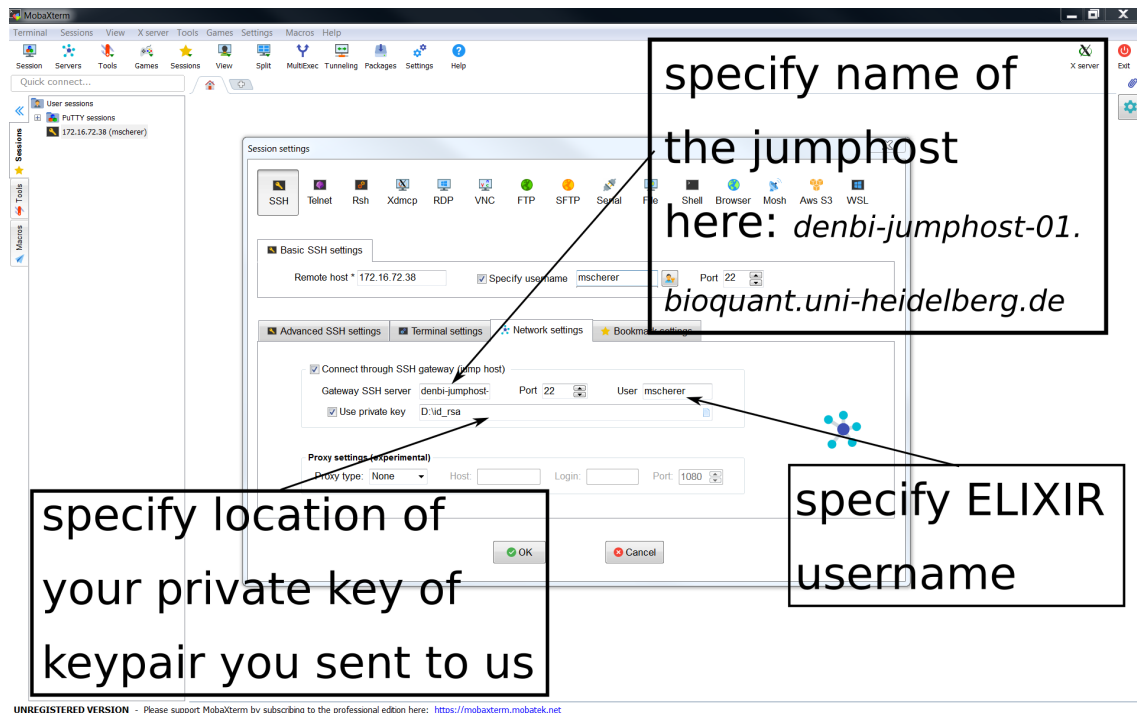
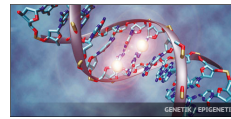


Figure 4: Network settings for MobaXterm.

2 Task2

2.1 Preparation questions

1. Read about basic Linux commands at <https://swcarpentry.github.io/shell-novice/>. How can you create *symbolic links* to all comma-separated value files (.csv) in a folder named files.

2. Read about basic R commands at <http://swcarpentry.github.io/r-novice-inflammation/>. What is a `data.frame` and what is the difference to a `matrix`?

3. Read about R-markdown at <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet> and here <https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf>. How can you specify in R-markdown that a specific code chunk is *not* to be executed during document rendering?



4. Explore the RNA pipeline, grape-nf, documented at <https://github.com/guigolab/grape-nf> Sketch an example for a sample annotation sheet (aka sample metadata sheet) here.

5. Read about IGV at <http://software.broadinstitute.org/software/igv/UserGuide>. Which data formats generated by grape-nf does IGV support?

6. Read about the Bioconductor-package edgeR: <https://bioconductor.org/packages/release/bioc/vignettes/tximport/inst/doc/tximport.html>. How is the basic R data object for storing data called?
