



## EyeTN workshop

March 30 - April 1 2015

## Quality Analysis

During this workshop we will run all programs on a server called *phix*. There we generated a folder with the data for this course under `/scratch/workshop`

### 1. Connect to the phix

- Open a terminal and connect to the phix (`ssh`)
- navigate to the folder `/scratch/workshop` (`cd`)
- create a folder with your name (`mkdir`)
- run the provided script (`script.sh`) and redirect the output to a file called “workshop.txt” in your folder.
- take a look into this file
- write your name into a new line in that file
- logout of the phix

### 2. Quality control

- create a folder in your home directory of your computer called “workshop”.
- mount the workshop directory of the phix on your computer:  
`sshfs user@phix:/scratch/workshop workshop`
- if this is not working, open “Computer” click on “file” “Connect to Server” use “ssh”  
server:“phix”,port:“22”
- connect to the phix
- the folder “data/1\_examples” contains a file called “example.fastq”. It contains reads generated from a NGS run.
- look into the file and try to understand the structure of this file (`less`)
- run a quality control program to assess the quality of this file (`fastqc`)
- look at the report, how is the quality of the reads? (on your computer)
- trim the reads so that every base has a quality of at least 30 (`fastq_quality_trimmer`)
- in your folder on the phix, create a folder for the new quality report
- again analyze the quality of the new fastq file

### 3. IGV

- logout of the phix (`exit`)
- the folder “data/examples” contains a fasta file. This is the reference genome in fasta format

- look at his file, try to understand the file format (`less`)
- start igv
- load the reference into igv
- load the gff file into igv
- load the file *assembly.bam* into igv
- load the file *transcription.bam* files into igv
- The two bam files contain a transcriptome assembly (*assembly.bam*) and a mapping of the reads against the reference (*transcripton.bam*). In the IGV view you can compare the real identified genes (from the gff file) to the assembled transcripts. And you can get an overview of some of the expressed genes. During the next days you will learn how you map the reads against such a reference and how you get real transcription levels out of these files.