

# BICF Nanocourse: Gene Expression and Regulation (06/06/2019)

**Your account and password are trainXX and passwordXX, respectively.**

Today we are going to:

- Align RNA sequencing reads to a reference genome
- Examine alignments using samtools
- Examine alignments in Integrative Genomics Viewer (IGV)

First, we will log into a compute node:

1. Open <https://portal.biohpc.swmed.edu/accounts/login/?next=terminal/webgui/> using a web browser
2. Input your training ID (trainXX) and password (passwordXX)
3. Go to Cloud Services, then click Web Visualization
4. Click the **Connect via web** link to launch your web visualization GUI

Note: The Linux commands shown in **red** need to be executed correctly in your home directory, while the other commands shown in black are optional for practice.

List of Linux commands

1. Display the current working directory  
`pwd`
2. Change directory to /usr  
`cd /usr`
3. Change directory to /archive/nanocourse/gene\_expr/trainXX , where trainXX is your account  
`cd /archive/nanocourse/gene_expr/trainXX`
4. Change directory to your home directory  
`cd ~`
5. Display files under the current directory  
`ls -l`

6. Display the files to be used during this session

```
ls -l /archive/nanocourse/gene_expr/shared/session1
```

7. Make a shortcut to the course files

```
ln -s /archive/nanocourse/gene_expr/shared/session1 session1
```

The file name and path to the real single-end sequencing reads for this workshop is:  
session1/reads/RNA.heart.e11.rep1.fastq.gz

To examine the file, you can optionally perform the following commands:

1. Extract a compressed read file and redirect to a text file

```
gzip -cd session1/reads/RNA.heart.e11.rep1.fastq.gz > RNA.heart.e11.rep1.fastq
```

2. Display the text file; to stop the command, type *ctrl+c*

```
cat RNA.heart.e11.rep1.fastq
```

3. Display the first 10 lines of the read file

```
head RNA.heart.e11.rep1.fastq
```

4. Display the first page of the read file; type *space* to go to the next page, and type *q* to quit

```
less RNA.heart.e11.rep1.fastq
```

5. Display the first page of the read file directly on the compressed file using pipe *|*

```
gzip -cd session1/reads/RNA.heart.e11.rep1.fastq.gz | less
```

6. Count the number of lines in a read file

```
wc -l RNA.heart.e11.rep1.fastq
```

7. Divide the number of lines we calculated above by 4 to get the number of reads

```
expr 113425724 / 4
```

Now we are going to run our alignment using the single-end reads against the mouse genome, mm10 (this alignment step using HISAT2 might take 15 to 20 minutes)

1. We will use HISAT2 with a graph index to align reads as follows:

```
session1/programs/hisat2 -p 8 -x session1/indexes/genome_snp_tran -U  
session1/reads/RNA.heart.e11.rep1.fastq.gz > heart.e11.sam
```

Next we will view the results of the alignment using samtools in two different formats - SAM (Sequence Alignment & Mapping) and BAM (Binary version of SAM):

1. Look at the SAM file (use *space* to go to the next page and *q* to quit)

```
less heart.e11.sam
```

## 2. Convert the SAM file into a BAM file

```
session1/programs/samtools view -@ 8 -bS heart.e11.sam > heart.e11.unsorted.bam
```

## 3. Create a sorted BAM file

```
session1/programs/samtools sort -@ 8 heart.e11.unsorted.bam -o heart.e11.sorted.bam
```

## 4. Make an index for the sorted BAM file

```
session1/programs/samtools index heart.e11.sorted.bam
```

## 5. Look at the sorted alignments

```
session1/programs/samtools view heart.e11.sorted.bam | less
```

## 6. Look at alignments between 61,989,341 and 61,990,361 on Chromosome 15

```
session1/programs/samtools view heart.e11.sorted.bam 15:61,989,341-61,990,361
```

## 7. Look at bases of reads located at a particular locus to identify variants

```
session1/programs/samtools mpileup -f session1/indexes/genome.fa -r 15:61,989,341-61,990,361 heart.e11.sorted.bam
```

Now we will look at the alignment using IGV (Integrative Genomics Viewer):

## 1. Run IGV

```
module add IGV/2.3.90
igv.sh
```

## 2. Load the genome using: Genomes -> Load from Server -> Select Mouse mm10

## 3. Open the BAM file using File -> Load From File in IGV and choose heart.e11.sorted.bam