

# homework\_03

Esteban Jorquera

24-02-2022

## Using IGV

### RNA-seq *P. chabaudi* AS

- -Load annotations for the genome (which one is the annotation file? What is this format describing?)

The annotation file is “PccAS\_v3.gff3”, the gff3 (General Feature Format) file format describes the coordinates of an annotation in this case a gene annotation

- What is the browser displaying?

The gene annotation track loaded from “PccAS\_v3.gff3” over the reference genome of *Plasmodium chabaudi* AS contained in the “PccAS\_v3\_genome.fa” file and its index “PccAS\_v3\_genome.fa.fai”

• What happens if you zoom in or out? If we zoom in we can see the individual genes including introns and exons, if we zoom out gene locations overlap showing the relative abundance of genes across the genome

• Load an alignment file for an RNA-seq experiment MT1 and MT2 MT1 and MT2 correspond to 2 different samples

• Is the data in the correct format? All files are raw sequencing reads in the fastq files, and therefore not alignments, requiring aligning using a program like hisat2

• What kind of data do you need? The data needed should be a sam or bam file which contains the reads already aligned to the reference genome

- Why are there two files per sample? There are 2 files per sample because the sequencing was pair-ended

### fastqc analysis of the *P. chabaudi* RNA-seq reads

```
#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
#You can edit the scriptsince this line
#
# Your job name
#$ -N Esteban_fastqc_plasmodium
```

```

# Send an email after the job has finished
#$ -m e
#$ -M ejorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads fastqc module,
# executes fastqc to analyze the 2 pair-ended samples of plasmodium chabaudi RNA-seq
(module load fastqc/0.11.3 ;
fastqc /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/MT1_1.fastq -o output ;
fastqc /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/MT1_2.fastq -o output ;
fastqc /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/MT2_1.fastq -o output ;
fastqc /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/MT2_2.fastq -o output)

```

### hisat2 alignment of the P. chabaudi RNA-seq reads

```

#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
#You can edit the scriptsince this line
#
# Your job name
#$ -N Esteban_hisat2_plasmodium

# Send an email after the job has finished
#$ -m e
#$ -M ejorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads hisat2 module,
# executes hisat2-build to index the plasmodium chabaudi genome
# executes hisat2 to align the 2 pair-ended samples of plasmodium chabaudi RNA-seq

(module load hisat2/2.0.0-beta ;
hisat2-build /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/PccAS_v3_genome.fa /mnt/Citosina/amedina/
hisat2 --max-intronlen 10000 -x /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/PccAS_v3_hisat2.idx -
hisat2 --max-intronlen 10000 -x /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/PccAS_v3_hisat2.idx -

```

### samtools sam to bam conversion, sorting and indexing of the P. chabaudi RNA-seq alignments

```

#!/bin/bash
# Use current working directory
#$ -cwd

```

```

#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
#$ -N Esteban_samtools_plasmodium

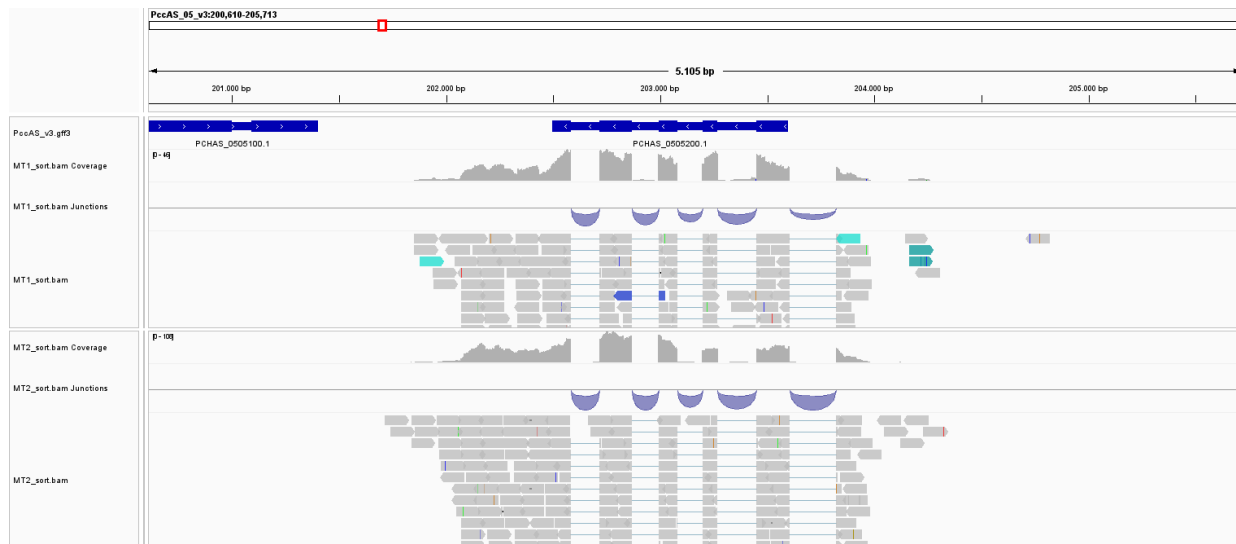
# Send an email after the job has finished
#$ -m e
#$ -M ejorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads samtools module,
# executes samtools view for sam to bam conversion of the plasmodium chabaudi AS RNA-seq alignment data
# executes samtools sort to sort the generated bam file
# executes samtools index to index the sorted bam file
(module load samtools/1.9 ;
samtools view -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT1.sam > /mnt/Citosina/amedina/
samtools view -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT2.sam > /mnt/Citosina/amedina/
samtools sort /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT1.bam -o /mnt/Citosina/amedina/
samtools sort /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT2.bam -o /mnt/Citosina/amedina/
samtools index -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT1_sort.bam ;
samtools index -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MT2_sort.bam)

#
scp ejorquera@dna.lavis.unam.mx:/mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/{MT1_1_fastqc
#
scp ejorquera@dna.lavis.unam.mx:/mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/{MT1_sort.bam

```

Now we can load the BAM files to IGV... Remember you need the index • What do you see, explore the genome. We see the individual reads of both samples aligned to the reference genome and we can see their alignment to the genes thanks to the annotation file

- Visualize loci: PCHAS\_0505200 and PCHAS\_1409500
- What do you see?



At the gene PCHAS\_0505200 we can see the alignment to the genome of the reads of both samples, reads align to the described exons in the gene, interestingly it seems that both samples are showing an extra exon or untranslated region upstream of the annotated 1st exon



At the gene PCHAS\_1409500 we see something similar than at PCHAS\_0505200, there also seems to be an extra exon upstream or untranslated region of the annotated 1st exon, however this time the gene is in the opposite strand

- Can you export the figure? Yes, the current browser view can be exported as either a png file or as svg file.

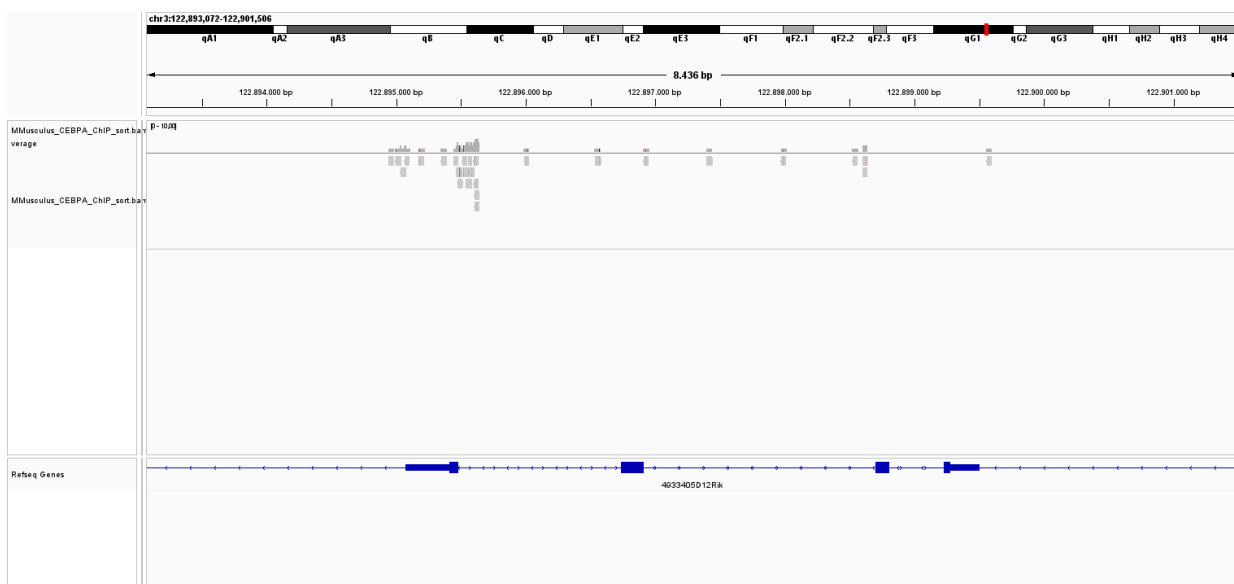
## CEBPA *M. musculus*

- Using IGV try to visualize the data for mouse ChIP-seq you generated. Is a liver data set, can you point to an interesting loci?

## samtools sam to bam conversion, sorting and indexing of the M. musculus CEBPA ChIP-seq alignment

```
#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
#$ -N Esteban_samtools_CEBPA

# Send an email after the job has finished
#$ -m e
#$ -M ejorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads samtools module,
# executes samtools view for sam to bam conversion of the mus musculus ChIP-seq alignment data
# executes samtools sort to sort the generated bam file
# executes samtools index to index the sorted bam file
(module load samtools/1.9 ;
samtools view -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_2/output/MMusculus_CEBPA_ChIP.sam > /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MMusculus_CEBPA_ChIP.bam
samtools sort /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MMusculus_CEBPA_ChIP.bam -o /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MMusculus_CEBPA_ChIP_sort.bam
samtools index -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MMusculus_CEBPA_ChIP_sort.bam)
```



## deeptools bam to bigwig conversion of the M. musculus CEBPA ChIP-seq alignment

```
#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
#$ -N Esteban_deeptools_CEBPA

# Send an email after the job has finished
#$ -m e
#$ -M ejorquera@uc.cl
#
# Line required if modules are to be used, source modules environment
. /etc/profile.d/modules.sh
#
# Loads deeptools module,
# executes bamCoverage for bam to bw conversion of the mus musculus ChIP-seq alignment data
(module load deeptools/2.5.3 ;
bamCoverage -b /mnt/Citosina/amedina/ejorquera/BioInfoII/Tarea_3/output/MMusculus_CEBPA_ChIP_sort.bam -
```

