

# ReporteT01

Axel R

2023-02-10

## Practica archivo BAM

```
# Entrar a un qlogin
qlogin
# Entrar a la carpeta donde se encuentra el archivo
cd /mnt/Timina/bioinfoII/format_qc
# Se copia el archivo a la carpeta donde se estara trabajando
cp NA20538.bam /mnt/Timina/bioinfoII/arodriguez
# Dirigirse a la carpeta donde se trabaja
```

## Practica bfc

## Practica estadísticas

### What is the total number of reads?

Usando el comando `samtools flagstat` se puede observar las estadísticas generales del archivo BAM. Si lo que se requiere saber es el **numero total de lecturas**, entonces se utiliza el comando:

```
samtools flagstat NA20538.bam | head -n 1
```

En donde el comando `head -n 1` funciona para imprimir solo la primera línea del output, donde se encuentran el número de lecturas totales.

```
347367 + 0 in total (QC-passed reads + QC-failed reads)
```

Como se observa, el número de *reads* totales fue de **347,367**.

### Otras alternativas

```
# Estadísticas generales
samtools flagstat NA20538.bam
# Lecturas totales ejemplo 2
samtools view -c NA20538.bam # 347367
```

```
# Lecturas totales ejemplo 3
samtools stats NA20538.bam | grep 'SN' | cut -f 2- # 347367
# Lecturas totales ejemplo 4
samtools stats NA20538.bam | grep 'raw total sequences' | cut -f 2- # 347367
```

- `view -c` cuenta los alineamientos e imprime el total.
- `samtools stats` muestra estadísticas relevantes sobre cada lane y grupo de lectura, así como información sobre las secuencias.
- `grep` busca e imprime las líneas coincidentes con un patrón (en el ej.3 'SN')
- La sección SN del comando `samtools stats` brinda un resumen con conteos, porcentajes y promedios, en un estilo similar al de `samtools flagstat`, pero más completo.
- `cut` se utiliza para seleccionar una columna del output. Con el parámetro `-f 2-` le indicamos que solo queremos las últimas dos columnas del output.
- `grep 'raw total sequences'` se utiliza para especificar que solo se requieren el total de lecturas, lo que nos da un output más ordenado.

## What proportion of the reads were mapped?

Con el mismo comando de `samtools flagstat NA20538.bam`, se observa que el porcentaje de **reads** mapeados fue de **93.26%** (323,966 lecturas).

```
# Filtramos los resultados que coincidan con 'mapped' e imprimimos
# todos los argumentos de la primera fila
samtools flagstat NA20538.bam | grep 'mapped' | awk '{print $0}' | head -n 1
323966 + 0 mapped (93.26% : N/A)
```

El comando `awk '{print $0}'` imprime la primera fila del output del comando `samtools flagstat NA20538.bam | grep 'mapped'`.

Generalmente `awk '{print $n}'` es utilizado para imprimir la  $n$ -ésima columna de un output. `awk 'NR==m {print $n}'` se utiliza para imprimir la  $m$ -ésima fila y la  $n$ -ésima columna.

## Otras alternativas

```
# Lecturas mapeadas (Mapped alignments)

## Ejemplo 2
samtools view -F 0x904 -c NA20538.bam # 323966
## Ejemplo 3
samtools view -c -F 260 NA20538.bam # 323966
## Ejemplo 4
samtools view -F 0x04 -c NA20538.bam # 323966
## Ejemplo 5
samtools stats NA20538.bam | grep 'SN' | grep 'reads mapped' | cut -f 2- #323966
```

## How many reads were mapped to a different chromosome?

```
samtools stats NA20538.bam | grep 'pairs on different chromosomes:' | cut -f 2-
```

De acuerdo a la sección SN del archivo, **4,055** lecturas fueron mapeadas a un cromosoma diferente.

### What is the insert size mean and standard deviation?

```
samtools stats -F SECONDARY NA20538.bam | grep "insert size" | cut -f 2-
```

Finalmente, en la sección de Summary Numbers (SN) generada por `samtools stats` se encuentra la información requerida.

- insert size average: **190.3**
- insert size standard deviation: **136.4**