

# Práctico 4

Curso de Bioinformática II, Biotecnología - ORT

31 de Marzo de 2016

## 1 SAMtools

SAMtools es un conjunto de herramientas que manipula los alineamientos en el formato BAM. Puede utilizar y exportar al formato SAM, ordena, indexa y permite recuperar los reads alineados de cualquier región del alineamiento con rapidez.

En este ejercicio vamos a trabajar con el archivo Tbovis6.bam (ver material extra) en una terminal de linux.

### 1.1 Visualización del alineamiento

#### **samtools view**

Sin parámetros devuelve todos los alineamientos de un archivo BAM en pantalla en formato SAM

```
samtools view Tbovis6.bam | less
```

Cuál es la diferencia de usar las opciones -h o -H ?

```
samtools view Tbovis6.bam -h |less
```

```
samtools view Tbovis6.bam -H |less
```

Qué programa se utilizó para alinear este archivo?  
Cuántos líneas de alinmiento contiene?  
Es un experimento single-end or paired end?

## 1.2 SAMtools flagstat

Devuelve información útil y general del mapeo, utilizando para ello el FLAG.

```
samtools flagstat Tbovis6.bam
```

Cuántos reads alinearon contra la referencia?  
Cuántos reads en total tiene el experimento?

## 1.3 SAMtools view

Los archivos SAM/BAM pueden o no incluir los reads no mapeados, en general los nuevos programas incluyen los reads no mapeados.

Para poder extraer sólo los reads mapeados de un archivo BAM se puede utilizar la opción **-F**

Dicha opción filtra los reads que cumplan con FLAG dado.

Para entender fácilmente un flag, o generar uno pueden ver:

<https://broadinstitute.github.io/picard/explain-flags.html>

Que flag debemos usar si queremos filtrar los reads no mapeados?  
Cuál es el total de reads mapeados?

hint: utilizar opción **-c**

Coincide el resultado con el obtenido con flagstat?

Cuántos FLAG distintos tiene el alineamiento?

hint: visualizar el archivo bam, extraer la segunda columna (que contiene el FLAG), ordenar y utilizar el comando **uniq** y por último contar las ocurrencias.

## 1.4 SAMtools order and index

Comunmente es necesario ordenar e indexar el archivo BAM, de forma que el acceso a él sea más eficiente. Para ello se utilizan las siguientes herramientas:

```
samtools sort Tbovis6.bam Tbovis6_sorted
samtools index Tbovis6_sorted.bam
```

## 1.5 Calidad del alineamiento

La quinta columna del archivo de alineamiento corresponde a la calidad del alineamiento (MAPQ)

Utilizando la siguiente línea de comando podemos ver las distintas calidades obtenidas en el mapeo:

```
samtools view Tbovis6.bam |cut -f5|sort |uniq -c |sort -nr
```

Observa las calidades y las veces que cada una aparece, qué nos indica acerca del mapeo y del programa elegido para el mismo?

Cuál es la calidad que más veces se obtuvo?

## 1.6 Filtro por MAPQ

Busca una opción de SAMtools view que permita descartar alineamientos de baja calidad.

Cuántos reads alinean con una calidad mayor de 40?