

# Preprocesado de Secuencias de NGS

## Sesión práctica

Trabajando con la muestra *mirna.fastq*

1. Realiza el control de calidad para la muestra.

¿Qué parámetros consideras indicativos de mala calidad?

Anota tus conclusiones

2. Realiza un **trimming** de calidad para la muestra, con un umbral mínimo de **calidad de 20**

¿Qué diferencias observas con el fichero original?

Anota tus conclusiones

3. Realiza un **trimming** de calidad para la muestra, con un umbral mínimo de **calidad de 28**

¿Ha mejorado la calidad con respecto al filtrado anterior?

Anota tus conclusiones

4. Realiza un **trimming** de calidad para la muestra, con un umbral mínimo de **calidad de 28**, eliminando lecturas con un **tamaño inferior a 30**.

¿Ha mejorado la calidad con respecto al filtrado anterior?

¿Cuántas lecturas se han eliminado?

Anota tus conclusiones

5. Realiza un **trimming** de calidad para la muestra, con un umbral mínimo de **calidad de 28**, eliminando lecturas con un **tamaño inferior a 35**

¿Ha mejorado la calidad con respecto al filtrado anterior?

¿Cuántas lecturas se han eliminado?

Anota tus conclusiones

Trabajando con la muestra *solid\_converted.fastq*

6. Realiza el control de calidad para la muestra.

¿Qué parámetros consideras indicativos de mala calidad?

Anota tus conclusiones

7. Realiza un *trimming* de calidad para la muestra, con un umbral mínimo de **calidad de 20**

¿Qué diferencias observas con el fichero original?

¿Consideras que ha sido efectivo el *trimming*?

Anota tus conclusiones

8. Realiza un *trimming* de calidad para la muestra, con un umbral mínimo de **calidad de 28**

¿Ha mejorado la calidad con respecto al filtrado anterior?

Anota tus conclusiones

9. Realiza un *trimming* de calidad para la muestra, con un umbral mínimo de **calidad de 28**, eliminando lecturas con un **tamaño inferior a 47**

¿Ha mejorado la calidad con respecto al filtrado anterior?

¿Cuántas lecturas se han eliminado?

Anota tus conclusiones

10. Elimina las lecturas con **menos de un 90%** de bases con **calidad superior a 20**

¿Ha sido efectivo el filtrado?

¿Cuántas lecturas se han eliminado?

Anota tus conclusiones

## Anexo 1: Comandos

**Trimming** de las secuencias con una **calidad mínima de 20**

```
<RUTA_FASTX_TOOLKIT>/fastq_quality_trimmer -t 20 -i <sample>.fastq -o <sample>_q20.fastq
```

**Trimming** de las secuencias con una **calidad mínima de 20** y una **longitud mínima de 30**

```
<RUTA_FASTX_TOOLKIT>/fastq_quality_trimmer -t 20 -l 30 -i <sample>.fastq -o  
<sample>_q20_l30.fastq
```

**Trimming** de las secuencias con una **calidad mínima de 20** y una **longitud mínima de 30**

```
<RUTA_FASTX_TOOLKIT>/fastq_quality_trimmer -t 20 -l 30 -i <sample>.fastq -o  
<sample>_q20_l30.fastq
```

**Eliminación** de lecturas con **menos de un 90%** de bases con calidad menor o igual a 20

```
<RUTA_FASTX_TOOLKIT>/fastq_quality_filter -q 20 -p 90 -i <sample>.fastq -o  
<sample>_q20_p90.fastq
```