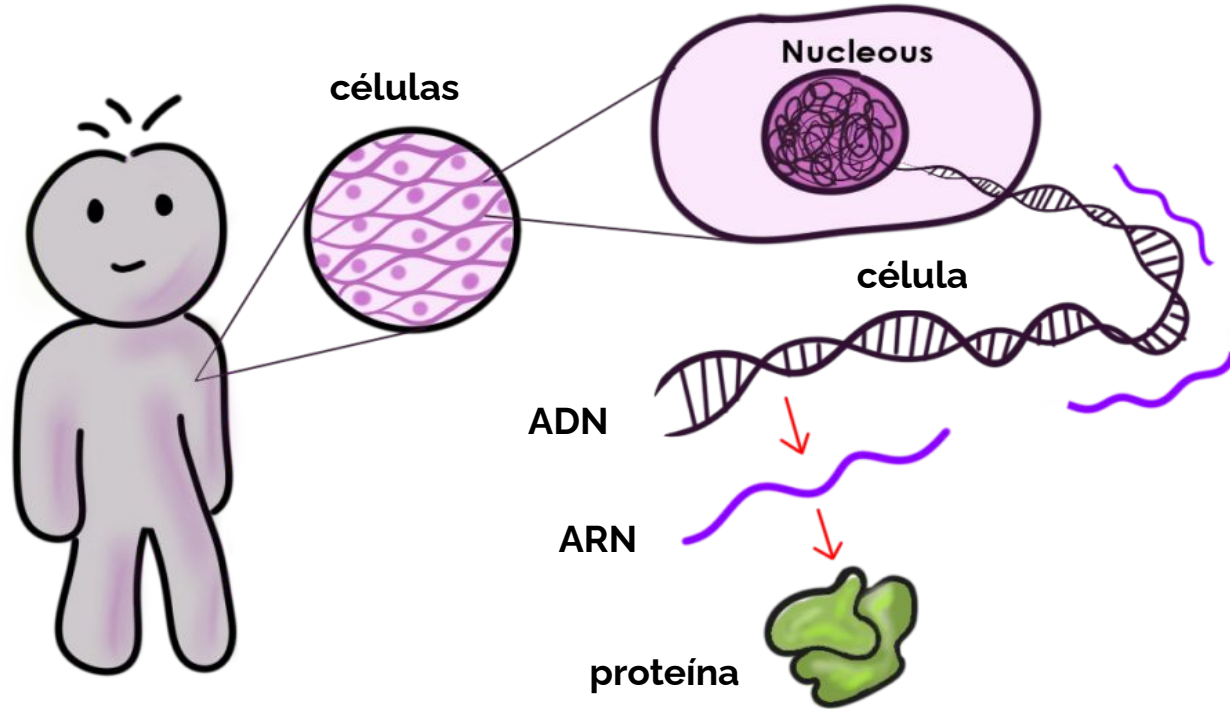


FastQ

Alejandra Cervera Taboada
acerverat@inmegen.gob.mx

Herramientas básicas de bioinformática
enero de 2024

Esto ya lo saben



RNA-seq



Fastq

```
--->gzip -cd L2I_S1_L001_R1_001.fastq.gz | head
@M00805:5:000000000-A0VLL:1:1101:16473:1320 1:N:0:1
NTTGTTCATCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTTTGCTAAT
TTGGAAGTCTAGGTCAGCTGAAGATCCTGTGAGCGAAGTTCCGGCAGTGTACACAGCAC
+
#55<<?BBDBDDDDDDFFFFFHFFFHHHHHFFHAFHHHHHHHHHHBHHHHHFFFHFFFHHHHHDGDGHC
AFHFHHHHHHHFGHDDHFBFHDFHFFFHHHHFFA=@BEEEEED)@<B?BE3==?EEEE
@M00805:5:000000000-A0VLL:1:1101:15023:1321 1:N:0:1
NAGAAATCACAGACATACAAAGCAGTCTGTGTCTTAGGTCCTGAGCAGCCTCCAGCACATTCT
AGCATCTGCCGTCACATTGTTCTGCACACACCGTCCTTGTCAGTGCAGAAGACAGA
+
#55???BBDEDDDDDDGGGGGGIIIIIIIIIIIIIIIIIIIIHIIHIIHFGHHHIIIIIIIIIIHIIII
HHHHHHHHHHHHHHHHHHHHHHHHHHHHGGFGEGGGGGGGGGGGGGGGGGEGGGGGCEGG>
@M00805:5:000000000-A0VLL:1:1101:14046:1321 1:N:0:1
NTTTCGTGGAAGTGGGTACCTGACAGTGTGCACGCCCCCAGCAGGTTCACAATATTCTCGTGG
ACATGAGTGCCTCTCTTTCAGAGCTGTCTGCTTTTTCTGTCAAAGAAAGGAGCATT
```

Fastq

ID lectura

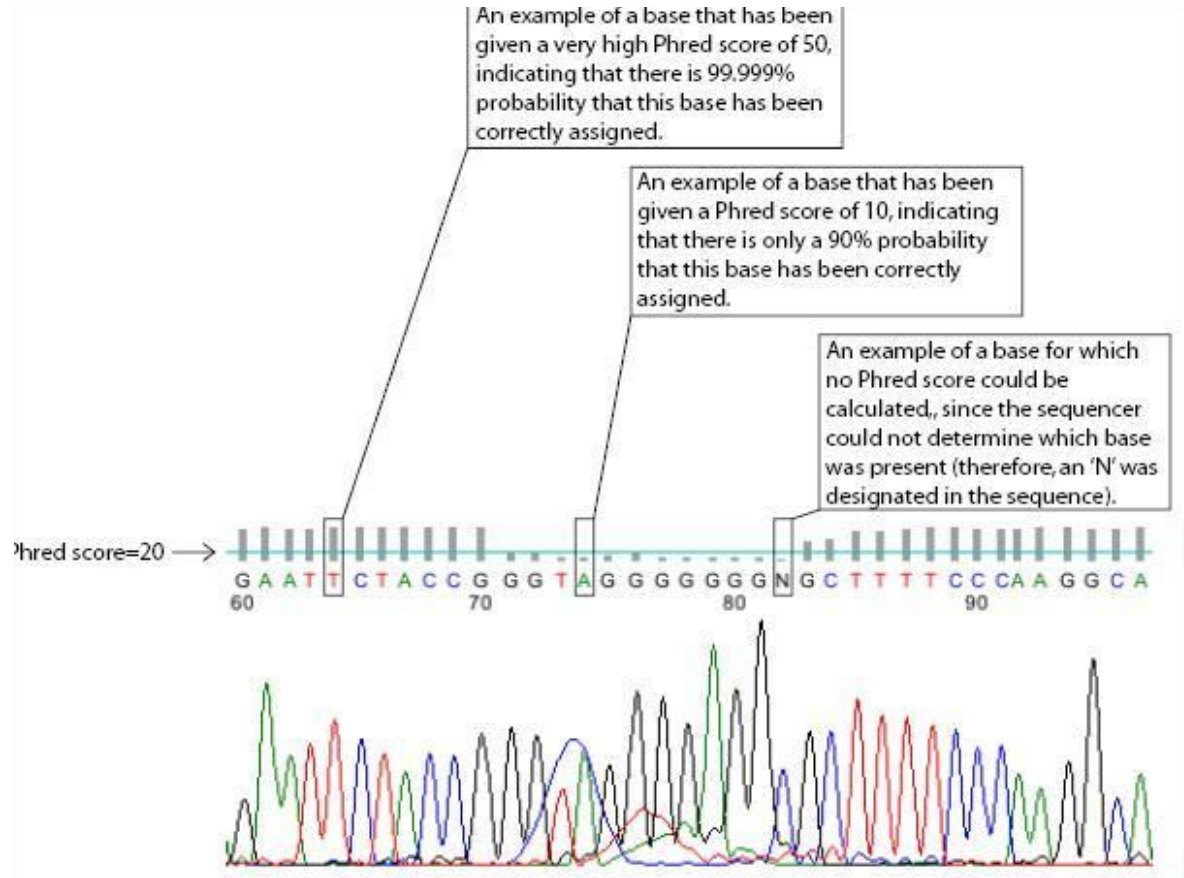
```
--->gzip -cd L2I_S1_L001_R1_001.fastq.gz | head
@M00805:5:000000000-A0VLL:1:1101:16473:1320 1:N:0:1
NTTGTCATCAGCTGAAGATGAAATAGGATGTAATCAGACGACACAGGAAGCAGATTTTGCTAAT
TTGGAAGTGGTCTGAGATCCTGTGAGCGAAGTTCGGCAGTGTACAGCAC
+
#55<<?BBDBDDDDDDFFFFFHFFFHHHHHFFHAFHHHHHHHHHBHHHHHFFHFFFHHHHHHHDGDGHC
AFHFHHHHHHHFGHDDHFBFHDFFHFFFHHHFFA=@BEEEEED)@<B?BE3==?EEEE
@M00805:5:000000000-A0VLL:1:1101:15023:1321 1:N:0:1
NAGAAATCACAGACATACAAAGCAGTCTGTGTCCTTAGGTCCTGAGCAGCCTCCAGCAGATTCT
AGCATCTGCCGTCACATTGTTCTGCACACACCGTCCTTGTCAGTGCAGAAAGACAGA
+
#55???BBDEDEDDDDGGGGGGIIIIIIIIIIIIIIIIIIIIIIHIIHIIHFGHHHIIIIIIIIIIHIIII
HHHHHHHHHHHHHHHHHHHHHHHHHHHHHGGFGEGGGGGGGGGGGGGGGGGGGEGGGGGCEGG>
@M00805:5:000000000-A0VLL:1:1101:14046:1321 1:N:0:1
NTTTCGTGGAAGTGGGTTACCTGACAGTGTGCACGCCCCCAGCAGGTTACAATATTCTCGTGG
ACATGAGTGCCTCTCTTTTCAGAGCTGTCTGCTTTTTCTGTCAAAGAAAGGAGCATT
```

Secuencia

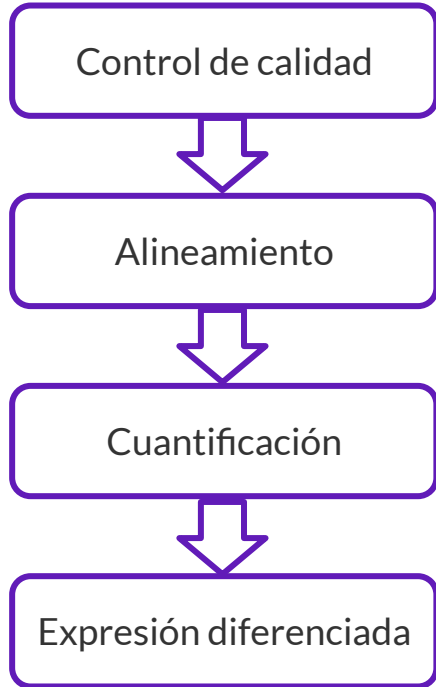
Score

Phred Score

- El cromatograma hasta abajo es producido por el secuenciador
- Cada color representa una base distinta (A,T,C,G)
- Las barras grises representan el Phred score para cada base



Pipeline básico



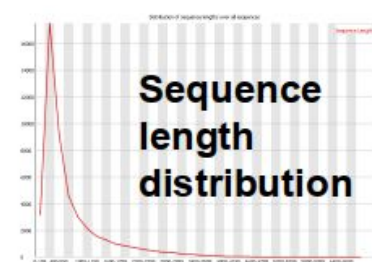
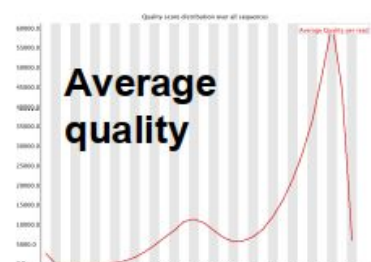
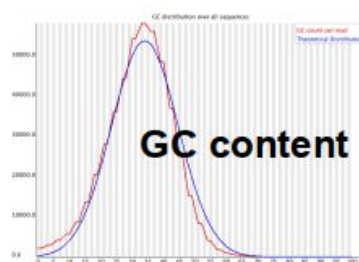
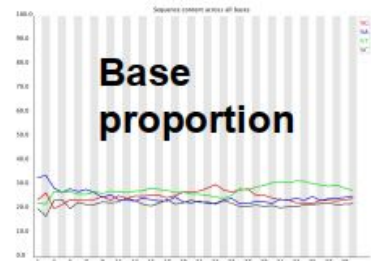
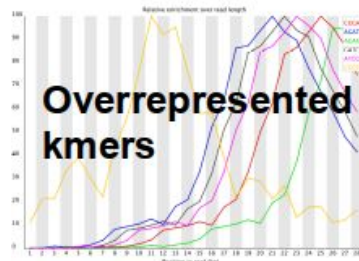
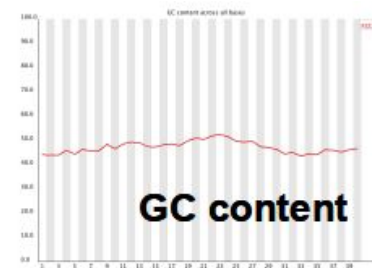
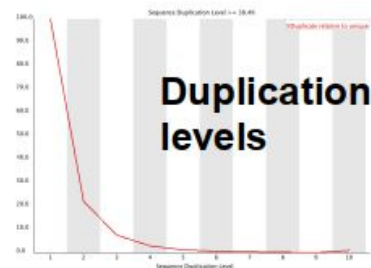
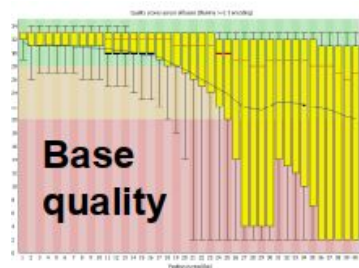
Pipeline básico

Control de calidad

Alineamiento

Cuantificación

Expresión diferenciada




```
$ cd datos.taller

# Creamos un directorio para nuestros datos
$ mkdir RNA_seq
$ cd RNA_seq

# Descargamos los datos para el curso
$ wget http://genomedata.org/rnaseq-tutorial/HBR\_UHR\_ERCC\_ds\_5pc.tar

# Los descomprimimos
$ tar xvf HBR_UHR_ERCC_ds_5pc.tar

# Inspeccionamos las primeras líneas de uno de los archivos
$ zcat HBR_Rep1_ERCC-Mix2_Build37-ErccTranscripts-chr22.read1.fastq.gz | head
```