

Aprende a utilizar los flujos de análisis bioinformáticos internos del Inmegen

Cuantificación y análisis de expresión diferencial

Profesores: Dra. Alejandra Cervera
Dra. Laura Gómez
Dr. Daniel Pérez



Repaso clase anterior

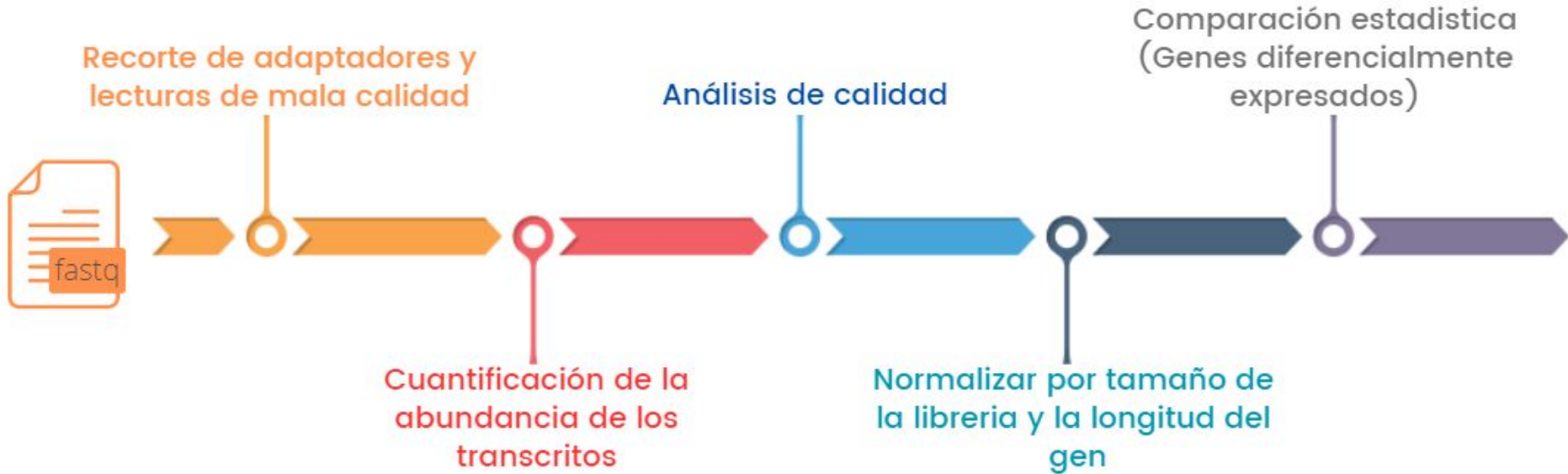
1. ¿Qué es NextFlow?
2. ¿Qué es Docker?
3. ¿Cuales son los principales archivos que componen un flujo de trabajo de NextFlow?
4. ¿Dudas?

Temario del curso

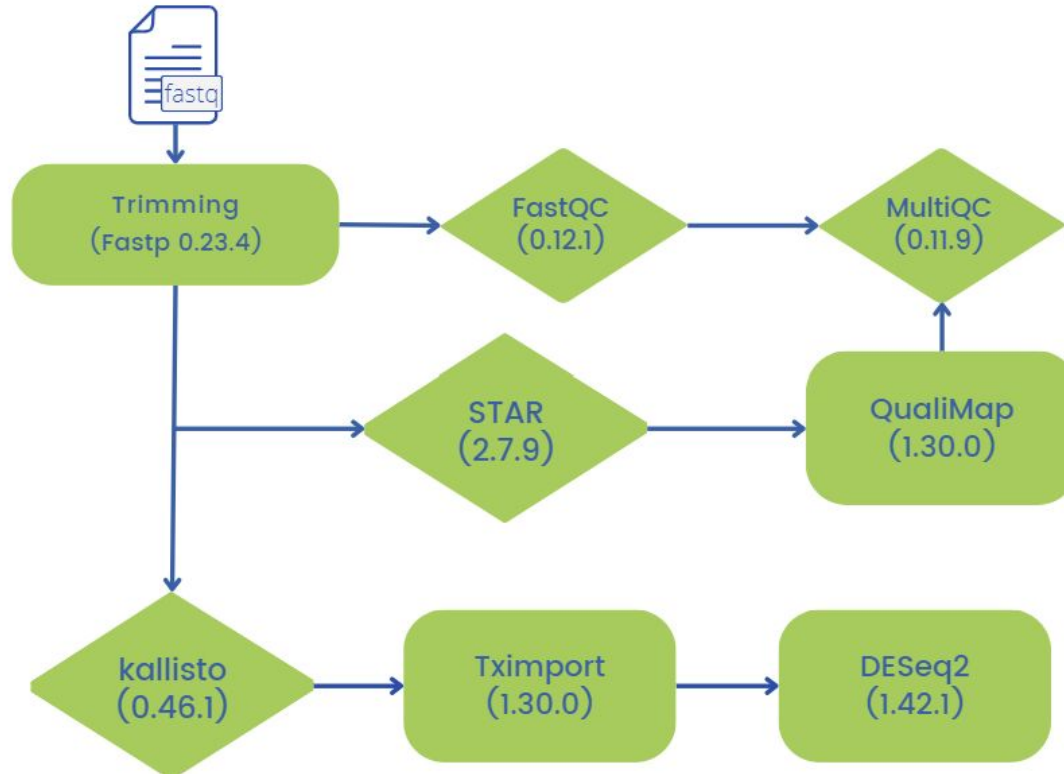
- Repaso de Bash
- Introducción a Nextflow y Docker
- **Cuantificación y análisis de expresión diferencial**
- Identificación automatizada de variantes germinales
- Identificación automatizada de variantes somáticas
- Identificación automatizada de variantes de datos de RNA-seq

Cuantificación y análisis de expresión diferencial

Flujo de trabajo del análisis de expresión diferencial (pipeline Q&DEA)



Herramientas del pipeline Q&DEA



Herramientas utilizadas

- **Recorte de adaptadores**
 - Fastp
- **Cuantificación de la abundancia de los transcritos**
 - Kallisto
- **Análisis de expresión diferencial**
 - R
 - Tximport
 - DESeq2
- **Análisis de calidad**
 - STAR
 - QualiMap
 - MultiQC

Flujo de trabajo de cuantificación y análisis de expresión diferencial (pipeline Q&DEA)

Explore el repositorio

https://github.com/INMEGEN/Pipelines_INMEGEN/tree/Principal

Particularmente el directorio **QDEA-RNAseq**

¿Qué elementos del código de NextFlow reconocen?

Archivo `sample_info.tsv`

Este archivo contiene:

- Nombre de las muestras (`Sample`)
- Ruta absoluta al archivo `R1` (`R1`)
- Ruta absoluta al archivo `R2` (`R2`)

Ejemplo:

```
Sample    R1      R2
ID_N1    /path/to/read_1/label1_R1.fastq.gz /path/to/read_2/label1_R2.fastq.gz
ID_N2    /path/to/read_1/label2_R1.fastq.gz /path/to/read_2/label2_R2.fastq.gz
```


Archivo `metadata.tsv`

Este archivo contiene:

- Nombre de las muestras (`Sample`)
- Nombre de las muestras (`SampleID`); nombre que aparece en los gráficos generados
- Condición de las muestras (`condition`); Describe brevemente la condición experimental de cada una de las muestras (normal, tratada, tumor, etc). Por el momento solo se puede comparar 2 condiciones a la vez.

Ejemplo:

<code>Sample</code>	<code>SampleID</code>	<code>condition</code>
<code>ID_N1</code>	<code>sample_label1</code>	<code>Normal</code>
<code>ID_N2</code>	<code>sample_labe2</code>	<code>Tratada</code>

Archivo [nextflow.config](#)

Contiene los parámetros necesarios para correr el flujo de trabajo.

Los parámetros que es imprescindible editar son:

- Directorio de salida
- Nombre del proyecto
- Ruta absoluta y nombre del índice de kallisto
- Ruta absoluta y nombre de la referencia de STAR
- Nombre del archivo GTF

Ejercicio

Con las muestras de RNA seq utilizadas en la clase 1 correr el flujo de trabajo cuantificación y análisis de expresión diferencial (pipeline Q&DEA)

¿dudas?