

Aprende a utilizar los flujos de análisis bioinformáticos internos del Inmegen

Identificación conjunta de variantes germinales (WES/WGS)

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



Repaso clase anterior

1. ¿Dudas del pipeline VC-RNAseq?

Repositorio de la clase

https://github.com/INMEGEN/Clase_pipelines/tree/main

Temario del curso

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

Identificación conjunta de variantes germinales (WES/WGS)

Variante genética

El Instituto Nacional del Cáncer (NIH) de Estados Unidos define una variante como:

Una alteración en la secuencia más común de nucleótidos del ADN. El término variante se usa para describir una alteración que puede ser benigna, patógena o de repercusión incierta. Este término se usa cada vez más en lugar del término mutación. También se llama variación de secuencia, variación genética, variante de secuencia y variante genética.

Variantes

Sequence Variants

SNV (Single Nucleotide Variant)

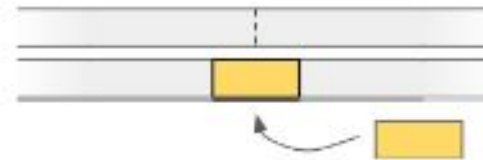
Ref	A	A	G	G	G	C	T	G
Query	A	A	G	G	A	C	T	G

INDEL (Insertion or Deletion)

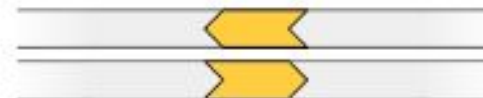
Ref	A	A	G	G	G	C	T	G	
Query	A	A	G	-	-	-	C	T	G

Structural Variants

Insertion



Inversion

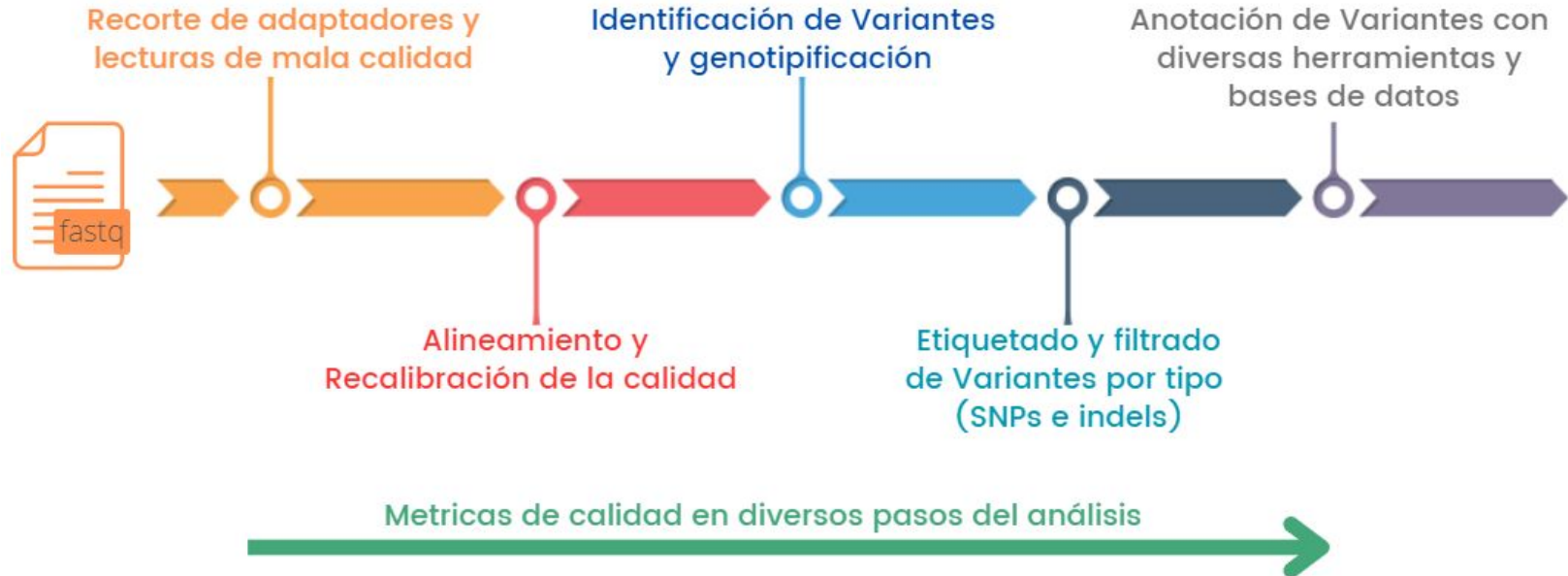


Variante de línea germinal

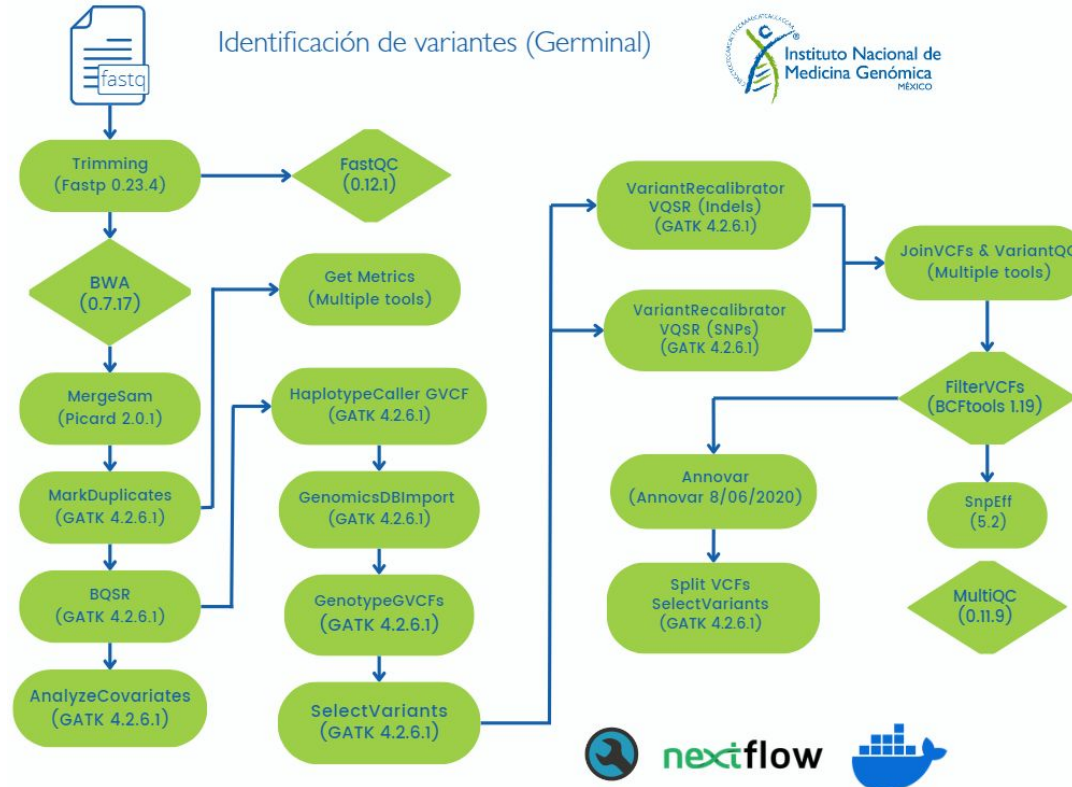
El Instituto Nacional del Cáncer (NIH) de Estados Unidos define una variante de línea germinal como:

Variante que se encuentra en una célula reproductora (óvulo o espermatozoide) y que está presente en el ADN de todas las células de la descendencia. Una variante de la línea germinal pasa del progenitor a su descendencia, por lo tanto, **es hereditaria**. También se llama mutación de la línea germinal y variante germinal.

Identificación conjunta de variantes germinales de datos de exoma o genoma completo



Herramientas del pipeline VC-Germline



Identificación conjunta de variantes germinales (VC-Germline)

Explore el repositorio

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

Particularmente el directorio **VC-Germline**

¿Qué elementos del código de NextFlow reconocen?

Archivo `sample_info.tsv`

Este archivo contiene:

- Nombre de la muestra (Sample_name)
- Identificador de la muestra (SampleID)
- Barcode de la flowcell y número de lane. (RG_PU)
- Tecnología de secuenciación (RG_PL)
- Barcode de la librería de secuenciación (RG_LB)
- Ruta absoluta al archivo R1 (R1)
- Ruta absoluta al archivo R2 (R2)

Ejemplo:

Sample_name	SampleID	RG_PU	RG_PL	RG_LB	R1	R2
IDS1 ID_L1 FLOWCELL.1	ILLUMINA	BARCODE			Path/to/fastq_S1_R1.fastq	Path/to/fastq_S1_R2.fastq
IDS2 ID_L2 FLOWCELL.2	ILLUMINA	BARCODE			Path/to/fastq_S2_R1.fastq	Path/to/fastq_S2_R2.fastq

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
- Si las muestras se distribuyeron en diferentes carriles (lanes)
- Ruta absoluta y nombre del índice de BWA
- Ruta absoluta de los archivos necesarios para BQSR
- Ruta absoluta de las bases de datos de Annovar

Ejercicio

Seguir las instrucciones del repositorio para correr el flujo de trabajo de NextFlow que se encuentra en el directorio VC-Germline

¿dudas?