



## Tercera prova d' avaluació contínua: Anàlisi de variants minoritàries amb Galaxy

1. Introducció .....	1
1.1 Enunciat de la PEC.....	1
2. Informe de l'anàlisi.....	3
2.1 Estructura bàsica de l' informe.....	3
2.2 Sobre el format de l' informe.....	3

**Data de publicació de l'enunciat: 11/01/2024**

**Data límit per presentar la PEC: 26/01/2024<sup>1</sup>**

### 1. Introducció

#### 1.1 Enunciat de la PEC

L' objectiu d' aquesta activitat és realitzar una anàlisi de variants minoritàries seguint una metodologia similar a l' exposada en els exemples proporcionats en els materials d' estudi del programa.

En aquesta ocasió, ens enfocarem en la identificació de petites variants (SNVs/indels) en l'ADN genòmic humà, utilitzant dades obtingudes del projecte dels 1000 genomes, accessible a través del repositori ISGR.

Les mostres amb les quals treballareu estan formades per dos arxius de seqüències aparellats "*Paired End reads*"

---

<sup>1</sup> La data de lliurament és la que s' indica en l' enunciat de la PEC. En cas de no coincidir amb la indicada a l'aula, aquesta (la de l'enunciat) serà la que predomini.



Per tal de no sobrecarregar computacionalment l'exercici, hem extret aleatòriament 10 conjunts de seqüències de cada parella d'arxius fastq. Amb ells hem generat deu parells d'arxius amb les mateixes seqüències, seleccionades també de forma aleatòria.

Aquests parells s'han denominat seqData\_1.fq/seqData\_2.fq, seqData1\_1.fq/seqData1\_2.fq, i així successivament fins seqData9\_1.fq/seqData9\_2.fq.

Cadascun de vosaltres ha d'escollir *els arxius de la parella d'arxius corresponent a l'última xifra o a l'últim codi numèric del vostre DNI o equivalent* (per exemple, passaport).

Podeu descarregar la parella d'arxius de la següent carpeta de google drive<sup>2</sup>:

[https://drive.google.com/drive/folders/1mbBfMRth-VGUqOd\\_1zgPj\\_EM5kde8P4L?usp=sharing](https://drive.google.com/drive/folders/1mbBfMRth-VGUqOd_1zgPj_EM5kde8P4L?usp=sharing)

Amb les dades seleccionades haureu de realitzar una anàlisi de variants "estàndard", seguint el procés que hem utilitzat en els exemples vistos a l'aula, tot i que no hi ha problema en què introduïu canvis, sempre que els justifiqueu. Bàsicament, aquest pipeline inclou:

1. Càrrega de les dades.
2. Control de qualitat<sup>3</sup>
3. Alineació dels "reads" al genoma.
4. Recerca de diferències entre les seqüències alineades i el genoma de referència
5. Visualització dels arxius BAM amb un visor de genoma integrat
6. Filtratge i Anotació de variants

Cal assenyalar que algunes etapes del procés estan definides de manera "difusa". Això és així, no per deixadesa sinó amb l'objectiu de proporcionar flexibilitat, permetent que, en explicar les decisions preses, es mantingui la coherència en l'anàlisi.

---

<sup>2</sup> Recordeu que, per accedir a la carpeta de drive que conté les dades haureu d'iniciar prèviament sessió en google amb el vostre usuari de la UOC.

<sup>3</sup> No cal que us estengueu en les explicacions dels resultats d'aquest apartat ja que ja el treballem en l'activitat 1.



## 2. Informe de l'anàlisi

Un cop realitzat l'anàlisi hauríeu d'elaborar un informe exposant què heu fet, com ho heu fet i quins resultats heu obtingut.

### 2.1 Estructura bàsica de l' informe

Bàsicament el vostre informe no ha de deixar de contenir les parts següents:

1. Abstract, amb un breu resum sobre el procés i els principals resultats.
2. Objectius de l' estudi
3. Materials i Mètodes: Naturalesa de les dades; procediment general d' anàlisi; utilitzades.
4. Resultats
5. Discussió i limitacions de l' estudi
6. Conclusió a la qual us porta la vostra anàlisi

### 2.2 Sobre el format de l' informe.

En treballar amb Galaxy, no és tan senzill integrar informe i resultats com quan es treballa amb Rmarkdown, per la qual cosa teniu llibertat a l'hora d'escollir el format del vostre informe. Això sí, penseu en quin és l'objectiu d'un informe, és a dir *comunicar*. És important que sigui fàcil de llegir, i que tant el procés com les conclusions i resultats importants estiguin ben visibles i localitzats. Utilitzeu un abstract al principi si voleu fer un resum del que heu fet i obtingut, taules de continguts i, opcionalment de figures. I col·loca allò que es pugui considerar secundari, addicional o opcional, en un apèndix.