

# Curso ALAB JC 2025

## Recursos de consulta

- Presentación del curso: [Curso MexVar ALAB - Juan Comas 2025 - Presentacion de Google](#)
- Repositorio guía: [GitHub - Curso\\_ALAB2025: Introducción](#)

```
1 #####
2 # 1. Quitar SNPs problemáticos
3 #####
4
5 # Buscar variantes duplicadas (mismo ID/posición) en el archivo bi
6 # --bfile poptest : usa poptest.bed/.bim/.fam
7 # --list-duplicate-vars : detecta variantes duplicadas
8 # ids-only suppress-first : solo imprime los IDs y marca para elim
9 plink --bfile poptest --list-duplicate-vars ids-only suppress-firs
10
11 # Contar cuántas variantes duplicadas se encontraron
12 wc -l plink.dupvar
13
14 # Eliminar las variantes duplicadas del dataset original y guardar
15 # --exclude plink.dupvar : lista de SNPs a excluir
16 # --make-bed : genera nuevos .bed/.bim/.fam
17 # --out poptest.t1 : prefijo del nuevo archivo
18 plink --bfile poptest --exclude plink.dupvar --make-bed --out pop
19
20 # Identificar INDELS (inserciones/deleciones) en el archivo .bim
21 # En el .bim, las columnas 5 y 6 son los alelos.
22 # Aquí buscamos alelos con I o D (I = insertion, D = deletion)
23 # y nos quedamos con la columna 2 (ID del SNP) para hacer una lista
24 awk '$5 ~/I/ || $5 ~/D/ || $6 ~/I/ || $6 ~/D/ ' poptest.t1.bim | c
25
26 # Contar cuántos INDELS se van a eliminar
27 wc -l indels.txt
28
29 # Eliminar INDELS del dataset y guardar como poptest.t2
30 plink --bfile poptest.t1 --exclude indels.txt --make-bed --out pop
31
32 #####
33 # 2. Missingness y heterocigosidad
34 #####
35
36 # Calcular missingness (por individuo y por SNP) para poptest.t2
```

```

37 # Esto genera archivos .imiss (por individuo) y .lmiss (por SNP)
38 plink --bfile poptest.t2 --missing --out poptest.t2
39
40 # Calcular tasa de heterocigosidad por individuo
41 # Esto genera archivo .het
42 plink --bfile poptest.t2 --het --out poptest.t2
43
44 # Crear carpeta para guardar los gráficos de QC
45 mkdir QCplots
46
47 # Correr script de R que hace las gráficas de QC
48 # Argumentos: <prefix PLINK> <etapa> <nombre del dataset>
49 # Aquí: usamos poptest.t2, etapa "pre" (antes de filtros fuertes) :
50 Rscript /data/data_lab/projects/AfroCaribbean/scripts_for_training
51
52 #####
53 # 3. Relacionados (IBD/parentesco)
54 #####
55
56 # Calcular IBD (identical by descent) entre todos los pares de ind
57 # Se usa después por un script en Perl (Anderson et al. 2010)
58 # para detectar parientes de primer grado.
59 plink --bfile poptest.t2 --genome --out poptest.t2
60
61 # El script de Anderson et al. usa poptest.t2.genome y genera fail
62 # con los individuos que tienen IBD > 0.5 (parientes cercanos).
63 # Aquí solo contamos cuántos individuos serán removidos.
64 wc -l fail-IBD-QC.txt
65
66 # Remover individuos emparentados (listados en fail-IBD-QC.txt)
67 # y guardar dataset como poptest.t3
68 plink --bfile poptest.t2 --remove fail-IBD-QC.txt --make-bed --out
69
70 #####
71 # 4. Filtros estándar de calidad (QC)
72 #####
73
74 # Aplicar filtros de:
75 # --mind 0.1 : elimina individuos con >10% de genotipos faltantes
76 # --geno 0.05: elimina SNPs con >5% de genotipos faltantes
77 plink --bfile poptest.t3 --mind 0.1 --geno 0.05 --make-bed --out po
78
79 # Filtrar SNPs que violan Hardy-Weinberg:
80 # --hwe 1e-5 : elimina SNPs con p(HWE) < 1x10^-5 (posibles errores
81 plink --bfile poptest.t4 --hwe 1e-5 --make-bed --out poptest.final
82
83 #####

```

```
84 # 5. Quedarse solo con autosomas y revisar QC
85 #####
86
87 # Conservar solo cromosomas autosómicos (1-22)
88 plink --bfile poptest.final --autosome --make-bed --out poptest.fi
89
90 # Recalcular heterocigosidad en autosomas
91 plink --bfile poptest.final.autosomes --het --out poptest.final.au
92
93 # Recalcular missingness en autosomas
94 plink --bfile poptest.final.autosomes --missing --out poptest.fina
95
96 # Correr de nuevo el script de R para hacer las gráficas de QC
97 # Ahora etapa "Post" (después de filtros) y nombre "Test"
98 Rscript /data/data_lab/projects/AfroCaribbean/scripts_for_training,
```