

Procesamiento de los datos del microarreglo MEGA

1. Filtrar los genotipos por call frequency (0.95) y remover sitios que no son SNPs (ej. indels).
2. Generar los archivos plink desde Genome Studio.
3. Migrar estos archivos al mismo directorio que contiene los siguientes archivos:
 - `list_snps_to_change_alleles.txt` (540,597 SNPs)
 - `new_positions_snps.txt`
 - `MEGA_Consortium_v2_15070954_A1_b138_rsids.txt`
 - `1KG_Omni.bed`; `1KG_Omni.fam`; `1KG_Omni.bim`
 - `convert_MEGA_alleles_sentli.pl` : este script remueve SNPs mapeados en el cromosoma 0, SNPs duplicados (por posición física y por ID), cambia los alelos, actualiza la posición física de dos SNPs, renombra los SNPs, y finalmente calcula la concordancia de genotipos de alguno de estos controles: HG01946, HG01935, NA21440, NA21317.
IMPORTANTE 1: Los controles se tienen que llamar de esta manera o habrá un error en el programa.
IMPORTANTE 2: La concordancia debe de ser mayor a 0.95.
4. Correr el script: `perl convert_MEGA_alleles_sentli.pl archivo.ped`
5. Este programa genera los siguientes archivos:
 - `archivo.flip.ped` y `archivo.flip.map` : archivos PLINK finales.
 - `duplicate_snps_archivo.txt` : ID de los SNPs que tienen el mismo nombre.
 - `duplicate_positions_archivo.txt` : ID de los SNPs que tienen la misma posición.
6. Aparte de los archivos procesados, hay dos archivos que se pueden dar a los usuarios:
 - `multiallelic_SNPs.txt` tiene la lista de los SNPs que no son bialélicos (9,421 SNPs).
 - `complementary_snps.txt` tiene la lista de SNPs cuyos alelos son C/G o A/T (2,036 SNPs).