Practical Day 4: Fine-mapping

In this practical we will use three different approaches to better characterize association signals within a genomic region, identify independent signals and perform fine-mapping. Details on how to do these have been provided in the demo.

Working directory - Day4_Finemap

LDassoc

Interactively visualize association p-value results and linkage disequilibrium patterns for a genomic region of interest (https://ldlink.nci.nih.gov/?tab=ldassoc)

1. Download the input file needed for running LDassoc online tool

    ```
    wget
    https://github.com/WCSCourses/HumanGenEpi/raw/main/manuals/Fine_m
    apping/LDLR.txt
    ```

2. Upload data - **LDLR.txt** ; SNP to test : **rs112552009**
3. Select AFR, SAS, and EAS super-populations, one at a time, and study LD between significant SNPs in the region and the lead signal
4. Export images for each super-population.
5. Which of these populations show fewer SNPs in high LD with the lead SNP?

Casual-DB

Open Causal-DB in a browser (http://www.mulinlab.org/causaldb/index.html)

Part A

1. Search for Cholesterol in Causal-DB (please search it in the correct category!)

2. Locate Prins et al. 2017 study for Total Cholesterol

3. Navigate to the first peak on chromosome 19

4. Which genes does the peak in this genomic region correspond to?

5. Check how many SNPs are there in the 95% credible set predicted by the 3 programs?

6. Check how many SNPs are there in the 99% credible set predicted by the 3 programs?

7. Sort SNPs by PAINTOR scores

8. Compare the top 6 variants to see which two SNPs have the strongest functional evidence?

Part B

1. Now go to this GWAS dataset

http://www.mulinlab.org/causaldb/block.html?d=5050&f=GD03664&b=5559

2. Go to the peak around *MAPK14* gene on chromosome 6?

3. Report the number of SNPs in the 95% credible set for the region

4. Sort SNPs by PAINTOR scores - which SNP has the best functional evidence?

GCTA-COJO

Multi-SNP based conditional and joint analysis using summary data

Run GCTA COJO on file provided (**ldlr.ma**)

```
gcta64 --bfile san.chr19 --cojo-file ldlr.ma --cojo-slct --out
san.chr19
```

How many independent associations can you see in the SAN population?