Assignment 8

Due: 21.04.2025 @ 23:59

Task 1: Colocalisation testing in R (2 points)

Learning objective: Understand how colocalisation testing can be used to assess if two different traits (e.g. gene expression and diseases risk) share a causal variant in the same genomic region.

This assignment is based on the Colocalisation tutorial. The tutorial compares genetic variants associated with CTLA4 expression in CD4+ T-cells from the OneK1K dataset (from Assignment 7) to genetic variants associated with rheumatoid arthritis (RA) from the same genomic region.

Based on the tutorial, answer the following questions:
1. Based on visual inspection of the association signals with the two traits (CTLA4 expression and rheumatoid arthritis) in the region, does it seem plausible to you that they might be driven by the same causal variant?
2. The coloc algorithm outputs five posterior probability values (PP.H0-PP.H4). What do these posterior probabilities represent? See the coloc paper for more details.
3. Which posterior probability is the highest in the tutorial example? How do you interpret that result?
3. Colocalisation testing can be sensitive to the prior probabilities p1, p2, and p12. What do these prior probabilities represent? How do the colocalisation results change if you decrease p12 from 1e-5 to 1e-6?

Task 2: Exploring genetic colocalisations in the Open Targets Platform (2 points)

Learning objective: Understand how the Open Targets Platform can be used to explore genetic colocalisation across thousands of traits.

A fine mapped credible set for rheumatoid arthritis in the FinnGen study can be found here: https://platform.opentargets.org/credible-set/6af70a4e033cd05c6eb39bb4001dae94

1. Which genetic variant has the largest posterior inclusion probability (PIP) and what is it? Is it the same variant that you previously identified for CTLA4?
2. How many variants are in the credible set?
3. Looking at the GWAS colocalization panel, what are some of the other diseases that colocalise with the RA GWAS signal?

4. Now look at the MolQTL colocalisation panel and focus on the eqtl and sceqtl (single-cell eQTL) data types. Which genes colocalise with the RA GWAS signal? In which cell types and/or tissues do these colocalisations occur? Is the OneK1K CD4+ T-cell dataset part of it?