GAW20 Data Analysis

# Introduction

This analysis applied four R scripts (described below) to filter multiple SNPs files (**snps** folder) and their corresponding annotation files (**annotation** folder) for 1105 participants (described in **COVAR.csv** metadata file).

# Data

1. **COVAR.csv:** a metadata of multiple phenotypes for 1105 participants in the GAW20 data sets
2. **snps**: a folder containing multiple genotype (SNPs) data
3. **annotation:** a folder for genotype information (annotations) for the SNPs in the **snps** folder

# R scripts

### GAW20\_Functions2022June06.R

This R script compiled the functions for filtering and classification of SNPs data.

### GAW20\_FilteringCode2022June06.R

This R script was applied for all the SNPs filtering including machine learning-based approaches (calling the **GAW20\_Functions2022June06.R** script)

### GAW20\_ML\_Classification2022June06.R

This script was applied to assess classification performance of the selected SPNs candidate sets using the Metabolic Disease Syndrome (IDF) phenotype (calling the **GAW20\_Functions2022June06.R** script).

### GAW20\_ML\_PerfSummary2022June06.R

This script was applied to summarize the outputs of the **GAW20\_ML\_Classification2022June06.**R scripts