Clinical Data Analytics Workshop (5 March 2024)

Instructor: Hyungwon Choi

In this workshop, we work with the data published in *Tan et al., Variability of the Plasma Lipidome and Subclinical Coronary Atherosclerosis, Atheroscler Thromb Vasc Biol, 2021* DOI: 10.1161/atvbaha.121.316847. The paper describes a lipidomics study of ~80 individuals at a high risk of coronary artery disease (CAD). The participants were invited monthly for blood sampling up to five times and had coronary artery plaque burden assessed using computed tomography coronary angiography (CTCA) at the end of the follow-up.

Using high-quality quantitative lipidomics data and other meta data of the participants, we will first inspect overall data trends via visualization and dimension reduction. Using custom R scripts, we synchronize the quantitative lipidomic data and the plaque burden data (outcome) for the same subjects and practice clustering of the subjects by different plaque subtypes. We fit linear mixed effects models on repeated measure data (lipids) and compute population-level properties of lipid species such as within-individual and between-individual variability (coefficients of variation). Finally, we identify lipid species in which the visit-to-visit variability is significantly associated with different plaque types.

Please ensure the latest versions of R and RStudio are installed on your computer. We will install R packages **before the start of the workshop** with assistance by teaching assistants.

- R (version 4.3.2. Download from https://cloud.r-project.org/.
- RStudio (newest version, at least 2022.02). Download from https://posit.co/download/rstudio-desktop/.

R markdown document and the data sets will be distributed in the beginning of the workshop.