**Clinical Data Analytics Workshop (5 March 2024)**

Instructor: Hyungwon Choi

In this workshop, we use 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](https://doi.org/10.1161/atvbaha.121.316847) as an example dataset for this workshop. The data set comes from a lipidomics study of individuals at a high risk of cardiovascular diseases, where the participants were invited monthly for blood sampling up to five times and their coronary artery plaque burden was assessed using computed tomography coronary angiography (CTCA) at the end of the follow-up. Using high quality quantitative lipidomics data and other meta data for subjects, we will learn to inspect overall trends via data visualization and dimension reduction. Using custom R code, we practice synchronizing the quantitative lipidomic data and their plaque burden data (outcome) and cluster subjects by different plaque types. Using the repeated measure data and linear mixed effects model, we compute population-level properties of lipid species such as within-individual and between-individual variability (coefficients of variation). Finally, we search for lipid species whose visit-to-visit variability is associated with different plaque types.

**Preparing for the workshop**

Please ensure you install R, RStudio and used R packages **before the start of the workshop**, as detailed below.

**R and RStudio**

* R (version 4.3.2. Download from <https://cloud.r-project.org/>. Check your R version by running following command in your console: R.Version()$version.string
* RStudio (newest version, at least 2022.02). Download from <https://posit.co/download/rstudio-desktop/>. Check your RStudio version by either looking clicking *About RStudio* under the menu *Help.*

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