THLBB2021_PUB005: Associations of healthy food choices with gut microbiota profiles -datasets

Cohort: FINRISK 2002
Data set accession number in THL Biobank: THLBB2021_PUB005

Publication information

Associations of healthy food choices with gut microbiota profiles
Kari K Koponen, Aaro Salosensaari, Matti O Ruuskanen, Aki S Havulinna, Satu Männistö, Pekka Jousilahti, Joonatan Palmu, Rodolfo Salido, Karenina Sanders, Caitriona Brennan, Gregory C Humphrey, Jon G Sanders, Guillaume Meric, Susan Cheng, Michael Inouye, Mohit Jain, Teemu J Niiranen, Liisa M Valsta, Rob Knight, Veikko V Salomaa

Correspondence: kari.koponen@thl.fi, veikko.salomaa@thl.fi

Published article's journal and reference information

Dataset availability information

Data described in the manuscript, code book, and analytic code are available from Finnish Institute for Health and Welfare Biobank based on a written application as instructed on the website of the Biobank (https://thl.fi/en/web/thl‑biobank). The data are not publicly available because they contain information that could compromise research participant privacy/consent. The source code for the analyses is openly available at 10.5281/zenodo.4671405.

Participant and method description

https://academic.oup.com/ajcn/article/114/2/605/6280094#282958852

Description of datasets

THLBB2021_PUB005 datasets contain all the data needed to replicate the analyses of the article "Associations of healthy food choices with gut microbiota profiles".

DOI to the latest version of the source code: 10.5281/zenodo.4671405

The following source files from the FINRISK 2002 cohort can be used to replicate the analyses contained in the paper with the code available at the DOI mentioned above and in the article.

These files need to be placed in the folder "data/" in the source code folder after which each analysis .Rmd file can be run independently. In addition to these files, some of the analysis .Rmds create several intermediary folders, files, tables and figures in the working directory. Some analyses may take a long time (up to hours) to complete depending on the computing power available.

Phyloseq files for the genus and species levels that both include the necessary phenotype data:
1. foodchmicrobpr_genus_060421.rds

2. foodchmicrobpr_species_060421.rds
- Phenotype data frames of the above phyloseq files include the following variables: "FID", "SEX", "AGE", "BMI", "SMOKING", "MEDICATION", "FOOD_RECOMMENDED_CHOICES_deciles", "FOOD_RECOMMENDED_CHOICES", "FIBER_TOTAL", "RED_MEAT", "BERRIES", "BREAD", "VEGETABLES", "FRUIT", "JUICES", "FISH", "POULTRY", "LOW_FAT_CHEESE", "DRESSINGS_OILS", "NUTS_SEEDS", "Stool.Barcode.name", "Stool.Barcode.ID".

Measurements of KEGG orthology metabolites:
3. foodchmicrobpr_kegg_060421.rds