THLBB2021_PUB003: Taxonomic signatures of cause-specific mortality risk in human gut microbiome -datasets

Cohort: FINRISK 2002
Datasets' accession number in THL Biobank: THLBB2021_PUB003

Publication information

Taxonomic signatures of cause-specific mortality risk in human gut microbiome
Aaro Salosensaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfthan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti & Teemu Niiranen

Correspondence: teemu.niiranen@utu.fi, leo.lahti@utu.fi

Published article's journal and reference information

Nat Commun 12, 2671 (2021)
DOI: https://doi.org/10.1038/s41467-021-22962-y

Dataset availability information

"The metagenomic data are available from the European Genome-Phenome Archive (accession number EGAD00001007035).

"The phenotype data contain sensitive information from healthcare registers and they are available through the THL biobank upon submission of a research plan and signing a data transfer agreement (https://thl.fi/en/web/thl-biobank/for-researchers/application-process)."

Please note that some of the phenotype variables were derived from national health registers and their use requires a separate data permit from Findata (https://findata.fi/en/data-permits/).

Participant and method description

https://www.nature.com/articles/s41467-021-22962-y#Sec10

Source code

DOI for the source code: 10.5281/zenodo.4306060

The source files listed in the next section from the FINRISK 2002 cohort can be used to replicate the analyses contained in the paper with the code included in the Supplementary Information of the article (https://www.nature.com/articles/s41467-021-22962-y#Sec27).

These files need to be placed in the folder input/ of the source code after which the pipeline can be run by sourcing code/main.R

In addition to these files, the analysis code creates several intermediary files, tables and figures which are saved in the folder output/.

Some computations can take a long time (several hours) to complete the first time the analysis is run, but the intermediary files created can be used subsequently to run the code much faster.

Description of datasets

The THLBB2021_PUB003 datasets include all the data needed to replicate the experiments of the article "Taxonomic Signatures of Cause-Specific Mortality Risk in Human Gut Microbiome."

Phenotype data file:
1. meta_data.txt
The meta data table includes the following variables for 6834 FINRISK 2002 participants: "Barcode", "BL_AGE", "BMI", "MEN", "CURR_SMOKE", "PREVAL_DIAB", "BL_USE_RX_L", "SYSTM", "BP_TREAT", "EAST", "DEATH", "DEATH_AGEDIFF", and "K_TPKS".

Description of meta data variables:
2. covariate_names.txt
- Description of the phenotype variables in the meta data.

NCBI referenced microbe counts aggregated at species level:
3. species_reads.txt

Table of taxonomic relations:
4. taxonomic_table.txt

Description of functions of the metabolites in the KEGG Orthology:
5. ko_terms.txt

Measurements of KEGG orthology metabolites:
6. ko.txt

Metabolite pathway measurements:
7. pathways.txt

Metabolic module measurements:
8. modules.txt