nature research

Corresponding author(s): Eran Segal

☐ Initial submission ☒ Revised version ☐ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

## Experimental design

1. Sample size

Describe how sample size was determined.

As this is an observational study based on voluntary participation, we used all the samples that were collected throughout the study collection phase. We verified that our sample was well powered to answer our research questions using extensive power analyses, as described in detail in the Supplementary Information.

2. Data exclusions

Describe any data exclusions.

Exclusion criteria consisted of: (i) pregnancy; (ii) usage of antibiotics within three months prior to participation; (iii) chronically active inflammatory or neoplastic disease in the three years prior to enrollment; (iv) chronic gastrointestinal disorder, including Inflammatory Bowel Disease and Celiac disease; (v) skin disease, including contact dermatitis, precluding proper attachment of the continuous glucose monitor; (vi) active neuropsychiatric disorder; (vii) myocardial infarction or cerebrovascular accident in the 6 months prior to participation; (viii) chronic immunosuppressive medication usage; (ix) pre-diagnosed type I or type II diabetes mellitus.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experimental replication was performed by repeating all analyses on an independent cohort, namely the LifeLines-DEEP cohort.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

As this is an observational study, randomization is not relevant and was not employed. We controlled for potential sources of confounding by (a) excluding participants who are at least third degree relatives with another participants (unless stated otherwise for specific analyses); (b) excluding participants self-reported to share a household with another participant (unless stated otherwise for specific analyses); (c) including covariates encoding gender, age, stool collection method, self-reported daily median caloric, fat, protein and carbohydrates consumption, and the top five principal component of the host genotypes (unless stated otherwise for specific analyses); (d) when testing for associations between specific SNPs and specific taxa, we used a linear mixed model, which controls for confounding due to population structure or subtle genetic relatedness.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

As this is an observational study, blinding is not relevant and was not employed.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☒ | ☐ Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Trimmomatic (v.0.32)
Bowtie2 (v.2.1.0)
MetaPhlan 2.2
RL-SKAT 1.0
FIESTA 1.0
QIIME 1.9.1
Pathoscope 2.0
R GENESIS package 2.4.0
Python package FaST-LMM 0.1
GCTA 1.26.0
plink 1.90b3.44
KING-robust 2.0
R vegan package 2.4-2
Python package scikit-learn 0.18.2
Python package scipy 0.19.1
R DBF.test function (https://wwwf.imperial.ac.uk/~gmontana/software/dbf/dbf_test.R)
Impute 2 2.3.2
Eagle 2  2.3.2
Python package xgboost 0.6

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

| | |
|---|---|
| a. State the source of each eukaryotic cell line used. | No eukrayotic cell lines were used |
| b. Describe the method of cell line authentication used. | No eukrayotic cell lines were used |
| c. Report whether the cell lines were tested for mycoplasma contamination. | No eukrayotic cell lines were used |
| d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | No eukrayotic cell lines were used |

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

| | |
|---|---|
| Provide details on animals and/or animal-derived materials used in the study. | No animals were used |

Policy information about studies involving human research participants

12. Description of human research participants

| | |
|---|---|
| Describe the covariate-relevant population characteristics of the human research participants. | Study participants were healthy individuals aged 18–70 able to provide informed consent and operate a glucometer. Anthropometric and blood pressure measurements were taken by a CRA or a certified nurse, as well as a blood test and weights. |